

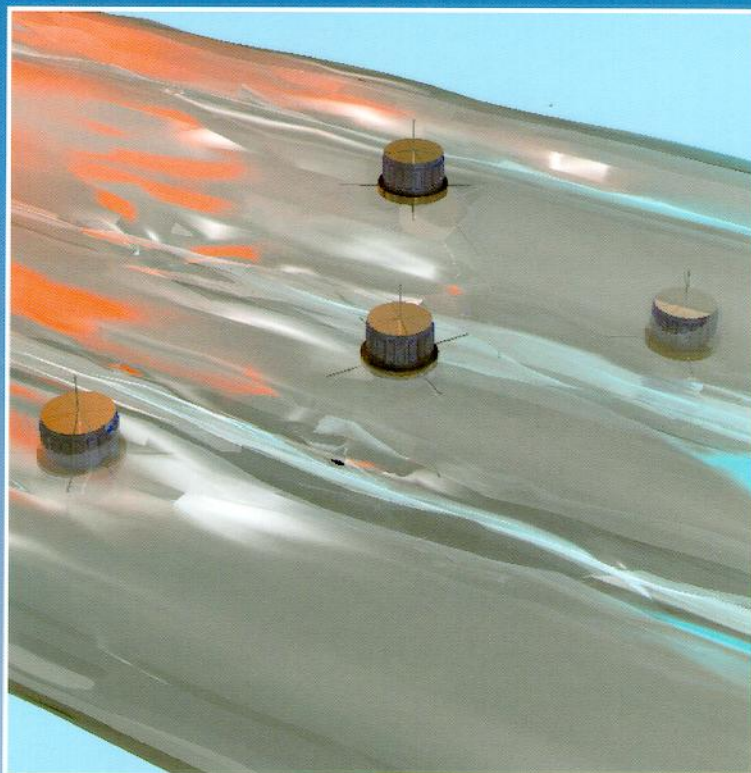


ISSI Scientific Report

SR-001

Analysis Methods for Multi-Spacecraft Data

Götz Paschmann and Patrick W. Daly (Eds.)



INTERNATIONAL
SPACE
SCIENCE
INSTITUTE



SR-001

Paper Edition, July 1998

Electronic Edition, July 2000

Analysis Methods for Multi-Spacecraft Data

Editors

Götz Paschmann

Max-Planck-Institut für extraterrestrische Physik
Garching, Germany

Patrick W. Daly

Max-Planck-Institut für Aeronomie
Katlenburg-Lindau, Germany



Cover: Four satellites traversing an unstable current sheet
(Büchner et al., page [472](#))

Electronic Edition 1.1

Published for: The International Space Science
Institute
Hallerstrasse 6, CH-3012 Bern,
Switzerland
www.issi.unibe.ch

by: ESA Publications Division
Keplerlaan 1, 2200 AG Noordwijk,
The Netherlands

Publication Manager: Bruce Battrick
Cover Design: Carel Haakman

Paper Edition Copyright: ©1998 ISSI/ESA

Electronic Edition Copyright: ©2000 ISSI

Contents

Foreword by the Directors of ISSI	xiii
Preface	xv
Introduction	1
Notation Conventions	4
1 Spectral Analysis	5
1.1 Introduction	5
1.2 Basic Concepts	6
1.2.1 Fourier Series	6
1.2.2 Parseval's Relation and Power Spectral Density (PSD)	6
1.2.3 Phase	7
1.2.4 Discrete Fourier Transform (DFT)	8
1.2.5 Normalisation	9
1.2.6 Time-Frequency Analysis	10
1.3 Fourier Techniques	15
1.3.1 Fast Fourier Transform (FFT)	15
1.3.2 Detrending	15
1.3.3 Windowing	16
1.3.4 Averaging and Stationarity	19
1.3.5 Overlapping Intervals	21
1.3.6 Zero-Padding	22
1.3.7 Welch Method for Time-Dependent PSD Estimation	24
1.4 Wavelet Techniques	25
1.4.1 Morlet Wavelets	25
1.4.2 Wavelets and Fourier Methods	27
1.4.3 Continuous Wavelet Transform (CWT)	28
1.4.4 Comparing WFT and MWT: an Example	29
1.4.5 Implementations of the Wavelet Transform	31
1.4.6 Wavelet Packet Analysis	32
1.5 Spectral Analysis of Multiple Signals	32
1.5.1 Cross-Spectral Analysis	32
1.5.2 Averaging and Coherence	33
1.5.3 Cross-Spectral Analysis: an Example	34
1.5.4 Cross-Wavelet Analysis	36
1.6 Parametric Methods	37
1.7 Final Comments	39

2	Time Series Resampling	43
2.1	Introduction	43
2.2	Definitions	44
2.3	Time-Domain Operations	46
2.4	Algorithms	47
2.5	Aliasing, Filtering and the Nyquist Frequency	48
2.5.1	The Sampling Factor	50
2.5.2	Analogue to Digital Conversion	50
2.5.3	Resampling and the Reduction Factor	51
2.5.4	Filtering and the Filter Factor	51
2.6	The Art of Joining Data Sets	52
2.6.1	Fuzzy Join	52
2.6.2	Averaging	53
2.6.3	Synchronisation	55
2.6.4	Interpolation	56
2.6.5	Determination of the Sampling Factor	56
2.6.6	Data Gaps	57
2.6.7	Irregularly Spaced Data	58
2.7	Metadata and Status Data	58
2.8	Discussion	59
2.8.1	Scientific Use of the Sampling Factor	59
2.A	The Bartlett Window	60
3	Filtering: General Framework	65
3.1	Introduction	65
3.2	Multi-Spacecraft Filtering Technique	65
3.2.1	Spectral Representation of the Wave Field	65
3.2.2	Notations and Data Preparation	66
3.2.3	Filter Bank Approach and $P(\omega, k)$ Estimation	67
3.3	Limitations Related to the Field	68
3.3.1	Time Stationarity and Space Homogeneity	69
3.3.2	Spatial Aliasing	69
3.4	Limitations Related to the Experimental Constraints	70
3.4.1	The Spacecraft Configuration	70
3.4.2	Inaccuracy in the Time Synchronisation	70
3.4.3	Inaccuracy in the Inter-Spacecraft Distances	71
3.5	Examples	71
3.5.1	$P(\omega, k)$ Estimator Related to the Electromagnetic Wave Field	71
3.5.2	$P(\omega, k)$ Estimator Related to the Magnetic Field	74
3.5.3	Electromagnetic Surface Wave Detector (SWD)	75
3.6	Summary	77
4	Filtering: Plasma Mode Recognition	79
4.1	Introduction	79
4.2	Mode Decomposition of MHD Wave Fields	79
4.3	Test of the Decomposition Method	82
4.4	The k Filtering and Generalised MVA	83

4.5	Model Validation	86
4.6	Conclusions	87
5	Measurement of Plasma Velocity Distributions	91
5.1	Introduction	91
5.2	General Principles of 3-D Velocity Distribution Measurements	92
5.2.1	Velocity Space Filter	92
5.2.2	Mass per Charge Analysis	105
5.2.3	The Detector and Counter	107
5.3	Calibration	116
5.3.1	Importance of Ground and In-Flight Calibrations	116
5.3.2	Special Considerations for Multi-Spacecraft Data Analysis	116
5.4	False Counts and Statistical Fluctuations	117
5.5	Distribution Data Products	118
5.5.1	Pitch Angle Distributions	119
5.5.2	Full Three Dimensional Distributions	120
5.5.3	Reduced Three Dimensional Distributions	121
5.5.4	Four Dimensional Distributions	121
5.5.5	Data Compression	122
5.6	Summary	122
6	Moments of Plasma Velocity Distributions	125
6.1	Introduction	125
6.2	Definitions	125
6.2.1	Single-Species Moments	125
6.2.2	One-Fluid Moments	126
6.3	Computations	127
6.4	Accuracy	128
6.4.1	Error Sources	130
6.4.2	Simulation Results	137
6.4.3	Overall Error Assessment	144
6.5	Interpretation	146
6.5.1	Single-Species Moments	146
6.5.2	One-Fluid Moments	150
6.6	Distribution Functions, Moments, Partial Moments	151
6.7	Spherical Harmonic Analysis	152
6.8	Applications	152
6.8.1	Identification of Regions and Boundaries	152
6.8.2	Conservation Equations	153
6.9	Summary	157
7	Analysis of Plasma Kinetics	159
7.1	Introduction	159
7.2	Liouville's Theorem	160
7.3	Liouville Mapping: Known Fields	161
7.4	Liouville Mapping: Unknown Fields	163
7.5	Remote Sensing of Boundaries and Non-Local Processes	164

7.6	Velocity Dispersion or “Velocity Filter”	164
7.7	Particle Anisotropies and Remote Sensing	165
7.7.1	The Gyro-Orbit	165
7.7.2	Particle Anisotropies	167
7.7.3	When is an Anisotropy not an Anisotropy?	171
7.7.4	Remote Sensing of Boundaries	172
7.7.5	Absorbing or Source Boundary?	175
7.8	Example Applications	175
7.8.1	1-D Mapping of Electrons at the Earth’s Bow Shock	175
7.8.2	2-D Electron Mapping in the Earth’s Foreshock	176
7.8.3	Remote Sensing of the Earth’s Bow Shock	177
7.8.4	Ions in the Cusp	178
7.8.5	Remote Sensing of a Flux Transfer Event	178
8	Minimum and Maximum Variance Analysis	185
8.1	Introduction	185
8.2	Theory	185
8.2.1	Elementary Considerations	185
8.2.2	Derivation of Minimum Variance Analysis on Magnetic Field	187
8.2.3	Hodogram Representation	189
8.2.4	Variance Ellipsoid	191
8.2.5	Degeneracy	192
8.2.6	Constraint $\langle B_3 \rangle = 0$	193
8.3	Error Estimates	196
8.3.1	Analytical Estimates of Statistical Errors	196
8.3.2	Bootstrap Error Estimates	198
8.3.3	Other Error Estimates	201
8.3.4	Stationarity	202
8.3.5	Data Filtering and Optimal Analysis	204
8.4	Other Applications	206
8.4.1	Minimum Variance Analysis on Mass Flux	206
8.4.2	Maximum Variance Analysis on Electric Field	209
8.4.3	Application to 2-D Structures	212
8.5	Discussion of AMPTE/IRM Event	212
8.6	Summary	215
9	DeHoffmann-Teller Analysis	221
9.1	Introduction	221
9.2	Existence and Properties of HT Frame	222
9.3	Determination of V_{HT} from Experimental Data	226
9.3.1	Minimisation of Residual Electric Field	226
9.3.2	Minimisation of Residual Cross-Field Velocity	228
9.3.3	Walén Relation in HT Frame	229
9.4	Uncertainty in V_{HT} Determination	230
9.4.1	Analytical Error Estimates	230
9.4.2	Bootstrap Error Estimates	233
9.4.3	Other Uncertainty Estimates	234

9.4.4	Stationarity	236
9.5	Accelerating HT Frame	240
9.6	Discussion	241
9.7	Summary	244
	Appendix	246
10	Shock and Discontinuity Parameters	249
10.1	Introduction	249
10.2	The Shock Problem: Rankine-Hugoniot Relations	249
10.3	Shock Parameters	250
10.3.1	Shock Geometry	250
10.3.2	Mach Numbers	253
10.3.3	Important Ratios	253
10.4	Determination of Shock and Discontinuity Normals	253
10.4.1	Variance Analyses	254
10.4.2	Coplanarity and Related Single Spacecraft Methods	254
10.4.3	Multi-Spacecraft Timings	256
10.4.4	Combined Approaches	257
10.4.5	Shock Jump Conditions	258
10.4.6	Model Boundary Equations	259
10.4.7	Tangential Discontinuities	263
10.4.8	Rotational Discontinuities	263
10.5	Determination of the Shock/Discontinuity Speed	263
10.5.1	Mass Flux Algorithm	264
10.5.2	Shock Foot Thickness Algorithm	264
10.5.3	Multi-Spacecraft Timing Algorithm	265
10.5.4	Smith and Burton Algorithm	265
10.5.5	Generalised deHoffmann-Teller Transformation	265
10.5.6	Velocity of a Tangential Discontinuity	266
10.6	Determining θ_{Bnu}	266
10.6.1	Application of the Shock Normal	266
10.6.2	Ensemble θ_{Bnu}	267
10.7	Application	267
11	Discontinuity Analysis: Orientation and Motion	271
11.1	Introduction	271
11.2	Methodology	273
11.2.1	General assumptions	273
11.2.2	Method outline	274
11.3	Pre-Selection Issues	276
11.3.1	Assumptions	276
11.3.2	Boundary Normal Analysis	276
11.3.3	Planarity	278
11.3.4	Stationarity	279
11.4	Discontinuity Analysis: Motional Properties	279
11.4.1	Planar-DA Technique	279
11.4.2	Basic Application	281

11.4.3	Noise: Error Combinations	283
11.4.4	Noise: Data Examples	283
11.4.5	Identification Of Non-Planar Structures	285
11.5	Discontinuity Analysis: Combined Motional and Curvature Analysis	287
11.5.1	Method Development	287
11.5.2	Non-Planar Analysis: Constant Motion	289
11.5.3	Non-Planar Analysis: Information Balance to Lowest Order	291
11.5.4	Consequences for Curvature Analysis	295
11.5.5	Data Analysis: Natural Noise	297
11.5.6	Self-Consistent Curvature and Normal Analysis	298
11.6	Conclusions	300
11.A	Model Examples	302
11.A.1	Planar Model	302
11.A.2	Non-Planar Models	302
12	Spatial Gradients and the Volumetric Tensor	307
12.1	Introduction	307
12.1.1	Timing	308
12.1.2	A Simple Boundary Crossing	308
12.2	Orientation and Motion of a Plane Discontinuity	310
12.2.1	A Symmetrical Treatment of the Simplest Case	310
12.2.2	When Timing is only Relative	311
12.3	Spatial Gradients	313
12.3.1	The Gradient of a Vector	313
12.3.2	The Gradient of a Solenoidal Vector	314
12.4	The Volumetric Tensor \mathbf{R}	315
12.4.1	A Simple Geometric Interpretation	315
12.4.2	Relation to the Inertia Tensor	316
12.4.3	The Special Case of a Tetrahedron	317
12.5	Discussion	319
12.A	Spacecraft Non-Coplanar: Alternative Proof for the Tetrahedron Volume	320
12.B	Spacecraft Coplanar	320
12.C	Spacecraft Colinear	322
13	Tetrahedron Geometric Factors	323
13.1	Introduction	323
13.2	Measurement Performance	324
13.3	The Shape of the Tetrahedron	325
13.3.1	The 1-Dimensional Geometric Factors	325
13.3.2	A Geometric Representation of the Size, Shape, and Orientation of a Polyhedron	327
13.3.3	Size, Elongation, and Planarity of a Polyhedron	328
13.4	Study of the 1-D Geometric Factors over the Tetrahedron Reservoir	330
13.4.1	The Five Types of Tetrahedra	330
13.4.2	Computation of a Reservoir of Five Types of Tetrahedra	330
13.4.3	The 1-D Geometric Factors and the Types of Tetrahedra	332
13.5	Study of the 1-D Geometric Factors with the E - P Parameters	335

13.5.1	Computing an Homogeneous Tetrahedra Reservoir	335
13.5.2	Cluster Orbit Tetrahedron in a Time Diagram	335
13.5.3	Cluster Orbit Tetrahedron in the E - P Diagram	338
13.5.4	E - P Diagram for 1-D Geometric Factors	338
13.6	Conclusions	343
13.A	Calculation of Geometric Factors Q_{GM} and Q_{RR}	344
14	Spatial Interpolation: Theory	349
14.1	Introduction	349
14.2	Estimation of Gradients	350
14.2.1	Linear Interpolation and Related Estimators	350
14.2.2	Quadratic Interpolation and Related Estimators	353
14.3	Physical and Geometrical Errors	355
14.3.1	Statistical Properties of the Reciprocal Vectors	355
14.3.2	Statistical Properties of the Linear Estimator \mathbf{LG}	357
14.3.3	Comparison of the Errors for a Regular Tetrahedron	358
14.3.4	Effect of the Orbital Motion on \mathbf{LG}	359
14.4	Truncation Errors for a One-dimensional Model	360
14.4.1	Truncation Errors for $\mathbf{LG}[\mathbf{B}]$	361
14.4.2	Truncation Errors for $\mathbf{QG}[\mathbf{B}]$	362
14.4.3	Truncation Errors for $\mathbf{L}_2\mathbf{G}[\mathbf{B}]$	363
14.5	Other Applications	364
14.5.1	Derivation of the Spatial Aliasing Condition	364
14.5.2	Characterisation of a Planar Discontinuity	365
14.6	Conclusions	368
15	Spatial Interpolation: Application	371
15.1	Introduction	371
15.1.1	Statement of the Problem	372
15.2	Relationship Between HLS and Barycentric Methods	372
15.2.1	Relationship Between the Reciprocal and Volumetric Tensors	372
15.2.2	Identity of the HLS and Linear Barycentric Estimators of the Gradient	373
15.3	Truncation Errors of \mathbf{LG} for a Magnetic Dipole	373
15.3.1	The Dipole Field and its Gradient	374
15.3.2	Linear Estimation of the Gradient of the Dipole Field	374
15.3.3	Truncation Errors for a Regular Tetrahedron	375
15.4	Comparison of the Methods Using Simulated Data	375
15.4.1	Evolution of the Tetrahedron along the Orbit	376
15.4.2	Estimations of $\mathbf{G}[\mathbf{B}]$	376
15.4.3	Quality of the Estimations	379
15.5	Future Developments	389
15.5.1	Filtering of the Data	389
15.5.2	The Spacecraft Position	391
15.6	Conclusions	391

16 Accuracy of Current Density Determination	395
16.1 Introduction	395
16.2 The Curlometer Technique	396
16.2.1 Background	396
16.2.2 Application: Cluster Context	398
16.2.3 Analysis Technique	403
16.3 Accuracy of Current Density Determination	404
16.3.1 Parameters Used to Define the Shape of a Tetrahedron	404
16.3.2 Simulation Method	405
16.3.3 The Tetrahedron Reservoir	405
16.3.4 The Current Structure Models	405
16.3.5 The Computation of \mathbf{J} and $\nabla \cdot \mathbf{B}$	407
16.4 Results	407
16.4.1 Influence of Tetrahedron Shape on $\Delta J/J$ Accuracy	407
16.4.2 Influence of Tetrahedron Shape on $\nabla \cdot \mathbf{B}/ \nabla \times \mathbf{B} $ Estimate	408
16.4.3 Relationship between $\Delta J/J$ and $\nabla \cdot \mathbf{B}/ \nabla \times \mathbf{B} $	408
16.4.4 Influence of Current Direction on Error $\Delta J/J$	410
16.4.5 Heterogeneous Current Profile	411
16.5 Conclusions	412
17 Accuracy of Plasma Moment Derivatives	419
17.1 Error Estimates from a Simplified Model	420
17.1.1 Partial Derivative Estimation Accuracy	420
17.1.2 Discussion of Competing Error Terms	421
17.1.3 Directional Derivative Estimation Accuracy	422
17.1.4 Time Derivative Estimation Accuracy	423
17.1.5 Divergence Estimation Accuracy	424
17.1.6 Curl Estimation Accuracy: The General Case	425
17.1.7 Curl Estimation Accuracy: $\Delta x_1 = \Delta x_2$	426
17.2 Arbitrary Satellite Configurations	430
17.2.1 Derivatives from Barycentric Coordinates	430
17.2.2 Error Estimation	432
17.3 Application to the Earth's Magnetosphere	434
17.A Details of Error Estimation: Simplified Geometry	438
17.A.1 Error Estimates for Partial Derivatives	438
17.A.2 Curl Estimation Accuracy: The General Case	442
17.A.3 Curl Estimation Accuracy: $\Delta x_1 = \Delta x_2$	443
17.B Details of Error Estimation: General Geometry	445
17.B.1 Directional Derivatives and Partial Derivatives	445
17.B.2 Divergence Estimation Accuracy	446
17.B.3 Curl Estimation Accuracy	446
18 Numerical Modelling and Simulation	449
18.1 Introduction	449
18.2 Test Particle Calculations	451
18.2.1 Non-adiabatic Particle Motion	452
18.2.2 Mass-dependent Energetic Particle Spectra	453

18.3	Magnetohydrodynamic Modelling	457
18.3.1	Identification of Wave Modes	458
18.3.2	Magnetopause Entry Mechanisms	459
18.3.3	Magnetotail Energy Release	462
18.4	Kinetic Plasma Simulations	468
18.4.1	Fokker-Planck Simulations	469
18.4.2	Guiding Centre Simulations	469
18.4.3	Hybrid Simulations	469
18.4.4	Fully Self-consistent Kinetic Plasma Simulations	470
	Notes	479
	Authors' Addresses	485
	Index	487

Foreword by the Directors of ISSI

This volume is the first in the series “ISSI Scientific Reports”, which is one of the two series of publications of ISSI. Whereas the “Space Sciences Series of ISSI” contains the results of the study projects and workshops that form a major part of ISSI’s scientific programme, the ISSI Scientific Report series is intended primarily for important results concerning methodology or techniques. The contents of the present volume are a very good example of what we hope will become a main line for ISSI Scientific Reports: a collection of advanced methods for analysing experimental data, many of them quite new, which will hopefully be very useful for the relevant parts of the scientific community, from graduate students to experienced solar system researchers.

When ISSI started its scientific programme, less than three years ago, the methodological line represented by this volume was not in our minds, but it was brought there by Götz Paschmann, Alain Roux and others, who told us that there existed no forum for works of this kind and that ISSI’s support would be important for the possibilities of the interested scientists to carry the work through. After consultation with ESA representatives, we decided to offer ISSI’s support to a working group with the task to produce a handbook on methods of analysing data from clusters of spacecraft. We are very pleased with the outcome of the project and we thank Götz Paschmann (chair), Malcolm Dunlop (deputy chair) and their colleagues in the working group for the extensive efforts that they have contributed to the production of this interesting and useful ISSI Scientific Report No. 1.

Johannes Geiss
Bengt Hultqvist
Bern, Switzerland
April 1998

Preface

Preface to the Paper Edition

The preparation of the Cluster mission included efforts and initiatives to develop methods and tools for the analysis of multipoint measurements, as documented in several workshop proceedings. But it was apparent that there was a need to considerably expand and improve on previous work, and then collect it in a Handbook that would present the analysis techniques in an elementary fashion and would include a careful assessment of accuracies and limitations, preferably demonstrated by applications to real or simulated data.

In this situation, ISSI kindly agreed to support a working group on “Advanced Analysis Methods for Data from Clusters of Spacecraft”. The group comprised 15 members, all of them associated with the Cluster mission. In alphabetical order, the members were J. Büchner, G. Chanteur, P. W. Daly, M. W. Dunlop (Deputy Chair), A. I. Eriksson, A. N. Fazakerley, K.-H. Glassmeier, M. Goldstein, C. C. Harvey, U. Motschmann, G. Paschmann (Chair), A. Roux, R. Schmidt (Cluster Project Scientist), S. J. Schwartz, B. U. Ö. Sonnerup, and J. Vogt.

The working group met at ISSI for several days each in March 1996, June 1996 and January 1997. The second meeting took place two days after the loss of the Cluster mission that resulted from the failure of the Ariane 501 launcher. Even though this event put into jeopardy the chances to actually obtain the appropriate multipoint measurements, the working group went ahead, particularly since it had decided earlier to include the description of advanced single-point analysis methods that had not been properly documented before. The April 1997 decision for a Cluster II recovery mission, to be launched in the year 2000, has provided new motivation for the work on analysis techniques.

At the first meeting, the group defined its program and tentatively agreed on a set of chapters and their authors. Authorship was not restricted to members of the working group. It was realised right at the beginning that substantial efforts were needed to test some of the methods and/or establish their validity and accuracy. This work was carried out between the meetings. Progress was reviewed, and the program finalised, at the second meeting. At the third meeting, drafts of the chapters were discussed and the remaining work and schedule defined.

All chapters in this book were written in \LaTeX and all figures, over 150, were provided as encapsulated PostScript files. This permitted not only a uniform appearance throughout the whole book, but also enabled extensive cross-referencing between the chapters and a comprehensive index of keywords.

In July 1997, when all manuscripts had been received and properly formatted by the production editor, P. W. Daly, an editorial committee, consisting of G. Chanteur, P. W. Daly, M. W. Dunlop, C. C. Harvey, G. Paschmann, and S. J. Schwartz, reviewed the chapters, and asked the authors for revisions. All chapters were then sent to outside referees. After another round of revisions, the book was finally submitted electronically to the ESA Publication Branch at ESTEC in April 1998.

The working group members and authors invested a very substantial amount of time and work into this project. We would like to thank them for this effort, and for the cooperative spirit in which the goal was achieved. We thank the editorial committee for careful

reading of the manuscripts, and the external referees, A. Balogh, G. Berge, D. Burgess, S. W. H. Cowley, I. Dandouras, R. C. Elphic, M. P. Gough, J. M. Quinn, N. Sckopke, and E. Whipple, for their helpful comments and criticisms. P. W. Daly did an excellent job as production editor and B. Batrick, of ESA's Publication Branch, guaranteed a speedy and professional production and delivery. Finally, the entire working group would like to thank the ISSI directors, J. Geiss and B. Hultqvist, and their staff for their generous support and warm hospitality that made this project possible and enjoyable.

Götz Paschmann
Garching, Germany
Malcolm W. Dunlop
London, United Kingdom

Preface to the Electronic Edition

Soon after the appearance of the first (paper) edition of this book, P. W. Daly began to experiment with an electronic version in PDF with navigation aids and internal links for cross-references. Coloured figures which were printed in black and white in the book could finally be properly reproduced.

The editors and authors then decided that an electronic edition would also provide an opportunity for correcting errors, making additional comments, and/or otherwise extending the original material based on new work.

To accomplish this, we decided to make use of *endnotes*, numbered by chapter and sequence numbers, and collected together in a separate section at the end of the book (page 479). At the corresponding location in the text itself, a marginal marker appears with the note number, this marker being a dynamic link to note text. ⇒0.0

Changes to the text have been made in the following ways:

1. Typographical or similar mistakes that do not change the content or meaning, have been simply fixed without comment.
2. Errors in equations and data have been corrected in the text, with an endnote and marker to point this out.
3. Comments by the authors have been added as endnotes with a marker at the appropriate location in the text.
4. Extensive new material, from other publications, have been made available as separate PDF files, attached at the end of the book. Dynamic links are provided to direct the reader immediately to these pages.

In this way, we can maintain a 100% correspondence between the pages of the paper and electronic editions. We consider this to be desirable since many readers will see the latter as a complement to the former and will want to use the two in parallel.

Götz Paschmann
Bern, Switzerland
Patrick W. Daly
Katlenburg-Lindau, Germany

Introduction

The Earth's magnetosphere is the bubble in which we live, buffeted by the solar output $\Rightarrow 0.1$ in the form of a collisionless, supersonic solar wind. Large and small variations in the solar wind give rise to dynamic, often dramatic responses of the magnetosphere due to, for example, magnetic reconnection at the dayside magnetopause (which locally ruptures the bubble) and reconnection in the geomagnetic tail (which is responsible for major magnetic storms, aurora, and related phenomena). These transient processes, together with the motion and internal structure of thin boundaries (bow shock, magnetopause, plasma sheet current layer), embody fundamental physical mechanisms representative of many astrophysical applications and are the mediators of "space weather". Their variability implies that they involve three-dimensional, time-dependent structures on a range of scales.

Single spacecraft measure the time series of physical parameters as seen at the spacecraft position. As such, these measurements cannot unambiguously separate spatial and temporal variability. The study of the structure and evolution of key physical processes therefore requires a multi-point approach. Two-spacecraft missions, such as ISEE-1 and -2 or AMPTE-IRM and -UKS, were an improvement, but still allowed only one-dimensional (spatial) resolution. In many cases this dimension did not correspond to the most important dimension of the phenomenon. Use of dual spacecraft, therefore, has been restricted largely to studies of special events. Four spacecraft, with adjustable separations, such as to be implemented for the Cluster mission, are the minimum needed to uniquely address the issue of resolving a three-dimensional structure, at least to lowest order in the gradients.

The term "adjustable", of course, hides to some extent the problem that, even with multiple spacecraft, the spatial sampling must be well matched to the spatial scales of importance. Suitable separation distances for such co-orbiting clusters of spacecraft are in the range between hundred km and tens of thousands of km, depending on the event sampled. These scales are very much smaller than the typical separations of spacecraft within the ISTP fleet of spacecraft which are intended for global investigations and require very different analysis techniques than those developed for a phased array of spacecraft, flying in formation. Coordinated measurements between spacecraft in ISTP studies require favourable conjunctions across often different regions of the magnetosphere, whereas closely spaced clusters probe small-scale structures within the same region, or across narrow boundaries.

Analysis of data from a closely-spaced cluster of spacecraft can be approached on several levels. On the lowest level, one would base the selection of events for further analysis on low-resolution summary data from one of the spacecraft. Folding in knowledge of the spatial configuration of the spacecraft array, relative to some (model) boundary normal or some dominant spatial direction, such as the magnetic field, can help test suitability for a given analysis objective (by comparing characteristic scales, anisotropy, or time dependence).

At the next level, simple inspection or analysis of high-resolution data from one, and eventually all, of the spacecraft, allows choice of events that meet specific selection criteria. At this level, it is also possible to check on whether the results at the different locations are mutually consistent with the nature of the selected event that was assumed in the applied analysis method. For example, whether they are consistent with assumptions of

stationarity or two-dimensionality, or particular spectral response. Having arrived at such a level of description, one could then decide whether the situation warrants application of truly multipoint techniques.

In accordance with this hierarchy of analysis levels, the book not only describes true multipoint techniques, but includes detailed descriptions of methods established for single-point measurements.

The book begins with a tutorial on Spectral Analysis (Chapter 1) because it must be realised that for any reasonable number of spacecraft in a cluster, the time domain is very much better covered than the spatial domain. For the comparison of multi-instrument and multipoint time series, time synchronisation is an important element. Therefore Chapter 2 (Time Series Resampling Methods) deals with the methods available for resampling of the data that are measured with different time resolution and the limitations that are involved.

Chapter 3 (Multi-Spacecraft Filtering: General Framework) describes a powerful technique to determine the \mathbf{k} vectors of a wave field from measurements of magnetic and electric fields at four points. Chapter 4 (Multi-Spacecraft Filtering: Plasma Mode Recognition) extends the scope of the analysis to infer the wave modes that are present, based upon fields and plasma measurements. Both chapters use synthetic data to test the techniques.

Chapters 5 (Measurement of Plasma Velocity Distributions) and 6 (Moments of the Plasma Velocity Distributions) are tutorials. They were included because the methods and limitations of plasma measurements and their expression in terms of moments are often not fully appreciated. Chapter 7 (Multi-Spacecraft Analysis of Plasma Kinetics) describes the use of particle distribution functions in inferring plasma kinetic effects, including remote sensing of boundaries.

Chapters 8–11 deal with the determination of the orientation and motion of plasma boundaries which is one of the main capabilities of multipoint studies. Chapter 8 (Variance Analysis) describes the estimators for the normal directions of current sheets based on variance analysis of magnetic and electric field data from a single spacecraft. Chapter 9 (DeHoffmann-Teller Analysis) deals with the determination of the coordinate system where the electric field vanishes, and describes the utility of such a system. Chapters 8 and 9 use the same satellite data to demonstrate the capability and limitations of the techniques. Chapter 10 (Shock and Discontinuity Parameters) is a general review of the methodology for the determination of shock and discontinuity parameters, primarily from single-point data. Chapter 11 (Multi-Spacecraft Discontinuity Analysis: Orientation and Motion) focuses on methods using multipoint magnetic field measurements.

Chapters 12 (Spatial Gradients and the Volumetric Tensor) and 13 (Tetrahedron Geometric Factors) describe various schemes to derive a quantitative measure of the configuration of the polyhedron formed by a cluster of co-orbiting spacecraft and its deformation along an orbit. Such measures are essential for pre-selecting data intervals for further analysis.

Chapters 14–17 describe the methods and accuracy with which spatial derivatives, such as the gradient, curl or divergence of some field, can be determined from multipoint measurements. These chapters are therefore fundamental for the prime purpose of multipoint missions. Chapter 14 (Spatial Interpolation for Four Spacecraft: Theory) describes the basic concepts. Chapter 15 (Spatial Interpolation for Four Spacecraft: Application to Magnetic Gradients) applies the technique to model magnetic fields. Chapter 16 (Accuracy of Current Density Determinations) focuses on the ability to infer currents from $\nabla \times \mathbf{B}$. Chapter 17 (Accuracy of Plasma Moment Derivatives) deals with the accuracy

that can be expected for spatial and time derivatives of plasma velocity moments.

Chapter 18, finally, is a review of numerical simulations related to plasma kinetic effects that are useful for guiding the analysis and interpretation of the measurements. This chapter thus falls into the category of mission-oriented theory that has been developed in support of NASA's ISTP missions, but also for Cluster, through the European Network for the Numerical Simulation of Space Plasmas.

Each chapter contains its own bibliography, which provides references to background material on the development of the subject and further clarification. Literature citations in the text have been deliberately minimised in order not to distract from the essentially tutorial nature of the book.

Although style and scope of the chapters differ, every effort has been made to ensure a reasonably uniform notation. Nevertheless, the authors have been allowed sufficient flexibility to avoid casting well-known equations into a form unfamiliar to experienced users.

In summary, the book is a collection of material on analysis techniques pertaining to multi-spacecraft data. The foundations of the methods are clearly provided, but only some of them have been tested with real or simulated data, and only a few are cast in the form of readily applicable tools. Moreover, the coverage is by no means complete; a noteworthy example is the lack of a chapter on multipoint analysis of fully-developed plasma turbulence. Nevertheless, it is hoped that the publication of this book will encourage further developments which could be incorporated in a future issue. $\Rightarrow 0.2$

Bibliography

Previous work on the subject has been published in several workshop proceedings:

Proceedings of the International Workshop on "Space Plasma Physics Investigations by Cluster and Regatta", Graz, 20.–22 Feb. 1990, ESA SP-306, European Space Agency, Paris, France, 1990.

Proc. International Conf. "Spatio-Temporal Analysis for Resolving plasma Turbulence (START)", Aussois, 31 Jan.–5 Feb. 1993, ESA WPP-047, European Space Agency, Paris, France, 1993.

Proceedings of Cluster Workshops, Braunschweig, 28–30 Sep. 1994, and Toulouse, 16–17 Nov. 1994, ESA SP-371, European Space Agency, Paris, France, 1995.

Proceedings of the Cluster-II Workshop Multiscale/Multipoint Plasma Measurements, Imperial College, London, 22–24 September 1999, ESA SP-449, European Space Agency, Paris, France, 2000.

See also the article

Song, P. and Russell, C. T., Time series data analyses in space physics, *Space Sci. Rev.*, **87**, 387–463, 1998.

The Cluster mission and its experiments are described in detail in

Escoubet, C. P., Russell, C. T., and Schmidt, R., editors, *The Cluster and Phoenix Missions*, Kluwer Academic, Dordrecht, Boston, London, 1997, reprinted from *Space Science Reviews*, Vol. 79, Nos. 1–2, 1997.

Notation Conventions

The conventions used in this book are as follows:

- Greek subscripts α, β, \dots apply to spacecraft; Latin subscripts i, j, \dots to cartesian coordinates.
- Vectors are indicated by boldface symbols, as $\mathbf{B}, \mathbf{v}, \boldsymbol{\omega}$. For purposes of matrix multiplication, these are considered to be *column vectors*; the corresponding row vectors are $\mathbf{B}^T, \mathbf{v}^T, \boldsymbol{\omega}^T$.
- Unit vectors are written as $\hat{\mathbf{n}}, \hat{\mathbf{b}}$.
- Matrices and tensors are represented by sans serif characters, e.g. \mathbf{M}, Π ; their transposed forms are \mathbf{M}^T, Π^T and their hermitian conjugates are $\mathbf{M}^\dagger, \Pi^\dagger$.
- Multiplication of vectors is marked with the standard operators for the dot ($\mathbf{a} \cdot \mathbf{b}$) and cross ($\mathbf{a} \times \mathbf{b}$) products.
- Matrix multiplication has no explicit operator; in this context, vectors are treated as column matrices, e.g.:

$$\mathbf{a}^T \mathbf{S} \mathbf{b} = \sum_{ij} a_i S_{ij} b_j$$

Thus the dyadic $\mathbf{a}\mathbf{b}^T$ represents the 3×3 tensor whose ij component is $a_i b_j$, while the product $\mathbf{a}^T \mathbf{b}$ is equivalent to $\mathbf{a} \cdot \mathbf{b} = \sum_i a_i b_i$.

— 1 —

Spectral Analysis

ANDERS I. ERIKSSON

*Swedish Institute of Space Physics
Uppsala, Sweden*

1.1 Introduction

Large amounts of data, like one or more time series from some spacecraft carried instruments, have to be reduced and presented in understandable quantities. As physical theories and models often are formulated in terms of frequency rather than time, it is often useful to transform the data from the time domain into the frequency domain. The transformation to frequency domain and application of statistics on the result is known as spectral analysis.

The literature on spectral analysis is voluminous. In most cases, it is written by experts in signal processing, which means that there are many texts available outlining the fundamental mathematics and elaborating the fine points. However, this is not the only background needed for a space physicist who is put to the task of actually analysing spacecraft plasma data. Simple questions on normalisation, physical interpretation, and how to actually use the methods in practical situations are sometimes forgotten in spectral analysis texts. This chapter aims at conveying some information of that sort, offering a complement to the textbooks rather than a substitute for them. The discussion is illustrated with idealised examples as well as real satellite data.

In order not to expand the chapter into a book in itself, we concentrate on the application of basic Fourier and wavelet methods, not treating interesting topics in time series analysis like stationarity tests, filtering, correlation functions, and nonlinear analysis methods. Higher order methods like bispectral analysis are also neglected. Fundamentals of such items are covered by many of the references to this chapter, and methods particularly suited for multipoint measurements are of course found throughout this book. Other introductions with multi-spacecraft applications in mind can be found in the paper by *Glassmeier and Motschmann* [1995] and in other papers in the same publication.

The disposition of the chapter is as follows. First, we introduce the basic concepts in Section 1.2, where we also discuss time-frequency methods for the analysis of non-stationary signals. The practical implementation of classical Fourier techniques is treated in Section 1.3, while the implementation of Morlet wavelet analysis is discussed in Section 1.4. In Section 1.5, we turn to the simultaneous analysis of two or more signals by the cross spectrum technique, particularly relevant for the analysis of multipoint measurements. Finally, we touch upon the use of parametric spectral methods in Section 1.6.

1.2 Basic Concepts

1.2.1 Fourier Series

Given any process, represented as a function $u(t)$ of time t , we can only study it for some finite time span $t_0 < t < t_0 + T$. Fourier's theorem states that if the signal is finite and piece-wise continuous, it can be written as

$$u(t) = \sum_{n=-\infty}^{\infty} \tilde{u}[n] \exp(-2\pi i f_n t) \quad (1.1)$$

where

$$f_n = n/T \quad (1.2)$$

and the complex Fourier coefficients $\tilde{u}[n]$ are given by

$$\tilde{u}[n] = \frac{1}{T} \int_{t_0}^{t_0+T} u(t) \exp(2\pi i f_n t) dt \quad (1.3)$$

The transformation 1.3 replaces a continuous infinity of values $u(t)$ on a finite interval $0 \leq t < T$ by an infinite series of values $\tilde{u}[n]$ for all integers n . Each term in the Fourier sum 1.1 corresponds to an oscillation at a frequency $f_n = n/T$. As such oscillations are the eigenmodes of small-amplitude perturbations in most physical systems, this kind of transformation is an obvious means for understanding the physics behind some signal.

One may note that the coefficients $\tilde{u}[n]$ and $\tilde{u}[-n]$ corresponds to oscillations with the same value of $|f_n|$, which is the quantity determining the time scale of the oscillation. Hence, the sum of the two terms for n and $-n$ in equation 1.1 will describe the part of the signal corresponding to a certain time scale $1/|f_n|$, so for a real signal, this sum must be real. Thus,

$$\tilde{u}[-n] = \tilde{u}^*[n] \quad (1.4)$$

where the star denotes complex conjugation. For a real signal, the two terms labelled by $\pm n$ may be considered to represent the same frequency $|f_n|$. While the $\tilde{u}[n]$ coefficients are sometimes called Fourier amplitudes, the sum of the two $\pm n$ terms in equation 1.1 evaluated using equation 1.4 shows that the amplitude of the sinusoidal wave of frequency $|f_n| > 0$ in equation 1.1 in fact is $2|\tilde{u}[n]|$.

1.2.2 Parseval's Relation and Power Spectral Density (PSD)

Central to the physical interpretation of the Fourier series is Parseval's relation

$$\frac{1}{T} \int_{t_0}^{t_0+T} u^2(t) dt = \sum_{n=-\infty}^{\infty} |\tilde{u}[n]|^2 \quad (1.5)$$

which in the case of a real signal, where equation 1.4 applies, becomes

$$\frac{1}{T} \int_{t_0}^{t_0+T} u^2(t) dt = \tilde{u}^2[0] + 2 \sum_{n=1}^{\infty} |\tilde{u}[n]|^2 \quad (1.6)$$

The left-hand side is an average of what we may call the average signal energy or signal power¹. This nomenclature is not without foundation: for instance, if $u(t)$ is a component of the electric field, magnetic field, or plasma velocity, this signal energy is related to the contribution to the physical energy density (SI unit J/m³) in the plasma from this field component by a factor of $\varepsilon_0/2$, $1/(2\mu_0)$, or $\rho/2$, respectively, where ρ denotes the mass density. Parseval's relation opens up the possibility of interpreting each of the terms in the sum to the right in equation 1.6 as the contribution to the signal energy from the corresponding frequency f_n .

The terms in the sum at the right of equation 1.6 depend on the length of the time interval T . If this interval is doubled to $2T$ by adding the same data set again, i.e. by taking $u(t+T) = u(t)$, $0 \leq t < T$, the left-hand side of 1.6 will stay constant, while there will be twice as many terms as before in the right-hand sum, the frequency spacing

$$\Delta f = 1/T \quad (1.7)$$

decreasing by half when T is doubled. On the average, the value of the terms $|\tilde{u}[n]|^2$ in the sum in 1.6 therefore must decrease by a factor of 2. Hence, the coefficients $|\tilde{u}[n]|^2$ depends on signal length, and they are therefore not suitable for describing the physical process itself. To describe the distribution of signal energy density in frequency space, we instead introduce a function S_u known as the *power spectral density (PSD)* by

$$S_u[n] = 2T |\tilde{u}[n]|^2 \quad (1.8)$$

for all non-negative integers n . Parseval's relation 1.6 then takes the form

$$\frac{1}{T} \int_0^T u^2(t) dt = S_u[0] \frac{\Delta f}{2} + \sum_{n=1}^{\infty} S_u[n] \Delta f \quad (1.9)$$

Having S_u defined in this way, its value at a particular frequency will not change if changing the record length T in the way outlined above. It is thus possible to picture the discrete set of PSD values as samples $S_u[n] = S_u(f_n)$ of a continuous function $S_u(f)$. In this picture, the Parseval relation 1.9 is the trapezoidal approximation to the integral relation

$$\frac{1}{T} \int_0^T u^2(t) dt = \int_0^{\infty} S_u(f) df \quad (1.10)$$

Our definition of the PSD has the virtue of having an immediate physical interpretation: $S_u(f_n)\Delta f$ is the contribution to the signal energy from the frequency interval Δf around f_n . It can only be used for real signals, since negative frequencies are ignored.

1.2.3 Phase

As the PSD (equation 1.9) is based on the squared magnitude of the complex Fourier coefficients $\tilde{u}[n]$, it obviously throws away half of the information in the signal. The other half lies in the phase spectrum $\varphi[n]$, defined by

$$\tilde{u}[n] = |\tilde{u}[n]| \exp(i\varphi[n]), \quad (1.11)$$

¹Signal processing texts sometimes distinguish between the energy and the power of a signal: see the discussion by *Champeney* [1973], chapter 4.

with $\varphi[-n] = -\varphi[n]$ and $\varphi[0] = 0$ due to equation 1.4 The absolute value of the phase depends on exactly when we start sampling the signal, so there is little physical information in one single phase value by itself. On the other hand, there is a lot of information to be gained by studies of the relative phase of two signals, as is further discussed in Section 1.5.

If we know the phase spectrum as well as the PSD, no information has been lost and the signal itself can be reconstructed by using equations 1.11 and 1.8. Constructing time series compatible with some given PSD is sometimes useful in numerical modelling and simulations of space plasma situations. The phase may be selected at random, or by some assumption motivated by the model. One should note that the random phase assumption may result in data with significantly different statistics in the time domain than had the original signal. The physics represented by the two signals can therefore be quite different. Methods for treating such problems exist [see *Theiler et al., 1992*, and references therein].

1.2.4 Discrete Fourier Transform (DFT)

Our principal interest is in data from instruments providing a sampled rather than continuous output: this replaces the continuous function $u(t)$ above by a discrete set of N measurements

$$u[j] = u(t_j) = u(t_0 + j \Delta t) \quad (1.12)$$

where Δt is the sampling spacing, whose inverse is the sampling frequency f_s , and $j = 0, 1, 2, \dots, N - 1$. From equation 1.2, we get

$$f_n = \frac{n}{T} = \frac{n}{N \Delta t} = \frac{n}{N} f_s \quad (1.13)$$

By replacing the integrals in the Section 1.2.1 above by sums, where dt is replaced by Δt , we define the discrete Fourier transform (DFT)

$$\tilde{u}[n] = \frac{1}{N} \sum_{j=0}^{N-1} u[j] \exp(2\pi i n j / N) \quad (1.14)$$

and its inverse

$$u(t_j) = u[j] = \sum_{n=-N/2}^{N/2-1} \tilde{u}[n] \exp(-2\pi i n j / N) \quad (1.15)$$

where we have assumed N to be even (generalisation to odd N is straightforward). The index n here runs from $-N/2$ to $N/2 - 1$, but it is customary to let it run from 0 to $N - 1$ instead, by defining $\tilde{u}[n + N] = \tilde{u}[n]$. This is possible since the only effect of replacing n by $n + N$ in the exponential in 1.15 is a multiplication by $\exp(2\pi i N j / N) = 1$. We thus write the inverse DFT as

$$u(t_j) = u[j] = \sum_{n=0}^{N-1} \tilde{u}[n] \exp(-2\pi i n j / N) \quad (1.16)$$

For a real signal, equation 1.4 told us that the negative frequency components carried no extra information, and this is the case for $n \geq N/2$ as well:

$$\tilde{u}[N - n] = \tilde{u}^*[n] \quad (1.17)$$

For the DFT of a real signal, Parseval's relation takes the form

$$\frac{1}{N} \sum_{j=0}^{N-1} u^2[j] = \sum_{n=0}^{N-1} |\tilde{u}[n]|^2 = \frac{1}{2} S_u[0] \frac{f_s}{N} + \sum_{n=1}^{\frac{N}{2}-1} S_u[n] \frac{f_s}{N} + \frac{1}{2} S_u[N/2] \frac{f_s}{N} \quad (1.18)$$

where the PSD estimate is

$$S_u[n] = \frac{2N}{f_s} |\tilde{u}[n]|^2, \quad n = 0, 1, 2, \dots, N/2 \quad (1.19)$$

If the signal is measured in some unit X and the frequency unit is hertz (Hz), the unit of the DFT as defined by 1.14 also will be X, while the PSD will be in units of X²/Hz. This is appropriate, as we may picture the Parseval relation 1.18 as an approximation of the integral relation 1.10.

1.2.5 Normalisation

The definitions 1.14 and 1.16 of the DFT and its inverse are not the only possible. For instance, the choice of in which of the exponentials in equations 1.14 and 1.16 we place the minus sign is quite arbitrary, but for a real time series, this does not change the physical results. The convention adopted here is the familiar one for a physicist, as harmonically oscillating properties normally are represented by functions of the form $\exp(-i\omega t)$ in the physics literature. On the other hand, works in signal processing usually use the opposite sign in the exponentials.

More important is that the factor of $1/N$ we have placed in 1.14 is sometimes placed in front of the sum in 1.16 instead, or even, for greater symmetry, split into two factors of $1/\sqrt{N}$ in front of each of the sums in 1.14 and 1.16. Other factors may also be introduced: see Table 1.1 for a sample of different conventions found in the literature and in some software packages. For a physicist, it is usually natural to think of the Fourier decomposition as a means of separating oscillations at various frequencies, which means that the inverse Fourier transform should be just a sum of the Fourier coefficients multiplied by complex exponentials, as in equation 1.16.

As a further complication, there is little agreement in the literature on how to define the PSD in terms of the Fourier coefficients, i.e., how to write equation 1.19. One reason for this is that many works on spectral analysis are written by mathematicians or experts in signal processing, for whom explicit appearance of sampling frequency in the PSD definition or the confinement to real signals are undesirable. From the point of view of a practising space plasma physicist, a normalisation of the PSD such that Parseval's relation is of the form 1.18 is most often (but not always) the natural choice. As the conventions vary between different algorithms and software packages, it is wise to check the normalisation in the particular routine one is using by straightforward evaluation of both sides of equation 1.18. When publishing spectra, a good service to the reader is to tell what normalisation convention has been used, for example by showing the form of Parseval's relation.

These issues of normalisation may seem dry and uninteresting, but are in fact quite important: somebody will look at the power spectra you publish, try to estimate wave amplitudes in a frequency band or field energy or something similar, and get wrong answers if she/he has misunderstood your normalisation. Trivial as it may seem, the normalisation of spectra is remarkably often a source of confusion.

Table 1.1: Samples of DFT definitions used in the literature and some software packages.

Normalisation conventions in Fourier analysis

Fourier transform: $\tilde{u}[n] = A \sum_{j=0}^{N-1} u[j] \exp(2\pi B i j n / N)$				
Inverse transform: $u[j] = C \sum_{n=0}^{N-1} \tilde{u}[n] \exp(2\pi D i j n / N)$				
	A	B	C	D
This work	$1/N$	+	1	-
IDL	$1/N$	-	1	+
Mathematica	$1/\sqrt{N}$	+	$1/\sqrt{N}$	-
Matlab	1	-	$1/N$	+
<i>Bendat and Piersol</i> [1971]	1	-	$1/N$	+
<i>Brockwell and Davis</i> [1987]	$1/\sqrt{N}$	-	$1/\sqrt{N}$	+
<i>Brook and Wynne</i> [1988]	$1/N$	-	1	+
<i>Kay</i> [1988]	1	-	$1/N$	+
<i>Marple</i> [1987]	T	-	$1/(TN)$	+
<i>Press et al.</i> [1992]	1	+	$1/N$	-
<i>Strang and Nguyen</i> [1996]	1	-	$1/N$	+
<i>Welch</i> [1967]	$1/N$	-	1	+

1.2.6 Time-Frequency Analysis

In their most straightforward application, the Fourier methods described above transform a function $u(t)$ into a function $\tilde{u}(f)$. Hence, one time series is transformed into one spectrum. This type of transform is sensible if the process one is studying is stationary. However, the space plasma physicist often concentrates on dynamical situations rather than truly stationary phenomena. In addition, the interest is often directed toward boundary layers or other inhomogeneous regions, where the record of data will be non-stationary, due to the motion of the spacecraft through the inhomogeneous medium, even if it results from processes stationary at any single point in the plasma. As an example, we may take a spacecraft crossing a magnetopause: the characteristic frequencies and the waves present are different inside and outside the magnetosphere, and in addition there are wave phenomena associated with the magnetopause itself.

In such cases, a more useful representation of the signal will result if we apply the concept of two time scales: one fast time scale, to be represented in the frequency domain, and one slow time scale, for which we keep the time domain representation. We then split the signal record of N samples into M shorter intervals, each of length L , and calculate the PSD for each of these. This technique is known as the short-time Fourier transform (STFT) or the windowed Fourier transform (WFT), as the selection of a particular part of the time series can be seen as a multiplication with a rectangular window function

$$w_k[j] = \begin{cases} 0, & j < kL \\ 1, & kL \leq j < (k+1)L \\ 0, & j > (k+1)L \end{cases} \quad (1.20)$$

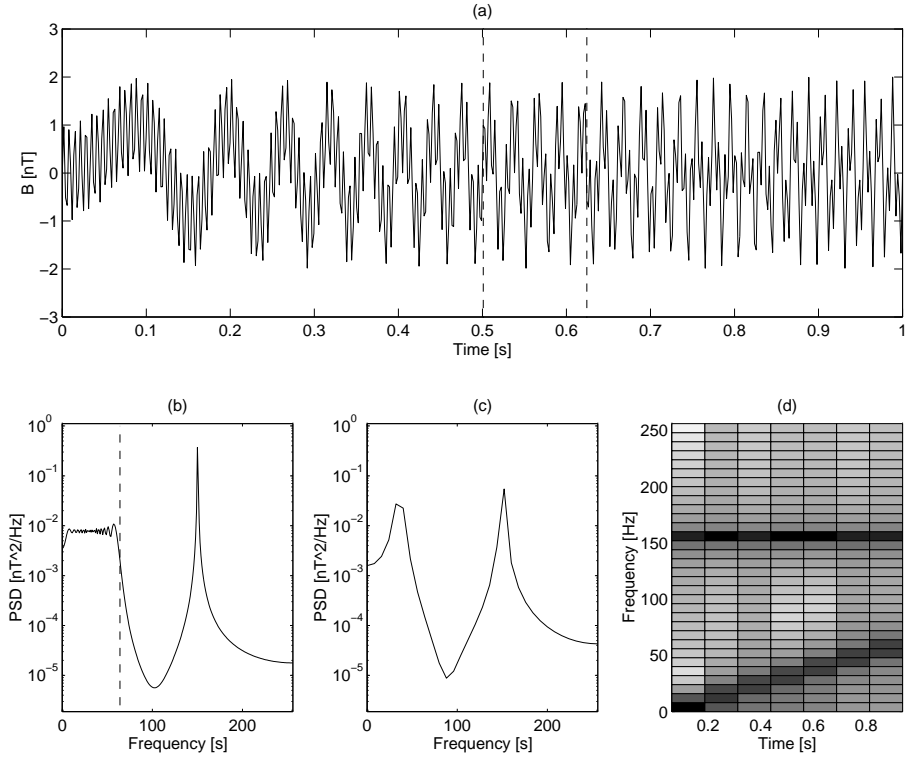


Figure 1.1: Time-frequency analysis. (a) One second of the time series of the signal in equation 1.21, sampled at 512 samples/s. (b) 512-point PSD of the signal in (a). (c) 64-point PSD of the part of the signal between the dashed lines above. (d) 64-point PSDs of 8 subintervals of the signal above. Dark is high and light is low values of log PSD.

Generally, the technique of representing fast variations in the frequency domain and slow variations in the time domain is known as time-frequency analysis. Data analysed in this way are normally presented as colour or grey-scale coded panels, which often are known as dynamic spectrograms. Examples using real data can be found in Figure 1.10 on page 30.

Another example, based on synthetic data, is shown in Figure 1.1. The synthetic signal in panel (a) is represented by $N = 512$ samples of

$$u(t) = B \sin 2\pi f(t)t + B \sin 2\pi f_0 t \quad (1.21)$$

where $0 < t < 1$ s, $B = 1$ nT, $f_0 = 150$ Hz, and $f(t)$ rises linearly from zero at $t = 0$ to $f_s/8$ at $t = 1$ s. This signal includes a non-stationary part (the first term in equation 1.21) as well as a stationary part (the second term). Panel (b) shows the PSD calculated from equation 1.19. The stationary part is well represented by a narrow spectral peak, while the energy of the non-stationary part is spread out in a frequency band extending up to about $f_s/8$ (the dashed vertical line) without further information. Panels (c) and (d) show

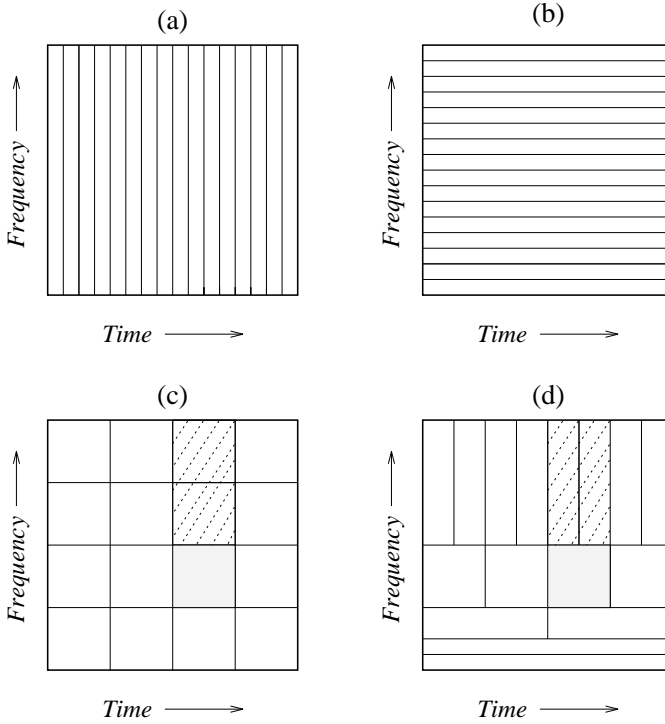


Figure 1.2: Some possible divisions of the time-frequency plane into 16 equal-sized rectangles. (a) Time domain representation. (b) Fourier representation. (c) Time-frequency Fourier representation (WFT). (d) Octave band analysis (wavelet transform). The meaning of the shading is explained in the text on page 28 just after equation 1.43. For an explanation of the dashed rectangles, see page 32, Section 1.4.6.

the effect of dividing the original record into $M = 8$ intervals of length $L = 64$ samples (0.125 s). In (c), we have plotted the PSD (as defined by 1.19) based on the $L = 64$ points between the dashed lines in panel (a). Finally, in (d) we see a time-frequency plot, where the PSD is represented by a logarithmic grey scale.

Obviously, the time-frequency analysis has several merits. In panel (d) of Figure 1.1 we can clearly see the linear trend in the frequency of the low frequency signal, and the stationary character of the high-frequency part. Also, in panel (c), the low frequency signal is represented by a peak with a maximum value of the same order as the peak value for the stationary high-frequency signal, which is reasonable. The low-frequency peak is of course broader, as the frequency of this signal varies from 32 Hz to 40 Hz during the analysed time interval. On the other hand, the representation of the stationary signal at 150 Hz is better in (b) than in (c), in the sense that its energy is more concentrated in a narrow spectral peak.

This illustrates the basic trade-off situation in time-frequency analysis: we can get better resolution in time only at the expense of resolution in frequency. Having a time series of N points sampled at frequency f_s , we get a frequency resolution $\Delta f = f_s/N$ if we calculate the Fourier transform of the full record. In that case, the time resolution is

equal to the length of the record, $\Delta t = N/f_s$, so that $\Delta f \Delta t = 1$. Dividing into M records increases the time resolution by a factor of M , while the frequency resolution goes down by the same factor, so that we still have $\Delta f \Delta t = 1$. In general,

$$\Delta f \Delta t \geq 1 \quad (1.22)$$

where the inequality results when averaging is introduced (Section 1.3.4). As our definitions of Δf and Δt are not in terms of standard deviation as in quantum mechanics [Landau and Lifshitz, 1977, p. 47], one should not exactly identify 1.22 with the Heisenberg uncertainty relation, although the two are related.

The relation 1.22 may be illustrated as in Figure 1.2, where we divide the time-frequency plane into 16 rectangles of unit area in four different ways. Panel (a) represents the untransformed time series,² and panel (b) the DFT of the full signal. Panel (c) is an example of a time-frequency Fourier analysis.

We could also think of less symmetric ways of dividing the time-frequency plane of Figure 1.2. Panel (d) shows a particularly useful example, known as octave band analysis. Here, the time resolution is different for different frequencies, giving better resolution in time for higher frequency components. This decreases the frequency resolution at high frequencies, but the relative resolution $\Delta f/f$ is kept constant. We could express this as having constant resolution in $\log f$. Implementing this kind of division of the time-frequency plane leads to what is known as wavelet analysis, to which we return in Section 1.4.

Which division of the time-frequency plane in Figure 1.2 is best? The answer clearly depends on the properties of the signal, and the questions we put to it. If the signal, or at least the component of it we are interested in studying, can be assumed to result from a truly stationary process, the resolution is better placed in the frequency than in time, as in panel (b). If, in addition, the signal can be assumed to be composed of a set of discrete frequencies, there may even be reason to consider parametric spectral models, which we briefly touch upon in Section 1.6. A Fourier implementation of time-frequency analysis (panel (c)) is useful for a situation with one slow time scale, represented in the time domain, modulating fast oscillations, represented in the frequency domain. An example of a situation where such a description is useful is a magnetospheric satellite flying through a density cavity, where the slow time scale is set by the size of the cavity divided by the spacecraft speed, while the waves present the fast variations. Finally, the wavelet division of the time-frequency plane in panel (d) is well suited for turbulent situations, with a more or less continuous hierarchy of time scales.

Wavelet analysis can be extended to include other unsymmetric partitions of the time-frequency plane than the one in Figure 1.2d. We touch briefly upon this in Section 1.4.6 (page 32).

As well as depending on the signal properties, the method to use is to some extent dependent on our interpretation of the signal, which in turn may depend on other data, our experience, or our prejudice. Consider the time series of synthetic data in panel (a) of Figure 1.3. The time-frequency analysis in panel (b) suggests a description in terms of a modulated wave at 300 Hz. On the other hand, the single Fourier spectrum of all points in panels (c) and (d) suggests a description as a superposition of two sinusoidal waves at 297 Hz and 303 Hz. Mathematically, these descriptions are of course equivalent. However,

²Formally, panel (a) should be interpreted as $M = 16$ DFTs, each based on $L = 2$ data points, from a total record of $N = 32$ samples, since the N -point PSD only gives information at $N/2$ frequencies.

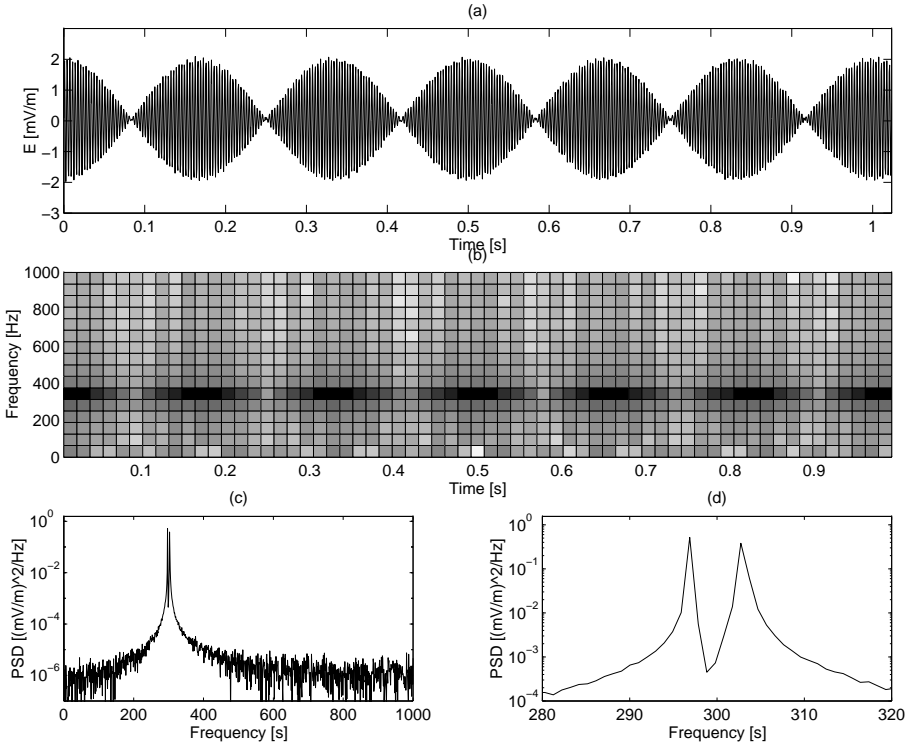


Figure 1.3: A modulated sine wave (a), with some spectral representations: (b) Fourier time-frequency analysis, (c) Fourier spectrum, (d) detail of Fourier spectrum.

one or the other may be more physically meaningful in a given situation. For example, if some other measurement shows that a likely source of the wave is an electron beam whose energy flux pulsates with a frequency corresponding to the observed modulation frequency, a description in terms of modulations, as in panel (b), is appropriate. On the other hand, if we have reason to believe in the presence of two distinct localised and stable sources with slightly different characteristics, the description offered by panels (c) and (d) is more natural.

In the 1990s, wavelet methods (Section 1.4) have become popular. Wavelets constitute a major advance in the spectral analysis, opening new approaches to signal handling, but there is no reason to throw the traditional Fourier based methods overboard. As outlined above, and as will further be discussed in Section 1.4.2, there are applications where a division of the time-frequency plane as in panel (c) of Figure 1.2 is more appropriate than the wavelet division in panel (d). In other circumstances, the wavelet division is the more advantageous.

1.3 Fourier Techniques

This section considers how to practically use Fourier methods for PSD estimation. Ideas and concepts are described, while detailed algorithms are left for textbooks like *Jenkins and Watts* [1968], *Bendat and Piersol* [1971], and *Kay* [1988]. For the reader interested in clear and concise treatments on a level between this section and the textbooks, the old but concise paper by *Welch* [1967] and the practically oriented text by *Press et al.* [1992] are recommended.

1.3.1 Fast Fourier Transform (FFT)

The fast Fourier transform (FFT) is an algorithm for efficient implementation of the DFT 1.14 and its inverse 1.16. This algorithm is the only implementation of the DFT that you are ever likely to see or use, as it is far more efficient than for example a direct summation of the series in equation 1.14. The FFT algorithm is described in most textbooks on signal processing [e.g. *Bendat and Piersol*, 1971] or mathematical computer methods [e.g. *Press et al.*, 1992], and there you will find out that while a direct implementation of equation 1.14 requires a number of calculations proportional to N^2 , the clever FFT algorithm only needs on the order of $N \log N$ calculations for doing the same job.

The FFT algorithm is included in virtually all software packages intended for data analysis, like Matlab or IDL, so you should never have to actually write the code yourself or even understand it in detail. However, one should be aware that the computational efficiency of the FFT is fully exploited only when the number of data points is an integer power of two ($N = 2^M$).

1.3.2 Detrending

The finite time series from a measurement is often to be regarded as a finite observation of a process, stationary or non-stationary, that was going on before the start of our observations and that will continue afterwards. Hence, it is very likely that our data contains components with periods longer than the length of the data record. Such components in the signal are called *trends*. As the trends represents time scales longer than those a spectral analysis of our data record can resolve, it is desirable to remove them from the data prior to spectral estimation. If not removed, trends may distort the low-frequency end of the estimated spectrum.

Detrending may often be done by fitting (by least-squares methods or otherwise) a linear or perhaps quadratic function to the data, and thereafter subtract this function. Sometimes, more advanced fits like sinusoidals may be called for.

An example of a signal with trends is shown in Figure 1.4. Panel (a) shows 3 seconds of measurements of a spin-plane component of the electric field by a magnetospheric satellite spinning around its axis with a period of 6 seconds. One trend is obvious: there is a sinusoidal variation at the spin frequency, of which we see half a period. This signal is due to spin modulation of the electric field induced by the spacecraft motion through the geomagnetic field, and is indicated by the dashed line. Removing this, we get a time series as in panel (b). Considering the part of the time series between the vertical lines in panel (b), we find another trend as well, due to a field variation at a time-scale longer than the length of this interval. Panel (c) shows an enlargement of this part of the total time series,

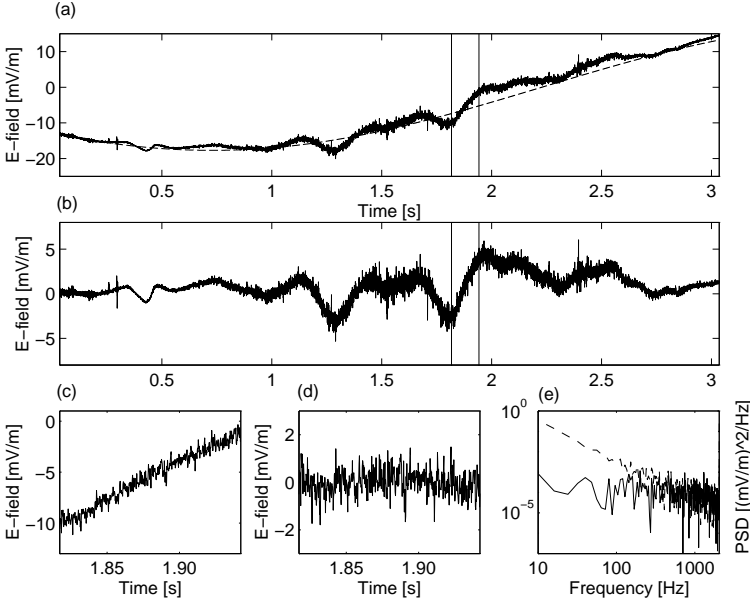


Figure 1.4: Measurement of an electric field component by the wave instrument on the Freja satellite. Zero time corresponds to UT 005416, April 12, 1994. The signal has been low-pass filtered at 1.3 kHz before sampling at 4000 samples/s.

before any of the two trends have been removed. Panel (d) shows the same signal after removal of a linear trend by subtracting a least-squares fit. Finally, panel (e) displays the power spectra of the original signal from panel (c) (dashed) and of the detrended signal from panel (d) (solid), calculated according to equation 1.19 with a 512 point DFT. It is clear that the linear trend carries energies to all frequencies, thereby drowning details in the signal we really are interested in. For example, the peak in the solid curve around 30 Hz cannot be discerned in the dashed curve, where it is completely drowned by the spectral energy from the linear trend.

Removing a trend is a major manipulation of the data. Hence, caution is needed, so that the detrending does not do more harm than good. A few non-physical outliers in the data could wreak havoc in a fitting algorithm, and thus cause the detrending to actually add a spurious component in the data. Before removing a trend, its physical origin should be understood. Visual inspection of the time series before and after detrending is a very good idea.

1.3.3 Windowing

In Section 1.2.6, we introduced the rectangular window function 1.20. The abrupt behaviour of the rectangular window at the edges causes it to leak quite a lot of energy from any main spectral peak to other frequencies. This can be understood by considering that when applying a data window to the time series, we will analyse the signal $w_k[j] u[j]$

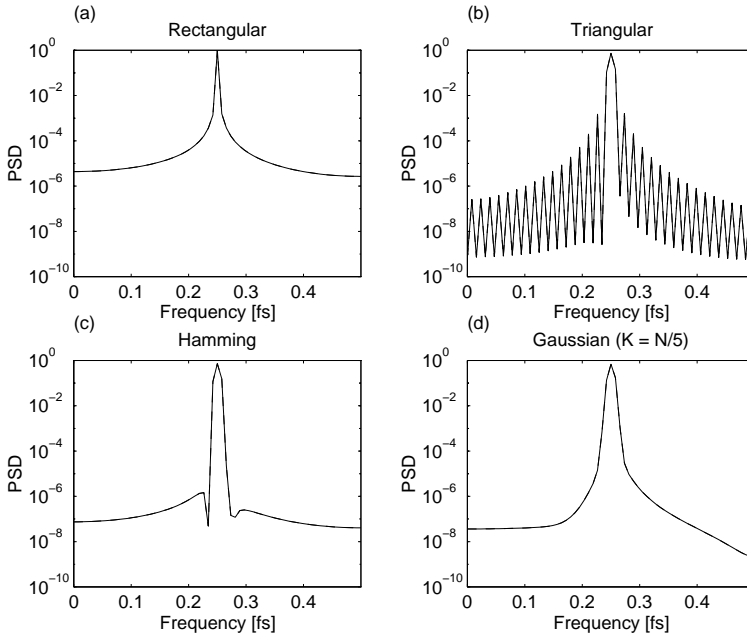


Figure 1.5: 128 point PSDs of a sinusoid at $f = f_s/5$ calculated using different data windows. The frequency is in units of the sampling frequency f_s . The PSD values are normalised so that the total integrated power for $0 < f < f_s/2$ is unity.

rather than the presumably infinite record of $u[j]$ itself. As a multiplication in the time domain is a convolution in the frequency domain, the effect of applying a window will be to convolve the “true” PSD with the PSD of the window. As the Fourier transform of a step function is spread out over all frequencies, this implies that part of the signal energy is moved from any frequency to all others. We illustrate this by an idealised example in Figure 1.5a. The 128 point PSD of a signal consisting of only a pure sinusoid at $f = f_s/5$ is calculated (using a rectangular window 1.20) and normalised so that the total integrated PSD in $0 < f < f_s/2$ is unity. It can be seen that signal energy leaks to the entire frequency interval. The detailed behaviour depends on the frequency of the signal: in the extreme example of an integer number of wave periods fitting into the data interval, there is no leakage at all, as the signal frequency exactly equals one of the Fourier frequencies 1.13 in this case. For a real signal, this is of course a zero-probability event, and a leakage as in Figure 1.5a will generally occur.

As a remedy for the frequency leakage behaviour, a variety of other window functions have been designed, a few of which are shown in Figure 1.6. As can be seen from the idealised example in Figure 1.5, these windows all decrease the leakage of energy to frequencies far away, at the cost of increased width of the main peak. That this price has to be paid is clear from the uncertainty relation 1.22: the window concentrates the signal in time, and hence must give a widened spectral peak.

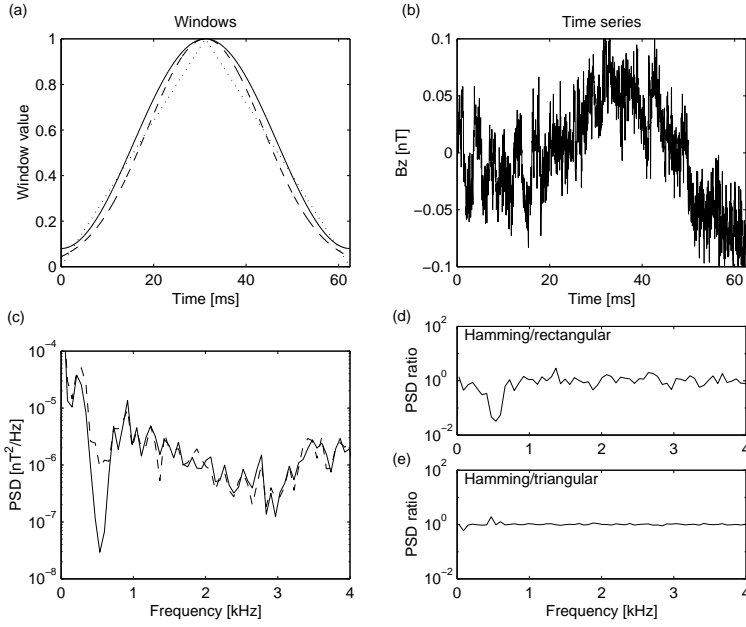


Figure 1.6: (a) Some often used data windows: triangular (dotted), Hamming (solid) and Gaussian ($K = L/5$, dashed). (b) Time series of magnetic field data from the Freja wave instrument. (c) PSD calculations using Hamming (solid curve) and rectangular (dashed) windows. (d) and (e) Ratios of PSDs calculated with different windows.

The simplest of these windows is the triangular (or Bartlett) window, given by

$$w[j] = 1 - \left| 1 - \frac{2j}{L-1} \right| \quad (1.23)$$

inside the interval of interest (which is the same as in equation 1.20 above) and zero outside. As is seen in Figure 1.5(b), it has a considerably better behaviour than the rectangular window in terms of energy spread. Even better in this respect and very widely used are the Hamming window

$$w[j] = 0.54 - 0.46 \cos\left(\frac{2\pi j}{L-1}\right) \quad (1.24)$$

and the Hann (often called Hanning) window (not shown)

$$w[j] = 0.5 - 0.5 \cos\left(\frac{2\pi j}{L-1}\right) \quad (1.25)$$

Both are zero outside the L -point interval of interest. The Gaussian window

$$w[j] = \exp\left(-\frac{(j - \frac{L-1}{2})^2}{2K^2}\right) \quad (1.26)$$

does not go exactly to zero, but may in practice be cut off at some distance from the centre. For an infinite time series, this window is optimal in terms of low leakage to frequencies, which makes it theoretically favoured [Rönnmark, 1990]. It differs from the others in that its width K is an adjustable parameter (chosen to $L/5$ in our examples). We will return to the Gaussian window when discussing Morlet wavelet analysis in Section 1.4.1.

The detailed properties of these and other windows have been studied by for example Harris [1978]. However, for most applications to data sampled in space, there is little point in putting much effort into the choice of data windows. In practice, the difference between spectra obtained by use of Hamming, Hann and Gaussian windows will be small: even the triangular window is tolerable for many purposes. This is illustrated by an analysis of magnetic field data from the Freja satellite shown in Figure 1.6(b). The PSDs of this time series (using averaging of 4 spectra, see Section 1.3.4) calculated with rectangular and Hamming windows are shown in panel (c). It can be seen that the Hamming window can resolve the minimum near 0.5 kHz much better than can the rectangular window. This difference is illustrated in panel (d), where the ratio of the PSDs in panel (c) is plotted. Finally, in panel (e) we plot the corresponding ratio for the Hamming and triangular windows, and find a much smaller difference. The important point is to use some reasonable window other than the rectangular, and preferably not the triangular.

In general, when multiplying the signal by window coefficients, which all are ≤ 1 , signal energy is lost. The exact amount of energy lost will depend on the signal: obviously, a record consisting of a single narrow spike in the middle of the interval, surrounded by zeroes, will lose no energy at all as the window functions all have the value one at the centre, while a record consisting of a spike at one of the end points and all other points being zero will be almost completely lost. By considering Parseval's relation 1.19, we find that the statistically expected decrease in the PSD value due to windowing should be the mean square value of the window,

$$W_{ss} = \frac{1}{N} \sum_{j=0}^{N-1} w[j]^2 \quad (1.27)$$

In order to compensate the PSD for the energy loss in the windowing, it should be divided by the W_{ss} value for the window at hand.

1.3.4 Averaging and Stationarity

In most texts on signal analysis [e.g. Kay, 1988; Welch, 1967], it is shown that when applying 1.19 with rectangular window to a time series resulting from a stationary random process, we get standard deviations in the PSD estimate equal to the PSD values themselves, so that the expected error is 100%. This situation is improved (and complicated) by the use of data windows other than the rectangular window (see, for instance, Welch [1967], Jenkins and Watts [1968], or Brockwell and Davis [1987]), but is still not satisfactory.

We emphasise the assumptions made above: (1) the time series is one particular realisation of a random process whose parameters we want to estimate, and (2) these parameters are constant or slowly varying, so that the process is almost stationary. It is only in these circumstances the question of standard deviation enters the problem. If we interpret our data as a unique observation of a deterministic non-stationary process whose details we

want to explore, there is no problem in the fact that consecutive spectra may show large variation—it simply reflects the changing physical situation. We will return to this point below, after briefly having considered the stationary case.

If we actually have an almost stationary random process, our first attempt to decrease the variance in the PSD estimates may be to increase the record length L . However, the effect of increasing the number of points in the DFT is only to get a PSD evaluation at more frequencies, not to improve the estimate at any single frequency, so this approach only gives us more points with the same bad variance. Increasing the record length is still wise, but we should not put all the samples into the DFT at once. Instead, we divide the total record of L samples into P shorter pieces of data, each of length K . After detrending and windowing each K -point interval, we compute its K -point PSD. Finally, we take the average of the P PSDs as our spectral estimate. Averaging reduces the variance with a factor of $1/P$, so in the case of rectangular windows, the normalised standard error will go down from the 100% mentioned above to

$$\epsilon = 1/\sqrt{P} \quad (1.28)$$

Some data analysis packets include means for calculating the confidence interval of the PSD estimate of a given time series, using methods described by e.g. *Jenkins and Watts [1968]*. If such means are not utilised, one could use the rule of thumb that spectral features below the $1/\sqrt{P}$ level should not be trusted to result from a stationary process.

Averaging can also be performed in the frequency domain rather than in time [e.g. *Bendat and Piersol, 1971*]. In that case, one calculates the PSD of all L data points, and then replace the $L/2$ PSD values one gets by $L/(2P)$ values, each being the average of P neighbouring frequency components in the original PSD. The effect on the standard deviation of the PSD estimate is similar to the averaging of spectra in the time domain.

By reducing the amount of information in the original signal by a factor of $1/P$, we increase the quality and comprehensibility of the remaining data. The loss of information in the averaging is described by the uncertainty relation 1.22, which for averaged spectra becomes

$$\Delta f \Delta t \geq P \quad (1.29)$$

Averaging is useful for PSD estimates, but not for estimates of the complex DFT 1.19 or the phase spectrum 1.11, as the absolute value of the phase angle is completely dependent on exactly when the data interval starts.

While averaging is useful for reducing noise from statistical fluctuations in a signal from a random process, it is not always justified for a deterministic signal. The idea of improving the PSD calculation is linked to the concept of our signal as a sample signal from a random process. In this view, the PSD is a statistical property of the signal, and we want to estimate it with as little noise as possible. This is often a very reasonable way of looking at the signal, but is not the only possible. For averaging to be physically justified, the signal must be assumed stationary over the time interval from which we construct the PSDs which we average. However, we are sometimes interested in the spectral properties of a non-stationary signal. For example, some sort of wave packet may be observed when passing a spacecraft. There is only this wave packet, so there is no ensemble to average over. Still, its spectral properties can be very interesting. It is more natural to look at this wave packet as a deterministic signal whose physical properties we can investigate in detail than to consider it a sample from a random process whose statistical parameters

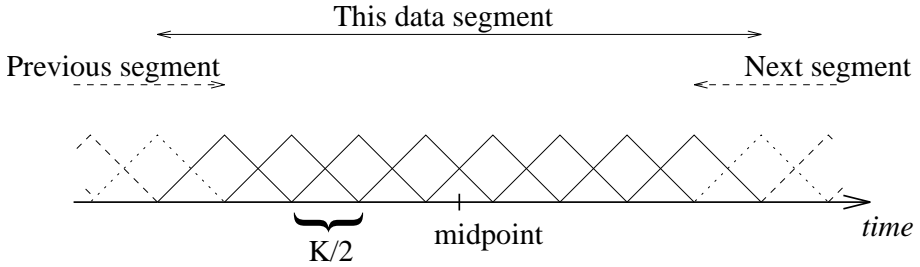


Figure 1.7: Example of use of overlapping time intervals for averaging of spectra. Eight overlapping triangular windows (solid triangles) of length K are applied to the data, the PSD of each corresponding time interval is calculated, and the average is constructed. This is then taken as the power spectrum for the data period denoted “This data segment”, which is time tagged by its midpoint. We then proceed to the next segment, whose first triangular windows are shown dashed.

we want to estimate. With this view, “irregularities” in the spectrum are not regarded as “noise” but as physical signals to be explained. A good discussion about conscious and tacit assumptions on stationarity in spectral analysis of space plasma waves is given by [Rönmark \[1990\]](#).

For the analysis of deterministic data with a strong component of high frequency random noise superimposed, averaging may be useful. If the deterministic signal is stationary under the length of the analysis, averaging can reduce the effect of added noise.

Most textbooks on spectral analysis take the statistical, random process approach [e.g. [Bendat and Piersol, 1971](#); [Brockwell and Davis, 1987](#); [Kay, 1988](#)]. A basic assumption in these works is that the signal is a sample of a stationary random process. In contrast, stationarity is not emphasised in modern texts on wavelet analysis. This is partly motivated by the ability of wavelets, localised in time as well as frequency, to model non-stationary phenomena (see Section 1.4).

1.3.5 Overlapping Intervals

In the discussion of averaging in Section 1.3.4 above, we took a time series of L points, divided it into segments of K points each, and applied a data window to each segment. Windowing implies giving the data points that are multiplied by the flanks of the data window a low statistical weight in the final result. It is therefore reasonable to use overlapping windows, as illustrated in Figure 1.7. An overlap of 50% is a natural choice, particularly for the triangular window. Except for the first data points of the first segment and the last points of the last segment, each data point will be used in two data windows, and hence in two PSD estimates before averaging. If the weighting of the point is w_1 in one of the windows, it will in the case of triangular windows be $1 - w_1$ in the other window, giving a total statistical weight of 1 for all data points.

In the case of triangular windows and 50% overlaps, the normalised standard deviation in the PSD will not be $1/P$ as suggested by equation 1.28, but [[Welch, 1967](#)]

$$\epsilon' \approx 1.2/P \quad (1.30)$$

The other commonly used windows 1.24–1.26 give ϵ values of the same order for 50% overlap. The factor of 1.2 may seem discouraging, but one should note that the P in equation 1.30 is not the same as in 1.28, since overlapping makes it possible to squeeze in more windows in a certain amount of time. For a given record which is separated into P non-overlapping subintervals of length K , we can squeeze in $P' = 2P - 1$ windows of length K if we allow a 50% overlap. Hence, P in 1.30 should be replaced by $2P - 1$ if we want to compare the performance of 50% overlapping triangular and non-overlapping rectangular windows on the same data record. Already for $P = 2$, giving $P' = 3$, we find that the ratio ϵ'/ϵ goes down to 0.8. For $P = 5$ ($P' = 9$), this ratio is 0.67. Hence, the use of overlapping windowed data increases the quality of the output in some sense, at the expense of more computation.

One may note that overlapping will never do any harm except for increasing the computation time. In signal processing literature, there is sometimes skepticism against the use of overlapping, which is justified if you want to construct computationally efficient routines. For the practising physicist, who is eager to extract as good information as possible from his/her data, it is often advisable to sacrifice some computational efficiency for optimal use of the data.

We have here discussed overlapping of intervals for which the PSDs are averaged. Another possibility in WFT analysis is to allow overlap of two consecutive data intervals for which averaged PSDs are calculated. By letting this overlap be almost complete, i.e. just skipping one data point between total PSD estimates, one can give the time-frequency analysis an apparent time resolution similar to the time resolution of the original time series. This gives a smoother appearance of a time-frequency plot, which would remove the square pattern from a display like Figure 1.3(b). No effective time resolution is gained, as this is governed by the uncertainty relation 1.29. If using this smoothing method, one should be aware that there may be different weighting of different data points unless the overlapping is done cleverly. We can see this from Figure 1.7, where this kind of smoothing would correspond to letting “previous segment” and the “next segment” move in toward the midpoint. This could result in some data points being used in two PSD calculations, others in three, and others in four PSD calculations.

1.3.6 Zero-Padding

Let us assume that we have a time series record of N samples, which we increase to length $2N$ by adding N zeroes. What happens to the Fourier transform? From the DFT definition 1.14, it is obvious that we get twice as many Fourier coefficients $\tilde{u}[n]$ as before. Also, it can be seen that coefficient number n in the length N expansion must be equal to coefficient number $2n$ in the new length $2N$ expansion, as $u[j]$ is zero for $j > N$. This illustrates that zero-padding, which is the process of lengthening a time series by adding zeroes at its end, makes no essential change to the shape of a spectrum [Marple, 1987]. At first, it may seem that zero-padding increases the frequency resolution, as the number of Fourier coefficients increase. However, the amount of information on the real physical signal is obviously the same with or without zero-padding, so the increased frequency resolution must be spurious. For the data from a given time interval, the frequency resolution will always be limited by equation 1.22.

An illustration of the spurious increase in frequency resolution is given in Figure 1.8. The power spectrum of a test signal consisting of two equal-amplitude sinusoids at fre-

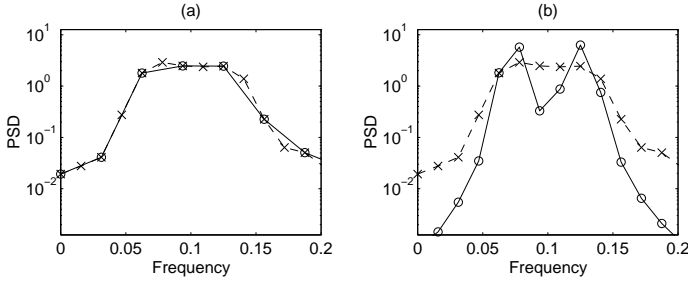


Figure 1.8: Effect of zero-padding. PSDs of a test signal consisting of two sinusoids at frequencies 0.075 and 0.125, in units of the sampling frequency. Dashed lines with crosses (in both panels): 32 data points, 32 zeroes padded. Solid lines with circles (in both panels): no zero-padding, (a) 32 data points, (b) 64 data points.

quencies 0.075 and 0.125 in units of the sampling frequency is calculated from 32 points without zero-padding (solid line in (a)), from 64 points without zero-padding (solid curve in (b)), and from 32 data points with 32 zeroes padded (dashed line in both panels). In (a), we can see that the PSDs of the signals with and without zero-padding coincide at the frequencies where both are evaluated. In (b), the PSD of the zero-padded signal is evaluated at the same frequencies as the 64-point signal, but it cannot resolve the two spectral peaks.

In general, if we pad a signal record of N data points with M zeroes, the Fourier coefficients of the resulting record are given by substituting $M + N$ for N in equation 1.14, where $u[n] = 0$ for $N < j < N + M$. Parseval's relation then takes the form

$$\frac{1}{N + M} \sum_{j=0}^{N-1} u^2[j] = \sum_{n=0}^{N+M-1} |\tilde{u}[n]|^2 \quad (1.31)$$

However, we do not want the padded zeroes to change the physics of the situation as expressed by the PSD estimate. By considering Parseval's relation, we find that the PSD expression 1.19 should be multiplied by a factor $(N + M)^2/M^2$, so that

$$S_u[n] = 2 \frac{(N + M)^2}{N} |\tilde{u}[n]|^2 / f_s \quad (1.32)$$

is the proper expression for the PSD from a zero-padded signal, evaluated at frequencies

$$f_n = \frac{n}{N + M} f_s, \quad n = 0, 1, \dots, N + M - 1 \quad (1.33)$$

If combined with windowing, zero-padding should obviously be applied after the windowing. Otherwise, we would not get the smoothness at the ends of the non-zero part of the time series which is one of the aims with windowing. Also, the correction factor 1.27 for the power loss in the windowing process would be erroneous if the non-zero data occupied only part of the extent of the window. When combined with extra overlapping discussed in Section 1.3.5 above, it results in smooth time-frequency spectra without the sharply defined squares of Figure 1.3.

Since zero-padding does not provide any new information, just a smoothing of the spectrum, it is not very useful in most cases. However, in cases where the length of the time series to be transformed by an FFT algorithm is not an integer power of two, zero-padding up to the next such number may actually speed up the FFT calculation.

Zero-padding is not the only way of “cheating” by evaluating the signal at more frequencies than the data really allow. Another variant is by what may possibly be called “continuous Fourier transform”, where one evaluates the DFT defined by 1.14 at arbitrary real values of $n < N/2$, not just integers. This scheme is rarely used in Fourier analysis, although it is common in wavelet applications (Section 1.4.3).

1.3.7 Welch Method for Time-Dependent PSD Estimation

Applying the techniques above, we get an algorithm for estimation of the time-dependent PSD. Having all the building blocks at hand, we summarise the resulting method below. This technique is due to and very well described by *Welch* [1967].

1. Divide the total time series

$$u[j], \quad j = 0, 1, \dots, N - 1$$

of N samples into M shorter intervals

$$u_{ml}[j], \quad j = p, p + 1, \dots, p + L - 1, \quad m = 0, 1, \dots, M - 1$$

of L samples each. The time resolution of the time-frequency analysis will be L/f_s . If there is an overlap of r points, we have $p = m L - r$.

2. Divide each of the intervals of length L into P segments

$$u_{ml}[j], \quad j = q, q + 1, \dots, q + K - 1, \quad l = 0, 1, \dots, P - 1$$

of length K . If these intervals overlap by s points, $q = p + l(K - s)$. A good choice is $s = K/2$.

3. Multiply each data segment term by term by a window function $w[j - q]$, $j = q, q + 1, \dots, q + K - 1$.
4. Calculate the DFT

$$\tilde{u}_{ml}[n] = \frac{1}{N} \sum_{j=q}^{q+K+Z-1} w[j - q] u[j] \exp(2\pi i n(j - q)/K), \quad n = 0, 1, \dots, K/2,$$

of the windowed time series, preferably using the FFT algorithm.

5. Calculate the PSD estimate, corrected for the windowing:

$$S_{ml}[n] = \frac{2K}{f_s W_{ss}} |\tilde{u}_{ml}[n]|^2$$

6. Average over the P short segments to get a PSD estimate with better variance:

$$S_m[n] = \frac{2KP}{f_s W_{ss}} \sum_{l=0}^{P-1} |\tilde{u}_{ml}[n]|^2$$

7. $S_m[n]$ is our resulting time dependent spectrum, evaluated at frequencies

$$f_n = \frac{n}{K} f_s$$

The midpoints of the time intervals are

$$t_m = t_0 + (m + 1/2) \frac{L - r}{f_s}$$

This scheme is frequently used in applications of spectral analysis, and is conveniently implemented in many software packages.

If desired for some reason, zero-padding could be put in after step 3, in which case a correction for the padding should be included in step 5. If the overlaps r and s are put equal, all data points except those at the very ends of the total time series will be used equally much in the analysis.

1.4 Wavelet Techniques

Wavelet analysis is a very rich field of techniques useful for many different applications, for example data analysis, theoretical electromagnetics, and data compression, as can be seen in any of the many texts on the subject [e.g. [Daubechies, 1990](#); [Kaiser, 1994](#); [Strang and Nguyen, 1996](#)]. We will here take a narrow-minded approach, only considering wavelet methods as an alternative to the Fourier methods for spectral estimations. Discussions of this aspect of wavelets can be found in the literature on applications to space data [e.g. [Holter, 1995](#); [Lagoutte et al., 1992](#)].

1.4.1 Morlet Wavelets

In Section 1.2.6, we found that the time-frequency plane may in principle be partitioned in many different ways. We now turn to the question of how to actually implement a partition of the type shown in panel (d) of Figure 1.2. The answer lies in wavelet analysis, which is unlike the traditional Fourier techniques in that it intrinsically relies on a time-frequency approach. The basic idea of wavelet analysis is to expand a signal in basis functions which are localised in time as well as frequency, so that they have the character of wave packets. This places wavelet methods somewhere between the Fourier techniques, where the basis functions $\exp(-2\pi i f t)$ are sharp in frequency but completely spread out in time, and the pure time series representation, which offers perfect localisation in time but includes all frequencies³.

³In practice, the basis functions of Fourier analysis are in fact localised in time by the data window, and the time series is localised in frequency by its finite sampling rate. However, all Fourier basis functions (all frequencies) are localised in the same time interval, which is not the case in wavelet analysis.

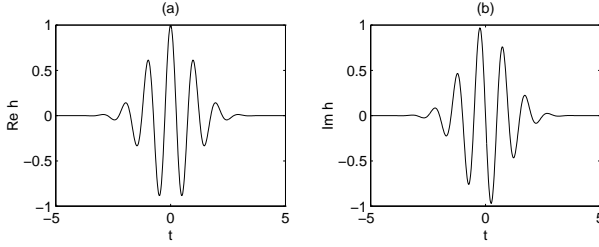


Figure 1.9: The Morlet mother wavelet, defined by equation 1.34, with $\omega_0 = 2\pi$. (a) Real part. (b) Imaginary part.

From a physical viewpoint, localisation in time as well as in frequency is an attractive perspective for the analysis of non-stationary signals. In particular, a basis consisting of localised packets of sinusoidal waves is appealing, as sinusoidal waves are the eigenmodes of a plasma. For the envelope of the wave packet, a Gaussian is a natural choice. As is well known from quantum mechanics, a Gaussian wave packet optimises localisation in both time and frequency as it is the only wave packet for which we get a ‘=’ rather than a ‘ \geq ’ in the uncertainty principle $\delta f \delta t \geq 1$ [Landau and Lifshitz, 1977, p. 48]. This leads us naturally to the Morlet wavelet

$$h(t) = \exp(-t^2/2) \exp(-i\omega_0 t) \quad (1.34)$$

As ω_0 is a free parameter, equation 1.34 defines a family of functions. The value of ω_0 determines the number of oscillations in a wave packet, and has to be sufficiently large for 1.34 to be useful as a wavelet, due to problems with non-vanishing mean value of the real part. For $\omega_0 = 2\pi$, which we will use here for reasons to be seen later (equation 1.36), this error is negligible in practice. In space plasma applications, our choice $\omega_0 = 2\pi$ has sometimes been used [e.g. Dudok de Wit et al., 1995], although $\omega_0 = 5$ seems more common [e.g. Holter, 1995; Lagoutte et al., 1992], probably because it is close to the value $\pi\sqrt{2/\ln 2} \approx 5.34$ originally used by Morlet et al. [1982]. For Morlet’s ω_0 , the amplitude decreases to half its maximum in one period of the wave.

Many other wavelet families than the Morlets are possible, and several classes can be found in any text on wavelet methods. For our purpose, which is the study of spectral properties of non-stationary time series, we take the view that the Morlet wavelet, with its clear physical interpretation as a modulated sinusoidal oscillation and good properties of localisation in frequency as well as in time, is the natural choice.

By stretching and translation of a wavelet like 1.34, called a mother wavelet, we can get a whole set of wavelets of the same shape, known as daughter wavelets. It is customary to denote the stretching and translation by two parameters a and τ known as scale (or dilation) and translation, respectively. The daughter wavelets $h_{a\tau}(t)$ are then written as

$$h_{a\tau}(t) = \frac{1}{\sqrt{a}} h\left(\frac{t - \tau}{a}\right) \quad (1.35)$$

The concept of scale is used here instead of the concept of frequency, which most often is associated with sinusoidal functions. However, for the specific case of the Morlet wavelets,

which are based on sinusoidal functions, it is reasonable to replace the scale a by the frequency $f = 1/a$, so we write the daughter wavelets as

$$h_{f\tau}(t) = \sqrt{f} h(f(t - \tau)) = \sqrt{f} \exp\left(-\frac{f^2(t - \tau)^2}{2}\right) \exp(-2\pi i f(t - \tau)) \quad (1.36)$$

This explains our choice $\omega_0 = 2\pi$: the Morlet wavelets become Gaussian envelopes of a carrier wave with frequency f .

The transformation above ensures that all daughter wavelets will look like their mother wavelet. Irrespective of f and τ , equally many periods of the oscillation will fit into the packet.

We can now define the Morlet wavelet transform (MWT) of the signal $u(t)$ by

$$C(\tau, f) = \int u(t) h_{f\tau}^*(t) dt \quad (1.37)$$

In principle, the limits of integration should be $\pm\infty$. However, as the wavelets $h_{f\tau}(t)$ are localised in time, little error is introduced by integrating over a finite time interval. Finally, by going from integrals to sums in a fashion similar to how we introduced the discrete Fourier transform in Section 1.2.4, we make possible the practical evaluation of wavelet coefficients.

1.4.2 Wavelets and Fourier Methods

The Morlet wavelets 1.36 can be written on the form

$$h_{f\tau}(t) = A w_\tau(t, f) \exp(-2\pi i f t) \quad (1.38)$$

where

$$w_\tau(t, f) = \exp\left(-\frac{f^2(t - \tau)^2}{2}\right) \quad (1.39)$$

and

$$A = \sqrt{f} \exp(2\pi i f \tau) \quad (1.40)$$

Hence, the Morlet wavelet transform 1.37 may be written as

$$C(\tau, f) = A^* \int w_\tau(t, f) u(t) \exp(2\pi i f t) dt \quad (1.41)$$

When we discussed the use of window functions for Fourier methods in Section 1.3.3, we only considered sampled time series. However, for a continuous signal $u(t)$ to which we have applied a window function centred at $t = \tau$, denoted $W_\tau(t)$, the Fourier integral 1.3 is

$$\tilde{u}_\tau(f) = \frac{1}{T} \int W_\tau(t) u(t) \exp(2\pi i f t) dt \quad (1.42)$$

Now assume the window function is a Gaussian. Apart from the factors in front of the integrals, the only difference between equations 1.41 and 1.42 then is that the window function $w_\tau(t, f)$ in 1.41 depends on frequency as well as on time, while the window

$W_\tau(t)$ in 1.42 is the same for all frequencies. However, for any frequency f_0 , we can always choose the Gaussian window so that $W_\tau(t) = w_\tau(t, f_0)$. At this frequency, it is possible to interpret the Morlet wavelet transform as a Fourier transform with a Gaussian window.

This result is useful for our understanding of wavelet methods. It enables us to apply many results and methods of Fourier analysis also to the wavelet transforms. We list a few of them here:

1. PSD estimation. The similarity of equations 1.41 and 1.42 indicates how to obtain a PSD estimate with physically meaningful normalisation with wavelet methods (equation 1.43).
2. Phase. We can also construct a wavelet phase spectrum corresponding to the Fourier phase spectrum 1.11.
3. Detrending. Frequency components below the lowest resolvable will affect the wavelet methods as well as the Fourier methods, so detrending (Section 1.3.2) may be useful.
4. Windowing is obviously inherent in the wavelet transform.
5. Averaging. For a random signal, the wavelet based PSD will also show statistical fluctuations, which in principle may be quenched by averaging in time, at the expense of temporal resolution⁴ (Section 1.3.4).
6. Zero-padding is not a useful technique for wavelet analysis, since the zeroes cannot be padded after the windowing (Section 1.3.6) in a wavelet transform, where the window is implicit in the basis wavelet itself.

For the PSD, by comparing the Morlet wavelet transform 1.41 to the Gaussian window Fourier transform 1.42, using the definition 1.8 of the power spectral density and correcting for the window (envelope) by the factor W_{ss} defined by 1.27, we conclude that the PSD definition for the MWT which gives PSD values equal to what we find for the corresponding Gaussian windowed DFT estimate is

$$S_u^{\text{MWT}}(t, f) = \frac{2}{\sqrt{\pi}} |C(t, f)|^2 \quad (1.43)$$

With this definition, the PSD values derived from Fourier analysis and wavelet transforms are comparable, in the sense that the PSD estimates for the shaded squares in panels (c) and (d) of Figure 1.2 are equal.

1.4.3 Continuous Wavelet Transform (CWT)

The wavelet transform is based on wave packets where the relation between frequency and packet width is constant. Hence, it naturally provides a means to implement the octave band analysis suggested by the division of the time-frequency plane in panel (d) of

⁴This somewhat unconventional viewpoint is further discussed below in Section 1.4.4.

Figure 1.2. As the t - f plane cannot be divided into more rectangles than half the number of samples in the time series, evaluation of one wavelet coefficient for each of these rectangles should give a complete description of the PSD.

In practical applications of wavelet transforms to spectral problems, one often evaluates more wavelet coefficients than is actually needed. This is known as doing a continuous wavelet transform (CWT), where the word continuous signifies that we evaluate 1.37 at freely chosen f and t . However, one should note that the time-frequency resolution of the CWT will be as depicted in panel (d) of Figure 1.2, even though the CWT may be evaluated at many points. We cannot get around the restrictions of the uncertainty relation 1.22.

⇒ 1.1

In the same way, one may also define a continuous Fourier transform, which simply amounts to extending the definition 1.14 to non-integer values of $n = f/f_s$, and letting the data windows overlap arbitrarily much. For an inverse transform back to the time domain, the non-integer values of n are of course entirely superfluous: no new information is gained by evaluating the DFT or the MWT for more time-frequency locations than $N/2$, and hence these extra coefficients never enter in an inverse transform. In fact, this sort of continuous Fourier transform is never used in practice, as it cannot be implemented with the FFT scheme. If one desires a smoother evaluation of the PSD, zero padding (Section 1.3.6) is used instead.

1.4.4 Comparing WFT and MWT: an Example

An example of a comparison of wavelet and Fourier methods for PSD calculation is shown in Figure 1.10. Panel (a) shows a time series of the electric field sampled at 32 000 samples/s by the wave instrument on the Freja satellite. The data shows lower hybrid waves of a bursty or modulated character [Eriksson *et al.*, 1994]. Panels (b)–(e) display different time-frequency spectral representations of the signal in (a). The total number of samples is $N = 13000$. The methods used for the spectral estimation all have different time resolution, and only the time interval for which all yield results is shown. For all the spectra, the number of displayed bins in the time-frequency plane are much higher than the number $N/2$ which defines the real information content in the spectra (Section 1.2.6), giving a smoother variation in time and/or frequency.

Panel (b) shows a Fourier time-frequency analysis based on averages of eight 64-point DFTs overlapping by 50%. The effective time resolution therefore is 512 samples or 16 ms, but the PSDs shown are separated only by one sample, giving a smoother time variation in the plot. As discussed in Section 1.3.4, the use of the averaging method is motivated if we interpret the signal as an almost stationary random process, whose slowly varying statistical parameters we want to estimate. The result of the averaging is to reduce the variations between consecutive spectra, and the spectrum is very clean. On the other hand, the resolution in the time-frequency plane goes down by a factor of eight as indicated by equation 1.29, and little trace of the bursty nature of the time series in panel (a) is retained.

In panel (c), we show a Fourier-based PSD estimate with the same time resolution (512 samples) as in (a). Here, no averaging has been used, and the effect of random fluctuations is therefore stronger. This results in larger variations within spectra, as is seen by the many horizontal structures in the plot. The frequency resolution in (c) is much better than in (b), and splittings of the spectral peak around 4 kHz can be clearly seen. On the other

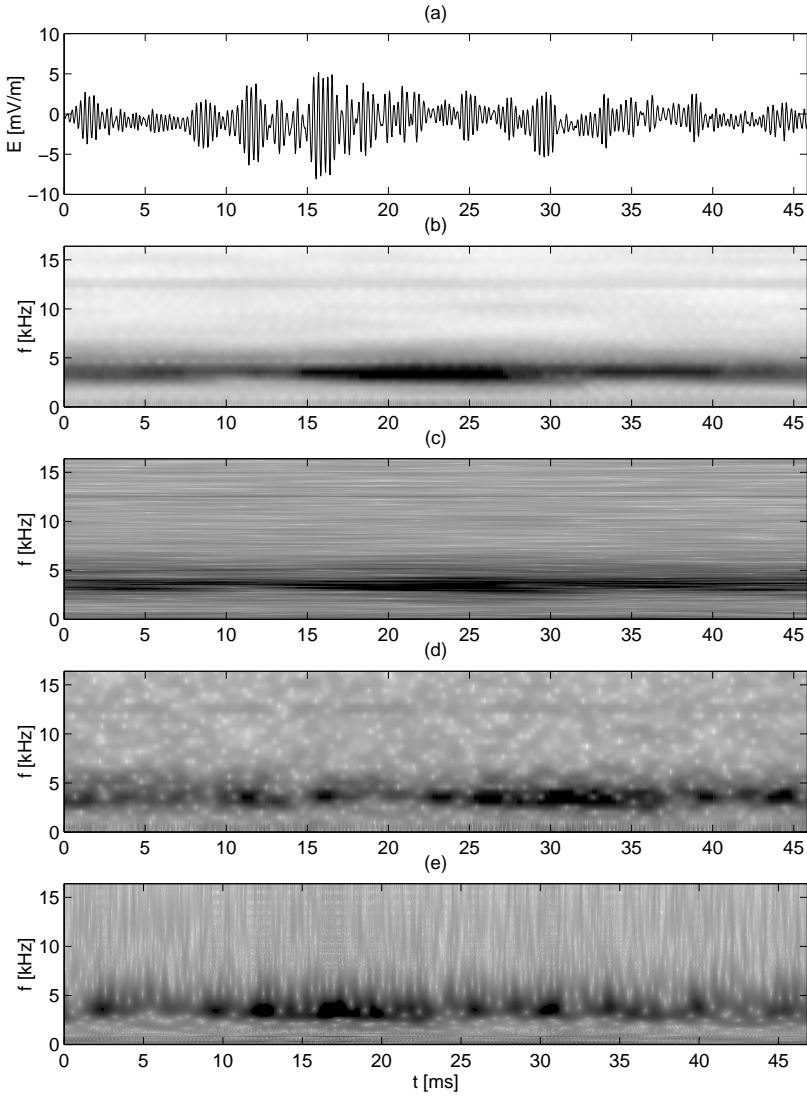


Figure 1.10: A time series (a) with time-frequency spectra obtained by different methods: (b) Averaged DFT analysis, 64 points, 8 averages. (c) High frequency-resolution DFT analysis, 512 points. (d) High time-resolution DFT analysis, 64 points. (e) Morlet wavelet analysis.

hand, little time variation can be discerned. The modulated nature of the time series is represented by frequency splittings rather than as time variations, as discussed in the text on page 13 and exemplified in Figure 1.3.

Panel (d) displays the result of a 64-point Fourier analysis without averaging. Hence, the frequency resolution is as for panel (b), while the time resolution is much better. Variations of the amplitude of the time series are now reflected in the time domain, as variations of the PSD with time. While the structures in (c) were almost horizontal, they are here fairly isotropic.

Finally, panel (e) shows the PSD based on Morlet wavelet analysis. The modulational pattern of the signal is represented as time variations of the PSD, as in panel (d). The structures to be seen in this plot are horizontal at the lowest frequencies (few kHz), and vertical at higher frequencies (above 5 kHz). This behaviour reflects that in the wavelet analysis, the time-frequency plane is partitioned as in panel (d) of Figure 1.2. This results in characteristic flame-like patterns in the plot: like the flames in a fire, the horizontal scale sizes are smaller at the top than in the base. These flame patterns are analogous to the large variance in PSD estimates by DFT methods, discussed in Section 1.3.4. They may very well reflect the random variations of a stationary process, in which case it may be justified to remove them by averaging in time. This is not normally discussed in wavelet texts, as one of the fundamental virtues of wavelet analysis is their usefulness for studying non-stationary phenomena. However, as such flames will turn up even if analysing a stationary process like Gaussian random noise, one should be careful not to put too much physical meaning in their appearance, and sometimes one may consider averaging in time.

All the Fourier-based plots (b, c, d) more or less clearly show a narrow spectral line near 12 kHz. This component cannot be seen in the wavelet plot (e). The reason for this is obvious from a glance at Figure 1.2. The partitioning of the time-frequency plane in wavelet analysis is such that all spectral details in the upper half plane (above $f_s/4$) are completely lost. In this particular case, the origin of the 12 kHz line is purely instrumental, so little harm is done by using the wavelet analysis here, but this is certainly not the general case.

Even though the plots in Figure 1.10 show continuous transforms evaluated at many more points than $N/2$, we see that the partitions of the t - f plane (c) and (d) in Figure 1.2 are inherent to the Fourier and wavelet methods, respectively, looking at the structuring of spectral detail. Evaluating the PSDs at very many points cannot hide these fundamental aspects of the methods, since there is no extra information gained by displaying more than $N/2$ points in the t - f plane. There is no way to fool relation 1.22.

1.4.5 Implementations of the Wavelet Transform

For some wavelet families, there are very efficient computation schemes. For others, including the Morlet wavelets, there are no algorithms as fast as the FFT routine of Fourier analysis. Discussions on algorithms can be found in standard texts, but fortunately these are already implemented in many commercial software packages, including the commonly used IDL and Matlab (as a separate toolbox), so you should not have to write them all yourself. Normalisation can be a problem when using such routines, since it is not always simple to deduce from the documentation what normalisation conventions are used, but the correspondence between Fourier and wavelet methods discussed in Section 1.4.2 can be used to derive formulas like 1.43 and for checking normalisation.

1.4.6 Wavelet Packet Analysis

As the wavelet analysis discussed above is based on wavelets with a similar number of wave periods in the wave packet for all carrier frequencies, it is naturally associated with the time-frequency partitioning shown in Figure 1.2d on page 12. However, it may very well be that the signal consists of a few transient pulses at long time scales, combined with stable narrow-band emissions at high frequencies. In that case, a partitioning corresponding to turning Figure 1.2d upside down could be advantageous. One could also think of a signal whose natural representation is in some completely unsymmetric division of the t - f plane, like what we get if we interchange the dashed rectangular areas in panels (c) and (d) of Figure 1.2. In terms of a Morlet wavelet framework, this would correspond to using basis wavelets 1.36 with varying values of ω_0 , i.e. with a varying number of wave periods fitting into a wave packet.

Schemes for wavelet analysis along these lines are known as wavelet packet analysis [e.g. *Strang and Nguyen, 1996*]. This includes automatic search for a best basis, in which as much signal energy as possible is concentrated into as few t - f -boxes as possible. This has mainly found applications for purposes like image compression, but clearly holds an interesting potential for analysis of space plasma time series. A brief treatment of applications to geophysics, with references to theoretical work, is included in the review article by *Kumar and Foufoula-Georgiou [1997]*.

1.5 Spectral Analysis of Multiple Signals

1.5.1 Cross-Spectral Analysis

Cross-spectral analysis of two time series is a means for looking at the relations between individual spectral components of two data records, usually sampled simultaneously. We define the cross-spectral density (CSD) of the two time series $u[j]$ and $v[j]$, with DFTs $\tilde{u}[n]$ and $\tilde{v}[n]$, respectively, by

$$G_{uv}[n] = \frac{2N}{f_s} \tilde{u}^*[n] \tilde{v}[n] \quad (1.44)$$

where $*$ indicates complex conjugation. With this normalisation, the PSD becomes a special case of the cross-spectral density,

$$S_u[n] = G_{uu}[n] \quad (1.45)$$

In this chapter, we have confined ourselves to the case when u and v are signals of the same kind sampled at different locations, for example by different satellites. Cross-spectral methods are also very useful for the analysis of several different signals sampled at one location in space. One then forms a spectral matrix consisting of the cross-spectra of the signals, whose diagonal elements are the PSDs of the signals. This matrix can then be used for the study of wave properties like wave normal directions and field polarisation. A nice example is the classical study by *Means [1972]*, and one may note that the STAFF instrument on the Cluster mission implements on-board spectral matrix analysis of five field components [*Cornilleau-Wehrlin et al., 1997*]. The multi-spacecraft filtering technique presented in Chapter 3 is a further development, where signals from each of several measurement locations are included.

Unlike the PSD, the cross-spectral density 1.44 of real signals will in general be complex-valued. One introduces the coincident spectral density C_{uv} and the quadrature spectral density Q_{uv} as the real and imaginary parts of G_{uv} , respectively:

$$G_{uv}[n] = C_{uv}[n] + i Q_{uv}[n] \quad (1.46)$$

The phase of the cross-spectral density is defined by

$$G_{uv}[n] = |G_{uv}[n]| \exp(i\varphi_{uv}[n]) \quad (1.47)$$

so that

$$\tan \varphi_{uv}[n] = \frac{Q_{uv}[n]}{C_{uv}[n]} \quad (1.48)$$

By considering the definition 1.11 of the phase spectrum of a single signal, it follows that

$$G_{uv}[n] = |G_{uv}[n]| \exp(i\varphi_{uv}[n]) = |\tilde{u}[n]| |\tilde{v}[n]| \exp(i\{\varphi_v[n] - \varphi_u[n]\}) \quad (1.49)$$

Therefore, the phase of the cross-spectrum is the phase difference between the two signals.

To see the physics of the coincident and quadrature spectral densities defined by 1.46 above, we note that $C_{uv} = |G_{uv}| \cos \varphi_{uv}$ and $Q_{uv} = |G_{uv}| \sin \varphi_{uv}$. Thus C_{uv} and Q_{uv} represent the in-phase and out-of-phase contributions from the two signals to the total CSD. A simple example is the case where v is the voltage variation (in volts) in some system and u is the current in the same system (in ampères). In this case, C_{uv} and Q_{uv} are the spectra of effective and reactive power, respectively, in watts per hertz⁵. The impedance spectrum is $Z[n] = G_{uv}[n]/S_v[n]$ (in ohms), where the real part $R[n] = C_{uv}[n]/S_v[n]$ is the resistance spectrum, while the imaginary part $X[n] = Q_{uv}[n]/S_v[n]$ is the reactance spectrum, with contributions from the inductive and capacitive reactance spectra.

The implementation of cross-spectral methods by Fourier methods is straightforward. In addition, high-level data processing packages like IDL and Matlab include functions for phase spectra, coherence and other cross-spectral quantities. The application of cross-spectral analysis to space plasma data will be exemplified in Section 1.5.3. Before that, we must have a look at how to implement the definitions of the cross-spectral quantities and interpret the results.

1.5.2 Averaging and Coherence

From the definitions above, it is clear that as soon as we have the Fourier transforms of two signals, we can calculate a phase spectrum. However, just doing so and using the result to find for example a wavelength spectrum is not always a good idea. First, if we just consider the phase spectrum, there is obviously no information on signal strength in it, so frequencies with a clear signal are treated in the same way as frequencies only containing instrumental noise. Therefore, it is necessary to compare the phase spectrum to the CSD magnitude or to the PSDs of the signals, to see that there really is significant signal strength at the frequencies we are considering.

Second, how do we know that there really is any relation between the two signals at the frequency we are considering? The definitions above allow us the possibility of always

⁵Conventionally volt-ampères per hertz for the reactive power.

calculating the phase spectrum, irrespective of any causal relation between the signals. In the case of a stationary process, one can define a coherence spectrum,

$$\gamma_{uv}^2[n] = \frac{|G_{uv}[n]|^2}{S_u[n] S_v[n]} \quad (1.50)$$

which is useful in this context. At first this may seem strange: from the definition of G_{uv} 1.44 and the relation $S_u = G_{uu}$ 1.45, one may conclude that $\gamma_{uv}^2[n] \equiv 1$, regardless of the signals u and v . This is certainly true, unless the spectral estimates G_{uv} , S_u , and S_v are calculated using the averaging method of Section 1.3.4. If averaging is included in the calculation of the spectral densities, the value of γ_{uv}^2 can end up anywhere between zero and one. As the denominator in 1.50 is positive, the coherence is essentially determined by the constancy of the phase of the numerator, i.e. by the stability of the phase spectrum $\varphi_{uv}[n]$. If, for a certain frequency f_n , the phase spectrum changes rapidly between the segments included in the averaging process, the averaged G_{uv} for this frequency can take any complex value. In the case of random phase fluctuations and sufficiently many averages, the value of the coherence will come close to zero. The coherence function 1.50 therefore indicates the stability of the phase spectrum estimate. A low value of the coherence at some frequency implies that the phenomenon behind the phase spectrum estimate at this frequency cannot be considered correlated between the two measurement points and stationary in time. Hence, a usual prescription for cross-phase methods is to include averaging in the calculation and only use points which turn out to show values of γ_{uv}^2 higher than some threshold value.

It is important to note that the coherence concept is meaningful only for stationary signals. Still, the phase concept may be useful also for non-stationary signals, for example a wave packet seen only briefly in the spacecraft frame. If we have some reason to believe that there is a relation between two signals u and v , it is perfectly legitimate to examine their cross-spectrum even if the observation time is so short that no averaging and hence no coherence estimate can be done. This is the situation we encountered on page 20, where we discussed averaging. If we interpret our signals as produced by a stationary random process, averaging is useful, and the coherence spectrum is appropriate for determining the quality of the phase spectrum. On the other hand, if we interpret the signals as a unique record of some non-stationary phenomenon, averaging and hence the coherence function are irrelevant concepts. In this case, one must resort to other means for judging the reliability of the phase spectrum. Checking the signal intensity from the PSD or CSD is an obvious approach. It may also help to inspect the phase spectrum itself: if it is fairly smooth, its interpretation is safer than if it is wildly and randomly fluctuating with changing frequency.

For non-stationary signals, it may also be a good idea to use wavelet rather than Fourier methods as a basis for the cross-spectral analysis. We return to this question in Section 1.5.4.

1.5.3 Cross-Spectral Analysis: an Example

As a simple example of the use of cross-spectral analysis in space applications, we show an analysis of Langmuir probe current data from the Freja wave instrument in Figure 1.11. Panel (a) shows the relative variations of the current $\delta I/I$ collected by two

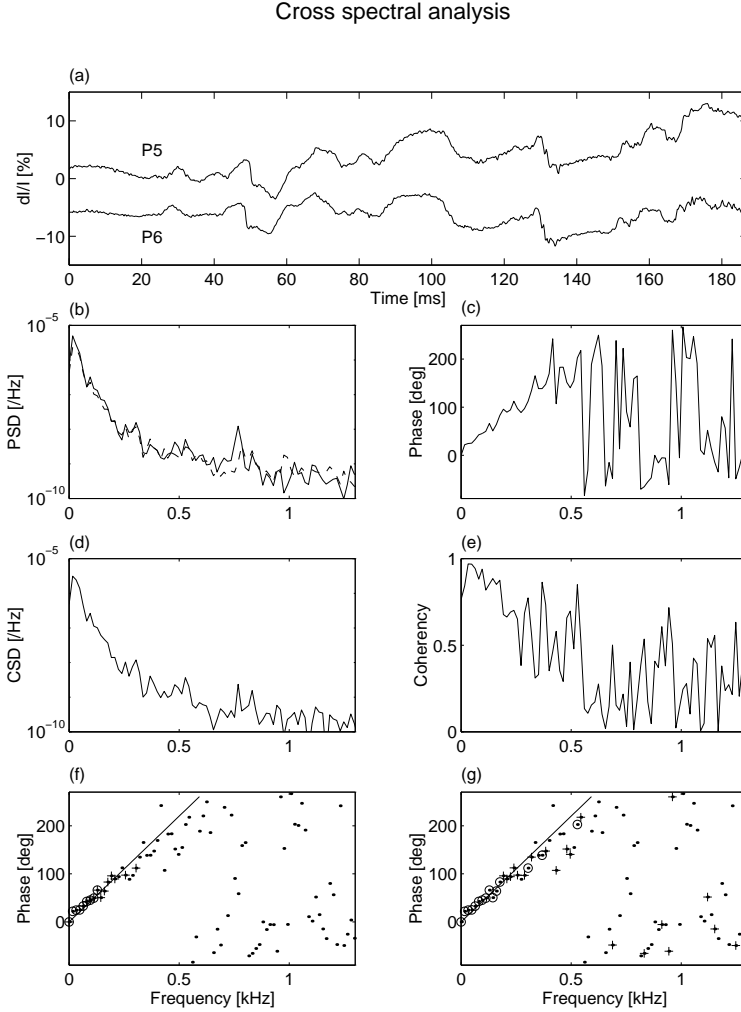


Figure 1.11: Langmuir probe data from the Freja wave instrument illustrating the use of cross-spectral analysis. All spectra calculated using averaging of five 256-point DFTs with 128 point overlap. (a) Time series of probe current fluctuations from two probes, sampled at 4 ksamples/s. (b) Power spectra of the P5 (solid) and P6 (dashed) signals. (c) Relative phase spectrum. (d) Magnitude of cross-spectral density. (e) Coherence spectrum. (f) Relative phase spectrum coded by CSD magnitude. Circles: $|CSD| > 10^{-7} \text{ Hz}^{-1}$. Crosses: $10^{-8} \text{ Hz}^{-1} < |CSD| < 10^{-7} \text{ Hz}^{-1}$. Dots: $|CSD| < 10^{-7} \text{ Hz}^{-1}$. (g) Relative phase spectrum coded by coherence. Circles: $\gamma^2 > 0.75$. Crosses: $0.5 < \gamma^2 < 0.75$. Dots: $\gamma^2 < 0.5$.

positively biased spherical probes P5 and P6 mounted on 5.6 meter booms at opposite sides of the spacecraft. The curves have been shifted along the ordinate in order to separate them. These two time series thus constitute a multipoint measurement in space. The two curves are obviously very similar.

Figure 1.11(b) shows the power spectra of the two $\delta I/I$ signals. As expected, they are very similar, both showing a decay with increasing frequency. The phase and amplitude of the CSD are shown in panels (c) and (d), respectively, and the coherence as defined by 1.50 is plotted in panel (e). As the spectra in (b) are similar to each other, it is not surprising that (d) looks about the same as well. The coherence spectrum in (e) shows a decay with frequency reminiscent of the power spectra, although there are some coherence peaks at higher frequencies. Panels (f) and (g) show the phase spectrum again, this time with different symbols used for plotting values corresponding to different ranges of CSD and γ^2 (see caption).

For multipoint measurements, an important feature of the phase difference is that it provides information on wavelengths. Consider a wave field varying as $\exp(i[kx - \omega t])$. Measurements at two points separated by a distance d will have a phase difference

$$\varphi = kd \quad (1.51)$$

which can be revealed by a cross-spectral analysis. Knowing the separation distance, the wave number k and hence the wavelength of the wave would be known. In the case of Figure 1.11, the points with high CSD and coherence at low frequencies in panels (f) and (g) lines up around a line $\varphi = af = a'\omega$, where $a \approx 440^\circ/\text{kHz} \approx 7.7 \cdot 10^{-3} \text{ rad/Hz}$ which gives $a' \approx 2.4 \cdot 10^{-3} \text{ s}$. Comparing this to the predicted phase difference 1.51 using the known probe separation $d = 11.2 \text{ m}$, we find that this corresponds to a phase velocity

$$v_\phi = \frac{\omega}{k} = \frac{d}{a'} \approx 4.6 \text{ km/s} \quad (1.52)$$

along the line of probe separation in the reference frame of the spacecraft. In this case, the satellite speed was 6.1 km/s and the angle between the probe separation distance and the spacecraft velocity was 40° , so the observed behaviour simply reflects the satellite flying through essentially stationary structures in the plasma. However, the principle for the calculations would be the same also for travelling waves, and the method have been used for several successful measurements of wavelengths in space. For further examples and discussions we refer to the review by *LaBelle and Kintner* [1989], to the application papers by for example *Holmgren and Kintner* [1990] or *Vago et al.* [1992], and to the interesting extension of this method to the estimation of frequency-wavenumber spectra by *Beall et al.* [1982].

1.5.4 Cross-Wavelet Analysis

Cross-spectral analysis does not have to be based on Fourier methods. The Morlet wavelet analysis discussed in Section 1.4 is also possible to use for phase spectrum estimations, which can be advantageous for non-stationary processes. This approach has quite recently been applied to frequency-wavenumber spectral estimation in space plasma data by *Dudok de Wit et al.* [1995] and *Pinçon et al.* [1997].

The advantage of wavelet methods is in fact more pronounced for the study of phase spectra than for the PSD estimations we discussed in Section 1.4. The reason for this is

that while the PSD is a positive definite additive quantity for which averaging is useful, the phase spectrum can have all signs, and the average of a changing phase has no physical meaning. The phase estimates we get from a Fourier implementation of cross-spectral analysis is based upon some fixed time interval, equal for all frequencies. Hence, these estimates are based on a few wave periods for a low frequency wave, but many wave periods for an oscillation at higher frequencies. We thus intrinsically put higher demands on the phase stability of waves at higher frequencies as compared to low-frequency fluctuations. For PSD estimates, this is no fundamental problem. Estimating the power of a changing signal simply gives the mean power in the time interval in question.

To illustrate this, we consider the idealised example in Figure 1.12. One artificial data record (dashed) consists of unit amplitude sinusoids at 0.1 Hz with some added noise, sampled at 1 sample/s. The other record (solid) is similar, but with a 180° phase shift halfway through the interval. The total interval of 192 points is divided into five overlapping 64 point sections, which are Hamming windowed before DFT analysis. The averaged PSDs over these five sections are shown in panel (b). The fact that one of the signals is not stationary and changes its phase during the interval does not change its PSD (solid curve) very much, although there is some broadening. Panels (c) and (d) show the phase and the amplitude of the CSD, and panel (e) the coherence spectrum. It is seen that the coherence at the signal frequency 0.1 Hz is very low, as it must be due to the changing phase of one of the signals. Finally, panels (f) and (g) show the phase spectra based only on the first and last 64 points of data, respectively. These tells us the true story: the phase at 0.1 Hz is 0 at the beginning of the record and 180° at the end.

If the signal to be analysed by cross-spectral methods is non-stationary but still narrow in its frequency content, a Fourier approach may work well. One then divides the time series into records of a length corresponding to a few periods of the dominant component, and can then keep track of the changing phase. However, if the non-stationary signal contains a wide range of frequencies, a Fourier based cross-spectral method is unlikely to provide useful phase information over all the frequency spectrum. A wavelet implementation, using wavelet coefficients 1.37 instead of the DFT in the definition of the cross-spectral density 1.44 is then likely a better choice. This unavoidably means losing spectral detail at high frequencies, as the wavelet decomposition of course gives less frequency resolution at high frequencies than does its Fourier counterpart.

1.6 Parametric Methods

The fundamental idea of these methods, which have names like AR (autoregressive), MA (moving average), and ARMA, are to estimate the power spectrum of signal by estimating the parameters of a model for the time series of the signal. Parametric methods are akin to the Fourier techniques in the sense that their division of the time-frequency plane is similar, but otherwise quite different in spirit. These methods may very well dominate in modern textbooks on spectral analysis [e.g. *Brockwell and Davis, 1987*; *Kay, 1988*; *Marple, 1987*], and sometimes one will hear that these methods gives “better” spectral estimates than classical Fourier methods. It is important to realise that this is true only if the process we are studying really is of the kind assumed in the parametric model. To illustrate this point we will briefly discuss one of the best known and most used approaches, the AR method of spectral estimation, which for a Gaussian random process is equivalent to the

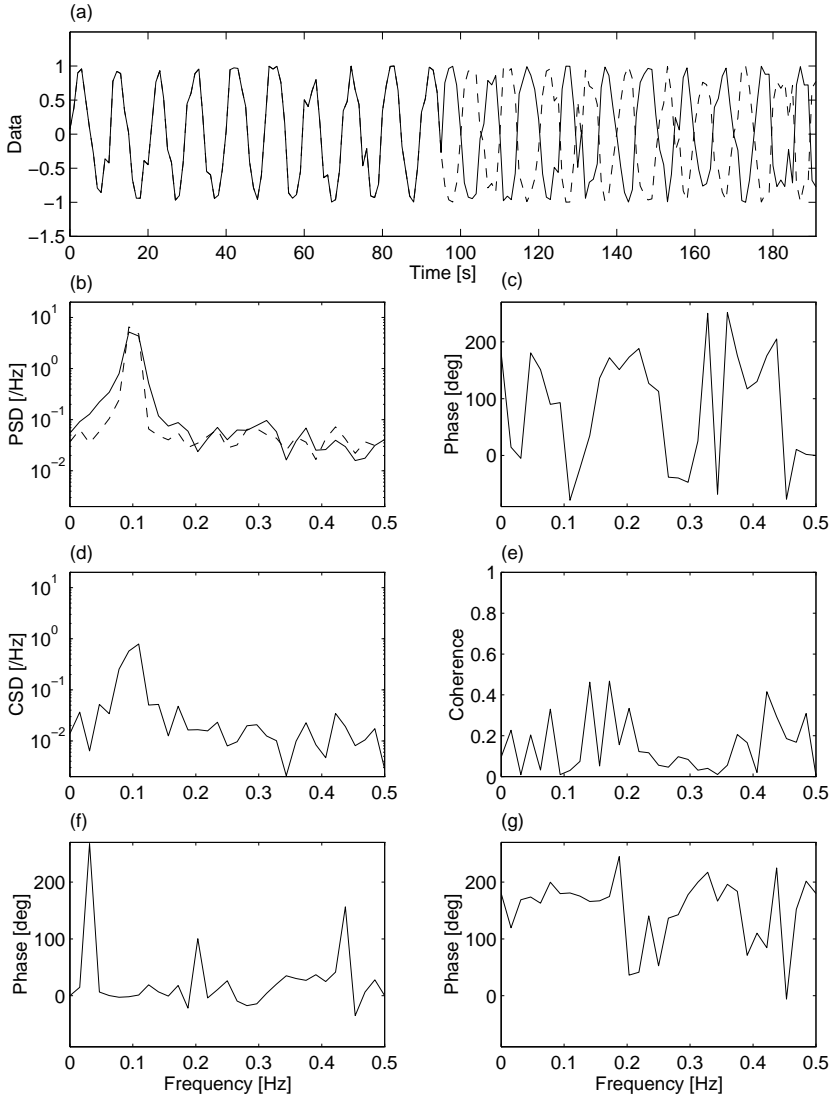


Figure 1.12: Synthetic signals with non-stationary relative phase, and cross-spectral analysis of these. autoregressive spectral methods. (a) Time series. (b) PSDs of the time series. (c) Phase spectrum based on the full record. (d) CSD of the full time series. (e) Coherence spectrum. (f) and (g) Phase spectra based on first and last thirds of the data record, respectively.

maximum entropy method (MEM) [Ables, 1974].

An AR model of order p models the time series as autoregressive, which means that the sample $u[j]$ is a linear function of the preceding samples:

$$u[j] = \sum_{k=0}^p a[k] u[j-k] + \eta[j] \quad (1.53)$$

where $\eta[j]$ represents uncorrelated noise. An AR model therefore is appropriate for a linear dynamical system. As linear systems are completely described by a set of eigenmodes with well defined frequencies, AR spectral estimation is good at representing spectra where the energy is concentrated in sharp peaks. Hence, if we have a situation where we have reason to expect well-defined eigenfrequencies, an AR algorithm is a good choice. For the type of problems facing the physicist studying processes in a turbulent space plasma, this is a rare case. Consider the example in Figure 1.13. A model PSD with a broad maximum superposed on a $1/f$ spectrum,

$$S(f) = \frac{1}{f} + 30 \exp\left(-\left(\frac{f-0.2}{0.05}\right)^2\right) \quad (1.54)$$

was used for creating the time series in panel (a) as described in Section 1.2.3 with random phase spectrum $\phi[n]$. In panel (b), two sinusoids with amplitude 5 and frequencies 0.10 and 0.31 have been added. Panels (c) and (d) shows the model spectrum $S(f)$ (dashed) together with the spectral estimates by the Welch method (averaged and windowed DFT) and an AR method with $p = 10$. The Fourier estimate is the ragged curve which lies almost on the dashed model curve, while the smoother AR estimate mostly lies below. The advantage of the AR method is that it represents the two sinusoids in Panel (d) as narrow lines. However, if we are interested in anything else than these discrete frequencies, the Fourier method estimate clearly is better. The AR estimator tries to represent all spectral features by narrow peaks, including the broad peak around frequency 0.2 and the $1/f$ spectrum at low frequencies. When studying plasma processes in space, we are generally interested in the full spectrum, and the AR method is therefore usually not optimal. The same can be said for the application of other parametric spectral methods, although there of course are exceptions. An example where parametric methods indeed are appropriate is the search for signatures of solar eigenoscillations in the solar wind plasma by Thomson *et al.* [1995]. The solar oscillations are small-amplitude eigenmodes of a system not much perturbed by its environment, like a ringing bell, and hence it is reasonable to expect that the signatures of these oscillations can be described as an AR process. The eigenfrequencies carry information on the sun, and estimating them as exactly as possible is a task not suited for the Fourier and wavelet methods we emphasise in this chapter.

1.7 Final Comments

When doing spectral analysis of time series data from several spacecraft, two fundamental approaches suggest themselves. The first is to apply spectral analysis to the time series from each spacecraft separately, and then compare the spectral data. This is appropriate for the study of how processes on a much shorter length scale are modulated

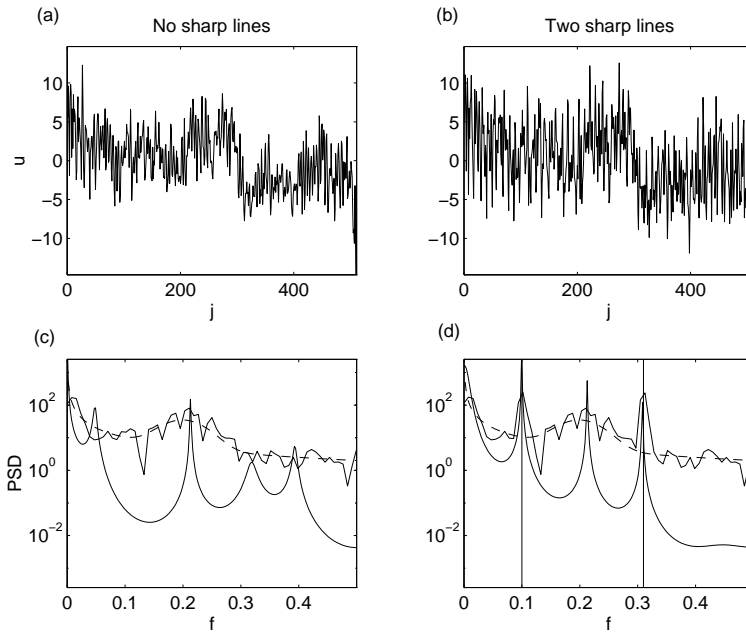


Figure 1.13: Synthetic signal and spectra calculated with Fourier and autoregressive spectral methods. (a) Signal constructed from the spectrum 1.54. (b) Two sinusoids at frequencies 0.10 and 0.31 added. (c) True spectrum of the signal in (a) (dashed), AR estimate (solid, sharply peaked) and DFT estimate (solid, following the dashed line). (d) As (c) but for the signal in (b).

at larger scales. The alternative approach, of an intrinsic multipoint nature, is to do spectral analysis of time series from two or more spacecraft together. The cross-spectral techniques described above in Section 1.5 belong to this class of methods, suitable for studying physical phenomena whose shortest spatial scale is on the order of the spacecraft separation distance. For another prime example of such a method, we refer the reader to the k -filtering technique described in Chapter 3 of this book.

Bibliography

- Ables, J. G., Maximum entropy spectral analysis, *Astron. Astrophys. Suppl. Series.*, **15**, 383–393, 1974.
- Beall, J. M., Kim, Y. C., and Powers, E. J., Estimation of wavenumber and frequency spectra using fixed probe pairs, *J. Appl. Phys.*, **53**, 3933–3940, 1982.
- Bendat, J. S. and Piersol, A. G., *Random data: Analysis and measurement procedures*, John Wiley & sons, 1971.
- Brockwell, P. J. and Davis, R. A., *Time series: theory and methods*, Springer, 1987.
- Brook, D. and Wynne, R. J., *Signal processing*, Edward Arnold, London, 1988.

- Champeney, D. C., *Fourier transforms and their physical applications*, Academic Press, 1973.
- Cornilleau-Wehrin et al., The Cluster spatio-temporal analysis of field fluctuations (STAFF) experiment, *Space Sci. Rev.*, **79**, 107–136, 1997.
- Daubechies, I., The wavelet transform, time-frequency localization and signal analysis, *IEEE Trans. Inform. Theory*, **36**, 961–1005, 1990.
- Dudok de Wit, T., Krasnosel'skikh, V. V., Bale, S. D., Dunlop, M. W., Lühr, H., Schwartz, S. J., and Wolliscroft, L. J. C., Determination of dispersion relations in quasi-stationary plasma turbulence using dual satellite data, *Geophys. Res. Lett.*, **22**, 2653–2656, 1995.
- Eriksson, A. I., Holback, B., Dovner, P. O., Boström, R., Holmgren, G., André, M., Eliasson, L., and Kintner, P. M., Freja observations of correlated small-scale density depletions and enhanced lower hybrid waves, *Geophys. Res. Lett.*, **21**, 1843–1846, 1994.
- Glassmeier, K.-H. and Motschmann, U., Comments on time series analysis, in *Proceedings of Cluster Workshops, Braunschweig, 28–30 Sep. 1994, Toulouse, 16–17 Nov. 1994*, ESA SP-371, pp. 7–14, European Space Agency, Paris, France, 1995.
- Harris, F. J., On the use of windows for harmonic analysis with the discrete Fourier transform, *Proc. IEEE*, **66**, 51–83, 1978.
- Holmgren, G. and Kintner, P. M., Experimental evidence of widespread regions of small-scale plasma irregularities in the magnetosphere, *J. Geophys. Res.*, **95**, 6015, 1990.
- Holter, Ø., Wavelet analysis of time series, in *Proceedings of Cluster Workshops, Braunschweig, 28–30 Sep. 1994, Toulouse, 16–17 Nov. 1994*, ESA SP-371, pp. 43–50, European Space Agency, Paris, France, 1995.
- Jenkins, G. M. and Watts, D. G., *Spectral analysis and its applications*, Holden Day, 1968.
- Kaiser, G., *A friendly guide to wavelets*, Birkhäuser, 1994.
- Kay, S. M., *Modern spectral estimation*, Prentice Hall, Englewood Cliffs, 1988.
- Kumar, P. and Foufoula-Georgiou, E., Wavelet analysis for geophysical applications, *Rev. Geophys.*, **35**, 385–412, 1997.
- LaBelle, J. and Kintner, P. M., The measurement of wavelength in space plasmas, *Rev. Geophys.*, **27**, 495, 1989.
- Lagoutte, D., Cersier, J. C., Plagnaud, J. L., Villain, J. P., and Forget, B., High-latitude ionospheric electrostatic turbulence studied by means of the wavelet transform, *J. Atmos. Terr. Phys.*, **54**, 1283–1293, 1992.
- Landau, L. D. and Lifshitz, E. M., *Quantum Mechanics (Non-Relativistic Theory)*, *Course of Theoretical Physics vol. 3*, 3rd edition, Pergamon, 1977.
- Marple, S. L., *Digital spectral analysis*, Prentice-Hall, Englewood Cliffs, 1987.
- Means, J. D., The use of the three-dimensional covariance matrix in analyzing the polarization properties of plane waves, *J. Geophys. Res.*, **77**, 5551–5559, 1972.
- Morlet, J., Arens, G., Fourgeau, I., and Giard, D., Wave propagation and sampling theory, *Geophysics*, **47**, 203–236, 1982.
- Pinçon, J. L., Kintner, P. M., Schuck, P., and Seyler, C. E., Observation and analysis of lower hybrid solitary structures as rotating eigenmodes, *J. Geophys. Res.*, **102**, 17 283–17 296, 1997.
- Press, W. H., Flannery, B. P., Teukolsky, S. A., and Vetterling, W. T., *Numerical recipes in C*, Cambridge University Press, 2nd ed., 1992.
- Rönnmark, K., Quantitative methods for waves in space plasmas, *Space Sci. Rev.*, **54**, 1–73, 1990.
- Strang, G. and Nguyen, T., *Wavelets and filter banks*, Wellesley-Cambridge Press, Welles-

- ley, MA., 1996.
- Theiler, J., Eubank, S., Longtin, A., Galdrikan, B., and Farmer, J. D., Testing for nonlinearity in time series: the method of surrogate data, *Physica D*, **58**, 77–94, 1992.
- Thomson, D. J., Maclennan, C. G., and Lanzerotti, L. J., Propagation of solar oscillations through the interplanetary medium, *Nature*, **376**, 139–144, 1995.
- Vago, J. L., Kintner, P. M., Chesney, S. W., Arnoldy, R. L., Lynch, K. A., Moore, T. E., and Pollock, C. J., Transverse ion acceleration by localized lower hybrid waves in the topside auroral ionosphere, *J. Geophys. Res.*, **97**, 16 935–16 957, 1992.
- Welch, P. D., The use of fast Fourier transform for the estimation of power spectra: A method based on time averaging over short, modified periodograms, *IEEE Trans. Audio. Electroacoust.*, **AU-15**, 70–73, 1967.

— 2 —

Time Series Resampling Methods

CHRISTOPHER C. HARVEY

*Observatoire de Paris-Meudon
Meudon, France*

STEVEN J. SCHWARTZ

*Queen Mary and Westfield College
London, United Kingdom*

2.1 Introduction

Space plasma physics data is almost entirely in the form of time series, and its analysis involves the manipulation of these time series. The sampling of each data set is synchronised with either some onboard clock, or the local spacecraft spin clock: and no two of these clocks will be synchronised. The comparison of non-synchronised time series data is thus fundamental to space physics. This is especially true for Cluster, as the purpose of this mission is to compare data from four different spacecraft.

If two time series are plotted one next to the other for visual comparison, then there is no problem. But as soon as they are correlated in any other way, it is essential that they be time-synchronous. For example, it may be thought useful to produce a scatter plot of observable x against observable y ; but this clearly makes sense only if x and y are sampled simultaneously. In reality, the sampling of x and y will generally fall into one of the following categories, with x and y sampled:

1. At different times with respect to a common clock, so that there is a constant offset in the time of sampling.
2. At different rates, but synchronised to the same clock. Then one sampling rate will be a rational fraction of the other.
3. In synchronisation with different clocks (e.g., on different spacecraft), so that there is no time synchronisation at all. The time difference between the successive samples of x and y grows regularly, without limit.
4. Both sampled at different rates, and synchronised to different clocks.

Clearly it is pointless to produce a scatter plot if the time offset falls into any the latter three categories. Even for category 1, the quality of the scatter plot may be seriously degraded.

Time synchronisation of segments of two (or more) time series is even more important prior to numerical comparison, such as cross-correlation or the calculation of the cross-spectrum. This also applies to the determination of spatial gradients, the primary objective \Rightarrow 2.1 of the Cluster mission (see Chapters 12 through 13 of this book).

The adjustment of a segment of time-series data set to produce a segment of data which is *scientifically equivalent* but with data sample timing strictly simultaneous with that of another data set is called “resampling”. In this chapter we set out the basic terminology, the scientific issues, and the functions which are associated with resampling.

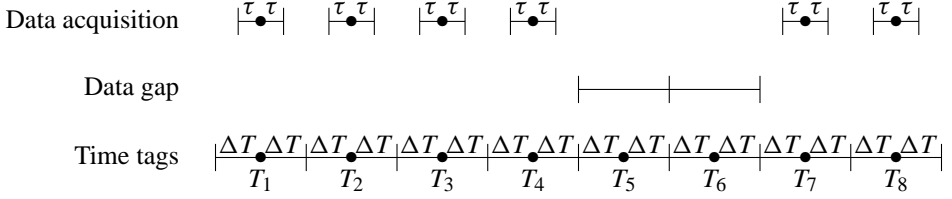


Figure 2.1: The various time tag-related parameters. In this illustration, $\tau = \Delta T/2$ so that, unless the input signal was filtered before being sampled, the sampling factor is $s = \frac{1}{4}$.

This chapter was originally motivated by the work done by the authors respectively for the data analysis system used by the Cluster Wave Experiment Consortium, and for the UK Science Analysis System. The authors are not aware of any specific treatment of the topics in this Chapter, but good general texts can be found amongst the references to Chapter 1. Note that resampling is discussed in a different context in Chapter 9, page 233.

2.2 Definitions

Most of the general terms used in this chapter are commonplace, but it is nevertheless useful to begin by recording a number of time-related concepts and definitions, some of which are illustrated in Figure 2.1.

1. **Data Set**, also called a time series data object. This is a set of values for a quantity X associated with time T , i.e., the set (T_i, X_i) , plus the metadata relating to both X and T . In other words, it includes all the information essential to be able to use the data. Note that the time tags do not necessarily have to be specified for each separate datum, it is enough to provide a method of obtaining the time tag of each individual datum.

In this paper we are concerned mainly with the time-series aspects of the data set; but the metadata will be mentioned briefly in Section 2.7.

2. **Datum**. the X_i 's represent values of a particular data quantity, which could be a scalar, vector, array or other form/collection of values. In this chapter, we shall represent all such quantities by a single X_i , without any explicit naming of the components thereof. Some operations, e.g., frame rotation, require specific types of data object. Each datum has associated with it metadata, including units, reference frame (for vectors and tensors), etc.
3. **Time Tag**. The times T_i are known as time tags. Each T_i tags a particular value X_i with a unique time to which that value is associated. The time tag could be the start, the middle, or the end of the interval over which measurement was performed to produce X_i . The tagging convention used in this chapter (and for Cluster) is that T_i is the middle of the interval. The format and units of the T_i must be established.

Note that at present there is no recognised unique standard convention for either time tagging or for the time format.

4. **Sample Interval.** The interval $2\Delta T$ is defined as the *nominal* interval between successive time tags. It is nominal in the sense that a single value is assumed to apply across the entire time covered by the time series. This sample interval is required in order to distinguish genuine data gaps (see Section 2.6.6) from regularly but sparsely sampled data. (For Cluster Prime and Summary Parameters, ΔT is held in the CDF variable “Half-Interval”.)

Some data sets may have irregularly spaced data samples. Then each data sample must be individually time-stamped. Treatment of such data objects is discussed in Section 2.6.7.

5. **Sample Duration.**

The sample duration 2τ defines the duration over which the measurement was made to determine the value of X . Thus, the datum X_i relates to measurements made during the time interval

$$T_i - \tau \leq t \leq T_i + \tau$$

The half-sample duration τ may be related to either the half-interval ΔT of the time tags or an averaging/integration interval used in determining/measuring X_i . Note in particular that τ can be less than ΔT as illustrated in Figure 2.1, e.g., if the quantity is not continuously sampled, or greater than ΔT , e.g., when a sliding boxcar average is performed. Also, the averaging may correspond to strictly linear integration as when counting particles, or may follow some other law, as when filters are placed before an analogue to digital converter.

6. **Data Gap.** A data gap is a time interval within which a time-tag and datum is expected, but for which there is no entry in the time series. The data gap extends over the interval defined by

$$T_i + \Delta T < t < T_{i+1} - \Delta T$$

7. **Sampling Factor.** The sampling factor s is the ratio of the Nyquist frequency to the highest spectral frequency component present in the data. Its importance is explained in Section 2.5. It is a “rule-of-thumb” parameter which nevertheless is very useful to describe the information content of the data: $s < 1$ (under-sampling) highlights that particular care is required when interpreting the data, while a value $s > 1$ (over-sampling) defines the effective upper frequency limit of any time-series analysis (Fourier transformation, correlation or wavelet analysis, etc.), which may be usefully performed.

The value of the sampling factor is determined by considerations which may be experimental (pre-digitiser anti-alias filtering), intrinsic to the data set (e.g., orbit data), or due to data processing (numerical filtering). In particular, the sampling factor may be changed when data sets are resampled; this described in Section 2.6.5. Its importance for data analysis is explained in Section 2.8.1.

2.3 Time-Domain Operations

In this section we describe briefly the different time-domain operations which can be performed on a data set. Some of these operations are described in greater detail later.

Note that in all these operations vector (and higher dimensional object) properties *must* be preserved, either by applying any algorithm simultaneously to all components, or else by decomposing and then recomposing the separate components.

1. **Resampling.** Resampling of a data set consists formally of interpolating a time series (T_i, X_i) onto a set of time tags (t_j) whose sample interval is $2\Delta t$ to produce a set of values for X at times t_j . We will call t_j the *target* time tags. The interpolation is performed using one of many possible algorithms, some of which are described in the next section. Thus

$$\left. \begin{matrix} (t_j) \\ (T_i, X_i) \end{matrix} \right\} \rightarrow (t_j, \chi_j) \quad (2.1)$$

where the χ_j 's are the interpolated values of the X_i 's.

As already mentioned, resampling of data is generally required when data comes from different sources. In particular, it is required:

- To compare any two data sets which are sampled at different frequencies, and in particular to compare high resolution data with summary data.
 - To compare data from the different spacecraft, because their sampling is not synchronised.
 - To compare spin-synchronous and clock-synchronous data from one spacecraft.
 - To compare satellite data with ground-based data and with data from other spacecraft.
2. **Joining.** The joining of two data sets is the operation of resampling one data set onto the time tags of the other data set, then presenting the two data sets together using a common set of time tags. By repeating this process, several data sets may be joined onto a common timeline.

Joining is the operation required for correlative data analysis.

3. **Merging.** When two data sets are merged both time lines are retained, and are simply put into a common format on a single file. Joining is different, because it places all the variables onto a single time line. Merging and joining are equivalent for two data sets already on the same time line.
4. **Concatenation.** This is the production of a single data set from two separate and consecutive data sets. Whereas joining concerns data from different sources acquired during the same time interval, concatenation concerns data from the same source for different, contiguous, time intervals. Concatenation is required:
 - To produce files which are continuous over different segments of the source data files: typical examples are continuity over midnight, or over the end of a year.

In this chapter we are primarily concerned with resampling and joining. Merging and concatenation are more straight-forward (and may employ resampling if required), although the treatment of the metadata may require care to ensure its applicability over the entire dataset.

2.4 Algorithms

The following basic techniques for joining data sets can be identified. Proper implementation requires consideration of the sampling factor, and is deferred until Section 2.6.

1. **Fuzzy join** Provided that $|T_i - t_j| \leq \Delta T_{\text{tol}}$ where ΔT_{tol} is the tolerance allowed, this selects the value of T_i closest to t_j , and sets $\chi_j = X_i$. Typically $\Delta T_{\text{tol}} = \tau + \Delta t$, i.e., the interval over which the measurement is valid must overlap the output time tag interval. That is $T_i \pm \tau$ must overlap $t_j \pm \Delta t$. Suitable recovery needs to be provided if this criterion is not satisfied, probably by treating it as a data gap (see below). Note that if the original time series measurements are valid for only a portion of the original time tag interval, i.e., $2\tau < T_{i+1} - T_i$, this fuzzy algorithm can fail to find a valid χ_j even if there is no data gap in the original time series. Another logical choice for ΔT_{tol} is therefore $\Delta T_{\text{tol}} = \Delta T + \Delta t$, which fails only at genuine data gaps.
2. **Linear interpolation** This algorithm interpolates linearly the data corresponding to the two nearest time-tags T_i . Thus, find i such that

$$T_i \leq t_j \leq T_{i+1} \quad (2.2)$$

then

$$\chi_j = X_i + \left(\frac{t_j - T_i}{T_{i+1} - T_i} \right) (X_{i+1} - X_i) \quad (2.3)$$

3. **(Cubic) Spline Interpolation** As above, but use a spline algorithm.
4. **Boxcar Averaging** Define a “boxcar” of width $2\Delta T_{\text{box}}$. Then

$$\chi_j = \langle X_i \rangle_{\text{box}} \quad (2.4)$$

where $\langle \dots \rangle_{\text{box}}$ denotes an average over all i 's satisfying

$$t_j - \Delta T_{\text{box}} \leq T_i \leq t_j + \Delta T_{\text{box}} \quad (2.5)$$

This algorithm is useful if the original time series has higher time resolution than the desired time tags. For reasons explained in Section 2.6.2, the relation

$$\Delta T_{\text{box}} \geq 2\Delta t$$

must be satisfied, preferably close to the condition of equality so as to optimise the sampling factor s .

5. **Weighted average** Linear interpolation is performed using a least-squares fit to the number of data points appropriate to the sampling interval of the new data set. The advantage over boxcar averaging is that this method allows account to be taken of the time at which the pre-averaged data points were acquired. This can make the averaged data resemble more closely data acquired, for example, from a sensor sampled using an analogue anti-alias filter. An analogue filter will generally produce an output which is an average with more weight attached to the input signal near the end of the sample duration than near the start. Weighted averaging allows a similar effect to be simulated when averaging high resolution data ready for comparison with lower resolution analogue-filtered data.

Methods 2 and 3 are used for increasing the time resolution of the data, and 4 and 5 for reducing it. Except for the fuzzy join which is so simple that it can only be accepted or be rejected, all these methods may be applicable, according to the circumstances. Examination of the scientific aspects of resampling in the following section shows that:

1. When the data sample interval is increased by a large factor, boxcar averaging is suitable provided that adjacent boxcars overlap in time by at least 50%. ⇒2.2
2. When the data sample interval is reduced, linear interpolation is suitable. ⇒2.2
3. When the data points are essentially shifted in time (for example, to synchronise two data sets with nearly the same sample period), the situation is more delicate. Linear interpolation of one of the data sets using only the two data points with time-tags nearest to the target time-tag leads to modulation of the spectrum of the resampled signal at the beat period. Use of a Bartlett window avoids this problem, but nevertheless changes the spectrum of the signal at frequencies approaching the Nyquist frequency.

These different procedures are described further in Section 2.6.

2.5 Aliasing, Filtering and the Nyquist Frequency

The dictionary defines “sample” as “a small portion to show the quality of the whole”. When a time series is sampled, the resulting series of data samples clearly contains less information than the original, and it is important to understand what information has been lost during sampling, and how it has been lost. Of fundamental importance in this respect is the *Nyquist frequency*, which is defined as half the sampling frequency, $f_N = f_s/2 = 1/(4\Delta T)$, of the corresponding data set.

At the risk of repeating what is well-known, consider two signals $x_1(t) = \cos 2\pi(f_N - f)t$ and $x_2(t) = \cos 2\pi(f_N + f)t$, both sampled instantaneously at times $t_n = 2n\Delta T$. The samples at time t_n are identical, $x_1(t_n) = x_2(t_n) = (-1)^n \cos 4\pi n f \Delta T$. Thus any spectral component with frequency f between f_N and $2f_N$ is “folded back” into the spectrum to appear at frequency $f_N - f$. Similarly, frequencies in the range $2nf_N < f < (2n+1)f_N$ are shifted, and between $(2n+1)f_N < f < (2n+2)f_N$ are shifted and folded, into the range $0 < f < f_N$. The effect is illustrated in Figure 2.2, which is for an input waveform which is a continuous function of time, as in the case of an analogue signal prior to digitisation.

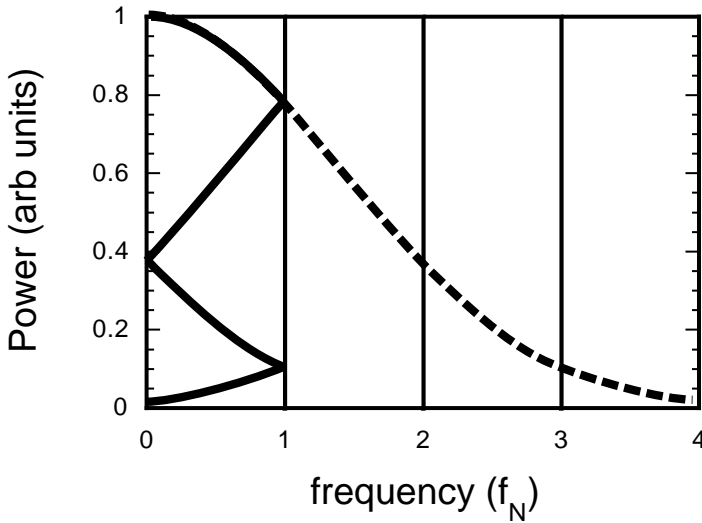


Figure 2.2: This figure shows how frequencies above the Nyquist frequency f_N (dashed) appear shifted and/or folded back to within the frequency range $0 \leq f \leq f_N$ (solid curves). For the purpose of illustration we have used a Gaussian power spectral distribution of width $\sqrt{2}f_N$, for which exactly 50% of the overall spectral power is aliased.

This shifting of higher frequencies so that they appear at frequencies below the Nyquist frequency is the well-known phenomenon of aliasing. It is the process by which the information content is reduced: all frequency components above the Nyquist frequency are shifted, and possibly folded back, to lie one on top of the other. This not only reduces the information content: it also makes the data well-nigh totally useless for spectral analysis (but see Section 2.8.1). Therefore, before being sampled, the data must be filtered to attenuate all but one of the frequency intervals nf_N to $(n+1)f_N$ until all significant out-of-band signals are reduced to a negligible level. Normally the value $n = 0$ is chosen, but this is not essential; it is, however, the only case considered here. Filtering above the Nyquist frequency should eliminate the dashed curve in the example of Figure 2.2 to leave only the uppermost continuous curve. A real filter will not have a precise cutoff, but a characteristic roll-off over a finite range of frequency; this roll-off should be chosen to obtain adequate rejection at frequencies above the Nyquist frequency.

Aliasing occurs whenever the sampling frequency is reduced, that is, whenever the interval between successive data samples is increased. In particular, it is associated with:

- the sampling of an analogue signal necessary to transform a physical sensor output into digital form, and
- the resampling of digital signals, which is the subject of this chapter.

To minimise the effects of aliasing, the data should be filtered *before the sampling frequency is reduced* so as to remove any frequency component above the Nyquist frequency

f_N of the resampled data. This anti-aliasing filter will be analogue prior to digitisation, or digital prior to resampling.

2.5.1 The Sampling Factor

We define the *sampling factor* s to be the ratio of the Nyquist frequency to the cutoff of the anti-aliasing filter used before the data was sampled or resampled. This filter will be either analogue or digital, depending upon whether the input data is analogue or digital. As mentioned in Section 2.2, s is approximately the ratio of the Nyquist frequency to the highest frequency present in the data.

The sampling factor is only a “rule-of-thumb” indicator of the data quality; it indicates the “information density”, but it does not describe the characteristics of the anti-aliasing filter. Some filters, such as analogue filters, or the Gaussian digital filter (page 55) do not have a well-defined cutoff. Note that more than one anti-aliasing filter may be involved: for example, one for the analogue-to-digital conversion, and a second one for subsequent numerical filtering; but in this case the second filter, provided that its filter factor q (see below) is sufficiently large, will dominate in the determination of s . The sampling factor is an essential guide to what can, and what can not, be done with the data, as will be explained in Section 2.8.1. As a general rule digital data should be produced with a value of s as close to unity as possible. But the primary purpose of resampling of data is to facilitate comparison with other data, and so it may well change the value of s .

2.5.2 Analogue to Digital Conversion

Although it is not really the subject of this chapter, it is instructive to consider analogue to digital conversion. Some experiments, for example wave experiments, can use analogue anti-aliasing filters to smooth the data before it is digitised; thus the problem can be completely mastered. For other experiments, such as particle experiments, this is not so easy. Very often the best that can be achieved is linear integration, or boxcar averaging, over an interval of time 2τ . In the frequency domain, the effect of this convolution in the time domain is to multiply the spectrum by

$$F(f) = \frac{1}{2\tau} \int_{-\tau}^{\tau} e^{2\pi i f t} dt = \frac{\sin(2\pi f \tau)}{2\pi f \tau}$$

The first zero of this function occurs at frequency $f_0 = 1/2\tau$. The condition $f_0 \leq f_N$ already requires $\tau \geq 2\Delta T$, which is often impracticable; and even when possible, the resulting protection against aliasing is rather mediocre due to the sidebands of $F(f)$. Assuming a white spectrum for the measured signal, the fractional contribution of aliased signals to the total power of the fluctuations is:

$$\int_{1/(4\Delta T)}^{\infty} F^2(f) df \bigg/ \int_0^{\infty} F^2(f) df = \frac{2}{\pi} \int_{1/(4\Delta T)}^{\infty} F^2(f) df$$

(using $\int_0^{\infty} F^2(f) df = \pi/2$). Values of this ratio for different values of $\tau/\Delta T$ are given in Table 2.1.

The values of s are only approximate, because the definition of s is somewhat imprecise. In Table 2.1 we have assumed that the “highest” frequency present is $f_0 = 1/2\tau$, the

ratio $\tau/\Delta T$	1/2	3/4	1	2	3	4	5
aliased power	53.3%	35.3%	22.6%	9.5%	6.6%	4.6%	3.6%
sampling factor	1/4	3/8	1/2	1	3/2	2	5/2

Table 2.1: Characteristics of a boxcar window. For different values of $\tau/\Delta T$ are tabulated the sampling factor s , and the total aliased (out-of-band) power for a white noise input signal.

first zero of $F(f)$. This is plainly incorrect; for $\tau = 2\Delta T$, we have $s = 1$, but nevertheless 9.5% of the power is from aliased out-of-band signals.

2.5.3 Resampling and the Reduction Factor

The same phenomenon of aliasing occurs when the sample rate of a digital signal is reduced. For example, if the sampling frequency (number of samples per second) is reduced from F_s to f_s , the new Nyquist frequency is $f_N = f_s/2$ and any spectral information in the original signal between f_N and F_N (including undesirable signals present due to earlier aliasing, for example, during analogue to digital conversion) is aliased into the frequency range below f_N . As before, frequencies between $2nf_N$ and $(2n+1)f_N$ are shifted, and frequencies between $(2n+1)f_N$ and $(2n+2)f_N$ are folded and shifted, into the frequency range 0 to f_N .

For a digital signal the power spectral density is not defined at frequencies above the Nyquist frequency. The figure equivalent to Figure 2.2 for a digital input would show the input power spectral density (the dotted line) to be identically zero above the Nyquist frequency F_N of the input signal. The ratio of the input to output Nyquist frequencies is precisely the ratio of the input to output sampling frequencies, $F_N/f_N = F_s/f_s = r$. We call the ratio r the *reduction factor*. In terms of the sample interval, $r = \Delta t/\Delta T$. Thus, the volume (number of data points) of the data set is multiplied by a factor of $1/r$.

2.5.4 Filtering and the Filter Factor

Clearly when $r > 1$ numerical anti-alias filtering of the input data set X_i is required to remove spectral components at frequencies above the Nyquist frequency $f_N = 1/(4\Delta t)$ of the time tags t_j . We define the *filter factor* q as the ratio of the Nyquist frequency F_N of the data which is being filtered to the cutoff frequency of the numerical filter. Filtering therefore increases the initial (before resampling) sampling factor S to some new value qS . The sampling factor s is preserved during the overall operation of filtering and resampling if $q = r$; this condition is not always satisfied.

Filtering is performed by convolution with a numerical window. Ideally the window should provide a filter with:

- minimal power in the sidelobes above the Nyquist frequency, to reduce aliasing; and
- a rapid cutoff close to the Nyquist frequency, to avoid attenuation of signals within the passband.

These are conflicting requirements, and so the actual filter must be chosen case by case. Several windows are suggested in Section 1.3.3 (equations 1.23 through 1.26), and their characteristics and efficacy are illustrated in Figures 1.5 and 1.6. The characteristics of the discrete boxcar and Bartlett windows are discussed respectively in Section 2.6.2 and Appendix 2.A.

2.6 The Art of Joining Data Sets

Joining consists of resampling a data set with interval $2\Delta T$, onto a set of time tags with interval $2\Delta t$. The time tags may either be associated with another data set, or be completely independent and generated, for example, synchronously with Universal Time. By taking the sets one after another, several data sets may be joined onto the same timeline.

The data set to be resampled will normally be the one with higher time resolution, $\Delta T < \Delta t$, so that joining reduces its time resolution (makes ΔT larger), that is, the reduction factor (Section 2.5.3) $r > 1$. It is not generally useful to increase the time resolution (reduce ΔT) of normal science data: this leads to a large sampling factor s , because physically meaningful information cannot be created by interpolation. But there may nevertheless be an occasional need to interpolate, for example, when joining several data sets of high resolution and only a few of lower resolution.

In either case care must be exercised:

- When $\Delta t > \Delta T$, the sample interval of the resampled data set is increased, and averaging must be performed to prevent under-sampling.
- When $\Delta t < \Delta T$, the sample interval of the resampled data set is reduced, which necessarily means that this data becomes over-sampled.

In the following sections we describe algorithms to join the primary data set

$$\dots \quad t_{j-1}, x_{j-1}, \quad t_j, x_j, \quad t_{j+1}, x_{j+1}, \quad \dots$$

and the secondary data set

$$\dots \quad T_{m-1}, X_{m-1}, \quad T_m, X_m, \quad T_{m+1}, X_{m+1}, \quad \dots$$

by resampling the secondary data set so that the resampled data point $\dots t_j, \chi_j \dots$ coincides in time with the primary data point $\dots t_j, x_j \dots$. We call t_j the “target” time for resampling at time t_j . We assume that the times t_j and T_j refer to the centres of their respective sampling intervals.

We now review each of the techniques mentioned in Section 2.4 in the light of the considerations of Section 2.5.

2.6.1 Fuzzy Join

There is no filtering, so the filter factor is $q = 1$. Therefore this method must not be used if the reduction factor is $r > 1$, as the data would be under-sampled ($s < 1$) and relatively useless. On the other hand, if $r < 1$, the data values obtained will be less good than those obtained by interpolation.

2.6.2 Averaging

When $\Delta T \ll \Delta t$ simple interpolation between the two data points T_i and T_{i+1} closest to the target time-tag t_j would result in an under-sampled and hence unusable secondary data set (see Section 2.5.3). The resampling algorithm must include numerical filtering.

Anti-alias filtering and resampling are two different operations which are described by two different parameters, the filter factor q and the reduction factor r . In terms of the sampling factor S of the secondary data, the sampling factor of the averaged data is

$$s = \frac{q}{r} S \quad (2.6)$$

The parameters q and r are not necessarily the same; q should normally be chosen to optimise s , that is, obtain $s \simeq 1$. In other words, q should be as small as possible while still satisfying

$$q \geq \frac{r}{S} \quad (2.7)$$

When the input data is over-sampled ($S > 1$), q may well be less than r as, for example, when resampling orbital trajectory data.

Exceptionally, it may be acceptable for equation 2.7 not to be satisfied if the reduction factor q is so large that, on physical grounds, the value of S may be considered to be irrelevant; then, the sampling factor is $s = q/r$. Caution must nevertheless be exercised during subsequent data analysis, as explained in Section 2.8.1.

Normally filtering and resampling are performed simultaneously by a single algorithm. Whatever the value of q , the spectrum of the resampled data will reflect the characteristics of the numerical filter as described by q and the nature (boxcar, Bartlett, etc.) of the filter employed, particularly at frequencies approaching the Nyquist frequency. When q is of the order of unity, the spectrum of the resampled data will also be influenced by the characteristics of earlier anti-alias filtering, performed prior to resampling.

The Boxcar Window

The boxcar window with filter factor q has a width of approximately $2q$ data points. The reasoning is analogous to that of Section 2.5.2. The analytic signal of frequency f sampled at intervals $2\Delta T$ has the discrete representation $X_k = e^{4\pi i f k \Delta T}$. The convolution of this with the N -point boxcar window extending from X_ℓ to $X_{\ell+N-1}$ is

$$F(f) = \frac{1}{N} \sum_{k=\ell}^{\ell+N-1} e^{4\pi i f (\ell+k) \Delta T} = \frac{\sin(2\pi f N \Delta T)}{N \sin(2\pi f \Delta T)} e^{4\pi i f (\ell + \frac{N-1}{2}) \Delta T} \quad (2.8)$$

where $e^{4\pi i f (\ell + \frac{N-1}{2}) \Delta T}$ is the phase of the signal at the centre of the window. The first zero of F occurs when $f = 1/(2N\Delta T)$, while the Nyquist frequency is $1/(4\Delta T)$; hence the filter factor is $q = N/2$. Thus to obtain a filter of factor q , the boxcar window is simply

$$w_k = \frac{1}{N} \quad \text{for} \quad 1 \leq k \leq N \quad \text{with} \quad N = 1 + \text{int}(2q) \quad (2.9)$$

where $\text{int}(2q)$ is the integral part of $2q$; note that q is generally not an integer, and $+1$ is added to ensure that $N \geq 2q$.

Filtering and resampling can be performed simultaneously by simply placing the boxcar to select the required number of data points of the secondary data set closest to the target time-tag t_j of the primary data set. Thus to obtain a filter factor q , for every data point t_j, x_j of the primary set we identify two data points of the secondary set, ℓ and n , such that

$$\text{and} \quad \begin{array}{ccc} t_j - 2\Delta T & < & T_\ell + \Delta T \text{int}(2q) \leq t_j \\ t_j & \leq & T_n - \Delta T \text{int}(2q) < t_j + 2\Delta T \end{array} \quad (2.10)$$

Thus equation 2.9 is satisfied by the N points closest to t_j . The calculation of the average is straightforward:

$$X'_j = \frac{1}{n - \ell + 1} \sum_{i=\ell}^n X_i \quad (2.11)$$

Note that (equation 2.10)

$$T_n - T_\ell \geq 4q\Delta T = \frac{4q}{r}\Delta t = \frac{2q}{r} \times \text{the sample interval of the primary data set.}$$

Adjacent boxcars *must overlap* by at least 50%. If they are simply contiguous the filter factor will be $q = r/2$, the output sampling factor will be $s = qS/r = S/2$, and the resampled data will be contaminated by aliased signals. The only possible exception is for over-sampled input data with $S > 2$, for which q may be less than r provided that $q \geq r/S$ (equation 2.7).

Boxcar averaging yields more than -20 dB rejection at the maximum of its first side-lobe, and an in-band attenuation of -3.9 dB at $f_N/2$, and -0.9 dB at $f_N/4$. It is entirely adequate when a large filter factor q is required. Then there are a large number of points within the window, which can be positioned with respect to the target time with reasonable precision. Compared to other windows, the boxcar window reduces both the computational requirements and the problems associated with end effects.

Problems arise, however, when the reduction factor r is small. Each time the target time t_j increments by $2\Delta T$ (one sample interval of the secondary data set), the phase increments (equation 2.8) by $4\pi f \Delta T$, as expected. But as t_j advances, this phase increase occurs discontinuously: both ℓ and n increment each time t_j reaches the value $\Delta T + (T_\ell + T_n)/2$ (equation 2.10). Simultaneously the phase of the filtered signal jumps by $4\pi f \Delta T$. For frequencies f approaching the Nyquist frequency $f_N = 1/(4\Delta t)$, the phase discontinuity of the resampled data is $\pi \Delta T / \Delta t = \pi / r$ (Section 2.5.3).

Equation 2.10 could be changed so that ℓ and n do not increment simultaneously: then smaller phase discontinuities are obtained (as small as 50 % of the above values), but they occur more frequently. Furthermore, N would alternate between two possible values, with consequent changes of the filter transfer function.

In summary, when the reduction factor r is close unity, the phase error of the resampled datum approaches to $\pm 90^\circ$, a value which can be reduced to $\pm 45^\circ$ if periodic changes in the transfer function can be tolerated. We conclude that for a reduction factor r of the order of unity, linear interpolation, and especially the Bartlett window, offer better solutions, as discussed in Section 2.6.3.

The Gaussian Window

The discrete equivalent of the Gaussian filter is the $(N+1)$ -point filter defined in terms of the binomial coefficients

$$w_r = {}^N C_r / 2^N \quad \text{for } 0 \leq r \leq N$$

the factor 2^N being to normalise the window, $\sum_{r=0}^N w_r = 1$. (It may be truncated on the wings, in which case the normalisation changes slightly.) This filter is not normally recommended; although it has no sidelobes, for this very same reason it is continuous across f_N and the overriding anti-aliasing requirement of negligible response above f_N implies negligible in-band response immediately below f_N , plus a rather gentle cutoff.

The other filters presented in Section 1.3.3 do have sharp cutoffs; to optimise the information content and obtain $s = 1$, the first cutoff should be placed at the Nyquist frequency of the primary data set.

The Bartlett Window

The triangular shape of the Bartlett window allows both the width and the position relative of the window relative to the target time-tag to be specified with some precision. In principle, this window is only slightly more complicated than the boxcar window; but, in practice, indexing and normalisation make it appear much more complicated. We have relegated the details to Appendix 2.A, where it is shown that the width of the window is approximately $4q$.

At all frequencies this window has approximately twice the attenuation (in dB) of the boxcar window. It has more than -40 dB rejection of the first sidelobe, -7.9 dB attenuation at $f_N/2$, and -1.8 dB at $f_N/4$. It is questionable whether this performance is worth the extra computation incurred by use of this window for $r \gg 1$: twice as many points in the window, and each point having to be multiplied by a different window coefficient w_r . If a symmetric window (equations 2.21 and 2.22, or 2.23 and 2.24, depending upon the value of q) is used, the phase error is at most 45° . But if the more general non-symmetric window of equation 2.15 is used, the phase errors are negligible; but the window coefficients must be evaluated for each resampled datum t_j, χ_j .

2.6.3 Synchronisation

When comparing two signals sampled at nearly, but not exactly, the same rate, “beating” will occur with a period

$$2\Delta T \Delta t / (\Delta T - \Delta t) \quad (2.12)$$

In particular, it will occur when comparing parameters measured by identical instruments on different spacecraft of the Cluster mission. Resampling is required to synchronise the data sets.

When resampling is performed with $r \simeq 1$, the Nyquist frequency does not change appreciably, there is no aliasing, and no filtering is required. The simple linear interpolation algorithm of equations 2.2 and 2.3 eliminates the phase problem associated with the algorithm of equation 2.11. Nevertheless, there remains an amplitude problem, as will now be shown.

When $t_j \simeq T_i$ (or T_{i+1}), the resampling algorithm 2.3 yields $\chi_j \simeq X_i$ (or X_{i+1}), and resampling has little effect on the data. But when $t_j \simeq (T_i + T_{i+1})/2$ it approximates the boxcar averaging described by equation 2.8, with $N=2$: the data is filtered with a factor $q=1$. While this does not affect the spectral cutoff frequency, it does affect the spectrum within the band with, for example, an attenuation of -3.9 dB at $f_N/2$ (section 2.6.2). As time advances there is a beating between the sampling periods $2\Delta t$ and $2\Delta T$ of the primary and secondary time-tags, so that the spectrum of the resampled data is modulated with the period of equation 2.12. It is possible to reduce this effect by using the sliding Bartlett window with $q=1$, that is, the 4-point window described by equations 2.15 through 2.20 of Appendix 2.A. This causes attenuation within the band which is greater than for algorithm 2.3, but constant (no beating).

This discussion indicates that the effects of resampling must not be overlooked as one approaches the Nyquist frequency, especially if data from one spacecraft is resampled, but not that from another. For symmetry of processing, it may be better to resample both data sets onto a common independent timeline, for example, one related in a simple way to Universal Time.

2.6.4 Interpolation

When $\Delta T \gg \Delta t$, the data X_i must be interpolated, using algorithms 2 or 3 (of Section 2.4).

Linear interpolation is the only type of interpolation which may be safely applied to science data. Higher order interpolation may be applied only to data which is over-sampled (sampling factor greater than unity). Even then, its application to science data is dangerous: experimental errors and noise in the original data set introduce completely spurious information into the resampled data. In any case, it is impossible to correlate two (or more) science data sets with a precision in time better than twice the length of the largest sampling interval of the data sets being correlated (Nyquist's theorem).

Interpolation increases the sampling factor, which becomes $s \times p$ where p is an *interpolation factor* equal to the number of interpolated points per interval of the original data. Note that the interpolation and reduction factors are reciprocal quantities, $p = 1/r$. Interpolation involves no filtering, $q = 1$, so that interpolated data necessarily has $s > 1$; the data is over-sampled (see Section 2.8.1).

Interpolation of higher order than cubic is almost certainly neither required nor desirable. It may be performed only on data which is over-sampled and which has negligible (or adequately little) experimental noise, and is thus known to contain no spectral information at frequencies above the Nyquist frequency: for example, orbit data. Nevertheless, standard over-sampled data products will normally be provided with sufficient time resolution for linear interpolation to be entirely adequate.

2.6.5 Determination of the Sampling Factor

Resampling will generally change the sampling factor associated with the data. The resultant sampling factor s depends upon the sampling factor S of the secondary (input) data, and upon the filtering and reduction (or interpolation) performed, as described by q and r (or $p = 1/r$). The sampling factor s of the resampled data is determined in one of three ways.

1. **The initial data and the resampling performed.** Provided that $S \geq 1$, the expression 2.6

$$s = \frac{q}{r} S$$

determines s in terms of the sampling factor S of the initial secondary data and the filtering and/or interpolation to which it has been subjected. If the original data is over-sampled, $S > 1$, the ratio q/r may be chosen to bring s closer to unity and thus optimise the information content of the resampled data.

2. **Leave unchanged.** If the secondary sampling factor S is less than unity, then it is best to leave it unchanged. Resampling neither increases nor reduces the effects of aliasing already present in the data.
3. **The resampling process alone.** When the filter and reduction factors q and r are both very large, s may be considered to be determined by $s = q/r$. The justification is that the numerical filtering extends to frequencies so much higher than the Nyquist frequency F_N of the secondary data set that all previously aliased power has been reduced to a negligible level, as explained in Section 2.8.1.

2.6.6 Data Gaps

The handling of data gaps is a problem. Correlative data analysis cannot be performed across a data gap but often, inside a data gap of one instrument, another instrument yields very interesting data which should not be lost when joining. Therefore any joining process must search, trap and deal with data gaps.

A data gap may be identified when

$$T_{i+1} > T_i + 2\Delta T + \Delta T_{\text{gap}}$$

where ΔT_{gap} is a parameter which may be specified, but whose default value is $\Delta T_{\text{gap}} = \Delta T$. Once a gap is encountered, there are several options for the interpolation algorithm to deal with it:

1. Apply the chosen join algorithm to fill the gap (i.e., ignore the fact that it is a gap). This will not help for fuzzy joining which is outside ΔT_{tol} (see joining algorithm 1). Similarly it will not help for boxcar averaging if the gap is too wide, i.e., $T_i + \Delta T + \Delta T_{\text{box}} < t_j$ and $T_{i+1} - \Delta T - \Delta T_{\text{box}} > t_j$. This can happen for gap widths $2\Delta T_{\text{gap}}$ as small as the boxcar width $2\Delta T_{\text{box}}$ and will certainly occur if the gap is more than twice this wide. In these cases, the algorithm has already failed, or will fail, to find a suitable value.
2. Apply a *different* algorithm to fill in the gap, e.g., linearly interpolate across the gap where in contiguous data a spline is used.
3. Remove (or fill with “missing” values) the entries (t_j, χ_j) for t_j lying within a gap in the T_i ’s.

4. Apply the above differently for small and large gaps by setting a threshold ΔT_{maxgap} . For example, linearly interpolate across small gaps and remove entries corresponding to large ones. Typically ΔT_{maxgap} might be related to one of the time tag half-interval ΔT , the measurement half-sample duration τ (if different), or the output time tag interval Δt .

In specifying data gap handling, there is the possibility of looping (i.e., choosing an alternative algorithm if the first one fails, then another, and so on). This is an implementation decision, but it is probably wise to avoid this possibility, so as to obtain a time data series of known homogenous quality.

2.6.7 Irregularly Spaced Data

The same general considerations apply to irregularly spaced data, although concepts such as the Nyquist frequency and the sampling factor s are conceptually rather vague. Note that the algorithms of Sections 2.6.2, 2.6.3 and 2.6.4 apply equally well to irregularly spaced time series data.

2.7 Metadata and Status Data

Mention must be made of the other information which forms an essential part of the data set, as mentioned in Section 2.2. With the gradual move towards an object oriented approach to data analysis, this metadata is required by client applications and must, therefore, be handled in the appropriate way when two data sets are joined.

There are two types of information, which are distinguished by the use which is made of them.

Metadata is information which is intimately attached to the physical data (units, coordinate system, possibly limit values). For example, it is imperative that any algorithm to calculate the Alfvén velocity know the units used for the two physical parameters involved, the density and the magnetic field.

The metadata changes so slowly that resampling is not required. But it must be joined when physical data is joined; how this is done is an implementation issue.

Status data is information which is more separated from the instrument data to which it refers, and is unlikely to be required by client applications using the instrument physical data. The instrument status (active/passive, or solar wind/magnetosphere) is a typical example. Such information may be used by other experiments to determine the suitability of the data for comparison with their own observations. It has an inherent time granularity, which is somewhat arbitrary.

The way in which status data is handled during joining is very dependent upon the architecture of the software implementation, and it is probably premature to discuss this further at this time.

2.8 Discussion

Any two data sets which are sampled at a different frequency will eventually become seriously out of step unless action is taken. The two data sets must be synchronous (i.e., any relative time offset must be constant) before any cross-correlation is attempted. Depending on the nature of the analysis, it may or may not be necessary for the sampling to be simultaneous; but, in practice, any action taken to ensure synchronisation can easily provide simultaneity.

Resampling and/or numerical filtering is also required for the determination of spatial gradients, where the time-series filtering is folded into the spatial filtering. This issue is discussed further in Chapter 15, Section 15.5.1.

The synchronisation of the timelines is achieved by resampling one or both data sets, so that they have identical, though perhaps shifted, timelines. Joining is similar, except that the two data sets share a common timeline.

The effects of averaging or interpolation are indicated by the sampling factor s , a rule-of-thumb parameter which indicates the information content of the data, and which should be appended to the metadata to reduce the possibility of data being used scientifically to draw erroneous conclusions. Ideally, this parameter should be passed to application programs and/or the scientists using them, to avoid fruitless searching for correlation where none is to be found, i.e., at frequencies above f_N/s , or inadvertent misuse of the data when $s < 1$.

Knowledge of the type (boxcar, Bartlett, etc.) of the anti-alias filter used is also essential when using the data for spectral analysis at frequencies approaching the Nyquist frequency. Neither the in-band attenuation, nor the possibility of (attenuated) aliased signals being present in the sidebands of the filter, should ever be overlooked.

2.8.1 Scientific Use of the Sampling Factor

The value of the sampling factor s is important, and must always be known before undertaking scientific analysis of the data.

When $s > 1$ the data is *over-sampled*, and it contains no usable spectral information at frequencies above f_N/s , where f_N is the Nyquist frequency. This information is clearly of primary scientific importance.

For the example of Table 2.1, a sampling factor of $\frac{1}{4}$ yields a power from out-of-band fluctuations which actually exceeds (53.3%) that from the in-band fluctuations if the input fluctuations have a white spectrum. This data is said to be *under-sampled*. Can such data be used at all for spectral analysis?

Spectral analysis may be possible at frequencies well below the Nyquist frequency $f_N = 1/(4\Delta T)$ by assuming that the spectrum of fluctuations is not white, but decreases rapidly with increasing frequency; in particular, many naturally occurring power spectra decrease as f^{-2} , or steeper. But it must never be forgotten that a significant enhancement of the fluctuations at a particular frequency, for example of a particle flux due to phase bunching by a large amplitude wave, may be aliased and cause strange effects to appear even at relatively low frequencies. Subject to this caveat, if $s < 1$ it is possible to smooth under-sampled data using, for example, a sliding boxcar average with a large reduction factor r , thus reducing the number of data points but not the amount of physically useful

information. The information which is eliminated is aliased, and therefore considered physically useless.

There is no way to remove an aliased signal once it is present in the data, and during data analysis this fact must never be overlooked. Consider filtering with a reduction ratio r . If the Nyquist frequency of the resampled data is f_N , the first aliased signal in the input which is not eliminated by this filtering would be a signal originally close to $(2r-1)f_N$ (because of the frequency folding around the Nyquist frequency rf_N of the input data). For a power spectrum varying as $f^{-\alpha}$, this first aliased signal will be $-10\alpha \log_{10}(2r-1)$ dB below the signal close f_N . Generally speaking, if s is initially less than unity, its value should not be changed during resampling unless the reduction factor r is large enough to render it extremely improbable (note that there always remains some subjectivity) that any spectral peak near $(2r-1)f_N$ can exceed the nominal $f^{-\alpha}$ spectrum by as much as $-10\alpha \log_{10}(2r-1)$ dB.

Another example of s being less than unity but not clearly defined concerns derived parameters, that is, ones which are not measured directly. A typical example concerns the moments of a particle distribution. Here the data set parameters (density, flow velocity, temperature, etc.) are derived from many individual measurements (of the particle flux, in different directions and/or energy range), collected sequentially during the interval between successive data set samples; each of these individual measurements has a very small sampling duration and consequently a small value of s . The value of s for the derived data set parameters is not clear, other than that it is significantly less than unity.

In summary, the sampling parameter s is important because:

- If $s < 1$, the data is *under-sampled*. Strictly speaking the data should not be used at all, because it may be contaminated by aliased signals of higher frequency. In reality, provided the data is first filtered and resampled, it may be used at frequencies $f \ll f_N$ provided that the user is well aware of the difficulties described above.
- If $s > 1$, the data is *over-sampled*. The data cannot be used for spectral studies at frequencies above f_N/s , where all significant spectral information has been removed by filtering. The information density is not optimum.

Appendix

2.A The Bartlett Window

Its triangular shape allows both the width and the position of the asymmetric Bartlett window to be specified with some precision. Consider the discrete Bartlett window with filter factor q centred on one of the time tags t_j of the primary data set; for each secondary time tag T_i lying in the interval,

$$t_j - 2q\Delta t < T_i < t_j + 2q\Delta t$$

The window function to be applied to the corresponding datum X_i is

$$w_i \propto 1 - \left| \frac{T_i - t_j}{2q\Delta t} \right|$$

normalised so that $\sum w_i = 1$. More specifically, we identify the values of ℓ , m and n which satisfy

$$\begin{array}{ccccc} T_{\ell-1} & \leq & t_j - 2q\Delta t & < & T_\ell \\ T_m - \Delta t & < & t_j & \leq & T_m + \Delta t \\ T_n & < & t_j + 2q\Delta t & \leq & T_{n+1} \end{array} \quad (2.13)$$

and introduce the fractional offset x of the time t_j of the centre of the Bartlett window with respect to the secondary time tag T_m ,

$$x = \frac{t_j - T_m}{2\Delta T} \quad \text{which satisfies} \quad -\frac{1}{2} < x \leq \frac{1}{2} \quad (2.14)$$

Note that as t_j advances with respect to the time tags T_i , the value of the index m increments by unity and x jumps from $x \simeq +1/2$ to $x \simeq -1/2$ whenever $t_j = (T_m + T_{m+1})/2$. The discrete window with filter factor q centred on a primary time tag with fractional offset x is defined by

$$\begin{aligned} w_{m-k} &= \frac{1}{S} \left(1 - \frac{k+x}{2q} \right) & \text{for} & \quad 1 \leq k \leq m-\ell \\ w_m &= \frac{1}{S} \left(1 - \frac{|x|}{2q} \right) \\ w_{m+k} &= \frac{1}{S} \left(1 - \frac{k-x}{2q} \right) & \text{for} & \quad 1 \leq k \leq n-m \end{aligned} \quad (2.15)$$

where S is the normalisation coefficient, chosen so that

$$\sum_{i=\ell}^n w_i = 1 \quad (2.16)$$

The samples w_i of the discrete Bartlett window are synchronous with the secondary data set, but the overall triangular form of the window is centred on time t_j of the primary data set; we call t_j the “target” time, it is the time tag of the resampled datum. The sample w_m of maximum amplitude occurs before or after t_j according to whether x is positive or negative; and the number of points before and after w_m are respectively

$$\begin{aligned} m - \ell &= \text{int}(2q - x) \\ n - m &= \text{int}(2q + x) \end{aligned} \quad (2.17)$$

where $\text{int}(2q+x)$ denotes the integral part of $2q+x$. Figure 2.3 shows the behaviour of the window as t_j moves relative to T_m (i.e., as x varies) for two typical values of q . The window has an odd number of points if $m - \ell = n - m$, that is, from 2.17, if either

$$|x| \leq \text{frpt}(2q) < \frac{1}{2} \quad \text{or} \quad |x| < 1 - \text{frpt}(2q) \leq \frac{1}{2} \quad (2.18)$$

where $\text{frpt}(2q)$ denotes the fractional part of $2q$.

- When $|x|$ is large enough for neither of equations 2.18 to be satisfied, the window has an even number of points. The normalisation coefficient S can be evaluated from equations 2.16 and 2.15,

$$S = S_e = 2p - \frac{p^2}{2q} \quad \text{where} \quad p = \text{int}\left(2q + \frac{1}{2}\right) \quad (2.19)$$

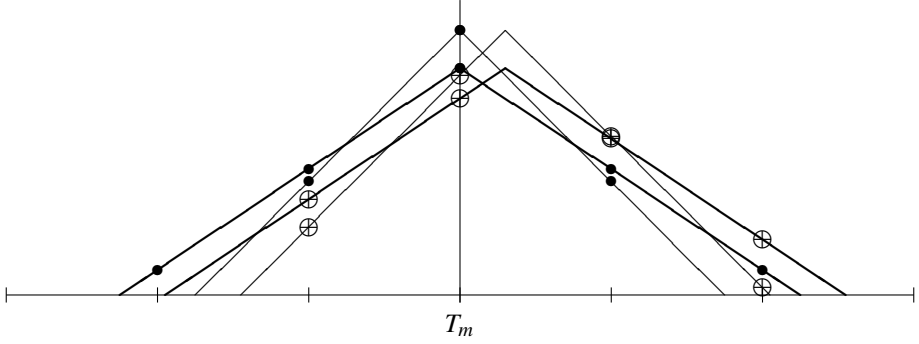


Figure 2.3: The variation of the discrete Bartlett window as the target time “slides” with respect to the time tags T_i , for two values of q close to $q = 1$. The windows are illustrated for $x = 0$ (bullets) and for $x = 0.3$ (crosses). They are not normalised, and the ordinate scale is arbitrary. For $q = 9/8$ (thick lines) when $x = 0$ the window has five points, one of which is “lost” when $x > 1/4$. For $q = 7/8$ (thin lines) the window has only three points when $x = 0$, but it gains an extra point when $x > 1/4$.

and is independent of x .

- Conversely, when $|x|$ is small enough to satisfy one or both of equations 2.18, the window has an odd number of points. The normalisation coefficient S varies with x , and may be expressed in terms of S_e of equation 2.19, thus

$$\begin{aligned} S &= S_e + \frac{\text{frpt}(2q) - |x|}{2q} & \text{if } \text{frpt}(2q) < \frac{1}{2} \\ S &= S_e + \frac{1 - \text{frpt}(2q) - |x|}{2q} & \text{if } \text{frpt}(2q) \geq \frac{1}{2} \end{aligned} \quad (2.20)$$

S has its maximum value when $x = 0$ and the target time of the centre t_j of the window is coincident with T_m .

Thus, whatever the value of q , the window has an odd number of points when $x = 0$ and an even number when $x = 1/2$, as required by considerations of symmetry.

Equations 2.17 and 2.18 show that when $4q$ is an integer the window has $4q$ points for all values of x except $x = \text{frpt}(2q)$. Furthermore, when $x=0$ or $x=\frac{1}{2}$ the points are distributed symmetrically with respect to the centre of the window, and equations 2.15 through 2.20 reduce to simple expressions which can also be readily derived from first principles. Thus when $4q$ is an even integer the window contains $4q$ points for all values of x except one, and is symmetric when $x=\frac{1}{2}$,

$$w_r = w_{4q-1-r} = \left(r + \frac{1}{2}\right) / 4q^2 \quad \text{for } 0 \leq r \leq 2q - 1 \quad (2.21)$$

For the exact value $x=0$ the window has only $4q - 1$ points, but is still symmetric,

$$w_r = w_{4q-2-r} = (r + 1) / 4q^2 \quad \text{for } 0 \leq r \leq 2q - 1 \quad (2.22)$$

When $4q$ is an odd integer the window also contains $4q$ points for all values of x except one, and is symmetric when $x=0$,

$$w_r = w_{4q-1-r} = \left(r + \frac{1}{2}\right) \bigg/ \left(4q^2 + \left(\frac{1}{2}\right)^2\right) \quad \text{for } 0 \leq r \leq 2q - \frac{1}{2} \quad (2.23)$$

For the $x=\frac{1}{2}$ precisely it has only $4q - 1$ points,

$$w_r = w_{4q-2-r} = (r + 1) \bigg/ \left(4q^2 - \left(\frac{1}{2}\right)^2\right) \quad \text{for } 0 \leq r \leq 2q - \frac{3}{2} \quad (2.24)$$

These windows serve both to illustrate equations 2.17 and 2.18, and being simple (compared with the general expressions 2.15 through 2.20) they can provide (page 55) an alternative to the boxcar window.

To obtain an expression equivalent to equation 2.8, we consider the analytic signal $e^{2\pi i f t}$ which yields, for the secondary datum sampled at time T_{m+k} , the value

$$X_{m+k} = e^{4\pi i f (m+k) \Delta T} = e^{i\alpha(m+k)} \quad \text{where } \alpha = 4\pi f \Delta T$$

When one or other of equations 2.18 is satisfied, the window has an odd number of points, and may be expressed by equations 2.15 with

$$1 \leq k \leq K = \text{int}(2q) \quad (2.25)$$

The convolution of this window with X_{m+k} is

$$F(f) = \frac{e^{i\alpha m}}{S} \left\{ \left(1 + \frac{|x|}{2q}\right) + \sum_{k=1}^K \left[\left(1 - \frac{k+x}{2q}\right) e^{-i\alpha k} + \left(1 - \frac{k-x}{2q}\right) e^{+i\alpha k} \right] \right\} \quad (2.26)$$

When neither of equations 2.18 is satisfied the window has an even number of points. It is convenient to introduce the flag

$$p = \frac{1}{2} \left(1 + \frac{x}{|x|}\right)$$

which takes the value 0 or 1 depending upon whether the “target” time t_j is before or after T_m . Then equations 2.15 may be expressed

$$\begin{aligned} w_{m-k+p} &= \frac{1}{S_e} \left(1 - \frac{k+x}{2q}\right) \\ w_{m+k+p-1} &= \frac{1}{S_e} \left(1 - \frac{k-x}{2q}\right) \end{aligned} \quad (2.27)$$

for

$$1 \leq k \leq K_e = \text{int}\left(2q + \frac{1}{2}\right) \quad (2.28)$$

The convolution of this window with X_{m+k} is

$$F(f) = \frac{e^{i\alpha(m+p)}}{S_e} \sum_{k=1}^{K_e} \left[\left(1 - \frac{k+x}{2q}\right) e^{-i\alpha k} + \left(1 - \frac{k-x}{2q}\right) e^{+i\alpha(k-1)} \right] \quad (2.29)$$

Equations 2.25 and 2.28 show that if $\text{frpt}(2q) < \frac{1}{2}$ then $K = K_e$ and the odd number of points forming the window when $|x|$ is small is greater than the even number of points for larger values of $|x|$; and conversely in the case $\text{frpt}(2q) > \frac{1}{2}$, for which $K = K_e - 1$.

Equations 2.26 and 2.29 show that, unlike the boxcar window, the Bartlett window produces no phase discontinuity as the “target” time t_j advances with respect to the time-tags T_k of the secondary data set. At the limits of validity of condition 2.18, that is when $x = \pm \text{frpt}(2q)$ or $\pm(1 - \text{frpt}(2q))$, the number of points in the window changes, nevertheless, because the term $\left(1 - \frac{k \pm x}{2q}\right)$ is zero for $k = K_e$, equations 2.26 and 2.29 become equivalent for this value of x , and there is no discontinuity. Furthermore, each time equation 2.13 causes m to increment, x jumps simultaneously from $+\frac{1}{2}$ to $-\frac{1}{2}$, causing p to jump from $+1$ to 0 , and consequently the phase of the multiplicative factor $e^{i\alpha(m+p)}$ in equation 2.29 suffers no discontinuity. The phase of the resampled datum increases continuously as t_j increases, for all values of t_j .

To examine the effects of the Bartlett window upon the spectrum Equations 2.26 and 2.29 may be evaluated.

Bibliography

For further reading, see the references in Chapter 1.

— 3 —

Multi-Spacecraft Filtering: General Framework

JEAN-LOUIS PINÇON
*LPCE/CNRS
Orléans, France*

UWE MOTSCHMANN
*Institute for Geophysics and Meteorology
Braunschweig, Germany*

3.1 Introduction

Measurements of the electric and magnetic fields in space plasmas commonly show fluctuations in time and space on all observed scales. Single satellite measurements generally do not allow disentangling spatial and temporal variations. With two satellites the ambiguity is removed only for simple motions of essentially one-dimensional structure. The determination of the shape and dynamics of three-dimensional structures requires a minimum of four spacecraft arranged in a tetrahedral configuration and equipped with instruments measuring fields and flows in three dimensions. Multi-satellite measurements open fundamental new possibilities to analyse spatial plasma structures in the magnetosphere and in the solar wind.

Although the analysis of multipoint data has been an increasingly important activity, there has been little effort to systematically analyse *in situ* data from multiple spacecraft except in the context of event studies. In this chapter we describe a multi-spacecraft data analysis technique allowing the identification of three-dimensional electromagnetic structures in the wave field. The organisation of the chapter is as follows: in Section 3.2 the technique is presented; in Sections 3.3 and 3.4 the limitations related respectively to the field and the experimental constraints are examined. Three applications using synthetic data are presented in Section 3.5.

3.2 Multi-Spacecraft Filtering Technique

3.2.1 Spectral Representation of the Wave Field

Hereafter we consider a multi-spacecraft experiment composed of N satellites. We denote as $A(t, \mathbf{r}_\alpha)$ the column vector consisting of the N field vectors measured at the satellite positions \mathbf{r}_α , $\alpha \in \{1, \dots, N\}$. $A(t, \mathbf{r})$ is assumed to be composed of L components. If we consider the case of electric and magnetic field measurements, then $L = 6$

and we have

$$\mathbf{A}(t, \mathbf{r}) = \begin{bmatrix} E_x(t, \mathbf{r}) \\ E_y(t, \mathbf{r}) \\ E_z(t, \mathbf{r}) \\ B_x(t, \mathbf{r}) \\ B_y(t, \mathbf{r}) \\ B_z(t, \mathbf{r}) \end{bmatrix} \quad (3.1)$$

The spectral representation of $\mathbf{A}(t, \mathbf{r})$ in the (ω, \mathbf{k}) domain is given by

$$\mathbf{A}(t, \mathbf{r}) = \int_{\omega} \int_{\mathbf{k}} \mathbf{A}(\omega, \mathbf{k}) e^{i(\omega t - \mathbf{k} \cdot \mathbf{r})} d\mathbf{k} d\omega \quad (3.2)$$

The high time resolution of the experiments generally allows the determination of the spectral amplitude with respect to the angular frequency $\mathbf{A}(\omega, \mathbf{r})$ by a Fourier transformation. Equation 3.2 becomes

$$\mathbf{A}(\omega, \mathbf{r}) = \int_{\mathbf{k}} \mathbf{A}(\omega, \mathbf{k}) e^{-i\mathbf{k} \cdot \mathbf{r}} d\mathbf{k} \quad (3.3)$$

The correlation matrix for two measurements is of particular importance for the characterisation of the field. It is constructed as the dyadic product of two measurement vectors $\mathbf{A}(t, \mathbf{r}_{\alpha})$ and $\mathbf{A}(t, \mathbf{r}_{\beta})$, where \mathbf{r}_{α} and \mathbf{r}_{β} refer to spacecraft positions. It is written

$$\mathbf{M}_A(\omega, \mathbf{r}_{\alpha}, \mathbf{r}_{\beta}) = E \left[\mathbf{A}(\omega, \mathbf{r}_{\alpha}) \mathbf{A}^{\dagger}(\omega, \mathbf{r}_{\beta}) \right] \quad (3.4)$$

where $E[\dots]$ stands for the mathematical expectation and indicates an ensemble average over a large number of distinct realisations of the data set. The superscript \dagger stands for hermitian adjoint (transpose and complex conjugation). Assuming the field is homogeneous in space which means the statistics of the fluctuating field are translation invariant in space, we have

$$\mathbf{M}_A(\omega, \mathbf{r}_{\alpha\beta}) = \int_{\mathbf{k}} \mathbf{S}_A(\omega, \mathbf{k}) e^{-i\mathbf{k} \cdot \mathbf{r}_{\alpha\beta}} d\mathbf{k} \quad \mathbf{r}_{\alpha\beta} = \mathbf{r}_{\alpha} - \mathbf{r}_{\beta} \quad (3.5)$$

$$\text{with} \quad \mathbf{S}_A(\omega, \mathbf{k}) = E \left[\mathbf{A}(\omega, \mathbf{k}) \mathbf{A}^{\dagger}(\omega, \mathbf{k}) \right] \quad (3.6)$$

and the trace of $\mathbf{S}_A(\omega, \mathbf{k})$ is the spectral energy density $P(\omega, \mathbf{k})$ we are looking for.

3.2.2 Notations and Data Preparation

In order to cope with multi-spacecraft measurements, and for the sake of simplicity, several definitions and notations are required. We define $\mathbf{A}(\omega)$ the $(NL \text{ by } 1)$ single column vector containing the $\mathbf{A}(\omega, \mathbf{r}_{\alpha})$ measured by the N spacecraft, namely

$$\mathbf{A}(\omega) = \begin{bmatrix} \mathbf{A}(\omega, \mathbf{r}_1) \\ \mathbf{A}(\omega, \mathbf{r}_2) \\ \vdots \\ \mathbf{A}(\omega, \mathbf{r}_N) \end{bmatrix} \quad (3.7)$$

then the relation between $A(\omega)$ and the spectral amplitude $A(\omega, \mathbf{k})$ is given by,

$$A(\omega) = \int_{\mathbf{k}} H(\mathbf{k}) A(\omega, \mathbf{k}) d\mathbf{k} \quad (3.8)$$

$$\text{with} \quad H(\mathbf{k}) = \begin{bmatrix} \mathbf{I} e^{-i\mathbf{k} \cdot \mathbf{r}_1} \\ \mathbf{I} e^{-i\mathbf{k} \cdot \mathbf{r}_2} \\ \vdots \\ \mathbf{I} e^{-i\mathbf{k} \cdot \mathbf{r}_N} \end{bmatrix} \quad (3.9)$$

\mathbf{I} is an (L by L) unit matrix and thus $H(\mathbf{k})$ is an (LN by L) matrix. Hereafter we use the symbol \mathbf{I} for all unit matrices whatever their rank. Now we introduce the (LN by LN) covariance matrix $M_A(\omega)$. It contains all the correlation matrices that can be estimated from the N spacecraft experiment and is defined by

$$M_A(\omega) = E \left[A(\omega) A^\dagger(\omega) \right] \quad (3.10)$$

This matrix contains all the measured data, and thus is a known quantity in our context. Theoretically, the determination of M_A requires the computation of the mathematical expectation, that is, the averaging of different realisations of the measured field vector $A(\omega)$. Assuming the process is ergodic, this ensemble average may be replaced by averaging in time, and one obtains

$$M_A(\omega) = \frac{1}{Q} \sum_{q=1}^Q A^q(\omega) A^{q\dagger}(\omega) \quad (3.11)$$

To get Q convenient measurements $A^q(\omega)$, the time series of measurements $A(t)$ is divided in Q subintervals. Fourier transformation in each subinterval q yields $A^q(\omega)$. According to the Fourier transformation theory, the length of a subinterval determines the frequency resolution.

From equation 3.5, the two matrices M_A and S_A are related by

$$M_A(\omega) = \int_{\mathbf{k}} H(\mathbf{k}) S_A(\omega, \mathbf{k}) H^\dagger(\mathbf{k}) d\mathbf{k} \quad (3.12)$$

The next section describes the multi-spacecraft filtering technique that allows us to obtain an optimal estimation of the field spectral energy density $P(\omega, \mathbf{k})$ from the matrix $M_A(\omega)$.

3.2.3 Filter Bank Approach and $P(\omega, \mathbf{k})$ Estimation

We shall adopt a filter-bank approach to find the appropriate combination of multi-spacecraft measurements for an optimum description of the field in the angular frequency and wave vector domains. Thus, the main objective is the determination of filters, each one being related to a different (ω, \mathbf{k}) pair and characterised by a matrix $F(\omega, \mathbf{k})$. For each (ω, \mathbf{k}) the purpose is to design the corresponding filter in such a way that having multi-spacecraft measurements $A(\omega)$ as input, it provides an optimum $A(\omega, \mathbf{k})$ estimation given by

$$A(\omega, \mathbf{k}) = F^\dagger(\omega, \mathbf{k}) A(\omega) \quad (3.13)$$

Consequently F^\dagger is an $(L \text{ by } LN)$ rectangular matrix. Taking the dyadic product of equation 3.13 with its hermitian adjoint, and then taking the expectation value, we connect the three matrices S_A , M_A , and F .

$$S_A(\omega, \mathbf{k}) = F^\dagger(\omega, \mathbf{k}) M_A(\omega) F(\omega, \mathbf{k}) \quad (3.14)$$

If no additional *a priori* information is available, a unique determination of the rectangular matrix $F(\omega, \mathbf{k})$ is achieved by requiring that the filter shall absorb all energy contained in M_A except that related with the angular frequency ω and the wave vector \mathbf{k} . This requirement is satisfied by minimising the trace of the matrix $S_A(\omega, \mathbf{k})$ with the constraint that any plane wave whose angular frequency is ω and wave vector is \mathbf{k} is passed undisturbed through the filter. Then the filter determination can be formulated as

$$\begin{aligned} P(\omega, \mathbf{k}) &= \text{Tr} \{ F^\dagger(\omega, \mathbf{k}) M_A(\omega) F(\omega, \mathbf{k}) \} = \text{minimum} \\ \text{with} \quad F^\dagger(\omega, \mathbf{k}) H(\mathbf{k}) A(\omega, \mathbf{k}) &= A(\omega, \mathbf{k}) \end{aligned} \quad (3.15)$$

This problem can be solved using the Lagrange multiplier technique. The detailed derivation of the spectral energy density $P(\omega, \mathbf{k})$ is presented in the Appendix. We obtain the following expression

$$P(\omega, \mathbf{k}) = \text{Tr} \left\{ \left[H^\dagger(\mathbf{k}) M_A^{-1}(\omega) H(\mathbf{k}) \right]^{-1} \right\} \quad (3.16)$$

If no *a priori* information is available, this expression is an optimum estimator of the field spectral energy density. It is optimum in the sense that it provides a maximum likelihood estimate of the frequency wave-vector power spectrum, provided the noise associated with the field measurements is a multi-dimensional Gaussian process. If any given specific knowledge about the signal is available, one can take it into account during the determination of the filter matrix F . This leads to a modification of the filter constraint and therefore a new expression of the optimum estimator. The derivation of $P(\omega, \mathbf{k})$ estimators, including such additional *a priori* information, is presented in Section 3.5 for three different applications.

Estimators obtained using the filter bank approach generally do not require much computation time. For instance, for the estimator corresponding to equation 3.16, once the $M_A(\omega)$ matrix has been estimated from the measured waveforms, the computations consist of multiplications and inversions of complex matrices. The largest matrix to invert is $M_A(\omega)$ whose rank is $L \times N$. For a multi-spacecraft mission consisting of $N = 4$ satellites and with $L = 6$ (3 electric plus 3 magnetic wave-field components), the rank of M_A is 24. Thus we are not in the case of a large matrix for which specific algorithms must be developed.

3.3 Limitations Related to the Field

The existence of the spectral energy density estimator is based on the hypotheses of time stationarity and space homogeneity of the measured field. Obviously, none of those hypotheses is completely fulfilled in a real experiment. The conditions under which they may be considered to be satisfied adequately are discussed below.

The validity of this estimator is also limited by an additional hypothesis requiring that the measured field be free of characteristic wavelengths smaller than the minimum inter-spacecraft distance. Otherwise an aliasing effect develops, which is similar to the one observed in the case of under-sampling of time series.

3.3.1 Time Stationarity and Space Homogeneity

Actually, an approximate temporal stationarity of the field is sufficient to provide reasonable data statistics. For all practical purposes only time stationarity during time intervals much longer than the maximum period studied in the field is necessary. A well-known statistical test [*Bendat and Piersol, 1971*] may be used to check the limited temporal stationarity of the data.

Strict conditions for homogeneity cannot be met in space, particularly in the vicinity of geophysical boundaries such as the bow shock or the magnetopause. In our case a statement of “limited homogeneity” is sufficient. It is fulfilled when the field is translation invariant over distances much larger than the maximum wavelength studied in the field. In the frame of a multi-spacecraft experiment, the translation invariance of the field can only be tested along the direction of the spacecraft trajectory. Data sets measured at close positions along the orbit may be used to check the invariance of the $P(\omega, \mathbf{k})$ solutions obtained.

3.3.2 Spatial Aliasing

This problem is also discussed in Section 14.5.1 of Chapter 14. The origin of the spatial aliasing comes from the fact that the spacecraft configuration does not distinguish two plane waves differing only by their wave vectors in such a way that

$$\Delta \mathbf{k} \cdot \mathbf{r}_\alpha = 2\pi n_\alpha + \phi \quad \forall \alpha \in \{1, \dots, N\} \quad (3.17)$$

where n_α are signed integers. For $N = 4$ it can be shown [*Neubauer and Glassmeier, 1990*] that $\Delta \mathbf{k}$ has the solutions

$$\Delta \mathbf{k} = \sum_{l=1}^{l=3} n_l \Delta \mathbf{k}_l \quad (3.18)$$

with the $\Delta \mathbf{k}_l$ given by

$$\begin{aligned} \Delta \mathbf{k}_1 &= \frac{2\pi}{V} \mathbf{r}_{31} \times \mathbf{r}_{21}; & \Delta \mathbf{k}_2 &= \frac{2\pi}{V} \mathbf{r}_{41} \times \mathbf{r}_{21}; & \Delta \mathbf{k}_3 &= \frac{2\pi}{V} \mathbf{r}_{41} \times \mathbf{r}_{31} \\ V &= \mathbf{r}_{41} \cdot (\mathbf{r}_{31} \times \mathbf{r}_{21}) \end{aligned} \quad (3.19)$$

The impossibility of distinguishing these plane waves is referred to as “spatial aliasing”. This spatial aliasing would be absent, and consequently the $P(\omega, \mathbf{k})$ estimation not distorted, if the characteristic lengths of the field correspond to wave vectors included inside the subvolume described by

$$\mathbf{k} = \sum_{l=1}^{l=3} e_l \Delta \mathbf{k}_l \quad \text{with} \quad -0.5 < e_l \leq 0.5 \quad (3.20)$$

This restriction corresponds roughly to the hypothesis of a field free of wavelengths smaller than the minimum inter-spacecraft distance. In the absence of any *a priori* information related to the minimum characteristic length of the field, there is no satisfactory way to identify spatial aliasing from multi-spacecraft data. Indeed, to insure the validity of the $P(\omega, \mathbf{k})$ estimator, the solution would be to remove from the data the part of the field related to the small wavelengths before performing the analysis. Unfortunately, such a filtering process would require a large number of measuring points not compatible with a multi-spacecraft mission. Actually, the motion of the satellites in space can be used to identify the presence of short wavelength disturbances parallel to the main velocity (for instance one can rely on the consistency of successive analyses or on comparisons between successive measurements at short distances).

3.4 Limitations Related to the Experimental Constraints

As real data were not available when we wrote this chapter, we used synthetic data generated by numerical simulation to study the effects of multi-spacecraft experimental constraints on the spectral energy density estimator. We applied the $P(\omega, \mathbf{k})$ estimator to data sampled synchronously at different satellite position $\mathbf{r}_\alpha, \alpha \in \{1, \dots, N\}$. The satellite positions and the wave vectors were tuned to avoid spatial aliasing. By varying the parameters of the simulations we were able to examine the effects related to the configuration geometry and to the accuracy in the measurements (distance between spacecraft, time synchronisation).

3.4.1 The Spacecraft Configuration

An unambiguous determination of three-dimensional structures requires, as a minimum, four spacecraft in a three-dimensional configuration. Using simulated data, several 4-spacecraft configurations were studied. The best solutions are obtained with a tetrahedral geometry [Pinçon and Lefeuvre, 1991, 1992]. The simulations demonstrate the necessity of checking for the shape of the spacecraft configuration before interpreting the obtained solutions. This can be easily done using parameters describing the tetrahedron geometry (we refer the reader to Chapter 13).

3.4.2 Inaccuracy in the Time Synchronisation

The accuracy in the time synchronisation between the measurements performed on the different spacecraft is of prime importance. For a given frequency ω , a time inaccuracy δt introduces a phase shift $\phi = \omega \delta t$ in the estimation of the power spectra. Phase shifts greater than a few degrees can theoretically distort a $P(\omega, \mathbf{k})$ estimation. Actually, an exact threshold in the required time accuracy is not easy to define. It depends on the other errors in the data, and on the satellite configuration. A series of simulations, not shown here, has been performed by Pinçon and Lefeuvre [1992] to provide guidelines. From them, it seems reasonable to fix an empirical threshold at 30° . At and above this value, the fit with the model becomes very poor. The practical consequence of this is to limit the validity domain of the $P(\omega, \mathbf{k})$ estimator to the low frequency range. For instance, on board Cluster the electric and magnetic field waveform data will be available with an

accuracy of about $50 \mu s$. For such a value, if we fix the maximum phase shift at 5° degree, the corresponding maximum frequency is equal to 280 Hz.

3.4.3 Inaccuracy in the Inter-Spacecraft Distances

The effect of errors related to the inaccuracy in the inter-spacecraft distances is similar to the one produced by the time synchronisation inaccuracy. For a given wave vector \mathbf{k} in the measured wave field, a distance inaccuracy $\delta \mathbf{r}_{\alpha\beta}$ introduces a phase shift $\mathbf{k} \cdot \delta \mathbf{r}_{\alpha\beta}$ in the estimation of the power spectra. As previously, this effect can only be empirically evaluated from simulations. It has been shown by *Pinçon and Lefeuvre* [1992] that the upper limit of the relative error in the distance has to be fixed between 10% and 20%. Cluster would easily meet this requirement since the relative distance error is expected to be less than or equal to 10%.

3.5 Examples

The data analysis technique described in this chapter can be used to identify a large class of three-dimensional structures. To illustrate this point, three different applications are presented in this section. In the first example we applied the $P(\omega, \mathbf{k})$ estimator to synthetic multi-spacecraft measurements composed of the three magnetic plus the three electric components. In the second example a similar analysis is performed but now the data are the three magnetic components. The last example demonstrates that relatively slight modifications of the $P(\omega, \mathbf{k})$ estimator allow us to deal with the detection of surface waves. For all examples the specific nature of the signal analysed is taken into account.

We used synthetic wave fields consisting of superpositions of plane waves and incoherent noise. All plane waves propagate with different wave vectors at one frequency ω . The satellite positions and the wave vectors are tuned to avoid spatial aliasing. For reasons of presentation, the wave vectors are chosen in the plane $k_z = 0$. To avoid a loss of generality, a wave-field simulation including wave vectors distributed over the whole three dimensional domain would have been preferable. But presentation of the corresponding $P(\omega, \mathbf{k})$ would have required several plots. Moreover, limited trials showed that the k_x and k_y resolutions of the solutions obtained in the $k_z = 0$ case are similar to the k_x and k_y resolutions obtained in the three dimensional case by integrating $P(\omega, \mathbf{k})$ over all k_z . Hereafter the solutions are represented in the plane $k_z = 0$ where the field energy is concentrated. The crosses indicate the location of the \mathbf{k} vectors related to the modes introduced in the simulation. The $P(\omega, \mathbf{k})$ solutions are expected to present significant peaks at these locations. The actual solutions, obtained from the synthetic multi-spacecraft data, are represented by contour lines linearly scaled between the maximum and the minimum value.

3.5.1 $P(\omega, \mathbf{k})$ Estimator Related to the Electromagnetic Wave Field

First we determined the optimum filter related to the identification of electromagnetic plane waves [*Pinçon and Lefeuvre*, 1991]. We assumed that the signals $A(\omega, \mathbf{r}_\alpha)$ measured

by the spacecraft are the 6 electromagnetic components of the wave field.

$$A(\omega, \mathbf{r}_\alpha) = \begin{bmatrix} \mathbf{E}(\omega, \mathbf{r}_\alpha) \\ \mathbf{B}(\omega, \mathbf{r}_\alpha) \end{bmatrix} \quad \forall \alpha \in \{1, \dots, N\} \quad (3.21)$$

From the Maxwell-Faraday relation, the electric and magnetic components in the (ω, \mathbf{k}) domain are related through \mathbf{k} and ω by $\mathbf{k} \times \mathbf{E}(\omega, \mathbf{k}) = \omega \mathbf{B}(\omega, \mathbf{k})$. As a consequence we have $A(\omega, \mathbf{k}) = \mathbf{C1}(\omega, \mathbf{k}) \mathbf{E}(\omega, \mathbf{k})$ with

$$\mathbf{C1}(\omega, \mathbf{k}) = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \\ 0 & \frac{-k_z}{\omega} & \frac{k_y}{\omega} \\ \frac{k_z}{\omega} & 0 & \frac{-k_x}{\omega} \\ \frac{-k_y}{\omega} & \frac{k_x}{\omega} & 0 \end{bmatrix} \quad (3.22)$$

Taking into account the *a priori* information coming from the electromagnetic nature of the field analysed and the Maxwell-Faraday relation, the new filter constraint is

$$\mathbf{F}^\dagger(\omega, \mathbf{k}) \mathbf{H}(\mathbf{k}) \mathbf{C1}(\omega, \mathbf{k}) = \mathbf{C1}(\omega, \mathbf{k}) \quad (3.23)$$

Solving the minimisation we find

$$P1(\omega, \mathbf{k}) = \text{Tr} \left\{ \mathbf{C1}(\omega, \mathbf{k}) \left[\mathbf{C1}^\dagger(\omega, \mathbf{k}) \mathbf{H}^\dagger(\mathbf{k}) \mathbf{M}_A^{-1}(\omega) \mathbf{H}(\mathbf{k}) \mathbf{C1}(\omega, \mathbf{k}) \right]^{-1} \mathbf{C1}^\dagger(\omega, \mathbf{k}) \right\} \quad (3.24)$$

Figure 3.1a shows the $P1(\omega, \mathbf{k})$ solution obtained with the four satellites arranged in a tetrahedral geometry. The wave vectors of the nine plane waves used for the simulation have been chosen in a way to avoid spatial aliasing. These wave vectors are quite well identified, the discrepancies between the peaks and the crosses are very small.

An illustration of the effects of multi-spacecraft experimental constraints on the $P1(\omega, \mathbf{k})$ solution is given by the Figures 3.1b, c, and d. Figure 3.1b shows the solution obtained with the four satellites arranged in a linear geometry parallel to the x axis. In this case the resolution following the y axis is very low and the fit with the model becomes very poor. Figure 3.1c shows the solution obtained with the four spacecraft arranged in a tetrahedral configuration. To simulate inaccuracy in time synchronisation, random phase shifts between -30° and $+30^\circ$ were introduced in the data. Figure 3.1d shows the solution obtained with the four spacecraft arranged in a tetrahedral configuration. The inaccuracy in the distance measurements is taken into account by imposing a relative error $(|\delta \mathbf{r}_{\alpha\beta}| / |\mathbf{r}_{\alpha\beta}|)$ equal to 20%.

Figure 3.2 demonstrates the effects of aliasing on the $P1(\omega, \mathbf{k})$ solution obtained with four satellites arranged in a tetrahedral geometry. The wave field analysed is the one used to obtain the Figure 3.1 with one more electromagnetic plane wave. The corresponding tenth \mathbf{k} vector has been deliberately chosen to introduce aliasing ($\mathbf{k}_{10}^\dagger = [-10, -1, 0]$). As pointed out in Section 3.3.2, the $P1(\omega, \mathbf{k})$ solution is not valid any more. This is clearly illustrated by the presence of a spurious peak located at $k_x = -0.7$ and $k_y = -1.2$.

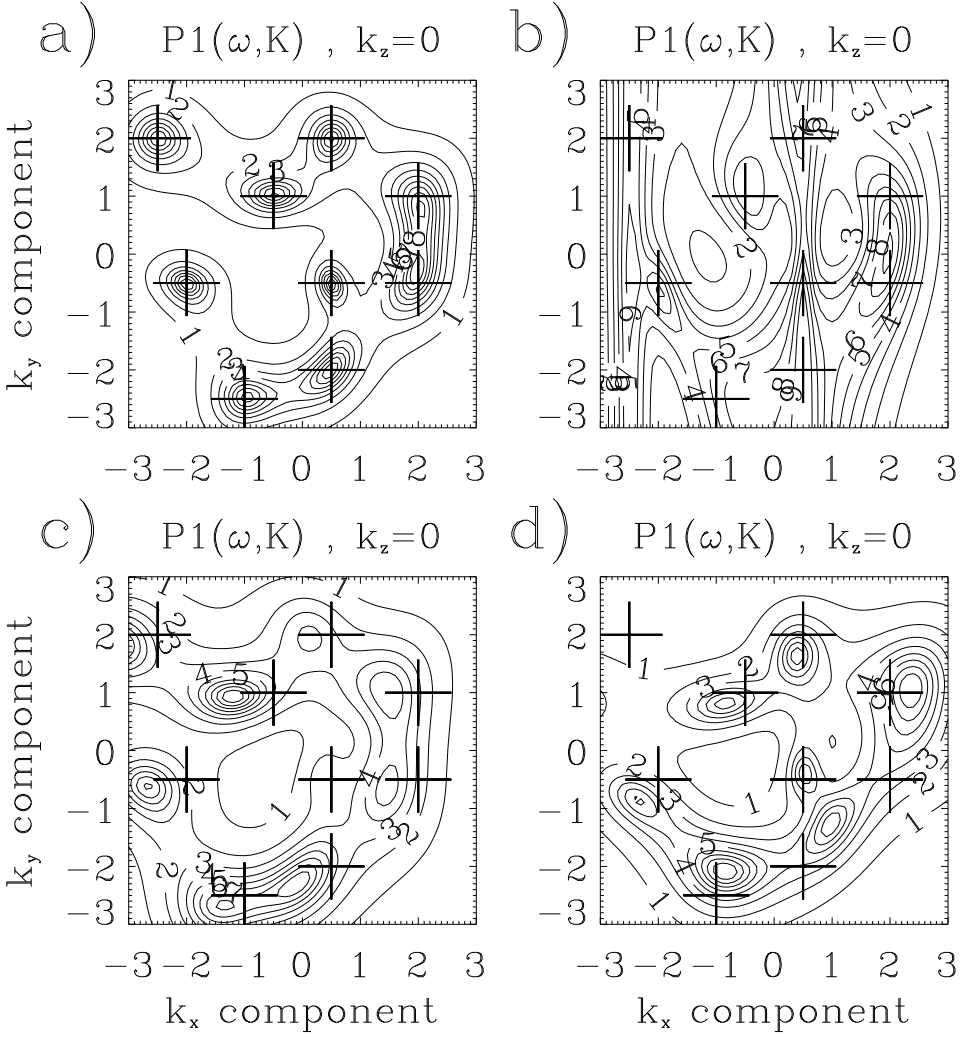


Figure 3.1: $P1(\omega, \mathbf{k})$ solutions obtained from four spacecraft for a wave field consisting of a superposition of nine electromagnetic plane waves. The crosses indicate the location of the nine \mathbf{k} vectors related to the plane waves. The satellite positions and the wave vectors are tuned to avoid spatial aliasing. The solutions are represented by contour lines linearly scaled between the maximum and the minimum. Solutions obtained using: a) a tetrahedral geometry; b) a linear geometry (parallel to the x axis; c) a tetrahedral geometry with random phase shifts between -30° and $+30^\circ$ to simulate time synchronisation inaccuracy; d) a tetrahedral geometry in the case of a 20% inter-spacecraft distance inaccuracy.

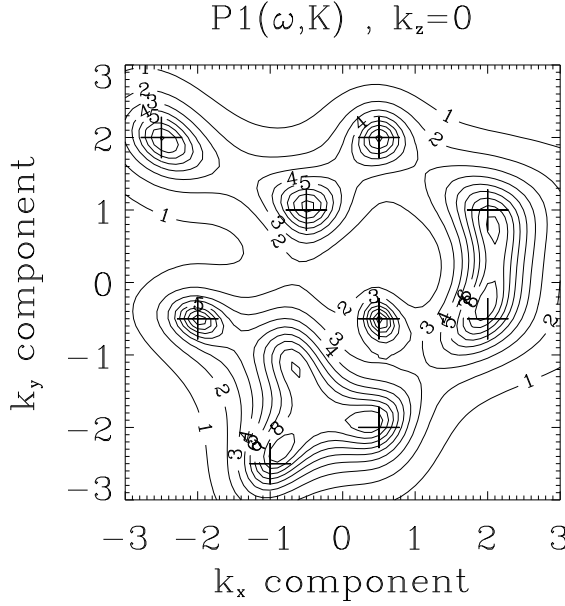


Figure 3.2: $P1(\omega, \mathbf{k})$ solutions obtained using a tetrahedral geometry for a wave field consisting of a superposition of the wave field of Figure 3.1 and one more electromagnetic plane wave. The additional \mathbf{k} vector has been chosen to introduce aliasing. The solution presents a spurious peak at $k_x = -0.7$ and $k_y = -1.2$.

3.5.2 $P(\omega, \mathbf{k})$ Estimator Related to the Magnetic Field

As a second illustration we determined the optimum filter related to the identification of magnetic plane waves [Motschmann et al., 1996]. We assumed that the signals $A(\omega, \mathbf{r}_\alpha)$ measured by the spacecraft are the 3 magnetic components.

$$A(\omega, \mathbf{r}_\alpha) = \mathbf{B}(\omega, \mathbf{r}_\alpha) \quad \forall \alpha \in \{1, \dots, N\} \quad (3.25)$$

Since the magnetic field is divergence-free, the magnetic components in the (ω, \mathbf{k}) domain are related to \mathbf{k} by $\mathbf{k} \cdot \mathbf{B}(\omega, \mathbf{k}) = 0$. As a consequence we have $\mathbf{B}(\omega, \mathbf{k}) = \mathbf{C2}(\omega, \mathbf{k}) \mathbf{B}(\omega, \mathbf{k})$, where

$$\mathbf{C2}(\omega, \mathbf{k}) = \mathbf{I} + \frac{\mathbf{k}\mathbf{k}^\dagger}{|\mathbf{k}|^2} \quad (3.26)$$

By taking into account the divergence-free nature of the measured wave field, the constraint can be rewritten as

$$\mathbf{F}^\dagger(\omega, \mathbf{k}) \mathbf{H}(\mathbf{k}) \mathbf{C2}(\omega, \mathbf{k}) = \mathbf{I} \quad (3.27)$$

With this new constraint the minimisation described by equation 3.15 yields

$$P2(\omega, \mathbf{k}) = \text{Tr} \left\{ \left[\mathbf{C2}^\dagger(\omega, \mathbf{k}) \mathbf{H}^\dagger(\mathbf{k}) \mathbf{M}_A^{-1}(\omega) \mathbf{H}(\mathbf{k}) \mathbf{C2}(\omega, \mathbf{k}) \right]^{-1} \right\} \quad (3.28)$$

The result of the application of $P2(\omega, \mathbf{k})$ to synthetic data is shown in Figure 3.3. We assumed a tetrahedron-like configuration, a perfect time synchronisation and no error in

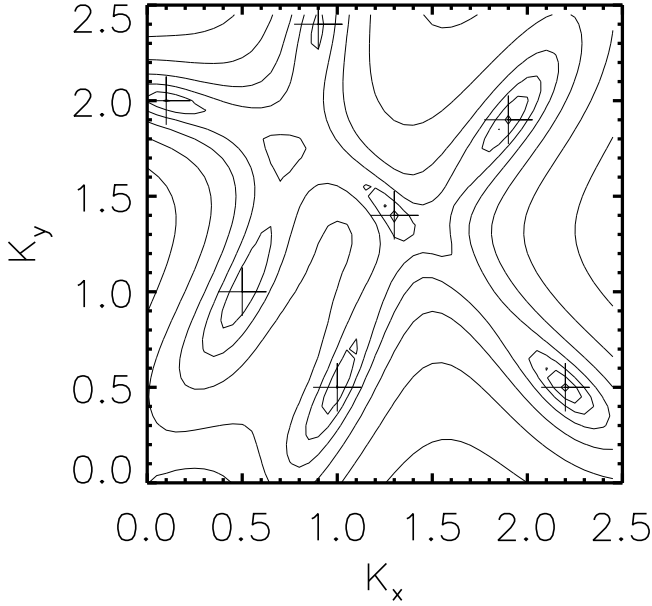


Figure 3.3: $P2(\omega, \mathbf{k})$ solution for the plane $k_z = 0$ obtained using a tetrahedral geometry. The wave field consists of a superposition of seven plane waves. [Reproduced from [Motschmann et al., 1996](#).]

the inter-spacecraft distances. The $P2(\omega, \mathbf{k})$ estimator has no problem finding the seven magnetic plane waves used in the simulation.

3.5.3 Electromagnetic Surface Wave Detector (SWD)

The last example is somewhat different from the two previous. Instead of a plane wave detector, we determined a filter allowing the identification of electromagnetic surface waves. We assumed that the signals measured by the spacecraft are the six electromagnetic components of the wave field.

$$\mathbf{A}(\omega, \mathbf{r}_\alpha) = \begin{bmatrix} \mathbf{E}(\omega, \mathbf{r}_\alpha) \\ \mathbf{B}(\omega, \mathbf{r}_\alpha) \end{bmatrix} \quad \forall \alpha \in \{1, \dots, N\} \quad (3.29)$$

A link with the previous $P1(\omega, \mathbf{k})$ estimator can be found by noting that an electromagnetic surface wave can be described as an electromagnetic plane wave with complex wave vector ($\mathbf{k} = \mathbf{k}_r + i\mathbf{k}_i$). The spatial inhomogeneity associated with the surface waves is included in the filter by choosing one of the satellite position as a reference and rewriting the H matrix as

$$\mathbf{H3}(\mathbf{k}) = \begin{bmatrix} \exp(-i\mathbf{k}_r \cdot \mathbf{r}_1) \exp(\mathbf{k}_i \cdot (\mathbf{r}_1 - \mathbf{r}_1)) \\ \vdots \\ \exp(-i\mathbf{k}_r \cdot \mathbf{r}_N) \exp(\mathbf{k}_i \cdot (\mathbf{r}_N - \mathbf{r}_1)) \end{bmatrix} \quad (3.30)$$

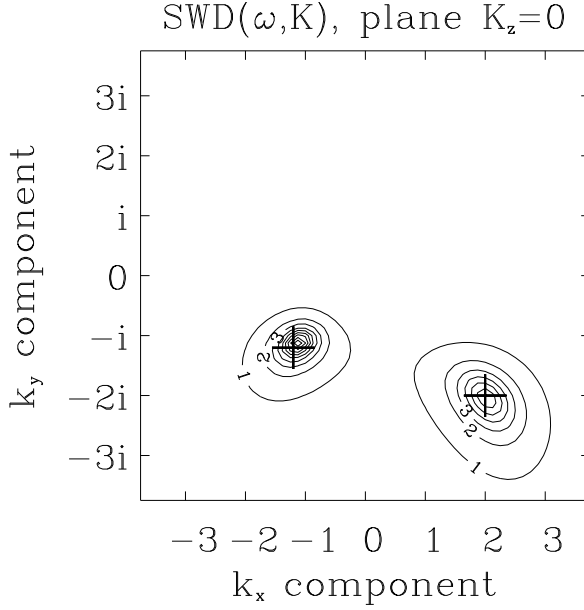


Figure 3.4: $SWD(\omega, \mathbf{k})$ solution for the complex plane ($k_x, ik_y, k_z = 0$) obtained using a triangular geometry; see text. [Reproduced from [Pinçon, 1995](#).]

This way the $H3$ matrix contains not only the relative phase information at all spacecraft positions, but also a model of the inhomogeneity between spacecraft. The previous hypothesis of spatial homogeneity is here replaced by the hypothesis of spatial inhomogeneity exclusively related to surface waves. Using the Maxwell-Faraday relation we still can relate the electric and magnetic components in the (ω, \mathbf{k}) domain through \mathbf{k} and ω . We have $\mathbf{A}(\omega, \mathbf{k}) = \mathbf{C3}(\omega, \mathbf{k}) \mathbf{E}(\omega, \mathbf{k})$ with $\mathbf{C3}$ similar to $\mathbf{C1}$ but \mathbf{k} is complex. The optimum (ω, \mathbf{k}) -filter is obtained by demanding that it absorbs all energy contained in $\mathbf{M}_A(\omega)$ except that corresponding to a surface wave characterised by the frequency ω and the complex wave vector \mathbf{k} .

$$SWD(\omega, \mathbf{k}) = \text{Tr} \left\{ \mathbf{F}^\dagger(\omega, \mathbf{k}) \mathbf{M}_A(\omega) \mathbf{F}(\omega, \mathbf{k}) \right\} = \text{minimum} \quad (3.31)$$

with $\mathbf{F}^\dagger(\omega, \mathbf{k}) \mathbf{H3}(\mathbf{k}) \mathbf{C3}(\omega, \mathbf{k}) = \mathbf{C3}(\omega, \mathbf{k})$

Solving the minimisation problem we find

$$SWD(\omega, \mathbf{k}) = \text{Tr} \left\{ \mathbf{C3}(\omega, \mathbf{k}) \left[\mathbf{C3}^\dagger(\omega, \mathbf{k}) \mathbf{H3}^\dagger(\mathbf{k}) \mathbf{M}_A^{-1}(\omega) \mathbf{H3}(\mathbf{k}) \mathbf{C3}(\omega, \mathbf{k}) \right]^{-1} \mathbf{C3}^\dagger(\omega, \mathbf{k}) \right\} \quad (3.32)$$

Figure 3.4 shows the SWD solution obtained using data generated by a 2-D MHD code. The simulated spacecraft are arranged in a triangular configuration, they are moving through the simulation domain with a constant velocity. We assumed a perfect time synchronisation and no errors in the inter-spacecraft distances. The simulated wave field is composed of two MHD electromagnetic surface waves with the same frequency and different complex vectors. The solution is represented in the complex plane (k_x real, k_y

imaginary, $k_z = 0$) where the energy of the simulated field is concentrated. The two surface waves of the simulation are clearly identified by the surface wave detector.

3.6 Summary

The multi-spacecraft data filtering technique described in this chapter allows the identification of three-dimensional structures in the wave field. The two important cases of plane waves- and surface waves-identification are examined. The characterisation is achieved through an estimation of the wave field spectral energy density in the angular frequency and wave vector domains. Using synthetic data we have validated the usefulness of this approach.

The list of applications presented in this chapter is not exhaustive. For instance it is also possible using this technique to develop a large scale current detector from magnetic field measurement. Having in mind a precise objective, and relevant multi-spacecraft data being available, the filter bank approach can be used to design filters specifically adapted to the problem. The performance of the estimators derived using the multi-spacecraft filtering technique is strongly related to the amount of *a priori* information included in the filter design. Not much can be said about this point, each case being a particular case.

The main limitation is related to the characteristic field lengths which have to be larger than the mean inter-spacecraft distance to avoid aliasing effects. To obtain a three-dimensional characterisation, simultaneous measurements on a minimum of four points in a tetrahedral configuration are required. Other limitations are related to the hypothesis of space homogeneity and the accuracy with which the multi-spacecraft configuration is defined in space and time.

The multi-spacecraft filtering technique does not require much computation time and can be used to identify a large class of different structures. The applications studied have shown that the resolution obtained in the \mathbf{k} domain is sufficient to provide an accurate characterisation of the wave field. This remains true even in the case of an experiment consisting of only four measuring points in space.

Appendix: Derivation of $P(\omega, \mathbf{k})$

The problem of the filter determination can be formulated as

$$\begin{aligned} P(\omega, \mathbf{k}) = \text{Tr} \{ \mathbf{F}^\dagger \mathbf{M} \mathbf{F} \} = \text{minimum} \\ \text{with } \mathbf{F}^\dagger \mathbf{H} = \mathbf{I} \end{aligned} \quad (3.33)$$

Since \mathbf{F} is a complex matrix each element has a real and an imaginary part, thus

$$\begin{aligned} F_{ij} &= X_{ij} + iY_{ij} \\ (F^\dagger)_{ji} &= X_{ij} - iY_{ij} \end{aligned} \quad (3.34)$$

X_{ij} and Y_{ij} are independent variables, consequently F_{ij} and $(F^\dagger)_{ji}$ are independent variables too. In what follows we treat \mathbf{F} and \mathbf{F}^\dagger as independent matrices. Then, using the Lagrange multiplier technique, to solve the filter determination problem formulated by equation 3.33 is equivalent to minimise

$$P(\omega, \mathbf{k}) = \text{Tr} \left\{ \mathbf{F}^\dagger \mathbf{M}_A(\omega) \mathbf{F} + \Lambda \left(\mathbf{I} - \mathbf{H}^\dagger \mathbf{F} \right) + \left(\mathbf{I} - \mathbf{F}^\dagger \mathbf{H} \right) \Gamma \right\} \quad (3.35)$$

where Λ and Γ are the Lagrangian multiplier matrices related to constraints involving \mathbf{F} and \mathbf{F}^\dagger , respectively. Equation 3.35 written in terms of components reads

$$P(\omega, \mathbf{k}) = (F^\dagger)_{ij} M_{jk} F_{ki} + \Lambda_{ii} - \Lambda_{ij} (H^\dagger)_{jk} F_{ki} + \Gamma_{ii} - (F^\dagger)_{ij} H_{jk} \Gamma_{ki} \quad (3.36)$$

Equating to 0 the partial derivative of this expression with respect to F_{ki} and $(F^\dagger)_{ij}$, yields the following results

$$\frac{\partial P(\omega, \mathbf{k})}{\partial F_{ki}} = 0 \longrightarrow \mathbf{F}^\dagger \mathbf{M} = \Lambda \mathbf{H}^\dagger \quad (3.37)$$

$$\frac{\partial P(\omega, \mathbf{k})}{\partial (F^\dagger)_{ij}} = 0 \longrightarrow \mathbf{M} \mathbf{F} = \mathbf{H} \Gamma \quad (3.38)$$

Then we multiply equation 3.37 by \mathbf{F} and equation 3.38 by \mathbf{F}^\dagger . From the definition of P and the constraint of equation 3.33, we get

$$\Lambda = \Gamma \quad \text{and} \quad P(\omega, \mathbf{k}) = \text{Tr} \{ \Gamma \} \quad (3.39)$$

Provided \mathbf{M} is not singular, the expression of Γ is obtained by multiplying equation 3.38 by $\mathbf{H}^\dagger \mathbf{M}^{-1}$, thus

$$\Gamma = [\mathbf{H}^\dagger \mathbf{M}^{-1} \mathbf{H}]^{-1} \quad (3.40)$$

Finally, substituting this into equation 3.39, we obtain an expression for the spectral energy density P

$$P(\omega, \mathbf{k}) = \text{Tr} \left\{ [\mathbf{H}^\dagger \mathbf{M}^{-1} \mathbf{H}]^{-1} \right\} \quad (3.41)$$

Acknowledgements

We thank Steven J. Schwartz, Götz Paschmann, and the referee for helpful comments and valuable suggestions.

Bibliography

- Bendat, J. S. and Piersol, A. G., *Random Data: Analysis and Measurement Procedures*, John Wiley, New York, 1971.
- Motschmann, U., Woodward, T. I., Glassmeier, K.-H., Southwood, D. J., and Pinçon, J.-L., Wavelength and direction filtering by magnetic measurements at satellite arrays: Generalized minimum variance analysis, *J. Geophys. Res.*, **101**, 4961–4965, 1996.
- Neubauer, F. and Glassmeier, K.-H., Use of an array of satellites as a wave telescope, *J. Geophys. Res.*, **95**, 19 115–19 122, 1990.
- Pinçon, J.-L. and Lefeuvre, F., Local characterization of homogeneous turbulence in a space plasma from simultaneous measurements of field components at several points in space, *J. Geophys. Res.*, **96**, 1789–1802, 1991.
- Pinçon, J.-L. and Lefeuvre, F., The application of the generalized Capon method to the analysis of a turbulent field in space plasma: Experimental constraints, *J. Atmos. Terr. Phys.*, **54**, 1237–1247, 1992.
- Pinçon, J.-L., Cluster and the k -filtering, in *Proceedings of Cluster Workshops, Braunschweig, 28–30 Sep. 1994, Toulouse, 16–17 Nov. 1994*, ESA SP-371, pp. 87–94, European Space Agency, Paris, France, 1995.

— 4 —

Multi-Spacecraft Filtering: Plasma Mode Recognition

UWE MOTSCHMANN AND KARL-HEINZ GLASSMEIER

*Technische Universität Braunschweig
Braunschweig, Germany*

JEAN-LOUIS PINÇON

*LPCE/CNRS
Orleans, France*

4.1 Introduction

Future multi-satellite space missions will provide us with a wealth of data requiring a very careful data analysis. However, data analysis is an art, the art of seeking hidden knowledge in a complex set of observations. New methods of pattern recognition are thus necessary to handle and interpret future missions data. It is the purpose of this chapter to outline a new method for the determination of the general polarisation properties of any given set of multi-instrument and multipoint measurements of wave fields in space plasmas. The method described is a mode decomposition technique suggested by *Glassmeier et al.* [1995]. As this method requires the knowledge of the wave propagation direction a generalisation of the wavelength and direction filtering technique described by *Pinçon and Lefeuvre* [1991] and *Motschmann et al.* [1996] is discussed, which also allows a validation of the plasma model used in the decomposition technique.

The method discussed is a *research* tool in that it makes use of *a priori* assumptions on the physical system to be studied and the models to describe it. In this the method resembles similarities with the classical Fourier analysis where the assumption of a time or space harmonic process is used to decompose any time or space series into its Fourier components. In a more general way, any signal can be decomposed into a prescribed set of basis patterns, which could be the MHD eigenmodes, for example. In this respect, the mode decomposition technique is a kind of generalisation of the Fourier analysis with the constraints applied allowing one to filter those parts of the signal to be analysed which can be described by the plasma model used.

4.2 Mode Decomposition of MHD Wave Fields

The mode decomposition method allows the calculation of the energy per eigenmode of a given wave field for a given position in the ω - k space. The method has similarities

with the pure state algorithm presented by *Samson* [1983]. The wave field to be analysed will be expanded in its set of eigenmodes. This requires the selection of a suitable plasma model. Here a magnetohydrodynamic description will be used. The corresponding eigenmodes are the Alfvén mode, the fast mode, and the slow mode, all travelling forward and backward, giving one six independent eigenmodes. The coefficients of the expansion represent the strength of the corresponding eigenmode with respect to the full wave field. In other words, the coefficients represent the degree of polarisation of the wave field with respect to a certain eigenmode.

The determination of the eigenvectors is analogous to the derivation of e.g. *Akhiezer and Akhiezer* [1975]. The basic equations are

$$\nabla \cdot \mathbf{b} = 0 \quad (4.1)$$

$$\frac{\partial \mathbf{b}}{\partial t} = \nabla \times (\mathbf{v} \times \mathbf{B}_0) \quad (4.2)$$

$$\rho_0 \frac{\partial \mathbf{v}}{\partial t} = \frac{\nabla \times \mathbf{b}}{\mu_0} \times \mathbf{B}_0 - c_s^2 \nabla \rho \quad (4.3)$$

$$\frac{\partial \rho}{\partial t} = -\rho_0 \nabla \cdot \mathbf{v} \quad (4.4)$$

Equation 4.1 is used to specify the frame of reference. It is chosen in such a way that the wave propagates parallel to the x axis ($\mathbf{k} \parallel \mathbf{x}$); thus $b_x = 0$. The background magnetic field \mathbf{B}_0 is in the xz plane with $\mathbf{B}_0 = B_0(\cos \theta, 0, \sin \theta)$, where θ is the angle between the x axis and the background magnetic field. The vectors \mathbf{b} , \mathbf{v} , and ρ are the magnetic field, the velocity, and the density variation of the wave field, c_s is the sound velocity and μ_0 the vacuum permeability. The six non-trivial components b_y , b_z , v_x , v_y , v_z , and ρ may be summarised in the dimensionless state vector

$$\mathbf{v} = \begin{pmatrix} v_1 \\ v_2 \\ v_3 \\ v_4 \\ v_5 \\ v_6 \end{pmatrix} = \frac{1}{v_A} \begin{pmatrix} v_y \\ b_y/\sqrt{\mu_0 \rho_0} \\ v_x \\ v_z \\ b_z/\sqrt{\mu_0 \rho_0} \\ \rho/\rho_0 c_s \end{pmatrix} \quad (4.5)$$

with the Alfvén velocity $v_A = B_0/\sqrt{\mu_0 \rho_0}$. The basic equations 4.1–4.3 may now be written in the form

$$\frac{\partial \mathbf{v}}{\partial t} + \mathbf{Z} \frac{\partial \mathbf{v}}{\partial x} = 0 \quad (4.6)$$

where the matrix \mathbf{Z} is defined as

$$\mathbf{Z} = v_A \begin{pmatrix} 0 & -\cos \theta & 0 & 0 & 0 & 0 \\ -\cos \theta & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & \sin \theta & \sqrt{\beta} \\ 0 & 0 & 0 & 0 & -\cos \theta & 0 \\ 0 & 0 & \sin \theta & -\cos \theta & 0 & 0 \\ 0 & 0 & \sqrt{\beta} & 0 & 0 & 0 \end{pmatrix} \quad (4.7)$$

where $\beta = c_s^2/v_A^2$ is the plasma beta. By means of a Fourier transformation one gets the algebraic equation

$$\mathbf{Z} \mathbf{v} = w \mathbf{v} \quad (4.8)$$

where the eigenvalue $w = \omega/k$ is the phase velocity. By means of an appropriate choice of weights for the components in equation 4.5, symmetry of \mathbf{Z} has been reached. This is convenient as further manipulations are easier since now the eigenvalues of \mathbf{Z} are always real and the eigenvectors are orthogonal. After some algebra we obtain the following eigenvalues w and the corresponding eigenvectors \mathbf{e} :

$$w_{(2)} = \pm v_A \cos \theta ; \quad \mathbf{e}_{(2)} = \begin{pmatrix} \mp \frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} \\ 0 \\ 0 \\ 0 \\ 0 \end{pmatrix} \quad (4.9)$$

These modes are the forward propagating (index 1) and the backward propagating (index 2) Alfvén modes.

$$w_{(4)} = \pm v_f ; \quad \mathbf{e}_{(4)} = \frac{1}{h_f} \begin{pmatrix} 0 \\ 0 \\ 1 \\ -\frac{v_A^2 \sin \theta \cos \theta}{v_f^2 - v_A^2 \cos^2 \theta} \\ \pm \frac{v_A v_f \sin \theta}{v_f^2 - v_A^2 \cos^2 \theta} \\ \pm \frac{v_A \sqrt{\beta}}{v_f} \end{pmatrix} \quad (4.10)$$

These modes are the forward propagating (index 3) and backward propagating (index 4) fast modes.

$$w_{(6)} = \pm v_s ; \quad \mathbf{e}_{(6)} = \frac{1}{h_s} \begin{pmatrix} 0 \\ 0 \\ 1 \\ -\frac{v_A^2 \sin \theta \cos \theta}{v_s^2 - v_A^2 \cos^2 \theta} \\ \pm \frac{v_A v_s \sin \theta}{v_s^2 - v_A^2 \cos^2 \theta} \\ \pm \frac{v_A \sqrt{\beta}}{v_s} \end{pmatrix} \quad (4.11)$$

These modes are the forward propagating (index 5) and backward propagating (index 6) slow modes.

In the above equations

$$v_{(f)} = v_A \sqrt{\frac{1+\beta}{2} \pm \frac{1}{2} \sqrt{1+\beta^2 - 2\beta \cos 2\theta}} \quad (4.12)$$

and the eigenvectors have been normalised using the normalisations

$$h_{(f)} = \sqrt{1 + \frac{v_A^2 \beta}{v_{(f)}^2} + \sin^2 \theta \frac{v_{(f)}^2 + v_A^2 \cos^2 \theta}{v_{(f)}^2 - v_A^2 \cos^2 \theta}} \quad (4.13)$$

Hence the set of eigenvectors $\mathbf{e}_1, \dots, \mathbf{e}_6$ forms an orthonormal base for the MHD wave state space.

Following work by *Samson and Olson [1980]* and *Samson [1983]* on pure states and their more general concept of polarisation, an arbitrary MHD wave field can be expanded in the set of eigenmodes via

$$\mathbf{v} = \sum_{i=1}^6 g_i \mathbf{e}_i \quad (4.14)$$

The complex coefficients g_i represent the contributions of the modes i to the total wave field \mathbf{v} . The total energy of the wave field is given by

$$|\mathbf{v}|^2 = \sum_{i=1}^6 |g_i|^2 \quad (4.15)$$

Thus, $|g_i|^2$ represents the energy contribution of mode i . The coefficients g_i can be determined from equation 4.14 via

$$g_i = \mathbf{e}_i^T \mathbf{v} \quad (4.16)$$

where T denotes the transposed vector.

Equation 4.16 may be regarded as describing the link between the actual observations and the plasma model chosen: \mathbf{v} is the observed wave field vector and the vectors \mathbf{e}_i are the eigenvectors of the model space. In traditional spectral density analysis the complex spectral density matrix is

$$\mathbf{S} = \mathbf{v} \mathbf{v}^\dagger \quad (4.17)$$

where \dagger stands for hermitian adjoint. With this definition of the spectral density matrix one gets

$$|g_i|^2 = \mathbf{e}_i^T \mathbf{S} \mathbf{e}_i \quad (4.18)$$

Finally, we may define normalised expansion coefficients via

$$g_i^N = \frac{g_i}{\sqrt{\sum_{j=1}^6 |g_j|^2}} \quad (4.19)$$

which represent the relative share of an eigenmode to the total wave field.

4.3 Test of the Decomposition Method

To test the applicability of the mode recognition method we simulate an MHD wave field. At the left boundary of a one-dimensional simulation box a wave transmitter of fixed frequency is operating. All components of the state vector \mathbf{v} will be excited. The simulation uses a hybrid code where the ions are described as particles and the electrons as a massless fluid (Section 18.4.3 in Chapter 18). In principle this also allows excitation of ion kinetic effects. However, these are suppressed by choosing a transmitter frequency much below the ion cyclotron frequency. Only magnetohydrodynamic modes will be excited in the simulation box. We use a transmitter frequency of $\omega = 10^{-2} \Omega_i$, and a length of the box of $10^3 c/\omega_i$. Here Ω_i and ω_i are the ion gyrofrequency and the ion plasma frequency, respectively, and c/ω_i is the ion inertial length. The background magnetic field B_0 has

an angle of $\theta = 30^\circ$ with the x axis which is the simulation axis. The plasma beta is $\beta = 1$. The resulting quasi-stationary and quasi-homogeneous wave field is displayed in Figure 4.1 where the b_y and v_x components are shown as examples.

The state vector as given from these simulations is transformed from the x - t space into the ω - k space with the resulting wave number spectrum displayed in Figure 4.2. It should be remembered that we run the simulation at a fixed transmitter frequency. The wave number spectral density maximises at $k/2\pi = 0.014, 0.023$, and $0.082 \omega_p/c$. For these maxima we determine the corresponding complex spectral density matrices according to equation 4.17. Subsequently, the eigenvectors $\mathbf{e}_1, \dots, \mathbf{e}_6$ are computed using $\theta = 30^\circ$ and $\beta = 1$. This allows one to determine the coefficients $|g_i^N|^2$ from 4.18 and 4.19 as listed in Table 4.1.

Table 4.1: Results of the simulation

$k/2\pi$	$ g_1^N ^2$	$ g_2^N ^2$	$ g_3^N ^2$	$ g_4^N ^2$	$ g_5^N ^2$	$ g_6^N ^2$
$0.014 \omega_i/c$	0.02	0.001	0.93	0.01	0.007	0.04
$0.023 \omega_i/c$	0.95	0.001	0.01	0.006	0.01	0.02
$0.082 \omega_i/c$	0.05	0.002	0.02	0.01	0.90	0.03

At $k/2\pi = 0.014 \omega_i/c$, the dominance of the third coefficient allows one to identify this mode as the forward propagating fast mode. At $k/2\pi = 0.023 \omega_i/c$, the first coefficient dominates, and the mode is identified as the forward propagating Alfvén mode. At $k/2\pi = 0.082 \omega_i/c$, the fifth coefficient dominates, and the mode is identified as the forward propagating slow mode, much as expected from our simulation.

We thus conclude that application of the discussed mode recognition mechanism to a simulated wave field yields a sufficiently precise mode identification, provided the wave propagation direction is known. To study the effect of any errors in k -determination, i.e., any uncertainties in the determination of the angle θ , we used the same spectral density matrix \mathbf{P} to determine the expansion coefficients. However, the angle θ has been varied by $\pm 20^\circ$ around the correct value. We found that only minor modifications of the expansion coefficients occur, i.e. only a small fraction of the wave energy is shifted between the different modes. In a similar way the influence of errors in the determination of the plasma β have been studied. Again it can be demonstrated that the suggested mode recognition mechanism is very robust against errors in β as well as θ . Determination of the respective modes is rather stable and unique.

4.4 The k Filtering and Generalised Minimum Variance Analysis

The mode decomposition as described above requires the knowledge of the wave propagation and the wave vector \mathbf{k} . A generalised minimum variance technique has been introduced into the field of space science by *Pinçon and Lefeuvre* [1991] and *Motschmann et al.* [1996], which allows the determination of wave propagation vectors with high precision. General aspects of this technique are described in Chapter 3. Here we shall outline only

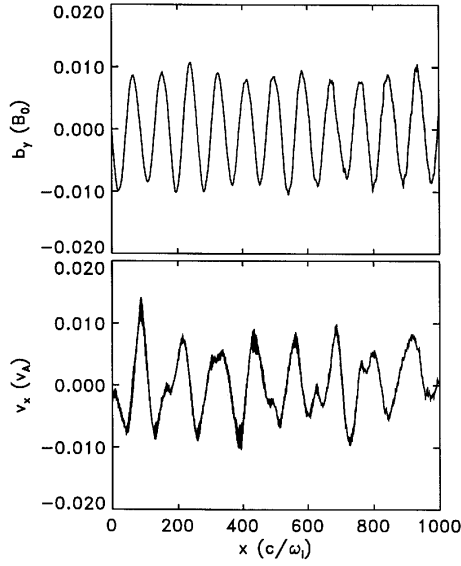


Figure 4.1: The b_y and v_x components of the simulated wave field in an one-dimensional simulation box.

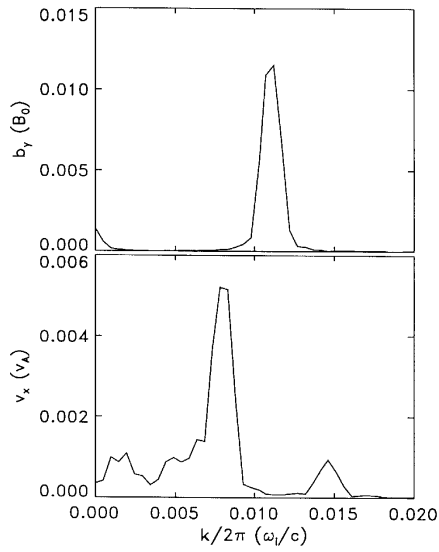


Figure 4.2: Fourier transformed b_y and v_x component of the simulated wave field.

those elements of the technique which are required to understand its potential for a wave mode analysis.

We assume a wave field consisting of an arbitrary number of partial waves. Each partial wave is assumed as a quasi-monochromatic plane wave. Thus it may be represented by its frequency ω_n , wave vector \mathbf{k}_n , and amplitude vector \mathbf{A}_n . With this one has

$$\mathbf{U}(\mathbf{r}, t) = \sum_{n=1}^N \mathbf{A}_n e^{i(\mathbf{k}_n \cdot \mathbf{r} - \omega_n t)} + \text{noise} \quad (4.20)$$

where the amplitudes \mathbf{A}_n are complex with random phase. Prior to any k -analysis, that is prior to the determination of propagation direction and wavelength, a frequency analysis must be done. This can be achieved with conventional techniques of spectral analysis. Any k -analysis will then be done for a fixed frequency ω .

To sketch the main steps of the analysis we start from the observed wave field amplitudes $\mathbf{U}(\mathbf{r}_\alpha)$, where $\alpha = 1, \dots, N$ and \mathbf{r}_α denote the position vectors of the measurement points. These are, for example, the locations of the four satellites as planned for the Cluster mission. This set of measurements can be arranged in the data matrix

$$\mathbf{M} = E \left(\begin{pmatrix} \mathbf{U}(\mathbf{r}_1) \\ \vdots \\ \mathbf{U}(\mathbf{r}_N) \end{pmatrix} \begin{pmatrix} \mathbf{U}(\mathbf{r}_1) \\ \vdots \\ \mathbf{U}(\mathbf{r}_N) \end{pmatrix}^\dagger \right) \quad (4.21)$$

where E stands for expectation value. This data matrix needs to be treated to extract the different waves, which built up the wave field vector \mathbf{U} . A detailed derivation of the required filter matrix \mathbf{F} is given by *Motschmann et al.* [1996]. The following expression is found

$$\mathbf{F}(\mathbf{k}) = \mathbf{M}^{-1} \mathbf{H}(\mathbf{k}) (\mathbf{H}^\dagger(\mathbf{k}) \mathbf{M}^{-1} \mathbf{H}(\mathbf{k}))^{-1} \quad (4.22)$$

This filter matrix has to be applied to the data matrix \mathbf{M} to receive the required spatial spectral information. The filter matrix describes an adaptive filter process with the filter depending on the data themselves. The matrix \mathbf{H} contains the sensor positions and is given by

$$\mathbf{H}(\mathbf{k}) = \begin{pmatrix} |e^{i\mathbf{k}\mathbf{r}_1} \\ \vdots \\ |e^{i\mathbf{k}\mathbf{r}_N} \end{pmatrix} \quad (4.23)$$

The spatial spectral information is contained in the spectral density matrix $\mathbf{P}(\mathbf{k})$ given by

$$\mathbf{P}(\mathbf{k}) = \mathbf{A}(\mathbf{k}) \mathbf{A}^\dagger(\mathbf{k}) \quad (4.24)$$

Now, applying \mathbf{F} to \mathbf{M} one obtains

$$\mathbf{P}(\mathbf{k}) = (\mathbf{H}^\dagger(\mathbf{k}) \mathbf{M}^{-1} \mathbf{H}(\mathbf{k}))^{-1} \quad (4.25)$$

and the spectral power density $\hat{P}(\mathbf{k})$ can be defined as the trace of this matrix:

$$\hat{P}(\mathbf{k}) = \text{Tr}[\mathbf{P}(\mathbf{k})] \quad (4.26)$$

The \mathbf{k} domain can now be scanned and those \mathbf{k} -positions be found where the spectral power density maximises. Results and further discussion of the k -filter technique are described in Chapter 3.

4.5 Model Validation

The filter process outlined above may be interpreted as a projection of the actual data onto the wave direction specified by the wave vector \mathbf{k} , i.e., the procedure outlined is a generalised projection method applied in \mathbf{k} space. In deriving equation 4.25 the only assumption used is that of a plane propagating wave. However, in general, other constraints on the wave field measured can be applied. This opens another application of the generalised minimum variance analysis, that of a more general validation of the model assumptions necessary for the mode decomposition method. We shall discuss two examples in the following, the case of selecting magnetic solenoidal wave fields and magnetohydrodynamic wave fields out of any given field.

Previously, the described filter process was applied to magnetic field data only. The amplitude A in this case is given by the magnetic field vectors, that is $A(\omega, \mathbf{k}) = \mathbf{B}(\omega, \mathbf{k})$. A natural constraint to be applied to the magnetic field is due to its solenoidal character. We thus have

$$\mathbf{k} \cdot \mathbf{B} = 0 \quad (4.27)$$

a constraint equivalent to the projection [*Motschmann et al., 1996*]

$$\mathbf{B} = \mathbf{C} \mathbf{B} \quad (4.28)$$

with

$$\mathbf{C} = \mathbf{I} + \kappa \mathbf{k}^\dagger \quad (4.29)$$

where κ is an arbitrary vector with the restrictions

$$\kappa \neq -\frac{\mathbf{k}}{|\mathbf{k}|^2} \quad \kappa \neq \mathbf{0} \quad (4.30)$$

Inserting 4.29 into equation 4.28 demonstrates that the constraints 4.28 and 4.27 are equivalent, and gives one

$$\mathbf{B} = \mathbf{B} + \kappa (\mathbf{k} \cdot \mathbf{B}) \quad (4.31)$$

from which equation 4.27 follows for $\kappa \neq \mathbf{0}$.

Application of constraint 4.28 modifies equation 4.24; the spectral density now reads

$$\mathbf{P}(\mathbf{k}) = (\mathbf{C}^\dagger(\mathbf{k})\mathbf{H}^\dagger(\mathbf{k})\mathbf{M}^{-1}\mathbf{H}(\mathbf{k})\mathbf{C}(\mathbf{k}))^{-1} \quad (4.32)$$

Note that equation 4.32 is valid for all constraints which can be written in the form 4.28. Determination of the spectral density matrix as given by 4.32 allows to separate solenoidal from non-solenoidal parts of the measured wave field. As any physically reasonable magnetic field must be solenoidal the special general mode filter outlined here is useful to suppress any non-solenoidal noise in the data.

In a similar way, any set of observations can be treated with respect to contributions which can be, for example, described as MHD and non-MHD variations. Such a model filter is useful as the mode decomposition technique requires a set of orthonormal eigenvectors to be found. These eigenvectors can be constructed from assuming e.g. an MHD model describing the observations with sufficient accuracy. However, any check of this assumption is very useful. And the constrained generalised minimum variance analysis opens such a possibility.

Here we shall discuss the case of plasma processes which can be described in the linearised MHD approximation. Observed parameters are the perturbed magnetic field vector \mathbf{b} , the perturbation of the bulk velocity \mathbf{v} , and the perturbed plasma mass density ρ . These perturbations are to be separated from the mean background parameters \mathbf{B}_0 , \mathbf{V}_0 , and ρ_0 . The wave vector amplitude is given by

$$\mathbf{A}(\omega, \mathbf{k}) = \begin{pmatrix} \mathbf{b}(\omega, \mathbf{k}) \\ \mathbf{v}(\omega, \mathbf{k}) \\ \rho(\omega, \mathbf{k}) \end{pmatrix} \quad (4.33)$$

with its components related via the following linearised MHD equations

$$\begin{aligned} -\omega \mathbf{b} + (\mathbf{V}_0 \cdot \mathbf{k}) \mathbf{b} &= (\mathbf{B}_0 \cdot \mathbf{k}) \mathbf{v} - \mathbf{B}_0 (\mathbf{k} \cdot \mathbf{v}) \\ -\omega \mathbf{v} + (\mathbf{V}_0 \cdot \mathbf{k}) \mathbf{v} &= \frac{\mathbf{B}_0 \cdot \mathbf{k}}{\mu_0 \rho_0} \mathbf{b} - \frac{\mathbf{k}}{\mu_0 \rho_0} \mathbf{B}_0 \cdot \mathbf{b} - \frac{c_s^2 \mathbf{k}}{\rho_0} \rho \\ -\omega \rho + (\mathbf{V}_0 \cdot \mathbf{k}) \rho &= -\rho_0 \mathbf{k} \rho \end{aligned} \quad (4.34)$$

Equations 4.34 can be represented in the form 4.28 after some algebra resulting into equation 4.35 (on page 88).

With this definition of \mathbf{C} , the spectral density matrix 4.32 the observations can be scanned in the \mathbf{k} domain. Due to the constraint applied only those contributions pass the filter which may be interpreted as waves in the linear MHD approximation. The mode decomposition method can then be used to determine the different MHD modes out of which the observed wave field is built up.

4.6 Conclusions

Techniques for the analysis of the dispersion and the polarisation of plane waves based on observations with small satellite arrays are presented. The major idea of the wave vector analysis is the construction and application of an adaptive filter selecting a partial wave at any \mathbf{k} and under constraints given by e.g. the MHD equations. By scanning the \mathbf{k} space the spectral density matrix is given for the entire \mathbf{k} domain. The presence of waves is determined for those positions in \mathbf{k} space where the spectral power maximises. This k -filtering method is exposed to the aliasing problem like any other spectral analysis method. To avoid ambiguities by spatial aliasing one has to demand $|\mathbf{k} \cdot \Delta \mathbf{r}| \leq 2\pi$ for any separation $\Delta \mathbf{r}$ between two sensor positions. When detecting a wave at least two sensors are needed within a wavelength.

A mode decomposition tool or generalised polarisation analysis procedure is introduced by defining a state vector in the state space spanned by the observed independent parameters, which number also defines the state space dimension. Whereas for the k -filter technique the number of observed parameters is less important a unique mode decomposition requires a full set of observable independent parameters. Thus, the state space can be spanned by the pre-selected eigenvectors, and any observed state may be expanded into a series of these eigenvectors. The coefficients of this expansion correspond to the power a specific mode contributes to the observed state.

A full mode recognition thus requires two major steps: First, the constrained minimum variance analysis is applied to check the data for the applicability of a particular plasma

$$\mathbf{C} = \begin{pmatrix} \frac{\mathbf{k} \cdot \mathbf{V}_0}{\omega} & 0 & 0 & -\frac{B_{0y}k_y + B_{0z}k_z}{\omega} & \frac{B_{0x}k_x}{\omega} & 0 \\ 0 & \frac{\mathbf{k} \cdot \mathbf{V}_0}{\omega} & 0 & -\frac{B_{0x}k_x + B_{0z}k_z}{\omega} & \frac{B_{0y}k_y}{\omega} & 0 \\ 0 & 0 & \frac{\mathbf{k} \cdot \mathbf{V}_0}{\omega} & \frac{B_{0z}k_z}{\omega} & \frac{B_{0x}k_x + B_{0y}k_y}{\omega} & 0 \\ -\frac{B_{0y}k_y + B_{0z}k_z}{\omega\mu_0\rho_0} & \frac{B_{0x}k_x}{\omega\mu_0\rho_0} & \frac{B_{0z}k_z}{\omega\mu_0\rho_0} & \frac{\mathbf{k} \cdot \mathbf{V}_0}{\omega} & 0 & \frac{c_s^2}{\omega\rho_0} \frac{k_x}{\omega} \\ \frac{B_{0x}k_y}{\omega\mu_0\rho_0} & -\frac{B_{0x}k_x + B_{0z}k_z}{\omega\mu_0\rho_0} & \frac{B_{0y}k_y}{\omega\mu_0\rho_0} & 0 & 0 & \frac{c_s^2}{\omega\rho_0} \frac{k_y}{\omega} \\ \frac{B_{0x}k_z}{\omega\mu_0\rho_0} & \frac{B_{0y}k_z}{\omega\mu_0\rho_0} & -\frac{B_{0x}k_x + B_{0y}k_y}{\omega\mu_0\rho_0} & 0 & 0 & \frac{c_s^2}{\omega\rho_0} \frac{k_z}{\omega} \\ 0 & 0 & 0 & \frac{\rho_0}{\omega} k_x & \frac{\rho_0}{\omega} k_y & \frac{\mathbf{k} \cdot \mathbf{V}_0}{\omega} \frac{k_z}{\omega} \end{pmatrix} \quad (4.35)$$

model, e.g. magnetohydrodynamics, and to determine the wave propagation vector \mathbf{k} . Second, the mode decomposition technique is applied to those wave number domains where the spectral density matrix maximises under the constrained k filter. This technique has been successfully tested with synthetic data. Anyone interested in the application of the techniques is invited to contact the authors for more information on the described tools.

Acknowledgements

The work of U. M. is supported by the Deutsche Forschungsgemeinschaft through a Werner-Heisenberg-Stipendium. Financial support by the Deutsche Agentur für Raumfahrtangelegenheiten DARA is gratefully acknowledged.

Bibliography

- Akhiezer, A. I. and Akhiezer, I., *Linear Theory*, vol. 1 of *Plasma Electrodynamics*, Pergamon Press, 1975.
- Glassmeier, K.-H., Motschmann, U., and vom Stein, R., Mode recognition of MHD wave fields at incomplete dispersion measurements, *Ann. Geophys.*, **13**, 76–83, 1995.
- Motschmann, U., Woodward, T. I., Glassmeier, K.-H., Southwood, D. J., and Pinçon, J.-L., Wavelength and direction filtering by magnetic measurements at satellite arrays: Generalized minimum variance analysis, *J. Geophys. Res.*, **101**, 4961–4965, 1996.
- Pinçon, J.-L. and Lefeuvre, F., Local characterization of homogeneous turbulence in a space plasma from simultaneous measurement of field components at several points in space, *J. Geophys. Res.*, **96**, 1789–1802, 1991.
- Samson, J. C., Pure states, polarized waves, and principal components in the spectra of multiple, geophysical time-series, *Geophys. J. R. Astron. Soc.*, **72**, 647–664, 1983.
- Samson, J. C. and Olson, J. V., Some comments on the descriptions of polarization states of waves, *Geophys. J. R. Astron. Soc.*, **61**, 115–129, 1980.

— 5 —

Measurement of Plasma Velocity Distributions

ANDREW N. FAZAKERLEY

*Mullard Space Science Laboratory
Holmbury St. Mary, United Kingdom*

STEVEN J. SCHWARTZ

*Queen Mary and Westfield College
London, United Kingdom*

GÖTZ PASCHMANN

*Max-Planck-Institut für extraterrestrische Physik
Garching, Germany*

5.1 Introduction

In this tutorial chapter, we discuss the measurement of plasma velocity distributions for particles belonging to the energy range from a few eV to a few tens of keV, sometimes referred to as the “hot plasma” regime. Instruments for the measurement of more energetic particles are not explicitly dealt with here, although some of the concepts reviewed here for hot plasma instruments also apply to energetic particle instruments. This chapter is a companion for Chapter 6. Plasma velocity moment determination is a procedure which manipulates the velocity distribution measurements; thus the quality of the plasma moments is affected by the quality of the distribution measurements. We will examine the general principles of typical modern instruments, showing the way the data is acquired, with emphasis on the capabilities and limitations of the instruments used. The chapter will illustrate the nature of the measured distribution data products and discuss their correct application and possible pitfalls in their interpretation.

A generic instrument for measuring plasma velocity distributions consists of a velocity space filter system, a detector and a counter. The velocity space filter controls access to the detector, so that it is only those charged particles from a pre-selected region (volume element) in velocity space which may reach the detector at a given instant in time. The detector is able to respond to an arriving charged particle by generating an electrical signal. The counter records the electrical signals and their time of arrival. The distribution is characterised by making measurements for a set of different regions in velocity space. Measurements of a multi-species plasma require the additional capability of distinguishing between ions of different masses and charge states. It is often important to be able to distinguish between particles of different masses, for example for studies seeking to identify the sources of particles, or studies where the mass density or higher moments of the distribution function are required (see Chapter 6).

5.2 General Principles of 3-D Velocity Distribution Measurements

There are a variety of instrument designs based on the “curved plate electrostatic analyser” which are able to acquire three-dimensional velocity distributions relatively rapidly. These can separate particles according to their direction of arrival and according to different values of energy per unit charge, E/q (and according to positive or negative charge), but do not distinguish between particles of different masses. Mass resolving instruments are built either by adding a “time of flight” (TOF) analyser section at the exit of an E/q analyser or by using a system which incorporates a deflection magnetic field.

In this section we will describe in general terms the process of measuring non-mass resolved velocity distributions with an E/q analyser, specifically a “Top Hat” spherical section analyser (an adaptation to provide mass resolving capability will also be mentioned). It will become clear that the capabilities of an instrument in terms of resolution and coverage of solid angle and E/q , and of time resolution, are not independent variables, and that an instrument is therefore usually tailored to its mission. Similarly, there are “dynamic range” constraints; limits to the count rates an instrument can deal with (also discussed below) which require that the analyser is designed to control the flux of particles reaching the detector to lie within acceptable limits in the plasmas of interest. The instrument may encounter plasmas other than those for which it has been optimised, in which case its performance may be poor. It is important that users of data acquired with such instruments have an awareness of the instrument capabilities and limitations, and in particular of when the data may not be providing a good description of the plasma the instrument encountered.

5.2.1 Velocity Space Filter

The velocity space filter must be able to select particles arriving from a restricted velocity space volume. It should be able to select a solid angle element from the unit sphere of look directions, and should be able to independently select particles of a chosen energy. A three-dimensional velocity space distribution is constructed by making measurements which sample the full 4π solid angle at a desired resolution, for each of a set of energy values which cover a required energy range, again at a desired resolution. Typical modern instruments are not able to simultaneously observe the full unit sphere, but instead cover a quasi-two-dimensional swathe following a great circle on the unit sphere, and rely on spacecraft spin or an appropriate device to scan the field of view around the full unit sphere. Simultaneous measurement of several distinct bands of E/q across a broad range of E/q is also not usually possible.

Instantaneous Solid Angle Coverage (Fixed E/q)

The “Top Hat” is a widely used instrument design which uses a curved plate analyser (CPA) as its velocity space filter. We discuss a spherical analyser here, but other analyser geometries which have different optical properties, such as the toroidal analyser, are also often used in Top Hat instruments. This is an example of the diversity in modern plasma instrument design which we discuss further below. We will refer to the curved plates in our example as “hemispheres” (although the turn angle is often less than 90° and the outer

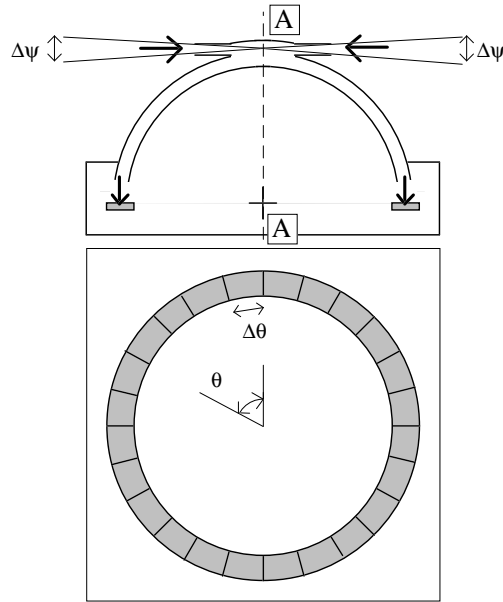


Figure 5.1: A schematic illustration of a Top Hat electrostatic analyser. The upper figure shows a cross-section through the analyser in a plane that contains the axis of rotational symmetry, which is marked by the dashed line labelled “AA”. Fine lines indicate the azimuthal angle acceptance range, $\Delta\psi$. Particles which do not have arrival paths lying within the fan bounded by these fine lines cannot reach the detector plane. The fan is defined by the parallel planar plates of the collimator section at the Top Hat entry aperture. The full range of directions from which particles may enter the aperture and go on to reach the detector plane is defined by rotating the acceptance fan about AA (and so includes directions into and out of the paper). As indicated by the bold arrows, the trajectories of particles which successfully pass through the analyser are rotated through about 90° as they pass between the “hemispheres” (see also Figure 5.3). Shortly after leaving the exit of the nested hemispheres, the particles strike the detector (indicated by the grey rectangles). The lower figure shows the detector in plan view (i.e. looking down the axis AA). We show a detector as an annular ring divided into several (the figure uses 24) equal sectors each associated with a different polar angle, θ . Each corresponds to a “polar zone” of width $\Delta\theta$ as marked. Other detector arrangements are also possible.

hemisphere has a hole in it!). Figure 5.1 provides a schematic illustration, showing that the analyser has a rotational symmetry about the axis AA. The analyser has a field of view of $[360^\circ \text{ by } \Delta\psi]$, where $\Delta\psi$ describes the aperture acceptance angle range and is usually less than 5° . A parallel plate collimator mounted at the entrance to the aperture (see Figure 5.1) is used to define $\Delta\psi$ for all particles whatever their energy. Particles approaching the analyser aperture from directions outside the aperture acceptance angle range will strike one of the collimator plates and so be brought to a halt (although they may generate secondary electrons, which must be controlled, but we will not discuss that problem here). Thus the collimator prevents any particles with arrival directions lying

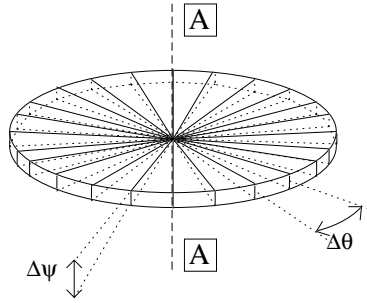


Figure 5.2: The instantaneous field of view of a Top Hat analyser consists of a rotation through 360° of the azimuthal angle acceptance fan, angular width $\Delta\psi$, about the symmetry axis AA. The field of view is shown divided up into separate sectors, of angular width $\Delta\theta$, each corresponding to a detector zone at a different polar angle.

outside the field of view from reaching the detector. The other aspect of velocity space filtering, E/q selection, is discussed later. The instantaneous field of view is illustrated in Figure 5.2.

We will work with a spherical polar coordinate system (θ, ϕ) where the polar angle, θ , is measured in the plane whose normal is parallel to the analyser symmetry axis, as illustrated in Figures 5.1 and 5.2. The azimuth angle is used later when we discuss scanning the field of view to cover the full unit sphere. The polar angle, θ , has a range of 360° , and is defined relative to an arbitrary reference direction. For an analyser on a rotating spacecraft, the reference direction is chosen to lie parallel to the spacecraft spin axis.

Particles arriving from anywhere in the field of view enter the analyser and are deflected away from the plane of the field of view, to an annular region at the analyser exit, where the detector is placed. A set of parallel particle trajectories from a given θ direction will be focused (i.e. the particle trajectories will be brought together in a small dot shaped region in the “focal plane”) irrespective of where they enter the aperture, as shown in Figure 5.3. If some of these trajectories are displaced from one another along the symmetry axis AA, the particles will arrive in the focal plane separated in the radial direction, so the focused trajectories form a narrow strip aligned radially. The detector dimensions are matched to the maximum possible radial extent of the focused beam, which is well defined by the analyser. The Top Hat’s desirable property of good focusing at all polar angles is not shared by all analyser designs.

Particles that enter the analyser from directions with different polar angles are transmitted to different parts of the detector and may be distinguished from one another according to where they strike the detector. The detector is usually subdivided into a number of zones (also often called sectors), and particles arriving in a given zone are grouped together by the counter. The analyser focusing will typically be quite precise (of order 1° in a well designed system) so the polar angle resolution $\Delta\theta$ is not usually constrained by the quality of the focus. Instead the chosen polar angle resolution is usually the coarsest resolution consistent with making an effective measurement of the plasmas the instrument is intended to study (e.g. see discussion of velocity space resolution requirements in Chapter 6, Section 6.4.1, page 132). Finer resolution brings with it the problems of more a

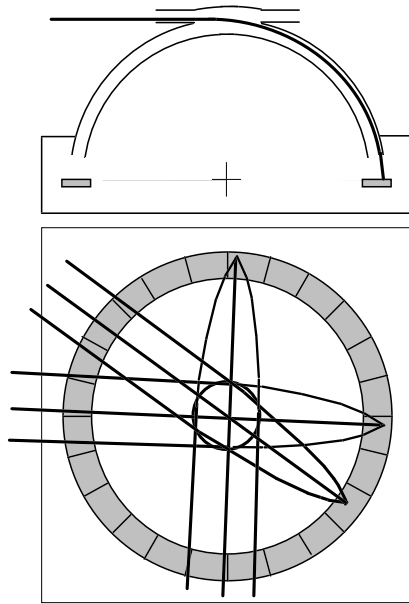


Figure 5.3: An illustration of particle trajectories through a Top Hat electrostatic analyser. Thick lines shows the trajectory of particles which have been accepted by the analyser. The lower figure shows a projection of particle trajectories onto the plane containing the detector. The small circle at the centre represents the opening in the outer hemisphere through which all detected particles must pass. The focusing of parallel beams arriving from (three) different polar angles is shown. The focusing properties of a Top Hat analyser are equally good whatever the polar angle of arrival.

complex detector, more counters, and how to reconcile a limited ability to transmit data with an increase in measured data volumes. Thus, finer resolution is only provided when it is a solid scientific requirement.

There is no discrimination between detected particles that arrive from directions sharing a given polar angle, but having different aperture entry angles in the range $\Delta\psi$. Thus the instrument samples a set of solid angles each of dimension $[\Delta\theta \text{ by } \Delta\psi]$.

Overlap of Solid Angle Coverage of Rotating Analyser (Fixed E/q)

In order to detect particles from the full unit sphere, the Top Hat field of view must be rotated through the full range of azimuth angle. A simple and commonly used approach is to mount the Top Hat on a spinning spacecraft with the analyser symmetry axis perpendicular to the spacecraft spin axis. The plane of the analyser field of view is then parallel to the plane containing the spin axis. (In the following discussion we make the reasonable assumption that we can neglect the offset of the spacecraft spin axis and the analyser field of view plane, and assume that the spin axis lies in the field of view plane.) Figure 5.4 illustrates the orientation of the analyser field of view in the (θ, ϕ) coordinate system. By convention, θ is defined so that the spin axis passes through $\theta = 0^\circ$ and $\theta = 180^\circ$.

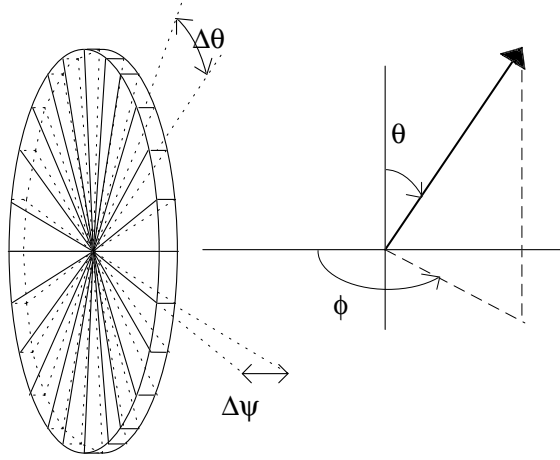


Figure 5.4: The instantaneous field of view of a Top Hat analyser mounted on a spinning spacecraft, shown in the (θ, ϕ) coordinate system where the spin axis is aligned with the $\theta = 0^\circ$ and 180° axis. The analyser symmetry axis AA lies in the spin plane, and rotates about the spin axis.

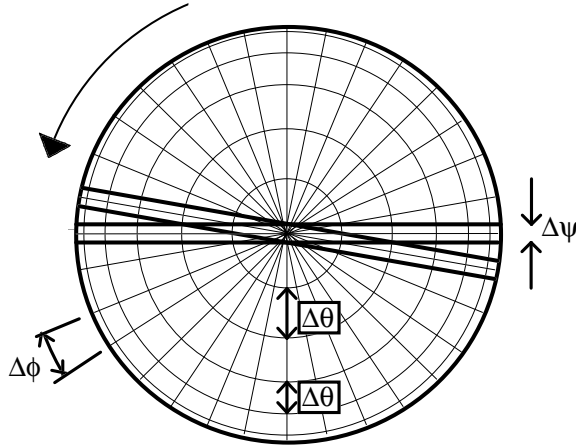


Figure 5.5: The instantaneous field of view shown at two separate instants, separated in azimuth angle by $\Delta\phi$, viewed looking along the spin axis direction. In order to illustrate the solid angle coverage of the field of view at these two instants, we have superimposed a unit sphere representing the full 4π solid angle range. Nested circles show boundaries between polar zones (shown $\Delta\theta = 15^\circ$ apart). Radial lines are centre lines of 32 azimuthal look directions ($\Delta\phi = 11.25^\circ$ apart). The illustrated azimuthal angle acceptance fan angular width $\Delta\psi = 5^\circ$. The solid angle associated with regions of a fixed angular width in polar and azimuth angle is $[\Delta\theta \text{ by } \Delta\psi]$. The overlap of solid angle coverage is obvious for polar zones near the spin axis (centre of the image).

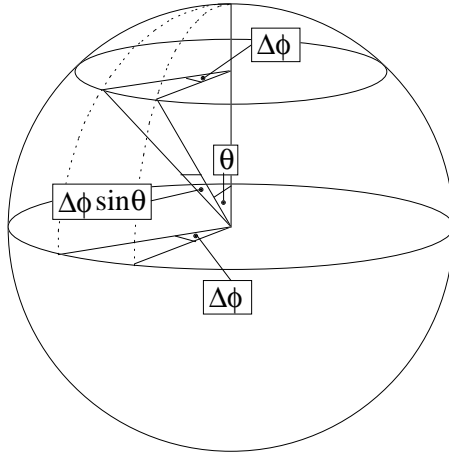


Figure 5.6: Solid angles are subtended from the centre of the unit sphere. Thus the solid angle subtended by an arc which subtends $\Delta\phi$ at the spin axis (in the plane parallel to the spin equatorial plane) and which virtually lies on a circle at constant θ , is given by $\Delta\phi \sin\theta$. For a Top Hat analyser, the solid angle associated with a fixed polar angular width is controlled by the azimuthal angle acceptance fan angular width $\Delta\psi$, rather than $\Delta\phi \sin\theta$, so we have an overlap between two measurements separated by $\Delta\phi$, of solid angle coverage near the spin axis, as illustrated in Figure 5.5.

Figure 5.5 shows a view looking down the axis of rotation of a steadily spinning analyser. Thin radial lines represent the plane (seen “edge on”) instantaneously occupied by the field of view at successive regularly spaced time intervals, Δt , which correspond to regularly spaced azimuthal look directions separated by angles, $\Delta\phi$. Nested circles represent the boundaries between polar zones, spaced out at intervals of $\Delta\theta$. The two bold rectangles represent the projection of the field of view of the analyser (imagine looking at Figure 5.2 from within the plane of the field of view) at two instants in time, separated by Δt . The width of each bold rectangle corresponds to the aperture acceptance angle $\Delta\psi$. In our example, $\Delta\phi > \Delta\psi$. It is clear from the figure that there is an overlap of the solid angle coverage of these two fields of view of the instrument (i.e. of the rectangles) near the spin axis, but not near the spin equator. The degree of overlap can be reduced by increasing Δt , but it can never be reduced to zero.

The reason for the overlap is that the solid angle associated with the field of view of a given polar zone is independent of the polar angle, in practice, whereas the solid angle would ideally become smaller as the look direction approaches the spin axis. Consider Figure 5.6, which shows a line of constant θ (i.e. a circle) marked on the unit sphere. Two meridian planes separated by an angle $\Delta\phi$ are indicated by dotted lines which show where they intersect the upper part of the sphere. These planes cut off an arc on the circle, which subtends an angle $\Delta\phi$ in the plane whose normal is along the vertical axis. The points marking the ends of this arc also identify a second arc (not explicitly shown) on the great circle which passes through both points, which subtends an angle $\Delta\phi \sin\theta$ at the centre of the sphere, as shown. Since solid angles are referred to the centre of the sphere, it is the angle subtended by the great circle arc rather than the angle between the meridian planes

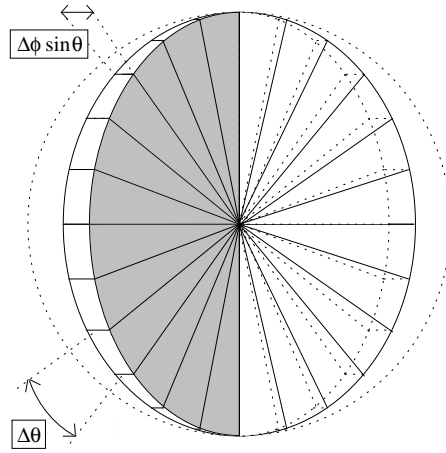


Figure 5.7: An illustration of the instantaneous field of view needed for an “ideal analyser” in which the solid angle coverage of observations spaced $\Delta\phi$ apart in azimuth could be seamlessly combined without overlap or gaps to provide complete 4π coverage of solid angle. Here the solid angle associated with regions of a fixed angular width in polar and azimuth angle would be $[\Delta\theta \text{ by } \Delta\phi \sin\theta]$.

which we use when defining solid angles. For complete and non-overlapping solid angle coverage, we would need a variable $\Delta\psi$ so that the field of view could conform to the pattern described by lines of constant θ and ϕ on the unit sphere, as indicated by thin lines in Figure 5.5 (see also Figure 5.11). Then the solid angle associated with a direction θ would be $[\Delta\theta \text{ by } \Delta\phi \sin\theta]$ as shown in Figure 5.7. In practice, the aperture acceptance angle $\Delta\psi$ is fundamentally independent of θ and does not decline as $\sin\theta$ as shown in Figure 5.4, so we have to accept that the solid angle coverage, given by $[\Delta\theta \text{ by } \Delta\psi]$, will overlap at the poles.

As shown in Figures 5.4 and 5.5, a “Top Hat” is able to observe simultaneously two polar angle fans, each of polar angle range 180° , of aperture angle width $\Delta\psi$, and these fans stare in directions separated by 180° in azimuth angle. Solid angle varies with polar zone, θ , and there is an inevitable overlap of solid angle coverage between measurements made near the spin axis at different times. The full range of azimuth angle is not simultaneously accessible, but the range is scanned as the satellite rotates so that particular azimuth directions may only be observed at particular times. A 360° field-of-view analyser can scan the full unit sphere in half a spin.

On some instruments the 360° field of view is sacrificed in favour of some other capability. The field of view may be divided into two parts, each playing a different role. For example, one part may retain a 180° field of view (from $\theta = 0^\circ$ and $\theta = 180^\circ$ to give 4π solid angle coverage in a spin) for studying magnetospheric and magnetosheath plasmas. For an ion instrument, the other part might be designed for solar wind measurements, with a reduced field of view of perhaps 135° , centred on $\theta = 90^\circ$, and a detector with fine polar resolution (i.e. small $\Delta\theta$). Alternatively, the two parts of the field of view could have different geometry factors (see Section 5.2.3) to give different degrees of sensitivity. This extends the instrument dynamic range, enabling it to make good measurements in a

wider variety of plasma conditions. A 360° field-of-view analyser must be mounted with the field of view plane tangential to the spacecraft surface in order that the spacecraft does not get in the way of arriving particles. However, stray electric fields at the spacecraft surface can affect the trajectories of the measured particles, to an undesirable degree in the case of low energy particles. Thus analysers intended to study low energy particles are sometimes mounted such that they look radially out from the spacecraft, sacrificing field of view in order that the trajectories of arriving particles are less affected by the spacecraft stray electric fields. In this case, a scan of the full unit sphere takes a full spin as the field of view is only 180°.

Energy Per Charge Sampling

Before discussing E/q sampling with a Top Hat analyser, we briefly mention another electrostatic analyser concept. The principle of the family of Retarding Potential Analysers (RPAs) is to direct particles through an electric potential that acts to oppose their motion. Particles with E/q values lower than the applied potential will be brought to a standstill and then sent back out through the entrance aperture without being detected. All particles with E/q in excess of the applied potential reach the detector and are counted. In order to assess the number of particles within a given band of E/q values, measurements are made for two values of the potential (corresponding to the upper and lower bounds of the E/q band). The difference in the number of counts then corresponds to the number of particles within the band of E/q values (provided that the plasma has not altered during the time between measurements). Note that the RPA must generate potentials equal to the E/q values of interest.

A curved plate analyser (CPA), directly measures the number of particles within a narrow band of E/q . The energy per unit charge of particles admitted by the analyser is controlled by virtue of an electric potential applied between the inner and outer hemispheres, which deflects the particle trajectories so that they closely follow the plate curvature, allowing particles to pass between the plates without striking them (and so being lost). The potential required to deflect a particle within a CPA is a fraction of the potential which would be required by an RPA to bring the same particle to rest (a technical advantage for the CPA). The CPA determines the number of particles within the selected band of E/q values in a single measurement (again, unlike the RPA). For a spherical geometry analyser with hemispheres of radii R_1 and R_2 (where $R_1 < R_2$), we can define the gap $\Delta R = R_2 - R_1$, the mean radius $R_M = (R_1 + R_2)/2$ and the radius at which the potential is midway between the values on the plates $R_V = \frac{R_2 R_1}{R_M}$. Particles travelling between the plates are subject to the radial electric field ($q dV/dR$) which provides a centripetal force. The field strength is controlled by the (variable) potential difference ΔV between the hemispheres and the dimensions of the hemispheres.

For simplicity we will consider particles which follow circular trajectories through the analyser (i.e. tangent to the hemispheres). For these particles, the required centripetal force for circular motion is provided by the electric field at all points along their path within the analyser. As the centripetal force is given by ($mv^2/R = 2E/R$) for a particle of kinetic energy E travelling at radius R , the balance of forces requires that

$$E/q = \frac{R}{2} \frac{dV}{dR} = \frac{1}{2} \frac{\Delta V}{\Delta R} \frac{R_2 R_1}{R} \quad (5.1)$$

The parameter $k = R_M/\Delta R$ is called the “analyser constant”. It is a useful figure of merit which is related to several properties of the analyser. For example, it can be used to characterise the ratio of the potential difference needed to bring the particle to rest, as in an RPA, and the CPA inter-hemisphere potential difference; $(E/q)/\Delta V = k/2$. This relationship is the result of applying equation 5.1 in the case of a circular path of radius R_V . Note that the analyser constant is purely controlled by the radii of the hemispheres. Typical values of k are in the range 5 to 10.

For a fixed potential, the analyser admits particles with energy per charge values which fall within the narrow range, or “intrinsic passband”, $\Delta E/q$, centred on a particular value of E/q , often termed the “passband centre energy”. The passband width is related to the range of E/q values accepted for a fixed ΔV which, according to equation 5.1, corresponds for circular trajectories to the range of values of R between R_1 and R_2 , i.e. ΔR . If we assume that the passband centre energy trajectory has radius R_V it is easy to use equation 5.1 to show for this special case of circular particle paths another important result involving k , which is that $\Delta E/E = 1/k$, i.e. $\Delta E/E$ is a constant.

In general, particle trajectories in a spherical section analyser can be eccentric ellipses as well as circles. The force balance expression used above does not hold along elliptical trajectories. The non-circular trajectories result in a passband which is not equally weighted at all energies within the acceptance range (as would be expected if circular trajectories alone are considered). For example, a particle with energy E which enters along a tangent at radial distance R will follow a circular path of radius R if E and R satisfy equation 5.1, but other particles with larger or smaller energies E' which enter along the same path will follow elliptical paths in the analyser, and will only reach the exit if their paths do not intersect a hemisphere. A particle will strike a hemisphere if the difference $E' - E$ exceeds the difference in the potentials of the equipotential surface of radius R and of the hemisphere towards which the particle moves. Thus, if R is near R_2 , then the range of transmitted energies which are smaller than E is greater than the range of transmitted energies larger than E . Similarly, a wider range of energies greater than E , as opposed to energies below E , are transmitted if R is near R_1 . The widest spread of energies to either side of E occurs when the path has radius R_V , at which distance the potential is midway between the potentials of the two hemispheres. Thus there is a peak in the energy response of the aperture as a whole, usually associated with particles entering on the path of radius R_V . This peak typically falls at the centre of the passband in a well designed analyser.

The particular expression given above for $\Delta E/E$ in terms of k applies only in the special case outlined there. For example, a CPA where the aperture dimensions are R_1 and R_2 , but where the gap between the plates is wider than ΔR , can accommodate additional particle trajectories which the analyser described above would not accept; these being paths which describe more eccentric ellipses than would be possible if the plate separation matched the aperture. Thus this second CPA will have a broader spread of accepted energies about the centre energy, and a larger $\Delta E/E$.

In real analysers, whether they are simple CPAs or use a Top Hat aperture, there are non-radial electric field components near the apertures. These “fringing fields” also lead to more complicated behaviour than we have outlined here, although instruments are often designed to minimise the effects of the fringing fields so that performance is close to the ideal discussed above. Experimental determination of the actual analyser transmission characteristics are always an important part of an instrument calibration, in order to validate both that the design has correctly addressed all the complexities hinted at here and

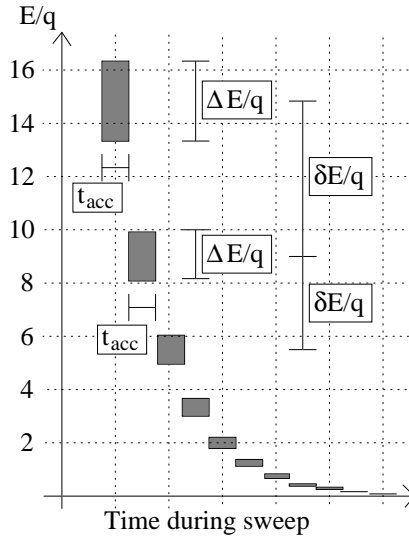


Figure 5.8: The figure illustrates the E/q step spacing (with centre values E/q separated by steps $\delta E/q$) and the change in E/q acceptance range $\Delta E/q$ with centre value under the condition that $\Delta E/E$ is constant. The axes are a linear energy per charge scale and a linear time scale (both in arbitrary units). In practice, energy steps are often spaced so that the steps are separated by $\delta E/q = \Delta E/q$ rather than the wider spacing shown here.

that the manufacture has correctly followed the design.

In fact, the property that $\Delta E/E$ is a constant applies generally for any given CPA design (i.e. even when elliptical paths are considered) so that if E/q is altered, $\Delta E/q$ also changes. Thus the intrinsic passband is narrower when the admittance energy is smaller, as shown in Figure 5.8. The intrinsic passband width is typically chosen so that $\Delta E/E$ is of order 10 to 20% for measurements of magnetospheric plasmas, but may be smaller where better E/q resolution (usually across a reduced E/q range) is required; the classic example being that of solar wind measurements.

An instrument is usually intended to study a broad range of E/q , but can only measure in a narrow range $\Delta E/q$ at any given moment. The desired coverage of E/q is achieved by making consecutive measurements for values of E/q spread across the required range. Typically these are spaced logarithmically, by $\delta E/q$, (chosen so that $\delta E/E$ is constant) and ideally $\delta E/q = \Delta E/q$ so that there are no gaps in coverage between the passbands of consecutive measurements, as illustrated Figure 5.9.

Data is usually gathered for a fixed time interval, the accumulation time t_{acc} (the duration of which is influenced by other instrument design parameters). In some instruments the inter-hemisphere potential is held constant for $> 90\%$ of t_{acc} and rapidly lowered to the next required value during the remainder of the interval. In this case, where we “step” through a range of measured E/q values, each measurement is essentially taken at a well-defined, fixed value of E/q and for a well defined passband $\Delta E/q$.

Another common practice is to “sweep” through a range of measured E/q values, which involves allowing the potential to continuously vary (usually declining logarithmi-

that a stepping analyser must use steps separated by gaps wider than the intrinsic bandpass ($\delta E/q > \Delta E/q$) in order to sample points throughout the full E/q range. Consequently, coverage of the full E/q range is not complete in this case. A sweeping analyser can cover the same E/q range in the same shorter time, taking data on a coarser E/q resolution, $(\Delta E/q)_{\text{eff}}$, but without gaps in E/q coverage. For an instrument on a spinning satellite, the azimuthal resolution is controlled by the ratio of T_{range} to the spin period. As a result, improved energy resolution and coverage are only achieved at the cost of worsened azimuthal resolution.

There is a short interval between sweeps (or a sequence of steps) during which the inter-hemisphere potential is raised from its minimum to its maximum values. Information acquired from the counters during these brief “flyback” intervals is usually discarded as it is difficult to interpret.

A typical instrument might admit particles from an E/q range with a maximum as high as several tens of keV/ q and a minimum of only a few eV/ q .

Solid Angle Coverage of Rotating Analyser (Varying E/Q)

We will consider a sweeping instrument in the following. A sweep through the full E/q range takes a finite time ($T_{\text{range}} = T_{\text{sweep}}$), during which the satellite rotates and the look direction azimuth angle changes by $\Delta\phi_{\text{sweep}}$. The duration of a measurement at a particular E/q value, t_{acc} , corresponds to a small change in azimuth $\Delta\phi_{\text{acc}}$. Instruments are usually designed with a sweep period which is an integer fraction of the spin period, so that there are an integer number of sweeps in a spin, and so that $\Delta\phi_{\text{sweep}}$ is consistent with the required azimuthal resolution. The consequence is that measurements taken at a particular E/q (during the short time t_{acc}) on consecutive sweeps are separated in time by T_{sweep} and are correspondingly separated in azimuth by $\Delta\phi_{\text{sweep}}$. Measurements taken at some other E/q on consecutive sweeps are likewise separated in time by T_{sweep} and thus in azimuth by $\Delta\phi_{\text{sweep}}$, but are displaced from the first set of measurements by the time taken for the sweep to move from one E/q to the other (some integer multiple of t_{acc}) and the look directions are displaced by a corresponding azimuth angle (an integer multiple of $\Delta\phi_{\text{acc}}$). These concepts are illustrated in Figure 5.10. There is a gap in velocity space coverage due to flyback intervals during which no useful E/q data is taken from the small azimuth angle range corresponding to the flyback time. This effect is illustrated in Figures 5.9 and 5.10.

The angle $\Delta\phi_{\text{acc}}$ is typically small compared to the aperture acceptance angle, $\Delta\psi$. Figure 5.9 uses representative values; the azimuth angle resolution $\Delta\phi (= \Delta\phi_{\text{sweep}})$ corresponds to 11.25° , or 32 sweeps per spin, and the value of $\Delta\psi$ is shown as 5° . For an energy sweep of duration $16 t_{\text{acc}}$ (15 accumulation intervals and one flyback) $\Delta\phi_{\text{acc}} = \Delta\phi/16$, corresponding to 0.7° , much smaller than $\Delta\psi$. For strict accuracy, we note that the solid angle coverage taking into account spin is $[\Delta\psi' \text{ by } \Delta\theta]$, where $\Delta\psi' = \Delta\psi + \Delta\phi_{\text{acc}} \sin \theta$.

As $\Delta\psi' \approx \Delta\psi$ is clearly a good approximation (usually), Figures 5.4 and 5.5 are a good representation of the solid angle coverage during an accumulation time, although we previously used them to discuss instantaneous solid angle coverage. Thus, since t_{acc} is usually a small fraction of T_{sweep} , our earlier discussion of how solid angle coverage overlaps near the poles applies here as well. The gaps in solid angle coverage for $\theta = 90^\circ$ occur for each E/q measurement and are illustrated in Figure 5.10 as black bands. These coverage gaps close up at a value of θ intermediate between 90° and 0° (180°) and in

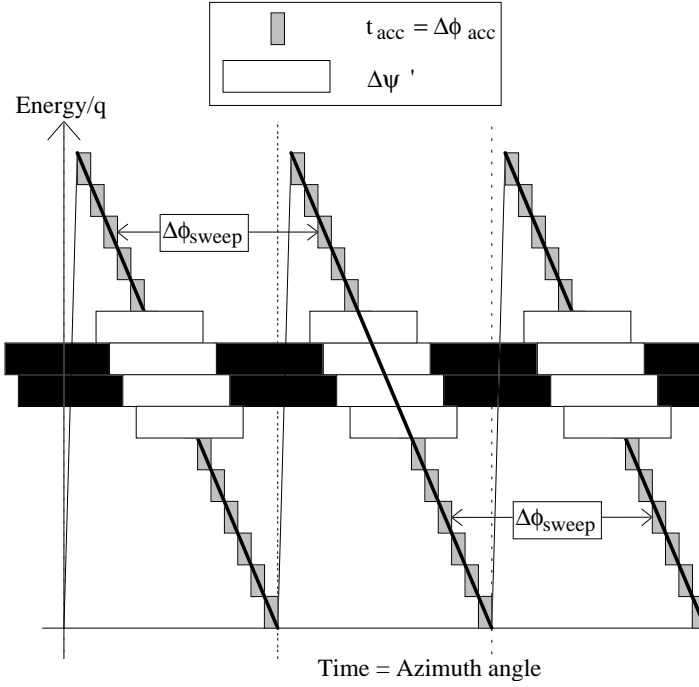


Figure 5.10: Demonstration of three consecutive measurement cycles across the E/q range, each identical to the cycle shown in Figure 5.9. For four values of E/q we show the azimuthal coverage $\Delta\psi'$. Note that $\Delta\psi' < \Delta\phi_{\text{sweep}}$ for the situation outlined in the caption to Figure 5.9 so that the azimuth angular acceptance is not wide enough to overlap with the coverage at the same value of E/q on the next sweep. Black areas represents regions of E/q -azimuth angle space which are not sampled. We would see an overlap if the figure represented the case of $\theta = 15^\circ$.

the polar regions we usually have overlapping coverage rather than gaps. If we had an analyser with a field of view which tapered towards the poles (Figure 5.7), and the value of $\Delta\psi$ were chosen to be equal to $\Delta\phi$, the sweeping (or stepping) of the analyser potential would ensure that samples of a given E/q were spaced $\Delta\phi$ apart and the instrument would naturally have perfect solid angle coverage at all values of E/q , with no overlap, nor any gaps (this ideal is shown in Figure 5.11). Unfortunately, there are a number of technical problems with the realisation of this design approach, so practical analysers tend to have both overlap and gaps in solid angle coverage.

We will now summarise the discussion in Section 5.2.1. We have seen that the direction of origin of the stream of particles arriving at a particular detector polar zone is continuously varying if the satellite is spinning. For zone n , the incident particles arriving at any given instant come from a solid angle $[\Delta\theta$ by $\Delta\psi]$, centred on the polar angle $\theta = (n - 0.5)\Delta\theta$ and on a continuously changing azimuth angle $\phi_{n(t)}$. During a time t_{acc} the azimuth angle changes by a small amount $\Delta\phi_{\text{acc}} (= \phi_{n(t+t_{\text{acc}})} - \phi_{n(t)})$ and the full solid angle scanned is $[\Delta\theta$ by $\Delta\psi']$. Only that fraction of the incident stream which the E/q selector admits, instantaneously given by the intrinsic bandpass $\Delta E/q$, is allowed to

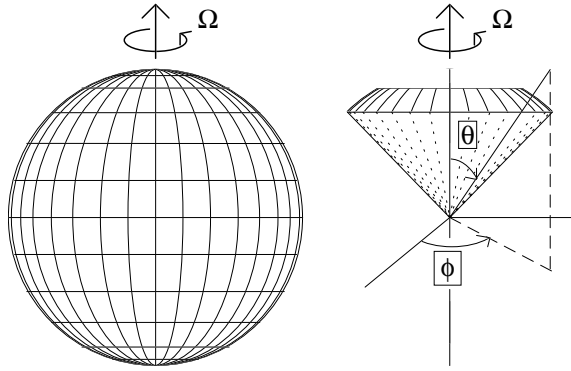


Figure 5.11: Solid angle coverage associated with the rotation of the field of view of an “ideal analyser” (see also Figure 5.7) in which the solid angle elements correspond to regions of a fixed angular width in polar and azimuth angle, $\Delta\theta$ and $\Delta\phi$. The right-hand figure shows coverage due to a single polar zone during one spin. The spin is subdivided into 32 azimuthal sectors.

reach the detector. The selected value of E/q changes continuously (for a sweeping analyser) so the range of E/q values admitted during t_{acc} is given by an effective bandpass, $(\Delta E/q)_{\text{eff}}$, larger than $\Delta E/q$. All polar angles can be continuously and simultaneously monitored, but continuous and simultaneous coverage of energy and azimuth is not possible with a Top Hat instrument. A typical instrument could have; 24 equal polar zones each of $\Delta\theta = 15^\circ$, $T_{\text{spin}} = 32T_{\text{sweep}}$ so there are 32 equal azimuth zones each of $\Delta\phi = 11.25^\circ$, and $T_{\text{sweep}} = 16t_{\text{acc}}$ (15 accumulation intervals and one flyback) so there are 15 E/q measurements per sweep. Such an instrument generates 11520 measurement values each spin. This is 2880 measurements per second if the spin period has a typical value of 4 seconds.

5.2.2 Mass per Charge Analysis

An E/q analyser is incapable of directly discriminating between particle of different masses. A data analyst may be able to identify different species by examining E/q analysed distributions, in certain special cases. An example is the case of solar wind proton and alpha particle populations. These have about the same bulk speed, but different masses. The kinetic energy of bulk motion of the two populations differs by the ratio of their masses (4:1). The spread of particle energies relative to the kinetic energy of bulk motion is characterised by the thermal speed of the population, and for the solar wind populations the thermal speeds are small compared to the difference in their kinetic energy of bulk motion. Thus the two populations occupy distinct and non-overlapping ranges of E/q . The same populations, slowed somewhat and greatly heated as they cross the bow shock into the magnetosheath, become indistinguishable using an E/q analyser. Their thermal energies now cover a much broader range, more than sufficient to bridge the energy per charge gap due to the difference in kinetic energy of bulk magnetosheath motion, and so the populations now occupy overlapping ranges of E/q . An M/q analyser will distinguish between protons (H^+ , $M/q = 1$), alpha particles (He^{2+} , $M/q = 2$) and even singly charged helium (He^+ , $M/q = 4$) enabling discrimination between these different

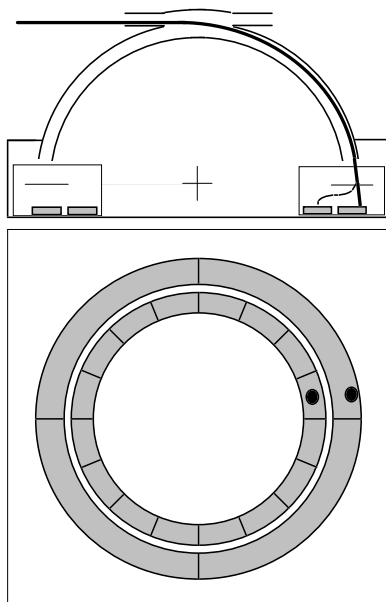


Figure 5.12: A schematic illustration of a Top Hat analyser with a Time of Flight section placed at the analyser exit. An incoming ion trajectory is shown as a thick line. The ion penetrates the foil and carries on to the outer detector (“stop detector”). Secondary electrons are generated at the foil and these travel very quickly to the inner detector (“start detector”) – their path is indicated by the thin “flattened S” shaped curve. Electrons arrive at the inner detector before the ion arrives at the outer detector. Ion and electron arrival sites are marked on the detector in the lower figure.

particle populations in both the solar wind and the magnetosheath.

There are several concepts for M/q analysis in which various combinations of electric and magnetic fields are used. Spacecraft magnetic cleanliness constraints vary from mission to mission, in some cases preventing the use of certain instrument designs. However, instruments have been built with a magnetic field geometry which minimises “leakage” of magnetic field beyond the instrument, for example by using a ring shaped magnetic field matched to the shape of the annular exit region of a velocity space analyser. Particles leaving the analyser are deflected radially to differing degrees according to their mass, while preserving their polar angle distribution. The position sensitive detector therefore needs to be able to simultaneously monitor not only polar zones but also (in essence) radial arrival position within each polar zone.

A different family of instruments uses only electrostatic fields for M/q analysis, and are based on the “time of flight” (TOF) principle. TOF designs are used in conjunction with velocity space analysers, so that particles entering the TOF system already have well-defined E/q values and have been sorted according to incident polar direction, e.g. see Figure 5.12. A common TOF technique involves a velocity measurement, which is usually achieved by determining the time taken for the particles (ions) to travel a known distance between a thin foil (through which they pass) and a detector, the “stop” detector. The foil

is chosen so as to impede the particle as little as possible, but must also generate electrons at the point through which the particle passes. These electrons are directed by a (relatively small) electric potential onto a second detector, the “start” detector. The system is arranged so that the time taken for the electrons to reach the start detector is well known and very small compared to the time for the particle to reach the stop detector. The signal from the start detector is used to infer when the particle crossed the foil, and this is compared to the time of arrival of the particle at the stop detector to determine the time of flight of the particle across the measured distance, and hence the particle speed. A linear electric field placed between the analyser exit and the foil is used to “post-accelerate” particles through a known potential. The post-acceleration provides the particle with sufficient energy to easily penetrate the foil, although by increasing the particle energy it reduces the time of flight which is thus harder to measure. The increase in kinetic energy per unit charge as the particle transits the post-acceleration region (given by the known potential U) plus the kinetic energy per unit charge of the particle as it left the electrostatic analyser (also known from the E/q analyser) gives the kinetic energy per charge of the particle as it enters the TOF section, E_{TOF}/q . The mass per charge is straightforwardly derived given E_{TOF}/q and the corresponding particle speed measured in the TOF section. Polar position information is usually measured by dividing the “start” detector into polar zones. In some instruments, the stop detector is designed to measure the energy of the arriving particle, in which case the instrument unambiguously determines E , M , and q . M/q resolution depends on timing accuracy, on the angular spread of particle trajectories passing through the TOF due to the dispersion of their exit trajectories from the E/q analyser, on the accuracy of the E/q measurement and in particular on the spread of energy per charge values $\Delta E/q$ admitted by the E/q analyser. The accuracy is also limited by the interaction of the particles with the foil, in which some will experience a degree of angular scattering and energy absorption. Pre-flight calibrations provide a proper understanding of the significance of these various effects and thus of the true instrument performance limits.

A more recent (positively charged particle specific) development in TOF design deserves a brief mention due to its superior M/q resolution. The attractive feature of this design is that the particle makes a timed journey in a parabolic trajectory under the influence of a linear electric field, for which the time of flight is not dependent on the particle E/q , but depends only on the M/q and the electric field strength. The M/q resolution is thus limited only by timing accuracy (and not for example by the properties of the E/q analyser).

The maximum count rate for which good M/q measurements can be made is limited by the ability of the instrument to associate signals correctly from the start and stop detectors without ambiguity. The larger the range of M/q to be measured, the larger the TOF could be and so the smaller the maximum count rate. Other effects which limit performance at higher count rates are discussed in the following section.

5.2.3 The Detector and Counter

The role of the detector was defined in Section 5.1 as being to respond to an arriving charged particle by generating an electrical signal. The overall detector-counter efficiency, ϵ_{DC} , may be defined as the ratio of the number of counted particles to the number of incident particles. It is often the case that the performance of detector polar zones varies from zone to zone, so it is usual to consider a separate efficiency $(\epsilon_{\text{DC}})_i$ for each polar zone

separately. A second important property is the dead time τ_i of each detector-counter zone, which is a parameter used to describe effects which cause a decline in ϵ_{DC} with increasing count rate.

Microchannel Plate Chevron Pair Detector

A commonly used detector is the microchannel plate chevron pair (MCP), with a bias electric potential difference across it, beneath which is placed a charge collecting anode, divided into several parts which define the detector zones. A microchannel plate consists of a large number of tiny hollow tubes (“pores”) packed together so that their open ends form the top and bottom surfaces of the plate. An incident particle arrives at the upper surface of the upper plate and may enter a pore in which case it will usually generate an avalanche of electrons through that plate, which in turn stimulates further avalanches in the lower plate. The result is a huge ($> 1 \times 10^6$) charge amplification that generates a sufficiently large charge to be registered by electronics attached to the anode below the lower surface of the MCP.

The Anode

As noted earlier, it is the design of the anode that controls polar resolution. The MCP is usually a continuous unbroken object and the analyser has good polar angle focusing properties. The “anode” we have been discussing actually consists of a set of equal sized Discrete Anodes, each with its own signal processing and counting electronics. The spatial resolution $\Delta\theta$ is controlled by the size of each discrete anode. Anodes do not need to be the same size as each other. A solar wind measuring instrument might use smaller anodes for that part of the polar angle range which is expected to see the solar wind. Similarly, anodes designed to have widths related to bins of constant pitch angle can be used in regions of well-defined and steady magnetic field orientation (e.g. sounding rocket applications). The discrete anode approach has the advantage that each anode/counter operates separately, so that the instrument can handle high count rates on many anodes at once, but there is also the disadvantage that greater angular resolution requires additional anodes and associated electronics. In practice, considerations related to electronic complexity limit the number of discrete anodes. Alternative anode designs allow higher angular resolution, at the expense of high count rate performance.

Another common design is the “charge dividing” Resistive Anode, which consists of a material of uniform resistance. A charge cloud deposited on the anode will be divided between charge collectors on opposite sides of the material in a ratio which indicates the location of the charge cloud on the shortest straight line between the collectors. With two sets of collectors and a quadrilateral anode, position sensing in two dimensions is possible. A more suitable approach for application to an annular anode is the “charge sharing” Wedge and Strip Anode which has three electrodes. Thinking first of a rectangular anode, two of the electrodes are reminiscent of the teeth of interlocking combs, the third fills in the gaps between the teeth. One of the two “combs” has rectangular teeth (“strips”). The strip thickness gets progressively smaller from one end of the comb to the other, and the other comb has identical teeth which narrow to a point (“wedges”). Thus the strips change thickness from one end of the comb to the other, while the wedges change thickness from base to tip, in the direction transverse to that in which strip length varies.

Consequently, the ratio of the electrode areas at any point is unique, and the ratios of the amounts of charge collected is also unique, allowing one to determine the location of the charge cloud. The design can be modified for use in an annular or semi-annular anode. Anodes with such two-dimensional position sensitivity are needed for the family of mass-resolving analysers that disperse particles of different mass in one dimension, while using an orthogonal dimension for angular position sensing.

Fraction of Particles which Generate Signals

Not all incident particles precipitate a charge avalanche. A proportion of incident particles strike the material between the pore voids, rather than entering a pore, and are thus undetected. Those particles that enter pores but do not generate charge avalanches are also undetected. The efficiency (“quantum efficiency”) of the process whereby particles generate a charge avalanche depends on the nature of the particle and on its energy (pre-acceleration techniques are often used to optimise performance). In summary, the peak MCP efficiency is typically of order 60 to 70% but the efficiency can be smaller according to the particle E/q .

The performance of an MCP may not be uniform across the area exposed to incident particles, due to manufacturing limitations (some blocked pores, variation of quantum efficiency from pore to pore, etc.). If the incident particle trajectories are not along the normal to the MCP, pore bias angle effects may also lead to variations of efficiency with position on the MCP surface. Typically the detector performance is therefore evaluated polar zone by polar zone.

All the above factors contribute to an MCP efficiency parameter, $(\epsilon_{\text{MCP}}(E))_i$, which is energy dependent and polar zone dependent. As we explain below, not all charge avalanches which are triggered are actually counted, so the detector-counter efficiency is smaller than the MCP efficiency.

Fraction of Signals that are Counted

The signal gain, i.e. the amount of charge liberated at the anode for an incident particle, exhibits a statistical variation from particle to particle, but follows a roughly Gaussian distribution (even for an ideal homogeneous MCP). For each charge avalanche, the charge deposited at the anode is electronically amplified and the corresponding signal is further processed before it is accepted as a counted particle. In a crucial processing step, the signal amplitude is compared to a predetermined threshold which it must exceed in order for the particle to be counted (so as to avoid reacting to electronic noise which can mimic an anode signal). A fraction of the real signals may generate a pulse which falls below the threshold, but this should be only a small proportion of the real signals for a correctly set threshold. If we take this cause of non-counting of an incident particle into account as well, together with $(\epsilon_{\text{MCP}}(E))_i$ we finally arrive at the detector-counter efficiency $(\epsilon_{\text{DC}}(E))_i$ applicable for low count rates.

Limits to Detector Performance at High Count Rates

A limitation on MCP performance can arise when the count rate is very high. The signals generated by the MCP in response to incident particles become weaker and the

mean signal gain falls. The effect can be thought of in terms of a pore requiring a finite time to recover after a charge avalanche, so that when a new particle arrives before pore “recharging” is complete, the resulting charge avalanche is smaller than if the pore had been fully recharged. An equivalent view is that there is a limit to the rate at which charge can be extracted from the MCP, and as it is approached, that current no longer increases with count rate, but instead the limiting current is divided between more events which, on average, become correspondingly weaker. Fewer particles generate signals exceeding the threshold, so a smaller proportion of the incident particles are counted. Detector-counter efficiency thus falls with increasing count rates, although in a well-designed system this gain depression effect should not cause serious efficiency loss within the range of count rates for which the system is specified. Note though, that the dynamic range required of instruments depends on their mission, and instruments are sometimes flown in the knowledge that their performance has been optimised for some, but not all of the environments they will encounter.

Limits to Counter Performance at High Count Rates

At high count rates we must also consider constraints due to the response of the counter electronics. We have briefly indicated that after the charge cloud caused by a particle interacting with the MCP reaches the anode, there is a certain amount of electronic amplification and other processing to produce a signal suitable for acceptance/rejection as a “count”. The processing electronics require a finite time, known as a “dead time” to deal with this activity and are unable to register subsequent counts while they are doing so, hence the term, “dead time”. It is common to divide counters into two groups, “non-paralysable” and “paralysable” counters. The non-paralysable counter is unresponsive to additional incident particles while it is occupied with processing an event, and thus has a fixed dead time. In contrast, further events may extend the dead time of the paralysable counter, with very severe consequences at high count rates, hence its name. Actual counter behaviour is often more complex than these two models suggest, but performance typically lies between the bounds set by the two models.

Poisson statistics are used to describe the frequency of arrival of particles at the detector. The probability Π that a new count will arrive during a fixed dead time τ is given by the negative exponential

$$\Pi = 1 - e^{-\lambda\tau} \quad (5.2)$$

when the incident count rate is λ . The probability Π is thus very low at low count rates, rising rapidly for count rates in the range 0.1τ to 10τ to values very close to 1.

A non-paralysable counter will respond to an incident count rate λ by registering events at the reduced rate λ' given by

$$\lambda' = \lambda / (1 + \lambda\tau) \quad (5.3)$$

while the expression for a paralysable counter is

$$\lambda' = \lambda e^{-\lambda\tau} \quad (5.4)$$

These expressions for registered count rates λ' are illustrated by curves in Figure 5.13 (for a typical electronic dead time of order $1 \mu s$) which also shows the incident count rate λ for comparison. Both types of counter work well at low count rates, i.e. $\lambda' \approx \lambda$, because at low count rates few counts arrive within a dead time of one another. The counters become

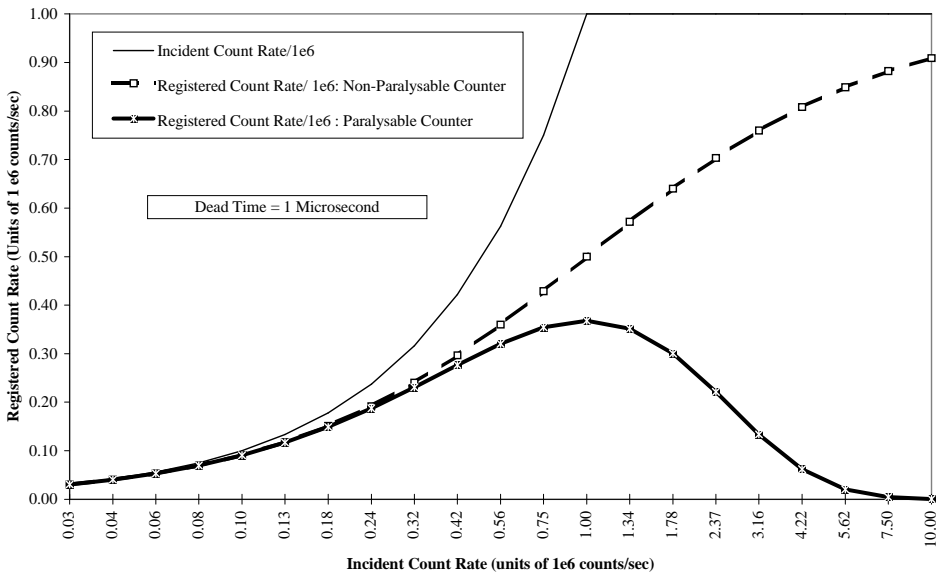


Figure 5.13: An illustration (using a log scale) of the variation of registered count rate with the rate at which particles arrive at the detector, for three cases. The thin solid line represents a perfect detecting and counting system, so all arriving particles are counted. The thicker lines represent two models of counting system performance, the paralysable and non-paralysable counters. Neither counter performs well at high incident count rates; limits to counting capability place an upper limit on the dynamic range of an instrument. The paralysable counter curve has a maximum value, which can cause problems as a registered count rate could be interpreted as one of two values, according to whether the instrument is believed to be measuring low or high particle fluxes. The non-paralysable counter curve tends asymptotically to a constant value. Once near this value, it is difficult to infer the incident count rate reliably using the registered count rate.

useless when the change in registered counts for a change in incident count rate becomes very small. The effective operating range of the paralysable counter is thus quite limited, as λ' reaches a peak value of $\lambda' \approx 0.37\lambda$ at $\lambda = 1/\tau$ after which λ' declines with increasing count rate.

Thus, for a paralysable detector, we are unable to reconstruct the incident count rate from the detected count rate once the incident count rate exceeds $\lambda = 1/\tau$. The non-paralysable counter is still useful at $\lambda = 1/\tau$, recording about 50% of incident counts, however as count rates increase, the registered count rate tends to a limiting count rate of $\lambda' = 1/\tau$. Nevertheless, the non-paralysable counter can operate over a count rate range at least a factor of 10 higher than the paralysable counter.

When characterising an instrument it is usual to classify it as one of the two types and determine a corresponding dead time τ . Dead time correction can then be performed on the registered events to reconstruct the incident count rate, using whichever of equations 5.3 or 5.4 is more appropriate. Often the system will be designed so that the count rates do not reach levels at which there is a significant difference between either of the two models,

or the real behaviour, so that a good correction is possible using whichever model is more convenient.

If the detector-counter is operated at count rates where MCP gain depression becomes significant, the registered count rates will cease to conform to the electronic dead time model so that the relationship between incident count rate and registered count rate is no longer so well described using τ together with one of equations 5.3 or 5.4. Thus the accuracy with which the incident count rate is recovered from the registered count rate will be reduced. Ideally, the system is designed to avoid MCP gain depression.

Conversion from Counts to Phase Space Density

The goal of the measurement process is to characterise the velocity space distribution of the plasma within which the instrument is operated. In the following discussion we deal with the detection by an E/q analyser of singly charged particles of a single species. Thus we refer to particle energy rather than E/q . The treatment of data from an analyser which also resolves M/q follows identical lines for each individual species in turn.

The counter registers a continuous stream of electrical signals in response to particles arriving at the detector front end. As discussed earlier, the instrument divides time up into a long series of accumulation intervals of duration t_{acc} and records the number of pulses from each anode during each such interval, together with a reference time for each interval. The reference time is used to identify both the azimuthal look direction and the selected particle energy corresponding to that interval. The polar look direction is defined by the polar zone (anode) from which the data is acquired.

Determination of Number of Counts that Reached the Detector

The number of particles, N , that actually reached the detector from the analyser, can be related to the number of particles, P' , registered by the detector-counter system during the accumulation time t_{acc} (so $\lambda' = P'/t_{\text{acc}}$) by taking into consideration the detector efficiency and the possibility of a dead time effect. For a non-paralysable counter, we adapt equation 5.3 to give P , the number of registered particles corrected for dead time effects (intended to represent the true number of particles which arrived during the accumulation time) using a polar zone dependent dead time τ_i

$$P = \frac{P'}{1 - P' \frac{\tau_i}{t_{\text{acc}}}} \quad (5.5)$$

We have neglected to account for the MCP gain depression effect. Thus this expression only applies if there is no significant MCP gain depression effect, or else in the case that the consequences of any effect are regarded as having been successfully subsumed within the electronic dead time correction. Now we correct for the energy dependent and polar zone dependent detector-counter efficiency, to give

$$N_{ijk} = P_{ijk}/(\epsilon_{\text{DC}}(E_k))_i \quad (5.6)$$

where i represents the polar zone, j the azimuthal sector, and k the centre energy (discussed further in the following paragraphs) associated with the accumulation interval.

Relating Counts Accepted by the Analyser and the Velocity Space Density in the Acceptance Volume

The role of the analyser (velocity filter) is to ensure that particles that reach the detector could only have come from a specific restricted velocity space volume, the “acceptance volume”, centred on the velocity space coordinate $\{\theta_i, \phi_j, E_k\}$. The acceptance volume has dimensions $[\Delta\theta \text{ by } \Delta\psi' \text{ by } \Delta E]$ for a stepping analyser or $[\Delta\theta \text{ by } \Delta\psi' \text{ by } \Delta E_{\text{eff}}]$ for a sweeping analyser. The filtering action can be represented mathematically by an “analyser response function”. The particle count rate (N_{ijk}/t_{acc}) depends on the number flux across the instrument aperture of those particles which the filter allows through, i.e. those particles within the phase space acceptance volume labelled by ijk .

The mathematical representation of the process relates the count rate to an integration over all of velocity space of the product of the phase space density, the number flux across the aperture and the analyser response function. In spherical polar velocity space coordinates, the infinitesimal volume element $d\mathbf{v} = v^2 dv \sin\theta d\theta d\phi$. Since $\Delta E/E = \text{constant}$, we can write $v^2 dv = v^3 dv/v = \text{constant} \cdot v^3$ so that $d\mathbf{v}$ is directly proportional to v^3 . The number flux varies as v . Thus we can expect that the count rate will be related to the phase space density by a term involving v^4 . In fact, the expression relating the mean phase space density in the acceptance volume, f_{ijk} , to the number of counts N_{ijk} determined to have reached the detector in the accumulation time t_{acc} is

$$f_{ijk} = N_{ijk}/(t_{\text{acc}} v_k^4 G_i)$$

where v_k is the velocity corresponding to the centre energy associated with the accumulation interval, and G_i is the analyser response function integrated over the range of velocities able to enter the aperture, a property of the analyser which is independent of the effective energy of the measurement. In principle, G is independent of polar angle for an ideal Top Hat analyser (though not for all analysers) however in practice there is the possibility of a variation of G with polar zone (see below) hence the suffix i . Thorough mathematical derivations of this expression, and of G , are available in texts listed in the bibliography.

Centre Energy

The meaning of the term “centre energy E_k ” (and its corresponding velocity v_k) depends on the energy coverage method used by the analyser. As discussed in Section 5.2.1, an analyser instantaneously accepts particles with energies lying within an energy passband ΔE associated with a passband centre energy E . For a stepping analyser, the accepted energy (and passband) dwell on these values throughout virtually all the accumulation time. In this case the centre energy $E_k = E$. For a sweeping analyser, the instantaneously accepted energy E declines during the accumulation. Then the value E_k represents a form of mean energy for the distribution of energies that are admitted during the accumulation.

Geometric Factor

In discussing the geometry factor it is useful to consider the question of what particles may reach a given polar zone when the analyser is held at constant potential, so the

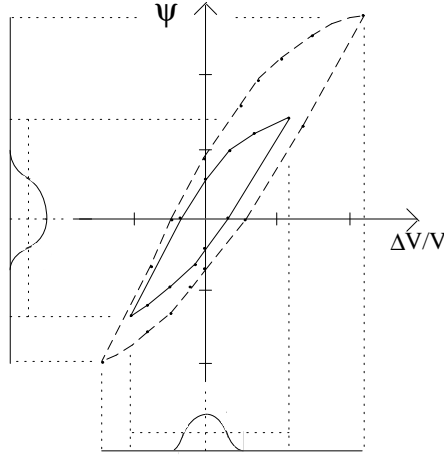


Figure 5.14: The figure shows representative transmission contours for a Top Hat analyser, in $\Delta v/v (= \Delta E/2E)$ versus azimuthal acceptance angle ψ space. The outer, dashed contour represents no transmission (and particles with values of $\Delta v/v$ and ψ outside this contour are not transmitted to the detector). The inner, solid contour defined the 50% transmission level. Also shown are sketches of the transmitted fraction versus $\Delta v/v$ for constant ψ and of the transmitted fraction versus ψ for constant $\Delta v/v$. Together with dotted guide lines indicating the extreme values of $\Delta v/v$ and ψ for which transmission is possible, these are intended to emphasise that the use of the labels $\Delta E/q$ and $\Delta\psi$ to describe the transmitted range of values of E/q and ψ is a major over-simplification. Consequently, precise prediction of geometric factor is difficult, and usually requires the use of numerical simulations of many particle trajectories.

centre energy is constant. In the earlier discussion, we have stated that the analyser will accept particles from a range of energy ΔE , and a range of aperture angles $\Delta\psi$ (these being defined by the analyser shape, collimator and potential) and from a range of polar angle $\Delta\theta$ defined by the detector zone size. In reality, the energy-azimuth angle response for a particular polar angle (e.g. for $\Delta\theta \approx 0$) is more complicated than we have so far suggested; we illustrate it in Figure 5.14. This figure is adapted from measured data (i.e. a real instrument) and shows two closed loops somewhat resembling hysteresis curves. In fact these loops are the contours of the normalised analyser transmission factor mapped in $(\Delta v/v, \psi)$ space, where the dashed line shows a transmission factor of zero and the solid line shows a transmission factor of 50%. (Note that $\Delta E/E = 2\Delta v/v$). The figure also provides rough sketches of the transmission factor versus $\Delta v/v$ for $\psi = 0$ and of the transmission factor versus ψ for $\Delta v/v = 0$, to reinforce the idea that the contours represent a surface.

It can be seen that the acceptance criteria are more restrictive than is suggested by the simplest picture (in which we consider only circular particle trajectories within the analyser) where we have full transmission within a rectangular region of dimensions ΔE , and $\Delta\psi$. Instead, the angle acceptance range will be slightly different for particles with different energies within the energy range ΔE . Similarly, the energy acceptance range varies according to ψ . These properties are the consequence of the way in which some

particle trajectories which are accepted by the collimator nevertheless follow trajectories which strike a hemisphere and thus go undetected. We discussed earlier how a subset of particles which follow elliptical paths are rejected according to their energy. Similarly, arrival at the aperture with non-tangential entry directions leads to rejection of a subset according to their entry angle. The area under the surface represents the number of counts reaching the detector if the analyser is uniformly illuminated from all aperture entry angles and energies, but only a single polar angle ($\Delta\theta \approx 0$). The number of particles reaching the detector under conditions of uniform illumination from the full unit sphere and all energies is naturally larger, as contributions from all polar angles within the range $\Delta\theta$ are included. This number can be characterised by an energy independent geometry factor G_i .

The parameter G will be different for analysers with different physical dimensions (e.g. more particles can enter if the gap between hemispheres is wider, or if the collimator is shortened). Similarly G_i would be different if the polar zone angular width were different, even if the analyser dimensions were not altered (e.g. instruments for solar wind and magnetosphere studies often have reduced polar zone widths in the range of polar angle expected to “see” the solar wind). A different cause for the variation of G_i with polar zone is a deviation from true concentricity of the analyser hemispheres, which alters the aperture acceptance angle as a function of polar angle (less particles can get through if the inter-hemisphere gap is reduced, and vice versa).

Key Issues

This section has explained how we convert from counts to phase space density. The expression for f_{ijk} tells us the average phase space density within the acceptance volume corresponding to the measurement labelled by the suffix ijk . We have stated the location and dimensions of the acceptance volume corresponding to a particular ijk . The measurement is unable to reveal any variations in phase space density within the acceptance volume during the acceptance time. Thus f_{ijk} represents the average value of $f(\mathbf{v})$ in the solid angle and energy range assigned to the coordinates ijk and the mean velocity associated with the measurement is v_k .

The combination of the measurements acquired during a spin consists of a set of samples f_{ijk} at discrete locations within velocity space. As discussed earlier, for many instruments the sampling is not simultaneous in the energy-azimuth regime, and coverage of azimuths at a given energy is incomplete for polar angles near 90° , but over-sampled at polar angles near the spin poles. The over-sampling brings with it the possibility that a feature existing in a narrow azimuth range can be sampled repeatedly with the result that its azimuthal extent is exaggerated. Gaps in solid angle coverage which occur at polar angles near 90° , and at every energy due to the sequential energy range sampling process, bring the possibility that features existing in a narrow energy or azimuth range will not be sampled at all, and so will go undetected. This has been illustrated in Figure 5.10. It is not uncommon practice for data analysts to treat the measurements as pseudo-continuous and as if they were acquired simultaneously. It should be clear from this discussion that the study of events in which the velocity distribution varies over time scales and/or velocity space volumes comparable to those involved in the sampling should be performed with caution and skepticism (if attempted at all).

Finally we note that a continuous distribution function is needed for moments calculations. Moments calculations implicitly treat the measurements as pseudo-continuous and

as if they were acquired simultaneously (just as we have cautioned against in a different context). Details of the process of determining moments from measured distributions are given in Chapter 6.

5.3 Calibration

5.3.1 Importance of Ground and In-Flight Calibrations

We have seen that several instrumental factors need to be properly accounted for in deriving velocity space distributions from recorded count rates. These properties are summarised in the parameters $\epsilon_{\text{DC}}(E)_i$, τ_i and G_i .

Ground calibration of a complete instrument will show how registered count rate (P'_{ijk}/t_{acc}) varies with the particle flux incident on the analyser. Calibration at low count rates allows one to determine performance in the absence of electronic dead time effects and MCP gain depression, but the results nevertheless reflect the combined effects of the geometry factor and the detector efficiency. Ideally the detector efficiency would be determined in the absence of the analyser, however this is not always possible in practice. In some cases the detector efficiency is assumed to be 1 and all effects, including energy dependence and detector performance variation with polar angle are subsumed within a (misnamed) term $G_i(E)$ for each detector zone. The dead time can be estimated by using increasingly large count rates. Ideally, the calibration would determine the bounds of the regime for which a simple dead time correction can be applied, although laboratory calibrations often cannot reach the high count rates expected in space. A very important aspect of the ground calibration is to determine the properties of each zone individually.

Monitoring of in-flight performance is usually required when working with MCP detectors, since the MCP gain tends to decrease as the total quantity of charge collected at the anode increases, i.e. the gain begins to fall as the total number of counted particles grows. Provided the performance variation is well characterised by studying the progression of the decline in gain, calibration factors can be altered accordingly, and the variation can be corrected for in the data analysis. Alteration of the MCP bias voltage during flight can increase the gain to compensate for a decline. In this case there will also be a modification of the calibration factors to be used when interpreting the data. One disadvantage of increasing MCP gain in this way is an increase in the level of noise intrinsic to the MCP. Note that MCP gain depression properties may be altered when the bias voltage is changed; this is usually only significant at very high count rates.

In summary, it is important to use the appropriate calibrations factors for the time at which a distribution is acquired when deriving quantities in scientific units, and to be aware that the pre-flight calibrations may not be the appropriate factors for data acquired several months into a mission. Sometimes there are additional special problems encountered in flight which do not fall into the general categories dealt with here and which require specific knowledge of, and experience with, the particular instrument.

5.3.2 Special Considerations for Multi-Spacecraft Data Analysis

Multi-spacecraft data analysis typically involves making detailed investigations of the differences between conditions at the different spacecraft. If the spacecraft are co-orbiting,

i.e. spacecraft separation distances are small, then measurements must be performed with the maximum possible relative accuracy so that even quite small differences in plasma properties at different spacecraft can be recognised (i.e. the relative error margin must be smaller than the plasma property differences if those differences are to be measured).

Distribution measurement with a set of identical instruments on co-orbiting spacecraft can actually be performed with more confidence than with instruments on solitary spacecraft, since frequent comparisons of performance in very similar plasmas can be made. For example, if the detector on one satellite ages faster than on another, this can be recognised at a very early stage allowing remedial action, and timely correction of calibration factors. It is usually difficult to recognise such variations on a single satellite until the performance is quite degraded. In principle, regular inter-comparison with other instruments which measure a common parameter using a different measurement principle (e.g. electron density can be measured by plasma analysers and various wave experiment techniques) can be a helpful performance check, but in practice, the opportunities for this approach will depend on the nature of the payload and on the adoption of cooperative procedures adopted by the relevant instrument teams.

The comparisons mentioned above, to track detector performance degradation, are only possible if the analyser performance of the various instruments is close to identical, so that near identical velocity space regions are sampled for easy comparison. As has been noted above, non-concentricity of the hemispheres will introduce variation of the geometry factor with polar angle (this could easily happen in a different way on each analyser). The degree of non-concentricity can usually be minimised by careful design and manufacture. Similarly, the system controlling the energy sweeps must be reproduced identically on each instrument, so that the same v_k are used.

5.4 False Counts and Statistical Fluctuations

A number of cautions concerning the application and interpretation of distribution data are provided in Section 6.4.1 of Chapter 6. We will provide a different perspective on some of those issues, and others, in this section.

So far we have discussed how an instrument measures particles from selected regions of velocity space. These particles might be considered to generate “real counts”. There are a number of causes of “false counts” as well as the possibility that the distribution of velocities entering the analyser can in some situations be significantly different to the natural distribution. The latter problem is due to spacecraft charging, in which the spacecraft potential becomes significantly positive (or sometimes negative) thereby accelerating arriving electrons and repelling arriving ions (so that the lowest energy ions do not reach the detector). Usually the spacecraft potential is relatively low, and only makes a significant change to the energies of low energy particles. Many modern spacecraft carry devices intended to control the spacecraft potential, holding it to levels near zero most of the time, for example by using ion emitters.

There are many sources of false counts, of differing levels of importance. The MCP spontaneously generates a few counts per spin—little can be done about this, but it is a minor problem. Photoelectrons can be generated by solar UV photons entering the aperture if they strike the analyser hemispheres. This can be a serious problem, but there are well established design methods for keeping it to acceptably low levels. Secondary electrons

due to energetic particles striking the analyser in the aperture region are also controlled by careful design. Penetrating radiation is a more serious issue. It affects instruments in all orbits of interest in a mild form (the relatively low fluxes of cosmic rays) and is more serious within the regions of trapped, very energetic particles often found in a magnetosphere (“radiation belts”). The particles are sufficiently energetic to penetrate the instrument structure and directly stimulate the MCP (i.e. without passing between the analyser hemispheres). The solution here is to add shielding mass, but the spacecraft mass budgets constrain the degree to which shielding can be carried out. A further source of false registered counts will be noise counts associated with electronic noise falsely triggering the counter electronics.

We have already seen that there are upper limits to the counting rate. There are also lower limits. Clearly, there is a need for the (real) count rate to exceed the noise count rate in order that a signal can be recognised. In radiation belt missions, it is common practice to include a “background” channel which never receives particles from the analyser, but which can provide an indication of the radiation induced false count rate, which can be subtracted from the signals in the other channels to improve signal recovery. A better alternative would be to make a regular measurement of the radiation induced counts in each anode, as the anodes are likely to be affected differently.

Another issue applies even for the case of a noise-free instrument. It is the need for there to be sufficient counts in a distribution that the observation has a low statistical error. The arrival of particles is usually treated as a random process which can be described by Poisson statistics. The impact of statistical fluctuations at low count rates on the calculation of plasma velocity moments is discussed in Chapter 6. When working with a distribution, rather than moments, it is important not to attribute too much significance to bins (i.e. particular ijk measurements) with only one or two counts in them. As we have seen, there are many possible sources of false counts which could be responsible for these data. In assessing the contribution of false counts to measured distributions it is often helpful to initially examine distributions using counts rather than phase space density. The energy weighting associated in converting to phase space density can mislead the unwary as to the significance of data, especially at lower energies.

5.5 Distribution Data Products

Chapters 6 and 7 provide a discussion of the scientific applications of velocity distributions. Here we briefly review some of the data products commonly used to transmit information about measured particle distributions, in order to emphasise the need to understand how these data products are assembled before working with them. Typically the data products are only a well chosen subset of the measured distribution. We emphasise how the properties of the data products are affected by the procedure by which they were acquired with a Top Hat analyser. An understanding of these properties can be important to the data user, particularly where case studies of plasma behaviour at high time resolution are concerned, as such studies usually involve detailed examination of a sequence of individual distributions.

5.5.1 Pitch Angle Distributions

Under conditions where a particle distribution is expected to be gyrotropically symmetric, a measurement of a 2-dimensional pitch angle distribution (if available) is sufficient to characterise the particle distribution. The three-dimensional distribution is reconstructed by simply rotating the pitch angle distribution about the magnetic field. The assumption of gyrotropic symmetry is considered to be often valid for plasma electron distributions, although not usually for ion distributions. Both electron and ion distributions can exhibit gyrophase structure during wave-particle interactions or when within a gyroradius of a plasma boundary (but observations of electron gyrostructure in regions only a gyroradius wide can be challenging as the plasma electron gyroradius is typically so small compared to the relative motion of the instrument and the plasma in an observation time).

Gyrotropy (when there is truly no gyrophase structure) is a property of the distribution function in a frame in which there are no cross-field drifts (e.g. the species rest frame, which does not usually coincide with the spacecraft/instrument rest frame). Electron drifts are typically much smaller than the electron thermal speeds for hot plasma distributions, so that the measurement frame is a good approximation to the species rest frame. Thus for electron measurements, a two dimensional distribution containing the magnetic field direction can simply be selected from a three-dimensional measurement and treated as an electron “pitch angle distribution”. Indeed, it is quite common practice for electron measuring instruments in particular to transmit pitch angle distributions in preference to full three dimensional distributions (there is a substantial reduction in the telemetry required per distribution, allowing a higher rate of transmission of pitch angle distributions than of more complete 3-dimensional distributions). Ion drifts are not usually negligible compared to ion thermal speeds for hot plasma distributions. The removal of ion cross-field drifts by on-board computations is not usually attempted. Instead, ion pitch angle distributions are usually reconstructed during ground-based data analysis working with three-dimensional distributions.

Onboard pitch angle selection requires some indication of the orientation of the magnetic field. Two approaches are possible. One may take data directly from the magnetic field experiment onboard the satellite or else infer the field direction by determining an axis of symmetry in the measured distribution. The former technique is subject to the risk that offsets and calibration factors available onboard for correcting the raw magnetic field data may be inaccurate, leading to an incorrect selection of the field direction and hence of the subset of the measured plasma data to be used as “pitch angle data”. The second technique is also vulnerable to the risk that the distribution does not have a well defined symmetry axis, as for example for an isotropic distribution (or that if there is one, it may not coincide with the magnetic field direction). Some instruments return information defining the direction assumed for the magnetic field, for later comparison with the true direction as determined by proper (ground based) analysis of the magnetic field data. It is also useful to verify whether the magnetic field is stable (slowly-varying with respect to the spin period) during the acquisition of pitch angle data, particularly if the pitch angle data seems unusual.

The magnetic field direction falls within the field of view of a 360° field-of-view analyser twice per spin (provided the direction is not varying rapidly compared to the spin period). At these two times during each spin the instrument can acquire a complete pitch angle distribution in a single energy sweep, since polar zone coverage is simultaneous.

The time taken can be a small fraction of the spin period sweep (e.g. 0.125 s for a 4 s spin and an instrument with 32 sweeps per spin). Thus the instrument can provide snapshots of the pitch angle distribution at a high rate.

Sometimes, a Top Hat will be used to collect data from a more limited 180° field of view. In this case, a complete pitch angle distribution cannot be acquired in a single energy sweep (except in the special case where the spin axis is aligned with the magnetic field direction). A full pitch angle distribution must usually consist of two parts acquired at times separated by half a spin period, each of duration an energy sweep. The pitch angle distribution acquired in this way may be misleading if the magnetic field changes during the interval between the collection of its two parts. Note that if the 180° field of view instrument is twinned with a second identical instrument on the opposite side of the spacecraft, which covers the same energy range, we can in principle revert to the 360° field of view situation.

Instruments with still more restricted fields of view cannot in general acquire full pitch angle distributions, although in special situations it is still possible. For example, the magnetic field direction is generally stable and predictable in the inner magnetosphere, so a spacecraft in an equatorial plane orbit can be arranged with its spin axis roughly normal to the mean magnetic field direction (a “cartwheel” configuration). An instrument with a field of view of say 120° , arranged so that the field of view includes the spin axis direction and the normal to the spin axis (roughly along the magnetic field direction), will look up the field and down the field at two intervals half a spin apart during each spin, in each case covering at least a 90° section of the pitch angle distribution. The full pitch angle distribution is reconstructed by combining the two sets of data from each spin.

Some instruments will return unprocessed counts data from the look directions judged to contain pitch angle data. Other instruments go further and re-sort data into a predetermined set of pitch angle bins. The benefit may be a further reduction in telemetry requirement per distribution, which is accompanied by the risk of some form of error being introduced in the onboard procedure. The data analyst on the ground has less control over the analysis when onboard processing is involved and should be fully aware of the nature of the onboard processing before interpreting the data.

5.5.2 Full Three Dimensional Distributions

A complete three dimensional distribution measurement usually generates a very large amount of data (e.g. see the Summary of Section 5.2.1). A typical instrument will be capable of transmitting such a distribution, but will only do so occasionally, to provide a spin resolution “snapshot”. One important role of the snapshot is its use to verify the proper working of any onboard software, which may be used to generate moments (or more usually the related “moments sums”), pitch angle distributions, etc. In addition, the provision of these snapshots, or at least of reduced three dimensional distributions, is essential for assessing the meaning of the moments. Section 6.6 in Chapter 6 provides a number of examples of this, showing that more than one distribution can give rise to a particular set of moments values. These snapshots are also important as they provide an opportunity to discover plasma behaviour which may be disguised in the process of forming other, more commonly transmitted data products. For example, narrow beams may be observable at the solid angle resolution of the measurement, but may not be apparent in a lower resolution distribution constructed to reduce telemetry loading or for moments generation.

Some reduction in the size of a full three dimensional distribution is possible without severe degradation of the information content. For example, some data may be discarded from polar zones near the spin axis, where over-sampling is known to occur (as discussed in Section 5.2.1).

It will be clear from the earlier discussion that a full three dimensional distribution can be acquired in half a spin period using a 360° field-of-view analyser or in a spin period using a 180° field of view analyser. Particle distributions are usually organised by the magnetic field direction. When this is so, variations of the magnetic field direction on time scales of order a spin or less can change the orientation of the field of view of the instrument with respect to the distribution it is trying to measure, leading to incomplete coverage of some parts of the distribution, and over-sampling of other parts (distinct from the over-sampling discussed in Section 5.2.1). If an examination of magnetic field data at a time resolution of many vectors per spin shows that the field is not static, but also shows that the field rotation is not too complicated, it is possible in principle to reconstruct a partial distribution from the acquired data by carefully determining which regions of velocity space were in fact sampled.

5.5.3 Reduced Three Dimensional Distributions

As full three dimensional distributions are so large, and two dimensional slices are not always useful (e.g. see earlier discussion in Section 5.5.1) a common approach is to produce reduced size distributions which may be transmitted more frequently. As already noted, size reduction can be achieved by discarding parts of the distribution, for example part of the azimuth sectors as for the pitch angle distribution case. In principle, instruments could discard data according to a variety of rules (do not transmit empty bins, do not use this part of the energy range when thought to be in this plasma environment, etc.) but in general instruments do not have the resources to carry out anything other than very simple selection processes, and there is always the risk of discarding important data. It is more common to retain the full energy and solid angle range, but to “decimate” the distribution by combining data in adjacent energy and/or solid angle regions, effectively reducing the energy and/or solid angle resolution of the data. Thus we have a further trade-off between energy/solid angle resolution and time resolution, on top of those enforced by the measurement technique, which we described earlier. Other techniques are adaptations of data compression techniques from imaging instruments. In general these become more “lossy” as the measured distribution becomes less smooth/uniform (and more interesting!) so the trade-off here is between compressibility and loss of accuracy in structured parts of the distribution. The data user should be aware of the process by which the distribution is decimated or otherwise compressed, as the compression process may destroy information about structure which is of interest. Cross-checking with full resolution distributions, which may be occasionally available, is recommended if this risk is likely to be important.

5.5.4 Four Dimensional Distributions

Our discussion of three dimensional distributions can be applied on a species by species basis to distributions from a mass-resolving ion instrument, however it is customary to think of four dimensional distributions as the data product in this case. The mass dimension usually increases the distribution size by a factor of only a few, say four. Nevertheless,

this places a higher demand on the telemetry link than a non-mass resolving instrument, so such distributions are more likely to be telemetered in compressed or decimated form.

5.5.5 Data Compression

So far we have discussed the compression of distribution data products. In order to make good use of telemetry, data compression techniques are usually also applied to individual counts values, to represent the values using fewer bits or bytes. The reconstruction of the original values during ground processing will usually introduce some difference between the onboard value and the ground value. The error will usually be a few per cent at most for elements of a distribution matrix, and will be smaller (by design) where accuracy is considered more important, for example for moments calculated onboard. In some data analysis applications, particularly involving differences between distributions measurements, the decompression error may be significant, however it is disregarded in most single spacecraft work.

5.6 Summary

In this chapter we have described in some detail how a modern plasma analyser measures velocity space density in discrete regions of velocity space. We also explain how these measurements are combined to provide a distribution representative of the plasma phase space distribution. The use of these measurements in the construction of plasma velocity moments and the uses of those velocity moments are the topics of Chapter 6.

We have tried to show that instrument characteristics such as energy per charge coverage and resolution, solid angle coverage and resolution, time resolution and dynamic range are all interdependent, and that a particular instrument is necessarily designed for a particular plasma environment. That instrument will often also be operated in other environments, in which its performance may be less satisfactory. Sometimes this will be obvious to a data analyst, but not always, so the analyst is strongly advised to spend some time verifying that the questions they wish to address can in fact be answered with the data collected by the instrument, while in the environment of interest to the analyst.

We have also briefly discussed the range of data products which may be produced by the instrument. Again it is worth noting that much of the telemetry from an instrument may not represent the best resolution (in energy per charge, solid angle or time) available. Less commonly used data products may be more appropriate for some studies.

The importance of calibration and in-flight monitoring of instrument characteristics is also highlighted. A data analyst can normally expect that data will be supplied from an instrument team with a reliable calibration, but should be aware that calibrations are sometimes reassessed as an increasingly large data set is gathered, so that data carries a greater risk of recalibration relatively soon after it has been acquired. We note that the possibility of inter-calibration between instruments on spacecraft which are co-orbiting at relatively small separations gives grounds for confidence that good relative accuracy can be achieved and maintained between the spacecraft.

Acknowledgements

The authors wish to acknowledge helpful discussions with Andrew Coates, David Walton and Adrian James, and useful suggestions made by the referee.

Bibliography

Detailed descriptions of the hot plasma and energetic particle detecting instrumentation (CIS, PEACE and RAPID) for the four spacecraft Cluster I mission, now to be rebuilt for the Cluster II mission are found in:

Escoubet, C. P., Russell, C. T., and Schmidt, R., editors, *The Cluster and Phoenix Missions*, Kluwer Academic, Dordrecht, Boston, London, 1997, reprinted from *Space Science Reviews*, Vol. 79, Nos. 1–2, 1997

Descriptions of the instrumentation onboard the Polar and Wind satellites are collected in:

Russell, C. T., editor, *The Global Geospace Mission*, Kluwer Academic, Dordrecht, Boston, London, 1995, reprinted from *Space Science Reviews*, Vol. 71, Nos. 1–4, 1995

And for the four spacecraft (two pairs) Interball mission:

Galperin, Y., Muliarchik, T., and Thouvenin, J.-P., Interball: mission and payload, published by IKI and CNES, 1995

The texts listed in the bibliography section of Chapter 6 (page 157) which refer to moments accuracy are also recommended for readers of this chapter. Similarly, the texts listed there concerning the use of spherical harmonic analysis may also be of interest to people working with particle distribution data from hot plasma instruments.

The original paper discussing the Top Hat analyser:

Carlson, C. W., Curtis, D. W., Paschmann, G., and Michael, W., An instrument for rapidly measuring plasma distribution functions with high resolution, *Adv. Space Res.*, **2**, 67, 1983. Figure 5.14 is based on a figure from this paper.

A review which discusses a wider range of composition measurement techniques than are addressed in this chapter, and which references further interesting review articles concerning measurement techniques for particular plasma environments, such as the solar wind, magnetospheres or cometary plasma:

Gloeckler, G., Ion composition measurement techniques for space plasmas, *Rev. Sci. Instr.*, **61**, 3613, 1990.

An example of in-flight cross-calibration between different kinds of instrument:

Kessel, R. L., Johnstone, A. D., Brown, C. C., and Woolliscroft, L. J. C., A comparison of the ion density measured simultaneously by a wave and a particle instrument, *J. Geophys. Res.*, **96**, 1833, 1991.

A paper that provides a more mathematical discussion of converting counts to phase space density than the chapter text, has an interesting example of data compression (transmitting only what is essential for solar wind measurements) and describes an unusual in-flight

calibration technique:

Johnstone, A. D., Coates, A. J., Wilken, B., W. Stüdemann, Weiss, W., Irelli, R. C., Formisano, V., Borg, H., Olsen, S., Winningham, J. D., Bryant, D. A., and Kellock, S. J., The Giotto three-dimensional positive ion analyser, *J. Phys. E; Sci. Instruments*, **20**, 795, 1987.

— 6 —

Moments of Plasma Velocity Distributions

GÖTZ PASCHMANN

*Max-Planck-Institut für extraterrestrische Physik
Garching, Germany*

ANDREW N. FAZAKERLEY

*Mullard Space Science Laboratory
Holmbury St. Mary, United Kingdom*

STEVEN J. SCHWARTZ

*Queen Mary and Westfield College
London, United Kingdom*

6.1 Introduction

This chapter is intended as a tutorial on plasma moments: their definition, computation, interpretation, accuracy, and use. The chapter builds upon Chapter 5 which describes the measurements of the particle distribution functions from which the moments are constructed. Velocity moments are the plasma parameters that most easily lend themselves to quantitative multipoint analysis. For a description of the analysis of multipoint moments the reader is referred to Chapter 17.

6.2 Definitions

6.2.1 Single-Species Moments

The moments of the velocity distribution function, $f(\mathbf{v})$, of a given particle species are defined as

$$M_n \equiv \int f(\mathbf{v}) \mathbf{v}^n d^3v \quad (6.1)$$

where \mathbf{v}^n is an n -fold dyadic product, and d^3v is the volume element in velocity space.

The moments commonly used are the zero-, first-, second-, and third-order moments, i.e., the number density, N , the number flux density vector, $N\mathbf{V}$, the momentum flux density tensor, Π , and the energy flux density vector, \mathbf{Q} , respectively:

$$N = \int f(\mathbf{v}) d^3v \quad (6.2)$$

$$N\mathbf{V} = \int f(\mathbf{v}) \mathbf{v} d^3v \quad (6.3)$$

$$\Pi = m \int f(\mathbf{v}) \mathbf{v} \mathbf{v} d^3v \quad (6.4)$$

$$\mathbf{Q} = \frac{m}{2} \int f(\mathbf{v}) v^2 \mathbf{v} d^3v \quad (6.5)$$

Higher moments can be calculated, but do not have a simple physical interpretation.

Multiplying N and $N\mathbf{V}$ with the particle mass m , one obtains the mass density, ρ , and the mass flux density vector, $\rho\mathbf{V}$, respectively. From the zero- and first-order moments one can compute the average velocity, \mathbf{V} , of the particles, referred to as the bulk velocity:

$$\mathbf{V} = (N\mathbf{V})/N \quad (6.6)$$

Converting the momentum flux tensor and the energy flux vector to the frame where the average velocity is zero, one obtains the pressure tensor, \mathbf{P} , and the heat flux vector, \mathbf{H} :

$$\mathbf{P} = \mathbf{\Pi} - \rho \mathbf{V} \mathbf{V} \quad (6.7)$$

$$\mathbf{H} = \mathbf{Q} - \mathbf{V} \cdot \mathbf{P} - \frac{1}{2} \mathbf{V} \text{Tr}(\mathbf{P}) \quad (6.8)$$

One can also write down definitions for \mathbf{P} and \mathbf{H} directly in terms of the velocity differences relative to the bulk velocity:

$$\mathbf{P} = m \int f(\mathbf{v}) (\mathbf{v} - \mathbf{V}) (\mathbf{v} - \mathbf{V}) d^3v \quad (6.9)$$

$$\mathbf{H} = \frac{m}{2} \int f(\mathbf{v}) |\mathbf{v} - \mathbf{V}|^2 (\mathbf{v} - \mathbf{V}) d^3v \quad (6.10)$$

Using the definition $\mathbf{P} \equiv N k \mathbf{T}$, where k is the Boltzmann constant (1.38×10^{-23} J/K), one can convert the pressure tensor into a temperature tensor. Note that the momentum flux, pressure and temperature tensors are symmetric tensors.

Scalar pressures and temperatures can be obtained from the trace of the associated tensors: $p \equiv \text{Tr}(\mathbf{P})/3$ and $T \equiv \text{Tr}(\mathbf{T})/3 = p/(Nk)$. Temperatures are converted to thermal energies kT , usually measured in eV, by noting that 1 eV corresponds to 1.16×10^4 K. Temperature and thermal energy are related to the thermal velocity, defined via $kT = (m/2)v_{\text{th}}^2$. For the case of electrons and protons the relations are $v_{\text{th},e} = 593\sqrt{kT_e} = 5.50\sqrt{T_e}$ and $v_{\text{th},p} = 13.8\sqrt{kT_p} = 0.128\sqrt{T_p}$, respectively, when the velocities are measured in km/s. For a discussion of the meaning of temperature and thermal velocity, see Section 6.5.

6.2.2 One-Fluid Moments

Separate sets of moments can be computed for each particle species. From these individual sets of moments one can derive single-fluid quantities. For illustration we choose the case of two species, e.g., electrons and protons, or two different ion species. The equations are easily generalised to more than two species.

$$\rho = N_e m_e + N_i m_i = \rho_e + \rho_i \quad (6.11)$$

$$N = N_e + N_i \quad (6.12)$$

$$\mathbf{V} = (\rho_e \mathbf{V}_e + \rho_i \mathbf{V}_i)/(\rho_e + \rho_i) \quad (6.13)$$

$$\mathbf{P} = \mathbf{P}_e^{CM} + \mathbf{P}_i^{CM} \quad (6.14)$$

It should be noted that in general the total pressure is not the sum of the partial pressures if the latter are computed in the bulk-velocity frames of the individual species. Instead the

individual pressures must be computed in their joint centre-of-mass frame, which we will distinguish from the individual bulk-velocity frames by the superscript CM. Only if the two species have identical bulk velocities is the total pressure simply the sum of the partial pressures computed in the individual bulk-velocity frames.

If the two species are electrons and protons, then $\rho \approx \rho_i$ and $\mathbf{V} \approx \mathbf{V}_i$, and there are two new quantities, namely space charge and current density:

$$\sigma = e(N_i - N_e) \quad (6.15)$$

$$\mathbf{j} = e(N_i \mathbf{V}_i - N_e \mathbf{V}_e) \quad (6.16)$$

where e is the elementary charge, $e = 1.6 \times 10^{-19}$ C.

6.3 Computations

As described in Chapter 5, modern plasma instruments which measure the three-dimensional velocity distribution function acquire thousands of samples per distribution. For example, an instrument which samples 30 energies E , 16 azimuth angles ϕ , and 8 elevation angles θ , will collect a total of 3840 count rate samples, $C_{ijk} = C(E_i, \phi_j, \theta_k)$, per spacecraft spin. Each of these samples is itself an integral over some range of energies and angles, the intrinsic acceptance volume of the instrument. The elevation angle samples are typically abutting, while the azimuth samples are usually spaced much wider apart than the intrinsic acceptance, thus leaving gaps that are not covered. Energy coverage is usually achieved by sweeping or stepping the intrinsic narrow energy window exponentially in time and accumulating the counts in a number (30 in our example) of bins. This way the energy channels are contiguous, with centres spaced equidistantly in $\log E$. Typical values for $\Delta E/E$ are 0.3 and the dynamic range in energy is ≈ 2000 . A $\Delta E/E$ of 0.3 corresponds to a $\Delta v/v$ of 0.15, while a 22.5° (i.e., 0.4 rad) angular resolution corresponds to a $\Delta v/v$ of 0.4. The resolution in velocity magnitude is thus usually better than the resolution in angle.

From the matrix of count rates the moments can then be computed according to the definitions given in Section 6.2, by noting that, for an electrostatic analyser, $f_{ijk} \propto C_{ijk}/v_{ijk}^4$, with a proportionality factor given by the detector response function (see Chapter 5). The integrations are replaced by summations and usually the assumption is made that $f(\mathbf{v})$ is constant over the integration volume elements, d^3v . Noting that $d^3v = v^2 dv = v^3 dv/v$, and dv/v is constant for these types of analysers, one obtains the expressions given in Table 6.1 where the net velocity exponents are one less than in the equations in Section 6.2.

Table 6.1 illustrates the large number of different cosine and sine weights in the integration over angle. These weights are computed as integrals over the angle channels. The C_{ijk} in the table are not the raw count rates, but have already been corrected for detector dead time and energy-dependent detection efficiency, and a factor $\cos \theta$ has been applied that removes the over-sampling at high elevation angles (see Chapter 5). Not shown in Table 6.1 are the factors involving the detector response function, the size of the velocity space elements, and the particle mass in front of the expressions that can be applied after the summations.

Instead of assuming that f is constant over the integration volume element, one can also use a trapezoidal method for the summations, thus allowing f to vary linearly between

adjacent measurements. Simulations have, however, shown that it makes a negligible difference (see Section 6.4.2).

A more sophisticated scheme has been implemented for the moments from the Hydra instrument on the Polar spacecraft. There the integration is done with cubic splines. Furthermore, integration over speed is effectively extended from 0 to ∞ . This is done by mapping the infinite interval onto a finite one chosen such that it is centred where the peak contribution to a particular moment occurs in the measured distribution. This guarantees that the actual measurements define the transformed integration within most of the finite interval. The scheme then exploits that $f \rightarrow 0$ for $v \rightarrow \infty$ and that the integral vanishes at $v = 0$. Cases where this scheme becomes questionable are identified by computing, separately for each of the moments, the fractional contribution to the integral from the energy range beyond the highest energy of the actual measurements. For electrons, where problems occur usually at low energies, care is taken to remove photoelectron counts, and f is then shifted in energy according to the spacecraft potential as measured by the double-probe instrument. Finally, f is defined at speeds lower than actually measured, by doing a Taylor expansion around the origin which matches the measurements at the first few energy steps. The Hydra method has been checked by simulations, as described in Section 6.4.2.

For completeness we note that in the early days, instruments did not have full three-dimensional (3-D) velocity-space coverage and/or resolution and measured only two-dimensional (2-D) distribution functions. Some measurements provided a cut through the 3-D distribution, i.e., $f_{ij} = f(E_i, \phi_j)$. Other instruments did not resolve θ either, but covered some range in θ up to some maximum value, e.g., 55° in the case of the 2-D detectors of the Los Alamos/Garching Fast Plasma Experiment on ISEE-1 and -2. Moment integration from 2-D distributions required some assumption on the behaviour of the distribution as a function of the elevation angle. For convected distributions that were nearly isotropic in the bulk-velocity frame, the computed density and 2-D vector and tensor quantities were reasonably accurate.

6.4 Accuracy

In addition to the errors they inherit from the underlying distribution functions (see Chapter 5), the moments suffer from limitations introduced by the implied integration procedures. Much of the difficulty arises from the extraordinary dynamic ranges in particle velocity (or energy), and in the flux levels and angular widths of the incident distribution that one has to cope with. Densities range from 0.01 cm^{-3} in the lobes of the geomagnetic tail to several 100 cm^{-3} in the (highly compressed) magnetosheath. The necessary energy range is indicated by the temperatures to be encountered which range from $<10 \text{ eV}$ ($<10^5 \text{ K}$) for solar wind protons to $>10 \text{ keV}$ ($>10^8 \text{ K}$) for plasma sheet protons. The dynamic range in fluxes is 10^7 for ions and 10^5 for electrons. The angular widths of the distributions are determined by the ratio of thermal and bulk speeds, and range between $\approx 20^\circ$ for highly supersonic distributions, such as the solar wind ions, to 360° for subsonic distributions found elsewhere. These numbers illustrate the formidable task a single instrument faces if it is to be used in all these regimes which is the usual situation. Note that in this paper we have ignored ionospheric (and plasmaspheric) plasmas that would increase these ranges even further, having densities of 10^3 cm^{-3} and higher, and temperatures of

Table 6.1: Moment summations as carried out on board AMPTE/IRM for ions and electrons. Factors involving detector geometry factor, integration-volume elements, or particle mass have been omitted because they can be applied on the ground.

Density:

$$N = \sum_k \frac{1}{V(E)} \sum_{\phi} \sum_{\theta} C(\theta, \phi, E)$$

Number flux density vector:

$$NV_x = \sum_E \sum_{\phi} \cos \phi \sum_{\theta} \cos \theta C(\theta, \phi, E)$$

$$NV_y = \sum_E \sum_{\phi} \sin \phi \sum_{\theta} \cos \theta C(\theta, \phi, E)$$

$$NV_z = \sum_E \sum_{\phi} \sum_{\theta} \sin \theta C(\theta, \phi, E)$$

Momentum flux density tensor:

$$P_{xx} = \sum_E V(E) \sum_{\phi} \cos^2 \phi \sum_{\theta} \cos^2 \theta C(\theta, \phi, E)$$

$$P_{yy} = \sum_E V(E) \sum_{\phi} \sin^2 \phi \sum_{\theta} \cos^2 \theta C(\theta, \phi, E)$$

$$P_{zz} = \sum_E V(E) \sum_{\phi} \sum_{\theta} \sin^2 \theta C(\theta, \phi, E)$$

$$P_{xy} = \sum_E V(E) \sum_{\phi} \cos \phi \sin \phi \sum_{\theta} \cos^2 \theta C(\theta, \phi, E)$$

$$P_{xz} = \sum_E V(E) \sum_{\phi} \cos \phi \sum_{\theta} \cos \theta \sin \theta C(\theta, \phi, E)$$

$$P_{yz} = \sum_E V(E) \sum_{\phi} \sin \phi \sum_{\theta} \cos \theta \sin \theta C(\theta, \phi, E)$$

Energy flux density vector:

$$H_x = \sum_E V^2(E) \sum_{\phi} \cos \phi \sum_{\theta} \cos \theta C(\theta, \phi, E)$$

$$H_y = \sum_E V^2(E) \sum_{\phi} \sin \phi \sum_{\theta} \cos \theta C(\theta, \phi, E)$$

$$H_z = \sum_E V^2(E) \sum_{\phi} \sum_{\theta} \sin \theta C(\theta, \phi, E)$$

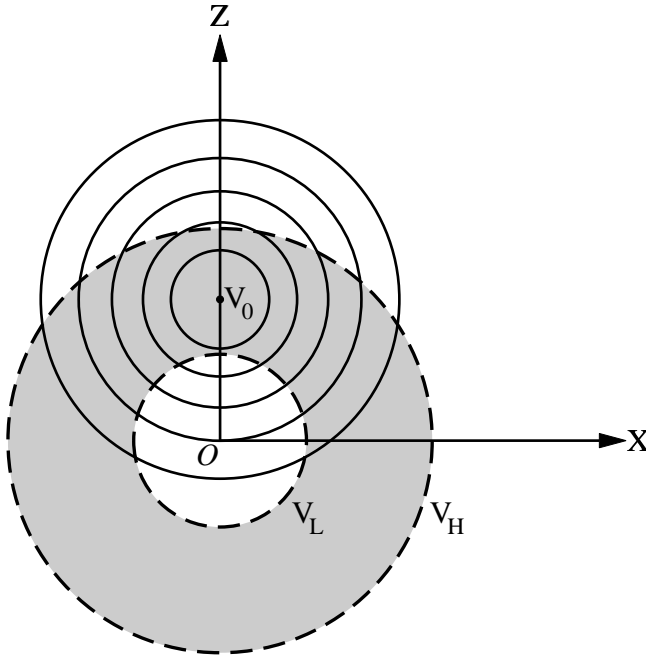


Figure 6.1: Illustration of the effect of a velocity range of the measurements (grey annulus) that does not completely cover the input velocity distribution, represented by the isocontours centred on the bulk velocity V_0 .

1 eV or less, and requiring dedicated measurement techniques.

In this section we will first discuss qualitatively all the factors that can affect the moments, and then look at the results of various studies that have quantitatively evaluated the resulting errors.

6.4.1 Error Sources

Energy Range

The measurements are made over an energy range that cannot be arbitrarily large. First, analyser deflection voltages (and thus the range of energies a detector can accept) are technically constrained to a limited dynamic range, typically a few thousand. Second, counting statistics sets a minimum dwell time per energy channel which limits the energy range that can be covered in a certain time. If the distribution function extends outside this finite range, as schematically illustrated in Figure 6.1, all moments are adversely affected, but to varying degrees because of their different velocity weights.

If the distribution extends to energies below the low-energy limit of the measurements, density and pressure are both underestimated, but the density more so than the pressure because of the higher velocity weighting of the pressure. As a consequence, the computed temperature, being proportional to p/N , is too high. The computed bulk velocity magnitude will be overestimated, but its direction will be unaffected, unless the temperature is

very anisotropic and magnetic field and bulk velocity are not aligned. If the distribution extends beyond the high-energy limit of the measurements, density and pressure are still underestimated, but now the pressure more so than the density, with the result that the computed temperature is too small. The bulk speed will also be underestimated in this case. Furthermore, any time a significant portion of the true distribution is excluded from the measurements, the distribution no longer is symmetric in the bulk-velocity frame, and false pressure and temperature anisotropies are introduced.

The effect of the limited energy range is particularly acute at low energies where a population of very cold (of order 1 eV) particles of ionospheric origin would be missed entirely by typical hot-plasma instruments. For electrons the problem occurs also in the solar wind where thermal energies are ≈ 10 eV (corresponding to $\approx 10^5$ K), close to the low-energy limit of typical hot-plasma instruments.

There are several ways to estimate the contributions from below the low-energy limit. If one has reason to believe that the low-energy portion of the distribution can be approximated by a Maxwellian (such as for the core of the solar wind electrons), one can estimate the contribution to the moments of the missed electrons by fitting a Maxwellian to the measurements at low energies, taking care not to include channels contaminated by photoelectrons. If one is only interested in improving the density, one can calculate the fraction of the density that resides at the energies not measured, under the assumption that the distribution is Maxwellian with a temperature equal to that computed from the moments. The measured density is then adjusted accordingly. This correction, which was employed on AMPTE/IRM for the electrons in the solar wind, is an underestimate because the temperature that is being used is too high. A third possibility is to approximate the missing part of the distribution by linearly interpolating $\log f$ through zero speed between oppositely directed azimuth-angle channels. This scheme seems plausible for magnetosheath electrons which often have a flat-topped distribution at the lowest energies.

For ions the problem with cold populations is less severe because even a small bulk velocity will shift the ion energies into a comfortable energy range. For example, an oxygen ion at 20 km/s has an energy of already 34 eV. The question of detection is then more one of energy resolution rather than energy range. (For electrons bulk motion is no help because even a 1000 km/s corresponds to only 3 eV.) For ions, problem often occur at high energies. Assuming a temperature of 2×10^8 K (corresponding to a thermal energy of ≈ 20 keV), not uncommon in the Earth's plasma sheet, protons with a Maxwellian distribution have $\approx 25\%$ of their density and $\approx 50\%$ of their pressure at energies above 40 keV, the upper energy limit of typical hot-plasma instruments. To catch 90% of the proton pressure would require measurements up to 80 keV. Because real distributions tend to fall off less steeply at high energies than a Maxwellian, the real situation is even worse, particularly if the plasma has a substantial bulk speed which shifts the distribution to even higher energies. The problem can be overcome to some extent by computing (and adding) the partial pressure of ions above 40 keV from spectra measured by an energetic particle instrument, as done on AMPTE/IRM.

Note that the scheme employed for the Hydra instrument on Polar technically integrates over energies between 0 and ∞ , as described in Section 6.3.

Velocity Space Resolution

Features in the distribution functions that are comparable with or narrower than the integration volume elements cannot be properly handled by the moment integration. Cold ion beams such as the solar wind are an example. For an instrument with 22.5° angular resolution, for example, a cold beam will have almost all counts concentrated in a single angle sample. The moment computation will then attribute half of the 22.5° angle-spacing, or 20% of the bulk speed, to the thermal speed. Assuming a bulk speed of 400 km/s, the apparent thermal speed is 80 km/s, equivalent to almost 35 eV or 4×10^5 K. At 700 km/s the apparent temperature would be $\approx 10^6$ K. These values are much in excess of the true solar wind ion temperature which is typically $\leq 10^5$ K.

The bulk speed of a cold beam will be determined fairly accurately, to within the thermal speed, as long as the beam is sampled at all, while the error in the density will strongly depend on where (relative to the peak) the distribution is sampled. On AMPTE/IRM the azimuth sampling pattern was rotated by one quarter of the sample spacing in successive spins. In the solar wind (or other cold-beam environments), the computed ion density therefore exhibited a strong modulation over a four-spin cycle. Averaging the moments over one complete four-spin cycle removes most of the density modulation, but at the expense of reduced time resolution. The computed temperature still is too large because the intrinsic angular acceptance of the instrument, particularly in polar angle, is itself already too large for the solar wind. To overcome these difficulties, some ion instruments (such as the AMPTE/UKS instrument), have used narrower intrinsic acceptance angles and added extra samples near the solar direction. A better solution is to add a separate instrument specifically designed for the measurement of narrow beams.

Time Variations

Moment integration assumes that no significant time variations occur while the part of velocity space with significant counts is being sampled. For a supersonic distribution, occupying only a small angular sector, this takes only a fraction of a spacecraft spin. For a “hot” distribution, it takes an entire spin. Either way, the time between successive samples of the same element in velocity space is much larger (typically by a factor of several hundred) than the duration of each sample. In the language introduced in Chapter 2, the moments are extremely under-sampled, and thus subject to severe aliasing.

Part of the stability argument concerns the magnetic field. If the distribution exhibits an anisotropy ordered by the magnetic field direction, the moments will be affected if the magnetic field rotates while the distribution function samples are still being collected. As the magnetic field is measured many times per spin, such occurrences can be recognised in the data and the moments flagged accordingly. The off-diagonal terms in the pressure (temperature) tensors are particularly sensitive to variations in the magnetic field direction, as discussed in Section 6.5.

An interesting case occurs if the spacecraft crosses a density boundary in the middle of the measurement cycle. In this case one segment of azimuth-angle samples have high counts, the other segment has low counts. Computing moments of this distribution produces an artificial bulk velocity as well as anisotropic temperatures, but only for the interval straddling the density change. Such an occurrence is easily recognisable.

Another problem can occur if the spacecraft crosses a boundary separating plasmas

with very different distribution functions, such as the bow shock. If for one reason or another the distribution is not properly measured on one side of the boundary, then fundamental quantities such as the density jump across it are not determined correctly.

Counting Statistics, Averaging

The total number of counts recorded in a distribution is proportional to the incident particle flux and to the geometric and efficiency factors of the instrument. The input flux in turn is proportional to density times thermal speed of the plasma, $N v_{\text{th}}$. For a given v_{th} the relative uncertainty in the density will vary as $1/\sqrt{N}$ because the counts obey Poisson statistics. For a density ranging from 0.01 to 100 cm^{-3} , the associated statistical error in the computed density thus varies by two orders of magnitude.

Along an orbit such as that of Cluster, the incident fluxes extend over a dynamic range that is approximately 7 orders of magnitude in the case of ions. This is more than a detector with a fixed geometry factor can handle. For this reason the ion instrument (CIS) employs detectors with different geometry factors, one each for low and high flux levels. For electrons the dynamic range in incident flux is less, approximately 5 orders of magnitude. Still, the PEACE electron instrument developed for Cluster consists of two separate detectors with different geometry factors.

The different velocity weights of the counts (see Table 6.1 on page 129) shift the emphasis in the moments summations from where the peak counts actually occur, thus reducing statistical accuracy, except for the number flux vector where no shift occurs. Still, one can take the total number of counts accumulated in a distribution as a crude measure of the statistical accuracy, as demonstrated in Figure 6.4. This parameter is, for example, recorded by the PEACE electron instrument on Cluster.

Averaging can overcome some of the counting statistics problems. The reason for this is that the moment computation is a linear operation and therefore $M(f^1) + M(f^2) + \dots = M(f^1 + f^2 + \dots)$, where the f^j are the distributions measured in a sequence of spins. Adding the raw moments from several spins is thus equivalent to computing the moments from the summed distribution, i.e., effectively adding up the counts from several spins. But it should be noted that one should average the raw moments before computing the derived quantities. For example, to improve the statistics on the bulk velocity, one must add the moments N and $N\mathbf{V}$ from a number of spins and then compute \mathbf{V} by dividing the summed $N\mathbf{V}$ by the summed N . Averaging the \mathbf{V} directly does not improve the statistics.

Background Counts

Another cause of error are counts which are not due to plasma particles. These “background” counts can have several sources, including (high-energy) penetrating particles, detector dark counts, solar UV induced counts, photoelectrons, or electronic noise. Penetrating particle background counts are only important where there is a high flux of energetic particles, e.g., in the radiation belts or during solar energetic particle events. UV-induced counts are only seen when the detector is facing the sun. A particularly severe problem are photoelectrons from spacecraft surfaces because they can dominate the count rates of electron instruments at low energies. To minimise errors, the magnitude of the various backgrounds must be estimated or empirically determined and subtracted before moments summation.

Spacecraft Charging

A spacecraft in a plasma will charge to an electrical potential that is determined by a balance of all the currents from electrons and ions that are incident on or emitted from its surface, including the photoelectron current if illuminated by the sun. Spacecraft potentials thus depend on particle fluxes and in magnetospheric plasmas are typically less than tens of volts in sunlight for an electrostatically clean spacecraft. When photoelectrons dominate, the spacecraft potential is positive. Spacecraft potentials change the velocity of the incident low-energy particles and can prevent particles from reaching the detectors altogether. If one knows the spacecraft potential from other measurements, or if one can estimate it from features in the measured distributions themselves, one can correct for the velocity change, but only for those particles that can still reach the detector. It is thus advantageous to reduce the spacecraft potential by some active control. Significant reduction of positive spacecraft potentials has been achieved with indium ion emitters strong enough to dominate the current balance. Holding the spacecraft potential near zero also allows the photoelectrons to disperse so that they are not measured along with the natural plasma electrons.

Effect of Thermal-to-Bulk-Speed Ratio

If the bulk velocity is large compared with the thermal velocity (“supersonic flow”), the particles (and thus $f(v)$) occupy a relatively small region in velocity space. If the measurements have adequate resolution, it is then no problem to determine the bulk velocity reliably, as illustrated above for the case of solar wind ions. If on the other hand the bulk speed is much lower than the thermal speed (extreme “subsonic” flow), the shift of the distribution caused by the bulk velocity introduces only small changes in the count rates at any position in velocity space, and is thus harder to measure accurately.

The problem is particularly acute with electrons which in space plasmas are always subsonic. Electron temperatures are similar to those of the ions, but at a given temperature the electron thermal speed is the square root of the mass ratio (≈ 43) times that of the protons. Furthermore, the maximum bulk speeds that occur in space plasmas are of order 1000 km/s, corresponding to only 3 eV for electrons, but to almost 6 keV for protons (and still higher for any heavier ions that might be present). By comparison, the electron thermal speed typically corresponds to tens to hundreds of eV. This explains why electron bulk velocities are harder to measure than ion bulk velocities. For the same reason electric currents, which are proportional to the difference between electron and ion bulk velocities, are difficult to directly determine from the ion and electron moments, too.

Ion-Composition

If ion species are not resolved, as is true for any energy-per-charge analyser, moments are usually computed under the assumption that all ions are protons. Although it is true that space plasmas usually are dominated by protons, even a small addition by number of heavier ions can drastically change the mass density. For example, the presence of only 6% oxygen ions by number causes the computed mass density to be off by a factor of almost two.

If two ion species are present at the same velocities, for example protons and singly-charged oxygen, an E/q analyser will record the two species at energies that differ by

their mass ratio, which is 16 in our example. Under the pure proton plasma assumption, this will be interpreted as a double-beam situation and the computed moments can be grossly in error, in particular the temperature. If the species involved are known, one can fit model distributions to each peak and compute separate moments for each species. Before the advent of mass-resolving instruments, this method has been employed for E/q ion measurements in the solar wind to separate the contribution of alpha particles (whose abundance is of order 5%) from those of the dominant protons.

Calibration

All the moments in equation 6.1 inherit from the measurements of f the proportionality to the geometry factor of the analyser and a factor representing the normalised detector efficiency (see Chapter 5). These factors can be applied after the moment summation. For microchannel plate detectors (MCP's), the detection efficiency depends on particle energy and the applied voltage, and can be spatially non-uniform. This efficiency corrections must already be included during moments summation. Geometric factors can be calibrated with particle beams in the laboratory, but can also be computed with ray-tracing techniques. Efficiencies are usually determined by measurement. Absolute uncertainties in instrument response functions are usually 10% or higher, while relative uncertainties (between instrument of the same design and build-status) are as low as a few per cent. As the bulk velocity involves a ratio between raw moments, common factors drop out. The bulk velocity is thus independent of the geometry factor.

On-board versus Ground Computations

Beginning with AMPTE/IRM, moment computations are now often done on board, such as in the ion and electron instruments developed for Cluster and for Equator-S. This is because a distribution function consists of several thousand samples and there is usually not enough telemetry to transmit it at the full time resolution. The moments up to the heat flux, on the other hand, amount to just 13 numbers per species and therefore require only a modest bitrate. Computing the moments already on board is mandatory if high time-resolution is required.

The disadvantages of on-board computation are limited numerical accuracy and less flexibility. On the ground it is easier to use more complicated integration schemes, or to adjust the integration to correct for photoelectron contamination, or to compute partial moments in case a distribution is composed of several distinct populations. On the ground one can also easily, and retrospectively, change the energy dependence of the detector efficiency, correct for any non-uniformity in detector response, and even deal with cases where excessively high incident fluxes have caused gain depressions in the MCP's. Full analysis of the errors from counting statistics is also easier on the ground.

To combine the advantages of on-board and ground computations, a reasonable compromise is a solution where moments are computed on board and transmitted with full time-resolution, but moments are then also computed on the ground from the fewer distribution functions that have been transmitted.

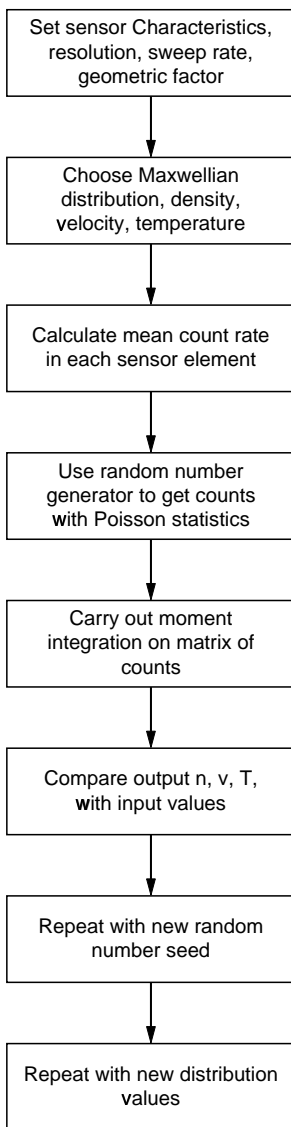


Figure 6.2: Simulation flowchart for moment computations.

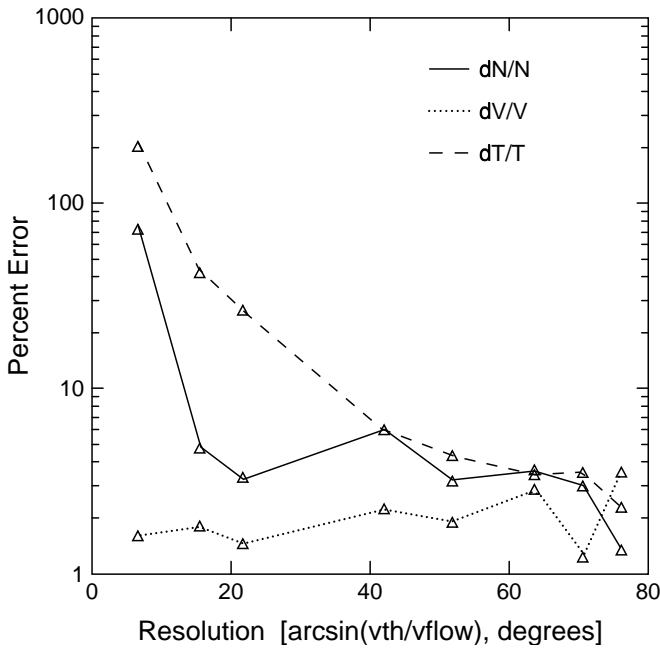


Figure 6.3: Effect of the angular resolution of the measurements on the accuracy of the moments for the AMPTE/UKS ion instrument. The density was assumed at 10 cm^{-3} , the bulk speed at 100 km/s , and the angular spread of the input distribution, measured by the ratio between thermal and bulk speeds, was varied as shown along the abscissa. The angular resolution of the measurements was fixed at 22.5° .

6.4.2 Simulation Results

Computer simulations of the effects discussed in the previous section have been carried out at various institutions involved in space plasma measurements. The method consists in choosing a distribution with a given density, bulk velocity and temperature; translating this input distribution into the count rates that a real instrument with specified response characteristics would produce; computing the moments from these “measured” distributions, and comparing them with the moments of the input distribution, i.e., those moments that a perfect instrument would produce. Figure 6.2 illustrates the process. In the following we will discuss some of the results obtained this way for instruments on previous and future missions. Because input distributions have commonly been taken as Maxwellians, the resulting moments had, by definition, no off-diagonal terms in the pressure (and temperature) tensors, nor a heat flux. Errors in those terms could therefore not be determined.

The first example refers to results obtained for the ion plasma instrument flown on AMPTE/UKS that in one standard operating mode covered an energy/charge range between 10 V and 20 keV 16 times per satellite spin, i.e., with 22.5° resolution in azimuth and with similar resolution in elevation angle. Figure 6.3 shows the effect of the velocity-space resolution on the accuracy of the moments. A distribution with fixed density and bulk speed was assumed, and the thermal spread of the distribution was varied by changing the

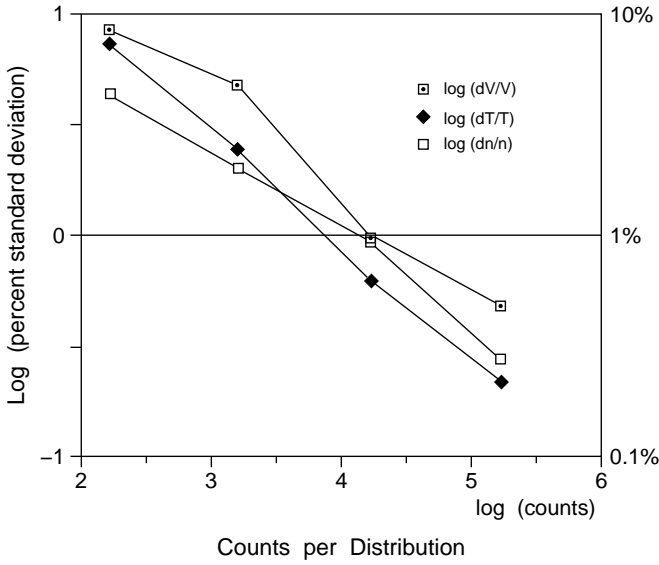


Figure 6.4: Effect of increasing the number of counts by increasing the input density from 10 to 10^4 cm^{-3} . The velocity and temperature were held constant at 100 km/s and 10^6 K , respectively.

temperature. The dramatic effect of the angular width of the distribution on temperature, discussed in the previous section, is clearly evident. As expected, the density is only affected if the thermal spread is close to the resolution of the instrument, and the bulk speed is hardly affected at all. Figure 6.4 shows the expected reduction in the statistical errors that result when the density is increased while keeping the other parameters constant.

The next example concerns results obtained for the CIS ion plasma instrument developed for Cluster. CIS comprises an energy-per-charge instrument named HIA that measures ions without mass resolution in the energy/charge range from 5 V to 32 kV . The other, named CODIF, resolves the most important ion species and covers a range between $\approx 15 \text{ V}$ and 40 kV . Both have two sections, one with a low geometry factor (LGF), the other with a much larger geometry factor (HGF). This is to get adequate counting statistics over the large dynamic range of fluxes that we alluded to earlier. ⇒6.1

Figures 6.5, 6.6, and 6.7 show the errors in computed density, temperature, and velocity, respectively, as a function of the density and temperature of the input Maxwellians that were varied between 0.1 and 100 cm^{-3} , and between 10 and 10^4 eV , respectively. The four panels refer to different combinations of instruments, geometry factors and assumed bulk velocity. The figures demonstrate the large range of errors that can occur, from less than 5% to more than 100% , as a result of the finite energy range and resolution, as well as from poor counting statistics at low densities or from overflow at high densities.

The effect of the limited energy range of the measurements on the accuracy of the moments can be considered separately from other aspects. Figure 6.8 shows the calculated errors in density as a function of the bulk speed, v_0 (vertical scale) and the minimum speed v_{\min} , i.e., the speed corresponding to the low-energy limit of the measurements

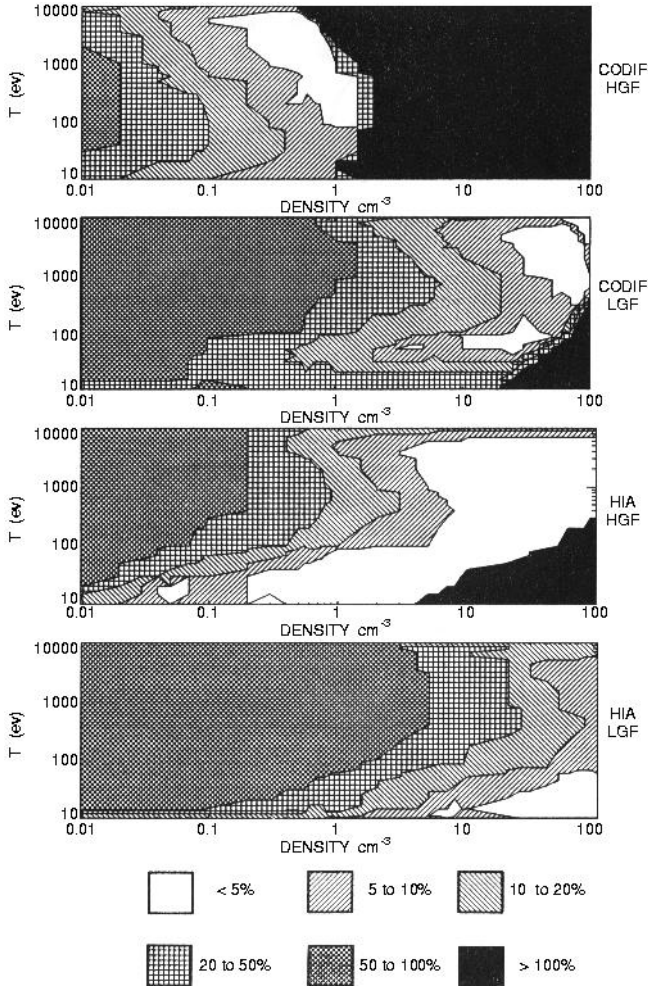


Figure 6.5: Errors in measured density for the HIA and CODIF ion instruments developed for Cluster, as a function of temperature (vertical scale) and density (horizontal scale), for assumed bulk speeds of (from top to bottom) 100, 300, 300, and 300 km/s, respectively.

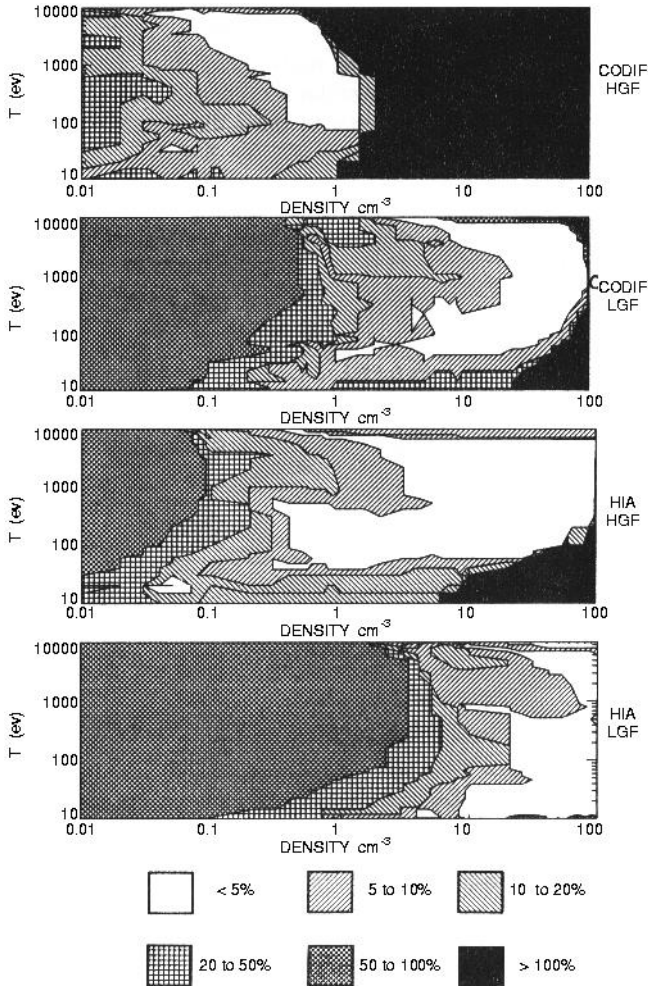


Figure 6.6: Same as Figure 6.5, but for the errors in temperature.

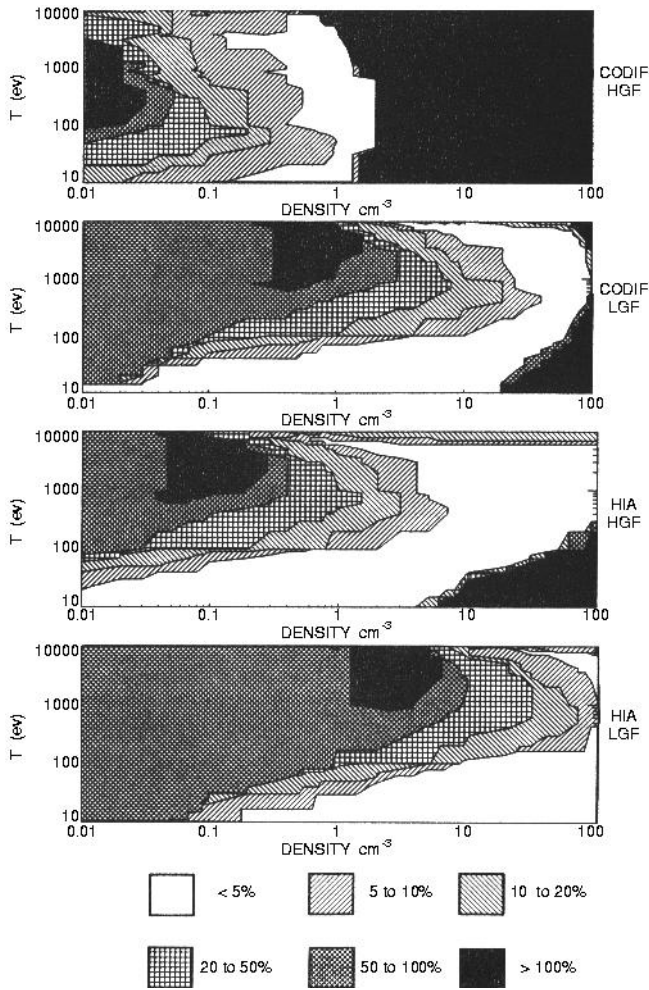


Figure 6.7: Same as Figure 6.5, but for the bulk velocity.

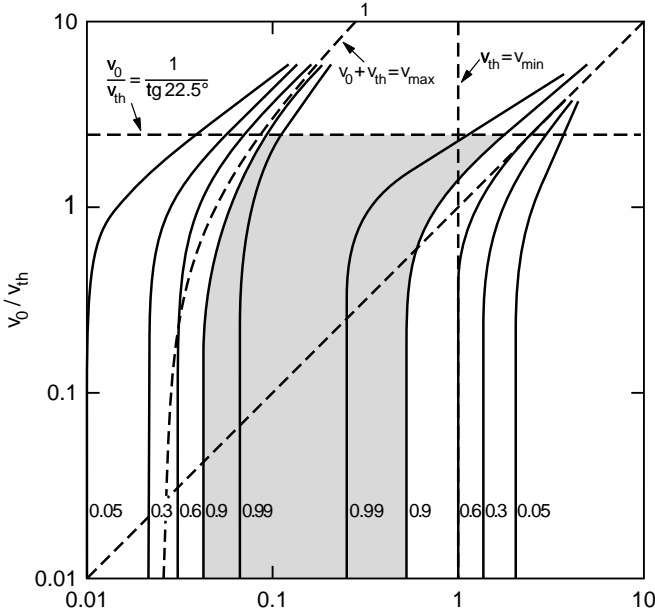


Figure 6.8: Errors in density as a function of the bulk speed (vertical axis) and the minimum speed detected by the instrument (horizontal axis), both normalised to the thermal speed. The numbers alongside the curves identify the fraction of the input density that the computed moments reproduce. The shaded region refers to fractions larger than 0.9, i.e., errors less than 10%. The computation assumed a ratio of 1500 between maximum and minimum energies covered by the instrument.

(horizontal axis), both normalised to the thermal speed of the plasma, v_{th} , assumed to have isotropic temperature. The figure, which was used when designing the AMPTE/IRM electron and ion instruments, demonstrates that the error increases sharply once v_{min}/v_{th} becomes larger than one. As discussed earlier, this effect is particularly significant for electrons. A low-energy limit of 10 eV is close to the typical thermal energy of electrons, i.e., $v_{min}/v_{th} \approx 1$, while for ions we typically have $v_{min}/v_{th} \ll 1$. The figure also shows the error that arises once the distribution extends above the upper limit, v_{max} of the measurements in case the thermal speed and/or the bulk speed become large. The calculations assumed a dynamic range in detected particle energies, E_{max}/E_{min} , of 1500, which corresponds to a ratio v_{max}/v_{min} of ≈ 38 .

Figure 6.9 shows in detail the effects of just the low energy limit on the accuracy of computed density, bulk speed, pressure, and temperature. The increase in error when the low energy limit, E_{min} , gets close to or exceeds the thermal speed of the distribution is clearly evident from these plots.

Distributions are often characterised by different temperatures in the directions parallel and perpendicular to the magnetic field. Such effects can be simulated by assuming a bi-Maxwellian as the input distribution. Figures 6.10 and 6.11 show the errors in density and T_{\parallel} , respectively, as a function of T_{\parallel} (vertical axis) and T_{\parallel}/T_{\perp} (horizontal axis), for the CESR electron instrument on the Interball mission. Figure 6.10 shows that the density is

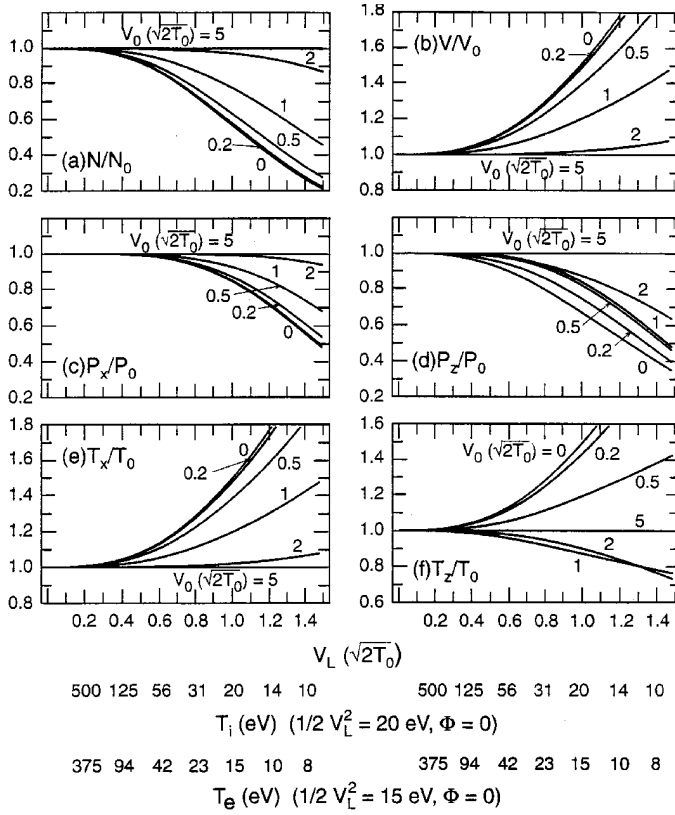


Figure 6.9: Computed moments, normalised to the true moments, as a function of the minimum particle speed detected by the instrument, V_L , with the different curves indicating different assumed bulk speeds, V_0 , directed along the z axis. All speeds are normalised to the thermal speed of the distribution. The two additional rows of numbers along the bottom indicate the ion and electron temperatures, respectively, that correspond to the normalised V_L if the lower energy limits are chosen as 20 eV and 15 eV, respectively. The maximum detected particle speed was taken as ∞ .

measured accurately as long as the thermal energy is well inside the energy range (10 eV to 25 keV) of the measurements, as expected. The density accuracy does not depend much on the temperature anisotropy, in contrast to the accuracy of the temperature (Figure 6.11) which depends strongly on the temperature anisotropy. Here the limited angular resolution drives the errors.

In the simulations presented so far, moment summation assumed that $f(\mathbf{v})$ is constant over the entire integration volume. But the simulations performed at MSSL have also tried a trapezoidal method for the summation that allows for linear variations in $f(\mathbf{v})$ between samples. But for the cases where the two methods have been compared the differences were negligible.

The more sophisticated integration scheme used for the Hydra measurements and described in Section 6.3 has also been checked with simulated data. For the published case

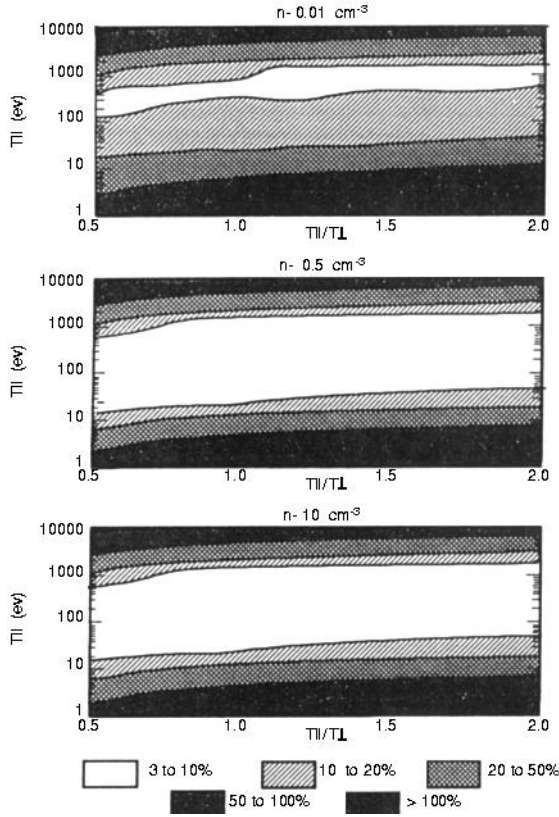


Figure 6.10: Error in density as a function of the parallel temperature T_{\parallel} (vertical scale) and the temperature anisotropy T_{\parallel}/T_{\perp} (horizontal scale) for three assumed input densities.

(a κ -distribution with $N=1 \text{ cm}^{-3}$ and $V=100 \text{ km/s}$), the agreement was found to be better than 1% for all moments up to and including the heat flux.

6.4.3 Overall Error Assessment

At the top of the list of error sources are inadequate range and resolution of the measurements. This effect is well-studied and quantified for model distributions by the computer simulations reported in the previous section. When the energy range properly matches the input distribution, errors are as small as a few %. When the input distributions get closer to the limits of instrument coverage or resolution, errors quickly become large. Because of the different velocity weighting, the effect will depend on the order of the moment. In actual data, such situations can often be recognised by inspection of the distribution functions when the latter are available. But even then the resulting error cannot be quantified, except when the distribution can be described by some model distribution (Maxwellian or kappa distribution) that is then fitted to the measured distribution, in which case the contributions to the moments from outside the measured energy range can be es-

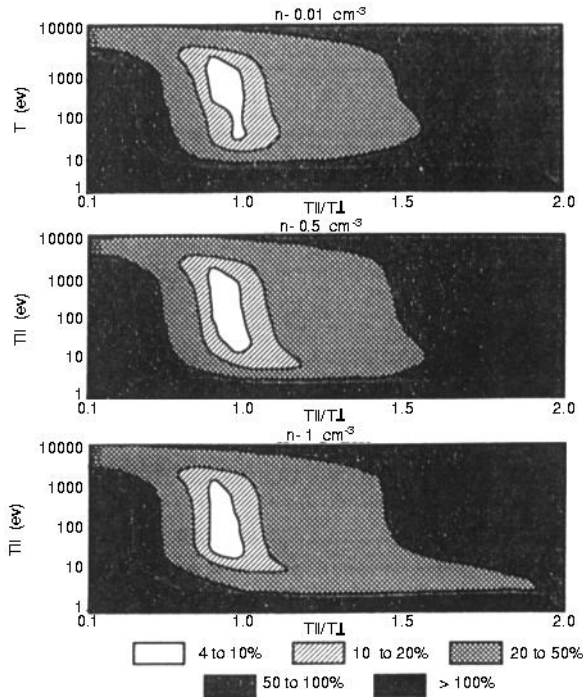


Figure 6.11: Same as Figure 6.10, but for the error in parallel temperature, T_{\parallel} .

timated accordingly. Needless to say, a population that falls entirely outside the energy range, or in the gaps between channels, cannot even be recognised in the distributions, let alone corrected for in the moments.

Uncertainties in absolute calibration are not considered in the simulations and must be added for those quantities that are not themselves ratios of moments, such as density and pressure. These uncertainties are estimated at no better than 10%. For bulk velocity and temperature such common factors drop out. When comparing instruments of identical design on different spacecraft, relative accuracies of a few % or better are achievable, at least when in-flight cross-calibrations are included.

The simulations do not deal with errors due to time variations. Such events can sometimes be recognised in the data by the occurrence of a single set of moments that differ strongly from those at adjacent times. Temporal variations in the magnetic field direction produce spurious anisotropies and off-diagonal terms in the pressure and temperature tensors.

Counting statistics are a major error source because the enormous dynamic range of fluxes invariably leads to insufficient count rates in the low-density regions that are encountered. The simulations have included the statistical errors by applying Poisson statistics to the received counts. For the actual measurements, the statistical errors for every moment can be computed if the moments are computed on the ground. Because of telemetry constraints one usually has the distributions available on the ground much less frequently than the moments, or with reduced resolution, and can thus compute the statistical errors

less frequently or less accurately. But one can always get a good measure of the statistical errors by transmitting the accumulated total counts in a distribution, together with the moments.

Overall, one can say that the uncertainties of the moments considered in the simulations (i.e., up to the second order) can be as low as a few % if the underlying distributions are well matched to energy range and resolution as well as to the geometry factor of the instrument, do not vary during one spin, and if the instrument is well calibrated.

Errors in the quantities not treated in the simulations (off-diagonal terms in the pressure or heat flux) are more difficult to judge. From AMPTE/IRM observations it appears that the “noise” in the off-diagonal terms for ions and electrons seems to range up to 10% of the diagonal terms, close to maximum magnitudes for these terms predicted by plasma simulations.

6.5 Interpretation

First we will discuss the meaning of the various moments when these are computed separately for each particle species, but we will also point out the special effects that arise if moments from a number of species are combined into one-fluid quantities.

6.5.1 Single-Species Moments

Number and Mass Densities

The meaning of number density and mass density is obvious from the definitions. Number densities in solar-terrestrial space plasmas range from $<0.1 \text{ cm}^{-3}$ in the magnetotail, to $\approx 100 \text{ cm}^{-3}$ in extreme conditions in the subsolar magnetosheath.

Bulk Velocity Vector

The bulk velocity \mathbf{V} is the average velocity of the particles and represents the motion of the plasma as a whole. Bulk velocities in space plasmas range from essentially zero to several thousand km/s.

Plasma bulk velocities are often better expressed in terms of dimensionless numbers, by scaling them to some characteristic velocity, such as the thermal speed or some wave speed. The solar wind ions, for example flow at bulk speeds much in excess of either their thermal speed or any wave speed the plasma can support. They thus represent a “supersonic” flow. The solar wind electrons, by contrast, although having the same bulk velocity as the ions and similar temperatures, are subsonic, because of their much higher thermal (and thus sound) speed.

Measured bulk velocities are often decomposed into the components perpendicular and parallel to \mathbf{B} to distinguish motion that does transport magnetic flux (“convection”) from motion that does not (though as noted below, there are some perpendicular drift motions that do not transport flux).

Momentum Flux and Pressure Tensors

The pressure tensor measures the momentum flux due to random particle motion, while the momentum flux tensor includes the momentum flux, $\rho \mathbf{V} \mathbf{V}$, associated with the motion of the plasma as a whole. The $\rho \mathbf{V} \mathbf{V}$ term is often referred to in the plasma literature as the dynamic pressure. It should be noted, however, that in ordinary fluid dynamics the term dynamic pressure is reserved for $(1/2) \rho \mathbf{V} \mathbf{V}$, a term which appears in the Bernoulli equation.

If $f(\mathbf{v})$ is spherically symmetric in the bulk-velocity frame, the pressure is isotropic and the pressure tensor will be diagonal, with the diagonal elements (P_{xx} , P_{yy} , P_{zz}) all equal to p , the scalar pressure. Scalar pressures typically range between 0.1 and 10 nPa.

If $f(\mathbf{v})$ is not isotropic but cylindrically symmetric (in the bulk-velocity frame) about the magnetic field direction, the pressure is different parallel and perpendicular to \mathbf{B} . In a coordinate system where the z axis is aligned with \mathbf{B} , the pressure tensor will still be diagonal, but with $P_{xx} = P_{yy} = p_{\perp}$ and $P_{zz} = p_{\parallel}$. \mathbf{P} still shows isotropy in planes perpendicular to \mathbf{B} and one can still define a scalar pressure via $p \equiv \text{Tr}(\mathbf{P})/3 = (2p_{\perp} + p_{\parallel})/3$. Distribution functions with such symmetry perpendicular to \mathbf{B} are referred to as gyrotropic.

In a coordinate system not aligned with \mathbf{B} , the pressure tensor for a distribution that is cylindrically symmetric in the bulk-velocity frame, becomes:

$$\mathbf{P} = p_{\perp} \mathbf{I} + (p_{\parallel} - p_{\perp}) \frac{\mathbf{B} \mathbf{B}^T}{B^2} \quad (6.17)$$

where \mathbf{I} is the unit tensor. As the trace of a tensor is invariant under coordinate transformations, the scalar pressure p stays the same. But the individual diagonal terms have changed, and the off-diagonal terms are no longer zero.

To separate the contributions of the non-spherically symmetric part of $f(\mathbf{v})$ from the symmetric part, the pressure tensor is sometimes split into two parts, one traceless, the other diagonal: $\mathbf{P} = \pi + p \mathbf{I}$. The part π not only contains the off-diagonal terms, but along its diagonal also the differences between the diagonal terms of \mathbf{P} and the scalar pressure, p .

The pressure tensor can always be diagonalised through a major-axes transformation, regardless of inherent symmetries, with diagonal elements P_{11} , P_{22} , P_{33} and zero off-diagonal terms. Unless $f(\mathbf{v})$ is spherically symmetric, or cylindrically symmetric with respect to \mathbf{B} , the diagonal elements will all be different, and none of the major axes will be aligned precisely with the magnetic field. These angle-offsets are an alternate way of quantifying the properties of the distribution associated with off-diagonal terms in the pressure.

Off-diagonal Terms

As illustrated by equation 6.17, non-zero off-diagonal terms in the pressure tensor already arise if in the bulk-velocity frame $f(\mathbf{v})$ is cylindrically symmetric around \mathbf{B} , but the chosen coordinate system is not aligned with \mathbf{B} . In this case the off-diagonal terms can be regarded as an artifact of the coordinate system and have no physical significance *per se*.

If non-zero off-diagonal terms occur in a field-aligned coordinate system, however, they indicate the presence of shear stresses in the plasma. P_{xy} , for example, measures the

flow of x-momentum by motion in the y-direction. The existence of such shear stresses requires a shear in bulk velocity, in our example $\partial V_x/\partial y \neq 0$. As \mathbf{P} is symmetric, $P_{xy} = P_{yx}$, which means there also must be a dependence on $\partial V_y/\partial x$. P_{xy} and P_{yx} must therefore each be proportional to $(\partial V_x/\partial y + \partial V_y/\partial x)$.

In ordinary fluids the off-diagonal terms are associated with viscosity, and the constants that relate the off-diagonal terms with the velocity derivatives are termed viscosity coefficients. As a result of collisions, particles carry momentum into other parts of the fluid, the larger the mean free path, the farther the momentum is carried. This tends to equalise the velocities and thus reduces the velocity shear. In a collisionless plasma a similar effect is introduced by the gyration of the particles around the magnetic field, with the gyroradius replacing the mean free path. This effect is often referred to as gyroviscosity.

Off-diagonal terms manifest themselves in the underlying $f(\mathbf{v})$ by the lack of cylindrical symmetry around \mathbf{B} in the bulk-velocity frame, i.e., by a deviation from gyrotropy. The reverse is, however, not true. One can construct (admittedly exotic) distribution functions that are not gyrotropic, but for which the off-diagonal elements in the pressure tensor are zero. Examples are distributions with $f \propto \sin n\phi$ with ϕ the gyrophase angle and $n > 2$.

Spurious off-diagonal terms (and anisotropies in general) in the computed pressure and temperature tensors result when the distribution function extends significantly beyond the energy range of the instrument, or when the magnetic field direction fluctuates while the distribution function is being measured. The latter effect can be illustrated with a case where the off-diagonal terms are assumed to be zero in a frame aligned with the instantaneous magnetic field. In the frame aligned with the magnetic field averaged over one spin, the tensor will have spurious off-diagonal terms that arise from the fluctuations in B_x, B_y (of order ϵB_z) around the average direction which is taken as z . According to equation 6.17, in a frame not aligned with \mathbf{B} the terms P_{xy}, P_{xz}, P_{yz} are proportional to the products $B_x B_y, B_x B_z, B_y B_z$, respectively. One thus obtains spurious components P_{xz}, P_{yz} that are of order ϵ , while P_{xy} is much smaller, of order ϵ^2 .

With regard to the expected magnitude, plasma simulations of current sheets typically show off-diagonal terms of no more than 10% or 20% of the diagonal terms, close to the detection limit of present measurements.

Temperature Tensor

The temperature tensor is obtained from the pressure tensor by noting that $\mathbf{T} \equiv \mathbf{P}/(Nk)$. Similarly, one can define a scalar temperature via $T \equiv \text{Tr}(\mathbf{T}/3) = p/(Nk)$. The temperature tensor has the same properties as were noted above for the pressure tensor. Temperatures in space plasmas range from 10^4 K to more than 10^8 K, and thermal energies from 1 eV to more than 10 keV.

The temperature determined this way is often referred to as kinetic temperature because it is not necessarily a temperature in the thermodynamic sense. The latter implies that the plasma is in thermal equilibrium, i.e., has a Maxwellian velocity distribution. The concept of a kinetic temperature can be illustrated with the two distributions compared in Figure 6.12, a single hot Maxwellian and two cold counter-streaming beams, respectively. They have the same parallel kinetic temperatures (and same density) although only one is a thermal distribution in the usual sense, meaning that a fixed fraction of particles is contained within a thermal velocity around the peak of $f(\mathbf{v})$. By contrast, in the case of the counter-streaming cold beams there are no particles close to the bulk velocity at zero.

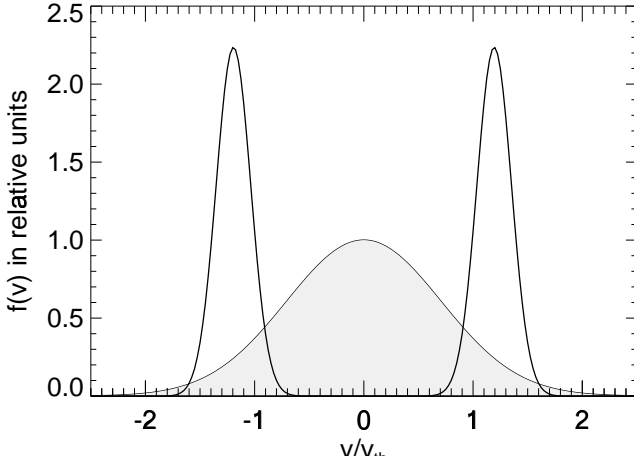


Figure 6.12: Comparison of two distribution functions, a hot Maxwellian and two counterstreaming cold beams, that have the same kinetic temperature and the same density.

Nevertheless, the kinetic temperature can be a useful quantity as a measure of the (random) velocity spread of the particles. Take the case of the bow shock. Immediately downstream of the quasi-perpendicular shock the distribution consists of the slowed down cold solar wind beam, surrounded by gyrating beams of ions that were first reflected at the shock, but eventually transmitted after gyrating back to the shock, having gained energy in the upstream electric field. The kinetic temperature computed for this distribution, which looks anything but “thermal” and is not even gyrotropic, is still an indication of what the temperature will be, once scattering in angle and energy has finally “thermalised” the distribution.

Heat Flux Vector

The heat flux vector measures the net flux of kinetic energy of random motion. It is commonly decomposed into H_{\parallel} , the component parallel to \mathbf{B} , and H_{\perp} , the component perpendicular to \mathbf{B} . H_{\parallel} is expected to be the dominant component as the magnetic field impedes heat flow in the perpendicular direction. Typical values of H_{\parallel} in space plasmas range up to 0.1 and 1 mW/m² for electrons and ions, respectively.

Heat flux can be a signature of fast particles escaping along the magnetic field from some source region. Examples are the electron heat flux in the solar wind that is a signature of magnetic connection to the hot corona. Energetic magnetospheric particles escaping along reconnected field lines across the magnetopause also constitute a heat flux. But heat flux is also taken as evidence for connection with a dissipation region. An example are the electron heat flux pulses observed when a spacecraft crosses the field lines that presumably connect to the so-called “diffusion region” where reconnection between interplanetary and terrestrial field lines occurs.

Even though strictly defined as an integral over the entire distribution, the heat flux is sometimes computed separately in the directions parallel and anti-parallel to the magnetic field. An example is the solar wind where the above-mentioned heat flux from escaping coronal electrons is partially reflected if the magnetic field line is connected to the Earth's bow shock. This "reverse" heat flux can be used as an indicator for magnetic connection to the bow shock.

6.5.2 One-Fluid Moments

From the definitions in the second part of Section 6.2, the interpretation of the one-fluid mass and number densities is straightforward. But with regard to the overall bulk velocity, it should be noted that it refers to the mass-weighted average of the velocities, and \mathbf{V} is thus often referred to as the centre-of-mass (CM) velocity. Use of a single bulk-velocity frame for all species introduces important changes for the pressure and temperature tensors which we will discuss below. In addition, there are the new quantities, charge density and current, if electrons are one of the species.

Pressure and Temperature Tensors

If one constructs single-fluid moments from the moments of several species, e.g., electrons and ions, the pressures and temperatures now refer to the overall centre-of-mass (CM) frame. If the CM frame is different from the bulk-velocity frames of the individual species, distribution functions that were symmetric in their own bulk-velocity frame, and thus had isotropic temperature, are no longer symmetric in the CM frame, with the result that pressure and temperature anisotropies are introduced. This happens, for example, if the species have the same V_\perp but different V_\parallel . Furthermore, individual distribution functions need no longer be gyrotropic in the CM frame, even though they might have been gyrotropic in their own bulk velocity frames. The one-fluid pressure and temperature tensors will then have acquired off-diagonal terms even though these were zero in the bulk-velocity frames of the individual species.

Charge Density

As long as one is dealing with low-frequency motions where electron inertia is not a factor, plasmas will be quasi-neutral in which case the charge density is zero.

Current Density

Assuming quasi-neutrality (i.e., $N_e = N_i = N$) and singly-charged ions, the current density is defined as $\mathbf{j} = eN(\mathbf{V}_i - \mathbf{V}_e)$. Thus \mathbf{j} can, in principle, be directly determined from the electron and ion moments. But this task can challenge the accuracy of the bulk velocity measurements. Take the case of the Earth's magnetopause: for a change in B of 30 nT over a thickness of 200 km, i.e., a current density j of $0.1 \mu\text{A}/\text{m}^2$, and a number density N of 30 cm^{-3} , the difference velocity is only 25 km/s. At the bow shock the plasma density is lower and the thickness of the current layer smaller, and the difference velocity is therefore much larger, e.g., 150 km/s for a jump in B of 15 nT, a thickness of 50 km, and a density of 10 cm^{-3} .

6.6 Distribution Functions, Moments, Partial Moments

From a representation of a distribution function in terms of contours in phase-space density, one can usually see immediately whether the particle population is moving with some bulk speed, has an isotropic or anisotropic temperature, or has some heat flux. Bulk motion is recognised by a shift of the contours' centres relative to the origin. Isotropic temperature means circular contours, at least in the vicinity of the peak, and temperature anisotropy is recognised by elongated contours. Superimposing the magnetic field direction clarifies whether the distribution is symmetric around \mathbf{B} and whether T_{\parallel}/T_{\perp} is larger or smaller than one. Heat flux is recognised by a "skewing" of the distribution at higher energies. A projection of the distribution onto a plane at right angles to \mathbf{B} shows whether the distribution is gyrotropic or not. In this sense the moments are a compact way to describe main features of the distribution function, provided $f(\mathbf{v})$ is not too complicated.

As the moments (up to the heat flux) are just 13 numbers per species, it is not surprising that there is much more information in the distribution functions that typically consist of several thousand samples in velocity space. It is therefore quite apparent that to fully utilise the information gathered by plasma instruments, the distribution functions should be transmitted with the highest possible time resolution.

The example of the bow shock ions noted earlier is a case in point. The superposition of the reflected ions appears in the moments as a reduction in bulk speed and a temperature increase, but only the inspection of $f(\mathbf{v})$ reveals that ions are being reflected, accelerated, returned to and transmitted across the shock. When concerned with wave-particle interactions, the moments are of only limited use as well. This is because it is the detailed structure of the distribution functions that indicates instability to the growth of waves.

Another good example occurs in the plasma sheet boundary layer (PSBL) in the Earth's magnetotail. At its outer edge there often is a single beam that moves earthward with large bulk speed. In the converging magnetic field closer to Earth the beam particles are mirrored, but will return deeper in the PSBL because of the inward convection induced by the prevailing dawn-dusk electric field. When superimposed on the original earthward beam, the return beam will partially compensate the earthward beam, causing a reduction in the average V of the distribution. Once earthward and return beams have attained almost equal densities, V will become almost zero, even though each beam may be moving at several 1000 km/s. The large velocity separation of the two beams along \mathbf{B} will simply manifest itself in the moments as a large parallel temperature. From the moments alone one would not be able to distinguish this situation from one with a single bi-Maxwellian with small or zero bulk velocity (see Figure 6.12). Only inspection of the distribution function will clarify what is happening.

Data compression is only one aspect of the utility of moments. More important is their role as the quantities entering the plasma fluid equations, in particular those concerned with conservation of mass, momentum, and energy. It should be stressed here that in these latter applications, the exact nature of the distribution function underlying the moments is irrelevant. When computing the pressure, in order to check pressure balance across some boundary for example, it is immaterial whether the distribution function is a simple Maxwellian or is made up of multiple beams moving at different speeds in different directions. Similarly, for tangential stress balance across a discontinuity that has both a magnetic field and a plasma mass flow across it, one must take the moments computed over the entire distribution function, even though the distribution on each side may be

made up of the particle population incident on the discontinuity, a portion that has been reflected at the discontinuity, as well as populations that were transmitted from the other side of the discontinuity. All these populations carry momentum towards or away from the discontinuity and therefore must be included in the balance.

The moments to be used in the conservation equations considered below require integration over the entire range of velocities. But there are questions that benefit from moments computed over just a portion of the distribution function (“partial moments”). Take the example of the region upstream of the Earth’s bow shock (referred to as the foreshock) that, in addition to the solar wind, contains a population of energetic ions accelerated at the shock. When moments over the entire distribution are computed, one notices that the bulk speed is reduced. Is this the result of the solar wind ions being slowed down, or is this an effect of the addition of the more energetic ions that have a small average velocity? In the distribution functions these energetic ions are usually well isolated from the solar wind ions. One can therefore answer that question by computing the bulk velocity of just the solar wind ions alone, and compare it with the wind velocity in the region outside the foreshock where the energetic particles are absent.

6.7 Spherical Harmonic Analysis

As pointed out in Section 6.4.1, if major portions of the plasma distribution function fall outside the instrument measurement range, moment analysis will not produce meaningful results. It is for this reason that the procedure is rarely applied to data sets from energetic particle experiments, where the lowest energy threshold might be several tens of keV. For example, the bulk velocity cannot be well estimated by the averaged measured velocity if the majority of the plasma particles are below the lowest threshold.

In such a situation, classical *spherical harmonic analysis* becomes useful. In this scheme, the measured particle fluxes $j(E, \theta, \phi)$ are fit to a series of spherical harmonic functions, with coefficients that contain all the angular information at a given particle energy. The coefficients are determined by means of a linear least squares fit of the measured data points.

Of particular significance for our purpose are the first-order coefficients that determine a vector, referred to as first-order anisotropy. This anisotropy vector has both magnitude and direction; it arises from particle density gradients and/or bulk flow.

Unlike moment analysis, spherical harmonic analysis is strictly a two-dimensional procedure over the surface of a sphere that must be carried out for each energy separately. For a detailed description of the spherical harmonic analysis procedure, see the bibliography.

6.8 Applications

6.8.1 Identification of Regions and Boundaries

The plasma moments, such as density, bulk speed, and temperature, are the principal means to distinguish the various plasma regimes, to describe their characteristic properties, and to identify crossings of boundaries between these regimes.

6.8.2 Conservation Equations

The plasma velocity moments appear in all the macroscopic equations that govern the behaviour of the plasma as a fluid, such as the continuity equations for mass, momentum and energy, or the generalised Ohm's law. These equations apply for time variations with characteristic frequencies smaller than the ion gyrofrequency and length scales larger than the ion gyroradius of the heaviest ion species present. Simplified equations, commonly referred to as MHD equations, result if further assumptions are made, as pointed out below. It is a widespread misconception that use of plasma moments implies validity of the assumptions underlying MHD.

The conservation equations contain spatial and temporal derivatives of scalar, vector and tensor quantities that require multipoint measurements of plasma moments, magnetic and electric fields. The equations depend on knowledge of the mass density and thus stress the importance of mass-resolved ion measurements.

We will consider only the two lowest-order equations, those associated with conservation of mass and momentum. Each equation introduces a new quantity whose evolution is described by the next higher equation. Thus the system is not closed, i.e., there are more unknowns than equations. The energy equation and the heat flux equation would be the next two levels. Without closure of this progression the equations are incomplete. Closure is obtained either by truncation (e.g., by setting the heat flux to zero), or by assuming some equation of state. The latter approach is easily recognised by a γ (the ratio of specific heats) appearing in the equations, but the value of γ is usually not known for a collisionless plasma.

Conservation of Mass

$$\frac{\partial \rho}{\partial t} = -\nabla \cdot (\rho \mathbf{V}) \quad (6.18)$$

The equation demands a balance between the time-derivative of ρ and the divergence of the mass-flux density vector. The equation is valid for each particle species separately, but also for the entire fluid, with ρ as defined in Section 6.2.2. The equation can be rewritten as:

$$\frac{d\rho}{dt} = -\rho \nabla \cdot \mathbf{V} \quad (6.19)$$

where $d/dt \equiv (\partial/\partial t + \mathbf{V} \cdot \nabla)$ is the convective or co-moving derivative. Both $d\rho/dt$ and $\nabla \cdot \mathbf{V}$ must be determined from multipoint measurements, as described in Chapter 17. If one finds $\nabla \cdot \mathbf{V} \approx 0$, one is dealing with incompressible flow, $d\rho/dt = 0$, and one can compare observations with results from theory or simulations that assume incompressibility and are therefore easier to do.

Conservation of Momentum

Each particle species, s , obeys an equation of motion:

$$\rho_s \frac{d\mathbf{V}_s}{dt} = -\nabla \cdot \mathbf{P}_s + \frac{q_s}{m_s} \rho_s (\mathbf{E} + \mathbf{V}_s \times \mathbf{B}) \quad (6.20)$$

If the plasma consists of electrons and one ion species, one can construct the equation of motion for the total fluid by adding the two corresponding equations of motion and using the single-fluid quantities defined in Section 6.2.2. Assuming zero space charge, one gets:

$$\rho \frac{d\mathbf{V}}{dt} = -\nabla \cdot \mathbf{P} + \mathbf{j} \times \mathbf{B} \quad (6.21)$$

Inserting $\mathbf{j} = \frac{1}{\mu_0} \nabla \times \mathbf{B}$, the $\mathbf{j} \times \mathbf{B}$ term can be expressed in terms of the gradient in magnetic pressure and the tension from magnetic field curvature:

$$\rho \frac{d\mathbf{V}}{dt} = -\nabla \cdot \mathbf{P} - \nabla \frac{B^2}{2\mu_0} + \frac{(\mathbf{B} \cdot \nabla) \mathbf{B}}{\mu_0} \quad (6.22)$$

If isotropic pressure is assumed and the $\mathbf{V} \cdot \nabla$ in the convective derivative is ignored, equation 6.21 reduces to the form used in MHD:

$$\rho \frac{\partial \mathbf{V}}{\partial t} = -\nabla \cdot \mathbf{p} + \mathbf{j} \times \mathbf{B} \quad (6.23)$$

Generalised Ohm's Law

Subtracting the equations of motion for electrons and ions, neglecting terms that are proportional to m_e/m_i , and assuming quasi-neutrality, one obtains:

$$\mathbf{E} + \mathbf{V} \times \mathbf{B} = \eta \mathbf{j} + \frac{1}{Ne} \mathbf{j} \times \mathbf{B} - \frac{1}{Ne} \nabla \cdot \mathbf{P}_e + \frac{m_e}{Ne^2} \frac{\partial \mathbf{j}}{\partial t} \quad (6.24)$$

In an ideal magnetohydrodynamic (MHD) fluid, all terms on the right are set to zero. In this case $\mathbf{E} = -\mathbf{V} \times \mathbf{B}$. In resistive MHD, only the $\eta \mathbf{j}$ term is retained. However, neglecting the second term on the right (the so-called Hall term) is not justified if significant currents are flowing, such as within plasma boundaries. Noting that $\mathbf{V} \approx \mathbf{V}_i$, and therefore $\mathbf{V} - \mathbf{V}_e \approx \mathbf{j}/Ne$, one can remove the Hall term in equation 6.24 by using \mathbf{V}_e instead of \mathbf{V} . Neglecting again the other terms on the right one gets $\mathbf{E} = -\mathbf{V}_e \times \mathbf{B}$. This shows that the magnetic field is more closely coupled to the electron bulk velocity rather than the ion bulk velocity. Therefore one should use the electron bulk velocity, rather than the ion velocity, if one is concerned with the motion of the magnetic field lines.

Plasma Drifts and Currents

If one neglects the convective derivative term in equation 6.20 and takes the cross product with \mathbf{B}/B^2 , one obtains the equation for the drift velocity of a species of the fluid under time-stationary conditions:

$$\mathbf{V}_s = \frac{\mathbf{E} \times \mathbf{B}}{B^2} + \frac{1}{q_s N_s B^2} (\mathbf{B} \times \nabla \cdot \mathbf{P}_s) \quad (6.25)$$

or if the pressure tensor can assumed to be given by equation 6.17,

$$\mathbf{V}_s = \frac{\mathbf{E} \times \mathbf{B}}{B^2} + \frac{1}{q_s N_s B^2} \mathbf{B} \times \nabla p_{s\perp} + \frac{1}{q_s N_s B^2} \mathbf{B} \times \nabla \cdot \left[(p_{s\parallel} - p_{s\perp}) \frac{\mathbf{B} \mathbf{B}^T}{B^2} \right] \quad (6.26)$$

Thus in addition to the $\mathbf{E} \times \mathbf{B}$ drift the fluid shows two other drifts: one caused by ∇p_\perp , another that depends on pressure anisotropy and field-line curvature. The pressure gradient drift is determined entirely by gyration of the particles without any net transport. Because of ∇p_\perp , there are more particles (or particles with higher energy) gyrating through a surface in the ∇p_\perp - \mathbf{B} plane from one side than from the other. This flux-imbalance mimics a flow, even though the particle gyrocentres do not move.

By comparing the perpendicular component of the bulk velocities, measured independently for electrons and various ion species, with the $\mathbf{E} \times \mathbf{B}$ velocity obtained from the electric and magnetic field measurements, one can obtain a measure of the importance of the extra terms in above equation.

Because the pressure gradient drift depends on the sign of the charge, the oppositely directed drift of ions and electrons gives rise to a current that is called diamagnetic because its magnetic field opposes the external field:

$$\mathbf{j}_\perp = \frac{\mathbf{B} \times \nabla p_\perp}{B^2} \quad (6.27)$$

where $p_\perp = p_{e\perp} + p_{i\perp}$.

Determination of ∇p_\perp from multipoint plasma measurements thus allows us to deduce the perpendicular current density, \mathbf{j}_\perp .

The full current vector, including \mathbf{j}_\parallel , can be estimated from the multipoint measurements of the magnetic field, via $\mathbf{j} = (1/\mu_0)\nabla \times \mathbf{B}$, or may be directly obtained from the measured ion and electron bulk velocities and densities, as discussed earlier on page 150 under current density..

Divergence of the Pressure Tensor

Several of the previous equations involve the divergence of the pressure tensor, \mathbf{P} . The equations are often simplified by replacing \mathbf{P} either by the expression given in equation 6.17, as in equation 6.26, or by a scalar pressure, p . In the latter case $\nabla \cdot \mathbf{P}$ becomes simply ∇p .

It is immediately evident that the assumptions on \mathbf{P} strongly affect the magnitude of the closure problem, because keeping the full pressure tensor requires six evolution equations, while one such equation is sufficient if only the scalar pressure p is kept.

In ordinary fluid dynamics one has

$$\nabla \cdot \mathbf{P} = \nabla p + \mu \nabla^2 \mathbf{V} \quad (6.28)$$

where μ is the (dynamic) viscosity. This ansatz is yet another way to simplify the equations, with some form of gyroviscosity replacing the viscosity of ordinary fluids.

By ignoring the off-diagonal terms altogether, one sometimes eliminates the most important terms, as illustrated by the following example. Consider a magnetic neutral sheet in the xy plane separating magnetic fields in the $+x$ and $-x$ directions, with an electric field in the y -direction. To see how the electric field force along y could be balanced, one needs the y -component of $\nabla \cdot \mathbf{P}$:

$$(\nabla \cdot \mathbf{P})_y = \frac{\partial P_{xy}}{\partial x} + \frac{\partial P_{yy}}{\partial y} + \frac{\partial P_{yz}}{\partial z}$$

If one can assume that the situation is two-dimensional, and thus well described by taking $\partial/\partial y = 0$, it follows that $\partial P_{yy}/\partial y = 0$, and thus only the off-diagonal terms matter in the force balance.

Vorticity

Another important quantity involving a gradient operation, and thus multipoint measurements, is the vorticity defined as $\boldsymbol{\omega} = \nabla \times \mathbf{V}$. If one takes the curl of equation 6.21 and inserts the expression for \mathbf{P} from equation 6.28, one can write the equation of motion in terms of the evolution of $\boldsymbol{\omega}$:

$$\frac{d\boldsymbol{\omega}}{dt} - (\boldsymbol{\omega} \cdot \nabla) \mathbf{V} = (\mathbf{B} \cdot \nabla) \mathbf{j} - (\mathbf{j} \cdot \nabla) \mathbf{B} + \mu \nabla^2 \boldsymbol{\omega} \quad (6.29)$$

where we have assumed incompressible flow, $\nabla \cdot \mathbf{V} = 0$. This equation is similar to the equation of motion for ordinary viscous fluids and is thus often the starting point for discussing MHD turbulence in relation to Navier-Stokes turbulence in ordinary fluids.

Jump Relations

Of particular importance is the application of the conservation equations, together with Maxwell's equations, to the case of plasma discontinuities, such as shocks, tangential (TD) and rotational (RD) discontinuities. When these are assumed to be time-stationary surfaces, where only the variations in the direction normal to the surface are important (1-D case), the equations can be converted to jump-conditions across the discontinuity that are commonly called Rankine-Hugoniot relations. They take on different forms for the different discontinuities and provide thus the main tool to distinguish observationally between the discontinuity types.

A familiar example where the R.-H. relations have been successfully applied is the Earth's bow shock, as discussed in Chapter 10, Section 10.2. Another example is the Earth's magnetopause. Without magnetic reconnection, the magnetosphere is "closed", and the magnetopause is expected to be a TD. When magnetic reconnection occurs, the magnetopause becomes an RD and the magnetosphere is "open". For a stationary RD, combination of the conservation of tangential momentum and the conservation of the tangential electric field, assumed to be given by the ideal Ohm's law, $\mathbf{E} = -\mathbf{V} \times \mathbf{B}$, yields

$$\left[\mathbf{V}_t - \frac{\mathbf{B}_t}{\sqrt{\mu_0 \rho}} \right] = \text{const} \quad (6.30)$$

For an anisotropic plasma pressure, equation 6.30 is modified:

$$\left[\mathbf{V}_t - \frac{(1 - \alpha) \mathbf{B}_t}{\sqrt{\mu_0 \rho}} \right] = \text{const} \quad (6.31)$$

where α , the pressure-anisotropy factor, is defined as $\alpha \equiv (p_{\parallel} - p_{\perp})/(B^2/\mu_0)$. For a TD the jump in \mathbf{V}_t across the discontinuity is arbitrary and thus unrelated to the jump in \mathbf{B}_t . Equations (6.30) or (6.31) are therefore powerful tools to distinguish TD's from RD's. In a related form (the Walén relation) they are discussed in Section 9.3.3 of Chapter 9.

We note in passing that a corollary of equation 6.31 is a jump relation involving mass density and pressure anisotropy,

$$\rho(1 - \alpha) = \text{const} \quad (6.32)$$

that has often been applied implicitly, but never truly tested, because of the lack of mass-resolved measurements with adequate time resolution.

If the assumption of a time-stationary, one-dimensional discontinuity is valid, then momentum conservation along the normal direction yields a jump condition that for the case of a TD and RD, and assuming isotropic pressure, takes the form

$$p + \frac{B^2}{2\mu_0} = \text{const} \quad (6.33)$$

Testing for normal momentum balance is a prerequisite for application of other jump conditions, such as tangential momentum balance.

If we take the case of a plasma consisting of electrons and one ion species, then $p \approx p_e + p_i = N(T_i + T_e)$, if quasi-neutrality is assumed. This illustrates that the normal momentum balance only fixes the jump in p but leaves open how this jump is split between the individual ion and electron pressures or temperatures.

6.9 Summary

The moments of the plasma velocity distributions functions and their spatial gradients play a key role in the quantitative description of plasma behaviour, as exemplified by the conservation equations for mass, momentum and energy. In this chapter we have described how the moments are defined, what they mean, how they are computed, and how their accuracy is affected by limitations in the measurements and in the integration schemes. The quantitative assessment of the errors forms the basis for the accuracy of the spatial gradients obtained from multipoint measurements.

Bibliography

Papers that discuss the computation and accuracy of the moments are:

- Curtis, D. W., Carlson, C. W., Lin, R. P., Paschmann, G., Rème, H., and Cros, A., Onboard data analysis techniques for space plasma particle instruments, *Rev. Sci. Instr.*, **60**, 372, 1989. *Table 6.1 is taken from this paper.*
- Kessel, R. L., Johnstone, A. D., Coates, A. J., and Gowen, R. A., Space plasma measurements with ion instruments, *Rev. Sci. Instr.*, **60**, 3750, 1989. *Figures 6.2, 6.3, and 6.4 are taken from this paper.*
- Martz, C., Sauvaud, J. A., and Rème, H., Accuracy of ion distribution measurements and related parameters using the Cluster CIS experiment, in *Proc. International Conf. "Spatio-Temporal Analysis for Resolving plasma Turbulence (START)"*, Aussois, 31 Jan.–5 Feb. 1993, ESA WPP-047, pp. 229–239, European Space Agency, Paris, France, 1993. *Figures 6.5, 6.6, and 6.7 are taken from this paper.*

Sauvaud, J. A., Martz, C., and Aoustin, C., Error estimation for a typical electron spectrometer: Application to the Interball electron detector and generalisation, in *Proc. International Conf. "Spatio-Temporal Analysis for Resolving plasma Turbulence (START)"*, Aussois, 31 Jan.–5 Feb. 1993, ESA WPP-047, p. 301, European Space Agency, Paris, France, 1993. Figures 6.10 and 6.11 are taken from this paper.

Song, P., Zhang, X. X., and Paschmann, G., Uncertainties in plasma measurements: Effects of lower cutoff energy and spacecraft charge, *Planet. Space Sci.*, **45**, 255, 1997. Figure 6.9 is taken from this paper.

Scudder, J. et al., Hydra—a three-dimensional electron and ion hot plasma instrument for the Polar spacecraft of the GGS mission, *Space Sci. Rev.*, **71**, 459, 1995. The moments integration procedure used for Hydra is described in this paper.

Introductions to magnetohydrodynamics with emphasis on space plasmas can be found in:

Siscoe, G. L., Solar system magnetohydrodynamics, in *Solar-Terrestrial Physics*, edited by R. L. Carovillano and J. M. Forbes, p. 11, D. Reidel Publ. Co., Dordrecht, 1983.

Baumjohann, W. and Treumann, R. E., *Basic Space Plasma Physics*, Cambridge University Press, Cambridge, UK, 1996.

For treatments of spherical harmonic analysis with energetic particles, see:

Sanderson, T. R. and Hynds, R. J., Multiple telescope measurements of particle anisotropies in space, *Planet. Space Sci.*, **25**, 799–807, 1977.

Sanderson, T. R. and Page, D. E., Spherical harmonic analysis of satellite anisotropy measurements, *Nucl. Inst. & Meth.*, **119**, 177–182, 1974.

Fränz, M., *Ionen mittlerer Energie in der Heliosphäre: Ergebnisse des Ulysses-EPAC-Experiments in der Ekliptik*, Ph.D. thesis, Technische Universität zu Braunschweig, Braunschweig, Germany, 1994, available as Report #MPAE-W-78-94-19 from the Max-Planck-Institut für Aeronomie, D-37191 Katlenburg-Lindau, Germany.

— 7 —

Multi-Spacecraft Analysis of Plasma Kinetics

STEVEN J. SCHWARTZ

*Queen Mary and Westfield College
London, United Kingdom*

PATRICK W. DALY

*Max Planck Institut für Aeronomie
Katlenburg-Lindau, Germany*

ANDREW N. FAZAKERLEY

*Mullard Space Science Laboratory
Holmbury St. Mary, United Kingdom*

7.1 Introduction

Handling data from charged particle analysers which measure phase space density requires some care and attention. Even with a single satellite, the errors and uncertainties introduced by energy level and response calibrations, incomplete sampling of phase space, and discrete integration down to plasma moments (density, velocity, pressure, etc.) results in basic errors typically larger than those in, say, magnetic field measurements (see Chapter 6 for more details concerning the computation of parameters from particle instruments). Additionally, the time required to complete a sampling interval is often seconds or more, comparable to many scales of interest and intrinsic variability, so that some time aliasing is often present. Multi-spacecraft comparisons compound these difficulties. Thus many of the methods (filtering, spatial gradient and other vector operators) introduced earlier with electromagnetic fields as examples are much more difficult to apply to particle data.

On the other hand, particle data holds a richness in phase space information which can be exploited to reveal the physical processes which govern the dynamics, and which can probe/remote sense non-local structures. Nearly all such work is based on applications of Liouville's Theorem. In this chapter we explore some of the ways in which this phase space information can be utilised. One major difference between these multi-spacecraft particle techniques and those discussed earlier with respect to lower dimensional, higher time resolution field data is that very often the particle techniques need to be adapted and/or designed with a single specific study in mind. Thus the techniques described below should be regarded as examples rather than off-the-shelf techniques.

Additionally, multi-species measurements provide another dimension which can be utilised. At its most basic level, measurements of both ions and electrons enables a direct measurement of the charge and current densities. Comparison of ions with different masses (or charge-to-mass ratios) probes different scalelengths and differentiates the relative contributions of different forces.

This chapter is organised as follows. In the next section, we review Liouville's Theorem, and other background information. This is followed by a discussion of techniques relating to basic moments of the particle distribution. Later sections cover various applications of Liouville's Theorem and related phase space aspects.

7.2 Liouville's Theorem

The Boltzmann equation describes the evolution of the single particle phase space distribution function $f(\mathbf{r}, \mathbf{v}, t)$:

$$\frac{\partial f}{\partial t} + \frac{\partial}{\partial \mathbf{r}} \cdot (\mathbf{v} f) + \frac{\partial}{\partial \mathbf{v}} \cdot (\mathbf{a} f) = \left(\frac{\partial f}{\partial t} \right)_{\text{collisions}} \quad (7.1)$$

where $\mathbf{a} = \mathbf{F}/m$ is the acceleration of a non-relativistic particle (\mathbf{F} is the force) and the right-hand side includes the effects of collisions or other processes which give rise to instantaneous changes in particle velocity or position (e.g., due to creation by ionisation, etc.). The Boltzmann equation is a statement about particle conservation, and is most easily interpreted by considering the flow into and out of a fixed volume of (\mathbf{r}, \mathbf{v}) phase space. [For relativistic particles f is written as a function of momentum \mathbf{p} instead of velocity. The third term in equation 7.1 becomes $\partial/\partial \mathbf{p} \cdot (\mathbf{F} f)$.]

The variables \mathbf{r} and \mathbf{v} are independent, so the \mathbf{v} can be moved outside of the derivative in middle term on the left-hand side. Additionally, if the velocity divergence of the acceleration is zero, the third term can be similarly rearranged. This is the case for the Lorentz force $q(\mathbf{E} + \mathbf{v} \times \mathbf{B})$. If the right-hand side of the Boltzmann equation can be neglected, the result is

$$\frac{Df}{Dt} \equiv \frac{\partial f}{\partial t} + \mathbf{v} \cdot \frac{\partial f}{\partial \mathbf{r}} + \frac{\mathbf{F}}{m} \cdot \frac{\partial f}{\partial \mathbf{v}} = 0 \quad (7.2)$$

[The relativistic form replaces $(\mathbf{F}/m) \cdot \partial f/\partial \mathbf{v}$ with $(\mathbf{F}) \cdot \partial f/\partial \mathbf{p}$.] The operator

$$\frac{D}{Dt} \equiv \frac{\partial}{\partial t} + \mathbf{v} \cdot \frac{\partial}{\partial \mathbf{r}} + \frac{\mathbf{F}}{m} \cdot \frac{\partial}{\partial \mathbf{v}} \quad (7.3)$$

represents the Lagrangian or convective derivative following a single particle trajectory $(\mathbf{r}(t), \mathbf{v}(t))$ in phase space. Thus equation 7.2 can be interpreted as a statement that phase space density is constant along particle trajectories in phase space, i.e.,

$$f(\mathbf{r}, \mathbf{v}, t) = f(\mathbf{r}_o, \mathbf{v}_o, 0) \quad (7.4)$$

where $\mathbf{r}(t)$ and $\mathbf{v}(t)$ are solutions of the particle equations of motion

$$\begin{aligned} \frac{d\mathbf{r}}{dt} &= \mathbf{v} \\ \frac{d\mathbf{v}}{dt} &= \frac{\mathbf{F}}{m} \end{aligned} \quad (7.5)$$

[or $d\mathbf{p}/dt = \mathbf{F}$] with initial condition $\mathbf{r}(0) = \mathbf{r}_o$ and $\mathbf{v}(0) = \mathbf{v}_o$. This is Liouville's Theorem, which reduces the task of solving equation 7.2 to one of solving single particle motion. Liouville's Theorem provides the basic tool for analysing multi-spacecraft particle data at the level of phase space density. The theorem is a very powerful, but easily misused, approach to a variety of kinetic problems. We discuss a few typical applications below, although individual problems often require specifically tailored techniques.

Most applications of Liouville's Theorem rely on the further assumptions of adiabatic particle motion in static or slowly varying fields. Magnetic moment conservation, for example, yields

$$\frac{v_{\perp}^2}{B} = \frac{v_{\perp o}^2}{B_o} \quad (7.6)$$

where v_{\perp} is the particle gyro-speed perpendicular to the magnetic field. Energy conservation in an electrostatic potential further yields

$$v_{\parallel}^2 + v_{\perp}^2 + 2q\phi/m = v_{\parallel o}^2 + v_{\perp o}^2 + 2q\phi_o/m \quad (7.7)$$

Equations 7.6 and 7.7 completely characterise the particle trajectory with these assumptions if the fields B and ϕ are given/known.

7.3 Liouville Mapping: Known Fields

If the (dc) electromagnetic fields are known, Liouville's Theorem can be used to study the extent to which scattering or other non-dc effects influence the particle response. Additionally, Liouville's Theorem provides a valuable tool for exploring boundaries in velocity space which can often be seen in particle data. These boundaries or features occur, e.g., as separators between different sources of plasmas. The shape and timing of such boundaries provides rich information about the plasma source, such as its location, extent, temporal variation, and properties of the electromagnetic fields and processes both at the source and during the subsequent particle motion to the spacecraft location.

Historically, applications of Liouville's Theorem have employed single spacecraft measurements by using measurements taken at different times, converting temporal variations to spatial gradients assuming time stationarity. Multi-spacecraft missions offer the possibility to map from one spacecraft location to another, thereby eliminating this assumption. However, since Liouville's Theorem deals with particle trajectories, the spacecraft in question must be connected by the trajectories of particles of interest. For particles with small gyroradii and large speeds, such as electrons, this requirement reduces to the spacecraft being connected by a magnetic field line. Slower particles and finite gyroradius effects complicate the matter, as trajectories which reach a particular spacecraft may diverge when traced backward in time, so that further assumptions of spatial homogeneity or symmetry may be required. Additionally, in a time-dependent situation, particles of different speeds arriving at the same time will have left and traversed the intervening fields at different times.

In the case where the fields are known and time stationary, the process is straightforward. Let us consider two spacecraft, numbered 1 and 2 and located at \mathbf{r}_1 and \mathbf{r}_2 respectively.

1. Solve equations 7.5 for $\mathbf{r}(t)$, $\mathbf{v}(t)$.
2. From these solutions, or from equations 7.6 and 7.7 if appropriate, deduce $\mathbf{v}_2(\mathbf{v}_1)$ corresponding to individual particle trajectories.
3. Given the measured distribution $f_1(\mathbf{v})$ construct the mapped distribution $f_2^m(\mathbf{v} = \mathbf{v}_2(\mathbf{v}_1)) \equiv f_1(\mathbf{v} = \mathbf{v}_1)$ as demanded by Liouville's Theorem. That is, assign to each velocity \mathbf{v}_2 the phase space density from the original distribution f_1 at the location in phase space $(\mathbf{r}_1, \mathbf{v}_1)$. If the original distribution f_1 is represented by contours in phase space, this amounts to taking points along the contour, mapping the motion of these points using the trajectory equations, and connecting them up with a contour whose height is identical to the original.

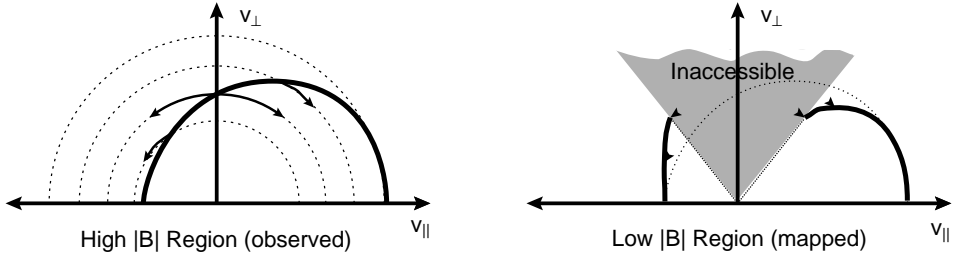


Figure 7.1: Sketch of a mapping using Liouville's Theorem to construct an expected distribution function. The solid line in the left is a contour of the observed phase space distribution in a high field region. The dotted lines are circles representing constant energy. In the absence of an electric potential, particle trajectories conserve energy and hence remain on such circles. When mapped to a low field region (assuming magnetic moment conservation), the final perpendicular velocity is given by equation 7.6. Four such trajectories are shown: two either side of 90° pitch angles, and two at intermediate pitch angles. Additionally, points at 0° and 180° pitch angles are unaltered in the mapping. The resultant mapped distribution is shown on the right (the arrowheads are left in to show the mapped points from the left diagram). Since these points started on a single contour of $f(\mathbf{v})$ they remain so, hence the mapped contour is found by connecting the arrowheads. Note that mapping from high to low fields, as shown here, leaves inaccessible regions in which the phase space density is filled in, if at all, by other processes or by trajectories which arrive there without passing through the high field region.

A sketch of this construction in the case of a simple magnetic field decrease and no potential difference is shown in Figure 7.1. Note that this mapping from \mathbf{r}_1 to \mathbf{r}_2 involves progressing some trajectories (e.g., $v_{\parallel} > 0$) forward in time while oppositely directed trajectories are advanced backward in time.

4. Compare the mapped distribution $f_2^m(\mathbf{v})$ with the observed one $f_2(\mathbf{v})$.

Discrepancies between observed and expected (mapped) distributions are indicators of one (or more) processes, e.g.,

1. Incorrect specification of the fields. This will result in systematic trends in the discrepancies in phase space. For example, an incorrect electric potential will shift all points by a fixed amount in energy.
2. Particle scattering between the two points. This will evidence itself by discrepancies localised to certain regions in phase space, with the observed distribution exhibiting generally smoother or more rounded features than the mapped ones.
3. Particle Mirroring/Inaccessible Regions of Phase Space. Some regions of phase space may not be connected by trajectories which pass through both locations. For example, if a field maximum lies between the two locations, particle trajectories around 90° will mirror and never reach the second location. The mapped distribution should

therefore have holes, although the size and shape of such holes depends on knowledge of the intervening structure. Holes also arise even in the case of spatially monotonic fields due to the same mirroring arguments as shown in the example sketched in Figure 7.1. All of these holes are connected to locations beyond the second location, and thus require specification of the phase space density there, rather than at the first location. A corollary to this statement is that it is always safer to map from low magnetic field regions to high ones rather than *vice versa*, since trajectories move toward 90° in this case and such holes are avoided or at least minimised. Electrostatic potentials also give rise to inaccessible regions in phase space.

Signatures of this behaviour include the appearance of holes in the mapped distribution, ridges along, e.g., lines of constant pitch angle which separate the accessible regions (connected to the first location) from the inaccessible ones (connected to points beyond the second location), parallel/anti-parallel symmetries in the observed distribution caused by mirroring, and other similar features.

4. Lack of sufficient connection of trajectories between the two locations, lack of sufficient time-synchronisation between the two measured distributions or knowledge of the temporal and spatial of the behaviour of the fields during the measurement interval, or some other aspect which makes the mapping inappropriate.

7.4 Liouville Mapping: Unknown Fields

In many cases, the intervening fields are not known, and are of interest. For example, the electrostatic potential can be very difficult to measure directly in space. In these circumstances, the mapping procedure can be used to determine the net field/potential changes by inverting the procedure. That is, treat the unknown fields as free parameters and adjust them to yield the best agreement between mapped and observed distributions. This can be done either by trial and error or via a formalised approach having fit the observed distributions with suitable functional forms and applying, e.g., a least squares algorithm to determine the functional coefficients/constants which yield the best fit.

Note that this method can not determine the detailed spatial variation of the fields between the two locations, but only the net changes between the two locations. However, the spatial variations could give rise to inaccessible regions in phase space, so that the mapped distributions may not fit everywhere in phase space. These mis-matches actually provide information about the intervening fields (e.g., magnetic field maxima, electric potential barriers, or trapping regions). These possible intervening structures imply that some considerable caution is required in applying simple mapping methods. Such problems may be overcome to some extent by assumptions of stationarity, etc., which then provides a whole sequence of distributions corresponding to relative motion between the spacecraft and plasma. This sequence should map from one distribution to the next and to all others, enabling some determination of the spatial variation of the fields corresponding to the individual measured distributions. The prospect of three or more multipoint measurements would allow, in perhaps fortuitous circumstances, a hybrid approach using multiple mappings.

7.5 Remote Sensing of Boundaries and Non-Local Processes

In addition to sensing intervening field structure, kinetic features can be used to infer global or distant morphology and events. For example, particles energised in a transient event require a finite amount of time to reach their point of observation. *A priori*, neither the event location nor initiation time are known. However, slower particles from the same event take longer to travel the same distance. In its simplest form, one can write down the trajectory as

$$x_{\text{obs}} - x_{\text{event}} = v \times (t_{\text{obs}}(v) - t_{\text{event}}) \quad (7.8)$$

Thus the time history of different velocity particle arrivals can be used to solve this set of equations for the unknown event position, x_{event} and event time t_{event} . In the sections which follow, we use similar arguments applied to more complicated situations. All of them rely on Liouville's Theorem, either explicitly or (as here) implicitly.

Additionally, such information is often used qualitatively to locate the observation point with respect to remote boundaries. For example, the Earth's bow shock is a copious source of suprathermal electrons, accelerated at the shock itself or escaping in the form of a broad heat flux from the hot magnetosheath. Thus the presence of such electrons implies that the observation point lies on a magnetic field line which is connected to the bow shock. In this regard, electrons have the advantage that their gyroradii are negligible and that their speeds are higher than characteristic bulk flows, so that they essentially travel along the magnetic field lines with, to lowest order, zero time delay.

In the realm of multi-satellite observations, each satellite then measures electrons on different field lines. Treating such measurements as simple on/off indicators of connection to the bow shock, the scientist could build an image of the bow shock with as many pixels as satellites. Such an image, or sequence thereof, would provide information on the shape and dynamics of the bow shock on scales which can be much larger than the spacecraft separation depending on the geometry. The fact that field lines are neither straight nor uniform complicates this analysis, but not fatally so in all cases.

7.6 Velocity Dispersion or "Velocity Filter"

Let us consider here a localised, time-stationary source of particles, by contrast with the transient discussed in the preceding section. As an analogy, consider a pier at one side of a body of water from which boats of different speeds travel toward the opposite shore. If the supply of boats is maintained then they will all arrive at the same point at the same rate with which they left. That is to say, the distribution in phase space at the arrival point will be identical to that at the departure pier, via Liouville's Theorem.

However, if the body of water is a flowing river, the boats will suffer a convective drift in addition to their own cross-river velocity. (In the plasma case, this drift is usually the $\mathbf{E} \times \mathbf{B}$ drift.) The fastest boats will still arrive close to their previous location, but slower boats will be swept further downstream. Thus boats leaving the same point will be "dispersed" along the opposite shore according to their velocities. At a given point on the opposite shore, the convective flow has acted as a filter to allow only particles of a single velocity to pass, hence the term "velocity filter" effect, which is also in common usage,

along with “time of flight” signatures. This concept is embodied in equation 7.8 together with a y -drift equation

$$y_{\text{obs}} - y_{\text{event}} = V_{\text{drift}} \times (t_{\text{obs}}(v) - t_{\text{event}}) \quad (7.9)$$

More common, perhaps, in the magnetospheric case is not a point source, but an extended one. Consider, then, the case of a semi-infinite line of piers extending downstream from the single pier invoked above, as sketched in Figure 7.2. If one approaches from the upstream direction on the opposite bank, the first boats to be seen will be the fastest ones which departed from the end pier. Downstream of this point, slightly slower boats from the end pier will arrive together with the fastest boats from the second pier. As one progresses further downstream, the fastest boats are always present and the distribution of arriving boats extends progressively to lower speeds as shown in the Figure.

This is what happens at the Earth’s bow shock, or for that matter any shock of finite extent, such as slow mode reconnection shocks in the geomagnetic tail, and is known as the “foreshock” region. Well upstream of the shock, the observer is disconnected magnetically and sees no shock-related particles. Just downstream of the tangent field line (in the case of curved shocks such as the bow shock) or separatrix (in the case of reconnection shocks) the fastest particles, typically energetic electrons, will be found. As one moves deeper into the foreshock, the electron distribution fills in to lower velocities down to some “cutoff” velocity. This cutoff velocity can be related, via the simple kinematics described above, to the geometry of the situation, to the extent that foreshock “coordinates” have been devised based on distance from the tangent point of contact (or “X-point”) along the magnetic field line to a point exactly upstream of the observer and a second distance from that upstream point to the observer. Deeper still lies an ion foreshock to which the same considerations, ignoring local acceleration processes, may be made. Similar reasoning has also been applied to the entry of particles into the cusp region and near the separatrix emanating from the X-line on the dayside magnetopause.

These ideas can be applied to a subset of the particle distribution (e.g., the field-aligned particles only) or to higher dimensional velocity space distributions, where magnetic moment conservation, or other trajectory considerations, must be included.

To date, the vast majority of applications have relied on data from a single spacecraft. Multiple spacecraft can be used as a collection of single spacecraft, to effectively determine the geometry of a physical region. Additionally, well-placed spacecraft can be used to apply some of the Liouville mapping techniques described above to map from source region to distant observation point, thereby shedding light on the intervening local processes which shape the distribution or provide local acceleration or scattering. Moreover, multiple spacecraft enable one to distinguish a localised source which is switched on at some time from a more extended source with a foreshock, as entering a foreshock yields a spacecraft time sequence which is very similar to that for a temporal switch on.

7.7 Particle Anisotropies and Remote Sensing

7.7.1 The Gyro-Orbit

Charged particles in a magnetic field describe a circular orbit perpendicular to the magnetic field vector \mathbf{B} , together with a motion parallel to the field producing a spiral

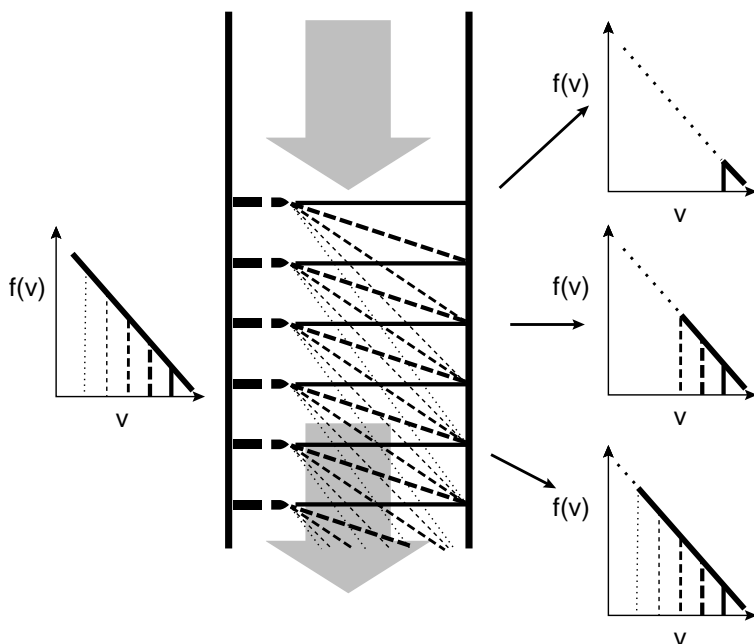


Figure 7.2: Velocity dispersion due to convective drift. A semi-infinite set of piers on the left all launch a range of boats whose velocity distribution function is shown to the left. The fastest boats (thick solid lines) travel in nearly horizontal lines, while slower boats (progressively thinner and more dotted lines) also suffer convection by the flow (broad grey arrows). On the opposite side of the river, the observed distribution of boats arriving depends on location. At the most upstream locations, no boats are seen. Moving downstream, first only the fastest boats are seen. Still further downstream slower boats from more upstream piers, together with faster boats from the opposite piers, are seen. Liouville's Theorem forces the phase space density to be constant along trajectories, enabling us to construct the observed distribution functions, as shown on the right for three locations. Note the cutoff at low velocities, and the way this cutoff systematically moves to lower velocities with downstream position.

trajectory. The frequency and radius of this “gyromotion” are given by

$$\omega = \frac{qB}{m} \quad (7.10)$$

$$\begin{aligned} R_g &= \frac{mv_{\perp}}{qB} \\ &= \frac{\sqrt{2mE}}{qB} \sin \beta \end{aligned} \quad (7.11)$$

where v_{\perp} is the particle velocity perpendicular to the magnetic field, and β is the pitch angle between \mathbf{v} and \mathbf{B} .

Equation 7.11 can also be written in vector form:

$$\mathbf{R}_g = \frac{m}{qB^2} \mathbf{v} \times \mathbf{B} \quad (7.12)$$

where \mathbf{R}_g is the vector from the particle’s position to its gyrocentre.

For protons, one can write

$$R_g = 4569.4 \text{ km} \cdot \frac{\sqrt{E/\text{keV}}}{B/\text{nT}} \sin \beta \quad (7.13)$$

Table 7.1 lists some sample results from equation 7.13, demonstrating that the gyroradius of energetic protons is quite comparable to the scale lengths of magnetospheric processes. For heavier ions, the gyroradius is even larger, scaling with \sqrt{m} (equation 7.11). On the other hand, electrons in these energy and field ranges possess gyroradii on the order of 100 km or less, too small for any observable effects normally.

Table 7.1: Sample proton gyroradii and periods

	10 nT	100 nT
10 keV	1400 km	140 km
100 keV	4600 km	460 km
Period	6.5 s	0.65 s

7.7.2 Particle Anisotropies

One of the most straightforward particle observations is a measure of the particle anisotropies, that is, the extent to which the distribution function deviates from isotropy in velocity space. Anisotropies can arise due to the relative motion of the observer with respect to the frame associated with an isotropic particle source, due to gradients in the intervening medium, or to additional sources, sinks, or scattering of the particles en route to the observer. First order directional anisotropies are most easily interpreted as net particle streaming. Here we outline the basic calculation of the expected anisotropies due to two effects: a gradient in the particle density and intrinsic first order anisotropies due to either constant velocity shifts (same for all particles) or more complex distributions. The

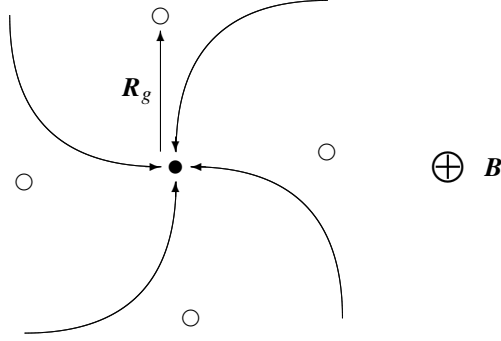


Figure 7.3: Ions impinging an observation point (black dot) from several directions; the gyrocentres of each orbit (open circles) are displaced by \mathbf{R}_g ; the magnetic field \mathbf{B} points into the paper.

first is related to the finite extent of the particle gyro-orbit and to spatial inhomogeneities. The second is purely a velocity-space property. If the mechanism(s) responsible for the anisotropy are known, a measurement of the local particle anisotropy provides direct inference of non-local characteristics (density gradients, source region velocity, etc.). We return later to the question of which effects are likely to be observable, as often there are competing processes which give rise to a reduced level of anisotropy.

The starting point for all such calculations is the basic conservation of particles. Let us assume that there is a spatial gradient in the number of particles. Due to the rapid gyromotion, the assumed spatial gradient will refer to particles' gyrocentre (or "guiding centre") as in Figure 7.3. Additionally, we assume that the particles' velocity distribution takes on a relatively simple, known form in a frame of reference moving with a bulk velocity \mathbf{V} with respect to the spacecraft frame.

Thus the problem commences by relating the phase space density $f(\mathbf{r}, \mathbf{v})$ in the spacecraft frame to the spatially-dependent, velocity distribution $f_{GC}(\mathbf{r}_{GC}, \mathbf{v}')$. The vector \mathbf{r}_{GC} is the position of the guiding centre of a particle whose instantaneous position and velocity are \mathbf{r} and \mathbf{v} . The velocity in the moving frame is denoted \mathbf{v}' . In practice, the spatial dependence of f_{GC} may be linked to the spatial dependencies of other parameters, such as B , which we shall ignore here. Equating the number of particles in an elemental volume in each coordinate system gives

$$f(\mathbf{r}, \mathbf{v}) d^3r d^3v = f_{GC}(\mathbf{r}_{GC}, \mathbf{v}') d^3r_{GC} d^3v' \quad (7.14)$$

This deceptively simple statement reduces the entire problem, and many similar to it, to one of coordinate transformation. In this case, the transformation laws are given by

$$\mathbf{r}_{GC} = \mathbf{r} + \mathbf{R}_g \quad (7.15)$$

$$\mathbf{v}' = \mathbf{v} - \mathbf{V} \quad (7.16)$$

Using equation 7.12 and assuming \mathbf{B} is uniform, it is straightforward to show that the jacobian relating $d^3r d^3v$ and $d^3r_{GC} d^3v'$ is unity in this case. It therefore follows from equation 7.14 that

$$f(\mathbf{r}, \mathbf{v}) = f_{GC}(\mathbf{r}_{GC}, \mathbf{v}') \quad (7.17)$$

Note that even if $f_{GC}(\mathbf{v}')$ is independent of gyrophase, $f(\mathbf{r}, \mathbf{v})$ contains a gyrophase dependence through the \mathbf{R}_g dependence of \mathbf{r}_{GC} .

We now Taylor expand the right-hand side of equation 7.17 about the point in phase space ($\mathbf{r}_{GC} = \mathbf{r}, \mathbf{v}' = \mathbf{v}$). This results in

$$\begin{aligned} f(\mathbf{r}, \mathbf{v}) &= f_{GC}(\mathbf{r}, \mathbf{v}) - \mathbf{V} \cdot \frac{\partial}{\partial \mathbf{v}} f_{GC}(\mathbf{r}, \mathbf{v}) + \mathbf{R}_g \cdot \nabla f_{GC}(\mathbf{r}, \mathbf{v}) \\ &\quad + \mathcal{O} \left[\left(\frac{R_g}{L} \right)^2, \frac{R_g}{L} \frac{V}{v}, \left(\frac{V}{v} \right)^2 \right] \end{aligned} \quad (7.18)$$

where L is the scale length of the spatial variation of f_{GC} . Now expand the velocity dependence of $f_{GC}(\mathbf{r}, \mathbf{v})$ in spherical harmonics, i.e.,

$$f_{GC}(\mathbf{r}, \mathbf{v}) = f_o(\mathbf{r}, v) \left[1 + \hat{\mathbf{v}} \cdot \boldsymbol{\varepsilon}_o + \hat{\mathbf{v}}^T \mathbf{S} \hat{\mathbf{v}} + \dots \right] \quad (7.19)$$

where $\boldsymbol{\varepsilon}_o$ is the first order anisotropy (i.e., the first degree term in the spherical harmonic expansion) and \mathbf{S} is the second order anisotropy. We shall assume that $|\mathbf{S}| \ll |\boldsymbol{\varepsilon}_o| \ll 1$ and restrict our calculations to first order results. The various derivatives of $f_{GC}(\mathbf{r}, \mathbf{v})$ are now easily calculated in terms of this expansion as

$$\nabla f_{GC}(\mathbf{r}, \mathbf{v}) = \nabla f_o(\mathbf{r}, v) [1 + \mathcal{O}(\boldsymbol{\varepsilon}_o)] \quad (7.20)$$

$$\begin{aligned} \frac{\partial f_{GC}(\mathbf{r}, \mathbf{v})}{\partial \mathbf{v}} &= \frac{\partial f_o(\mathbf{r}, v)}{\partial v} [1 + \mathcal{O}(\boldsymbol{\varepsilon}_o)] \\ &= \hat{\mathbf{v}} \frac{\partial f_o(\mathbf{r}, v)}{\partial v} [1 + \mathcal{O}(\boldsymbol{\varepsilon}_o)] \end{aligned} \quad (7.21)$$

Substituting the results from equations 7.19–7.21 into equation 7.18 brings us finally to $\Rightarrow 7.1$

$$\begin{aligned} f(\mathbf{r}, \mathbf{v}) &= f_o(\mathbf{r}, v) \left[1 + \hat{\mathbf{v}} \cdot \boldsymbol{\varepsilon}_o + \mathbf{R}_g \cdot \nabla \ln f_o(\mathbf{r}, v) - \mathbf{V} \cdot \hat{\mathbf{v}} \frac{\partial \ln f_o(\mathbf{r}, v)}{\partial v} \right. \\ &\quad \left. + \mathcal{O} \left(S, \frac{R_g}{L} \varepsilon_o, \frac{V}{v} \varepsilon_o, \left(\frac{R_g}{L} \varepsilon_o \right)^2, \frac{R_g}{L} \frac{V}{v}, \left(\frac{V}{v} \right)^2 \right) \right] \end{aligned} \quad (7.22)$$

Using equation 7.12 for \mathbf{R}_g , swapping dot and cross product, and general tidying leads to

$$f(\mathbf{r}, \mathbf{v}) \approx f_o(\mathbf{r}, v) [1 + \hat{\mathbf{v}} \cdot \boldsymbol{\varepsilon}_o + \hat{\mathbf{v}} \cdot \boldsymbol{\varepsilon}_{C-G} + \hat{\mathbf{v}} \cdot \boldsymbol{\varepsilon}_{\nabla n}] \quad (7.23)$$

where

$$\boldsymbol{\varepsilon}_{C-G} \equiv -\frac{\mathbf{V}}{v} \frac{\partial \ln f_o}{\partial \ln v} \quad (7.24)$$

is the Compton-Getting anisotropy and

$$\boldsymbol{\varepsilon}_{\nabla n} \equiv \frac{mv}{qB^2} \mathbf{B} \times \nabla \ln f_o \quad (7.25)$$

is the density gradient-induced anisotropy.

The above forms are not the way these anisotropies are usually presented, because the energetic particle detectors on which most observations are based do not measure f directly. The detector count rate N_{ijk}/t_{acc} is proportional to the differential intensity, dJ/dE ,

which is the number of particles per unit area per second per steradian per unit energy (E) travelling in the direction $\hat{\mathbf{v}}$ (see Chapter 5, particularly Section 5.2.3). Relating dJ/dE to f is easily accomplished by returning to an expression for the number of particles in an elemental phase space volume, which in this case becomes

$$f(\mathbf{r}, \mathbf{v}) d^3r d^3v = \frac{dJ}{dE} dA dt d\Omega dE \quad (7.26)$$

Noting $dA v dt = d^3r$, $dE = mv dv$ (for non-relativistic particles), and $d^3v = v^2 dv d\Omega$ leads to the conclusion that f and dJ/dE are related by

$$\frac{v^2}{m} f(\mathbf{r}, \mathbf{v}) = \frac{2E}{m^2} f(\mathbf{r}, \mathbf{v}) = \frac{dJ}{dE} \quad (7.27)$$

Thus the expansion to first order anisotropies for dJ/dE is just that obtained by multiplying equation 7.23 by the (isotropic) factor v^2/m . In particular, all the first-order Compton-Getting and density gradient anisotropies ϵ_{C-G} and $\epsilon_{\nabla n}$ are unchanged. However, we need to express them in terms of the measured parameter dJ/dE . The spatial variation is simply $\nabla \ln f_o = \nabla \ln(dJ/dE)$. Typically, dJ/dE is represented by a power law in energy of the form

$$\frac{dJ}{dE} \propto E^{-\gamma}$$

For such a form and non-relativistic particle speeds

$$\frac{\partial \ln f_o}{\partial \ln v} = \frac{v}{f} \left[-\frac{2m}{v^3} \frac{dJ}{dE} + \frac{m}{v^2} \frac{dE}{dv} \left(\frac{-\gamma}{E} \right) \frac{dJ}{dE} \right] = -2(\gamma + 1) \quad (7.28)$$

so that

$$\epsilon_{C-G} = 2(\gamma + 1) \frac{\mathbf{V}}{v} \quad (7.29)$$

$$\epsilon_{\nabla n} = \frac{mv}{qB^2} \mathbf{B} \times \nabla \ln \left(\frac{dJ}{dE} \right) \quad (7.30)$$

The Compton-Getting anisotropy arises because particles of fixed energy in the spacecraft frame correspond to different energies in the moving frame, depending on their direction. So although the distribution in the moving frame may be isotropic, different parts of the spectrum are being sampled at a single acceptance energy (spacecraft frame) and only the direction in the spacecraft frame is scanned. The resulting anisotropy reflects the spectral shape in the frame of bulk flow: if the spectrum were completely flat ($\partial f_o / \partial v = 0$) we have $\gamma = -1$ and $\epsilon_{C-G} = 0$; whereas if the slope of the spectrum goes positive, the anisotropy becomes negative, meaning it is opposite to the bulk flow \mathbf{V} . Normally the slope is negative and $\epsilon_{C-G} \parallel \mathbf{V}$. Note that particles with $v \lesssim |\mathbf{V}|$ are excluded from the present expansion. Treatment of these particles requires use of equation 7.17 without any approximations.

The intrinsic first order anisotropy represented by ϵ_o could also contain a component which would correspond likewise to a bulk velocity shift, and would incorporate the same Compton-Getting factor $2(\gamma + 1)/v$. It is more natural to include this component in the bulk velocity \mathbf{V} and reserve ϵ_o for first order anisotropies which are related to the internal structure of the velocity distribution function.

The density gradient anisotropy arises because there are more particle guiding centres (for a given energy) on one side than the other. For example, if there are more particles with guiding centres above the observation point shown in Figure 7.3 than below, there will be a greater flux of particles arriving from the left than from the right at the observation point; hence the observer will record an anisotropy. The anisotropy is perpendicular to both the density gradient and the magnetic field.

7.7.3 When is an Anisotropy not an Anisotropy?

The derivation given in the preceding section assumed that the only effects present were a frame transformation (and/or intrinsic first order anisotropy) and a density gradient. In fact, with the exception of the Compton-Getting anisotropy, most effects are counter-balanced by others, so that, e.g., an isolated density gradient anisotropy is rarely measured.

The fundamental argument rests with Liouville's Theorem, which requires that phase space density be constant along particle trajectories. If a distribution of particles is isotropic at some location, and particle trajectories connect the phase-space regions measured by the observer with that location, then the distortions in f are restricted to those implied by the particle trajectory solutions. A simple bulk frame shift produces the Compton-Getting effect derived above. Magnetic forces do not affect a particle's energy, and thus move particles along constant energy surfaces in velocity space which therefore cannot induce any anisotropy. And conservative electric forces act oppositely but symmetrically on trajectories coming from/going to the source location. That is, suppose particles of energy E are accelerated by such a field in coming from the source, so they appear at a higher energy $E + \Delta E$ when observed. Particles of the same observed energy $E + \Delta E$ travelling toward the source will lose an identical amount of energy, and will arrive at the source with an energy E . Since the source region corresponds, by assumption, to an isotropic velocity distribution, these two sets of particles will have the same phase-space density at the source and hence also, by Liouville's Theorem, at the observer's location. Thus the observer will also see an isotropic distribution.

Therefore, before applying the density gradient anisotropy given in Section 7.7.2 it is important to examine the process(es) which have given rise to the density gradient in the first place; such processes may themselves also lead to other anisotropies. Indeed, the above invocation of Liouville's Theorem insists that they will, and that the density gradient anisotropy will not be observed at all.

While this is a valuable lesson in the power of applying Liouville's Theorem, its naïve interpretation would suggest that first order anisotropies should be rarely observed and limited to the Compton-Getting values. In fact, first order anisotropies are frequently observed, and the reason is related to the restrictions necessary for Liouville's Theorem to hold. Some circumstances under which Liouville's Theorem is violated include:

1. when there are sources or sinks of particles;
2. when there are collisional, dissipative, or other forces for which $(\partial/\partial \mathbf{v}) \cdot \mathbf{F} \neq 0$;
3. when boundaries lead to particle trapping or exclusion, so that only portions of the distribution can be mapped from source to the spacecraft;
4. when spatial inhomogeneities lead to velocity filtering (see Section 7.6);

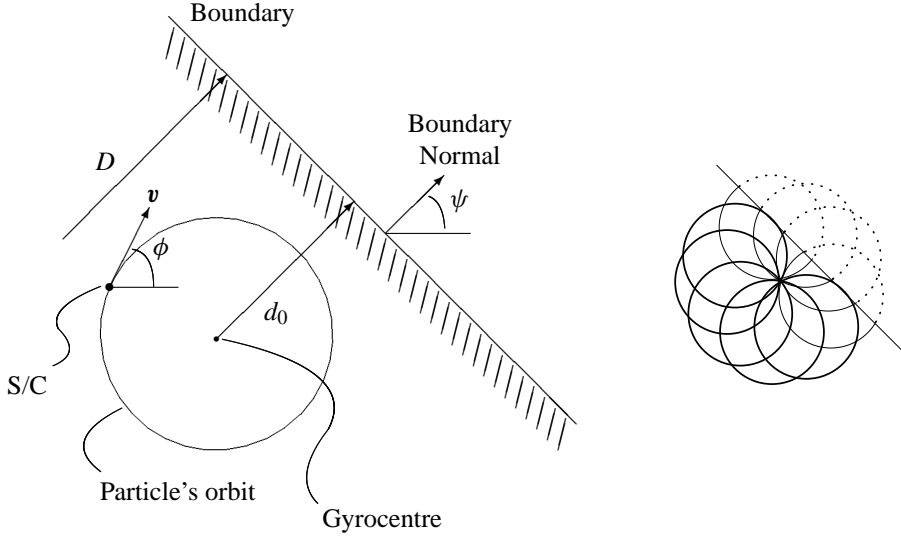


Figure 7.4: Left: the gyro-orbit of a particle whose gyrocentre is more than one gyroradius from an “absorbing” boundary. Right: snapshot of the boundary position, thick circles represent ion orbits that do not cross the absorbing boundary and therefore exhibit high intensities, while thin arcs are those parts of the trajectory connecting to the boundary, and dotted arcs are non-existent trajectories on the other side of the boundary.

5. when temporal variability at the source or elsewhere similarly leads to non-simultaneous observation of oppositely-directed trajectories.

In the next sections we explore some applications of particle kinetic effects at physical boundaries corresponding to some of these circumstances. Despite the breakdown of applicability of Liouville’s Theorem (or indeed because of it), it is still possible to use mapped particle trajectories in a quantitative way to infer remote plasma sources and structure.

7.7.4 Remote Sensing of Boundaries

In Section 7.7.2, we show how a large-scale density gradient can produce a first order anisotropy in the ion distributions. However, if there are very sharp gradients, such as the step function types at particle boundaries, the simple anisotropy formula of equation 7.30 no longer applies; instead, we can obtain snapshots of the moving boundary.

The left side of Figure 7.4 illustrates the gyro-orbit of an ion detected on the spacecraft (S/C) while approaching an absorbing boundary. The spacecraft is a distance D from the boundary, measured parallel to its normal, which is oriented at an angle ψ to some arbitrary reference azimuth. The ion is detected at velocity v , at phase angle ϕ ; its gyrocentre is a distance d_0 from the boundary.

$$d_0 = D - R_g \sin(\phi - \psi) \quad (7.31)$$

High fluxes are measured only if the entire orbit is on the particle-rich side of the boundary, for once an ion crosses the boundary on any part of its orbit, it is lost. In any real situation,

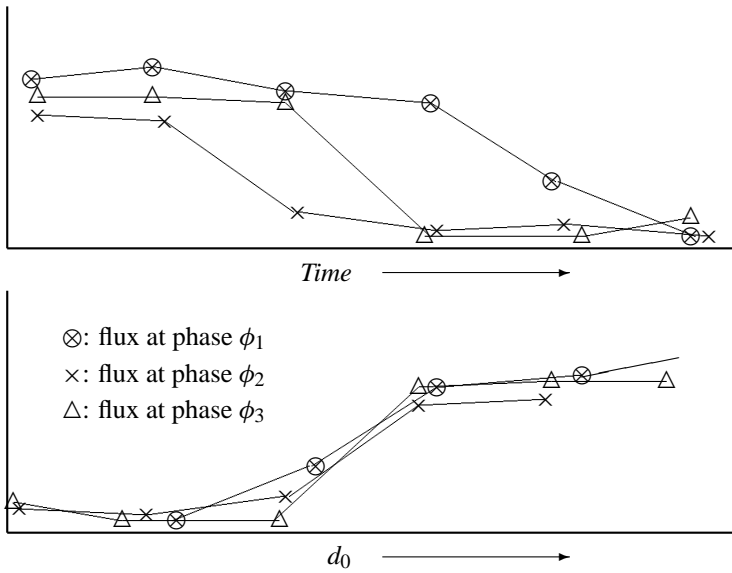


Figure 7.5: Ion fluxes plotted for 3 different phase angles as a function of time (upper) and gyrocentre distance d_0 from boundary (lower).

the boundary may also be a source of detected ions but at different flux levels related to the conditions on the other side of the boundary. For example, the magnetopause is an example of an absorbing boundary, since the magnetic field (usually) changes its configuration from one side to the other. This means that pure gyromotion cannot be maintained if an ion crosses the boundary. The ion finds itself in a field of a different orientation and adopts a new trajectory, which will not return to the spacecraft. Other ions from the far side of the boundary may cross over in such a way as to adopt a gyration that brings them to the spacecraft, giving rise to, e.g., lower flux measurements from certain arrival directions. This is illustrated in the right side of Figure 7.4, where the dotted arcs indicate non-existent particle trajectories, thin arcs indicate trajectories which take particles to or from the boundary, and thick circles indicate particle trajectories which do not intercept the boundary and so do not have reduced fluxes. As the spacecraft approaches the boundary, fluxes of particles on trajectories returning from the boundary reduce, beginning with the higher energy particles due to their larger gyroradii (see also Figure 7.6). The arrival directions at which these reduced fluxes appear span an increasing angular range as the boundary is approached. Thus, the ion flux, at a given energy and summed over all look directions, changes progressively from high to low as the spacecraft (together with the gyro-orbits arriving at the spacecraft) approaches and passes through the boundary. It is convenient to take the location of the gyrocentre itself as the indicator for high or low flux.

Displaying the boundary motion with a time series of snapshots may be very illustrative, but it would be desirable to have a more analytical method that can be reasonably automated. Particle fluxes, when plotted against time, exhibit different profiles for different phase angles, as shown in the upper plot in Figure 7.5. This “azimuthal asymmetry” results from the displacement of the various gyro-orbits. For each phase ϕ , determine the

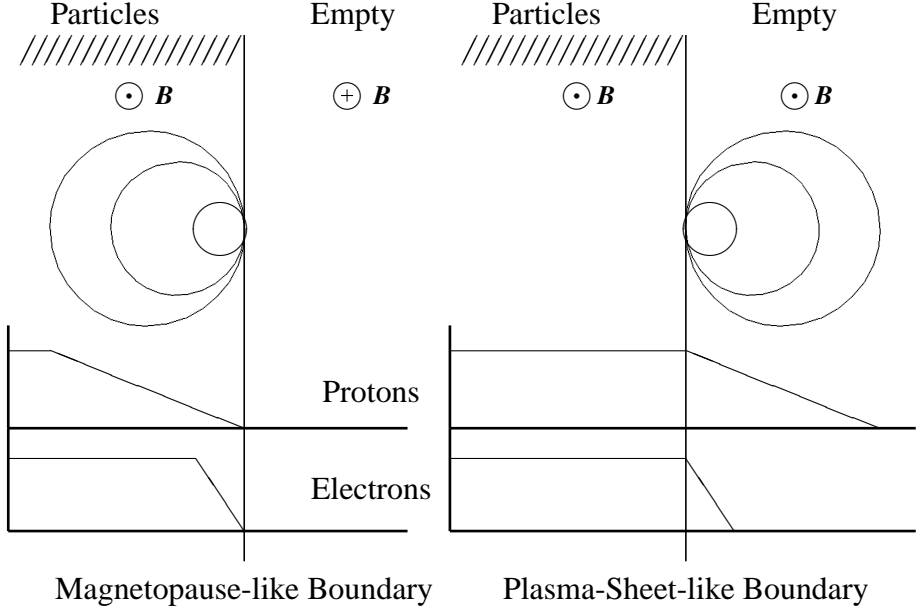


Figure 7.6: Different time profiles for flux variations while passing through a magnetopause-like, absorbing boundary (left) and a plasma-sheet-like, source boundary (right).

time $t_x(\phi)$ when its flux changes between high and low. We can now fit these switching times to a model of the boundary motion.

To do this, represent the boundary motion as $D = U(t - t_0)$; for each phase ϕ , solve equation 7.31 for the time t_x when $d_0 = 0$, the time when the gyrocentre for that phase crosses the boundary.

$$\begin{aligned}
 t_x(\phi) &= t_0 + \frac{R_g}{U} \sin(\phi - \psi) \\
 &= t_0 + \frac{\cos \psi}{U} R_g \sin \phi - \frac{\sin \psi}{U} R_g \cos \phi
 \end{aligned} \tag{7.32}$$

Equation 7.32 is linear in the three unknowns t_0 , $(1/U) \cos \psi$, and $(1/U) \sin \psi$. A linear least-squares fitting procedure may be applied to find these unknowns from the actual measurements of the $\{t_x(\phi)\}$.

Once the solution has been found, one can apply it to equation 7.31 to determine d_0 for each flux measurement. Plotting flux against d_0 , as in the lower panel of Figure 7.5, should demonstrate that this is the parameter that best orders the high-low transitions for all phases. Furthermore, snapshots like those on the right of Figure 7.4 can be drawn with the solution for the boundary motion, as graphic evidence for its correctness.

Note: in the above example, U is negative, meaning D and d_0 decrease with time, which is why the flux increases to the right in the lower panel.

7.7.5 Absorbing or Source Boundary?

In the above example, we have demonstrated remote sensing with a boundary that empties gyro-orbits when the centres cross it, something that sounds very unphysical. After all, the ion itself is never at the gyrocentre. Instead, one can imagine two types of discontinuities that can lead to particles vanishing, shown in Figure 7.6.

Absorbing boundary where the magnetic field changes configuration, so gyro-orbits are disturbed; this type is described above. Ideally, if any part of the orbit crosses the boundary, it must be empty. The magnetopause is such a boundary. There must of course be a source of the population on the particle-rich side.

Source boundary where the magnetic field is unchanged across it, but the field line on the boundary is the last one connecting to some remote source; ions can be injected onto this last field line at such a phase that the gyrocentre is outside the source region. Thus ions can be observed up to two gyroradii beyond the boundary. This corresponds to plasma sheet boundaries, or to the division between flux transfer events and their surroundings.

As long as one restricts oneself to a single gyroradius, there is no way remote sensing can distinguish the two types: the derived boundary locations will be shifted by two gyroradii depending on the model. However, with different gyroradii, either due to other pitch angles or particle species, comparison of their behaviour could resolve this, as shown in the plots in Figure 7.6. (These are meant to be omnidirectional fluxes plotted against position.)

In reality, there may be less difference between the two models than one thinks. The absorbing boundary is not perfectly solid; the randomness of the magnetic field rotation means many ions can indeed return from a boundary crossing, or there can be those re-entering after scattering from other gyro-orbits. The source boundary too is not absolute, for those ions populating the most external gyro-orbits would be originating from only a limited range of gyrophases. When gyro-averaged, such orbits would show a reduced density.

It is therefore best to maintain the gyrocentre itself as the determining criterion for full or empty orbits, not because it is physical, but because it is an average of the fuzziness of both models. One should, however, remain aware of the true causes of the changes in flux levels at boundary crossings.

7.8 Example Applications

Numerous applications of the basic ideas presented in this chapter can be found in the literature. A small subset are shown below by way of illustration.

7.8.1 1-D Mapping of Electrons at the Earth's Bow Shock

The Earth's bow shock represents a well-studied example of a collisionless shock. The internal structure of the shock layer has received considerable attention. Thermal electrons respond adiabatically to the changes in the magnetic field and to the (field-aligned) electric field (e.g., as measured in the deHoffmann-Teller frame in which the bulk flow outside the

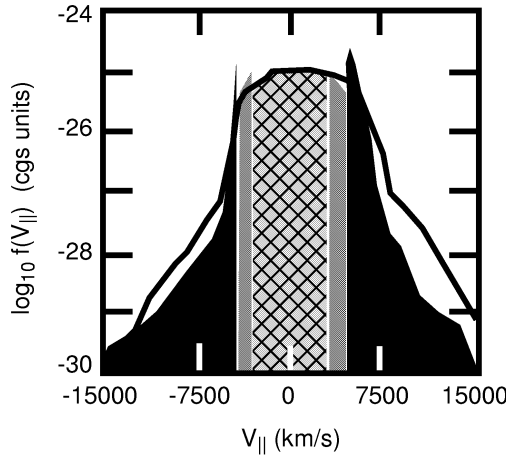


Figure 7.7: Observed (solid) and mapped (shaded) field-aligned electrons at a collisionless shock in which the fields have been determined experimentally. The innermost hatched region is inaccessible from either direction, while the intermediate region is populated by electrons originating from and returning to the downstream plasma. The outermost shaded regions correspond to electron trajectories which connect from far upstream to far downstream, and show the gross effect of the shock electrostatic field in accelerating the incident electrons to positive values (note the shifted peak). [After *Scudder et al.*, 1986, Figure 2B].

shock layer is field aligned). Figure 7.7 is taken from *Scudder et al.* [1986], and shows a cut of the electron distribution corresponding to field-aligned electrons. These electrons respond only to the electrostatic potential which, in this case, is determined observationally. The solid curve shows the measured distribution while the various shaded pieces are the result of Liouville mapping in the measured fields. The outer segments correspond to electron trajectories which connect from the asymptotic upstream to downstream states. Note the shift in the peak to positive velocities due to the acceleration by the potential. The innermost region is inaccessible from both the upstream and downstream regions (i.e., these trajectories are trapped near the vicinity of the shock), while the intervening regions correspond to electrons which start in the asymptotic downstream region, have insufficient energy to overcome the shock potential, and thus return to the downstream region. The mapping shows how well the overall features of the distribution are the result of the shock dc fields (the upstream distribution is much narrower in velocity, comparable to the width of the shifted peak).

7.8.2 2-D Electron Mapping in the Earth's Foreshock

An example of 2-D electron distributions is shown in Figure 7.8. The points represent measured electron distributions at several phase space densities taken in high field regions (top) and low field regions (bottom) within ULF waves present in the Earth's foreshock. These measured points are fit with smooth contours, in this case ellipses shown as the

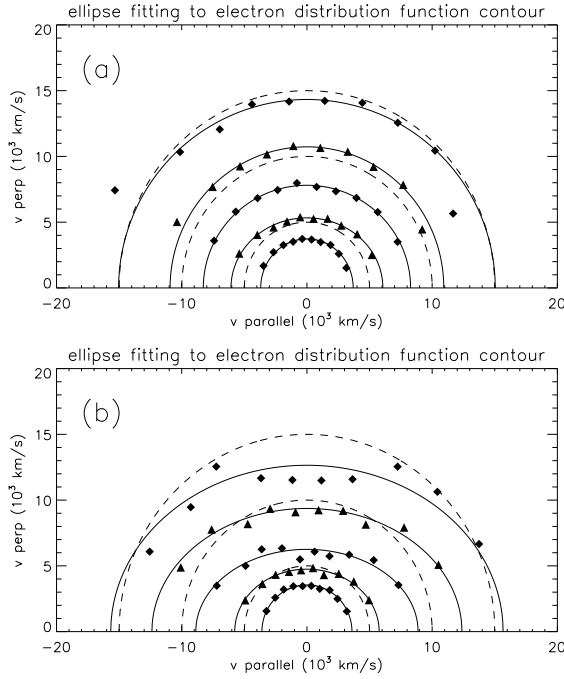


Figure 7.8: Observed contours of constant phase space density (symbols) and fitted ellipses (solid curves) in the field maxima (top) and minima (bottom) of electrons in the ULF wave field upstream of the Earth's bow shock. Dashed curves are circles to guide the eye. Adiabatic electron behaviour results in nearly isotropic contours in the high field regions, and a quantitative comparison between the fitted ellipses yields estimates of the intervening magnetic field and electric potential. [From *Chisham et al.*, 1996, Figure 6].

solid curves. Assuming adiabatic electron behaviour, equations 7.6 and 7.7 can be used to show that ellipsoidal contours map to ellipses. Moreover, if the contours at the minimum in the magnetic field are used, there are no regions inaccessible to mapped trajectories for monotonic fields and potentials (although not all trajectories starting at the minimum reach these regions). Thus the parameters of the ellipses at the same phase space density at field minimum and maximum (or anywhere in between, if desired) can be used to deduce the electrostatic field. This, together with the time variation throughout several wave cycles, is reported by *Chisham et al.* [1996]. The results show which range of electron energies (i.e., phase space densities) participate in a relatively simple adiabatic response and which regions where some other process(es) or breakdown in the assumptions occur. In principle, this method could be used when neither the electric nor magnetic field were known.

7.8.3 Remote Sensing of the Earth's Bow Shock by Field-Aligned Energetic Electrons

The Earth's bow shock is a copious emitter of suprathermal and energetic electrons into the upstream solar wind. As discussed in Section 7.6, dispersion in the foreshock re-

gion due to the velocity filter effect results in only particles above a certain (field-aligned) velocity being observed at any location behind the tangent line. The total electron population is then the superposition of the solar wind thermal population and, above this local cutoff velocity, the shock-associated energetic electrons. An example of such an electron distribution is shown in Figure 7.9. Since the cutoff velocity is related only to the geometry of the situation and the “convection” velocity, a direct measurement of this cutoff velocity and the local $\mathbf{E} \times \mathbf{B}$ convective drift enables one to reconstruct the geometry, that is, to locate the observation point in relation to the field line which is tangent to the shock. By assuming an empirical shape for the bow shock (see Section 10.4.6 on page 259), its position and scale can thus be determined. Multiple satellite studies of this kind open up the possibility to provide a more detailed, time-dependent picture of the bow shock position and shape, including local deviations from the model shape. Since these electron beams are unstable to Langmuir oscillations, direct observation of the electron beams and knowledge of the bow shock geometry provide wave analysts with an estimate of the (resonant) wave vector. This quantity can not be determined by direct wave measurements, and is crucial in theories of the nonlinear wave development.

7.8.4 Ions in the Cusp

Dayside reconnection at the magnetopause leads to an injection of energetic particles which travel along field lines and penetrate to low altitudes in the polar regions. The reconnection is associated with an $\mathbf{E} \times \mathbf{B}$ convection which results in a velocity filtering as the reconnected field lines convect poleward from the dayside. The consequence is that only ions above a cutoff energy dictated by such time-of-flight considerations will be observed at any location in the cusp. Assuming the reconnection yields a Maxwellian distribution moving along the field at the Alfvén speed in the rest frame of the field line (the deHoffmann-Teller frame), the distributions observed in the magnetosphere will be drifting Maxwellians truncated at the deHoffmann-Teller frame speed, referred to as Cowley-D distributions. An example of such a distribution is shown in Figure 7.10.

Moreover, an equatorward-moving spacecraft at low altitudes over the poles will see ions down to a cutoff velocity determined by the ion velocity and the time of flight from the reconnection site or, equivalently, distance behind the most recently reconnected field line. This is just the velocity filter effect in the curved dipole geometry of the Earth’s magnetic field rather than the straight geometry depicted in Figure 7.2. An example of such a set of observations together with a sketch of the velocity filter effect in curved geometry is shown in Figure 7.11.

7.8.5 Remote Sensing of a Flux Transfer Event by Finite Gyroradius Effects

The passage of a flux transfer event (FTE) over the ISEE-2 spacecraft as observed by the medium energy particle spectrometer has been extensively analysed by *Daly and Keppler* [1983]. Figure 7.12 shows the proton intensities in various pitch angle ranges during the passage into the FTE, from low to high intensity regimes: in the left diagram, the data are plotted against time, in the right one, by the distance of the gyrocentre from the deduced boundary. This diagram is the equivalent of Figure 7.5 with real data. The FTE boundary orientation and its speed are found from the best-fit solution of equation 7.32.

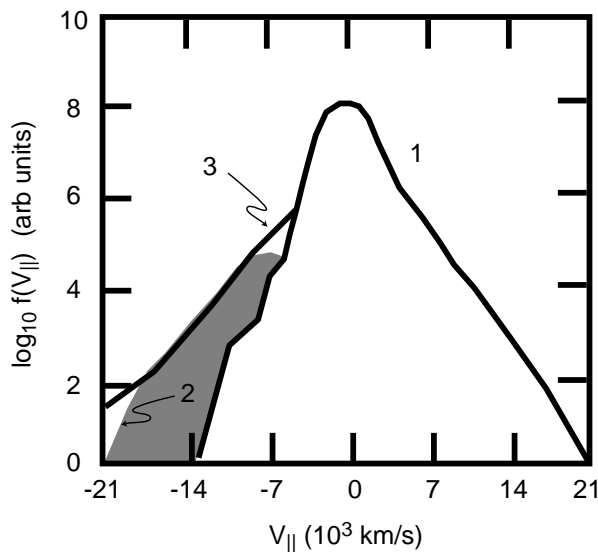


Figure 7.9: Parallel electron distributions observed with increasing penetration (1-3) into the Earth's foreshock. Note the thermal electron distribution and the superimposed electron beam (shaded) above the cutoff velocity as the foreshock is entered (2) and the extension to lower $|v_{||}|$ in 3. [After *Feldman et al., 1983*, Figure 1].

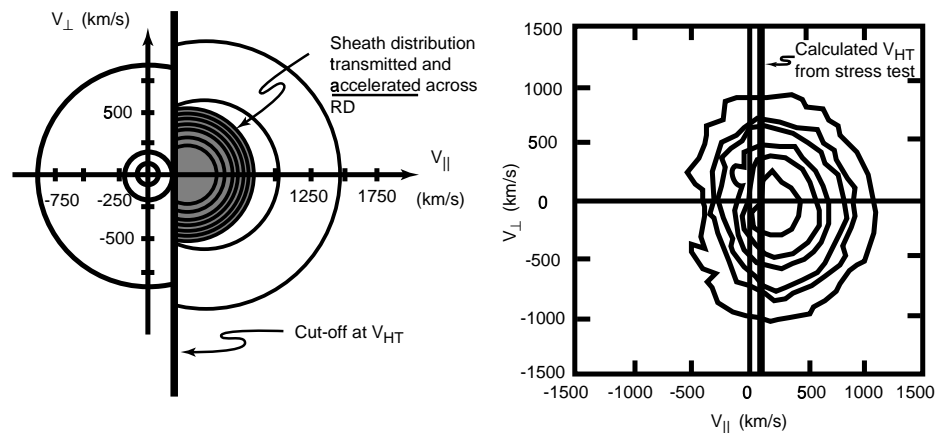


Figure 7.10: Predicted (left) and measured (right) Cowley-D distributions inside the Earth's magnetopause. The transmitted magnetosheath distributions have been accelerated by the reconnection process, and the D-shaped distribution is a consequence of the mapping and accessibility arguments discussed in this chapter. [After *Smith and Lockwood, 1996*, Figure 7].

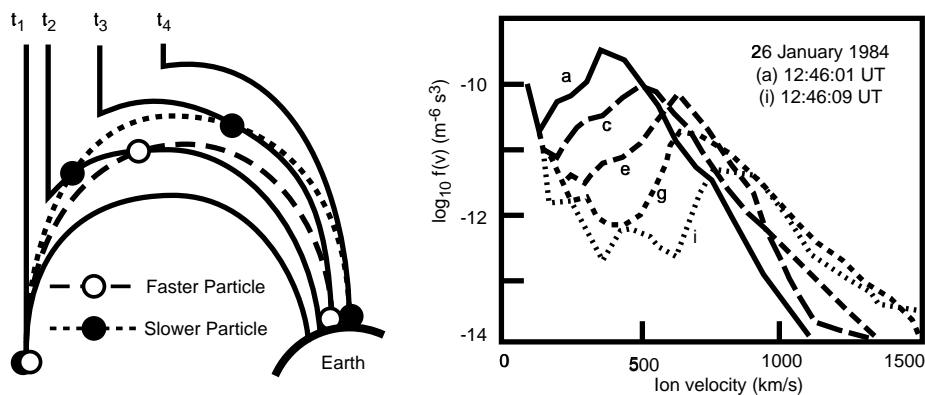


Figure 7.11: Velocity filtering in the Earth’s dipole geometry (left) showing a newly reconnected field-line at successive times (solid) and the actual trajectories of particles with different speeds (dashed). Right: observations of ion distributions (right) made by a low-altitude satellite moving equatorward from (a) to (i). Note the progressive loss of lower velocity particles as the satellite moves toward the equator. Some distributions have been omitted here for clarity. [After *Onsager and Elphic, 1996*, Figure 3 (left) and *Lockwood et al., 1994*, Figure 2 (right)].

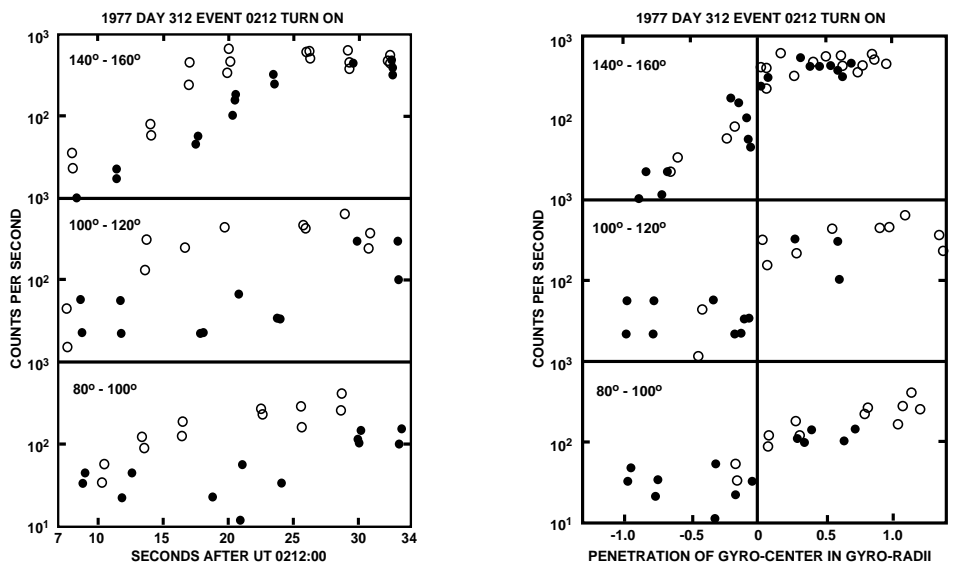


Figure 7.12: Left: plot of proton intensities against time in 3 pitch angle ranges during the entrance to the FTE; open and closed circles distinguish particles of different azimuths. Right: the same data plotted against d_0 , distance of the gyrocentre inside the particle-rich side of the boundary, calculated from the best-fit solution. [From *Daly and Keppler, 1983*, Figure 4].

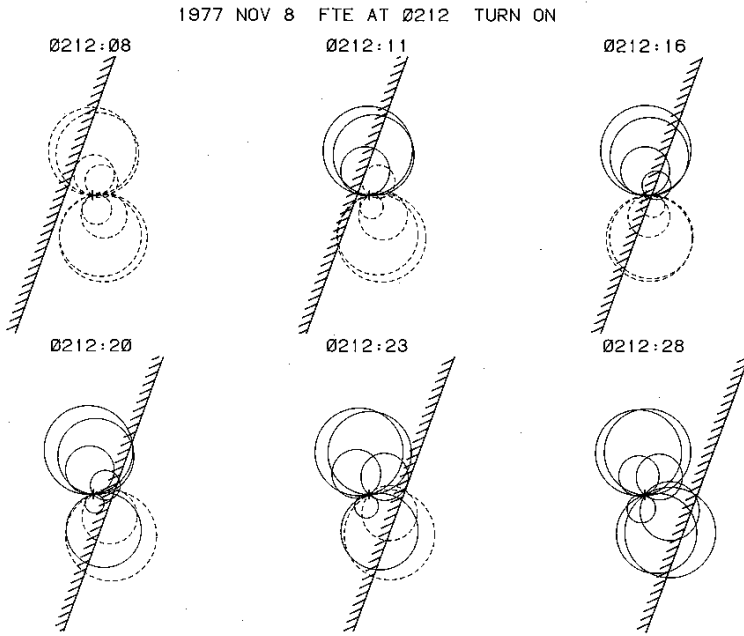


Figure 7.13: The proton gyro-orbits for 8 sectors over 6 spacecraft spins during the entrance to the FTE. The central time of each spin is printed above each plot as UT (hhmm:ss). Dashed circles are for particles of low intensities, solid ones for those of high intensity. The boundary drawn is the best-fit solution of equation 7.32. The high-intensity side of the boundary is marked by shading. [From *Daly and Keppler, 1983*, Figure 5].

The motion of the boundary is illustrated in snapshots over 6 spacecraft spins (≈ 4 s) in Figure 7.13, where solid and dashed circles are used to indicate gyro-orbits with high and low intensities respectively. This diagram confirms that the proton intensities switch from low to high when the gyrocentre crosses the boundary. The speed of the boundary was found to be 44 km/s along its normal, which translated into a speed for the FTE of 94 km/s, assuming it moves in the plane of the magnetopause.

Multi-spacecraft missions open up the possibility to extend such analyses from the simplifying planar assumption to more complicated boundary shapes.

Bibliography

Most basic plasma textbooks include a discussion of Liouville's Theorem. For example:

Bittencourt, J. A., *Fundamentals of Plasma Physics*, chap. 5, Pergamon Press, Oxford, 1986.

Papers that discuss Vlasov mapping at shocks are:

Scudder, J. D., Mangeney, A., Lacombe, C., Harvey, C. C., Wu, C. S., and Anderson, R. R., The resolved layer of a collisionless, high β , supercritical, quasi-perpendicular shock wave, 3. Vlasov electrodynamics, *J. Geophys. Res.*, **91**, 11 075–11 097, 1986. The shock electron results shown in Figure 7.7 are taken from this paper.

Schwartz, S. J., Thomsen, M. F., Feldman, W. C., and Douglas, F. T., Electron dynamics and potential jump across slow mode shocks, *J. Geophys. Res.*, **92**, 3165, 1987.

Schwartz, S. J., Thomsen, M. F., Bame, S. J., and Stansberry, J., Electron heating and the potential jump across fast mode shocks, *J. Geophys. Res.*, **93**, 12 923–12 931, 1988.

The 2-D electron mapping reproduced in Figure 7.8 is taken from:

Chisham, G., Schwartz, S. J., and Burgess, D., The anisotropy variations of electron distributions in the terrestrial ion foreshock, *J. Geophys. Res.*, **101**, 445–455, 1996.

Two-D mapping has also been invoked to explain ion behaviour in mirror mode waves in the Earth's magnetosheath, e.g.:

Leckband, J. A., Burgess, D., Pantellini, F. G. E., and Schwartz, S. J., Ion distributions associated with mirror waves in the Earth's magnetosheath, *Adv. Space Res.*, **15**, (8/9)345–(8/9)348, 1995.

Kivelson, M. G. and Southwood, D. J., Mirror instability II: The mechanism of non-linear saturation, *J. Geophys. Res.*, **101**, 17 365–17 371, 1996.

The foreshock coordinate system mentioned briefly in the text, and which has since been used by many other authors, was introduced by:

Filbert, P. C. and Kellogg, P. J., Electrostatic noise at the plasma frequency beyond the Earth's bow shock, *J. Geophys. Res.*, **84**, 1369, 1979.

The electron distribution shown in Figure 7.9 is taken from:

Feldman, W. C., Anderson, R. C., Bame, S. J., Gary, S. P., Gosling, J. T., McComas, D. J., Thomsen, M. F., Paschmann, G., and Hoppe, M. M., Electron velocity distributions near the Earth's bow shock, *J. Geophys. Res.*, **88**, 96, 1983.

The behaviour of particles in the cusp regions are reviewed by:

Smith, M. F. and Lockwood, M., Earth's magnetospheric cusps, *Rev. Geophys.*, **34**, 233–260, 1996. Figure 7.10 is reproduced there from its original source:

Smith, M. F. and Rodgers, D. J., Ion distributions at the dayside magnetopause, *J. Geophys. Res.*, **95**, 11 617, 1991. This figure has been used by many authors since.

The low-altitude satellite measurements shown at the right of Figure 7.11, together with the sketch at the left, are taken from, respectively:

Lockwood, M., Onsager, T. G., Davis, C. J., Smith, M. F., and Denig, W. F., The characteristics of the magnetopause reconnection X-line deduced from low-altitude satellite observations of cusp ions, *Geophys. Res. Lett.*, **21**, 2757–2760, 1994.

Onsager, T. G. and Elphic, R. C., Is magnetic reconnection intrinsically transient or steady-state? the Earth's magnetopause as a laboratory, *Trans. Am. Geophys. Union (EOS)*, **77**, 241–250, 1996.

Velocity filtering has been applied to good effect near the separatrices associated with the dayside X-line and in the geomagnetic tail by, respectively:

Gosling, J. T., Thomsen, M. F., Bame, S. J., Onsager, T. G., and Russell, C. T., The electron edge of the low latitude boundary layer during accelerated flow events, *Geophys. Res. Lett.*, **17**, 1833–1836, 1990.

Onsager, T. G., Thomsen, M. F., Gosling, J. T., and Bame, S. J., Electron distributions in the plasma sheet boundary layer: Time-of-flight effects, *Geophys. Res. Lett.*, **17**, 1837–1840, 1990.

Finite gyroradius effects were first intensively exploited at the magnetopause for remote sensing purposes by:

Kaufmann, R. L. and Konradi, A., Speed and thickness of the magnetopause, *J. Geophys. Res.*, **78**, 6549–6568, 1973.

Seminal work on particle anisotropies and Compton-Getting factors include:

Stern, D., The cosmic ray anisotropy, *Planet. Space Sci.*, **12**, 973–978, 1964.

Forman, M. A., The Compton-Getting effect for cosmic-ray particles and photons and the Lorentz-invariance of distribution functions, *Planet. Space Sci.*, **18**, 25–31, 1970.

See also Gleeson, L. J., The equations describing the cosmic-ray gas in the interplanetary region, *Planet. Space Sci.*, **17**, 31–47, 1969.

Example papers for the use of finite gyroradius effects to analyse the motion of a flux transfer event by the graphics method, to measure the density gradients in the plasma sheet of the geomagnetic tail, and to determine the orientation of the current sheet in the tail of comet Giacobini-Zinner, are respectively:

Daly, P. W. and Keppler, E., Remote sensing of a flux transfer event with energetic particles, *J. Geophys. Res.*, **88**, 3971–3980, 1983.

Daly, P. W., Sanderson, T. R., and Wenzel, K.-P., Survey of energetic ($E > 35\text{keV}$) ion anisotropies in the deep geomagnetic tail, *J. Geophys. Res.*, **89**, 10 733–10 739, 1984.

Daly, P. W., Sanderson, T. R., Wenzel, K., Cowley, S. W. H., Hynds, R. J., and Smith, E. J., Gyroradius effects on the energetic ions in the tail lobes of comet P/Giacobini-Zinner, *Geophys. Res. Lett.*, **13**, 419–422, 1986

— 8 —

Minimum and Maximum Variance Analysis

BENGT U. Ö. SONNERUP AND MAUREEN SCHEIBLE

*Dartmouth College
Hanover, NH, U.S.A.*

8.1 Introduction

The main purpose of minimum or maximum variance analysis (MVA) is to find, from single-spacecraft data, an estimator for the direction normal to a one-dimensional or approximately one-dimensional current layer, wave front, or other transition layer in a plasma. In the present chapter, we will develop the method in the context of determination of such a unit normal vector, \hat{n} , from minimum variance analysis of magnetic field vector data (MVAB) acquired by a spacecraft as it traverses a current sheet. This is the application where the method was first employed [Sonnerup and Cahill, 1967]. Other applications, for example the use of maximum variance analysis of measured electric field vectors (MVAE) for determination of a normal direction, will be discussed briefly toward the end of the chapter. In the context of current-layer traversals by several clustered spacecraft, MVA can be used on data from each individual spacecraft as an initial step. The results can then be used to establish stationarity of current-layer orientation and as a partial benchmark test for more sophisticated analysis tools, e.g., those in Chapters 11 and 15.

The presentation is organised as follows: in Section 8.2, the physical and mathematical basis of MVA is reviewed; in Section 8.3 error estimates are presented; in Section 8.4 other applications are examined, and in Section 8.5 an overview and discussion of results from a test case observed by the spacecraft AMPTE/IRM is provided. Finally, Section 8.6 contains a summary of the main items and precautions that pertain to MVA.

8.2 Theory

8.2.1 Elementary Considerations

The minimum variance analysis technique, applied to magnetic field vector data measured during a spacecraft traversal of a transition layer, is based on an idealised one-dimensional (1-D: $\partial/\partial x = 0$, $\partial/\partial y = 0$) model of the layer so that only one of the three terms remains in the cartesian expression for the divergence of \mathbf{B} :

$$\nabla \cdot \mathbf{B} = \partial B_z / \partial z = 0 \tag{8.1}$$

In other words, B_z is independent of z . Here (x, y, z) is a local cartesian coordinate system—unknown *a priori*—with its z axis pointing along the sought-after vector, \hat{n} , nor-

mal to the layer. It follows from Faraday's law, $\nabla \times \mathbf{E} = -\partial \mathbf{B} / \partial t$, that the field component B_z must also be time independent, $\partial B_z / \partial t = 0$, in such an idealised structure so that a spacecraft traversing it would observe a strictly constant value of B_z . In that case, only three distinct vector measurements, $\mathbf{B}^{(1)}$, $\mathbf{B}^{(2)}$, and $\mathbf{B}^{(3)}$, are needed to determine $\hat{\mathbf{n}}$. Typically, $\mathbf{B}^{(1)}$ and $\mathbf{B}^{(3)}$ would be measured on opposite sides of the layer, and $\mathbf{B}^{(2)}$ would be measured somewhere near the middle of the layer. Since $\Rightarrow 8.1$

$$\mathbf{B}^{(1)} \cdot \hat{\mathbf{n}} = \mathbf{B}^{(2)} \cdot \hat{\mathbf{n}} = \mathbf{B}^{(3)} \cdot \hat{\mathbf{n}} \quad (8.2)$$

the vectors $(\mathbf{B}^{(1)} - \mathbf{B}^{(2)})$ and $(\mathbf{B}^{(2)} - \mathbf{B}^{(3)})$ are tangential to the layer so that their cross product, assuming it is not zero, is along $\hat{\mathbf{n}}$:

$$\hat{\mathbf{n}} = \pm \frac{(\mathbf{B}^{(1)} - \mathbf{B}^{(2)}) \times (\mathbf{B}^{(2)} - \mathbf{B}^{(3)})}{|(\mathbf{B}^{(1)} - \mathbf{B}^{(2)}) \times (\mathbf{B}^{(2)} - \mathbf{B}^{(3)})|} \quad (8.3)$$

Several features of this simple example are of interest:

1. The calculation is not based on the assumption $\mathbf{B} \cdot \hat{\mathbf{n}} = 0$, which is sometimes made, but it allows determination of the actual value of the normal field component

$$B_n = \mathbf{B} \cdot \hat{\mathbf{n}} = \pm \frac{\mathbf{B}^{(1)} \cdot (\mathbf{B}^{(2)} \times \mathbf{B}^{(3)})}{|(\mathbf{B}^{(1)} - \mathbf{B}^{(2)}) \times (\mathbf{B}^{(2)} - \mathbf{B}^{(3)})|}$$

2. Exactly three vectors are needed to obtain a unique determination of $\hat{\mathbf{n}}$ and B_n , provided the difference vectors $(\mathbf{B}^{(1)} - \mathbf{B}^{(2)})$ and $(\mathbf{B}^{(2)} - \mathbf{B}^{(3)})$ in equation 8.3 are not aligned. This means that the three vector components tangential to the current layer cannot be arranged as shown in Figure 8.1a but must be as shown in Figure 8.1b. In the former case, the line ABC itself lies in the tangent plane of the layer but any vector $\hat{\mathbf{n}}$ perpendicular to ABC satisfies equation 8.2. In such a situation an additional condition, e.g., $\mathbf{B} \cdot \hat{\mathbf{n}} = 0$, is needed in order to obtain a unique $\hat{\mathbf{n}}$ vector (e.g., $\hat{\mathbf{n}} \propto \pm \mathbf{B}^{(1)} \times \mathbf{B}^{(3)}$). Note that in Figure 8.1a the electric current in the layer is unidirectional and perpendicular to the line ABC. In Figure 8.1b, the current vectors in the layer, which are perpendicular to the difference vectors, are not unidirectional and a unique (except for sign) $\hat{\mathbf{n}}$ vector is obtained from equation 8.3.
3. If the difference vectors are small, i.e., if two or all three of the measured vectors are nearly the same, then equation 8.3 approaches the form zero over zero so that no reliable normal vector is obtained. In practice, this situation is avoided, to the maximum extent that it can, by using one measured vector on each side of the layer and one somewhere near its centre, as mentioned above.
4. If only two (non-aligned) vectors, $\mathbf{B}^{(1)}$ and $\mathbf{B}^{(3)}$ say, are measured then an additional assumption, such as $\mathbf{B} \cdot \hat{\mathbf{n}} = 0$ for tangential discontinuities or coplanarity for shocks (see Chapter 10), is needed in order to obtain a normal vector. But it is noted that $(\mathbf{B}^{(1)} - \mathbf{B}^{(3)})$ still provides a vector tangential to the layer.
5. If more than three vectors are measured and the current is not unidirectional, then more than one $\hat{\mathbf{n}}$ vector determination can be made and, except in ideal circumstances, the resulting vectors are not exactly the same. This is the case to be addressed below.

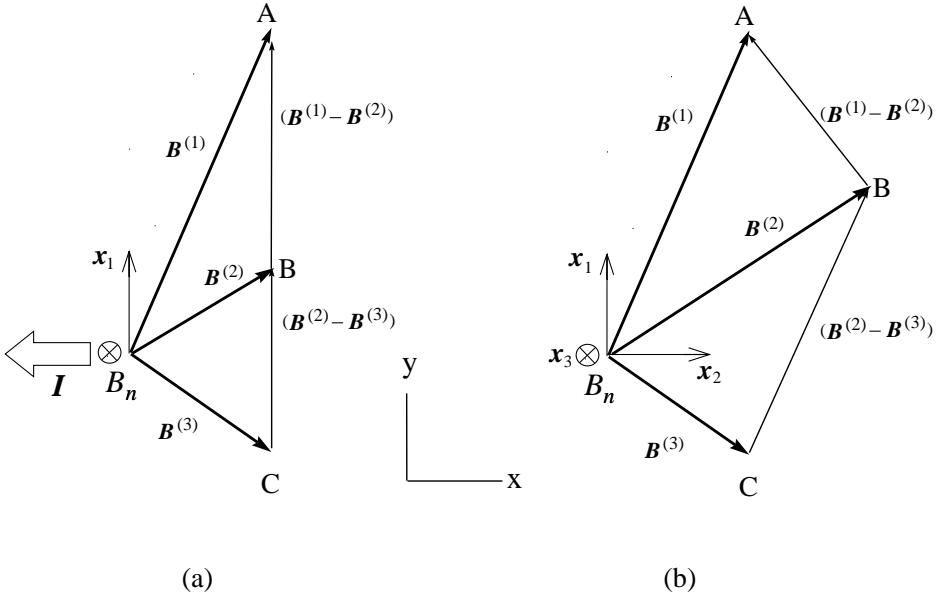


Figure 8.1: Projection onto the magnetosheath tangent xy plane of three \mathbf{B} vectors measured during spacecraft traversal of a 1-D current sheet. Field $\mathbf{B}^{(1)}$ is measured on one side, $\mathbf{B}^{(3)}$ on the other side, and $\mathbf{B}^{(2)}$ somewhere in the middle of the sheet. Each field vector has the same component B_n pointing along the normal vector, i.e., pointing into the paper. Difference vectors such as $(\mathbf{B}^{(1)} - \mathbf{B}^{(2)})$ and $(\mathbf{B}^{(2)} - \mathbf{B}^{(3)})$ are therefore tangential. (a) Difference vectors are colinear so that equation 8.3 fails to yield a normal vector, $\hat{\mathbf{n}}$; electric current is unidirectional and perpendicular to line ABC. (b) Difference vectors are not colinear, current is not unidirectional, and equation 8.3 yields a unique $\hat{\mathbf{n}}$ vector. The eigenvector triad $(\mathbf{x}_1, \mathbf{x}_2, \mathbf{x}_3)$, where $\mathbf{x}_3 = \hat{\mathbf{n}}$, is shown.

6. A constant off-set vector in the measured vectors $\mathbf{B}^{(1)}$, $\mathbf{B}^{(2)}$, and $\mathbf{B}^{(3)}$ does not influence the calculated normal direction. However, it influences the value of B_n , except in the special case where the off-set vector lies in the tangent plane.

8.2.2 Derivation of Minimum Variance Analysis on Magnetic Field (MVAB)

For real transition layers observed in space there are usually more or less pronounced deviations from the ideal 1-D model described in the previous section. The layer is likely to have 2-D or 3-D internal structures which evolve in time and to have temporal fluctuations in the orientation of its normal as well. In some cases, a systematic temporal change in the normal direction may occur during the spacecraft traversal time. To these effects must be added random as well as systematic measurement errors. For modern magnetometers, the former are usually negligible compared to other uncertainties but the latter can sometimes

arise in the form of a zero-level offset of the magnetometer measuring the field component along the spacecraft spin axis. As pointed out already, a constant offset of this type does not influence the determination of $\hat{\mathbf{n}}$ from equation 8.3 and this property will be seen to apply to MVAB as well. Another feature of the real situation is that the high time resolution available in many magnetometer experiments allows many vector measurements, $\mathbf{B}^{(m)}$ ($m = 1, 2, 3 \dots M$), to be made during a traversal.

The minimum variance technique is designed to deal with the situation where some or all of the non-ideal effects mentioned above, except a systematic temporal change in the normal direction, $\hat{\mathbf{n}}$, are present. As the estimate of $\hat{\mathbf{n}}$, the method identifies that direction in space along which the field-component set $\{\mathbf{B}^{(m)} \cdot \hat{\mathbf{n}}\}$ ($m = 1, 2, 3 \dots M$) has minimum variance. In other words, $\hat{\mathbf{n}}$ is determined by minimisation of ⇒ 8.2

$$\sigma^2 = \frac{1}{M} \sum_{m=1}^M \left| (\mathbf{B}^{(m)} - \langle \mathbf{B} \rangle) \cdot \hat{\mathbf{n}} \right|^2 \quad (8.4)$$

where the average $\langle \mathbf{B} \rangle$ is defined by

$$\langle \mathbf{B} \rangle \equiv \frac{1}{M} \sum_{m=1}^M \mathbf{B}^{(m)} \quad (8.5)$$

and where the minimisation is subject to the normalisation constraint $|\hat{\mathbf{n}}|^2 = 1$. Using a Lagrange multiplier, λ , to implement this constraint, one then seeks the solution of the set of three homogeneous linear equations

$$\begin{aligned} \frac{\partial}{\partial n_X} \left(\sigma^2 - \lambda (|\hat{\mathbf{n}}|^2 - 1) \right) &= 0 \\ \frac{\partial}{\partial n_Y} \left(\sigma^2 - \lambda (|\hat{\mathbf{n}}|^2 - 1) \right) &= 0 \\ \frac{\partial}{\partial n_Z} \left(\sigma^2 - \lambda (|\hat{\mathbf{n}}|^2 - 1) \right) &= 0 \end{aligned} \quad (8.6)$$

where σ^2 is given by equation 8.4 and $\hat{\mathbf{n}}$ is represented in terms of its three components (n_X, n_Y, n_Z) along the cartesian coordinate system X, Y, Z (e.g., GSE or GSM) in which the field data $\{\mathbf{B}^{(m)}\}$ are given. When the differentiations in equation 8.6 have been performed, the resulting set of three equations can be written in matrix form as

$$\sum_{v=1}^3 M_{\mu v}^B n_v = \lambda n_\mu \quad (8.7)$$

where the subscripts $\mu, v = 1, 2, 3$ denote cartesian components along the X, Y, Z system and

$$M_{\mu v}^B \equiv \langle B_\mu B_v \rangle - \langle B_\mu \rangle \langle B_v \rangle \quad (8.8)$$

is the magnetic variance matrix. It is seen from equation 8.7 that the allowed λ values are the eigenvalues $\lambda_1, \lambda_2, \lambda_3$ (given here in order of decreasing magnitude) of $M_{\mu v}^B$. Since $M_{\mu v}^B$ is symmetric, the eigenvalues are all real and the corresponding eigenvectors, $\mathbf{x}_1, \mathbf{x}_2$, and \mathbf{x}_3 , are orthogonal. The three eigenvectors represent the directions of maximum,

intermediate, and minimum variance of the field component along each vector. Note that the sense and magnitude of the eigenvectors remain arbitrary so that, for example, \mathbf{x}_i , $k\mathbf{x}_i$, $-\mathbf{x}_i$, and $-k\mathbf{x}_i$ ($i = 1, 2, 3$) all are valid eigenvectors. The corresponding λ values represent the actual variances in those field components and are therefore non-negative. This point becomes clear by writing the matrix $M_{\mu\nu}^B$ in the eigenvector basis $(\mathbf{x}_1, \mathbf{x}_2, \mathbf{x}_3)$ where it is diagonal with diagonal terms given by

$$M_{ii}^B = \langle B_i B_i \rangle - \langle B_i \rangle \langle B_i \rangle = \lambda_i \quad (8.9)$$

In summary, the minimum variance analysis consists of constructing the matrix $M_{\mu\nu}^B$, defined by equation 8.8 in terms of the measured field data and the cartesian coordinate system in which the measured data are represented, and then finding the three eigenvalues λ_i , and corresponding eigenvectors \mathbf{x}_i , of the matrix. The eigenvector \mathbf{x}_3 corresponding to the smallest eigenvalue, λ_3 , is used as the estimator for the vector normal to the current sheet and λ_3 itself represents the variance of the magnetic field component along the estimated normal. The eigenvectors \mathbf{x}_1 and \mathbf{x}_2 , corresponding to maximum and intermediate variance, are then tangential to the transition layer and the set $\{\mathbf{x}_1, \mathbf{x}_2, \mathbf{x}_3\}$ arranged as a right-handed orthonormal triad provides suitable basis vectors for the local coordinates (x, y, z) discussed in connection with equation 8.1. More generally, for any measured set of vectors $\{\mathbf{B}^{(m)}\}$, not necessarily obtained from a spacecraft traversal of a transition layer or wave front, the eigenvector set of the variance matrix $M_{\mu\nu}^B$ derived from the data provides a convenient natural coordinate system in which to display and analyse the data. Note also that the matrix $M_{\mu\nu}^B$ is independent of the temporal order of the measured vectors.

If the data set used for the minimum variance calculation has its minimal size, i.e., if it consists of exactly three vectors, as in Figure 8.1, then one finds $\lambda_3 = 0$. For the case in Figure 8.1a one would also find $\lambda_2 = 0$, $\lambda_1 \neq 0$, and \mathbf{x}_1 would be parallel to ABC. In this case the variance matrix $M_{\mu\nu}^B$ is said to be degenerate and all that can be said about \mathbf{x}_2 and \mathbf{x}_3 is that they are perpendicular to ABC. This situation is discussed in detail in Section 8.2.5. In Figure 8.1b, one would find $\lambda_2 \neq 0$, $\lambda_1 \neq 0$ and the triad $(\mathbf{x}_1, \mathbf{x}_2, \mathbf{x}_3)$ would be oriented as shown qualitatively in the figure.

8.2.3 Hodogram Representation

The magnetic hodograph is a curve in space constructed by drawing vectors from the origin, the lengths and directions of which represent the members of the measured set $\{\mathbf{B}^{(m)}\}$, and then connecting the arrowheads of those vectors by line segments, following the time sequence in which they were measured. It has become common practice to display the hodograph in two projections, called magnetic hodograms. The projection onto a plane tangential to the layer is a plot of the component $B_1 = \mathbf{B}^{(m)} \cdot \mathbf{x}_1$ versus $B_2 = \mathbf{B}^{(m)} \cdot \mathbf{x}_2$ and the side view projection is a plot of B_1 versus $B_3 = \mathbf{B}^{(m)} \cdot \mathbf{x}_3$, i.e., versus the normal field component. Such a hodogram pair is shown in Figure 8.2 for a particular AMPTE/IRM magnetopause crossing that will serve as our test case in this chapter and also in Chapter 9. The data given below the plots consist of the eigenvalues $(\lambda_1, \lambda_2, \lambda_3)$, the eigenvector components (x_{iX}, x_{iY}, x_{iZ}) , $i = 1, 2, 3$, along the original (GSE) coordinate system, and the average field components $(\langle B_1 \rangle, \langle B_2 \rangle, \langle B_3 \rangle)$ along $(\mathbf{x}_1, \mathbf{x}_2, \mathbf{x}_3)$. It is seen that the normal field component for this event fluctuates with standard deviation $\sqrt{\lambda_3} = \sqrt{7.08} =$

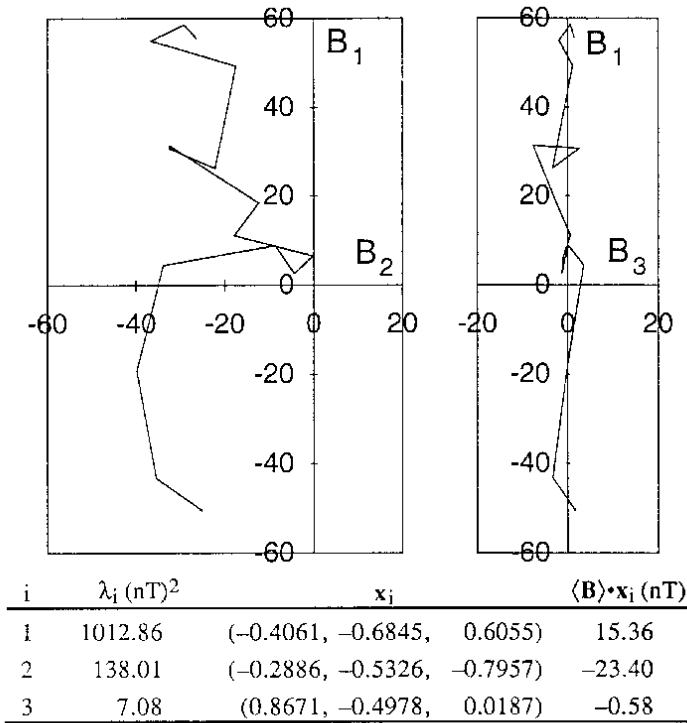


Figure 8.2: Hodogram pair for outbound magnetopause traversal by AMPTE/IRM on October 19, 1984, 05:18:20–05:19:26 UT. Units on axes are nT and high-resolution data have been averaged over spacecraft spin period (~ 4.35 s). Eigenvalues $\lambda_i \text{ (nT)}^2$, eigenvectors \mathbf{x}_i (GSE) and average field components $\langle \mathbf{B} \rangle \cdot \mathbf{x}_i \text{ (nT)}$ are given in the order of decreasing λ value: \mathbf{x}_3 serves as estimator for $\hat{\mathbf{n}}$.

2.66 nT around an average value that is near zero ($\langle B_3 \rangle = -0.58 \text{ nT}$) so that the current layer appears to be a tangential discontinuity (TD), albeit one having substantial 2-D or 3-D internal substructures which produce the fluctuations in the normal component. These substructures have been analysed in detail by *Sonnerup and Guo [1996]*. As discussed in Chapter 9, there are also other reasons to believe that our test case is indeed a TD.

Illustrations of the convenience of using of the eigenvector basis for analytical purposes are found in Sections 8.2.4, 8.2.6, and 8.3.1.

In the application to the dayside magnetopause given here, we have chosen \mathbf{x}_1 to correspond to the maximum variance eigenvector and \mathbf{x}_3 to the minimum variance eigenvector, directed northward and outward from the Earth, respectively. This choice is convenient because the ordered set $\{\mathbf{x}_1, \mathbf{x}_2, \mathbf{x}_3\}$ is then similar to the boundary normal coordinate axes, ordered as (L, M, N) , which were introduced by *Russell and Elphic [1979]*. However, the choice is not unique: if \mathbf{x}_1 and \mathbf{x}_3 are interchanged and if $\mathbf{x}_2 \rightarrow -\mathbf{x}_2$, the resulting eigenvector basis corresponds qualitatively to the GSE or GSM systems (X, Y, Z) for

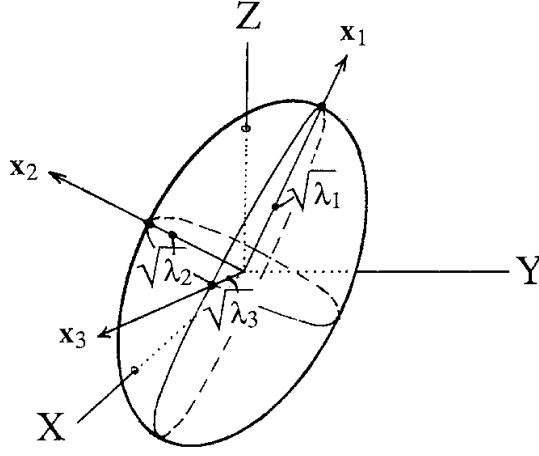


Figure 8.3: The variance ellipsoid. Eigenvectors, \mathbf{x}_1 , \mathbf{x}_2 , and \mathbf{x}_3 , are shown relative to the system (X, Y, Z) (e.g., GSE) in which the magnetic data are given. The normal direction is estimated to be along \mathbf{x}_3 .

magnetopause crossings in the subsolar region.

8.2.4 Variance Ellipsoid

The variance, σ^2 , of the magnetic field component along an arbitrarily chosen direction, defined by the unit vector $\hat{\mathbf{k}}$, say, can be written in terms of the variance matrix as

$$\sigma^2 = \sum_{\mu\nu} k_\mu M_{\mu\nu}^B k_\nu \quad (8.10)$$

This result follows from equation 8.4 with $\hat{\mathbf{n}}$ replaced by $\hat{\mathbf{k}}$. If the expression 8.10 is transformed to the eigenvector basis, it becomes

$$\sigma^2 = \lambda_1 k_1^2 + \lambda_2 k_2^2 + \lambda_3 k_3^2 \quad (8.11)$$

where $k_i = \hat{\mathbf{k}} \cdot \mathbf{x}_i$, $i = 1, 2, 3$. Equation 8.11 invites the definition of a “variance space” in which the coordinates along the eigenvectors \mathbf{x}_1 , \mathbf{x}_2 , and \mathbf{x}_3 are $\sigma_i = \sqrt{\lambda_i} k_i$, $i = 1, 2, 3$, and in which the variance, $\sigma^2 = \sigma_1^2 + \sigma_2^2 + \sigma_3^2$, is the distance from the origin. In terms of these coordinates the normalisation condition, $|\hat{\mathbf{k}}|^2 = k_1^2 + k_2^2 + k_3^2 = 1$, becomes

$$\frac{\sigma_1^2}{\lambda_1} + \frac{\sigma_2^2}{\lambda_2} + \frac{\sigma_3^2}{\lambda_3} = 1 \quad (8.12)$$

This expression defines the variance ellipsoid, as shown in Figure 8.3, the principal-axes half lengths of which are $\sqrt{\lambda_1}$, $\sqrt{\lambda_2}$, and $\sqrt{\lambda_3}$. The distance along an arbitrarily chosen direction in variance space, from the origin to the intersection of a radial line along that direction with the ellipsoid surface, represents the standard deviation of the magnetic field

component along the chosen direction. However, note that because the transformation from physical space to variance space involves unequal stretching along the three axes ($\sigma_i = \sqrt{\lambda_i} k_i$), directions in the two spaces do not agree, except for the principal directions. If $\hat{\mathbf{k}}$ is chosen along one of the principal axes, $\hat{\mathbf{k}} = \mathbf{x}_i$, then $\sigma^2 = \lambda_i$ as expected.

8.2.5 Degeneracy

When the three eigenvalues of the variance matrix are distinct, the matrix and the variance ellipsoid it represents are said to be non-degenerate. This is the most common situation in practice but there is also a significant number of cases where near degeneracy occurs. Three types of degeneracy are possible: $\lambda_1 \simeq \lambda_2$; $\lambda_2 \simeq \lambda_3$; and $\lambda_1 \simeq \lambda_2 \simeq \lambda_3$. The first of these cases, $\lambda_1 \simeq \lambda_2$, corresponds to a discus-shaped (oblate) variance ellipsoid; the minimum variance direction, \mathbf{x}_3 , which is along the axis of the discus, remains well determined but any pair of vectors perpendicular to \mathbf{x}_3 , i.e., any vectors lying in the equatorial plane of the discus, may serve as \mathbf{x}_1 and \mathbf{x}_2 . This degeneracy, therefore, does not limit the utility of MVAB for normal-vector and normal-field-component determinations, provided $\lambda_3 \ll \lambda_2 \simeq \lambda_1$.

The second type of degeneracy, $\lambda_2 \simeq \lambda_3$, corresponds to a cigar-shaped (prolate) variance ellipsoid with the axis of the cigar along \mathbf{x}_1 and with the intermediate and minimum variance directions, \mathbf{x}_2 and \mathbf{x}_3 , being constrained to be perpendicular to \mathbf{x}_1 but otherwise arbitrary. In this case, no valid direction normal to the layer is obtained from MVAB although \mathbf{x}_1 remains a good vector tangential to the layer, provided $\lambda_1 \gg \lambda_2 \simeq \lambda_3$. Near degeneracy of this type has been found to be rather common in analyses of magnetopause and geotail current layer crossings. A physical situation where it will occur is when the electric current in the layer is unidirectional or nearly unidirectional so that not only the normal ($B_1 B_3$) hodogram, but also the tangential ($B_1 B_2$) hodogram is a vertical line ($B_3 = \text{const.}$, $B_2 = \text{const.}$) or approximately a vertical line, i.e., when $\lambda_2 \simeq \lambda_3 \simeq 0$. This situation is the generalisation to many measured vectors of the degenerate case of three measured vectors with a common tangential component, shown in Figure 8.1a, where equation 8.2 fails to provide a normal vector. By contrast, Figure 8.1b corresponds to the non-degenerate case where $\lambda_1 > \lambda_2 > \lambda_3 = 0$ (or in an exceptional case to $\lambda_1 = \lambda_2 > \lambda_3 = 0$). In practice, near degeneracy with $\lambda_2 \simeq \lambda_3 \neq 0$ occurs more frequently than with $\lambda_2 \simeq \lambda_3 \simeq 0$.

The third case of degeneracy $\lambda_1 \simeq \lambda_2 \simeq \lambda_3$ corresponds to a spherical or nearly spherical variance ellipsoid. In this case no information about directions normal or tangential to a layer is obtained. If $\lambda_1 \simeq \lambda_2 \simeq \lambda_3 \simeq 0$, all measured magnetic field vectors are equal while for $\lambda_1 \simeq \lambda_2 \simeq \lambda_3 \neq 0$ the measured field consists of a uniform component (which could be zero) plus magnetic fluctuations having isotropic variance.

When two eigenvalues, λ_2 and λ_3 say, are nearly the same, the uncertainty in the corresponding eigenvectors is large with respect to rotation about the remaining eigenvector, \mathbf{x}_1 . In extreme cases, the orientation of the intermediate and minimum variance directions may trade places, corresponding approximately to a 90° rotation of \mathbf{x}_2 and \mathbf{x}_3 around \mathbf{x}_1 , in response to a minor adjustment of the number of data points or the data filtering used in the analysis. Such behaviour seriously or completely compromises the use of \mathbf{x}_3 as a predictor of the direction normal to the layer. A case in point is a layer of nearly unidirectional current which contains a string of tearing mode islands. In this situation both λ_2 and λ_3 are small but non-zero. Depending on where, and at what angle, the spacecraft trajectory intersects the layer and on how the data interval used for the MVAB is selected,

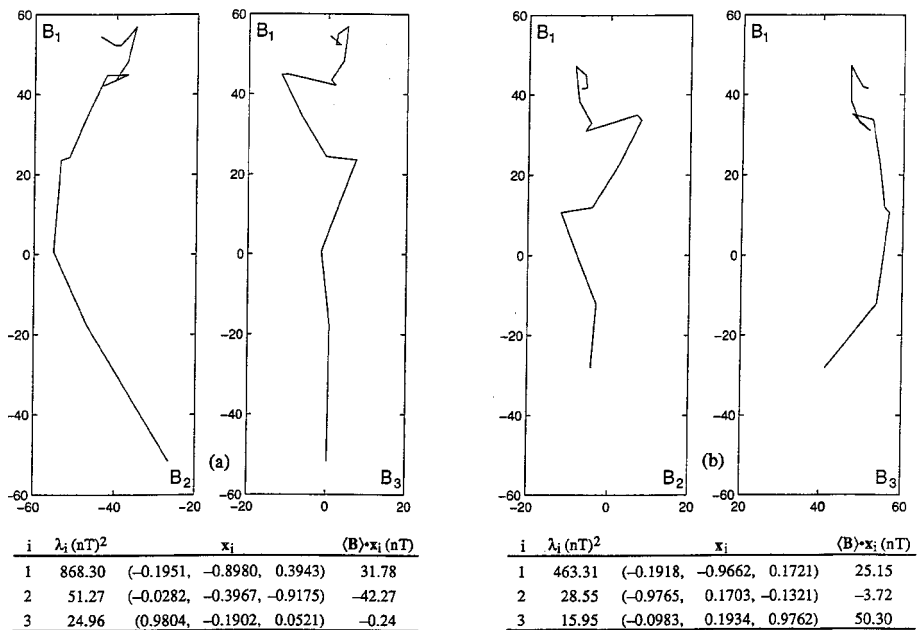


Figure 8.4: Eigenvector flip. Eigenvectors \mathbf{x}_2 and \mathbf{x}_3 are interchanged, approximately, when 1 data point is deleted (in right hodogram pair) at each end of the original data interval (shown in left hodogram pair). Data are from magnetopause traversal by AMPTE/IRM on September 25, 1984, 05:58:21–05:59:44 UT.

the minimum variance direction, \mathbf{x}_3 , may lie either along the normal to the layer or along the average current vector which is tangent to the layer. An example of actual eigenvector flipping for a noisy magnetopause traversal is shown in Figure 8.4. We see that the eigenvalue ratio in part (a) of the figure is substantial: $\lambda_2/\lambda_3 = 2.05$. Nevertheless, removal of only one point at each end of the data interval leads to a flip and to the result in part (b).

In summary, near degeneracy should result in large error estimates for the corresponding eigenvectors and for the field components along those eigenvectors. Quantitative error estimates having this property will be provided in Section 8.3.

8.2.6 Constraint $\langle \mathbf{B}_3 \rangle = 0$

In some circumstances, information may be available from other measurements or from theory to indicate that the current layer should be a tangential discontinuity, i.e., that the average field component, $\langle \mathbf{B} \rangle \cdot \hat{\mathbf{n}}$, along its normal should be zero. In particular for the degenerate case $\lambda_2 \simeq \lambda_3$, it may be necessary to implement a constraint of this type in order to obtain a useful normal vector prediction. The method often used is to calculate the average of a set of field vectors measured on one side of the layer and the average of another set measured on the opposite side and then use the cross product of the resulting two vectors as the predictor of the normal vector, $\hat{\mathbf{n}}$. However, there is an alternate approach [Sonnerup and Cahill, 1968] in which all field vectors measured within the layer

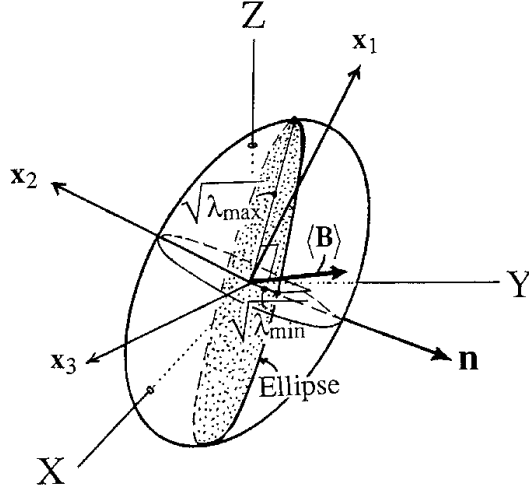


Figure 8.5: Intersection of variance ellipsoid with the plane $\langle \mathbf{B} \rangle \cdot \hat{\mathbf{n}} = 0$ (shown shaded). The normal direction is estimated to be along the minor axis $\left(\lambda_{\min}^{1/2}\right)$ of the ellipse marking the intersection.

as well as vectors measured on its two sides are utilised: $\hat{\mathbf{n}}$ is chosen such that the variance of the field component along it is a minimum, subject to the constraint $\langle \mathbf{B} \rangle \cdot \hat{\mathbf{n}} = 0$. Geometrically, this condition requires $\hat{\mathbf{n}}$ to lie in a plane perpendicular to $\langle \mathbf{B} \rangle$ (with the directions of $\hat{\mathbf{n}}$ and $\langle \mathbf{B} \rangle$ converted to corresponding directions in variance space). The intersection of this plane with the variance ellipsoid is an ellipse, as shown in Figure 8.5, the minor axis of which is used as the predictor for $\hat{\mathbf{n}}$.

The quantitative analysis is carried out most conveniently by minimising the expression $[\sigma^2 - \lambda(|\hat{\mathbf{n}}|^2 - 1) - 2\Gamma\langle \mathbf{B} \rangle \cdot \hat{\mathbf{n}}]$ where λ and 2Γ are Lagrange multipliers used to implement the constraints $|\hat{\mathbf{n}}|^2 = 1$ and $\langle \mathbf{B} \rangle \cdot \hat{\mathbf{n}} = 0$, respectively. Partial differentiation with respect to each of the three components of $\hat{\mathbf{n}}$ then leads to the following set of three linear non-homogeneous equations

$$\sum_{v=1}^3 M_{\mu v}^B \cdot n_v - \lambda n_{\mu} = \Gamma \langle B_{\mu} \rangle \quad \mu = 1, 2, 3 \quad (8.13)$$

where $M_{\mu v}^B$ is given by equation 8.8. Using the eigenvectors of $M_{\mu v}^B$ as basis vectors and assuming $\lambda \neq \lambda_i$, we find

$$n_i = \Gamma \langle B_i \rangle / (\lambda_i - \lambda) \quad i = 1, 2, 3 \quad (8.14)$$

where (n_1, n_2, n_3) and $(\langle B_1 \rangle, \langle B_2 \rangle, \langle B_3 \rangle)$ are the components of $\hat{\mathbf{n}}$ and $\langle \mathbf{B} \rangle$ along the maximum, intermediate, and minimum variance eigenvectors, respectively. The normalisation

constraint, $|\hat{\mathbf{n}}|^2 = 1$, then gives

$$\Gamma = \pm \left[\sum_{i=1}^3 \langle B_i \rangle^2 / (\lambda_i - \lambda)^2 \right]^{-1/2} \quad (8.15)$$

and, using equation 8.14, the constraint $\langle \mathbf{B} \rangle \cdot \hat{\mathbf{n}} = 0$ gives

$$\sum_{i=1}^3 \langle B_i \rangle^2 / (\lambda_i - \lambda) = 0 \quad (8.16)$$

which is a quadratic equation for the Lagrange multiplier λ . The two roots of this equation are denoted by λ_{\max} and λ_{\min} . The major and minor semiaxes of the ellipse of intersection shown in Figure 8.5 are $\sqrt{\lambda_{\max}}$ and $\sqrt{\lambda_{\min}}$, respectively, so that λ_{\max} and λ_{\min} represent the maximum and minimum values of the variance of the set $\{\mathbf{B}^{(m)} \cdot \hat{\mathbf{n}}\}$ when the constraint $\langle \mathbf{B} \rangle \cdot \hat{\mathbf{n}} = 0$ has been implemented. The sought-after normal vector, $\hat{\mathbf{n}}$, is obtained by substituting λ_{\min} into equations 8.15 and 8.14; the desired sense of $\hat{\mathbf{n}}$ is obtained by properly selecting the sign of Γ in equation 8.15.

In a more general formulation [A. V. Khrabrov, private communication], one can show that a constraint of the form $\hat{\mathbf{n}} \cdot \hat{\mathbf{e}} = 0$, where $\hat{\mathbf{e}}$ is a known unit vector, leads to the eigenvalue problem $\mathbf{P} \cdot \mathbf{M}^B \cdot \mathbf{P} \cdot \hat{\mathbf{n}} = \lambda \hat{\mathbf{n}}$. Here \mathbf{P} is the matrix describing projection of a vector onto the plane perpendicular to $\hat{\mathbf{e}}$, i.e., $P_{ij} = \delta_{ij} - e_i e_j$. By putting $\hat{\mathbf{n}} = \hat{\mathbf{e}}$ in the eigenvalue equation, it is then seen that $\hat{\mathbf{e}}$ is an eigenvector corresponding to $\lambda = 0$. In the application above, the other two eigenvalues are the same as λ_{\min} and λ_{\max} from equation 8.16 with their corresponding eigenvectors being the same as those from equation 8.14.

An alternative approach to the problem of finding the vector normal to a tangential discontinuity was described by *Siscoe et al.* [1968] who minimised the sum of the squares of the individual normal field components, subject to the constraint $|\hat{\mathbf{n}}|^2 = 1$. In other words, they extremised

$$\frac{1}{M} \sum_{m=1}^M \left(\mathbf{B}^{(m)} \cdot \hat{\mathbf{n}} \right)^2 - \lambda \left(|\hat{\mathbf{n}}|^2 - 1 \right) \quad (8.17)$$

which leads to the problem of finding the eigenvalues and eigenvectors of the matrix

$$\tilde{M}_{\mu\nu}^B = \langle B_\mu B_\nu \rangle$$

Table 8.1 contains a comparison, for the AMPTE/IRM event in Figure 8.2, of normal vectors and normal field components obtained (i) from minimum variance analysis without constraint; (ii) from minimum variance analysis with constraint $\langle \mathbf{B} \rangle \cdot \hat{\mathbf{n}} = 0$; and (iii) from minimisation of $\langle (\mathbf{B} \cdot \hat{\mathbf{n}})^2 \rangle$. It is seen that, in this particular case where the actual normal field component was very small or zero, all three calculations give very similar results, the maximum angular deviation being less than 1.4° . In cases where the actual normal-field component is substantial, the results from the three methods differ significantly and only method (i) is appropriate.

The method (iii) developed by *Siscoe et al.* has an alternate important application that will be illustrated later on. It can be used to characterise a set of nearly aligned normal-vector estimates in terms of their average direction (*Siscoe et al.* eigenvector corresponding to the maximum eigenvalue) and the elliptical cross-section of their error cone (*Siscoe et al.* eigenvectors and square roots of intermediate and minimum eigenvalues).

Table 8.1: Predictions of normal vector $\hat{\mathbf{n}}$ for magnetopause of tangential-discontinuity type observed by AMPTE/IRM on October 19, 1984 (05:18:20–05:19:26 UT).

	Method	n_x	n_y	n_z	$\langle \mathbf{B} \cdot \hat{\mathbf{n}} \rangle$ (nT)
(i):	Min. variance (MVAB)	0.8671	-0.4978	0.0187	-0.58
(ii):	MVAB + $\langle \mathbf{B} \cdot \hat{\mathbf{n}} \rangle \equiv 0$	0.8728	-0.4865	0.0386	0.00
(iii):	Min. $\langle \mathbf{B} \cdot \hat{\mathbf{n}} \rangle^2$	0.8718	-0.4887	0.0349	-0.11

8.3 Error Estimates

Uncertainties in the orientations of the eigenvectors ($\mathbf{x}_1, \mathbf{x}_2, \mathbf{x}_3$) and in the values of the average field components $\langle \mathbf{B} \rangle \cdot \mathbf{x}_i$ can be either purely statistical in nature or can be caused by lack of stationarity or quasi-one-dimensionality of the structure being studied. The statistical errors can be estimated by use of standard analytical tools or by use of computational techniques such as the bootstrap method. These approaches are illustrated and discussed below. A comparison of normal vectors and of errors is provided in Section 8.5. Uncertainties associated with lack of stationarity or other systematic effects are more difficult to assess but can be addressed, at least in part, by use of nested data segments: this approach is also presented.

8.3.1 Analytical Estimates of Statistical Errors

As mentioned already, it is difficult to estimate the actual error in the normal vector obtained from MVA because, in reality, the error may have a systematic part in addition to a part caused by finite sampling of a stationary noise component in the measured field vectors $\mathbf{B}^{(m)}$. In what follows we assume that no systematic errors are present: we are therefore concerned only with the noise component for which it is expected that the error should decrease as $M^{-1/2}$ as the number of measured field vectors, M , used in the analysis, increases. Following a recent development by *Khrabrov and Sonnerup [1998a]* (which was motivated by the preparation of this chapter), the uncertainties in the directions of the eigenvectors ($\mathbf{x}_1, \mathbf{x}_2, \mathbf{x}_3$) of the variance matrix, $\mathbf{M}^B \equiv \mathbf{M}$ (the superscript B is suppressed for brevity), calculated from the measured field vectors, are estimated by performing perturbation analysis on the eigenvector equation 8.7 around the unknown noise-free state which is denoted by an asterisk:

$$(\mathbf{M}^* + \Delta \mathbf{M}) \cdot (\mathbf{x}_i^* + \Delta \mathbf{x}_i) = (\lambda_i^* + \Delta \lambda_i)(\mathbf{x}_i^* + \Delta \mathbf{x}_i)$$

Here $i = 1, 2, 3$, correspond to maximum, intermediate, and minimum variance associated with \mathbf{M}^* , respectively. The linearised version of this equation becomes

$$\Delta \mathbf{M} \cdot \mathbf{x}_i^* + \mathbf{M}^* \cdot \Delta \mathbf{x}_i = \Delta \lambda_i \mathbf{x}_i^* + \lambda_i^* \Delta \mathbf{x}_i \quad (8.18)$$

Equation 8.18 is now written in the unperturbed eigenbasis in which \mathbf{M}^* is diagonal. Using subscript notation (but not the summation convention), the j th component of equation 8.18 becomes, after simple rearrangements,

$$(\lambda_j^* - \lambda_i^*) \Delta x_{ij} = -\Delta M_{ij} - \Delta \lambda_i \delta_{ij} \quad (8.19)$$

where Δx_{ij} is the j^{th} component of the vector $\Delta \mathbf{x}_i$ and where use has been made of $x_{ij}^* = \delta_{ij}$. Since \mathbf{M} and \mathbf{M}^* , are both symmetric matrices, we have $\Delta M_{ij} = \Delta M_{ji}$. From equation 8.19, one then concludes that $\Delta x_{ij} = -\Delta x_{ji}$ which expresses the fact that the perturbed eigenvectors must form an orthonormal triad. The perturbed eigenvectors must also retain unit length so that in the linear approximation $\Delta x_{ii} = \mathbf{x}_i^* \cdot \Delta \mathbf{x}_i = 0$. For $j = i$, equation 8.19 gives $\Delta \lambda_i = \Delta M_{ii}$ while the other components ($j \neq i$) give

$$\begin{aligned}\Delta x_{31} = -\Delta x_{13} &= -\Delta M_{13}/(\lambda_1^* - \lambda_3^*) \\ \Delta x_{32} = -\Delta x_{23} &= -\Delta M_{23}/(\lambda_2^* - \lambda_3^*) \\ \Delta x_{21} = -\Delta x_{12} &= -\Delta M_{21}/(\lambda_1^* - \lambda_2^*)\end{aligned}\quad (8.20)$$

In the linear approximation, the quantities Δx_{31} and Δx_{32} also represent the angular rotations (in radians) of the eigenvector \mathbf{x}_3 towards \mathbf{x}_1 and \mathbf{x}_2 , respectively. Similarly, Δx_{21} represents the angular rotation of \mathbf{x}_2 towards \mathbf{x}_1 . Since the noise-free state represents an ideal one-dimensional current layer of fixed orientation we have $\lambda_3^* = 0$ in equation 8.20.

The next step is to evaluate the ensemble averages, denoted by the double bracket $\langle\langle \dots \rangle\rangle$, of $(\Delta x_{31})^2$, $(\Delta x_{32})^2$ and $(\Delta x_{21})^2$. As is seen from equation 8.20 these averages are proportional to $\langle\langle \Delta M_{13}^2 \rangle\rangle$, $\langle\langle \Delta M_{23}^2 \rangle\rangle$ and $\langle\langle \Delta M_{21}^2 \rangle\rangle$, respectively. By ensemble average we mean the average over a large number of realisations of the noise component of the measured field. In reality, only one such realisation is available to us, namely the one contained in the measured set of vectors, $\{\mathbf{B}^{(m)}\}$. Furthermore, for this set we do not know *a priori* what part of $\mathbf{B}^{(m)}$ is noise. Nevertheless, we can obtain a formal expression for ΔM_{ij} by replacing $\mathbf{B}^{(m)}$ by $(\mathbf{B}^{(m)*} + \Delta \mathbf{B}^{(m)})$ in the definition 8.8 of \mathbf{M} , where $\Delta \mathbf{B}^{(m)}$ is the noise component. The noise is assumed to be stationary, isotropic, and spatially uncorrelated (the final error formulas may in fact work under less restrictive conditions) and, as a result, can be shown to have the following properties:

$$\begin{aligned}\langle\langle \Delta B_i^{(m)} \rangle\rangle &= 0 \quad i = 1, 2, 3 \\ \langle\langle \Delta B_i^{(m)} \Delta B_j^{(n)} \rangle\rangle &= \delta_{ij} \delta_{mn} \langle\langle (\Delta B_3^{(m)})^2 \rangle\rangle \\ \langle\langle \Delta B_i^{(m)} \Delta B_j^{(n)} \Delta B_k^{(p)} \rangle\rangle &= 0 \\ \langle\langle \Delta B_i^{(m)} \Delta B_j^{(n)} \Delta B_k^{(p)} \Delta B_l^{(q)} \rangle\rangle &= \left(\delta_{ij}^{mn} \delta_{kl}^{pq} + \delta_{ik}^{mp} \delta_{jl}^{nq} + \delta_{il}^{mq} \delta_{jk}^{np} \right) \cdot \langle\langle (\Delta B_3^{(m)})^2 \rangle\rangle^2\end{aligned}\quad (8.21)$$

In the fourth property, we have used the notation $\delta_{ij}^{mn} = \delta_{ij} \delta_{mn}$; we have also assumed the noise to be normally distributed. The second property indicates that the variance of each of the three vector components of the noise has been assumed to be the same (isotropy) and that it is represented by the variance of the component along \mathbf{x}_3^* . Furthermore, the assumption of time stationarity means that $\langle\langle (\Delta B_3^{(m)})^2 \rangle\rangle$ is, in fact, independent of m . As shown by *Khrabrov and Sonnerup* [1998a], it is also equal to $\langle\langle \lambda_3 \rangle\rangle M / (M - 1)$; similarly, they show that $\langle\langle \lambda_2 \rangle\rangle = \lambda_2^* + \langle\langle \lambda_3 \rangle\rangle$ and that $\langle\langle \lambda_1 \rangle\rangle = \lambda_1^* + \langle\langle \lambda_3 \rangle\rangle$. Although the ensemble averages, $\langle\langle \lambda_1 \rangle\rangle$, $\langle\langle \lambda_2 \rangle\rangle$, and $\langle\langle \lambda_3 \rangle\rangle$, deviate somewhat from the corresponding eigenvalues, λ_1 , λ_2 , and λ_3 , calculated from the actually measured field vectors, $\mathbf{B}^{(m)}$, we will use the set $(\lambda_1, \lambda_2, \lambda_3)$ to replace the ensemble averages in the error estimates. The errors associated with this replacement can be estimated by evaluating $\langle\langle (\Delta \lambda_i)^2 \rangle\rangle = \langle\langle (\Delta M_{ii})^2 \rangle\rangle$ for $i = 1, 2, 3$.

We may finally express the perturbed matrix elements, ΔM_{ij} , in terms of $\Delta \mathbf{B}^{(m)}$ by use of the definition of \mathbf{M} , then square ΔM_{ij} , perform the ensemble average and do the sums

over m . In this development the properties 8.21 are used. The detailed steps [Khrabrov and Sonnerup, 1998a] are straight-forward but too lengthy to be given here. Neglecting terms of order $\varepsilon^2 \equiv [\lambda_3/(\lambda_2 - \lambda_3)]^2/(M - 1)^2$ compared to unity, the result is

$$\langle\langle(\Delta M_{ij})^2\rangle\rangle = \langle\langle\lambda_3\rangle\rangle [\langle\langle\lambda_i\rangle\rangle + \langle\langle\lambda_j\rangle\rangle - \langle\langle\lambda_3\rangle\rangle] (1 + \delta_{ij})/(M - 1) \quad (8.22)$$

so that the angular error estimates (in radians) become

$$\begin{aligned} |\Delta\varphi_{ij}| = |\Delta\varphi_{ji}| &= \langle\langle(\Delta x_{ij})^2\rangle\rangle^{1/2} = \langle\langle(\Delta x_{ji})^2\rangle\rangle^{1/2} \\ &= \sqrt{\frac{\lambda_3}{(M - 1)} \frac{(\lambda_i + \lambda_j - \lambda_3)}{(\lambda_i - \lambda_j)^2}}, \quad i \neq j \end{aligned} \quad (8.23)$$

Here $|\Delta\varphi_{ij}|$ denotes the expected angular uncertainty of eigenvector \mathbf{x}_i for rotation toward or away from eigenvector \mathbf{x}_j . It is noted that, except for very small values of $\lambda_3/(M - 1)$, the uncertainty becomes large for the nearly degenerate case, $\lambda_i \simeq \lambda_j$, discussed in Section 8.2.5. This is an expected and desirable property but it must be remembered that the linear analysis used in producing the error estimate then breaks down. Such cases are nevertheless of interest. For example, if $\lambda_1 \simeq \lambda_2$ then \mathbf{x}_3 remains a good normal vector provided $|\Delta\varphi_{31}|$ and $|\Delta\varphi_{32}|$ are small. And if $\lambda_3 \simeq \lambda_2 \ll \lambda_1$ then \mathbf{x}_1 remains a good tangent vector to the current sheet provided $|\Delta\varphi_{12}|$ and $|\Delta\varphi_{13}|$ are small. An important use of this tangent vector is discussed in Sections 8.4.1 and 8.4.2. Another situation where $\lambda_2 \simeq \lambda_3$ is likely to occur is in maximum variance analysis of electric field data (MVAE). In this application, the maximum variance direction, \mathbf{x}_1 , is normal to the current layer, as discussed further in Section 8.4.2. In reality, the ordering $\lambda_3 < \lambda_2 \ll \lambda_1$, or even $\lambda_3 \ll \lambda_2 \ll \lambda_1$, is often found in MVAE, indicating a lack of isotropy of the noise (defined as the deviations from a 1-D time-independent structure); in that case the error estimate $|\Delta\varphi_{12}|$ should be calculated with λ_2 replacing λ_3 in equation 8.23. Additionally it is noted that non-isotropic noise may lead to a bias (a systematic error) in the normal vector estimate. ⇒8.3

The statistical uncertainty in the component of the average magnetic field along the eigenvector \mathbf{x}_3 is composed of three parts: the uncertainty in the average associated with the corresponding variance λ_3 and the two uncertainties associated with the angular error estimates for \mathbf{x}_3 . Assuming these errors to be independent, we can then write the composite statistical error estimate for $\langle \mathbf{B} \rangle \cdot \mathbf{x}_3$ as

$$|\Delta\langle \mathbf{B} \cdot \mathbf{x}_3 \rangle| = \sqrt{\frac{\lambda_3}{M - 1} + (\Delta\varphi_{32}\langle \mathbf{B} \rangle \cdot \mathbf{x}_2)^2 + (\Delta\varphi_{31}\langle \mathbf{B} \rangle \cdot \mathbf{x}_1)^2} \quad (8.24)$$

Similar expressions can be written for the uncertainties in $\langle \mathbf{B} \rangle \cdot \mathbf{x}_1$ and in $\langle \mathbf{B} \rangle \cdot \mathbf{x}_2$ but these error estimates are usually of less interest.

As an example, the error estimates for the magnetopause crossing depicted in Figure 8.2 have been calculated from equations 8.23 and 8.24, with the following results: $\Delta\varphi_{32} = \Delta\varphi_{23} = \pm 0.062$ rad = $\pm 3.6^\circ$; $\Delta\varphi_{31} = \Delta\varphi_{13} = \pm 0.022$ rad = $\pm 1.3^\circ$; $\Delta\varphi_{12} = \Delta\varphi_{21} = \pm 0.027$ rad = $\pm 1.5^\circ$; $\Delta\langle \mathbf{B} \cdot \mathbf{x}_3 \rangle = \pm 1.63$ nT. A variety of other applications of equations 8.23 and 8.24 will be given later on in this chapter.

8.3.2 Bootstrap Error Estimates

It has been proposed recently [Kawano and Higuchi, 1995] that the so-called bootstrap method may provide superior error estimates for the minimum variance analysis in the

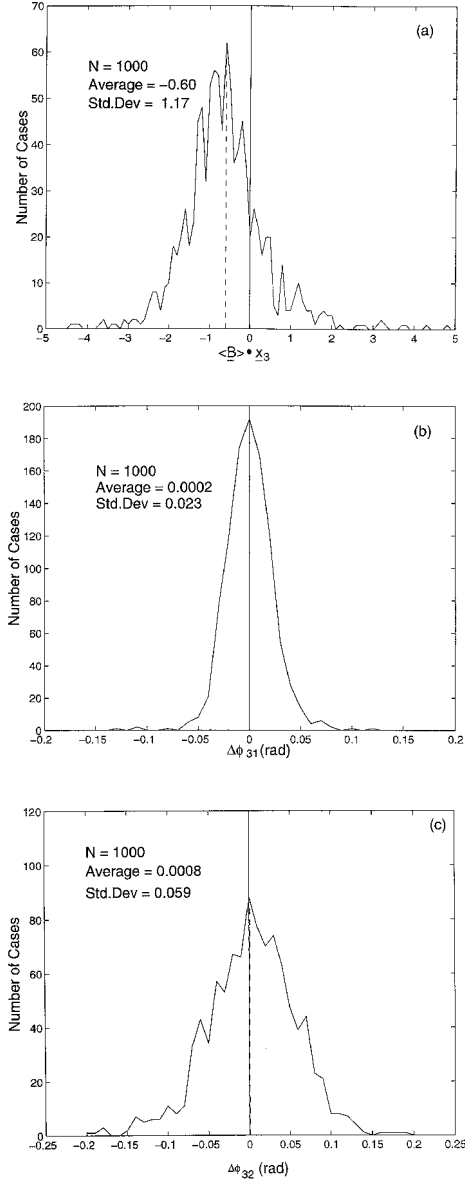


Figure 8.6: Bootstrap distributions for $N = 1000$ bootstrap sets $\{\mathbf{B}^{(m)}\}^{(i)}$ where $m = 1, 2, \dots, 16$ and $i = 1, 2, \dots, 1000$: (a) distribution of the normal field component, $\langle \mathbf{B}^{(m)} \rangle \cdot \mathbf{x}_3^{(i)}$; (b) distribution of angular deviation (radians) of $\mathbf{x}_3^{(i)}$ from vector \mathbf{x}_3 toward vector \mathbf{x}_1 (rotation about \mathbf{x}_2), with $(\mathbf{x}_1, \mathbf{x}_2, \mathbf{x}_3)$ given in Figure 8.2; (c) same for angular deviation of $\mathbf{x}_3^{(i)}$ from \mathbf{x}_3 toward \mathbf{x}_2 (rotation about \mathbf{x}_1).

sense that fewer assumptions are needed and that errors associated with time variance, non-planar effects, etc., are automatically incorporated. Although the utility of the method has not been tested extensively with real magnetopause data sets, it seems important to describe the bootstrap procedure briefly.

As described by *Kawano and Higuchi* [1995], the method is based on performing a very large number of minimum variance calculations, using for each calculation a bootstrap data sample generated from the measured vectors as follows. One vector is drawn at random, with replacement, from the measured set of M vectors. A second drawing is made in the same fashion and the process is continued until M drawings have been made. The resulting bootstrap sample thus consists of M vectors, all of which were measured but some of which may be identical so that a sample usually does not contain all the measured vectors. A large number ($N \geq 10^3$, say) of bootstrap samples is generated in this fashion and each sample is subjected to MVA thus producing a set of N minimum variance eigenvectors $\{\mathbf{x}_3\}$ and corresponding normal field components $\{\langle \mathbf{B} \rangle \cdot \mathbf{x}_3\}$. The distribution of these quantities can then be characterised in terms of averages, variances, skewness factors, etc. The square root of the variance (the standard deviation) is the quantity that can be directly compared to the error estimates in Section 8.3.1.

As an illustration, we have analysed the 16 vector data samples that form the basis of the minimum variance calculation illustrated in Figure 8.2. A total of $N = 10^3$ bootstrap samples were generated from this data set and the MVA was performed on each. The resulting distribution of normal field components and of the angular deviations of the minimum variance eigenvectors from \mathbf{x}_3 toward the eigenvectors \mathbf{x}_1 and \mathbf{x}_2 , with $(\mathbf{x}_1, \mathbf{x}_2, \mathbf{x}_3)$ coming from the measured set and given in Figure 8.2, are shown in Figure 8.6 along with the averages and standard deviations. These quantities are found to remain relatively insensitive to N , the number of bootstrap samples, at least for $N \geq 10^3$ (rough estimates can be obtained for substantially smaller N values). The bootstrap averages are close to the single-sample averages in Figure 8.2 and the bootstrap standard deviations for $\hat{\mathbf{n}}$ are about the same as the estimates in Section 8.3.1. It can be shown, based on [*Khrabrov and Sonnerup*, 1998a], that this agreement is not coincidental. The bootstrap standard deviation for $\{\langle \mathbf{B} \rangle \cdot \mathbf{n}\}$ (1.17 nT, see Figure 8.6a) is smaller than the value obtained analytically (1.63 nT, see Section 8.3.1). Because questions remain concerning the proper application of the bootstrap technique to MVA, we recommend use of the analytical results for which the underlying assumptions are transparent. Further discussion and application of the bootstrap technique may be found in Chapter 9. ⇒8.4

An estimator for the normal vector, $\hat{\mathbf{n}}$, can be obtained as the normalised component-by-component average of the individual bootstrap normals $\hat{\mathbf{n}}^{(i)}$ ($i = 1, 2, \dots, 1000$). The result is $\hat{\mathbf{n}} = (0.8669, -0.4982, 0.0182)$ which very nearly agrees with the vector \mathbf{x}_3 from Figure 8.2, the angle between the two vectors being only 0.04° . It follows from [*Khrabrov and Sonnerup*, 1998a] that this agreement is not accidental but is correct including terms of order ε (but not ε^2 ; ε was defined in connection with equation 8.22). The bootstrap average of the normal field component is -0.60 nT. An alternate estimator for the average normal, $\langle \langle \hat{\mathbf{n}} \rangle \rangle$, can be obtained from the N bootstrap normals by maximising the sum of the squares of the individual components $\hat{\mathbf{n}}^{(i)} \cdot \langle \langle \hat{\mathbf{n}} \rangle \rangle$. This procedure reduces to the *Siscoe et al.* least-squares problem (equation 8.17) with $\mathbf{B}^{(m)}$ replaced by $\hat{\mathbf{n}}^{(i)}$ and with the eigenvector of the largest rather than the smallest eigenvalue representing $\langle \langle \hat{\mathbf{n}} \rangle \rangle$. For our event, this procedure leads to very nearly the same $\hat{\mathbf{n}}$ vector as above. The square root of

the intermediate eigenvalue and of the smallest eigenvalue from the least-squares problem considered by *Siscoe et al.* and the corresponding eigenvectors define an uncertainty cone of elliptic cross section for the bootstrap normal vector. The size and orientation of this cone is found to be consistent with the standard deviations shown in Figures 8.6b and 8.6c and also with the results from equation 8.23 (for further discussion, see Section 8.5). It is concluded that the most convenient way to characterise the ensemble of bootstrap normal vectors is to apply the procedure of *Siscoe et al.* to them.

8.3.3 Other Error Estimates

Three other error estimates for MVAB can be found in the literature and will be mentioned briefly here.

Sonnerup [1971] proposed angular error estimates for \mathbf{x}_3 consisting of the angular change produced by assuming the variance ellipsoid to remain unperturbed but changing the minimum variance away from λ_3 by the amount $\Delta\lambda_3$ given by

$$|\Delta\lambda_3|^2 = \langle (\Delta M_{33})^2 \rangle = \frac{1}{(M-1)M} \sum_{m=1}^M \left[(\mathbf{B}^{(m)} \cdot \mathbf{x}_3 - \langle \mathbf{B} \rangle \cdot \mathbf{x}_3)^2 - \lambda_3 \right]^2 \quad (8.25)$$

This expression for $(\Delta\lambda_3)^2$ is simply the variance of λ_3 divided by $(M-1)$. From the two conditions

$$\begin{aligned} (\mathbf{x}_3 + \Delta\mathbf{x}_3) \cdot \mathbf{M}^B \cdot (\mathbf{x}_3 + \Delta\mathbf{x}_3) &= \lambda_3 + \Delta\lambda_3 \\ (\mathbf{x}_3 + \Delta\mathbf{x}_3)^2 &= 1, \end{aligned} \quad (8.26)$$

both carried to second order in $\Delta\mathbf{x}_3$, one then finds the expression

$$|\Delta\lambda_3| = \sum_{j=1}^3 (\Delta x_{3j})^2 (\lambda_j - \lambda_3) \quad (8.27)$$

The term $j = 3$ is equal to zero in the sum so that equation 8.27 defines a quadratic relationship between the two components of $\Delta\mathbf{x}_3$ that are perpendicular to \mathbf{x}_3 . It describes an error cone of elliptical cross section with major axis of half length $|\Delta x_{32}| = |\Delta\varphi_{32}| = [|\Delta\lambda_3|/(\lambda_2 - \lambda_3)]^{1/2}$ along \mathbf{x}_2 and minor axis of half length $|\Delta x_{31}| = |\Delta\varphi_{31}| = [|\Delta\lambda_3|/(\lambda_1 - \lambda_3)]^{1/2}$ along \mathbf{x}_1 . In other words, we can express the principal semiaxes of the cone of uncertainty as

$$|\Delta\varphi_{3i}| = |\Delta\varphi_{i3}| = \sqrt{|\Delta\lambda_3|/(\lambda_i - \lambda_3)} \quad i = 1, 2 \quad (8.28)$$

The values $|\Delta\lambda_{31}|$ and $|\Delta\lambda_{32}|$ from equation 8.28 can then be used in 8.24 to obtain $|\Delta(\mathbf{B} \cdot \mathbf{x}_3)|$. For the magnetopause crossing depicted in Figure 8.2 the resulting error estimates are $\Delta\lambda_3 = 3.06 \text{ (nT)}^2$, $\Delta\varphi_{32} = \pm 8.8^\circ$; $\Delta\varphi_{31} = \pm 3.2^\circ$; $\Delta(\mathbf{B} \cdot \mathbf{x}_3) = \pm 3.74 \text{ nT}$. It is seen that the uncertainties predicted from this model are more than twice as large as those given in Sections 8.3.1 and 8.3.2.

Kawano and Higuchi [1995] have pointed out that when the field component $\mathbf{B}^{(m)} \cdot \mathbf{x}_3$ is normally distributed the expression 8.25 for $|\Delta\lambda_3|^2$ can be replaced by the simpler formula

$$|\Delta\lambda_3|^2 = \langle (\Delta M_{33})^2 \rangle = 2\lambda_3^2/(M-1) \quad (8.29)$$

which can be obtained from equation 8.22 by putting $i = j = 3$. For the magnetopause crossing in Figure 8.2, this expression leads to $\Delta\lambda_3 = 2.59(\text{nT})^2$ which should be compared to $\Delta\lambda_3 = 3.06(\text{nT})^2$ from equation 8.25. By comparison of the combination of equations 8.28 and 8.29 to the estimate 8.23 with $j = 3$ we see that the error obtained from equations 8.28 and 8.29 is larger than that from 8.23 by the factor $(1 - \lambda_3/\lambda_1)^{1/2}[2(M - 1)]^{1/4}$. While the errors predicted by equation 8.23 exhibit the $M^{-1/2}$ dependence expected from stationary random noise, those predicted by 8.28 have a weaker $M^{-1/4}$ dependence. The latter behaviour may perhaps be justified when significant systematic errors (assumed independent of M) are present in addition to the noise. But for stationary isotropic noise, equation 8.23 has better theoretical justification than 8.28. Accordingly, with the caveats provided in Section 8.3.5, we recommend that, in the future, equation 8.23 be used. However, there is a need to perform benchmark tests of this equation by use of a variety of synthetic data sets.

Hoppe et al. [1981] used qualitative arguments to motivate the following formulas for the angular uncertainties in \mathbf{x}_3 :

$$\begin{aligned}\Delta\varphi_{31} &= \pm \tan^{-1} [\lambda_3/(\lambda_1 - \lambda_3)] \\ \Delta\varphi_{32} &= \pm \tan^{-1} [\lambda_3/(\lambda_2 - \lambda_3)]\end{aligned}\tag{8.30}$$

These formulas do not contain M , the number of data points used in the analysis. Therefore, in contrast to the estimates in Sections 8.3.1 and 8.3.2, these error estimates would not approach zero as $M \rightarrow \infty$, as expected for time-stationary statistical fluctuations. They are directly inconsistent with equation 8.23 or with equations 8.28 and 8.29. For the event in Figure 8.2, they produce the following error estimates: $\Delta\varphi_{31} = \pm 0.4^\circ$; $\Delta\varphi_{32} = \pm 3.1^\circ$. It appears that $\Delta\varphi_{31}$ is unrealistically small. Accordingly, we recommend against the use of equation 8.30.

A third, more extensive study of minimum variance errors was performed by *Lepping and Behannon* [1980]. It was based on a large number of simulated current sheet crossings in which the unperturbed tangential magnetic field rotated by an angle ω , but maintained constant magnitude while the normal field component was strictly constant (either non-zero or zero). To this unperturbed configuration was added isotropic noise of different amplitudes. The number of data points used in the MVA was proportional to the angle change ω , an assumption that is perhaps reasonable for solar-wind discontinuities but not necessarily for other current sheets. They encoded their error estimate for the normal field component, $\langle \mathbf{B} \rangle \cdot \mathbf{x}_3$, in a complicated purely empirical formula that, because of the special nature of the simulated structures, would seem to be of limited general use. If that formula is nevertheless used to estimate the error in $\langle \mathbf{B} \rangle \cdot \mathbf{x}_3$ for the event in Figure 8.2, the result is $\Delta\langle \mathbf{B} \cdot \mathbf{x}_3 \rangle \equiv \pm 1.9 \text{ nT}$, which is comparable to the error estimate in Section 8.3.1. However, the reader may wish to take note of Lepping and Behannon's claim that the formulas 8.25 and 8.28 usually underestimate the errors for the type of discontinuities they studied.

8.3.4 Stationarity

A partial check on time-stationarity and further testing of the error estimates can be obtained by doing MVAB on sets of nested data segments centred at or near the middle of the current sheet. Each larger member of a nest is generated by adding one data point at each end of the preceding segment. The smallest member of a nest has three data points.

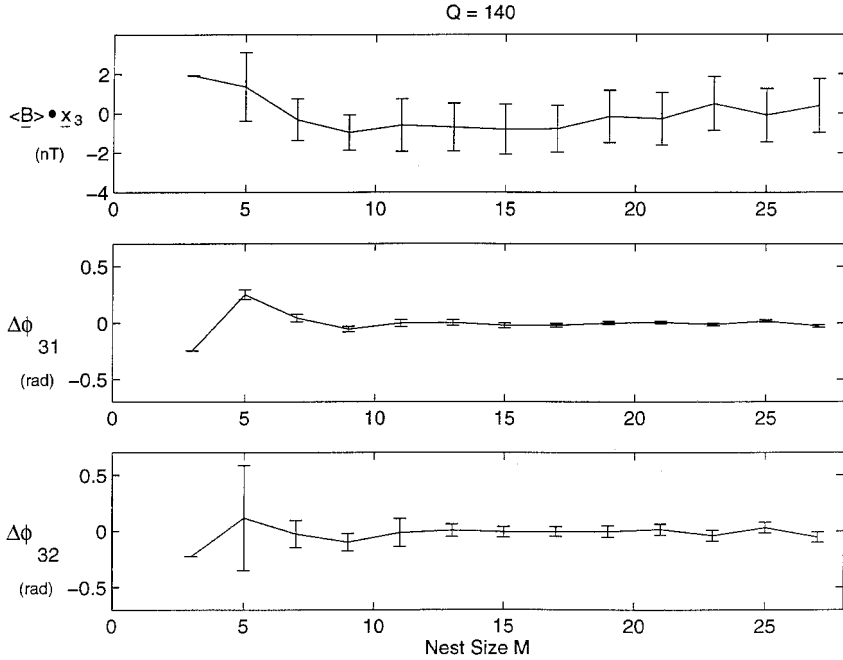


Figure 8.7: Results of minimum variance analysis for nested data segments starting with $M = 3$ vectors and then increasing to $M = 27$. The nest is centred at 05:18:50 UT. Each vector is the (4.35 s) average of $Q \simeq 140$ high-resolution \mathbf{B} measurements. Top panel shows normal magnetic field components and error bars for each M value. The quantities $\Delta\phi_{31}$ and $\Delta\phi_{32}$ indicate angular deviations (radians) from the reference normal \mathbf{x}_3 , toward \mathbf{x}_1 and \mathbf{x}_2 , respectively, where the set $(\mathbf{x}_1, \mathbf{x}_2, \mathbf{x}_3)$ is specified in Figure 8.2. Error bars are calculated from equations 8.23 and 8.24. Plateau used for composite error cone in Section 8.5 consists of the $K = 9$ largest segments.

The size of the largest member is limited by practical constraints such as the presence of magnetic structures, e.g., unrelated wave modes or turbulence, adjoining the current layer under study. If the normal vector and normal field component are strictly time stationary, then the results from all the different nested segments should be the same. In reality a group of the shortest and longest segments often gives results that are significantly different from those obtained for segments of intermediate duration. Within the intermediate range, the results of MVAB should be the same, or nearly the same, regardless of segment duration if the crossing is to be considered approximately time stationary. In other words, if the results of the MVAB ($\langle \mathbf{B} \rangle \cdot \mathbf{x}_3$ and the direction of \mathbf{x}_3) are plotted as functions of the segment length M (the segment duration is $T = (M - 1)\tau$, τ being the sampling interval), a plateau region should exist for intermediate M values: the wider and flatter the plateau the more nearly does the crossing exhibit time stationarity. The situation is illustrated in Figure 8.7 for the AMPTE/IRM event in Figure 8.2. Also shown are the statistical error bars calculated as described in Section 8.3.1. These error bars define a lower and an upper envelope curve for the expected values of the variable. Note that the error-bar lengths for

$\langle \mathbf{B} \rangle \cdot \mathbf{x}_3$ and $\Delta\varphi_{32}$ do not decrease as M increases, indicating that λ_3 is not constant but increases with M .

This result indicates that the noise is not time stationary. On the other hand, the near constancy of the three variables plotted in Figure 8.7 for $M \geq 9$ suggests that approximate time stationarity of the average magnetopause structure is at hand. The plateau average for the normal vector, calculated from the *Siscoe et al.* least-squares method, is $\hat{\mathbf{n}} = (0.8616, -0.5074, 0.0096)$ and the average field along it is $\langle \mathbf{B} \rangle \cdot \hat{\mathbf{n}} = 0.90$ nT. The vector $\hat{\mathbf{n}}$ forms an angle of 0.97° with the \mathbf{x}_3 vector from Figure 8.2; the error cone associated with the plateau normals will be discussed in Section 8.5.

8.3.5 Data Filtering and Optimal Analysis

Current layers observed in space often contain ELF magnetic fluctuations and various short-duration substructures of substantial amplitude. Such fluctuations influence the results of MVAB by increasing λ_3 and sometimes also by introducing a bias (systematic error) in the estimate for the normal vector, $\hat{\mathbf{n}}$. For this reason, it is desirable to examine the results of applying MVAB to data that have been subjected to various amounts of low-pass filtering. Such filtering can be performed in different ways and at various levels of sophistication but for present purposes the simplest approach suffices. It consists of calculating a set of consecutive non-overlapping means, each over Q high-resolution data points say, thus generating a new smoothed data set containing $\tilde{M} = M/Q$ (or, more precisely, the integer part of M/Q) points. MVAB is then performed for each choice of Q and the results for $\hat{\mathbf{n}}$ are compared.

Figure 8.8 shows the results of such a study for the AMPTE/IRM event in Figure 8.2. The calculations are based on a data interval containing 2048 magnetic field vectors measured at 32 samples per second and centred at 05:18:53:03 UT. The Q values used are $Q = 2^n$, $n = 0, 1, 2 \dots 9$, for which the corresponding \tilde{M} values are $\tilde{M} = 2^m$, $m = 11, 10, 9 \dots 2$. The figure shows the resulting ten normal vectors as dots on a nominal magnetopause tangent plane, defined by the vectors \mathbf{x}_1 and \mathbf{x}_2 in figure 8.2. Thus the reference normal vector in Figure 8.8, based on the usual 4.35 s averages ($Q \simeq 140$) is the eleventh dot, located at the origin and surrounded by its error ellipse (shown as a dashed curve). It is seen that the normals for $Q \leq 32$ are located in a tight cluster which is not centred at the origin but is displaced by approximately one degree from it. For $Q > 32$ the normal directions start to scatter increasingly widely and without an apparent pattern. All of the normals are consistent with the reference normal at the origin and its surrounding error ellipse. However, the reverse statement does not hold: the reference normal at the origin falls significantly outside the error cones associated with the smallest Q values (ellipse for $Q = 1$ is shown in the figure). We note that the tight clustering of the normal vectors for $Q \leq 32$ indicates that for instrumental and other reasons little noise is present in the data at the corresponding frequencies. It follows that for this event the $M^{-1/2}$ dependence of the error estimates is inappropriate for the smallest Q values, i.e., for the largest M values: it leads to analytical as well as bootstrap errors that are unrealistically small. For $Q = 32$ the error ellipse has grown to include the point at the origin while remaining significantly smaller than the (dashed) ellipse of the reference normal. Nevertheless, because of the possible bias mentioned above it remains unclear whether the normal for $Q = 32$ and its ellipse represent better estimates than the reference normal and its ellipse.

Experiments of the type described above are useful as a way to establish how robust

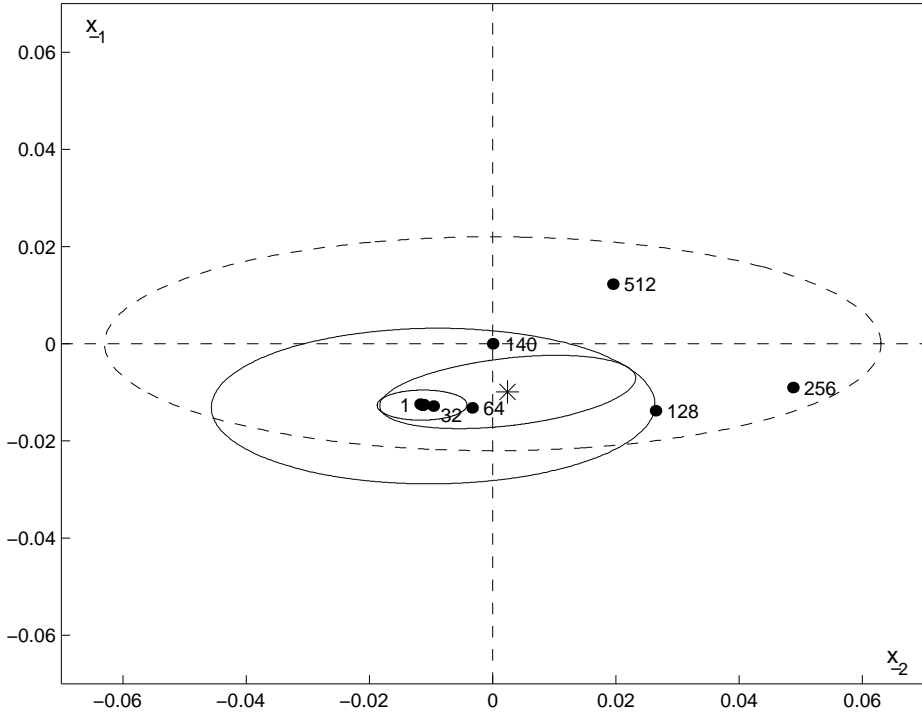


Figure 8.8: Influence of different averaging intervals, Q , on normal vectors obtained from MVAB: Q is the number of high-resolution vector samples (32 samples/s) averaged to obtain one of the \mathbf{B} vectors used in the minimum variance analysis ($Q \simeq 140$ corresponds to the standard averaging over 4.35 s used in Figure 8.7 and in all previous figures). Normal vectors for different Q values are shown as dots in the $\mathbf{x}_2\mathbf{x}_1$ plane with \mathbf{x}_2 and \mathbf{x}_1 from Figure 8.2 and with \mathbf{x}_3 from that figure represented by the dot at the origin ($Q = 140$). Axes indicate angular deviations, $\Delta\varphi_{32}$ and $\Delta\varphi_{31}$ (radians), of a normal toward or away from \mathbf{x}_2 and \mathbf{x}_1 , respectively. Analytically derived error cones, shown as elliptical projections in the $\mathbf{x}_2\mathbf{x}_1$ plane, are given for $Q = 1, 32$ and 140 . Average of all normals (except $Q = 140$) is shown as a star surrounded by its ellipse of uncertainty.

is the normal vector estimate obtained from MVAB. Indeed, a possible “best” normal and error cone may be obtained by applying the analysis of *Siscoe et al. [1968]* to all of the ten normal vectors obtained in the experiment. The result is shown by the star-shaped dot in Figure 8.8 and the ellipse centred at it. The error cone obtained in this manner provides an alternate result which can be used instead of, or in addition to, the analytical error estimates from Section 8.3.1 or the bootstrap errors from Section 8.3.2. From what has been said, it is clear that the latter two estimates must be applied to high-resolution data with extreme caution. It may also be desirable to extend the experimentation to include the use of nested data segments, perhaps with various choices for the nest centre and including the possibility of excising data segments taken within a current layer where there is evidence

that a substructure was sampled that had orientation different from that of the main layer. The hodogram pair provides an indispensable tool for the identification of such intervals and also for the choice of optimal beginning and end points of the total data interval used for MVAB.

8.4 Other Applications

The variance analysis technique developed in the previous sections also has applications to fields other than the magnetic field. In the context of finding the orientation of a quasi-one-dimensional layer, the important requirement that must be satisfied is that either the normal component or the two tangential components of the field analysed remain constant across the layer. Two examples are given below, along with a brief discussion of MVAB as it applies to two-dimensional structures, such as surface waves on a current sheet or magnetic flux ropes.

8.4.1 Minimum Variance Analysis on Mass Flux (MVA $\rho\mathbf{v}$)

The mass flux across a layer such as a shock or a rotational discontinuity should be independent of the coordinate z perpendicular to the layer provided the structure of the layer is time independent. In a frame moving with the layer, we then have

$$\nabla \cdot \rho \mathbf{v}' = 0 \quad (8.31)$$

which, assuming as before that $\partial/\partial x = 0$ and $\partial/\partial y = 0$, yields

$$\frac{\partial}{\partial z} \rho \mathbf{v}'_z = 0 \quad (8.32)$$

Here ρ is the mass density and \mathbf{v}' is the velocity vector in a frame moving with the layer, i.e., for each measured velocity, $\mathbf{v}^{(m)}$, in the spacecraft frame we have

$$\mathbf{v}^{(m)'} = \mathbf{v}^{(m)} - u_n \hat{\mathbf{n}} \quad (8.33)$$

where $u_n \hat{\mathbf{n}}$ is the unknown velocity of the moving frame. It is important to note that current layers and other discontinuities in space are almost always in a state of motion: they are observed as they move past an observing spacecraft at speeds that usually greatly exceed the spacecraft velocity. Because there is a net mass flow across a rotational discontinuity (RD) or a shock we have $\langle \mathbf{v} \rangle \cdot \hat{\mathbf{n}} \neq u_n$ so that u_n and $\hat{\mathbf{n}}$ are both unknown; only for a tangential discontinuity (TD) do we have $u_n = \langle \mathbf{v} \rangle \cdot \hat{\mathbf{n}}$. An additional complication that we do not address here is that the current sheet may be in a state of acceleration or deceleration so that u_n is not constant during the crossing.

Because of the fact that the unknown normal vector, $\hat{\mathbf{n}}$, enters in the expression for the vector field $\rho \mathbf{v}'$ from which the variance matrix, $\mathbf{M}^{\rho\mathbf{v}}$, is constructed, one must now use MVA in an iterative fashion. For example, one may first put $u_n = 0$ in order to obtain an initial $\hat{\mathbf{n}} = \mathbf{x}_3$ vector from MVA applied to the measured set $\{\rho^{(m)} \mathbf{v}^{(m)}\}$; this $\hat{\mathbf{n}}$ vector is then used with a chosen value for u_n (for RDs and shocks) or with $u_n = \langle \mathbf{v} \rangle \cdot \hat{\mathbf{n}}$ (for TDs) in order to produce a new set $\{\rho^{(m)} \mathbf{v}^{(m)'}\}$. MVA applied to this set will then yield a new $\hat{\mathbf{n}}$ vector to replace the initial one. The calculation is then repeated until $\hat{\mathbf{n}}$ no longer changes:

usually only a few iterations are needed but for large values of $|u_n|$ convergence problems may arise. In that case, a stepwise approach can be adopted in which one first iterates to obtain the \hat{n} vector corresponding to a smaller $|u_n|$ value and then uses that vector as the initial guess for a larger $|u_n|$, and so on. Rapidly converging iterative schemes for finding the smallest or largest eigenvalue of a matrix also exist and can be adapted to the present situation. Since the problem is nonlinear, difficulties with uniqueness may arise. For example, two iterations using the same value of u_n but starting from \mathbf{x}_3 and $-\mathbf{x}_3$, respectively, generally do not converge to the same final \hat{n} vector. However, it is usually easy to decide which answer makes physical sense.

⇒8.5

For a TD, normal vectors for different u_n values are calculated in the fashion described above until the u_n value has been found for which the condition $u_n = \langle \mathbf{v} \rangle \cdot \hat{n}$ is satisfied. Making the proper choice of u_n for RDs and shocks is a non-trivial task which requires observational information beyond $\rho^{(m)}$ and $\mathbf{v}^{(m)}$. If a good minimum variance direction has been obtained from MVAB, one may determine u_n by maximising error cone overlap for the MVAB normal and the MVA $\rho\mathbf{v}$ normal and then choosing the normal at the centre of the overlap region. If the MVAB normal is poorly determined but the maximum variance eigenvector, \mathbf{x}_1 , from MVAB has small errors, one may determine u_n from the condition $\hat{n} \cdot \mathbf{x}_1 = 0$. Other possibilities also exist. For RDs u_n could be chosen such that $\langle \mathbf{v}' \rangle \cdot \hat{n} = \pm \langle \mathbf{v}_A \rangle \cdot \hat{n}$, \mathbf{v}_A being the measured Alfvén velocities; for shocks one could use the Rankine-Hugoniot conditions, as discussed in detail in Chapter 10. Alternatively, timing differences from dual or multiple spacecraft measurements can be used to obtain u_n . Still another possibility (incompletely explored to date) is that one or more of the eigenvalues, $\lambda_1, \lambda_2, \lambda_3$, may exhibit an extremum at the correct value of u_n .

Provided $\lambda_2 \gg \lambda_3$, the process described above will provide a normal vector derived from plasma data alone but the direction of this vector depends somewhat on the choice of u_n , the magnitude of which remains unknown (unless there is other evidence to specify it as discussed above). Note that the sign of u_n is usually known. For example, a magnetopause crossing from the magnetosphere to the magnetosheath, such as the one in Figure 8.2, must have $u_n < 0$.

As an illustration, the minimum variance analysis on $\rho\mathbf{v}$ described in this section has been applied to the plasma data from the AMPTE/IRM magnetopause crossing in Figure 8.2. The results, for current-layer normal velocity $u_n = 0$ and for $u_n = -5$ km/s, are given in Table 8.2; hodograms for $u_n = -5$ km/s are shown in Figure 8.9. In this figure, the mass fluxes have been converted to weighted velocities, $\rho\mathbf{v}'/\langle\rho\rangle$ where $\langle\rho\rangle = 19.4$ particle masses per cm^3 . Information about plasma composition is required in order to determine the mass density but was not obtained by the AMPTE/IRM plasma instrument. If the composition ratios remain the same throughout the layer, then the normal-vector determination from $\rho^{(m)}\mathbf{v}^{(m)}$ does not depend on the actual values of those ratios. However, gradients in composition, which might be present across the magnetopause, would corrupt the determination.

From the results in Table 8.2 one can calculate that, for $u_n = 0$, the \mathbf{x}_3 vector derived from MVA $\rho\mathbf{v}$ forms an angle of 5.9° with the minimum variance direction for \mathbf{B} given in Figure 8.2 while for $u_n = -5$ km/s the angle has increased to 7.2° . Thus an attempt to minimise the angle would lead to $u_n > 0$; this would be inconsistent with the fact that the magnetopause was traversed in the direction from the magnetosphere to the magnetosheath which requires $u_n < 0$. This is also the case if the TD condition $u_n = \langle \mathbf{v} \rangle \cdot \hat{n}$, which should be applicable to this event, is implemented: it leads to $\hat{n} = (0.8584, -0.5096, -0.0577)$,

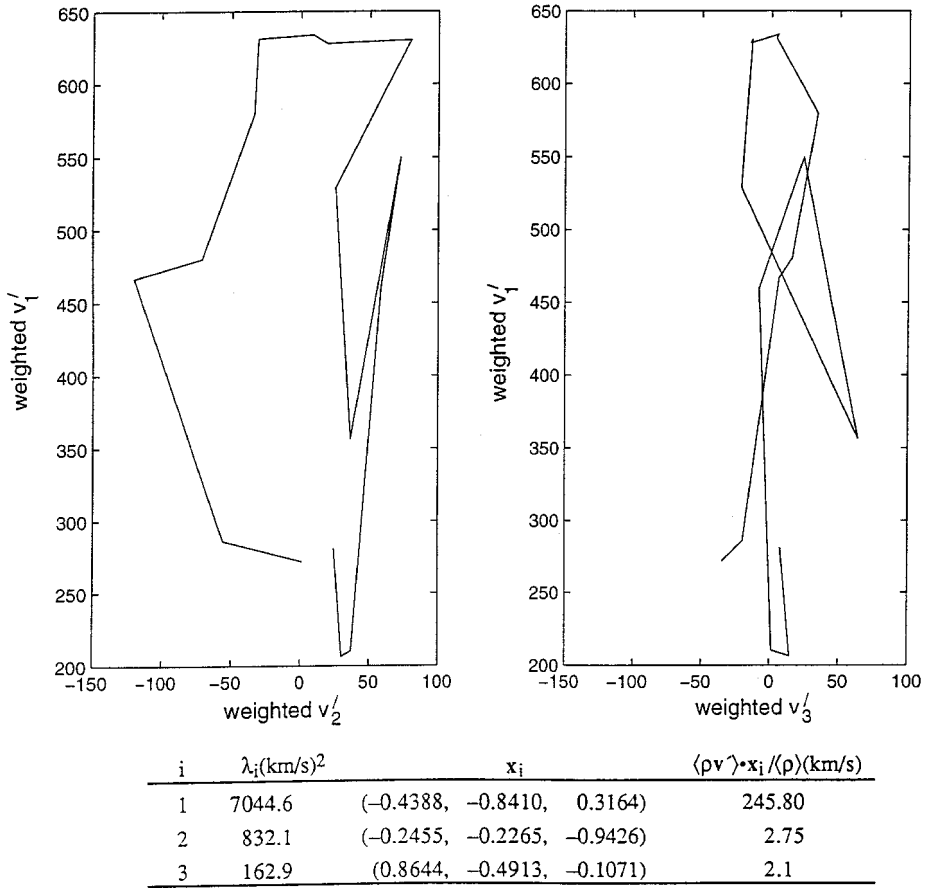


Figure 8.9: Hodograms and minimum variance results for the weighted velocity, $\rho \mathbf{v}' / \langle \rho \rangle$ (km/s), for event in Figure 8.2 with $u_n = -5$ km/s. The average density $\langle \rho \rangle$ corresponds to 19.4 particles/cm³; the range is 9.2–26.2 particles/cm³.

$u_n \simeq +6$ km/s and to a decrease of the angle between the MVAB and the MVA $\rho \mathbf{v}$ normals to 4.5°. Further discussion of the inconsistent sign for u_n is given in Section 8.5.

Because it is more difficult to accurately measure ρ and \mathbf{v} of the plasma (see Chapter 6) than to measure \mathbf{B} , the measurement errors in \mathbf{v} and, in particular, in ρ are at present sufficiently large so as to limit the ability of the analysis techniques described above to produce an accurate normal vector based on $\rho \mathbf{v}$ or a reliable value of u_n . It is emphasised that by requiring MVAB and MVA $\rho \mathbf{v}$ error-cone intersection to occur in order for u_n to be determined and by placing the final choice in the overlap region, this type of analysis will lead to rejection of poor cases and will place more emphasis on the normal vector determination having the smaller error cone. In our example, this means placing more emphasis on \mathbf{B} data and less on $\rho \mathbf{v}$ data. A more sophisticated technique which incorporates several conservation laws in one grand optimisation has been proposed recently [Kawano and Higuchi, 1996] but remains untested. Such approaches hold considerable promise but they

Table 8.2: Predictions from MVA ρv for the event in Figure 8.2

Quantity	$u_n = 0$ km/s	$u_n = -5$ km/s
\mathbf{x}_3	(0.8620, -0.4998 , -0.0847)	(0.8644, -0.4913 , -0.1071)
$\Delta\varphi_{31}$	0.0407 rad	0.0402 rad
$\Delta\varphi_{32}$	0.1447 rad	0.1421 rad
$\Delta\varphi_{21}$	0.0472 rad	0.0466 rad
$[\langle\rho\mathbf{v}'\rangle \cdot \mathbf{x}_3 \pm \Delta\langle\rho\mathbf{v}'\rangle \cdot \mathbf{x}_3]/\langle\rho\rangle$	0.8 ± 10.6 km/s	2.1 ± 10.4 km/s

may also encounter difficulties related to the limited quality of the plasma information. It is important that any such procedure should allow greater weight to be placed on quantities that are well determined with less weight on quantities that are less well determined.

Methods that utilise data on the two sides of the current layer but not in its interior are discussed in Chapter 10.

8.4.2 Maximum Variance Analysis on Electric Field (MVAE)

Faraday's law requires that the two electric field components tangential to a one-dimensional time independent layer remain constant throughout the layer and on its two sides. It is important to note that the electric field is to be measured in a frame of reference moving with the layer. There are certain classes of current layer for which this constancy of the tangential electric field can be used for determination of a vector normal to the layer. What is required is that the electric field component normal to the layer undergoes a large change as it often does at the magnetopause where the tangential flow and/or the tangential \mathbf{B} field components usually change direction and/or magnitude by substantial amounts across the layer. In that case, λ_1 is large and the maximum variance direction will serve as a good predictor of the normal to the layer. The variance ellipsoid would ideally be cigar shaped with $\lambda_1 \gg \lambda_2 \simeq \lambda_3$ (in practice one is more likely to find $\lambda_2 > \lambda_3$ or even $\lambda_2 \gg \lambda_3$ rather than $\lambda_2 \simeq \lambda_3$).

The electric field observed in a frame moving with the layer is

$$\mathbf{E}^{(m)'} = \mathbf{E}^{(m)} + u_n \hat{\mathbf{n}} \times \mathbf{B}^{(m)} \quad (8.34)$$

where $\mathbf{E}^{(m)}$ is the field measured in the spacecraft frame. The maximum variance eigenvector, \mathbf{x}_1 , of the matrix

$$\mathbf{M}_{\mu\nu}^E = \langle E'_\mu E'_\nu \rangle - \langle E'_\mu \rangle \langle E'_\nu \rangle \quad (8.35)$$

then becomes our predictor of the normal $\hat{\mathbf{n}}$. It is noted that, as in Section 8.4.1, the resulting normal vector depends on the choice of u_n , which may itself be a function of time, and that the calculation must be done iteratively since \mathbf{E}' itself contains the unknown normal $\hat{\mathbf{n}}$.

To date, this type of maximum variance analysis has not been performed on actual measured electric field vectors $\mathbf{E}^{(m)}$. Rather, the convection electric field $\mathbf{E}_c^{(m)} = -\mathbf{v}^{(m)} \times \mathbf{B}^{(m)}$, calculated from measured plasma velocities and magnetic fields, has been used as a proxy for $\mathbf{E}^{(m)}$. (A variant of MVAE that remains untested to date is to use $-\mathbf{v}_e^{(m)} \times$

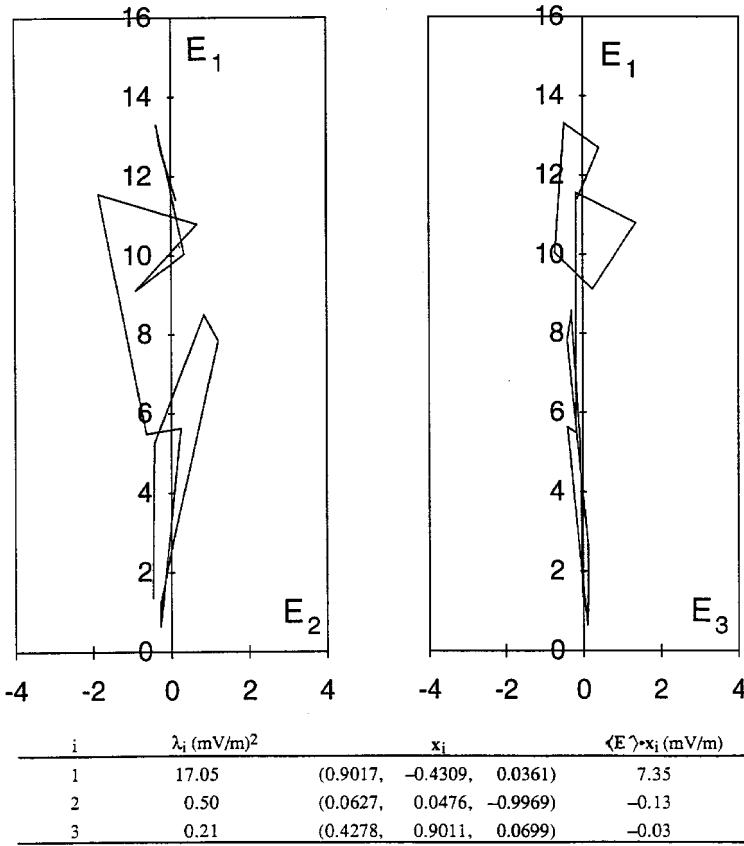


Figure 8.10: Hodograms and maximum variance results for $\mathbf{E}_c^{(m)} = -\mathbf{v}^{(m)} \times \mathbf{B}^{(m)}$ (mV/m) for the event in Figure 8.2 with $u_n = -5$ km/s.

$\mathbf{B}^{(m)}$, $\mathbf{v}_e^{(m)}$ being the measured electron velocities, in place of $\mathbf{E}_c^{(m)}$.) It has been found [Paschmann *et al.*, 1986; Sonnerup *et al.*, 1987, 1990] that the maximum variance analysis often provides a robust predictor of the magnetopause normal even in some cases where minimum variance analysis on \mathbf{B} gives $\lambda_2 \simeq \lambda_3$ and therefore fails to produce a reliable normal. Because the mass density ρ is at present less accurately determined from the plasma measurements than is the component of \mathbf{v} perpendicular to \mathbf{B} , the determination of $\hat{\mathbf{n}}$ and u_n from MVAE may be expected to be considerably better than that from MVA $_{\rho v}$. The comments, given in Section 8.4.1, concerning the determination of u_n , apply to MVAE as well.

We point out that, even when the correct value of u_n is known, the normal vectors obtained from MVAE are (usually weakly) frame dependent: if MVAE is performed for the same event in two frames that move relative to each other along the current sheet, the resulting normal vectors are not identical. This effect, which is caused by the noise, may become large if one moves to a frame of reference in which the noise in \mathbf{E}' dominates the convective field, $-\mathbf{v} \times \mathbf{B}$.

Table 8.3: Predictions from MVAE for the event in Figure 8.2.

Quantity	$u_n = 0$ km/s	$u_n = -5$ km/s
\mathbf{x}_1	(0.8911, -0.4533, 0.0197)	(0.9017, -0.4309, 0.0361)
$\Delta\varphi_{31}$	0.0289 rad	0.0290 rad
$\Delta\varphi_{32}$	0.2943 rad	0.2913 rad
$\Delta\varphi_{21}^*$	0.0451 rad	0.0453 rad
$\langle \mathbf{E} \rangle \cdot \mathbf{x}_2 \pm \Delta \langle \mathbf{E} \rangle \cdot \mathbf{x}_2$	-0.25 ± 0.38 mV/m	-0.13 ± 0.38 mV/m
$\langle \mathbf{E} \rangle \cdot \mathbf{x}_3 \pm \Delta \langle \mathbf{E} \rangle \cdot \mathbf{x}_3$	0.02 ± 0.25 mV/m	-0.03 ± 0.25 mV/m

*From equation 8.23 with λ_3 replaced by λ_2 .

As an illustration of MVAE, we have applied it to the AMPTE/IRM magnetopause crossing in Figure 8.2 with the results shown in Figure 8.10 and in Table 8.3. Note that \mathbf{x}_1 from MVAE with $u_n = 0$ forms an angle of 2.9° with \mathbf{x}_3 from MVAB in Figure 8.2 while for $u_n = -5$ km/s that angle has increased to 4.4° . Therefore an attempt to minimise the angle between the maximum variance eigenvector from \mathbf{E} and the minimum variance eigenvector from \mathbf{B} (Figure 8.2) would again lead to $u_n > 0$, in conflict with the actual direction of traversal of the magnetopause which requires $u_n < 0$. The TD condition, $u_n = \langle \mathbf{v} \rangle \cdot \hat{\mathbf{n}}$, leads to $u_n \simeq -48$ km/s but the resulting normal vector (0.9598, -0.2205 , 0.1736) is not believable: it forms an angle of 19.0° with \mathbf{x}_3 from MVAB.

An alternate approach to MVAE, in which a suitably defined residue of the integrated Faraday equation is minimised to yield both $\hat{\mathbf{n}}$ and u_n , has been described recently by *Terasawa et al.* [1996]. The method is based on the constancy of $\mathbf{B}^{(m)} \cdot \hat{\mathbf{n}}$ as well as of $\mathbf{E}^{(m)'} \times \hat{\mathbf{n}}$; minimisation of the variance of the tangential electric field plays a key role in the determination of u_n . As stressed in Section 8.4.1, caution should be exercised in accepting the results from such combined calculations because it is not clear that they properly weight information of high and less high quality. We have applied the method of *Terasawa et al.* to the AMPTE/IRM crossing in Figure 8.2 with the result $\hat{\mathbf{n}}_T = (0.8945, -0.4465, 0.0247)$ and $u_n = -1.5$ km/s. We also find $\langle \mathbf{v} \rangle \cdot \hat{\mathbf{n}}_T = -5$ km/s and $\langle \mathbf{B} \rangle \cdot \hat{\mathbf{n}}_T = -0.24$ nT. These results are reasonable: $\langle \mathbf{B} \rangle \cdot \hat{\mathbf{n}}_T$ is near zero and both u_n and $\langle \mathbf{v} \rangle \cdot \hat{\mathbf{n}}_T$ are negative, as required for an earthward moving TD. An inconsistency is that $u_n \neq \langle \mathbf{v} \rangle \cdot \hat{\mathbf{n}}_T$ but this discrepancy is well within uncertainties. However, we find that $\hat{\mathbf{n}}_T$ is close to the $\hat{\mathbf{n}}$ vector from MVAE with $u_n = 0$ (see Table 8.3): the two vectors form an angle of only 0.4° whereas the angle between $\hat{\mathbf{n}}_T$ and the MVAB vector \mathbf{x}_3 in Figure 8.2 is 3.3° . In this example, the method evidently puts most of the weight on the constancy of the tangential electric field, $\mathbf{E}^{(m)'} \times \hat{\mathbf{n}}_T$, and relatively little weight on the constancy of the normal magnetic field, $\mathbf{B}^{(m)} \cdot \hat{\mathbf{n}}_T$.

A number of items relative to MVAE and to the method of *Terasawa et al.* remain unexplored or incompletely understood at present. Included are the details of the inter-relationship between the two methods and their mutual relationship to the existence and quality of a so-called deHoffmann-Teller frame (see Chapter 9), i.e., a moving frame of reference in which the electric field is absent or small.

8.4.3 Application to 2-D Structures

We now discuss briefly a quite different use of MVAB, introduced by *Walthour and Sonnerup* [1995]. They applied this technique to magnetic field data obtained from remote sensing of two-dimensional bulges, such as flux transfer events or surface waves on the magnetopause, moving past an observing spacecraft. During the event, the spacecraft is assumed to be located entirely on one side of the current layer; it cannot penetrate the bulge or the layer. The details of this application are too lengthy to be presented here but some of the results are of interest. Linear ideal MHD theory can be used to show that the smallest eigenvalue of $M_{\mu\nu}^B$ (defined by equation 8.8) should be zero, i.e., $\lambda_3 = 0$, and that the corresponding eigenvector \mathbf{x}_3 should be tangential to the unperturbed layer, forming a certain angle, θ_N , with the invariant direction (the axis) of the bulge. The angle θ_N is obtainable from the direction of motion of the bulge relative to the unperturbed field and the so-called stretching factor, R , which enters into the MHD perturbation analysis of flow over small-amplitude bulges [*Sonnerup et al.*, 1992]. A curl-free magnetic field corresponds to $R = 1$ and $\theta_N = 0$; in that special case the field component along the invariant direction is exactly constant so that \mathbf{x}_3 is along the bulge axis. In the general case, it can also be shown [*Khrabrov and Sonnerup*, 1998b] that either the intermediate variance eigenvector, \mathbf{x}_2 , or the maximum variance eigenvector, \mathbf{x}_1 , should be normal to the unperturbed layer and that the ratio of maximum to intermediate eigenvalue $\lambda_1/\lambda_2 = R^2/\cos^2\theta_N$ in the former case and $\lambda_2/\lambda_1 = R^2/\cos^2\theta_N$ in the latter case. Several actual magnetopause events have been analysed for which these predicted properties of $M_{\mu\nu}^B$ were verified.

Another illustration is the case of a spacecraft traversing a force-free flux rope in which the axial current, along the z axis say, is uniformly distributed over the circular cross section of the rope. It is simple to show that magnetic data taken along a straight-line trajectory through the interior of the rope will produce $\lambda_3 = 0$ and a minimum variance direction that is not along the rope axis but is instead perpendicular to it and is, in fact, along the projection of the trajectory onto the perpendicular xy plane. Application of MVAB to flux ropes is discussed further by *Elphic and Russell* [1983] and *Lepping et al.* [1990].

From these examples the following lesson can be drawn. A nearly one-dimensional current layer will produce $\lambda_3 \ll \lambda_1$ but the converse is not necessarily true: if one finds $\lambda_3 \ll \lambda_1$ for a measured data set, it does not automatically follow that a 1-D current layer has been traversed. Furthermore, if one seeks an interpretation in terms of a 2-D structure instead, one cannot necessarily conclude that the minimum variance direction coincides with the axis of the structure.

8.5 Discussion of AMPTE/IRM Event

In this chapter we have used one particular magnetopause crossing by the AMPTE/IRM spacecraft in order to illustrate various applications of the minimum/maximum variance analysis and associated error estimates. For benchmarking purposes, we provide, in the Appendix, the 16 samples of field \mathbf{B} , number density N^* , and velocity \mathbf{v} for this event that were used in most of the calculations. We now compare the various results for the normal vector estimate, the normal field and flow component estimates, and the error estimates

for this event.

An overview of the normal vector orientations is given in Figure 8.11 where each normal vector is shown as a dot in the $(\mathbf{x}_1, \mathbf{x}_2)$ plane. The axes of this plane are provided by the maximum and intermediate eigenvectors in Figure 8.2 and the distances along the \mathbf{x}_1 and \mathbf{x}_2 axes represent angular deviations, $\Delta\varphi_{31}$ and $\Delta\varphi_{32}$, in radians. Thus the normal vector \mathbf{x}_3 from Figure 8.2 appears as point 1 at the origin with its error-cone cross section (from equation 8.23) being the dashed ellipse with major axis along \mathbf{x}_2 and minor axis along \mathbf{x}_1 . The minimum variance calculation with constraint $\langle \mathbf{B} \rangle \cdot \hat{\mathbf{n}} = 0$ gives point 2 and the normal vector of *Siscoe et al.* (from the expression 8.17) gives point 3 (no error cones shown). The average bootstrap estimate of $\hat{\mathbf{n}}$, calculated as the eigenvector corresponding to the maximum λ value in the analysis of *Siscoe et al.*, applied to the bootstrap normals, is point 4 which is very nearly coincident with point 1. It is surrounded by its slightly tilted error-cone cross section, the major and minor axes of which are along the eigenvectors corresponding to intermediate and minimum λ values in the analysis of *Siscoe et al.* and the semiaxis lengths of which are the square root of those λ values. This ellipse is nearly the same as the dashed ellipse for point 1. Point 5, surrounded by its error-cone cross section (or, more precisely, the projection of that cross section onto the x_1x_2 plane), is similarly the normal vector obtained by applying the analysis of *Siscoe et al.* to the $K = 9$ normal vectors on the plateau in Figure 8.8. The ellipse is small and falls inside the dashed error ellipse associated with point 1, suggesting that the latter error estimate is reasonable. The clustering of all of points 1–5 within the dashed ellipse reinforces the view that the error estimate from 8.23 is valid.

Points 6, 7 and 8, which are located along a straight line, are normal vectors from MVAE with $u_n = 0$ km/s, $u_n = +12$ km/s (point of maximum error-cone overlap with MVAB) and $u_n = +20$ km/s (from condition $\hat{\mathbf{n}} \cdot \mathbf{x}_1 = 0$ with \mathbf{x}_1 from MVAB), respectively. The elliptical error-cone cross section refers to point 6. For MVAE, the TD condition $u_n = \langle \mathbf{v} \rangle \cdot \hat{\mathbf{n}}$ yields $u_n = -48$ km/s; the corresponding normal vector is located outside of the diagram at $\mathbf{x}_1 = -0.13$, $\mathbf{x}_2 = -0.30$. Points 9, 10 and 11, which are also located along a straight line, are normal vectors from MVA ρ v with $u_n = 0$ km/s, $u_n = +6$ km/s (from the TD condition, $u_n = \langle \mathbf{v} \rangle \cdot \hat{\mathbf{n}}$) and $u_n = +15$ km/s (maximum error-cone overlap and also $\hat{\mathbf{n}} \cdot \mathbf{x}_1 \simeq 0$). The large elliptical error-cone cross section refers to point 10. Finally, point 12 is the normal from Terasawa's method. The relations $\langle \mathbf{B} \rangle \cdot \hat{\mathbf{n}} = 0$ and $\langle \mathbf{v} \rangle \cdot \hat{\mathbf{n}} = 0$, calculated from the 16 data points in the appendix, are also shown in Figure 8.11 as lines B–B and V–V, respectively. To the right of B–B $\langle \mathbf{B} \rangle \cdot \hat{\mathbf{n}} < 0$ and to the left of it $\langle \mathbf{B} \rangle \cdot \hat{\mathbf{n}} > 0$. Similarly, $\langle \mathbf{v} \rangle \cdot \hat{\mathbf{n}} > 0$ above V–V and $\langle \mathbf{v} \rangle \cdot \hat{\mathbf{n}} < 0$ below it.

Point 6 ($u_n = 0$ km/s, $\langle \mathbf{v} \rangle \cdot \hat{\mathbf{n}} = -3.6$ km/s) represents the best determination of $\hat{\mathbf{n}}$ from MVAE that is approximately consistent with $u_n \leq 0$ and with the structure being that of a TD. The result has difficulties: (i) the true value of u_n must in fact be negative and must coincide with $\langle \mathbf{v} \rangle \cdot \hat{\mathbf{n}}$; (ii) the error cone for point 6 does not contain points 1–5 and has only a small overlap with the error cone surrounding point 1. To bring about maximal error-cone overlap $u_n = +12$ km/s is needed. Similarly, point 10 ($u_n = \langle \mathbf{v} \rangle \cdot \hat{\mathbf{n}} = +6$ km/s) represents the best determination of $\hat{\mathbf{n}}$ and u_n from MVA ρ v for this TD. Again u_n has the wrong sign. The error cone for this point is large enough to include points 1–5 but optimal overlap would require $u_n \simeq +15$ km/s. These results all suggest a systematic deviation of the normals based on plasma measurements from the more reliable normals obtained from the magnetic data alone: they indicate that the plasma velocity component along $\hat{\mathbf{n}}$ may be too large. If that component is decreased by 12 km/s, say, point 6 from MVAE

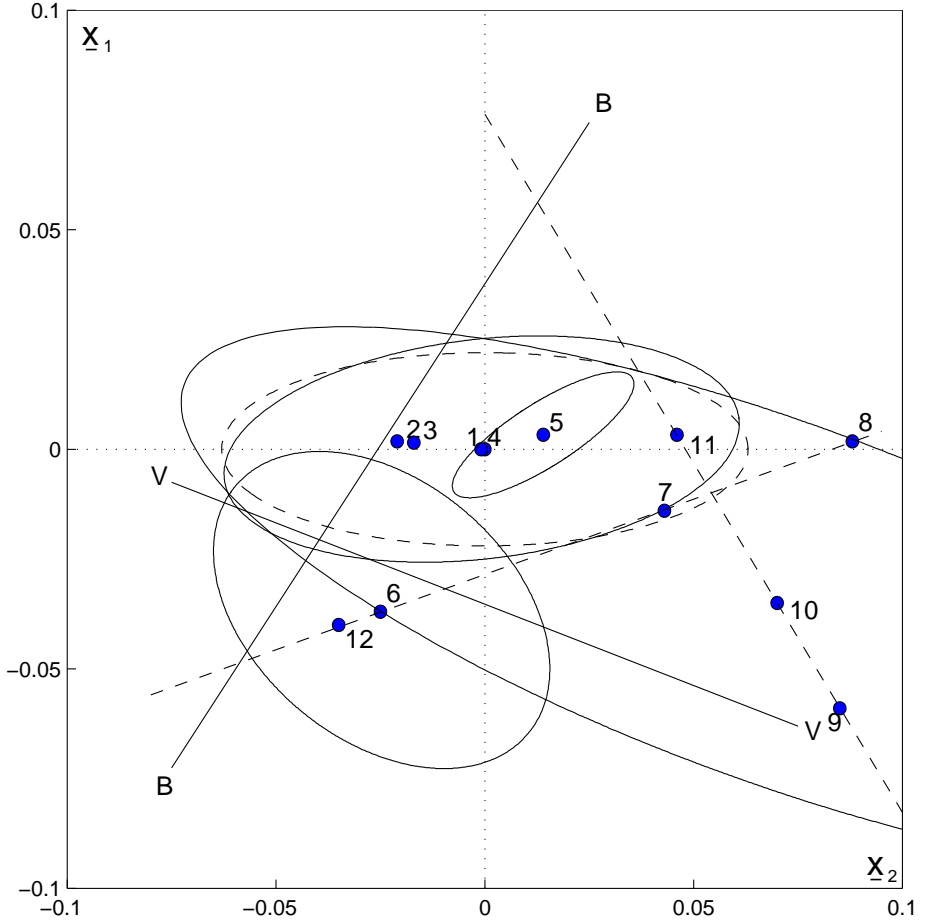


Figure 8.11: Overview of normal-vector determinations and errors for event in Figure 8.2. The x_2x_1 plane is shown with x_1 and x_2 given in Figure 8.2. The reference normal vector \mathbf{x}_3 from Figure 8.2 is the point at the origin; deviations $\Delta\varphi_{31}$ and $\Delta\varphi_{32}$ (radians) are shown as distances along x_1 and x_2 , respectively. The normal vectors, numbered from 1 to 12 are described in Section 8.5. The line B–B corresponds to $\langle \mathbf{B} \rangle \cdot \hat{\mathbf{n}} = 0$; on the left of that line $\langle \mathbf{B} \rangle \cdot \hat{\mathbf{n}} > 0$ and on the right $\langle \mathbf{B} \rangle \cdot \hat{\mathbf{n}} < 0$. Similarly, the line V–V corresponds to $\langle \mathbf{v} \rangle \cdot \hat{\mathbf{n}} = 0$; above the line $\langle \mathbf{v} \rangle \cdot \hat{\mathbf{n}} > 0$ and below it $\langle \mathbf{v} \rangle \cdot \hat{\mathbf{n}} < 0$. Both B–B and V–V are calculated from the data set in the Appendix. Ellipses represent error-cone cross sections, at unit distances from the cone vertex, projected onto the x_2x_1 plane.

would move to the location of point 7 where maximum error-cone overlap with the MVAB normal (point 1) occurs. Similarly, the TD point 10 from MVA ρv would move to the vicinity of point 11, again providing maximum error-cone overlap with the MVAB normal. However, the actual values of u_n for the TD points at their new locations would remain at 0 km/s and +6 km/s, respectively which is inconsistent with the requirement $u_n < 0$. Another indication of a systematic error in the plasma velocities is that the MVAB normal vector (point 1) has $\langle v \rangle \cdot \hat{n} = \langle v \rangle \cdot x_3 = +8$ km/s, i.e., it is well above the line V–V at which $\langle v \rangle \cdot \hat{n} = 0$. This result suggests that the measured normal velocity component of the plasma is too large by at least 8 km/s and perhaps more, depending on how rapid was the actual inward motion of the magnetopause past the spacecraft, i.e., depending on how negative u_n was. In Chapter 9, the magnetopause velocity u_n is calculated as the component of the deHoffmann-Teller frame velocity V_{HT} along x_3 , the MVAB reference normal. When the data set in the Appendix is used to obtain V_{HT} , the result is either $u_n = V_{HT} \cdot x_3 = (5.9 \pm 3.9)$ km/s or $u_n = (8.1 \pm 4.2)$ km/s, depending on certain details of the calculation. These results again suggest the presence of a small systematic error (although, as noted in Chapter 9, somewhat longer data segments for the calculation of V_{HT} do in fact yield the required negative values of u_n).

Because of the likely presence in the plasma velocity data of such a systematic error, perhaps caused by the plasma instrument being in its half-sweep mode as discussed below, it does not appear possible to determine the actual negative value of u_n for this event. However, it seems to have been small, probably in the range $0 > u_n > -10$ km/s.

The reason for the discrepancy between the normals based purely on B (points 1–5 in Figure 8.11) and those based partially or wholly on the plasma measurements (points 6 and 10, say, in Figure 8.11) is not understood. But it may be related, at least in part, to the fact that the plasma analyser was in its so-called half-sweep mode during the crossing. In this mode the ion energy range sampled is reduced from $20 \text{ eV/q} < E < 40 \text{ keV/q}$ to $150 \text{ eV/q} < E < 5.3 \text{ keV/q}$. Simulations of the instrument under typical magnetopause conditions indicate that systematic directional errors of the plasma flow velocity of a few degrees may occur in the half-sweep mode, along with an overestimate of the velocity magnitude and an underestimation of the density [G. Paschmann, private communication, 1996]. Some time aliasing is also likely to be present. Keeping these effects in mind, we conclude that the approximate agreement, illustrated in Figure 8.11, between the plasma-based normal vectors from MVA ρv and MVAE and the various magnetically-based normal vectors from MVAB is rather remarkable.

8.6 Summary

We summarise the material presented in this chapter as follows.

1. The main application of minimum variance analysis (MVA) is to the task of estimating \hat{n} , the direction perpendicular to approximately one-dimensional structures, such as current layers and plane wave fronts, from a set of magnetic field data measured by a single spacecraft during traversal of the structure.
2. MVA applied to magnetic field data (MVAB) consists of constructing the variance matrix (equation 8.8) and then finding its eigenvalues ($\lambda_1, \lambda_2, \lambda_3$) and corresponding eigenvectors (x_1, x_2, x_3). The direction of minimum variance, x_3 , is used as an

estimator of $\hat{\mathbf{n}}$, the corresponding eigenvalue, λ_3 , being the variance of the field component along \mathbf{x}_3 . A small value of λ_3 , compared to λ_2 and λ_1 , generally signals a good determination of $\hat{\mathbf{n}}$. Extreme caution must be exercised in cases where the separation between the minimum eigenvalue, λ_3 , and the intermediate eigenvalue, λ_2 , is small: the matrix is then nearly degenerate and small changes in the data interval used for the analysis may lead to an interchange of the eigenvectors \mathbf{x}_3 and \mathbf{x}_2 (Section 8.2.5). The normal vector obtained is unreliable in such cases. A simple rule of thumb is $\lambda_2/\lambda_3 \geq 10$ for a good application of MVA to a relatively small data set ($M < 50$, say).

3. Use of the *Siscoe et al.* formulation (equation 8.17) in place of 8.8 for determination of $\hat{\mathbf{n}}$ is not recommended, except perhaps for tangential discontinuities when λ_2/λ_3 is not large; however, for such cases MVA with constraint $\langle B_3 \rangle = 0$ (Section 8.2.6) $\Rightarrow 8.6$ gives similar results. Nevertheless, the *Siscoe et al.* analysis has been found useful for characterising a cluster of normal vectors obtained from MVA bootstrap calculations or from analysis of nested data segments.
4. We recommend (Section 8.3.1) that statistical errors be estimated by use of equations 8.23 and 8.24 and that a number of earlier estimates that have appeared in the literature be discarded. For small errors, the formula 8.23 gives results that nearly duplicate those obtained from the bootstrap technique (Section 8.3.2). In that case, the bootstrap average normal is also essentially the same as the minimum variance normal obtained by a single application of MVA to the measured data set. Thus there is no clear reason to perform the more time consuming bootstrap calculations to obtain $\hat{\mathbf{n}}$ and its angular uncertainties. However, in cases where the signal-to-noise ratio is small (λ_3 is comparable to λ_2 and perhaps λ_1) and the number of data points, M , used is not correspondingly large or when systematic errors are present, the bootstrap calculation may possibly give more realistic error estimates than equation 8.23. There are modest discrepancies between the analytical error estimate 8.24 for the average normal-field component and the corresponding bootstrap result, calculated as described by *Kawano and Higuchi* [1995]. The reasons for these deviations are not clearly understood but, at present, our recommendation is to use the results from equation 8.24. It is also noted that the estimates given by equations 8.23 and 8.24 are inversely proportional to $M^{1/2}$ so that very small errors may result when large data sets are used. Although this behaviour is expected for random noise, it may lead to the underestimation of uncertainties when data at high time resolution are used. It must also be remembered that in many applications additional, systematic errors, not described by equations 8.23 and 8.24, may be present as well. $\Rightarrow 8.3$ $\Rightarrow 8.4$
5. It is recommended that MVA be performed on nested sets of data intervals, centred at or near the middle of the structure being measured. Time stationarity can be checked in this manner (Section 8.3.4) but care must be taken to avoid including adjoining structures or turbulence that may degrade the quality of the normal vector determination. Similarly, it may sometimes be important to perform low-pass filtering of the data before MVA is applied (Section 8.3.5) in order to remove higher frequency wave activity in the interior of the layer. Such waves may have propagation vectors very different from $\hat{\mathbf{n}}$ and therefore may degrade the determination of

$\hat{\mathbf{n}}$. In most cases averaging the data over consecutive non-overlapping time intervals will provide filtering of sufficient quality.

6. We have pointed out (Section 8.4.3) that certain two-dimensional structures will lead to a magnetic variance matrix having λ_3 equal to zero, ideally, and in reality having $\lambda_3 \ll \lambda_2, \lambda_1$. It must be remembered, therefore, that the existence of a direction of low variance for a data set does not guarantee that the structure sampled by the spacecraft was a one-dimensional current sheet or wave.
7. There are also relevant applications of MVA to vector fields other than \mathbf{B} . In particular, mass flow conservation ($\text{MVA}_{\rho\mathbf{v}}$; see Section 8.4.1) across a one-dimensional layer implies that the minimum variance direction for $\rho\mathbf{v}'$, where $\mathbf{v}' = \mathbf{v} - u_n\hat{\mathbf{n}}$ is the plasma velocity in a frame moving with the layer, should be a good normal direction. A difficulty with this application is that the value of the speed, u_n , of the layer must be obtained separately from other measurements or conditions which is a simple task only for tangential discontinuities. Similarly, the maximum variance direction for the convection electric field $\mathbf{E}' = -(\mathbf{v} - u_n\hat{\mathbf{n}}) \times \mathbf{B}$ in the frame moving with the layer should be normal to the layer as a consequence of Faraday's law (MVAE ; see Section 8.4.2). Again, the value of u_n must be obtained separately. Experience indicates that this maximum variance analysis on \mathbf{E}' sometimes gives a reasonably reliable normal vector, the requirement being that $\lambda_1 \gg \lambda_2, \lambda_3$, even where minimum variance analysis on \mathbf{B} does not. The error estimate 8.23 for $|\Delta\varphi_{21}|$ should be modified by replacing λ_3 by λ_2 in the application to the MVAE normal. ⇒8.5
8. Recently, an analysis technique has been proposed in which the constancy of the normal component of \mathbf{B} and of the tangential components of \mathbf{E} are both satisfied in a single optimisation that yields $\hat{\mathbf{n}}$ as well as u_n [Terasawa et al., 1996]; an analytical solution to this optimisation problem and associated error estimates have been developed by Khrabrov and Sonnerup [1998c]. Although the utility of the method by Terasawa et al. has not yet been widely tested, it performed well in our AMPTE/IRM sample event. However, concerns exist that it does not always provide the proper relative weighting of quantities that are measured or calculated more accurately and less accurately. This concern has been addressed to some extent in a recent generalisation of MVA to include a variety of additional conservation laws [Kawano and Higuchi, 1996]. However, this method remains untested. At present, we are therefore in favour of separate use of MVAB and MVAE (plus perhaps $\text{MVA}_{\rho\mathbf{v}}$). The former yields a single optimal normal vector and an associated error cone. The latter yields a distinct normal vector and error cone for each chosen value of u_n . For tangential discontinuities, the proper value of u_n can in principle be found from the condition $u_n = \langle \mathbf{v} \rangle \cdot \hat{\mathbf{n}}$ but there is no guarantee that the result is reasonable. In other cases, the optimal choice for u_n should be that which leads to maximal overlap between the MVAB error cone and that of MVAE (and perhaps $\text{MVA}_{\rho\mathbf{v}}$); the optimal $\hat{\mathbf{n}}$ could then be taken at the centre of the overlap region. If the normal vector from MVAB has large errors, an alternate choice for u_n may be such that the corresponding normal vector from MVAE or $\text{MVA}_{\rho\mathbf{v}}$ is perpendicular to the maximum variance direction from MVAB. Or, following Terasawa et al., one may search for the u_n value that gives a minimum in the sum $(\lambda_2 + \lambda_3)$, obtained from MVAE. One advantage of the separate use of MVAB, MVAE, and $\text{MVA}_{\rho\mathbf{v}}$ is that

inconsistencies in the various determinations of \hat{n} may be revealed, as illustrated in Section 8.5. Such inconsistencies, which point to possible systematic errors in some of the measured quantities, may remain hidden if a single combined merit function is optimised.

Appendix

The 16 AMPTE/IRM data points, from 05:18:20–05:19:26 UT on October 19, 1984, used in Figures 8.2, 8.6, 8.9, and 8.10, are given in Table 8.4 for use in benchmark tests of MVA and other programs. It is noted that these data differ slightly from those used in *Sonnerup et al. [1990]*, probably as a result of round-off errors during coordinate transformations. Therefore normal vectors given in that paper differ slightly from those reported here.

Table 8.4: The 16 AMPTE/IRM data points from 05:18:20–05:19:26 UT on October 19, 1984 used in various demonstrations in this chapter

T	B_x	B_y	B_z	v_x	v_y	v_z	N^*
0.00	−13.6	−24.7	54.6	−111.0	−211.0	57.0	12.18
4.38	−14.8	−24.9	58.7	−102.0	−213.0	41.0	9.22
8.75	−13.4	−17.2	62.4	−111.0	−196.0	34.0	9.92
13.12	−14.0	−25.0	43.8	−135.0	−229.0	54.0	18.08
17.49	−7.1	−4.5	33.5	−128.0	−252.0	54.0	20.39
21.87	−0.9	−5.0	44.4	−82.0	−237.0	51.0	15.00
26.24	−10.0	−0.4	44.6	−139.0	−228.0	77.0	20.19
30.61	−6.1	−4.8	21.1	−143.0	−241.0	57.0	23.53
34.98	1.2	1.6	21.0	−132.0	−226.0	80.0	24.31
39.36	−3.4	−3.9	4.1	−117.0	−217.0	79.0	25.91
43.73	−0.9	1.2	5.0	−112.0	−210.0	93.0	26.17
48.10	−1.0	−1.5	12.3	−98.0	−212.0	92.0	24.49
52.47	11.0	13.2	29.7	−90.0	−186.0	104.0	22.20
56.85	19.1	34.4	20.1	−83.0	−168.0	121.0	22.86
61.22	24.9	50.1	1.9	−82.0	−129.0	88.0	17.56
65.59	29.2	47.1	−10.6	−93.0	−123.0	53.0	17.86

Units: B (nT); v (km/s); N^* (particles/cm³);
time T (sec), starting at 05:18:20.49 UT.

Acknowledgements

We thank G. Paschmann and H. Lühr for use of AMPTE/IRM plasma and magnetic field data. We are grateful to A. V. Khrabrov for developing the error analysis and for many other valuable

contributions to this chapter. We also thank H. Kawano for helpful comments. The research was supported by the National Aeronautics and Space Administration under grants NAGW-4023, -5137 and NAG 5-3031 and by the National Science Foundation, Atmospheric Sciences Division, under grant ATM-9422918 to Dartmouth College.

Bibliography

- Elphic, R. C. and Russell, C. T., Magnetic flux ropes in the Venus ionosphere, *J. Geophys. Res.*, **88**, 58, 1983.
- Hoppe, M. M., Russell, C. T., Frank, L. A., Eastman, T. E., and Greenstadt, E. W., Upstream hydromagnetic waves and their association with back streaming ion populations: ISEE 1 and 2 observations, *J. Geophys. Res.*, **86**, 4471, 1981.
- Kawano, H. and Higuchi, T., The bootstrap method in space physics: error estimation for the minimum-variance analysis, *Geophys. Res. Lett.*, **22**, 307, 1995.
- Kawano, H. and Higuchi, T., A generalization of the minimum variance analysis method, *Ann. Geophys.*, **14**, 1019, 1996.
- Khrabrov, A. V. and Sonnerup, B. U. Ö., Error estimates for minimum variance analysis, *J. Geophys. Res.*, **103**, 6641–6651, 1998a.
- Khrabrov, A. V. and Sonnerup, B. U. Ö., Magnetic variance analysis for small-amplitude waves and flux transfer events on a current sheet, *J. Geophys. Res.*, **103**, 11 907–11 918, 1998b.
- Khrabrov, A. V. and Sonnerup, B. U. Ö., Orientation and motion of current layers: Minimization of the Faraday residue, *Geophys. Res. Lett.*, **25**, 2372–2376, 1998c.
- Lepping, R. P. and Behannon, K. W., Magnetic field directional discontinuities: 1. minimum variance errors, *J. Geophys. Res.*, **85**, 4695, 1980.
- Lepping, R. P., Jones, J. A., and Burlaga, L. F., Magnetic field structure of interplanetary magnetic clouds at 1 AU, *J. Geophys. Res.*, **95**, 11 957, 1990.
- Paschmann, G., Papamastorakis, I., Baumjohann, W., Sckopke, N., Carlson, C. W., Sonnerup, B. U. Ö., and Lühr, H., The magnetopause for large magnetic shear: AMPTE/IRM observations, *J. Geophys. Res.*, **91**, 11 099, 1986.
- Russell, C. T. and Elphic, R. C., ISEE observations of flux transfer events at the dayside magnetopause, *Geophys. Res. Lett.*, **6**, 33, 1979.
- Siscoe, G. L., Davis, L., Coleman, Jr., P. J., Smith, E. J., and Jones, D. E., Power spectra and discontinuities of the interplanetary magnetic field: Mariner 4, *J. Geophys. Res.*, **73**, 61, 1968.
- Sonnerup, B. U. Ö., Magnetopause structure during the magnetic storm of September 24, 1961, *J. Geophys. Res.*, **76**, 6717, 1971.
- Sonnerup, B. U. Ö. and Cahill, Jr., L. J., Magnetopause structure and attitude from Explorer 12 observations, *J. Geophys. Res.*, **72**, 171, 1967.
- Sonnerup, B. U. Ö. and Cahill, Jr., L. J., Explorer 12 observations of the magnetopause current layer, *J. Geophys. Res.*, **73**, 1757, 1968.
- Sonnerup, B. U. Ö. and Guo, M., Magnetopause transects, *Geophys. Res. Lett.*, **23**, 3679, 1996.
- Sonnerup, B. U. Ö., Papamastorakis, I., Paschmann, G., and Lühr, H., Magnetopause properties from AMPTE/IRM observations of the convection electric field: Method development, *J. Geophys. Res.*, **92**, 12 137, 1987.

- Sonnerup, B. U. Ö., Papamastorakis, I., Paschmann, G., and Lühr, H., The magnetopause for large magnetic shear: Analysis of convection electric fields from AMPTE/IRM, *J. Geophys. Res.*, **95**, 10 541, 1990.
- Sonnerup, B. U. Ö., Hau, L.-N., and Walthour, D. W., On steady field aligned double-adiabatic flow, *J. Geophys. Res.*, **97**, 12 015, 1992.
- Terasawa, T., Kawano, H., Shinohara, I., Mukai, T., Saito, Y., Hoshino, M., Nishida, A., Machida, S., Nagai, T., Yamamoto, T., and Kokubun, S., On the determination of a moving MHD structure: Minimization of the residue of integrated Faraday's equation, *J. Geomagn. Geoelectr.*, **48**, 603, 1996.
- Walthour, D. W. and Sonnerup, B. U. Ö., Remote sensing of 2D magnetopause structures, in *Physics of the Magnetopause*, edited by P. Song, B. U. Ö. Sonnerup, and M. F. Thomsen, Geophysical Monograph 90, pp. 247–255, Amer. Geophys. Union, Washington D.C., 1995.

— 9 —

DeHoffmann-Teller Analysis

ALEXANDER V. KHRABROV AND BENGT U. Ö. SONNERUP

*Thayer School of Engineering, Dartmouth College
 Hanover, NH, U.S.A.*

9.1 Introduction

In experimental and theoretical studies of space plasmas, the term “deHoffmann-Teller frame”, hereafter referred to as the HT frame, has the established meaning of a galilean frame of reference in which the electric field vanishes in the plasma. It was named after the authors of the original work [*deHoffmann and Teller, 1950*] in which the concept was utilised in analysing jump conditions across MHD shock waves. The existence of an HT frame or, in the experimental context, an approximate HT frame, indicates that a coherent quasi-stationary pattern of magnetic field and plasma velocity such as a wave or current layer, is present. To first order, the observed time variation in such events would be due to the steady motion of the pattern relative to the instrument frame. The utility of an HT frame is that it allows one to identify the passage of such a moving quasi-static structure and that it facilitates further analysis and interpretation of the data. We will be referring to the procedure of identifying an HT frame as HT analysis. Its aim is to find the frame velocity vector (HT velocity), \mathbf{V}_{HT} , that best agrees with the set of measured values of magnetic field, \mathbf{B} , and electric field, \mathbf{E} , or magnetic field, \mathbf{B} , and plasma bulk velocity, \mathbf{v} , in cases where the convection electric field, $-\mathbf{v} \times \mathbf{B}$, can be used as a proxy for \mathbf{E} . Thus, the HT analysis is an attempt to characterise an experimental data set by a model that has one vector as a parameter. A natural extension of the procedure, also to be discussed here, is to allow for a constant acceleration of the HT frame as another parameter, in order to obtain a better fit. HT analysis of experimental data was first performed by *Aggson et al. [1983]* by use of an iterative method aimed at optimising the agreement between measured electric fields and the field $-\mathbf{V}_{\text{HT}} \times \mathbf{B}$. The purpose of the present chapter is to review a non-iterative, least-squares analysis technique developed by *Sonnerup et al. [1987, 1990]* and its potential extensions. Also, a discussion will be given of how the error in the experimentally determined value of \mathbf{V}_{HT} can be estimated. The techniques of the analysis, introduced step by step, will be illustrated using the same set of data as in Chapter 8 from a crossing of the magnetopause by the spacecraft AMPTE/IRM. In view of the context of this handbook, it ought to be said explicitly that, in its existing form described here, the HT analysis applies to sets of measurements acquired by a single spacecraft. However, to be eventually able to gain in-depth, quantitative understanding of the capabilities and limitations of various multiple-spacecraft techniques, it is essential to gain such understanding for single-spacecraft methods first. This chapter provides a step towards that goal.

The chapter is organised as follows. In Section 9.2, the concept of an HT frame and the circumstances in which it may exist, are outlined. Section 9.3 contains a review of the procedure for finding an approximate value of \mathbf{V}_{HT} from a set of measurements of magnetic field and plasma bulk-velocity vectors or magnetic and electric field vectors. The procedure is based on minimising the residual electric field in the least-squares sense. The use of electron fluid velocities in place of the ion velocities is also discussed briefly. In Section 9.4, where much of the practical work with actual data is performed, an assessment is made of the uncertainty margins of the deHoffmann-Teller velocity vector obtained from the minimisation procedure. The determination of uncertainty margins constitutes a new and not yet firmly established element of the HT analysis. Analytical error estimates are explored as well as those offered by the bootstrap approach that utilises multiple resamplings of the data set. The discussion of errors is also coupled with use of the HT analysis on multiple, nested data intervals to test time stationarity. The choice of an optimal interval is one of the topics addressed there. In Section 9.5, an extension of the basic procedure is presented that allows for a steadily accelerating HT frame. The mean square of the residual electric field is minimised with respect to two parameters, \mathbf{V}_{HT0} and \mathbf{a}_{HT} , which accounts for a linear time-variation of the frame velocity. This version of the method is relevant, for example, when the motion studied is oscillatory, as is frequently the case for the magnetopause. Sections 9.6 and 9.7 contain discussion and summary.

The following connections exist between this and other chapters in this book. There is a close commonality between the methodology of this chapter and Chapter 8 on minimum/maximum variance analysis. Also, in relation to shock fronts, the HT frame is discussed in Chapter 10, where, essentially, the vector component of \mathbf{V}_{HT} tangential to the front is considered.

9.2 Existence and Properties of HT Frame

If the electric field measured in the instrument frame is \mathbf{E} , then the electric field in the HT frame, assuming such a frame exists, is

$$\mathbf{E}' = \mathbf{E} + \mathbf{V}_{\text{HT}} \times \mathbf{B} = 0 \quad (9.1)$$

It follows from Faraday's law, expressed in the HT frame, that $\nabla \times \mathbf{E}' = -(\partial \mathbf{B} / \partial t)' = 0$, where the prime denotes quantities evaluated in the HT frame. In other words, the existence of the HT frame implies that the magnetic field structure sampled is stationary when viewed in that frame. However, the reverse claim cannot be made: many time-independent structures exist for which no frame of reference can be found in which $\mathbf{E}' \equiv 0$. In particular, a strictly perpendicular shock does not possess an HT frame: no single frame of reference can be found in which the electric field vanishes on both sides of the shock. The same situation arises for a tangential discontinuity having the special property that the magnetic fields on its two sides are parallel or anti-parallel when the velocities transverse to these fields are different. In both examples, a "proper" (but not unique) frame of reference exists in which the current layer is stationary but these frames are not HT frames and cannot be obtained by use of the methods developed in the main part of this chapter. All other one-dimensional time-independent current layers, whether they are shocks, tangential or rotational discontinuities, can be transformed to a "proper" frame of reference in which the electric fields in the regions of uniform field and flow on the two

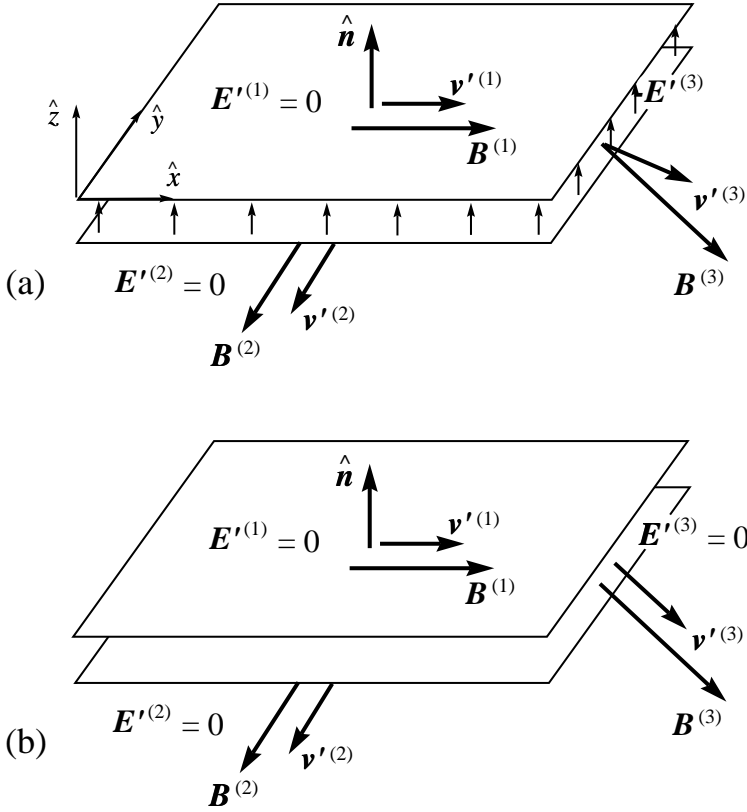


Figure 9.1: Magnetic field and velocity in the “proper” frame of a time-independent tangential discontinuity structure (a) when an intrinsic electric field $E'(3)$ exists in the interior of the layer; (b) when $E'(3) = 0$ so that the proper frame is also a true HT frame.

sides of the layer now vanish but where, nevertheless, an intrinsic electric field component along the normal to the layer may remain in its interior. This proper frame is unique (except in certain special cases) and is useful for observational and theoretical studies of the layer. It can be viewed as the HT frame for the layer when its structure is left unresolved. But, strictly speaking, it is not an HT frame when the structure is included. A simple illustration in the form of a tangential discontinuity is shown in Figure 9.1. It can be argued that, for discontinuities having an intrinsic electric field along the normal, it makes sense to perform HT analysis in which data points from the interior of the layer are excluded.

We emphasise that the existence of an HT frame is by no means limited to one-dimensional discontinuities. It has been found to exist for data collected during distant as well as close encounters with flux transfer events at the magnetopause [Papamastorakis *et al.*, 1989; Walthour and Sonnerup, 1995] and during crossings of magnetopause current layers having significant internal 2-D or 3-D structure [Sonnerup and Guo, 1996]. Applications to travelling waves, plasmoids, current filaments and vortices appear feasible as well.

From an observational viewpoint it is interesting to ask what is the minimal amount of measured information needed in order to exclude the existence of an HT frame. A single measured electromagnetic vector pair, $\{\mathbf{E}^{(1)}, \mathbf{B}^{(1)}\}$ is not sufficient because, from equation 9.1, one may then determine that

$$\mathbf{V}_{\text{HT}} = \frac{\mathbf{E}^{(1)} \times \mathbf{B}^{(1)}}{|\mathbf{B}^{(1)}|^2} + V_{\parallel} \frac{\mathbf{B}^{(1)}}{|\mathbf{B}^{(1)}|} \quad (9.2)$$

where V_{\parallel} is an arbitrary velocity component along $\mathbf{B}^{(1)}$. Thus an HT frame exists but is not unique. If two distinct measured electromagnetic vector pairs $\{\mathbf{E}^{(1)}, \mathbf{B}^{(1)}\}$ and $\{\mathbf{E}^{(2)}, \mathbf{B}^{(2)}\}$ are available, equation 9.1 gives two conditions:

$$\left. \begin{aligned} \mathbf{E}^{(1)} + \mathbf{V}_{\text{HT}} \times \mathbf{B}^{(1)} &= 0 \\ \mathbf{E}^{(2)} + \mathbf{V}_{\text{HT}} \times \mathbf{B}^{(2)} &= 0 \end{aligned} \right\} \quad (9.3)$$

These two equations are generally inconsistent, as can be seen from the following argument. Without loss of generality, we may assume that the vectors $\mathbf{B}^{(1)}$ and $\mathbf{B}^{(2)}$ lie in the xy plane. Then the dot products of $\hat{\mathbf{z}}$, the unit vector perpendicular to that plane, with equations 9.3 give

$$\left. \begin{aligned} \mathbf{E}^{(1)} \cdot \hat{\mathbf{z}} &= \mathbf{V}_{\text{HT}} \cdot (\hat{\mathbf{z}} \times \mathbf{B}^{(1)}) \\ \mathbf{E}^{(2)} \cdot \hat{\mathbf{z}} &= \mathbf{V}_{\text{HT}} \cdot (\hat{\mathbf{z}} \times \mathbf{B}^{(2)}) \end{aligned} \right\} \quad (9.4)$$

Assuming that $\mathbf{B}^{(1)}$ and $\mathbf{B}^{(2)}$ are not parallel, these two equations allow the two components of \mathbf{V}_{HT} in the xy plane to be uniquely determined. The cross product of $\hat{\mathbf{z}}$ with equation 9.3 gives

$$\left. \begin{aligned} \mathbf{B}^{(1)} (\mathbf{V}_{\text{HT}} \cdot \hat{\mathbf{z}}) &= \hat{\mathbf{z}} \times \mathbf{E}^{(1)} \\ \mathbf{B}^{(2)} (\mathbf{V}_{\text{HT}} \cdot \hat{\mathbf{z}}) &= \hat{\mathbf{z}} \times \mathbf{E}^{(2)} \end{aligned} \right\} \quad (9.5)$$

These two equations require $\mathbf{E}^{(1)} \cdot \mathbf{B}^{(1)} = 0$, and $\mathbf{E}^{(2)} \cdot \mathbf{B}^{(2)} = 0$, i.e., the two measured electric field vectors must be perpendicular to their corresponding measured magnetic field vectors, otherwise an HT frame does not exist (see also equation 9.1). Furthermore, assuming these conditions on \mathbf{E} to be met, the two equations 9.5 give a single value of the third component of \mathbf{V}_{HT} , namely

$$\mathbf{V}_{\text{HT}} \cdot \hat{\mathbf{z}} = \frac{\hat{\mathbf{z}} \cdot (\mathbf{E}^{(1)} \times \mathbf{B}^{(1)})}{|\mathbf{B}^{(1)}|^2} = \frac{\hat{\mathbf{z}} \cdot (\mathbf{E}^{(2)} \times \mathbf{B}^{(2)})}{|\mathbf{B}^{(2)}|^2} \quad (9.6)$$

if, and only if, the two right-hand members of equation 9.6 are the same. This compatibility condition expresses an additional requirement for the existence of an HT frame, namely that the z component of the electric drift velocity must be the same for the two measured vector pairs $\{\mathbf{E}^{(1)}, \mathbf{B}^{(1)}\}$ and $\{\mathbf{E}^{(2)}, \mathbf{B}^{(2)}\}$.

The above conditions for the existence of an HT frame are restrictive but, with the exceptions noted earlier, they are always satisfied if the vector pairs are measured in the two uniform-field regions on opposite sides of a one-dimensional discontinuity having time-independent structure but being perhaps in a state of motion. For a tangential discontinuity $\hat{\mathbf{z}}$ is along the normal, $\hat{\mathbf{n}}$, to the discontinuity; for a shock it is tangential to the shock surface, $\hat{\mathbf{z}} \cdot \hat{\mathbf{n}} = 0$, as a result of the coplanarity condition ($\hat{\mathbf{n}}, \mathbf{B}^{(1)}$ and $\mathbf{B}^{(2)}$ are coplanar); for a rotational discontinuity $\hat{\mathbf{z}}$ is at an intermediate angle to the normal. The component

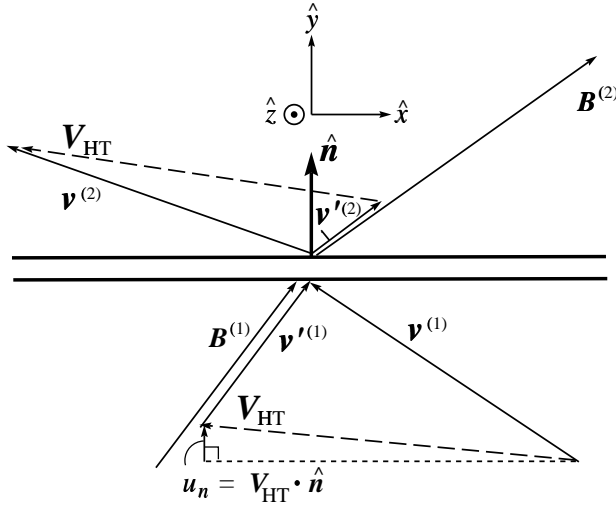


Figure 9.2: Planar flow and field configuration at a shock, showing measured magnetic fields, $\mathbf{B}^{(1)}$ and $\mathbf{B}^{(2)}$, as well as measured velocities, $\mathbf{v}^{(1)}$ and $\mathbf{v}^{(2)}$, in the spacecraft frame. Also shown is the HT-frame velocity, \mathbf{V}_{HT} , and its normal component, $\mathbf{V}_{\text{HT}} \cdot \hat{\mathbf{n}} = u_n$, which represents the motion of the shock. The plasma velocities $\mathbf{v}'^{(1)}$ and $\mathbf{v}'^{(2)}$ are evaluated in the HT frame.

of \mathbf{V}_{HT} along $\hat{\mathbf{n}}$, $\mathbf{V}_{\text{HT}} \cdot \hat{\mathbf{n}}$, represents the velocity of motion of the discontinuity along its normal. A simple planar illustration for a shock-like structure is shown in Figure 9.2.

We note that $\hat{\mathbf{n}}$ itself can in general not be obtained from the two measured vector pairs: the constancy of the normal magnetic field component, $\mathbf{B}^{(1)} \cdot \hat{\mathbf{n}} = \mathbf{B}^{(2)} \cdot \hat{\mathbf{n}}$, plus the normalisation $|\hat{\mathbf{n}}|^2 = 1$ provides only two equations for the three unknown components of $\hat{\mathbf{n}}$. The constancy of the tangential electric field in a frame moving with velocity $u_n \hat{\mathbf{n}}$ cannot be used to provide the missing equation because it is guaranteed by the existence of \mathbf{V}_{HT} . As discussed in Chapter 8, at least three distinct measurements of \mathbf{B} are needed to obtain $\hat{\mathbf{n}}$, except for non-perpendicular shocks, where coplanarity (see Section 10.4.2 of Chapter 10) provides the missing condition, and for tangential discontinuities, where $\mathbf{B} \cdot \hat{\mathbf{n}} = 0$ provides the missing information. We conclude that, with the exceptions noted already, two pairs of vectors \mathbf{E} and \mathbf{B} , measured on opposite sides of a discontinuity and satisfying equation 9.6, permit determination of \mathbf{V}_{HT} . But because $\hat{\mathbf{n}}$ is not obtained, the nature of the discontinuity cannot be established without additional information.

When more than two distinct electromagnetic vector sets $\{\mathbf{E}, \mathbf{B}\}$ are used in the analysis, the conditions discussed above must hold pairwise for all possible combinations of measured vector sets. This is a rather restrictive set of requirements that is unlikely to be satisfied as a chance occurrence. For real measured data sets, it is indeed unlikely ever to be precisely satisfied; this is the situation to be dealt with in the remainder of this chapter. If it is found to be reasonably well met, then it is likely that a quasi-stationary magnetic structure moving past the observing spacecraft has been sampled. When the data to be analysed have been collected within regions and structures where ideal MHD is approxi-

mately valid so that

$$\mathbf{E} + \mathbf{v} \times \mathbf{B} \cong 0 \quad (9.7)$$

then the convective electric field, $-\mathbf{v} \times \mathbf{B}$, calculated from measured values of the plasma velocity, \mathbf{v} , and magnetic field, \mathbf{B} , can be used as a proxy for the electric field. This approach can be taken when \mathbf{E} is not available or is not measured with sufficient accuracy. To date, this has been the most common situation and it is therefore the one we will address first, in Section 9.3.

9.3 Determination of V_{HT} from Experimental Data

9.3.1 Minimisation of Residual Electric Field

To obtain an approximation to V_{HT} from a set of experimental measurements of plasma bulk velocity, $\mathbf{v}^{(m)}$, and magnetic field, $\mathbf{B}^{(m)}$, $m = 1, 2 \dots M$, one may seek a reference frame in which the mean square of the electric field is as small as possible for the given set of M measurements. Denoting this quantity by $D(\mathbf{V})$, which is given by

$$D(\mathbf{V}) = \frac{1}{M} \sum_{m=1}^M \left| \mathbf{E}'^{(m)} \right|^2 = \frac{1}{M} \sum_{m=1}^M \left| (\mathbf{v}^{(m)} - \mathbf{V}) \times \mathbf{B}^{(m)} \right|^2 \quad (9.8)$$

the HT velocity is the value of the frame velocity, \mathbf{V} , that minimises D . We will use the symbol V_{HT} to designate this approximate value. As a function of its vector argument, the quantity D is a non-negative quadratic form which therefore must have a unique minimum. The minimisation condition $\nabla_{\mathbf{V}} D = 0$ leads, after straight-forward analysis, to the following linear equation for V_{HT} :

$$\mathbf{K}_0 \mathbf{V}_{\text{HT}} = \left\langle \mathbf{K}^{(m)} \mathbf{v}^{(m)} \right\rangle \quad (9.9)$$

Assuming \mathbf{K}_0 to be non-singular (see Section 9.4.1), the solution is

$$\mathbf{V}_{\text{HT}} = \mathbf{K}_0^{-1} \left\langle \mathbf{K}^{(m)} \mathbf{v}^{(m)} \right\rangle \quad (9.10)$$

In these expressions, each $\mathbf{K}^{(m)}$ is the matrix of projection, $\mathbf{P}^{(m)}$, into a plane perpendicular to $\mathbf{B}^{(m)}$, multiplied by $B^{(m)2}$:

$$\mathbf{K}_{\mu\nu}^{(m)} = B^{(m)2} \left(\delta_{\mu\nu} - \frac{B_{\mu}^{(m)} B_{\nu}^{(m)}}{B^{(m)2}} \right) \equiv B^{(m)2} \mathbf{P}_{\mu\nu}^{(m)} \quad (9.11)$$

The angle brackets $\langle \dots \rangle$ denote an average of an enclosed quantity over the set of M measurements, and $\mathbf{K}_0 \equiv \langle \mathbf{K}^{(m)} \rangle$. The formula 9.10 gives V_{HT} in terms of the measured quantities, $\mathbf{v}^{(m)}$, and $\mathbf{B}^{(m)}$.

Now we apply the procedure to an experimental data set, namely the magnetopause crossing by the AMPTE/IRM spacecraft on October 19, 1984, 05:18:53 UT. This event was originally examined by *Sonnerup et al. [1990]* and is also discussed in detail in Chapter 8, where Table 8.4 (page 218) lists the plasma and magnetic data for benchmarking purposes. Here, we concentrate on the HT analysis. The mid-plane of the current layer

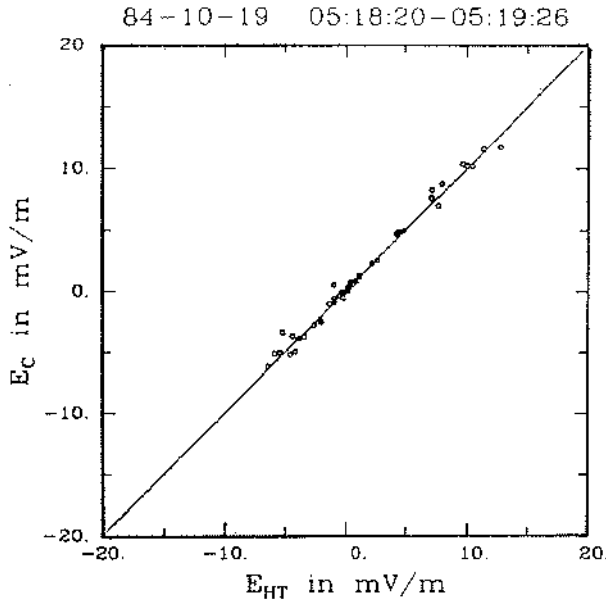


Figure 9.3: Scatter plot of the GSE components of $\mathbf{E}^{(m)} = -\mathbf{v}^{(m)} \times \mathbf{B}^{(m)}$ versus the corresponding components of $\mathbf{E}_{HT}^{(m)} = -V_{HT} \times \mathbf{B}^{(m)}$ for AMPTE/IRM magnetopause crossing. The correlation coefficient is 0.994.

corresponds to approximately 05:18:53 UT. The time interval chosen for the analysis of the event by *Sonnerup et al.* was from 05:18:20 to 05:19:26 UT. It comprised 16 measurements with an average sampling interval of 4.35 seconds, which was the time resolution of the plasma data. For the magnetic field, sampled with much higher time resolution than the particle distribution functions, spin-averaged values corresponding to the plasma data were used. Applying the formula 9.10 to this interval gives the following components of V_{HT} in GSE coordinates:

⇒9.1

$$V_{HT} = (-122.8, -223.0, 76.0) \text{ km/s}$$

This frame velocity is nearly tangential to the magnetopause: its component along the outward directed magnetopause normal, $\hat{\mathbf{n}} = (0.8671, -0.4978, 0.0187)$ in GSE, taken from Figure 8.2 (page 190) of Chapter 8, is $u_n = +5.9 \text{ km/s}$. It is noted that the true value of u_n must in fact be negative because the spacecraft made a transition from the magnetosphere to the magnetosheath. Only very small corrections in the estimated value of V_{HT} and/or $\hat{\mathbf{n}}$ are required in order to remove this discrepancy.

To give an impression of the quality of the HT frame, the two electric fields $\mathbf{E}_c^{(m)} = -\mathbf{v}^{(m)} \times \mathbf{B}^{(m)}$ and $\mathbf{E}_{HT}^{(m)} = -V_{HT} \times \mathbf{B}^{(m)}$ are plotted against each other, component by component, in Figure 9.3. The correlation between these two fields is seen to be very good: the correlation coefficient is $c = 0.994$. Another measure of the quality of the HT frame is the ratio $D(V_{HT})/D(0)$ which is found to have the value 0.012 for this event. Note that, since these quality measures are frame dependent, they are not suitable for inter-comparison of the quality of the HT frames of different events. The quality parameters c

and $D(V_{\text{HT}})/D(0)$ are related by the formula

$$c^2 + \frac{D(V_{\text{HT}})}{D(0)} = 1$$

which follows from equation 9.10.

If directly measured electric field data, $\mathbf{E}^{(m)}$, are used in place of $-\mathbf{v}^{(m)} \times \mathbf{B}^{(m)}$ for determination of V_{HT} , the formula 9.10 is replaced by

$$V_{\text{HT}} = K_0^{-1} \left\langle \mathbf{E}^{(m)} \times \mathbf{B}^{(m)} \right\rangle \quad (9.12)$$

Finally, for structures described by ideal Hall MHD, the generalised form of Ohm's law yields

$$\frac{1}{n_e e} \nabla p_e + \mathbf{E} + \mathbf{v}_e \times \mathbf{B} = 0 \quad (9.13)$$

where p_e is the electron pressure, n_e is the number density of electrons, e is the magnitude of the electron charge, and \mathbf{v}_e is the electron fluid velocity. It is seen from this expression that $-\mathbf{v}_e \times \mathbf{B}$ can now be used as a proxy, not for \mathbf{E} itself, but for $\tilde{\mathbf{E}} \equiv (\mathbf{E} + (\nabla p_e / n_e e))$. Thus, if accurate measurements of \mathbf{v}_e are available, the calculation of a frame velocity by replacing \mathbf{v} by \mathbf{v}_e in equation 9.10 or 9.14 becomes meaningful. It leads to a modified HT frame in which $\langle \tilde{\mathbf{E}}'^2 \rangle$ is made as small as possible. Such minimisation reflects the fact that in a simple two-fluid model, the magnetic field is frozen into the electron fluid, rather than the ion fluid. For a one-dimensional discontinuity having $\langle \mathbf{B} \rangle \cdot \hat{\mathbf{n}} \neq 0$, the intrinsic normal electric field mentioned in Section 9.2 comes exclusively from the ∇p_e term in equation 9.13. Therefore, an HT-frame determination based on \mathbf{v}_e will remove the influence of the intrinsic electric field on the frame velocity. It will produce the velocity of the ‘‘proper’’ frame of the discontinuity, mentioned in Section 9.2.

We mention the following link to Chapter 10. Suppose that one applies equation 9.10 to two ($M = 2$) \mathbf{E} , \mathbf{B} pairs measured on opposite sides of either a shock or a tangential discontinuity with a known orientation. Then, equation 10.6 (page 252 of Chapter 10) can be obtained by projecting the two sides of our equation 9.10 onto the discontinuity plane.

9.3.2 Minimisation of Residual Cross-Field Velocity

In principle, the best-fit approach to finding an approximation to V_{HT} can employ any merit function that is non-negative and whose value would be zero for a perfect HT frame. For instance, one can minimise the mean square of the cross-field plasma velocity, rather than the electric field, in the moving frame:

$$Q(\mathbf{V}) = \frac{1}{M} \sum_{m=1}^M \left(v_{\perp}^{(m)} \right)^2 = \left\langle \left(\frac{E_{\perp}^{(m)}}{B^{(m)}} \right)^2 \right\rangle = \left\langle \frac{|(\mathbf{v}^{(m)} - \mathbf{V}) \times \mathbf{B}^{(m)}|^2}{B^{(m)2}} \right\rangle \quad (9.14)$$

The process of minimising Q with respect to \mathbf{V} differs from that for D in a straightforward way: the factor multiplying the orthogonal projection operator in each matrix $K^{(m)}$ (see equation 9.11) would be unity instead of $|\mathbf{B}^{(m)}|^2$. In other words, $K^{(m)}$ and K_0 are replaced by $\mathbf{P}^{(m)}$ and $\mathbf{P}_0 \equiv \langle \mathbf{P}^{(m)} \rangle$ in equation 9.10. For the case at hand, the minimisation of Q yields

$$V_{\text{HT}} = (-120.9, -224.1, 72.9) \text{ km/s}$$

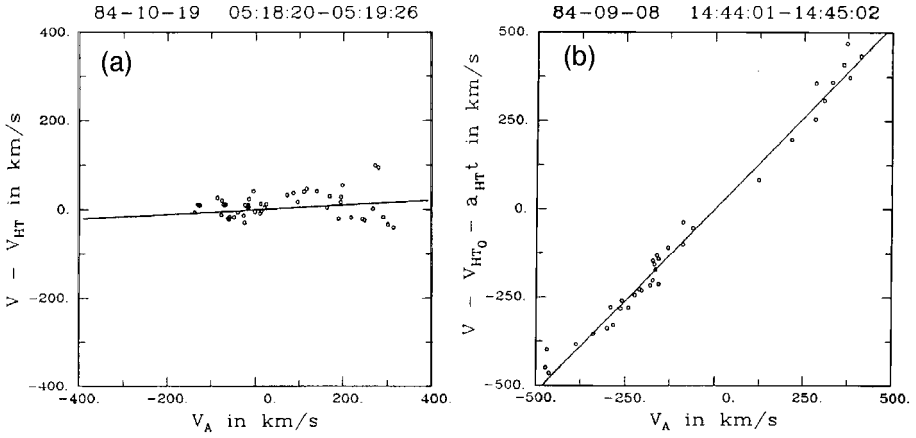


Figure 9.4: (a) Walén scatter plot of the GSE components of $\mathbf{v}'^{(m)} = (\mathbf{v}^{(m)} - \mathbf{V}_{HT})$, the plasma velocity in the HT frame, versus the corresponding components of the Alfvén velocity for AMPTE/IRM magnetopause crossing on October 19, 1984; (b) same for crossing on September 8, 1984 but now including acceleration of HT frame as discussed in Section 9.5.

and $u_n = +8.1$ km/s. Since u_n must be negative for this event, the latter result is somewhat further removed from the true value of u_n than is the outcome of the D minimisation: $u_n = +5.9$ km/s.

If a good HT frame exists for an observed MHD structure, one would expect different merit functions to produce approximations to \mathbf{V}_{HT} that agree within their error bounds. The HT frame is never perfect for any observed phenomenon, and the question arises how error bounds can be placed on the value of \mathbf{V}_{HT} obtained from the minimisation procedure. This issue will be examined and illustrated in Section 9.4.

9.3.3 Walén Relation in HT Frame

To illustrate the utility of the HT frame for purposes of data interpretation, we now examine the plasma velocities, $\mathbf{v}'^{(m)} \equiv \mathbf{v}^{(m)} - \mathbf{V}_{HT}$, in the HT frame and their relation to the local measured Alfvén velocities $\mathbf{V}_A^{(m)} = \mathbf{B}^{(m)}(\mu_0\rho^{(m)})^{-1/2}$, where $\rho^{(m)}$ is the measured mass density (assuming all particles to be H^+). A component-by-component scatter plot of these two velocities, referred to as a Walén plot, is shown in Figure 9.4a. It is seen that the plasma velocities remaining in the HT frame are relatively small compared to $|\mathbf{V}_{HT}|$ itself and, except near $V_A = 0$, compared to $|\mathbf{V}_A|$. This behaviour is in sharp contrast to magnetopause reconnection events where $\mathbf{v}'^{(m)}$ is typically proportional to $\mathbf{V}_A^{(m)}$ with a constant of proportionality that is often in the range $\pm(0.8 - 1.0)$. Such proportionality is interpreted as an indication that the magnetopause had the structure of a large-amplitude Alfvén wave or rotational discontinuity for which $\langle \mathbf{B}^{(m)} \rangle \cdot \hat{\mathbf{n}} \neq 0$ and $\mathbf{v}'^{(m)} = \pm \mathbf{V}_A^{(m)}$, where the sign on the right-hand side is the same as the sign of the product $(\langle \mathbf{v}'^{(m)} \rangle \cdot \hat{\mathbf{n}})(\langle \mathbf{B}^{(m)} \rangle \cdot \hat{\mathbf{n}})$. For such structures, the data points in the scatter plot fall along one of the diagonals, as illustrated in Figure 9.4b. The comparison of measured

data to this predicted behaviour has become known as the Walén test. For the data set in Figure 9.4a, the Walén test is not successful: instead of falling along one of the diagonals, the data are clustered around $\mathbf{v}'^{(m)} = 0$. The conclusion can be drawn that the magnetopause structure examined here was that of a tangential discontinuity, $\langle \mathbf{B}^{(m)} \rangle \cdot \hat{\mathbf{n}} \cong 0$, rather than that of a rotational discontinuity (or shock). But the current layer did not have purely one-dimensional structure. As shown in Figure 8.2, $\langle \mathbf{B}^{(m)} \rangle \cdot \hat{\mathbf{n}}$ was indeed nearly zero but substantial fluctuations in the individual values, $\mathbf{B}^{(m)} \cdot \hat{\mathbf{n}}$, were present. Because a good HT frame exists, the majority of these fluctuations can be interpreted as being caused by 2-D or 3-D quasi-stationary structures moving past the spacecraft with the HT velocity. Furthermore, since $\mathbf{v}'^{(m)}$ is small, this motion is approximately describable as frozen convection of magnetohydrostatic structures with the moving plasma. Details of these structures have been examined recently by *Sonnerup and Guo* [1996]. It is noted that the tangential discontinuity structure observed for the October 19, 1984, event (see Figures 8.2, 9.3, and 9.4a) is a good illustration of a point first emphasised by *Paschmann* [1985]: from the existence of a good HT frame for a magnetopause crossing, it does not follow that the structure observed is that of a rotational discontinuity.

9.4 Uncertainty in V_{HT} Determination

9.4.1 Analytical Error Estimates

To produce a simple quantitative statement about the uncertainty of V_{HT} calculated from the data, one can start by adopting the hypothesis that the imperfection of the HT-frame is due to some kind of random fluctuations whose probabilities are time-stationary and statistically independent between different measurements in the data series. A natural approach would be to treat the residual electric field $\mathbf{E}'^{(m)} = -(\mathbf{v}^{(m)} - \mathbf{V}_{\text{HT}}) \times \mathbf{B}^{(m)}$ itself as such a random quantity. In other words, we assume that the values of the residual electric field, calculated for a given data set, are samples of stationary noise with the following properties: (i) $\mathbf{E}'^{(m)}$, which is perpendicular to $\mathbf{B}^{(m)}$, has isotropic probability distribution in the plane normal to $\mathbf{B}^{(m)}$, and (ii) there is no correlation between values of $\mathbf{E}'^{(m)}$ at different locations and times. Given (i) and (ii), the average expected electrical noise level $\langle \langle E'^2 \rangle \rangle$ is approximated by the minimum value of D , computed for a sufficiently large data set. Here the double angle brackets $\langle \langle \cdot \cdot \rangle \rangle$ indicate an ensemble average over a large number of imagined repetitive realisations of the data set. Further, under the assumptions made, the uncertainty in the experimental determination of V_{HT} can be characterised by the variance matrix \mathbf{S} , defined by $S_{\mu\nu} = \langle \langle \Delta V_{\text{HT}\mu} \Delta V_{\text{HT}\nu} \rangle \rangle$, where the estimate of the fluctuations, $\Delta \mathbf{V}_{\text{HT}}$, for each member of the ensemble is obtained by differentiating equation 9.10. In this operation, variations in $\mathbf{B}^{(m)}$ and \mathbf{K}_0 are not included but only those in $\mathbf{E}^{(m)}$. This restriction is not strictly justified but is motivated by a desire to obtain a sufficiently simple result so as to be of practical use. We then obtain

$$\Delta \mathbf{V}_{\text{HT}} = \mathbf{K}_0^{-1} \left\langle \Delta \mathbf{E}^{(m)} \times \mathbf{B}^{(m)} \right\rangle \quad (9.15)$$

where $\Delta \mathbf{E}^{(m)} = \mathbf{E}'^{(m)} = -(\mathbf{v}^{(m)} - \mathbf{V}_{\text{HT}}) \times \mathbf{B}^{(m)}$. The process of evaluating the ensemble average $S_{\mu\nu} = \langle \langle \Delta V_{\text{HT}\mu} \Delta V_{\text{HT}\nu} \rangle \rangle$, using equation 9.15 to express $\Delta \mathbf{V}_{\text{HT}}$, is somewhat

lengthy; it is summarised in the Appendix. The result is simple, namely

$$S_{\mu\nu} = \frac{D(V_{\text{HT}})}{2M-3} K_0^{-1}{}_{\mu\nu} \quad (9.16)$$

For any chosen spatial direction, specified by a unit vector $\hat{\mathbf{n}}$, the variance of the velocity component $V_{\text{HT}} \cdot \hat{\mathbf{n}}$ is

$$\sigma_n^2 = n_\mu S_{\mu\nu} n_\nu \quad (9.17)$$

The eigenvectors, \mathbf{e}_i , of the variance matrix \mathbf{S} are the same as those of \mathbf{K}_0 , and the respective eigenvalues are

$$\Lambda_i = \frac{D(V_{\text{HT}})}{(2M-3)\lambda_i}$$

where λ_i are the eigenvalues of \mathbf{K}_0 . The eigenvalues Λ_i are the variances of the components of V_{HT} along \mathbf{e}_i .

It is seen from equation 9.16 that the matrix \mathbf{K}_0 , which appears in connection with minimising D , also characterises the directional properties of the uncertainty of the HT velocity. As an example, let us think of a situation where the observed magnetic field \mathbf{B} consists of a constant field, \mathbf{B}_0 , and a small perturbation, a situation that arises in the remote sensing of travelling bulges on the magnetopause [Walthour and Sonnerup, 1995]. We recall that each individual matrix $\mathbf{K}^{(m)}$ is proportional to the operator of projection on a plane perpendicular to the corresponding $\mathbf{B}^{(m)}$. Therefore, for each matrix $\mathbf{K}^{(m)}$, the field vector $\mathbf{B}^{(m)}$ is an eigenvector with its corresponding eigenvalue $\lambda_1^{(m)} = 0$. Further, for the slightly perturbed uniform field we consider, the directional variation of \mathbf{B} is small. Hence, one of the eigenvectors of the resulting matrix $\mathbf{K}_0 = \langle \mathbf{K}^{(m)} \rangle$ would be close in direction to \mathbf{B}_0 , and the corresponding eigenvalue, λ_1 , would be much smaller than the other two (both of which would have values close to B_0^2). In other words, \mathbf{K}_0 would be nearly singular. Consequently, the largest standard deviation of V_{HT} would be in the direction nearly along \mathbf{B}_0 : $\Lambda_1 = D(V_{\text{HT}})/(2M-3)\lambda_1$. This large uncertainty in the component of V_{HT} along \mathbf{B}_0 is directly related to the arbitrary parallel velocity component, V_{\parallel} , in equation 9.2. The described directional distribution of the uncertainty should exist also when the magnetic field is of a varying magnitude but still approximately unidirectional, like in a one-dimensional current layer with no field rotation. This situation might be encountered in the HT analysis of moving structures in the magnetotail.

To further describe the domain of uncertainty for V_{HT} , we reduce the quadratic form D to the principal-axes representation. First, we notice that $D(V)$ can be written as

$$D = D(V_{\text{HT}}) + \delta V \cdot \mathbf{K}_0 \cdot \delta V$$

where $\delta V = V - V_{\text{HT}}$. Hence the matrix of the quadratic form D is \mathbf{K}_0 , whose eigenvectors specify the principal axes. In the eigenbasis, $\{\mathbf{e}_i\}$, the form becomes

$$D = D(V_{\text{HT}}) + \sum_{i=1}^3 \lambda_i (\delta V_i)^2 \quad (9.18)$$

In V space, we define an ellipsoid with principal axes directed along \mathbf{e}_i , and with the semiaxes being the corresponding standard deviations, σ_i , for V_{HT} :

$$\sigma_i = \sqrt{\Lambda_i} = \sqrt{\frac{D(V_{\text{HT}})}{(2M-3)\lambda_i}} \quad (9.19)$$

Table 9.1: HT velocities and principal axes of error ellipsoids in velocity space for magnetopause crossing by AMPTE/IRM on October 19, 1984 (05:18:20–05:19:26 UT)

GSE Components					MVAB* Components		
	$\mathbf{V}_{\text{HT}}^{\dagger}$	$(V_{\text{HT}x}$	$V_{\text{HT}y}$	$V_{\text{HT}z})$	$(\mathbf{V}_{\text{HT}} \cdot \mathbf{x}_1$	$V_{\text{HT}} \cdot \mathbf{x}_2$	$\mathbf{V}_{\text{HT}} \cdot \mathbf{x}_3)$
i	$\sqrt{\Lambda_i}^{\dagger}$	e_{xi}	e_{yi}	e_{zi}	$\mathbf{e}_i \cdot \mathbf{x}_1$	$\mathbf{e}_i \cdot \mathbf{x}_2$	$\mathbf{e}_i \cdot \mathbf{x}_3$
Min. D:	\mathbf{V}_{HT}	(−122.8	−223.0	76.0)	(248.5	93.8	5.9)
1	7.62	−0.2457	−0.3726	0.8949	0.8967	−0.4427	−0.0108
2	4.59	0.4225	0.7897	0.4448	−0.4428	−0.8965	−0.0184
3	3.94	0.8724	−0.4874	0.0366	0.0015	−0.0213	0.9998
Min. Q:	\mathbf{V}_{HT}	(−120.9	−224.1	72.9)	(246.6	96.2	8.1)
1	8.19	−0.1870	−0.2206	0.9616	0.7956	−0.6043	−0.0443
2	4.85	0.4771	0.8371	0.2675	−0.6048	−0.7964	0.0020
3	4.21	0.8587	−0.5089	0.0609	0.0364	−0.0252	0.9990

† km/s

*MVAB components refer to principal axes derived from magnetic variance matrix

$M_{\mu\nu}^B = \langle B_{\mu} B_{\nu} \rangle - \langle B_{\mu} \rangle \langle B_{\nu} \rangle$ (See Figure 8.2 on page 190 of Chapter 8)

The equation of this ellipsoid of uncertainty is

$$\sum_{i=1}^3 \frac{(\delta V_i)^2}{\Lambda_i} = 1 \quad (9.20)$$

or, using the relation between the eigenvalues, Λ_i , of \mathbf{S} and the corresponding eigenvalues λ_i of \mathbf{K}_0 ,

$$\sum_{i=1}^3 \lambda_i (\delta V_i)^2 = \frac{D(\mathbf{V}_{\text{HT}})}{2M-3} \quad (9.21)$$

We see from equations 9.18 and 9.21 that the ellipsoid of uncertainty belongs to the family of ellipsoids $D = \text{constant}$, the particular value of the constant being $D(\mathbf{V}_{\text{HT}})[1+1/(2M-3)]$.

To illustrate the arbitrariness involved in the analysis of errors outlined here, we notice that instead of \mathbf{E}' , another vector—for instance, the residual cross-field plasma velocity \mathbf{v}'_{\perp} —can be treated as a stationarily distributed random quantity. In that case, the magnitude of $\mathbf{E}'^{(m)}$ would not be independent of $\mathbf{B}^{(m)}$, but would scale proportionally to $B^{(m)}$. Recalling the discussion in Section 9.3, a calculation of this type appears natural when Q , the mean square of the residual perpendicular velocity (equation 9.14) is chosen for minimisation. In analogy with the preceding development for D , the variance matrix of values of \mathbf{V}_{HT} , computed by minimising Q , is

$$S_{\mu\nu} = \frac{Q(\mathbf{V}_{\text{HT}})}{2M-3} P_{0\mu\nu}^{-1} \quad (9.22)$$

where $\mathbf{P}_0 \equiv \langle \mathbf{P}^{(m)} \rangle$. Although the analytical error estimates depend on somewhat arbitrary hypotheses concerning the nature of the underlying fluctuations, such estimates should

be useful when a good HT-frame exists. We expect all estimates of the kind presented above to produce error margins of comparable magnitude; this is indeed observed in our worked example which is summarised in Table 9.1. It is also seen from this table that e_3 , the direction of minimum uncertainty, is nearly aligned with $x_3 = \hat{n}$, the magnetopause normal from MVAB. Therefore the two values of $u_n = V_{\text{HT}} \cdot \hat{n}$ obtained from minimisation of $D(V_{\text{HT}})$ and $Q(V_{\text{HT}})$ can now be given the following ranges of uncertainty: $u_n = (5.9 \pm 3.9)$ km/s from $D(V_{\text{HT}})$ and $u_n = (8.1 \pm 4.2)$ km/s from $Q(V_{\text{HT}})$. Thus the two estimates of u_n are mutually consistent but both are inconsistent with the true value of u_n , which must be negative for this crossing.

The V_{HT} vectors and analytical error estimates in Table 9.1 will also be compared to V_{HT} and uncertainties predicted by the bootstrap method in Section 9.4.2.

9.4.2 Bootstrap Error Estimates

Another means of estimating the error bounds is offered by resampling methods. Unlike the analytical estimates, methods of this class are aimed at recovering the actual probability distribution of the parameter being estimated (the HT velocity vector in the present case) by creating a large number of simulated realisations of the data set. Also, these direct methods do not rely on linearising the problem in question. However, when specifying how the resampling is done, one still needs to make some assumptions about the statistical properties of the noise. The need for such assumptions is therefore common for both direct simulation and analytical approaches. For their implementation, the resampling methods rely on computing power.

Among existing resampling techniques, the bootstrap has been shown to achieve the best theoretical performance [Efron, 1979]. Its application to spacecraft data has been discussed recently by Kawano and Higuchi [1995], who performed minimum variance analysis on an artificial set of magnetic field measurements and compared the error bounds obtained with analytical estimates by Sonnerup [1971]. The noise in the artificial magnetic data was Gaussian, isotropic, stationary and uncorrelated. Here, as in Chapter 8, the bootstrap will be applied to actual spacecraft data. We now describe the method from a practical standpoint. The mathematical underpinnings and general discussion of the technique can be found in the original paper by Efron [1979], who is credited with inventing the method and giving it the name, and in modern textbooks such as Hjorth [1994]. The bootstrap algorithm produces a simulated discrete approximation of the probability distribution for a chosen parameter in the following manner. Suppose one has a prescription (called an estimator) for obtaining an estimated value of some parameter for a given set of measurements $\{X^{(m)}\}$ with $m = 1, 2 \dots M$. This underlying rule itself is outside the scope of the resampling method. In the present context, the parameter is V_{HT} , the measurements $X^{(m)}$ are pairs of vectors $(\mathbf{B}^{(m)}, \mathbf{v}^{(m)})$, and the prescription to determine an estimated value of V_{HT} is that this value should minimise $D(V)$. To simulate the ensemble of realisations of the data set, a random sequence of M integers $i_1, i_2 \dots i_M$ is generated N times, where N is a large number. The integers are uniformly distributed in the range from 1 to M . Each such sequence, $i_1, i_2 \dots i_M$, specifies a resample, $m = i_1, i_2 \dots i_M$, of the measured data set $\{X^{(m)}\}$. For each resample, an estimated value of the parameter is obtained from the available rule (equation 9.10 in the problem at hand). Then, a stepwise probability distribution can be constructed by choosing an appropriate bin size in parameter space (the V_{HT} space in our context) and counting the number of realisations for each bin. Also, one can

Table 9.2: Bootstrap determination of HT velocities and principal axes of error ellipsoids in velocity space for magnetopause crossing by AMPTE/IRM on October 19, 1984 (05:18:20–05:19:26 UT)

GSE Components					MVAB* Components		
	V_{HT}^\dagger	$(V_{HTx}$	V_{HTy}	$V_{HTz})$	$(V_{HT} \cdot \mathbf{x}_1$	$V_{HT} \cdot \mathbf{x}_2$	$V_{HT} \cdot \mathbf{x}_3)$
i	$\sqrt{\Lambda_i}^\dagger$	e_{xi}	e_{yi}	e_{zi}	$\mathbf{e}_i \cdot \mathbf{x}_1$	$\mathbf{e}_i \cdot \mathbf{x}_2$	$\mathbf{e}_i \cdot \mathbf{x}_3$
Min. D:	V_{HT}	(−122.7	−223.2	74.5)	(247.7	95.0	6.1)
1	7.88	−0.3188	−0.3688	0.8731	0.9106	−0.4063	−0.0765
2	5.29	0.6677	0.5664	0.4831	−0.3663	−0.8788	0.3060
3	4.62	−0.6727	0.7370	0.0657	−0.1915	−0.2507	−0.9489
Min. Q:	V_{HT}	(−120.9	−224.0	72.1)	(246.1	96.8	8.0)
1	8.18	−0.3437	−0.2569	0.9032	0.8623	−0.4827	−0.1532
2	5.37	0.6686	0.6085	0.4275	−0.4292	−0.8572	0.2848
3	3.41	−0.6595	0.7508	−0.0374	−0.2687	−0.1798	−0.9463

† km/s

*MVAB components refer to principal axes derived from magnetic variance matrix

$$M_{\mu\nu}^B = \langle B_\mu B_\nu \rangle - \langle B_\mu \rangle \langle B_\nu \rangle \text{ (See Figure 8.2 on page 190 of Chapter 8)}$$

evaluate the moments of the distribution. The simplest way to proceed is to compute the moments directly over the discrete set of samples; this will be done here for the average and the variance matrix, $S_{\mu\nu}$, of the set of N simulated HT vectors. In this fashion, the prediction for V_{HT} and its uncertainty offered by the bootstrap method can be obtained and compared with the outcome of a single calculation (no resampling) and the corresponding analytically estimated error bounds, as given in Table 9.1. Further comments on the bootstrap technique and on the analytical error estimates are given in section 9.7.

Table 9.2 shows bootstrap results ($N = 5000$) for the AMPTE/IRM event studied here, using both D minimisation and Q minimisation. By comparison with the analytical results in Table 9.1, it can be seen that the V_{HT} values from bootstrap are close to those obtained from the deterministic calculation. The error ellipses are also similar, in terms of size and, to a lesser extent, axes orientation (although \mathbf{e}_3 remains somewhat close to the eigenvector \mathbf{x}_3 from MVAB). The bootstrap ellipsoid is slightly larger than the analytically derived ellipsoid. The values of the component $u_n = V_{HT} \cdot \hat{\mathbf{n}}$ are (6.1 ± 4.7) km/s for the $D(V_{HT})$ calculation and (8.0 ± 3.5) km/s for the $Q(V_{HT})$ calculation. Within estimated uncertainties, these u_n values are mutually consistent and are also consistent with the results in Table 9.1.

The actual distributions of the V_{HT} components along the vectors \mathbf{e}_i , given in the top half of Table 9.2, are shown in Figure 9.5 for the D calculation.

9.4.3 Other Uncertainty Estimates

Previously, the issue of the errors in the V_{HT} determination was addressed by *Walthour et al.* [1993] and *Terasawa et al.* [1996]. The former employed the HT analysis in connection with identifying two-dimensional quasi-stationary structures moving along the magnetopause. Their argument for an error estimate was, in effect, intuitively based on equa-

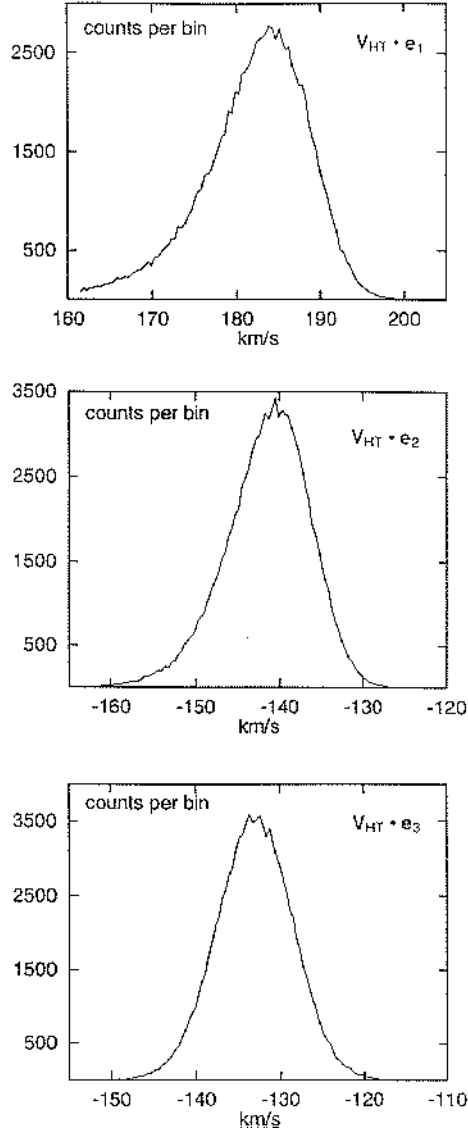


Figure 9.5: Bootstrap distributions ($N = 10^5$ minimisations of D) for the components of V_{HT} along the principal axes, e_1 , e_2 , e_3 , of the variance matrix describing the spread in the simulated V_{HT} vectors.

tion (9.18), and can be presented as follows. Owing to the noise, the minimum value of D fluctuates, and the uncertainty in the value of the minimum of the quadratic form corresponds to a region of uncertainty in V_{HT} according to

$$\Delta D = \Delta V_{\text{HT}} \cdot \mathbf{K}_0 \cdot \Delta V_{\text{HT}}$$

For a fixed ΔD , this equation specifies an ellipsoid in V_{HT} space. For ΔD , *Walthour et al.* chose one estimated standard deviation of D :

$$\sigma_D = \sqrt{\frac{1}{M(M-1)} \sum_{m=1}^M \left(E'^{(m)2} - D(V_{\text{HT}}) \right)^2} \quad (9.23)$$

Thus, the uncertainty intervals for V_{HT} along the principal axes of \mathbf{K}_0 used by *Walthour et al.* are

$$\sigma_i = \sqrt{\frac{\sigma_D}{\lambda_i}} \quad (9.24)$$

where, as before, λ_i are the eigenvalues of \mathbf{K}_0 . Thus their ellipsoid of uncertainty has the same proportions and the same orientation of its principal axes as the error ellipsoid in Section 9.4.1 but the size of the ellipsoid of *Walthour et al.* scales, to the leading order, like $M^{-1/4}$, rather than $M^{-1/2}$. The latter scaling is expected for uncorrelated stationary noise (implied, in fact, in the formula 9.23 for σ_D). Another conceptual difficulty with the estimate by *Walthour et al.* is that it predicts zero error for an imperfect HT frame in the (unlikely) event where the residual field $E'^{(m)}$ has the same magnitude for all m values. For the AMPTE/IRM event studied here, *Walthour et al.* standard deviations are larger than ours by a factor of 3.0 (for $M = 16$). We consider their values to be too large.

Terasawa et al. [1996] also assigned error margins to their estimation of V_{HT} . They developed a new technique for inferring orientation and velocity of a one-dimensional MHD structure, and needed to know the uncertainty in order to compare the predicted velocity against the value obtained by minimising D . The ellipsoid of uncertainty adopted by *Terasawa et al.* was defined so that, on its surface, $D(V)$ equals twice the minimum value. This ellipsoid too has the same geometrical shape and orientation as the one introduced in Section 9.4.1. However, with the definition used by these authors, the size of the ellipsoid does not decrease in proportion to the inverse square root of M , the number of data points in the set, which should be the case, at least in theory, when the errors are independent and identically distributed random quantities. In practice, the errors are not purely statistical and M may not be sufficiently large for the error estimates from Sections 9.4.1 and 9.4.2 to display this asymptotic inverse square root behaviour. For the AMPTE/IRM event, with $M = 16$, the standard deviations by *Terasawa et al.* are $\sqrt{2M-3} \cong 5.4$ times larger than the $\Lambda_i^{1/2}$ values given in Table 9.1 for the D calculation. We consider their values to be excessive.

9.4.4 Stationarity

A partial check on the time stationarity of the estimated value of V_{HT} and a further check of the error estimates can be obtained by applying the HT analysis to a set of nested data segments, centred at or near the middle of the current sheet. This process is also of help in searching for an optimal data interval for the analysis. The approach is to examine

whether a plateau exists in the predicted values of V_{HT} as a function of the interval length, M . The occurrence of a plateau, where the variations of the V_{HT} values fall within the estimated error bars, indicates that approximate time stationarity of the event itself may be at hand; an optimal choice of V_{HT} may be the average value on the plateau or may be chosen at the M value, on the plateau, where the uncertainty is a minimum.

Figure 9.6 displays the results of the calculations, using minimisation of D . The three components of V_{HT} with error bars are plotted versus M , the number of data points included. The value $M = 16$ corresponds to the interval used up to this point in the present chapter. The chosen fixed basis vectors onto which the components of V_{HT} are projected are the magnetic variance-matrix eigenvectors ($\mathbf{x}_1, \mathbf{x}_2, \mathbf{x}_3$) for this original interval. This reference basis is given in Figure 8.2 of Chapter 8 (page 190). It allows one to visualise the orientation of V_{HT} relative to the magnetopause current layer: the vectors \mathbf{x}_1 and \mathbf{x}_2 are tangential to the magnetopause and the minimum variance eigenvector \mathbf{x}_3 approximates the normal. The error bars for each component of V_{HT} in Figure 9.6 show plus/minus one standard deviation calculated as $(\mathbf{x}_i \cdot \mathbf{S} \cdot \mathbf{x}_i)^{1/2}$, $i = 1, 2, 3$, with \mathbf{S} given by equation 9.16. The three deviations are not uncorrelated, since the eigenbasis of \mathbf{S} is not the reference basis. However, as noted already, the minimum variance directions for V_{HT} and \mathbf{B} almost coincide for the event studied as a consequence of its property of having $\mathbf{B}^{(m)} \cdot \hat{\mathbf{n}} \cong 0$.

The sizes of the error bars in Figure 9.6 appear plausible: except for small M values, there is significant overlap between the regions of uncertainty for values of V_{HT} corresponding to adjacent M values. Statistical as well as systematic errors are likely to contribute to the large scatter for $M < 10$. The presence of, not one, but two plateau regions is evident in the component of V_{HT} along the magnetopause normal, \mathbf{x}_3 . It is in this component that the uncertainties in V_{HT} are the smallest. The first plateau occupies the range $20 \leq M \leq 32$ and the second plateau the range from $M = 34$ to the end of the plot (at $M = 50$) or beyond. The transition from the first to the second plateau is abrupt in $V_{\text{HT}} \cdot \mathbf{x}_3$ and is accompanied by an abrupt increase in the error-bar lengths. In the component $V_{\text{HT}} \cdot \mathbf{x}_1$, the first plateau is less flat and the transition to the second plateau is gradual rather than abrupt. In the remaining component, $V_{\text{HT}} \cdot \mathbf{x}_2$, there is no clear distinction between the two plateau levels. If only the error-bar lengths are considered, the V_{HT} components on the first plateau might be considered representative for the event; in fact, the errors are smallest for $M = 26$. However, in an overall sense the second plateau, appears flatter and could be selected as optimal on that basis. Note, however, that the existence of two plateau levels in $V_{\text{HT}} \cdot \mathbf{x}_3$ is associated with temporal changes in the orientation and perhaps other properties of the actual magnetopause that become incorporated for $M \geq 33$.

As mentioned already, the sense of the magnetopause transition (from the magnetosphere to the magnetosheath) requires $V_{\text{HT}} \cdot \mathbf{x}_3 < 0$. It is seen in Figure 9.6 that this requirement is not satisfied for $M = 16$, where $V_{\text{HT}} \cdot \mathbf{x}_3 = 5.9 \text{ km/s}$, but that it is satisfied for $M \geq 20$: the average value on the first plateau is $V_{\text{HT}} \cdot \mathbf{x}_3 \cong -1.6 \text{ km/s}$ and on the second plateau $V_{\text{HT}} \cdot \mathbf{x}_3 \cong -10.8 \text{ km/s}$. Even though all error bars on the first plateau extend to positive values of $V_{\text{HT}} \cdot \mathbf{x}_3$, the consistently negative individual estimates of $V_{\text{HT}} \cdot \mathbf{x}_3$ on it suggest that the actual value of $V_{\text{HT}} \cdot \mathbf{x}_3$ was negative.

Stationarity of the noise is not evident on the first plateau, where the lengths of the error bars for $V_{\text{HT}} \cdot \mathbf{x}_3$ do not decrease with increasing M , indicating that $D(V_{\text{HT}})$ itself increases with increasing M . On the second plateau, the error bars are longer but they do decrease approximately as $M^{-1/2}$ with increasing M . In an overall sense, the results

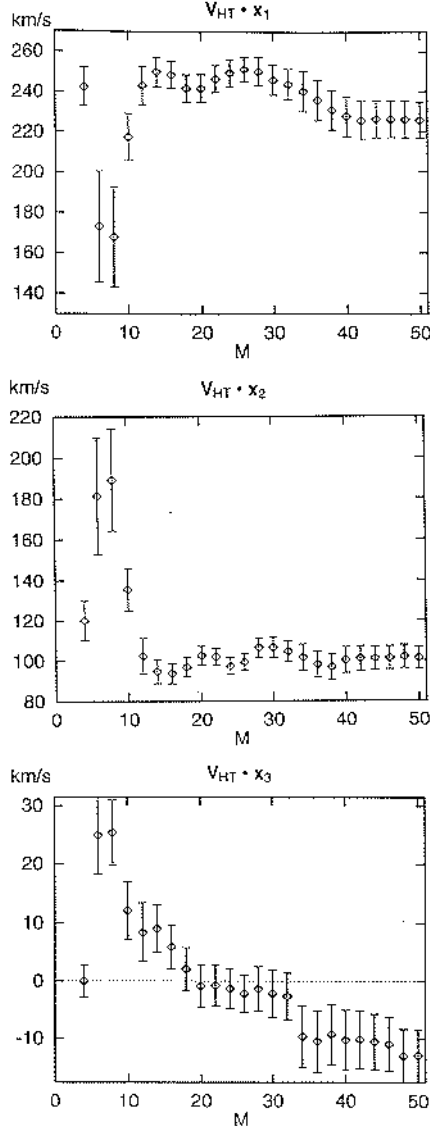


Figure 9.6: Analysis results for nested data segments in the range $4 \leq M \leq 50$, using minimisation of D and the error estimate in equation 9.16. The nest is centred at 05:18:53 UT. Components, $V_{HT} \cdot x_i$, $i = 1, 2, 3$, along the principal axes from MVAB with $M = 16$ (given in Figure 8.2 in Chapter 8) are shown. Note that the magnetopause normal $\hat{n} = x_3$. First plateau in $V_{HT} \cdot x_3$ plot is in the range $20 \leq M \leq 32$; a second plateau occurs for $M \geq 34$.

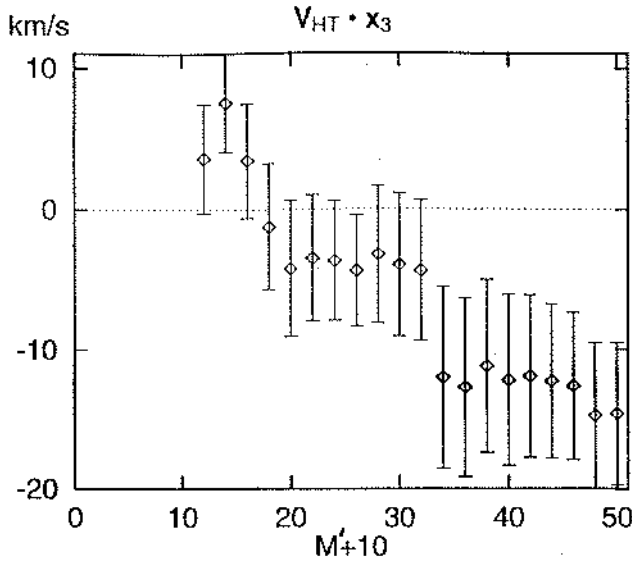


Figure 9.7: Analysis results for nested data segments, with the central 10 points of the nest excluded. Format is the same as in Figure 9.6.

in Figure 9.6 suggest that, in order of magnitude, equation 9.16 provides a reasonably realistic basis for estimation of the uncertainties in the determination of V_{HT} . The presence of a single well-defined plateau of substantial length would permit a reduction of the error estimate below the value given by equation 9.16 but, in the present case, the appearance of two plateau levels casts doubt on such a reduction.

The results of nested data-segment analysis, using minimisation of Q , are qualitatively similar to those based on D , except for the fact that $V_{HT} \cdot x_3$ is positive (+2.6 km/s) on the first plateau and about zero on the second. Also, the error bars are slightly shorter than in Figure 9.6 and the quantity $Q(V_{HT})$ displays a plateau for $10 \leq M \leq 28$, indicating stationarity of the noise. But no generalisations can be made as to whether the velocity noise is more likely to be stationary than the electric noise, or whether either of the two should be stationary; the nature of the noise has to be addressed individually for each experiment. Also, even for the specific event we study, the quasi-stationarity of $v_{\perp}^{(m)}$ does not mean that minimising Q is better than minimising D . Preference for one procedure over another may depend on physical factors, and such factors could be different from those responsible for the fluctuations in the system. For instance, in minimising D , the data points with large $|B|$ are more emphasised than those with weaker magnetic field, while in minimising Q , only the direction of each $B^{(m)}$ is important. Since in weak-field regions the MHD description is more likely to be violated, D might prove preferable to Q .

In Section 9.2, it was suggested that, for current-layers that have an intrinsic electric field as part of their internal structure, it may be desirable to exclude internal data points from the HT analysis. The results in Figure 9.6 also invite this approach. If the central 10 points of the event, where the scatter is large, are excluded, the results for $V_{HT} \cdot x_3$, shown in Figure 9.7, are obtained from minimisation of D . The error estimates in the

figure are calculated from equations 9.8 and 9.16 with M replaced by $M' = M - 10$ and $\mathbf{x}_3 = \mathbf{x}_3(M = 16)$. The errors shown in Figure 9.7 are slightly larger than those in the corresponding panel of Figure 9.6. This is a consequence of using $(M - 10)$ rather than M : the actual $D(V_{\text{HT}})$ values for Figure 9.7 are somewhat smaller than those for Figure 9.6. Two plateaus remain present in Figure 9.7 but the average value of $V_{\text{HT}} \cdot \mathbf{x}_3$ has decreased somewhat, compared to Figure 9.6, and is now $V_{\text{HT}} \cdot \mathbf{x}_3 \cong -3.9 \text{ km/s}$ and -12.7 km/s on the first and second plateau, respectively.

9.5 Accelerating HT Frame

In this section, we review the extension of the minimisation procedure to account, to the lowest order, for a possible non-steady motion of the HT frame. In the expression 9.8 for $D(V)$, one can formally allow for a time-dependent velocity, $V = V(t)$, whereupon D becomes a functional. Then, with a suitable parametrisation of the functional space, minimisation of D can be carried out to obtain a corresponding approximate function $V(t) = V_{\text{HT}}(t)$. In practice, going beyond a simple two-term linear approximation would seldom be justified, given the experimental uncertainties and the usually short duration of the spacecraft traversal of a quasi-stationary structure. The linear approximation means fitting the data with a steadily accelerating HT frame: $V(t) = V_0 + \mathbf{a}t$. The parameter values $V_0 = V_{\text{HT}0}$ and $\mathbf{a} = \mathbf{a}_{\text{HT}}$ that minimise D , are found from the following set of linear equations [Sonnerup *et al.*, 1987]:

$$\mathbf{K}_0 V_{\text{HT}0} + \mathbf{K}_1 \mathbf{a}_{\text{HT}} = \langle \mathbf{K}^{(m)} \mathbf{v}^{(m)} \rangle \quad (9.25)$$

$$\mathbf{K}_1 V_{\text{HT}0} + \mathbf{K}_2 \mathbf{a}_{\text{HT}} = \langle \mathbf{K}^{(m)} \mathbf{v}^{(m)} t^{(m)} \rangle \quad (9.26)$$

The matrices \mathbf{K}_q , $q = 0, 1, 2$ on the left-hand side are defined as

$$\mathbf{K}_q = \langle \mathbf{K}^{(m)} [t^{(m)}]^q \rangle$$

where $\mathbf{K}^{(m)}$ is given by equation 9.11. The time $t^{(m)}$ is the time at which the m -th measurement is performed, with $t^{(1)} = 0$. As before, the brackets $\langle \dots \rangle$ designate an average over the specified interval of observation, $m = 1, 2 \dots M$. The explicit solution for $V_{\text{HT}0}$ and \mathbf{a}_{HT} is

$$\mathbf{a}_{\text{HT}} = \left(\mathbf{K}_0^{-1} \mathbf{K}_1 - \mathbf{K}_1^{-1} \mathbf{K}_2 \right)^{-1} \left(\mathbf{K}_0^{-1} \langle \mathbf{K}^{(m)} \mathbf{v}^{(m)} \rangle - \mathbf{K}_1^{-1} \langle \mathbf{K}^{(m)} \mathbf{v}^{(m)} t^{(m)} \rangle \right) \quad (9.27)$$

$$V_{\text{HT}0} = \mathbf{K}_0^{-1} \left(\langle \mathbf{K}^{(m)} \mathbf{v}^{(m)} \rangle - \mathbf{K}_1 \mathbf{a}_{\text{HT}} \right) \quad (9.28)$$

For the purpose of numerical treatment, it is practical to write the above linear equations as a single set characterised by a 6×6 matrix. This representation, written in transparent block form, is

$$\begin{pmatrix} \mathbf{K}_0 & \mathbf{K}_1 \\ \mathbf{K}_1 & \mathbf{K}_2 \end{pmatrix} \begin{pmatrix} V_{\text{HT}0} \\ \mathbf{a}_{\text{HT}} \end{pmatrix} = \begin{pmatrix} \langle \mathbf{K}^{(m)} \mathbf{v}^{(m)} \rangle \\ \langle \mathbf{K}^{(m)} \mathbf{v}^{(m)} t^{(m)} \rangle \end{pmatrix} \quad (9.29)$$

The (symmetric) 6×6 matrix on the left-hand side and the 6-component vector on the right-hand side can be fed directly into a numerical solver. From the representation (9.29) it is also seen how the procedure can be generalised to obtain a least-squares polynomial approximation of $\mathbf{V}_{\text{HT}}(t)$ of an arbitrary degree. For the AMPTE/IRM event studied here, the numerical results for $M = 16$ are $\Rightarrow 9.2$

$$\mathbf{V}_{\text{HT}0} = (-102, -204, 21) \text{ km/s}$$

$$\mathbf{a}_{\text{HT}} = (-0.62, -0.54, 0.88) \text{ km/s}^2$$

which leads to $D(\mathbf{V}_{\text{HT}})/D(0) = 0.010$ rather than the value 0.012 from Section 9.3. It is evident that, for this event, the introduction of a non-zero, constant acceleration produces only a modest improvement in the least-squares fit. However, in some other events much larger improvements were found: for the case in Figure 9.4b $D/D(0)$ was reduced from 0.040 for $\mathbf{a}_{\text{HT}} = 0$ to 0.020 with \mathbf{a}_{HT} determined as described above.

From $\mathbf{V}_{\text{HT}} = \mathbf{V}_{\text{HT}0} + \mathbf{a}_{\text{HT}}t$, the HT velocity at the end of the data interval ($t = 66$ s) is found to be

$$\mathbf{V}_{\text{HT}}(66 \text{ s}) = (-143, -240, 79) \text{ km/s}$$

The average HT velocity is then

$$\bar{\mathbf{V}}_{\text{HT}} = (-123, -222, 50) \text{ km/s}$$

which should be compared to the result from Section 9.3:

$$\mathbf{V}_{\text{HT}} = (-123, -223, 76) \text{ km/s}$$

The component of the acceleration in the direction of the magnetopause normal, $\hat{\mathbf{n}} = \mathbf{x}_3$, is $\mathbf{a}_{\text{HT}} \cdot \hat{\mathbf{n}} = -0.25 \text{ km/s}^2$, indicating that the magnetopause is accelerating inwards (towards Earth). The initial and final normal velocities are $\mathbf{V}_{\text{HT}0} \cdot \hat{\mathbf{n}} = +13.5 \text{ km/s}$ and $\mathbf{V}_{\text{HT}}(66 \text{ s}) \cdot \hat{\mathbf{n}} = -3.1 \text{ km/s}$ with an average value $\bar{\mathbf{V}}_{\text{HT}} \cdot \hat{\mathbf{n}} = +4.8 \text{ km/s}$. These numbers should be compared to $\mathbf{V}_{\text{HT}} \cdot \hat{\mathbf{n}} = +5.9 \text{ km/s}$ from Section 9.3. As discussed already, $\mathbf{V}_{\text{HT}0} \cdot \hat{\mathbf{n}}$, $\mathbf{V}_{\text{HT}}(66 \text{ s}) \cdot \hat{\mathbf{n}}$, and $\bar{\mathbf{V}}_{\text{HT}} \cdot \hat{\mathbf{n}}$ must all be negative so that a systematic offset in the normal magnetopause velocities obtained for $M = 16$ appears to be present.

9.6 Discussion

In this chapter, we have used one particular AMPTE/IRM magnetopause crossing to illustrate the least-squares approach to determining the velocity, \mathbf{V}_{HT} , of the deHoffmann-Teller frame and to illustrate various error estimates for the resulting frame velocity. We now compare and discuss the results from these calculations.

An overview of the directions of the various \mathbf{V}_{HT} vectors arrived at in the chapter is given in Figure 9.8 where each \mathbf{V}_{HT} appears as a numbered dot in the plane of the figure. The orientation of this plane is defined by its two axes, (V_x, V_y) . The vertical axis, V_y , is along the magnetopause normal, $\hat{\mathbf{n}} = \mathbf{x}_3 (M = 16)$, determined for $M=16$ and given in Figure 8.2 of Chapter 8; the horizontal axis, V_x , is along $\mathbf{x}_3 \times \mathbf{V}_{\text{HT}1}$, where $\mathbf{V}_{\text{HT}1}$ is the deterministic estimate of the HT frame velocity shown as point 1 in the figure and obtained from minimisation of D with $M = 16$ (data are in upper half of Table 9.1). Thus point 1

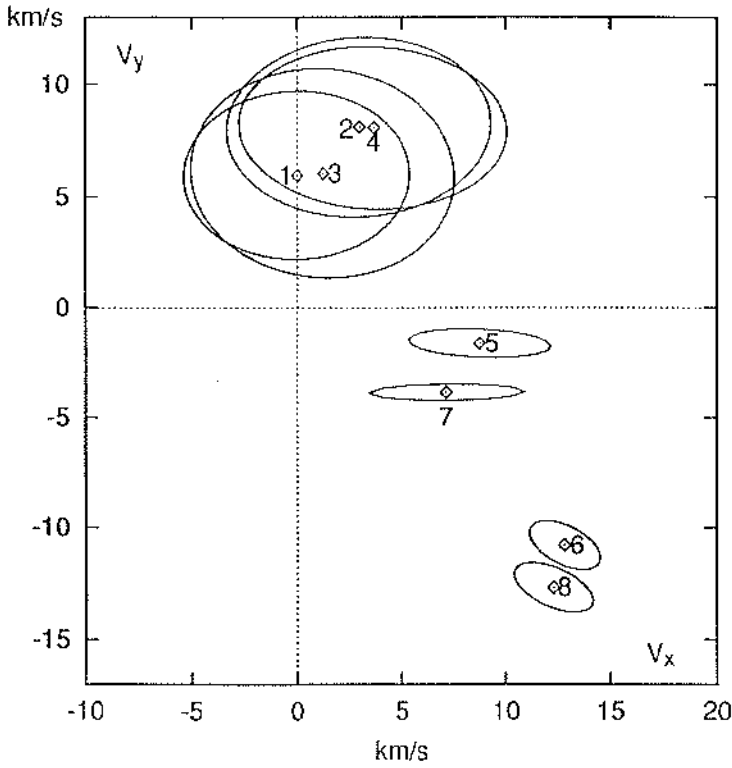


Figure 9.8: Summary of V_{HT} orientations and error ellipses. Each numbered dot, with its centred surrounding ellipse, represents one particular determination of V_{HT} and its region of uncertainty. For details, see text. The V_y coordinate of a dot represents $V_{HT} \cdot \hat{n}$ (km/s) with $\hat{n} = x_3(M = 16)$.

is located on the V_y axis at $V_y = V_{HT} \cdot \hat{n} = +5.9$ km/s. It is surrounded by an elliptical region of uncertainty, obtained as the intersection of its variance ellipsoid (set at one-standard deviation) with the plane of the figure. Similarly, point 2, with its surrounding ellipse, corresponds to the deterministic result of the Q minimisation for $M = 16$ (data are in lower half of Table 9.1). The corresponding statistical (bootstrap) estimates (data in Table 9.2) are represented by points 3 (from D) and 4 (from Q) and their corresponding ellipses. It is seen that the deterministic and the statistical-simulation error ellipses agree well and that each of the points 1, 2, 3, and 4 lies inside all four ellipses. The conclusion to be drawn is that the four determinations of V_{HT} are consistent and that the error estimates are consistent as well as reasonable. In other words, the determination of V_{HT} from the 16 data points, which describe the main part of the magnetopause, appears robust. But it is also seen that all four V_{HT} vectors are inconsistent with the requirement $V_{HT} \cdot \hat{n} < 0$, obtained from the fact that the spacecraft crossed the magnetopause in the direction from the magnetosphere to the magnetosheath. This discrepancy could of course be caused by errors in \hat{n} as well as in V_{HT} .

Point 5 in Figure 9.8 represents the average over the first plateau region in Figure 9.6,

i.e., it corresponds to using M values between 20 and 32 for determination of V_{HT} from minimisation of D ($\hat{n} = \mathbf{x}_3(M = 16)$ is retained). The small ellipse surrounding this point represents, not the total estimated uncertainty, but only the spread in the individual V_{HT} values on the plateau. The plateau result appears to be significantly different from the results for $M = 16$: There is now a small, but persistently negative value of $V_{\text{HT}} \cdot \hat{n}$, as required. The results for the second plateau ($M \geq 34$) are shown as point 6 in Figure 9.8. They are again significantly different from those for $M = 16$ as well as from those for $20 \leq M \leq 32$. In particular, the value of $V_{\text{HT}} \cdot \hat{n}$ is now even more negative. The presence of two plateaus must be the result of temporal changes in the system. The corresponding results when 10 points at the centre of the magnetopause have been removed are shown as points 7 and 8 for the first and second plateau, respectively.

We conclude that optimal separate determinations of \hat{n} and V_{HT} may require use of different M values for the minimum variance analysis on \mathbf{B} , to obtain \hat{n} , and for the HT analysis. In Chapter 8, evidence is presented, indicating that a systematic error may be present in the plasma velocity measurements during this event. The predicted sense of this error is such as to produce a spurious positive (outward) addition to the magnetopause velocity, $u_n = V_{\text{HT}} \cdot \hat{n}$, so that the data analysis may lead to a positive value of u_n even though its true value is negative. It must be concluded that, for the event studied here, the actual negative value of u_n cannot be established with certainty from the data. We believe this statement to apply, not only in the context of separate V_{HT} and normal vector calculations, as discussed here, but also for the combined data analysis technique proposed by *Terasawa et al.* [1996]. However, we are in a position to conclude that the magnitude of u_n was relatively small: the data are not consistent with rapid inward magnetopause motion.

Finally, we mention briefly how a “proper” frame could exist and be determined for two-dimensional time-independent structures for which no true HT frame exists. By “proper” frame we mean a frame that moves with the structure. In such a frame, the electric field component along the invariant direction of the structure, the z direction say, is constant, while strongly varying electric field may remain in a plane perpendicular to \hat{z} . The frame velocity, V_{HT} , is not unique in this situation because any arbitrary velocity component along \hat{z} can be added to it without changing the property $E'_z = \text{const} = E_0$. Thus it is convenient to put the frame velocity along the invariant direction equal to zero. Except for this constraint, the orientation and magnitude of \tilde{V}_{HT} , as well as E_0 and the orientation of the invariant axis, \hat{z} , are unknown, *a priori*. But \hat{z} , \tilde{V}_{HT} , and E_0 can in principle be found by minimisation of

$$\begin{aligned} \tilde{D}(\tilde{V}, \hat{z}, E_0) &= \frac{1}{M} \sum_{m=1}^M \left(\mathbf{E}'^{(m)} \cdot \hat{z} - E_0 \right)^2 \\ &= \frac{1}{M} \sum_{m=1}^M \left([(\mathbf{v}^{(m)} - \tilde{V}) \times \mathbf{B}^{(m)}] \cdot \hat{z} - E_0 \right)^2 \end{aligned} \quad (9.30)$$

where we impose the conditions $|\hat{z}|^2 = 1$ and $\tilde{V} \cdot \hat{z} = 0$. Because the GSE components of \tilde{V} are multiplied by those of \hat{z} in the expression for $\mathbf{E}'^{(m)}$, the minimisation of \tilde{D} does not produce linear equations. Therefore, the extremum may be found most conveniently by some trial-and-error process. This type of calculation could be of potential importance for 2-D structures where there is in fact no need for a true HT frame to exist: it is suffi-

cient to be able to find a frame which moves with the structure and in which the structure appears time-independent. An application of the technique of finding a proper frame of 2-D structures is currently under study. Proper frames of one-dimensional structures with intrinsic electric fields (i.e., not removable by galilean transformations) can be found by the methods developed by *Kawano and Higuchi [1996]* and *Terasawa et al. [1996]*. Such structures include perpendicular shocks. We have also shown that, within the framework of the HT analysis, all these structures can be accommodated by allowing for a constant, possibly non-zero, residual field E'_0 on the right-hand side of equation 9.1, and minimising D with respect to both V and E'_0 . This extension of the HT method will be presented in detail elsewhere.

9.7 Summary

1. DeHoffmann-Teller (HT) analysis is a means to identify, from first principles, the frame of reference (HT frame) in which the electric field associated with an observed structure in a magnetised plasma is as small as possible (ideally zero) and in which the structure therefore appears as close as possible to being stationary. In other words, the method fits the data with a model in which the observed temporal variations and electric fields are caused entirely by the motion of the sampled structure with a velocity $V = V_{HT}$ relative to the instruments. Steady motion of the HT frame is considered in the main part of the chapter but an extension to the case of steady acceleration is also given. An HT frame may exist, not only for one-dimensional current layers but for other moving two- and three-dimensional structures as well. However, for perpendicular shocks and other structures possessing a strong intrinsic electric field, the HT frame concept is not useful and the HT frame velocity obtained as described in the main part of this chapter does not coincide with the “proper” frame in which the structure is stationary. Such cases are not dealt with here.
2. As described in Section 9.3, the deHoffmann-Teller (HT) frame velocity, V_{HT} , can be determined from measured plasma velocity vectors, $v^{(m)}$, and magnetic field vectors, $B^{(m)}$, $m = 1, 2, \dots, M$, by minimisation of either $D(V_{HT})$ or $Q(V_{HT})$. Here $D(V_{HT})$ and $Q(V_{HT})$, defined by equations 9.8 and 9.14, are the averages of the square of the convection electric fields, $E_c^{(m)}$, and of $v_{\perp}^{(m)}$, the velocities transverse to $B^{(m)}$, respectively, remaining in the HT frame. The minimisation of $D(V_{HT})$ leads to the explicit expression 9.10 for V_{HT} , where the matrices $K^{(m)}$ and K_0 are defined by equation 9.11. Small modifications of these expressions are needed when $Q(V_{HT})$ is minimised. If actual measured electric field vectors, $E^{(m)}$, are used in place of the convection field $E_c^{(m)} = -v^{(m)} \times B^{(m)}$, then equation 9.10 is replaced by 9.12.
3. The validity of the HT velocity as a prediction of the “proper” frame velocity of a current layer or other structure becomes violated if an electrostatic field is present in this proper frame. However, provided that the large-scale behaviour is MHD, the non-ideal effects causing the residual field can manifest themselves only in thin layers. The measurements acquired inside the layers comprise but a small portion of the total data set and then should not strongly influence the result. Also, one

can eliminate the effect of such transition layers by excluding the corresponding measurements from the analysis (Section 9.4.4). An attempt to account for non-ideal effects in the model can be made by use of electron velocity measurements $\mathbf{v}_e^{(m)}$ in place of $\mathbf{v}^{(m)}$ (see end of Section 9.3.1), but this latter approach has not been practically tested. For two-dimensional or three-dimensional structures possessing an intrinsic electric field in their proper frame, the situation is more complicated. An, as yet untested, approach to two-dimensional cases is discussed in Section 9.6.

4. A useful measure of the quality of the HT frame thus determined is provided by the relative residual $D(\mathbf{V}_{\text{HT}})/D(0) = 1 - c^2$ (or the corresponding expression for $Q(\mathbf{V}_{\text{HT}})/Q(0)$), c being the correlation coefficient between $\mathbf{E}_{\text{HT}}^{(m)} = -\mathbf{V}_{\text{HT}} \times \mathbf{B}^{(m)}$ and $\mathbf{E}_c^{(m)} = -\mathbf{v}^{(m)} \times \mathbf{B}^{(m)}$ (between $\mathbf{V}_{\text{HT}\perp}^{(m)}$ and $\mathbf{v}_{\perp}^{(m)}$ for the $Q(\mathbf{V}_{\text{HT}})$ minimisation or between $\mathbf{E}_{\text{HT}}^{(m)}$ and $\mathbf{E}^{(m)}$ when actual measured electric fields are used). High quality HT frames will have $c \geq 0.99$. However, these quality measures are frame dependent, i.e., they depend on the actual magnitude of \mathbf{V}_{HT} ; for this reason they are not suitable for inter-comparison of different events.
5. The utility of the HT frame for the observational study of internal structures in current layers and other moving field configurations is discussed and is illustrated by a test of the so-called Walén relation, i.e., the extent to which the plasma flow speed in the HT frame agrees with Alfvén speed in a magnetopause current layer (Figure 9.4).
6. When an accurate normal direction $\hat{\mathbf{n}}$ can be derived for a current layer, e.g., as described in Chapter 8, then $\mathbf{V}_{\text{HT}} \cdot \hat{\mathbf{n}}$ represents the velocity of motion of the layer along the normal. This velocity allows determination of the thickness of the layer. Similarly, the components of \mathbf{V}_{HT} tangential to the layer permit determination of the spatial scales of two-dimensional and three-dimensional structures within it.
7. A simple analytical estimate of the error ellipsoid for \mathbf{V}_{HT} (at one standard-deviation level) is provided by equation 9.16 for the $D(\mathbf{V}_{\text{HT}})$ minimisation and by a small modification thereof, equation 9.22, for the $Q(\mathbf{V}_{\text{HT}})$ minimisation. These error estimates are frame independent and are therefore suitable for inter-comparison of different events. In special cases, the error ellipsoid may be very elongated.
8. At least in the illustrative example presented in this chapter, the above analytical error estimates are very similar to those obtained from a simplified application of the numerically intensive bootstrap method, as discussed in Section 9.4.2. A systematic inter-comparison of the two techniques, as applied to \mathbf{V}_{HT} analysis, is not available at present.
9. A few words of warning are in order concerning the error estimates. The analytical approach is based on a number of standard assumptions concerning the nature of the noise, as well as on linearisation. These simplifications permit analytical formulas to be derived and they lead to the usual $M^{-1/2}$ scaling as the number of the data points, M , increases. This behaviour is not always observed and it appears likely that in such cases one or more of the assumptions concerning the error sources are invalid. In particular, systematic errors caused by temporal and spatial variations are likely to be present in some cases. The bootstrap technique, while based on the

same statistical assumptions as the analytical ensemble-averaging, can in principle recover the entire probability distribution of the parameter in question and obtain error estimates without linearisation (on the other hand, an approximation inherent in bootstrap but not in analytical averaging is the Monte Carlo approximation of the statistical ensemble). But, for the V_{HT} determination, it is not entirely clear how the technique should be applied: in principle, not the signal but only the noise should be resampled [H. Kawano and T. Higuchi, private communication], a rule that is not observed in the bootstrap application presented here (or in Chapter 8). ⇒9.3

10. Because of the potential difficulties with the error estimation techniques mentioned above, it is particularly important to check the stationarity of the V_{HT} results and of the associated noise. Such checks can be performed by analysis of nested data segments (including shifts of the centre of the nest) as discussed in section 9.4.4. The scatter of V_{HT} vectors obtained from such experiments provides an additional basis from which uncertainties in the results can be estimated. A significant conclusion from the sample event studied in the chapter is that the optimal data segment for determination of V_{HT} does not necessarily coincide with the optimal segment for MVAB, MVAE, or MVA ρv analyses discussed in Chapter 8.
11. The method for determining an initial velocity, $V_{\text{HT}0}$, and a constant acceleration, a_{HT} , of the HT frame is described in Section 9.5 in terms of the inversion of a 6×6 matrix equation. Because the Earth's magnetopause is often in a state of oscillatory inward-outward motion, acceleration effects may be important in that application. Experience indicates that their inclusion can in some cases lead to substantial improvements of the quality of the HT frame, as measured by $D(V_{\text{HT}})/D(0)$, as well as to significant changes in the outcome of the Walén test.

Appendix

Here, we document the calculation of the matrix $S_{\mu\nu} = \langle\langle \Delta V_{\text{HT}\mu} \Delta V_{\text{HT}\nu} \rangle\rangle$, utilised in Section 9.4 to characterise the uncertainty of V_{HT} , under the assumptions stated in that section. The matrix \mathbf{S} is the variance matrix of all possible realisations of V_{HT} . Proceeding from equation 9.10, we write

$$V_{\text{HT}} = K_0^{-1} \left\langle K^{(m)} v^{(m)} \right\rangle = K_0^{-1} \frac{1}{M} \sum_{m=1}^M B^{(m)2} v_{\perp}^{(m)}$$

Therefore, using the summation convention, we have

$$S_{\mu\nu} = \langle\langle \Delta V_{\text{HT}\mu} \Delta V_{\text{HT}\nu} \rangle\rangle = \frac{1}{M^2} K_{0\mu\gamma}^{-1} K_{0\nu\xi}^{-1} \sum_{m,n=1}^M B^{(m)2} B^{(n)2} \langle\langle \Delta v_{\perp\gamma}^{(m)} \Delta v_{\perp\xi}^{(n)} \rangle\rangle \quad (9.31)$$

The (frame-independent) velocity fluctuations Δv_{\perp} are related to the electric field fluctuations ΔE , which we identify with E'_{\perp} , the residual field in the HT frame:

$$\Delta v_{\perp}^{(m)} = \frac{E'_{\perp}{}^{(m)} \times B^{(m)}}{B^{(m)2}} \quad (9.32)$$

The \mathbf{E}'_{\perp} field is postulated to possess a stationary probability distribution, isotropic in the plane normal to \mathbf{B} , and uncorrelated from one measurement to another:

$$\langle\langle E'_{\perp\mu}(m) E'_{\perp\nu}(n) \rangle\rangle = \frac{\langle\langle E'^2 \rangle\rangle}{2} \delta_{mn} \left(\delta_{\mu\nu} - \frac{B_{\mu}^{(m)} B_{\nu}^{(m)}}{B^{(m)2}} \right) \quad (9.33)$$

We use equations 9.32 and 9.33 to calculate the ensemble average $\langle\langle \Delta v_{\perp\gamma}^{(m)} \Delta v_{\perp\xi}^{(n)} \rangle\rangle$ needed in equation 9.31. From 9.32, it is seen that, if \mathbf{E}'_{\perp} is distributed isotropically in the plane normal to $\mathbf{B}^{(m)}$, then so is $\Delta \mathbf{v}_{\perp}^{(m)}$, and thus the velocity correlation writes in the same form as equation 9.33:

$$\begin{aligned} \langle\langle \Delta v_{\perp\gamma}^{(m)} \Delta v_{\perp\xi}^{(n)} \rangle\rangle &= \frac{\langle\langle E'^2 \rangle\rangle}{2B^{(m)2}} \delta_{mn} \left(\delta_{\gamma\xi} - \frac{B_{\gamma}^{(m)} B_{\xi}^{(m)}}{B^{(m)2}} \right) \\ &= \frac{\langle\langle E'^2 \rangle\rangle}{2B^{(m)4}} \delta_{mn} K_{\gamma\xi}^{(m)} \end{aligned} \quad (9.34)$$

Substituting equation 9.34 into 9.31 gives

$$\begin{aligned} S_{\mu\nu} &= \frac{\langle\langle E'^2 \rangle\rangle}{2M^2} K_{0\mu\gamma}^{-1} K_{0\nu\xi}^{-1} \sum_{m=1}^M K_{\gamma\xi}^{(m)} = \frac{\langle\langle E'^2 \rangle\rangle}{2M} K_{0\mu\gamma}^{-1} K_{0\nu\xi}^{-1} K_{0\gamma\xi} \\ &= \frac{\langle\langle E'^2 \rangle\rangle}{2M} K_{0\mu\nu}^{-1} \end{aligned} \quad (9.35)$$

Next we calculate the ensemble average of $D(\mathbf{V}_{\text{HT}})$. From equations 9.8 and 9.10, it follows that

$$\langle\langle D(\mathbf{V}_{\text{HT}}) \rangle\rangle = \langle\langle E'^2 \rangle\rangle - K_{0\mu\nu} S_{\mu\nu} = \langle\langle E'^2 \rangle\rangle \left(1 - \frac{3}{2M} \right) \quad (9.36)$$

where the second of the two equalities is obtained with the aid of equation 9.35. From 9.35 and 9.36, it follows that

$$S_{\mu\nu} = \frac{\langle\langle D(\mathbf{V}_{\text{HT}}) \rangle\rangle}{2M - 3} K_{0\mu\nu}^{-1}. \quad (9.37)$$

Finally, the average expected value of $D(\mathbf{V}_{\text{HT}})$ is approximated by the one obtained for the actual realisation of the data set, whereupon the uncertainty estimate 9.16 results.

Acknowledgements

We thank Drs. G. Paschmann and H. Lühr for use of AMPTE/IRM plasma and magnetic field data. The research was supported by the National Aeronautics and Space Administration under grants NAGW-4023, -5137 and NAG 5-3031, and by the National Science Foundation, Atmospheric Sciences Division, under grant ATM-9422918 to Dartmouth College.

Bibliography

Aggson, T. L., Gambardella, P. J., and Maynard, N. C., Electric field measurements at rotational magnetopause discontinuities, *J. Geophys. Res.*, **88**, 10000, 1983.

- deHoffmann, F. and Teller, E., Magneto-hydrodynamic shocks, *Phys. Rev.*, **80**, 692, 1950.
- Efron, B., Bootstrap methods: Another look at the jackknife, *Annals of Statistics*, **7**, 1, 1979.
- Hjorth, J. S. U., *Computer Intensive Statistical Methods*, Chapman and Hall, New York, 1994.
- Kawano, H. and Higuchi, T., The bootstrap method in space physics: error estimation for the minimum-variance analysis, *Geophys. Res. Lett.*, **22**, 307, 1995.
- Kawano, H. and Higuchi, T., A generalization of the minimum variance analysis method, *Ann. Geophys.*, **14**, 1019, 1996.
- Papamastorakis, I., Paschmann, G., Baumjohann, W., Sonnerup, B. U. Ö., and Lühr, H., Orientation, motion and other properties of flux transfer event structures on September 4, 1984, *J. Geophys. Res.*, **94**, 8852, 1989.
- Paschmann, G., Comment on "Electric field measurements at the magnetopause 1. Observation of large convective velocities at rotational magnetopause discontinuities", by T. L. Aggson, D. J. Gambardella, and N. D. Maynard, *J. Geophys. Res.*, **90**, 7629, 1985.
- Sonnerup, B. U. Ö., Magnetopause structure during the magnetic storm of September 24, 1961, *J. Geophys. Res.*, **76**, 6717, 1971.
- Sonnerup, B. U. Ö. and Guo, M., Magnetopause transects, *Geophys. Res. Lett.*, **23**, 3679, 1996.
- Sonnerup, B. U. Ö., Papamastorakis, I., Paschmann, G., and Lühr, H., Magnetopause properties from AMPTE/IRM observations of the convection electric field: Method development, *J. Geophys. Res.*, **92**, 12 137, 1987.
- Sonnerup, B. U. Ö., Papamastorakis, I., Paschmann, G., and Lühr, H., The magnetopause for large magnetic shear: Analysis of convection electric fields from AMPTE/IRM, *J. Geophys. Res.*, **95**, 10 541, 1990.
- Terasawa, T., Kawano, H., Shinohara, I., Mukai, T., Saito, Y., Hoshino, M., Nishida, A., Machida, S., Nagai, T., Yamamoto, T., and Kokubun, S., On the determination of a moving MHD structure: Minimization of the residue of integrated Faraday's equation, *J. Geomagn. Geoelectr.*, **48**, 603, 1996.
- Walthour, D. W. and Sonnerup, B. U. Ö., Remote sensing of 2D magnetopause structures, in *Physics of the Magnetopause*, edited by P. Song, B. U. Ö. Sonnerup, and M. F. Thomsen, Geophysical Monograph 90, pp. 247–255, Amer. Geophys. Union, Washington D.C., 1995.
- Walthour, D. W., Sonnerup, B. U. Ö., Paschmann, G., Lühr, H., Klumpar, D., and Potemra, T., Remote sensing of two-dimensional magnetopause structures, *J. Geophys. Res.*, **98**, 1489, 1993.

— 10 —

Shock and Discontinuity Normals, Mach Numbers, and Related Parameters

STEVEN J. SCHWARTZ
Queen Mary and Westfield College
London, United Kingdom

10.1 Introduction

Shocks and other boundaries provide key sites for mediating the mass, momentum, and energy exchange in space plasmas, and have thus been the subject of considerable research. The transient, non-planar nature of the Solar-Terrestrial interaction, however, complicates any interpretation of data from a single spacecraft. The importance of such complications, together with the importance of the boundary layer physics, is a prime motivation behind any multi-spacecraft mission. In this chapter we review the basic terminology and methodology for the determination of the most basic of parameters, such as the Mach number and the shock/discontinuity orientation. These parameters need to be established before any detailed investigation can proceed, and before any of the results thereof can be set into their proper context.

While we concentrate mainly on shocks, the methods described below apply to most “discontinuities” encountered in space plasmas, such as rotational or tangential discontinuities.

10.2 The Shock Problem: Rankine-Hugoniot Relations

The overall shock problem consists of a surface through which a non-zero mass flux flows and which effects an irreversible (i.e., entropy-increasing) transition via dissipation of some sort. At the macroscopic level, the shock must conserve total mass, momentum, and energy fluxes, together with an obedience of Maxwell’s equations. Adopting a particular macroscopic framework, such as time-stationary ideal MHD, the governing equations can be written in conservation form to reveal expressions for these fluxes in the planar (1-D) case. For example, starting from the mass continuity equation

$$\frac{\partial \rho}{\partial t} + \nabla \cdot (\rho \mathbf{V}) = 0 \quad (10.1)$$

only the $\hat{\mathbf{n}} d/dx_n$ operator is non-zero. In a frame in which the shock is at rest, this implies conservation of the normal mass flux and is given by

$$\rho_u (\mathbf{V}_u \cdot \hat{\mathbf{n}}) = \rho_d (\mathbf{V}_d \cdot \hat{\mathbf{n}}) \quad (10.2)$$

where \hat{n} is the shock normal and subscripts u and d denote quantities measured up- and downstream of the shock. Similar expressions can be derived for the three components of momentum carried through the shock, the energy flux, the normal magnetic field, and the two components of the tangential electric field. To these are added the frozen in field relations for the two tangential electric field components. The resulting system is 10 equations in the 10 parameters ρ , V , P (the plasma thermal pressure), B , and $E_{\text{tangential}}$ on either side of the shock. This set of relations between upstream and downstream parameters is known as the Rankine-Hugoniot relations (sometimes referred to as the “shock jump conditions” although they apply across any discontinuity). If the upstream state is completely specified, these relations can be solved for the downstream state. In the case of a more complicated system, e.g., two-fluid or multi-species, there are in general fewer equations than parameters, and further assumptions concerning energy partition, etc., are required to close the system. Even in MHD, the energy conservation equation requires some closure assumptions (e.g., zero heat flux) which at best mimic the consequences of the dissipation at the shock.

10.3 Shock Parameters

There are several basic plasma parameters which characterise the media on either side of a shock. These are usually based on plasma fields and moments. This section introduces the essential parameters and shock nomenclature.

10.3.1 Shock Geometry

In a plasma permeated by a magnetic field, there are several vectors which play a role in the analysis of a shock transition. These include the bulk flow velocity, V , the magnetic field, B , and the shock normal, \hat{n} . It is customary to orient the normal vector so that it points into the unshocked medium. In a frame in which the shock is at rest, this normal points “upstream” and the term upstream is often interchanged with “unshocked”. This use of upstream and downstream can cause confusion in the case of interplanetary shocks, however, since in the spacecraft rest frame a shock propagating in the anti-sunward direction (a “forward” shock) has its unshocked medium down-wind of the shock location. Nonetheless, we will follow common practice and use the terms upstream and downstream as seen in a frame in which the shock is at rest. Accordingly, subscripts u and d will be used to denote quantities measured in the upstream and downstream media respectively.

The basic shock geometry is sketched in Figure 10.1. The various angles between V_u , B_u , and \hat{n} are typically denoted by a subscripted θ , e.g., θ_{Bnu} is the angle between B_u and \hat{n} . Usually only the acute angles for θ_{Bnu} and θ_{Vnu} are required, but the user must be careful in analysing any given situation or in writing a generalised algorithm to cope with a rotation through 180° of the magnetic field or a normal vector which points into the downstream direction. The vector forms of the transformation velocities given below avoid such problems.

The transformation of a velocity measured in an arbitrary frame, V^{arb} , (e.g., the spacecraft frame of reference) to one in which the shock is at rest is accomplished by subtracting the shock velocity in the arbitrary frame of interest. Only the normal component of the

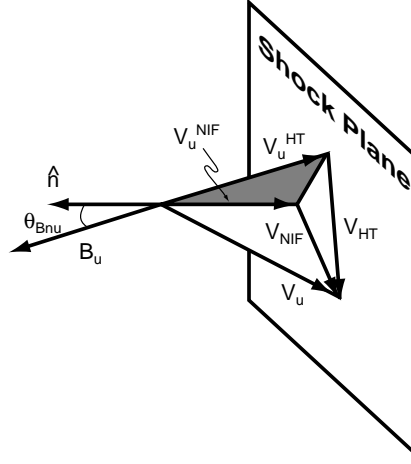


Figure 10.1: Sketch of the various vectors which enter into shock analyses. These are shown in a shock rest frame.

shock velocity, V_{sh}^{arb} , is unique, so we shall use

$$\mathbf{V}_{Rest}^{Shock} = \mathbf{V}^{arb} - V_{sh}^{arb} \hat{\mathbf{n}} \quad (10.3)$$

All the velocities subscripted u shown in Figure 10.1 are such shock rest frame velocities. Section 10.5 provides methods to determine V_{sh}^{arb} . There are, in fact, a multiplicity of shock rest frames, since any translation along a (planar) shock surface leaves the shock at rest. Two frames are useful:

The Normal Incidence Frame (NIF) is such that the upstream flow is directed along the shock normal. As we shall see below, it is only this component of the flow which enters into the shock Mach number.

The deHoffmann-Teller (HT) frame is such that the upstream flow is directed along the upstream magnetic field. This frame has the advantage that the $\mathbf{V}_u \times \mathbf{B}_u$ electric field vanishes. Additionally, due to the constancy of the tangential electric field across any plane layer, the flow and field are also aligned in the downstream region.

The transformation into the NIF frame from another shock rest frame is achieved via a frame velocity

$$\mathbf{V}_{NIF} = \hat{\mathbf{n}} \times (\mathbf{V}_u \times \hat{\mathbf{n}}) \quad (10.4)$$

so that in the NIF frame the upstream bulk velocity is

$$\mathbf{V}_u^{NIF} = \mathbf{V}_u - \mathbf{V}_{NIF} \quad (10.5)$$

It is straightforward to verify that \mathbf{V}_u^{NIF} is parallel to $\hat{\mathbf{n}}$ and that the transformation velocity \mathbf{V}_{NIF} lies in the shock plane. These vectors are shown in Figure 10.1.

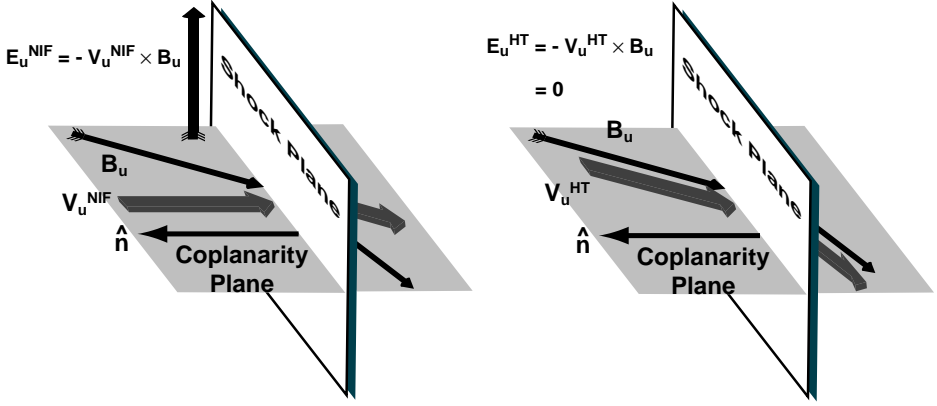


Figure 10.2: Sketch showing the various field, flow, and normal vectors in the Normal Incidence Frame (NIF) (left) and deHoffmann-Teller (HT) frame (right).

The transformation into the HT frame from another shock rest frame is achieved via a frame velocity

$$V_{HT} = \frac{\hat{n} \times (V_u \times B_u)}{B_u \cdot \hat{n}} \quad (10.6)$$

so that in the HT frame the upstream bulk velocity is

$$V_u^{HT} = V_u - V_{HT} \quad (10.7)$$

Note that this prescription for the velocity V_{HT} is restricted to transformations only from another shock rest frame, whereas V_{HT} in Chapter 9 provides the transformation from an arbitrary (e.g., spacecraft) frame. It is straightforward to verify that V_u^{HT} is parallel to B_u and that the transformation velocity V_{HT} lies in the shock plane. These vectors are also shown in Figure 10.1. For highly oblique magnetic fields (i.e., $\theta_{Bnu} \approx 90^\circ$) the transformation into the HT frame requires relativistic treatment, and cannot be achieved at all when $\theta_{Bnu} = 90^\circ$.

The resulting geometries in these two frames are shown in Figure 10.2. Note that the flow and magnetic field vectors in the HT frame are aligned with one another in both the upstream and downstream regions. Also note that by the coplanarity theorem (see below) there is a single plane, the Coplanarity Plane shown in the figure and defined by the magnetic field and normal vectors, which contains the flow, magnetic field, and normal vectors on both sides of the shock surface. This plane is also shaded in Figure 10.1.

This discussion of various shock frames reveals that there is only one geometric parameter which enters the shock problem, namely θ_{Bnu} . Shocks with large values of θ_{Bnu} are called quasi-perpendicular, while those with values near zero are quasi-parallel. In the case of fast mode shocks, the separation between quasi-perpendicular and quasi-parallel is usually taken at $\theta_{Bnu} = 45^\circ$, as this value divides the behaviour of reflected ions which participate in the shock dissipation.

10.3.2 Mach Numbers

The Mach number of a shock in an ordinary fluid is the ratio of the speed of the shock along the shock normal (i.e., $|\mathbf{V}_u \cdot \hat{\mathbf{n}}|$ in our notation) to the speed of sound in the medium upstream of the shock. In a magnetised plasma, there are three low frequency modes: the fast and slow magnetosonic waves and the intermediate (Alfvén) wave. The intermediate mode is incompressible and might not be expected to lead to a shock solution, although there has been some discussion about possible intermediate shocks in the theoretical literature. The intermediate mode does give rise to non-compressive sharp transitions, known as rotational discontinuities, in which the field (and flow) are rotated through some angle about the normal. Both the fast and slow magnetosonic waves give rise to shock solutions, known as fast and slow shocks.

Thus there are several Mach numbers of interest. The plasma counterparts to the sonic Mach number in fluids are obviously the fast Mach number, M_f , and slow Mach number, M_s , which are the ratios of the normally incident flow speed to the fast and slow MHD wave speeds in the upstream medium. These speeds are complicated by the non-isotropic nature of the MHD modes, so that the wave speeds depend on propagation direction (i.e., θ_{Bnu}). Thus an additional Mach number, the Alfvén Mach number, M_A , is often used to characterise a shock. This Mach number is calculated without regard to propagation direction, i.e.,

$$M_A = \frac{|\mathbf{V}_u \cdot \hat{\mathbf{n}}|}{|\mathbf{B}_u| / \sqrt{\mu_0 \rho_u}} \quad (10.8)$$

The intermediate Mach number, $M_I \equiv M_A \sec \theta_{Bnu}$, is useful as it represents an upper limit to the slow Mach number. There is no upper limit to the fast Mach number, although in this case there is a “critical” Mach number, M_c , above which simple resistivity cannot provide the total shock dissipation. M_c is a function of the various shock parameters, but is at most 2.7 and usually much closer to unity. Thus many fast mode shocks in space are supercritical.

10.3.3 Important Ratios

In addition to θ_{Bnu} and the relevant shock Mach number, two more ratios are useful in parameterising shocks. One is the upstream plasma β , i.e., the ratio of plasma to magnetic pressure. The value of β controls the relative importance of the magnetic field and the level of turbulence amongst other things. At the MHD level, θ_{Bnu} , M_A , and β_u completely specify the shock problem. M_A can be replaced by any other Mach number, as they are all related via θ_{Bnu} and β_u .

The other ratio of interest is the electron to ion temperature ratio, as this controls the expected micro-instabilities.

10.4 Determination of Shock and Discontinuity Normals

There are numerous techniques aimed at determining shock normals, shock speeds, and the values of upstream and downstream plasma and field parameters which best describe the shock. Many of them rely heavily on magnetic field data, which generally

provides good time resolution together with small experimental uncertainties. The field alone, however, cannot provide the shock Mach number, heating, and other parameters, and solutions which include plasma observations are thus required. In the case of travelling interplanetary discontinuities the shocks are often weak, implying that only small changes occur in the plasma properties. This, coupled with the rather short transit time of the spacecraft in relation to the shock, increases the difficulty of the task. In the case of standing planetary bow shocks, the upstream state is the solar wind flow, which appears to the ion instruments as a collimated beam, while the downstream state is a much broader, heated population. Under such circumstances, it can be difficult to resolve both states within a single instrument to the necessary precision, and combining data from two separate instruments imposes severe cross-calibration problems which must be overcome.

In the sections which follow, we describe a variety of approaches to this problem. There is a continual trade-off amongst ease of applicability, completeness, and accuracy which must be balanced.

10.4.1 Variance Analyses

A single spacecraft passing through a 1-D structure will see variations in the magnetic field. Since $\nabla \cdot \mathbf{B} = 0$, the normal component of the field must remain constant. It follows that if a unique direction can be found such that the variations in magnetic field along that direction are zero (or at least minimised to a sufficient extent), then this direction corresponds to the normal direction. This method fails for pure MHD shock solutions or other cases where the variance direction is degenerate (see Chapter 8). When considering the electric field, an opposite argument holds: the tangential components of \mathbf{E} should be continuous through such a layer, so the normal direction will correspond to that of maximum variance in \mathbf{E} . These variance techniques are described in greater detail in Chapter 8 and extended in Chapter 11 to multiple spacecraft encounters of curved surfaces, and will not be repeated here.

It is worth remembering here that, unlike all the other methods described below, the variance techniques deal with variations *within* the transition rather than the observations taken well up- and downstream of the transition. There may well be fluctuations in the upstream or downstream regions which do not lie along the shock normal (such as waves propagating along the magnetic field direction) which can give rise to difficulties if the interval selected for variance analysis is not carefully chosen and tested.

10.4.2 Coplanarity and Related Single Spacecraft Methods

The normal to a planar surface can be determined if two vectors which lie within the surface can be found. Several subsets of the Rankine-Hugoniot relations can be used to determine suitable vectors. The most widely used method relies on the Coplanarity Theorem which insists that, for compressive shocks, the magnetic field on both sides of the shock and shock normal all lie in the same plane. A corollary to this, since the only possible tangential stresses arise from the magnetic tension, is that the velocity jump across the shock also lies in this plane (as shown in Figure 10.2). The dominant change in velocity is usually along the shock normal, especially at moderate and higher Mach numbers.

Thus there are a variety of vectors which lie in the shock plane. These include the change in magnetic field (which also lies in the coplanarity plane), the cross-product of

the upstream and downstream magnetic fields (which is perpendicular to the coplanarity plane), and the cross-product between the upstream or downstream magnetic field or their difference with the change in bulk flow velocity. These vectors give rise to constraint equations for the shock normal:

$$(\Delta \mathbf{B}) \cdot \hat{\mathbf{n}} = 0 \quad (10.9)$$

$$(\mathbf{B}_d \times \mathbf{B}_u) \cdot \hat{\mathbf{n}} = 0 \quad (10.10)$$

$$(\mathbf{B}_u \times \Delta \mathbf{V}^{\text{arb}}) \cdot \hat{\mathbf{n}} = 0 \quad (10.11)$$

$$(\mathbf{B}_d \times \Delta \mathbf{V}^{\text{arb}}) \cdot \hat{\mathbf{n}} = 0 \quad (10.12)$$

$$(\Delta \mathbf{B} \times \Delta \mathbf{V}^{\text{arb}}) \cdot \hat{\mathbf{n}} = 0 \quad (10.13)$$

We have introduced the Δ notation to indicate the jump (downstream minus upstream) in any quantity, e.g., $\Delta \mathbf{B} \equiv \mathbf{B}_d - \mathbf{B}_u$. In equations 10.11, 10.12 and 10.13 we have used the superscript *arb* (see equation 10.3) on the velocity jump to indicate that this jump can be measured in any frame (e.g., in the spacecraft frame) and not just in a shock rest frame. In principle, any pair of vectors which are dotted with $\hat{\mathbf{n}}$ in the above constraints can be used to find the shock normal. (A third constraint to uniquely determine $\hat{\mathbf{n}}$ is to make it a unit vector. The sign of $\hat{\mathbf{n}}$ is arbitrary and can be adjusted if required to make $\hat{\mathbf{n}}$ point upstream.) For example, the magnetic coplanarity normal uses the vectors in equations 10.9 and 10.10 to give

$$\hat{\mathbf{n}}_{MC} = \pm \frac{(\mathbf{B}_d \times \mathbf{B}_u) \times (\Delta \mathbf{B})}{|(\mathbf{B}_d \times \mathbf{B}_u) \times (\Delta \mathbf{B})|} \quad (10.14)$$

Magnetic coplanarity is easy to apply, but fails for $\theta_{Bnu} = 0^\circ$ or 90° . Three mixed mode normals requiring both plasma and field data are also commonly used:

$$\hat{\mathbf{n}}_{MX1} = \pm \frac{(\mathbf{B}_u \times \Delta \mathbf{V}^{\text{arb}}) \times \Delta \mathbf{B}}{|(\mathbf{B}_u \times \Delta \mathbf{V}^{\text{arb}}) \times \Delta \mathbf{B}|} \quad (10.15)$$

$$\hat{\mathbf{n}}_{MX2} = \pm \frac{(\mathbf{B}_d \times \Delta \mathbf{V}^{\text{arb}}) \times \Delta \mathbf{B}}{|(\mathbf{B}_d \times \Delta \mathbf{V}^{\text{arb}}) \times \Delta \mathbf{B}|} \quad (10.16)$$

$$\hat{\mathbf{n}}_{MX3} = \pm \frac{(\Delta \mathbf{B} \times \Delta \mathbf{V}^{\text{arb}}) \times \Delta \mathbf{B}}{|(\Delta \mathbf{B} \times \Delta \mathbf{V}^{\text{arb}}) \times \Delta \mathbf{B}|} \quad (10.17)$$

There is also an approximate normal, the velocity coplanarity normal, given by

$$\hat{\mathbf{n}}_{VC} = \pm \frac{\mathbf{V}_d^{\text{arb}} - \mathbf{V}_u^{\text{arb}}}{|\mathbf{V}_d^{\text{arb}} - \mathbf{V}_u^{\text{arb}}|} \quad (10.18)$$

which is an approximation, valid at high Mach numbers and for θ_{Bnu} near 0° or 90° for which magnetic stresses are unimportant, to an exact relationship based on the Rankine-Hugoniot relations.

Applying the Algorithms

All of these single spacecraft normals can be evaluated using the same overall procedures:

1. Select data intervals in both the upstream and downstream regions and find the “average” values of the required quantities.
2. Compute the normal using the formula. Adjust the sign of \hat{n} to point upstream if desired.
3. [Optional but recommended] Compare normals computed from different methods, using different averaging intervals, etc.

An alternative approach, which has been applied to both minimum variance analysis and computations of θ_{Bnu} , is to compute normals, θ_{Bnu} ’s, or whatever quantity of interest from pairs of individual upstream and downstream data points and then to average the result over an ensemble of such pairs. The data points for any given pair can be chosen at random from the relevant upstream or downstream set, with subsequent replacement prior to the next choice. This approach has the advantage of providing, via the ensemble statistical deviation, an error estimate of the result.

Caveats

There are several pitfalls here, including:

1. All single spacecraft methods rely on time stationarity by assuming that upstream and downstream quantities measured at different times correspond to the same shock conditions.
2. Selecting different intervals for “average” values can lead to different results. Care should be taken to ensure that the shock layer itself is entirely excluded from these intervals.
3. Many methods fail when close to the singular cases $\theta_{Bnu} = 0^\circ$ or 90° .
4. All methods assume planar, 1-D shock geometry.

10.4.3 Multi-Spacecraft Timings

If the same boundary passes several spacecraft, the relative positions and timings can be used to construct the boundary normal and speed, since

$$(V_{sh}^{arb} t_{\alpha\beta}) \cdot \hat{n} = \mathbf{r}_{\alpha\beta} \cdot \hat{n} \quad (10.19)$$

where $\mathbf{r}_{\alpha\beta}$ is the separation vector between any spacecraft pair and $t_{\alpha\beta}$ the time difference between this pair for a particular boundary. Thus given 4 spacecraft, the normal vector and normal propagation velocity $V_{sh}^{arb} \equiv V_{sh}^{arb} \cdot \hat{n}$ are found from the solution of the following system:

$$\begin{pmatrix} \mathbf{r}_{12} \\ \mathbf{r}_{13} \\ \mathbf{r}_{14} \end{pmatrix} \cdot \frac{1}{V_{sh}^{arb}} \begin{pmatrix} n_x \\ n_y \\ n_z \end{pmatrix} = \begin{pmatrix} t_{12} \\ t_{13} \\ t_{14} \end{pmatrix} \quad (10.20)$$

This problem is also addressed elsewhere within this book, e.g., in Chapters 12 (Sections 12.1.2 and 12.2), 14 (Section 14.5.2) and the entire Chapter 11.

Algorithm

Solve the system 10.20 for $\hat{\mathbf{n}}/V_{\text{sh}}^{\text{arb}}$ (to within the \pm sign arbitrariness for $\hat{\mathbf{n}}$) by any standard linear algebra technique, e.g., by inverting the matrix on the left containing the separation vectors.

Caveats

1. If the separation vectors are large, the assumption of planarity may breakdown.
2. Spacecraft positions are often given in non-stationary systems, such as GSE. Since the origin of such systems moves (especially y_{GSE}) with time, spacecraft positions should be placed onto a common coordinate system.
3. The method fails if the spacecraft are nearly coplanar. This is discussed further in Chapters 12 and 14, and illustrated numerically in Chapter 15.

10.4.4 Combined Approaches

Provided some multi-spacecraft timings are available, it is possible to add to the system 10.20 any (or all) of the constraints given in Section 10.4.2. For example, consider the system

$$\mathbf{A} \cdot (\hat{\mathbf{n}}/V_{\text{sh}}^{\text{arb}}) \equiv \begin{pmatrix} \mathbf{r}_{12} \\ \mathbf{r}_{13} \\ \mathbf{r}_{14} \\ \Delta \mathbf{B} \\ \Delta \mathbf{B} \times \Delta \mathbf{V}^{\text{arb}} \end{pmatrix} \cdot \frac{1}{V_{\text{sh}}^{\text{arb}}} \begin{pmatrix} n_x \\ n_y \\ n_z \end{pmatrix} = \begin{pmatrix} t_{12} \\ t_{13} \\ t_{14} \\ 0 \\ 0 \end{pmatrix} \quad (10.21)$$

which makes use of equations 10.9 and 10.13.

Algorithm

The system 10.21 of equations is over-determined. The least squares solution which minimises the residuals on the right-hand side can be obtained by multiplying on the left throughout by the transpose of the matrix of coefficients \mathbf{A} and solving the resulting 3×3 square system for $\hat{\mathbf{n}}/V_{\text{sh}}^{\text{arb}}$ (to within the \pm sign arbitrariness for $\hat{\mathbf{n}}$). This approach has the advantage in that it can fold in more information and can be used when not all the quantities are known (e.g., when one spacecraft is missing).

Caveats

1. The least squares solution as described above takes no account of the relative errors or confidence in the various coefficients contained in \mathbf{A} . More general inversion techniques (e.g., singular value decomposition) provide some error analyses, and can weight the different constraint equations differently to reduce the residual errors. There are probably methods which can include the possibility of different error estimates for individual components of \mathbf{A} , but they have not yet been applied to these kinds of problems.

2. Some caution should be taken to ensure that the same information is not included many times in extending the system of equations to be solved.
3. The nature of the information contributed by multi-spacecraft timing depends upon the spatial geometry of the polyhedron defined by the spacecraft (see Chapter 12). The effect of this upon the combined approach merits further study.

10.4.5 Shock Jump Conditions

The previous methods use a small subset of the Rankine-Hugoniot relations and/or multi-spacecraft timings to determine the shock normal. A potentially more reliable approach is to take more of the Rankine-Hugoniot relations into account in order to establish a full set of upstream and downstream quantities (including the shock normal direction) which best satisfy these physical laws. Since the thermal properties of the shock processes often involve kinetic and anisotropic processes, multi-species, etc., it is probably best to avoid as much as possible those relations that involve the plasma pressure, namely the \hat{n} -momentum equation and the energy flux equation. The pressure jump can then be used to verify that the resulting solution does in fact represent a compressive, entropy-producing shock.

Lepping and Argentiero first put such a scheme together in 1971, although their method still relied on magnetic coplanarity to establish the shock normal direction. Viñas and Scudder (VS) overcame this difficulty in 1986. Although the method is too lengthy to be repeated here, we describe the overall philosophy and approach for reference.

VS begin with a set of pairs of measurements of the plasma parameters (ρ , \mathbf{V} , \mathbf{B}) on either side of the shock in an arbitrary frame of reference. Equation 10.3 is used to write the Rankine-Hugoniot relations for mass flux, normal magnetic field, tangential stress, and tangential electric field in an arbitrary frame of reference. The shock speed, V_{sh}^{arb} , enters linearly in the mass flux balance equation and can be eliminated. Treating the plasma parameters as known (from the observations) leads to a system of 7 equations in which the only unknowns are the two angles which define the shock normal direction. This nonlinear system is solved via a least squares method for these two angles. Once the normal is found, the shock speed can be found from the mass flux equation as the average of the shock speed inferred from individual pairs of up- and downstream measurements. This solution minimises in a least squares sense the residuals from the shock speed as inferred by the individual pairs.

Next, in a manner identical to that used to find V_{sh}^{arb} , the conservation constants for the mass flux, normal magnetic field, tangential stress, and tangential electric field can be evaluated as the averages of their values deduced from individual pairs of observations.

Next, using these conservation constants, VS set up a least squares problem to establish the self-consistent asymptotic states via two vector equations for \mathbf{V} and \mathbf{B} parameterised in terms of the mass density. The value of ρ which minimises the residuals in these equations then determines the asymptotic values of the various parameters.

Finally, the normal momentum equation can be used to test that the inferred jump in thermal pressure, as demanded by the asymptotic values of the other parameters, has the correct sign.

Algorithm

See the VS reference for a more detailed account of the algorithm.

Caveats

1. These methods also rely on the selection of suitable upstream and downstream intervals of data, and on pairing an upstream observation with a downstream one.
2. The usual caveats concerning stationarity and planarity apply.
3. The potential advantage of bringing more plasma parameters to bear on the problem has the potential disadvantage that it places more emphasis on quantities, such as the plasma mass density, which may not be particularly accurate.

10.4.6 Model Boundary Equations

In the case of the Earth's bow shock and magnetopause and, to perhaps a lesser extent, those at other planets, the large number of spacecraft encounters enables us to define the shape of the boundary surface on a statistical basis. Such statistical data sets can be fit by simple geometrical forms, from which the normal direction can be computed analytically. This method is straightforward, and the algorithm is described below. In many instances, such normals are likely to be as accurate as any of those computed above.

To begin, choose an appropriate model. Most are cylindrically symmetric conic sections which can thus be represented in the form

$$\frac{L}{r^{abd}} = 1 + \epsilon \cos \theta^{abd} \quad (10.22)$$

in which L is the semilatus rectum and ϵ the eccentricity of the conic. The variables r^{abd} and θ^{abd} are polar coordinates in the natural system for the conic. This natural system is aberrated by an angle α from GSE by, e.g., the Earth's orbital motion (30 km/s so that $\tan \alpha = 30 \text{ km/s} / V_{\text{solar wind}}$) and then perhaps displaced from the Earth's centre. Thus the relevant variable transformation is

$$\begin{pmatrix} x^{abd} \\ y^{abd} \\ z^{abd} \end{pmatrix} = \begin{pmatrix} \cos \alpha & -\sin \alpha & 0 \\ \sin \alpha & \cos \alpha & 0 \\ 0 & 0 & 1 \end{pmatrix} \cdot \begin{pmatrix} x \\ y \\ z \end{pmatrix} - \begin{pmatrix} x_o \\ y_o \\ z_o \end{pmatrix} \quad (10.23)$$

where \mathbf{r}_o is the displacement of the focus of the conic in the aberrated frame and \mathbf{r} is a position vector in the relevant observational frame (e.g., GSE). This configuration is sketched in Figure 10.3.

Different models give different values for ϵ , L , and \mathbf{r}_o . Some popular models are tabulated in Table 10.1. These models are also shown in Figure 10.4 for comparison. In any application, it may also be necessary to scale the distances (L and \mathbf{r}_o) which appear in the models. For example, the bow shock and magnetopause respond to changes in the dynamic pressure in the solar wind. The expected spatial variation is proportional to the dynamic pressure to the power $-1/6$, due to the balance with the magnetic pressure which the Earth's dipole field is able to exert. An alternative approach is to scale the distances to make the model pass through an observed location.

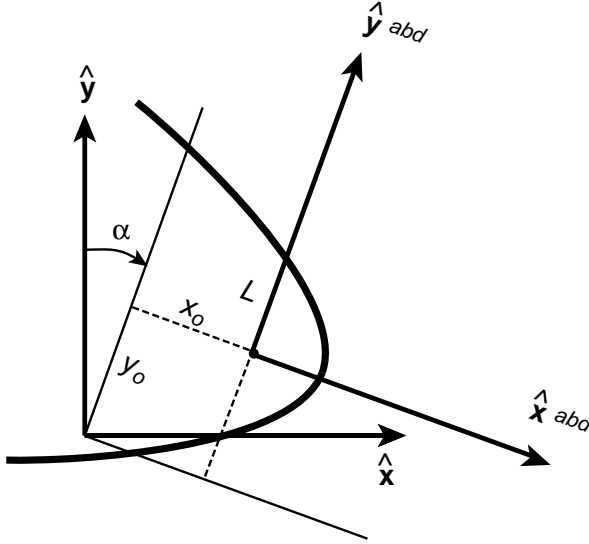


Figure 10.3: Observational (e.g., GSE) coordinate system (x, y) and aberrated-displaced system (x^{abd}, y^{abd}) for model boundaries. For simplicity, only a two-dimensional system is shown. Note that the direction of α is defined such that positive values correspond to the aberration due to the Earth's orbital motion, which is in the $-\hat{y}$ direction.

Algorithm

1. Choose an appropriate model. For measurements under unusual circumstances or at high latitudes, consider the parameterised 3-D models by *Peredo et al.* and *Roelof and Sibeck* for the bow shock and magnetopause respectively.
2. Calculate the aberration angle as given in the model or via a measurement of the solar wind speed. If solar wind data is not present, a typical value of 450 km/s, corresponding to $\alpha = 3.8^\circ$, is usually adequate.
3. Scale L and r_o as necessary. One way is by the $-1/6$ power of the solar wind ram pressure, normalised to the model mean as given in Table 10.1. This process is not particularly accurate and produces far less variation in, say, the bow shock position than is actually observed. Thus well upstream of the bow shock the absolute position of the bow shock can be uncertain by several R_E . Alternatively scale these parameters so that the model passes through a given position vector $\mathbf{r}_{\text{crossing}}$. This process can be reduced to the substitutions $L \rightarrow \sigma L$, $r_o \rightarrow \sigma r_o$, and $\mathbf{r} \rightarrow \mathbf{r}_{\text{crossing}}$ in equations 10.22 and 10.23 and solving the resulting quadratic equation for σ . In the case of hyperbolic ($\epsilon > 1$) models, the larger of the two roots corresponds to the correct branch of the hyperbola.
4. Calculate the gradient, ∇S , to the surface given by the model. This surface may be

Table 10.1: Parameter values for various model surfaces

Source Units	ϵ	L R_E	x_o R_E	y_o R_E	α $^\circ$	ρV_{sw}^2 nPa
<i>Terrestrial Bow Shock Models</i>						
Peredo et al., $z = 0$	0.98	26.1	2.0	0.3	$\alpha_o - 0.6$	3.1
Slavin and Holzer mean	1.16	23.3	3.0	0.0	α_o	2.1
Fairfield Meridian 4°	1.02	22.3	3.4	0.3	4.8	?
Fairfield Meridian No 4°	1.05	20.5	4.6	0.4	5.2	?
Formisano Unnorm. $z = 0$	0.97	22.8	2.6	1.1	3.6	3.7
Farris et al.	0.81	24.8	$\equiv 0$	$\equiv 0$	3.8	1.8
<i>Terrestrial Magnetopause Models</i>						
Roelof and Sibeck (F_{00} only)	0.91	11.2	4.82	$\equiv 0$	α_o	2.1
Fairfield Meridian 4°	0.79	13.1	3.6	0.4	-0.3	?
Fairfield Meridian No 4°	0.80	12.8	3.9	0.6	-1.5	?
Farris et al.	0.43	14.7	$\equiv 0$	$\equiv 0$	3.8	1.8
Petrinec et al., ($B_z > 0$)	0.42	14.6	$\equiv 0$	$\equiv 0$	α_o	~ 2.5
Petrinec et al., ($B_z < 0$)	0.50	14.6	$\equiv 0$	$\equiv 0$	α_o	~ 2.5
Formisano						
Unnorm. $z = 0$	0.82	12.5	4.1	0.1	4.2	3.7
Norm. $z = 0$	0.69	13.5	0.9	-0.4	6.6	3.7

Notes: The above table shows aberration angles positive when in the nominal sense for the Earth's orbital motion, as shown in Figure 10.3. The value α_o indicates that each data point was aberrated by the amount corresponding to the prevailing solar wind speed. The average solar wind dynamic pressure, where available, is also shown. All models have $z_o \equiv 0$. Non-axially symmetric models have been reduced to polar form after setting $z = 0$ in the model equation.

written

$$S(\mathbf{r}^{abd}(x, y, z)) \equiv \left(r^{abd} + \epsilon x^{abd} \right)^2 - L^2 = 0 \quad (10.24)$$

This gradient, expressed in terms of the aberrated coordinates \mathbf{r}^{abd} but rotated back into the unaberrated coordinate frame can be written

$$\left(\frac{r^{abd}}{2L} \nabla S \right) = \begin{pmatrix} [x^{abd}(1 - \epsilon^2) + \epsilon L] \cos \alpha + y^{abd} \sin \alpha \\ -[x^{abd}(1 - \epsilon^2) + \epsilon L] \sin \alpha + y^{abd} \cos \alpha \\ z^{abd} \end{pmatrix} \quad (10.25)$$

5. The normal $\hat{\mathbf{n}}$ is parallel to this gradient, i.e.,

$$\hat{\mathbf{n}} = \pm \frac{\nabla S}{|\nabla S|} \quad (10.26)$$

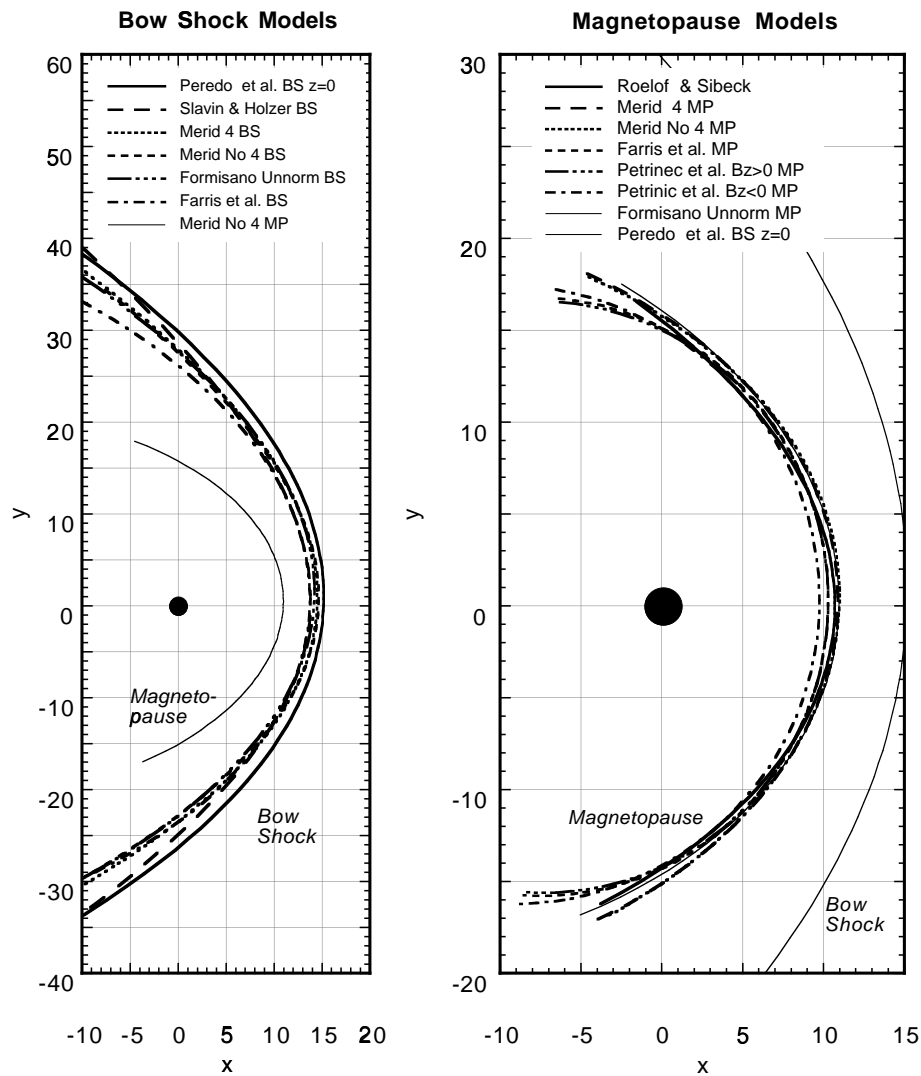


Figure 10.4: Dayside portions of model terrestrial bow shocks (left) and magnetopauses (right) based on the parameters given in Table 10.1. The models have not been scaled and have been aberrated by the amount shown in the table with $\alpha_o = 3.8^\circ$. For the purposes of estimating shock normals, most models agree to a fair degree. The magnetopause tailward of the terminator ($x = 0$) and/or at high latitudes is more complex, and varies most with interplanetary conditions. Some models include non-axially symmetric terms. These are shown in the ecliptic plane only (see the bibliography for more details).

Caveats

1. There may be ripples or transients which distort \hat{n} from the model value.
2. Scaling the distances can be imprecise when the spacecraft is not at an actual crossing.
3. There are a multiplicity of models although, as shown in Figure 10.4, there is not a large variation in the normal direction at a given position, at least for the dayside portions and under nominal interplanetary conditions.
4. At greater tailward distances the bow shock models need to be modified to asymptote to the fast mode Mach cone (see Slavin, J. A., Holzer, R. E., Spreiter, J. R., and Stahara, S. S., Planetary Mach cones: Theory and observation, *J. Geophys. Res.*, **89**, 2708–2714, 1984).

10.4.7 Tangential Discontinuities

In the case of pure tangential discontinuities, it is possible to find the normal to the discontinuity by simply noting that both upstream and downstream magnetic field vectors are parallel to the shock plane and, unlike the case of a perpendicular shock, are not in general parallel to one another. Thus in this case the normal is given by

$$\hat{n} = \pm \frac{\mathbf{B}_u \times \mathbf{B}_d}{|\mathbf{B}_u \times \mathbf{B}_d|} \quad (10.27)$$

Caveats

The problem here is to use sufficient plasma and field variations, particularly the pressures, to establish that the discontinuity in question is indeed a tangential discontinuity and not a rotational discontinuity or a slow shock. Since all the pressures rise across a fast shock, it is usually easier to distinguish these from the pressure balance structures required by tangential discontinuities, although weak shocks may again prove difficult.

10.4.8 Rotational Discontinuities

If sufficient field resolution is present, minimum variance analysis of the magnetic field, or maximum variance of the electric field, provides an estimate of the normal direction. The caveats given above concerning the identification of tangential discontinuities also apply here.

10.5 Determination of the Shock/Discontinuity Speed

The shock speed along the normal, $V_u \cdot \hat{n}$, is a vital parameter, as it determines the shock Mach number. There are a variety of methods in use to calculate this speed from observational data, or to calculate the shock speed relative to an arbitrary observational frame, $V_{sh}^{arb} \hat{n}$, which are summarised here. These velocities are related by equation 10.3. Most of the following methods also require knowledge of the shock normal.

10.5.1 Mass Flux Algorithm

Writing the shock mass flux conservation equation in terms of quantities measured in an arbitrary frame yields

$$\rho_u(\mathbf{V}_u^{\text{arb}} - V_{\text{sh}}^{\text{arb}} \hat{\mathbf{n}}) \cdot \hat{\mathbf{n}} = \rho_d(\mathbf{V}_d^{\text{arb}} - V_{\text{sh}}^{\text{arb}} \hat{\mathbf{n}}) \cdot \hat{\mathbf{n}} \quad (10.28)$$

This equation can be solved for $V_{\text{sh}}^{\text{arb}}$ to give

$$V_{\text{sh}}^{\text{arb}} = \frac{\Delta(\rho \mathbf{V}^{\text{arb}})}{\Delta\rho} \cdot \hat{\mathbf{n}} \quad (10.29)$$

Caveats

The only difficulty applying equation 10.29 is its reliance on good plasma density measurements on both sides of the shock.

10.5.2 Shock Foot Thickness Algorithm

Quasi-perpendicular supercritical collisionless shocks initiate their dissipation process by reflecting a portion of the incoming ion distribution. Under these geometries, such reflected ions gyrate around the magnetic field in front of the main shock ramp and return to the shock. The extent of this foot region in front of such shocks is directly related to these reflected ion trajectories which, in turn, are simply related to the incident normal velocity (in a shock rest frame), the strength of the field, and the shock geometry. The shock foot is clearly visible as a gradual rise in the magnetic field, and can also be found in the commencement of ion-acoustic-like noise upstream of the shock ramp. Assuming the incident ions are specularly reflected at the shock, and neglecting their thermal motion, the reflected ions reach their maximum upstream excursion and turn around after a time t_{turn} which is the solution to

$$\cos(\Omega t_{\text{turn}}) = \frac{1 - 2 \cos^2 \theta_{Bnu}}{2 \sin^2 \theta_{Bnu}} \quad (10.30)$$

where Ω is the ion gyrofrequency. At this time, their distance, d_{foot} , along the normal from the shock ramp is

$$\frac{d_{\text{foot}}}{(\mathbf{V}_u \cdot \hat{\mathbf{n}})/\Omega} \equiv f(\theta_{Bnu}) = \Omega t_{\text{turn}} \left(2 \cos^2 \theta_{Bnu} - 1 \right) + 2 \sin^2 \theta_{Bnu} \sin(\Omega t_{\text{turn}}) \quad (10.31)$$

Knowing the time Δt_{foot} taken for the foot to pass over a point in an arbitrary observational frame (e.g., the spacecraft frame) then provides the additional information required, resulting in

$$\mathbf{V}_u \cdot \hat{\mathbf{n}} = \frac{\mathbf{V}_u^{\text{arb}} \cdot \hat{\mathbf{n}}}{1 \mp (f(\theta_{Bnu})/\Omega \Delta t_{\text{foot}})} \quad (10.32)$$

The upper (−) sign is used when the observed transition is from upstream to downstream and the lower (+) when the observed sequence is downstream to upstream.

Algorithm

1. Measure the foot passage time Δt_{foot} . Determine θ_{Bnu} from the data by either finding \hat{n} and \mathbf{B}_u or by an algorithm given below in Section 10.6.
2. Compute $f(\theta_{Bnu})$.
3. Apply equation 10.32.

Caveats

1. This is quite a good method for these shocks, but requires a good methodology for identifying the foot region.
2. The method requires that the relative shock/observer motion be steady during the passage of the foot.

10.5.3 Multi-Spacecraft Timing Algorithm

This approach for determining shock speeds is detailed above in Sections 10.4.3 and 10.4.4.

10.5.4 Smith and Burton Algorithm

Smith and Burton have derived an algorithm to determine the shock speed which does not require an explicit calculation of the shock normal. Their algorithm is derived from the Rankine-Hugoniot relation which represents continuity of the tangential electric field [$-\hat{n} \times (\mathbf{V}_u \times \mathbf{B}_u) = -\hat{n} \times (\mathbf{V}_d \times \mathbf{B}_d)$ in our notation]. Some manipulation and implicit use of the coplanarity theorem yields

$$|\mathbf{V}_u \cdot \hat{n}| = \frac{|\Delta \mathbf{V}^{\text{arb}} \times \mathbf{B}_d|}{|\Delta \mathbf{B}|} \quad (10.33)$$

Caveats

1. This algorithm requires a good vector determination of $\Delta \mathbf{V}^{\text{arb}}$.
2. The algorithm works for all shock geometries, including parallel “switch on” shocks, but breaks down for parallel acoustic shocks, which have $\mathbf{B}_d = \mathbf{B}_u$.

10.5.5 Generalised deHoffmann-Teller Transformation

Chapter 9 shows how the transformation velocity from an arbitrary spacecraft frame to the deHoffmann-Teller frame can be found if there are good determinations of the flow velocity and field on both sides of the shock transition. The shock velocity along the normal $V_{\text{sh}}^{\text{arb}}$ is simply the normal component of the generalised HT-frame velocity, as shown in Figure 9.2 (page 225) of Chapter 9.

Algorithm

1. Use the methods in Chapter 9 to determine the velocity of the deHoffmann-Teller frame with respect to an arbitrary (e.g., spacecraft) frame.
2. Take the normal component of this velocity

10.5.6 Velocity of a Tangential Discontinuity

Since a tangential discontinuity has, by definition, zero mass flux through the discontinuity layer, in a rest frame moving with the discontinuity the flow velocities on either side have only tangential components. Thus in an arbitrary frame of reference, the component of the flow velocity normal to the discontinuity must equal the speed of the discontinuity, i.e.,

$$V_{TD}^{\text{arb}} = \mathbf{V}^{\text{arb}} \cdot \hat{\mathbf{n}} \quad (10.34)$$

This should produce the same result regardless of whether the upstream or downstream flow velocity is used.

Algorithm

1. Apply equation 10.34 using a measured flow velocity and a normal vector determined by, e.g., equation 10.27.
2. [Optional] Apply equation 10.34 for flow velocities measured both upstream and downstream of the discontinuity to provide some estimate of the error.

Caveats

1. The difference between values computed using upstream and downstream values may be due to either errors in the measured flow velocities, in the discontinuity normal determination, or both.

10.6 Determining θ_{Bnu} **10.6.1 Application of the Shock Normal****Algorithm**

1. Determine $\hat{\mathbf{n}}$ via one of the algorithms given in Section 10.4.
2. Determine \mathbf{B}_u in a manner consistent with that used for $\hat{\mathbf{n}}$. That is, if averages of upstream parameters are used to find $\hat{\mathbf{n}}$, use the same \mathbf{B}_u .
3. $\theta_{Bnu} = \cos^{-1}(\hat{\mathbf{n}} \cdot \mathbf{B}_u / |\mathbf{B}_u|)$.

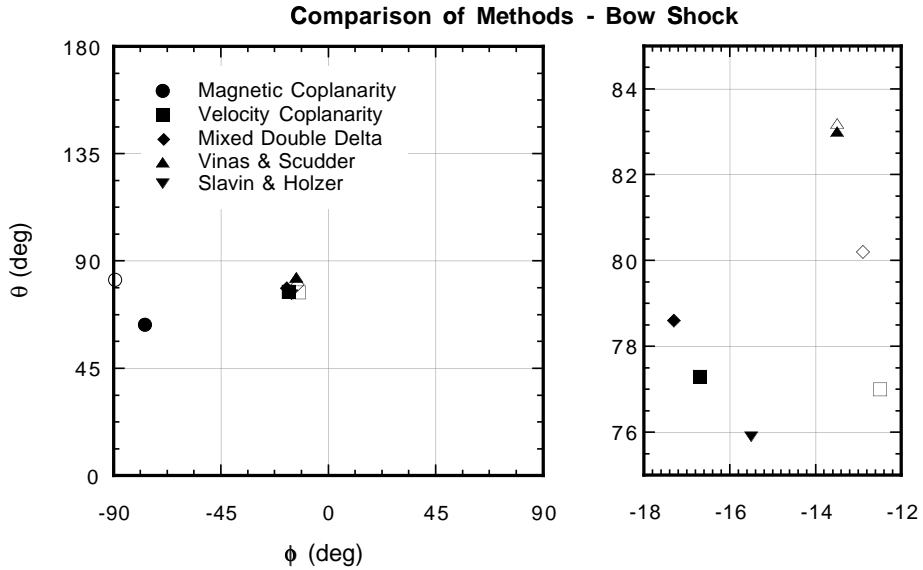


Figure 10.5: Normal directions based on several of the methods given in this chapter as applied to a particular crossing of the Earth’s bow shock. The normals are given in terms of their spherical polar angles in the GSE frame: $\theta = 0^\circ$ is along \hat{z} while $\phi = 0^\circ$ is the sunward direction. Open symbols use the same methods as their solid counterparts but use input up- and downstream parameters averaged over fewer observational data points. Note the failure of magnetic coplanarity in this case due to the near perpendicular shock geometry ($\theta_{Bnu} \sim 80^\circ$ in this case). The right panel shows in detail the spread of the other methods and reveals a typical uncertainty of $5 - 10^\circ$ or more.

10.6.2 Ensemble θ_{Bnu}

Algorithm

1. Apply the previous algorithm (e.g., using \hat{n}) to pairs of upstream and downstream data points.
2. Ensemble average the results to provide $\langle \theta_{Bnu} \rangle$ together with its standard deviation.

10.7 Application

As an example of many of the above methods, we present here the results of normal determination for a particularly well-studied example of the Earth’s bow shock. Figure 10.5 shows the results of several coplanarity-like techniques, the Rankine-Hugoniot solution of Viñas and Scudder, and a bow shock model normal. Most methods agree to within $5 - 10^\circ$, although the magnetic coplanarity technique fails badly here due to the near perpendicular orientation of the upstream field in this case. Of course, the determination of θ_{Bnu} itself

varies with the different normals, and thus has a similar uncertainty. The choice of method may depend on the particular case. Ideally, different methods should be compared with one another.

Bibliography

Two AGU Monographs provide a wealth of information relating to shock physics:

Kennel, C. F., A quarter century of collisionless shock research, in *Collisionless Shocks in the Heliosphere: A Tutorial Review*, edited by R. G. Stone and B. T. Tsurutani, AGU Monograph 34, pp. 1–36, American Geophysical Union, Washington, D. C., 1985, covers in some detail the various shock modes, critical Mach numbers, and the like.

Tsurutani, B. T. and Stone, R. G., editors, *Collisionless Shocks in the Heliosphere: Reviews Of Current Research*, AGU Monograph 35, American Geophysical Union, Washington, D. C., 1985, is the companion monograph containing much useful material.

A good basic introduction to collisionless shocks and discontinuities, including the Rankine-Hugoniot relations, shock frames and geometry, and shock structure/processes can be found in:

Burgess, D., Collisionless shocks, in *Introduction to Space Physics*, edited by M. G. Kivelson and C. T. Russell, pp. 129–163, Cambridge University Press, Cambridge, U. K., 1995.

The seminal work on the subject of shock normal determination by multiple satellite techniques, and a work that applies it in detail, are:

Russell, C. T., Mellott, M. M., Smith, E. J., and King, J. H., Multiple spacecraft observations of interplanetary shocks: Four spacecraft determination of shock normals, *J. Geophys. Res.*, **88**, 4739–4748, 1983.

Russell, C. T., Gosling, J. T., Zwickl, R. D., and Smith, E. J., Multiple spacecraft observations of interplanetary shocks: ISEE three-dimensional plasma measurements, *J. Geophys. Res.*, **88**, 9941–9947, 1983.

The Rankine-Hugoniot approach to shock normal and parameter determination is detailed in:

Viñas, A. F. and Scudder, J. D., Fast and optimal solution to the “Rankine-Hugoniot problem”, *J. Geophys. Res.*, **91**, 39–58, 1986.

Scudder, J. D., Mangeney, A., Lacombe, C., Harvey, C. C., Aggson, T. L., Anderson, R. R., Gosling, J. T., Paschmann, G., and Russell, C. T., The resolved layer of a collisionless, high β , supercritical quasi-perpendicular shock wave, 1. Rankine-Hugoniot geometry, currents, and stationarity, *J. Geophys. Res.*, **91**, 11 019–11 052, 1986, contains the data for the comparison shown in Figure 10.5.

A partial approach to the Rankine-Hugoniot problem was presented originally by:

Lepping, R. P. and Argentiero, P. D., Single spacecraft method of estimating shock normals, *J. Geophys. Res.*, **76**, 4349, 1971

The above approach was modified further in:

Acuña, M. H. and Lepping, R. P., Modification to shock fitting program, *J. Geophys. Res.*, **89**, 11 004, 1984.

The velocity coplanarity method is presented in:

Abraham-Shrauner, B., Determination of magnetohydrodynamic shock normals, *J. Geophys. Res.*, **77**, 736, 1972.

Abraham-Shrauner, B. and Yun, S. H., Interplanetary shocks seen by Ames probe on Pioneer 6 and 7, *J. Geophys. Res.*, **81**, 2097, 1976, *applied it further*

Smith, E. J. and Burton, M. E., Shock analysis: 3 useful new relations, *J. Geophys. Res.*, **93**, 2730–2734, 1988, *who proposed additional relations.*

The boundary models shown in Table 10.1 have been derived from the following references:

Peredo, M., Slavin, J. A., Mazur, E., and Curtis, S. A., Three-dimensional position and shape of the bow shock and their variation with Alfvénic, sonic and magnetosonic Mach numbers and interplanetary magnetic field orientation, *J. Geophys. Res.*, **100**, 7907–7916, 1995. [*The parameters shown in Table 10.1 correspond to their full M_A range $< 2 - 20 >$, p -normalised and GIPM rotated.*]

Slavin, J. A. and Holzer, R. E., Solar wind flow about the terrestrial planets, 1. modelling bow shock position and shape, *J. Geophys. Res.*, **86**, 11 401–11 418, 1981. [*The values in Table 10.1 correct errors in the original paper. Also, there is an unrelated typographical error there: the k'_3 equation should have a “+” sign in front of the middle (k_2) term. (J. A. Slavin, private communication, 1997).*]

See also Slavin, J. A., Holzer, R. E., Spreiter, J. R., and Stahara, S. S., Planetary Mach cones: Theory and observation, *J. Geophys. Res.*, **89**, 2708–2714, 1984.

Fairfield, D. H., Average and unusual locations of the Earth’s magnetopause and bow shock, *J. Geophys. Res.*, **76**, 6700–6716, 1971.

Formisano, V., Orientation and shape of the Earth’s bow shock in three dimensions, *Planet. Space Sci.*, **27**, 1151–1161, 1979.

Farris, M. H., Petrinec, S. M., and Russell, C. T., The thickness of the magnetosheath: Constraints on the polytropic index, *Geophys. Res. Lett.*, **18**, 1821–1824, 1991.

Roelof, E. C. and Sibeck, D. C., Magnetopause shape as a bivariate function of interplanetary magnetic field b_z and solar wind dynamic pressure, *J. Geophys. Res.*, **98**, 21 421–21 450, 1993.

Roelof, E. C. and Sibeck, D. C., Correction to “magnetopause shape as a bivariate function of interplanetary magnetic field b_z and solar wind dynamic pressure”, *J. Geophys. Res.*, **99**, 8787–8788, 1994.

See also Holzer, R. E. and Slavin, J. A., Magnetic flux transfer associated with expansions and contractions of the dayside magnetosphere, *J. Geophys. Res.*, **83**, 3831, 1978, and Holzer, R. E. and Slavin, J. A., A correlative study of magnetic flux transfer in the magnetosphere, *J. Geophys. Res.*, **84**, 2573, 1979.

Petrinec, S. P., Song, P., and Russell, C. T., Solar cycle variations in the size and shape of the magnetopause, *J. Geophys. Res.*, **96**, 7893–7896, 1991.

The shock foot thickness method for determining shock speeds is described in:

Gosling, J. T. and Thomsen, M. F., Specularly reflected ions, shock foot thickness, and shock velocity determination in space, *J. Geophys. Res.*, **90**, 9893, 1985.

The ensemble approach to θ_{Bnu} determination was introduced by:

Balogh, A., González-Esparza, J. A., Forsyth, R. J., Burton, M. E., Goldstein, B. E., Smith, E. J., and Bame, S. J., Interplanetary shock waves: Ulysses observations in and out of the ecliptic plane, *Space Sci. Rev.*, **72**, 171–180, 1995.

— 11 —

Multi-Spacecraft Discontinuity Analysis: Orientation and Motion

MALCOLM W. DUNLOP AND TIMOTHY I. WOODWARD

*Imperial College of Science, Technology and Medicine
London, United Kingdom*

11.1 Introduction

In this chapter we discuss the shape, orientation, and motion of discontinuities in the magnetic field, as sampled by an array of closely separated spacecraft. Our central purpose is to introduce an established analysis method, designed to determine these properties. We describe the use of the technique and its application to simulated as well as real data. We first discuss the methodology adopted and then briefly present the application to both planar and non-planar event models in situations of linearly accelerating motion. The ability of the analysis to distinguish between non-constant motion and surface curvature is discussed. It is stressed that this chapter concentrates on the analysis of magnetic field data alone, since methods which incorporate multi-instrument data are predominantly limited to single spacecraft techniques at the present time (see, for example, Chapters 8 and 9).

Discontinuity analysis is used generically to describe the set of multi-spacecraft magnetometer analysis methods, which are best used in situations when the dominant, or most interesting, event scale length is much shorter than the spacecraft separation distances (the boundary structure is thin). As we discuss below, for such events where the spatial sampling is well below the effective Nyquist sampling rate, one first seeks parameters describing macroscopic properties of the event. As a secondary aim, one can assess the possibility of a more detailed analysis of structure. In ordering the data set, the analysis relies on an interpretation of the spatial content in the individual time series of each spacecraft in terms of its variance, and via the solenoidality of the field [Sonnerup and Cahill, 1967]. The thin boundary regime allows this coordination, within each spacecraft data set, to be made independently from comparisons across the spacecraft array. Additionally, boundary structure here is typically assumed to be quasi-static in its own frame, but moving with respect to the spacecraft (we refer to this as “the degree of stationarity”). Thus, macroscopic analysis requires two things: firstly, a limit on the degree of time dependence in the data to be useful, and secondly, that the data can be characterised in terms of a “thin boundary”.

It is interesting to note in passing that an opposing data regime occurs when the event scale length of variation is well in excess of the spacecraft separations. Direct differencing of inter-spacecraft measurements can then usefully approximate the differential changes in the field locally between spacecraft. In other words, the spatial content in the time series

for any one spacecraft is then less critical and coordination of the data is primarily across spacecraft (and is purely a multipoint analysis for each quantity).

In fact, *Dunlop et al.* [1988] have identified three broad characterisations of magnetic field analysis pertinent to a Cluster-like mission; namely, the Curlometer, the Wave Telescope and the Discontinuity Analyser. The curlometer method (covered in detail in Chapter 16, for example) is an analysis technique best used when high spatial sampling rates can be achieved. On the other hand, when event scales are of the order of the spacecraft separations, Fourier domain analysis techniques play an important role, such as the wave telescope or mode filter (see, for example, Chapters 3 and 4, for current analysis methods applying to space data). Such techniques rely on successful coordination of the data across both spacecraft positions and channels (instruments and components) simultaneously. Together with the discontinuity analyser, these techniques correspond to the three ranges of spacecraft separation distances relative to scale lengths of the physical structure sampled.

The Discontinuity Analyser formerly existed only in outline [*Dunlop et al.*, 1988] until *Mottez and Chanteur* [1994] described the possibility of using the Cluster array to determine local geometric properties relating to the surface of a discontinuity. Although there are few other direct references to a discontinuity analyser for a phased array such as Cluster, an early design of discontinuity analysis was made by *Russell et al.* [1983], who describe a method of accurate determination of shocks using four widely separated spacecraft, averaging the normals determined at each different spacecraft while assuming constant velocity of the planar discontinuity. (See section 10.4.3 for a discussion of this method.)

Boundary structures, including shocks, current sheets and other forms of discontinuities at plasma boundaries, have been studied extensively in the past (particularly, the Earth's bow shock and magnetopause), predominantly using single spacecraft data and sometimes by using multiple data sets from different instruments. A few dual spacecraft studies have been carried out, namely the ISEE-1 and -2 probes and the AMPTE/UKS and IRM pair, and some for special events covered fortuitously by largely separated and uncoordinated spacecraft.

It is known from single and dual spacecraft studies that the complexity of the physical structure (microscopic properties), together with its evolution in time, limits the macroscopic features which can be obtained from data on the boundary (its shape, orientation and motion). Typically, under assumptions of local planarity and some degree of stationarity [*Bendat and Piersol*, 1986] for the sampled event, little more than the orientation of the discontinuity, or perhaps simple motion for particular events, can be obtained with single spacecraft.

Additionally, the quality of these estimates depends upon the detailed microstructure in the sampled discontinuity. These include the presence of natural noise, wave or other properties, which can confuse simple model assumptions for the boundary unless either the structure is sampled fortuitously or judicious spectral filtering is possible. The latter will depend upon the degree to which the properties are in some sense conflicting in their effect.

Furthermore, no unambiguous check on consistency, if the assumptions are wrong, is available from single spacecraft information, except in special circumstances. Single spacecraft analysis usually combines data across measurement channels (field and plasma quantities and components). For example, minimum variance analysis (MVA) is over the components of the magnetic field through the requirement of solenoidality, and when

planarity is assumed MVA yields the boundary normals (Chapter 8).

The effect of extra anomalous physical structure in the boundary therefore provides a context for the analysis and needs to be known at some level of detail. The general problem posed in extending the analysis to multi-spacecraft data is how to combine the information self-consistently when the single spacecraft assumptions on the macroscopic behaviour are relaxed.

11.2 Methodology

11.2.1 General assumptions

The specific technique applied in Section 11.4 is currently built on a method which seeks to find a planar discontinuity (local to the spacecraft configuration), or to determine the degree of non-planarity, based on limited knowledge of the boundary motion. This application of discontinuity analysis is termed the planar-DA. The method emphasises, and can determine, motional properties when the boundary is found to be planar. Otherwise, only qualitative indications of curvature result, in the limit of small deviations from planarity (on the spacecraft separation scale).

Whether any analysis is able to distinguish between non-constant motion and surface curvature is a more subtle issue and needs some interpretation (this point is discussed later in Section 11.5). A determination of the surface properties, curvature and orientation, in fact requires an assumption of boundary motion; typically of constant velocity, and is the approach discussed in Section 11.5.2. Such an assumption is extremely difficult to check, however, and is one reason the planar-DA approach is taken.

Another important aspect driving the cautious approach adopted here is that low noise levels on the data signature (arising from either instrument effects or additional signal) are required to show the significance of either non-constant motion or curvature. The role of noise is demonstrated and discussed below in Section 11.4.3 and 11.5.5.

This form of the Discontinuity Analyser has been tested with simple event modelling tools, which reflect a number of structural properties of the field and form part of the software tools. We also employ a number of simple, geometric models for static fields. These are functions in space, in the form of: discontinuities with or without constant field magnitude, current sheets and 2-D curved boundaries based on scalar potential functions. We have simulated data from these using a number of trajectories of up to four spacecraft, and with a constant acceleration on the motion. In the absence of truly coordinated, multi-spacecraft data, such event models are a key test for the technique, but are not integral to the method. We describe, in Section 11.A, the models used here to simulate data, together with (in Section 11.4) results arising from a variety of runs with these.

In addition, during real event selection, it may enhance performance to filter any anomalous (to the basic boundary structure under investigation) signal if possible. Any filtering applied to the data will depend on the event properties (or assumptions of these), but may not necessarily help analysis quality. This, and other forms of preparation of the data to enhance the analysis quality, is important, but is only briefly treated here. We regard it as a single spacecraft analysis for which there is much description in Chapter 2 and elsewhere.

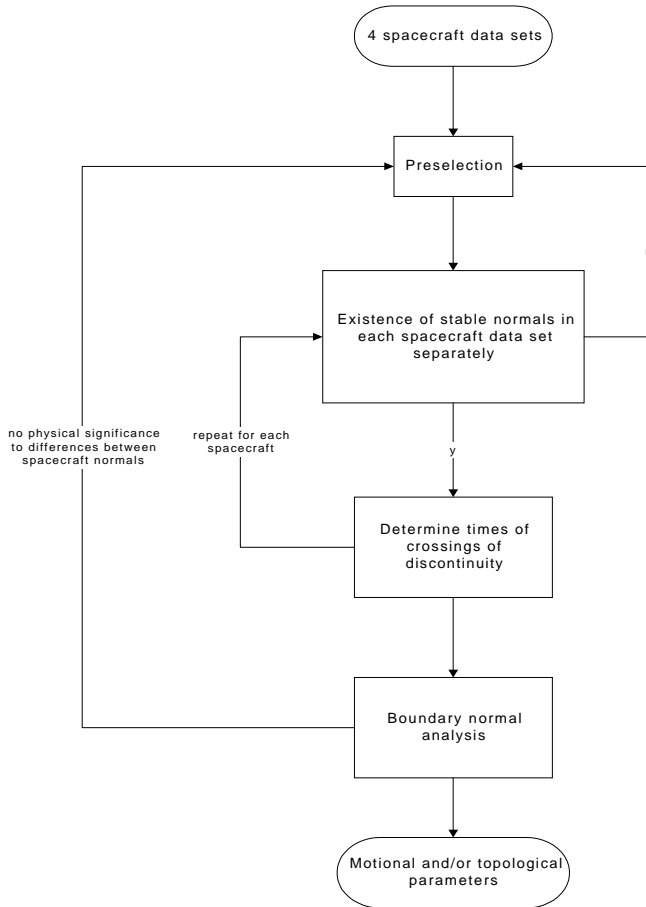


Figure 11.1: Conceptual flow for the boundary analysis.

11.2.2 Method outline

In view of the above assumptions in Section 11.2.1, we have developed a hierarchically-based methodology, consisting of four principal stages. These are an initial data pre-selection phase, a normal stability analysis, determination of crossing times, and finally a combined analysis of boundary normals. The principal components of the analysis procedure are identified in Figure 11.1 and are described below:

1. Data Pre-selection:

This process, represented by the second (from the top) box in the flow chart, serves to identify candidate data intervals which may possess the signature of a discontinuity detected on all spacecraft present. Predominantly, this depends on recognition of the event in the individual spacecraft data and selection is little different when only one spacecraft samples the structure. Clearly, however, the character of the event should show similarities amongst the spacecraft and this can often help distinguish between the signatures in complex encounters, such as the case of multiple

crossings. Since here we concentrate on demonstrations of the technique using simulated events, and most event identification methods are commonly known from single spacecraft studies, we do not discuss these explicitly now. Routines such as variance or spectral analysis, cross-correlations, tests for planarity [Farrugia *et al.*, 1990] and stationarity [Chapman and Dunlop, 1993] are used here.

2. *Normal Stability:*

The third box in Figure 11.1 represents a set of single spacecraft methods which potentially can be used to determine the principal directions (local to each spacecraft) for the boundary structure; the intention being to reveal the boundary normal directions. Since we analyse only magnetic field data and hence current based discontinuities, the stability of the boundary (surface) normals is determined, by initial choice, using minimum variance analysis (MVA) on a set of nested data intervals. Furthermore it is assessed for each spacecraft data set separately. Clearly, this analysis on the individual spacecraft, allows very few initial assumptions on the nature of the crossings to be made. Equivalent inspection of the boundary (for example, computation of tangential discontinuity normals) and analysis of the degree of planar ordering is then contained within a secondary diagnostic analysis. These diagnostics allow the stability of the normals to be traced over repeated computation when the selected interval is varied. Failure to detect a stable normal requires a return to the data pre-selection process. Either further pre-selection or pre-filtering is then carried out, or rejection of that particular interval of data. A wave propagating along a boundary, for example, will affect the result obtained from normal determination, but may often be suppressed or the effect on MVA clarified, by suitable high or low pass filtering [Dunlop *et al.*, 1996].

3. *Crossing Times:*

The fourth box in Figure 11.1 represents the procedure which seeks times at which the boundary, identified in step 2, crossed each spacecraft. In order to identify these times, the centre of the boundary signature for each spacecraft data set has to be located by curve fitting, or otherwise (for example, by inspection). Note that the success of MVA provides a canonical (natural) coordinate system which best represents the boundary in the data and this is taken advantage of in the current procedure. If the boundary is revealed clearly, inspection often provides the most accurate method of identification, if intensive, since it does not require presumptions on the form of the signature in the data. Modern visualisation tools are sufficient to make this user interaction very feasible, and are used in the software tools described in this chapter.

4. *Boundary Properties:*

The bottom two boxes in Figure 11.1 presume that, in principle, the motion and/or topology of the boundary can be analysed. The core, multi-spacecraft technique is represented by this part of the flow chart, with the details of application hidden, since they are described in detail below for each key encounter situation. As mentioned above, this core technique is currently based on an assumption of planarity, but where deviations from planarity can be identified. In general, any interpretation of the spatial structure has to be made in the context of self-consistency between the normal directions and the motion or topology of the boundary. The form of any check on consistency depends upon the implied properties of the boundary (whether

it is planar or non-planar, or has constant or non-constant motion). A planar boundary is consistent with colinear normals (within the implied error from the variance analysis), but non-planarity is not necessarily the only source for significantly non-colinear normals. For instance, the differences between the spacecraft normals may arise from other than simple boundary structure, and then further pre-filtering must be attempted, or there may exist no physical significance to the differences between the spacecraft normals, and then the analysis is ambiguous. The interpretation may also be obscured when the relative timing in step 3 arises from a combined effect of non-constant motion and curvature.

The procedural loop(s) implied by the checks on consistency in steps 3 and 4, assume some interaction between the interpretation of the normal analysis in step 2 (single spacecraft) and the choice of combined analysis in step 4. As stated there, strong, consistent and closely colinear normals imply good planarity, and then allow more detail on the boundary motion to be determined in principle (noise permitting). Non-planar boundaries are implied by significantly non-colinear (stable) normals; but the significance needs to be established (see section 11.4.1). Quantitative analysis in the presence of curvature will be addressed in section 11.5.

If the whole process is successful, the results will yield parameters characterising the boundary. Note, however, that the independent determination of the normals at each spacecraft is what allows the other macroscopic properties (shape, orientation and motion) to be checked during the subsequent analysis, rather than assumed. These two stages are pursued rather independently in the current technique, requiring caveats on their interpretation to be considered (as above). In principle they could be combined self-consistently, but then a problem of weighting each stage arises (as is considered in Section 11.5.6 and is touched on in Chapter 10).

11.3 Pre-Selection Issues

11.3.1 Assumptions

In the issues discussed below, stationarity and planarity are addressed in terms of the particular event properties and as part of the pre-selection procedure, although both of these properties enter into the development of the discontinuity analysis. For example, the basic planar-DA technique described below allows a degree of motional analysis to be made, but strict planarity of the field is not required for multi-spacecraft analysis. On the other hand, stationarity is generally assumed as a basic assumption of the analysis. For instance, non-dispersive structures, at least, are required for planar analysis, and for quantitative determination of parameters some assumptions of convective motion of a quasi-static structure must be made. General assumptions governing methodology have been presented above in Section 11.2.

11.3.2 Boundary Normal Analysis

Since we limit the analysis here to magnetic field data in the technique described below, it is natural to make use of the variance properties of magnetic vector. The behaviour of the magnetic field is constrained by the condition of solenoidality $\nabla \cdot \mathbf{B} = 0$. In terms

of the magnetic field variations, this constraint effectively links the direction of minimum variance to the boundary normal for simple, 1-D structures [Sonnerup and Cahill, 1967]. Chapter 8 discusses the various detailed forms of this method, which essentially differ in their approach to the statistical analysis which is employed to find the minimum variance direction. We briefly recall here the governing equations for MVAB, defined there, in order to compare to the procedure for the planarity test which is introduced in the next section.

The variance (computed over some data interval in time, represented by the sum over n in the equations below) of the component of the magnetic field vector along some direction, \hat{x} , is defined by

$$\sigma_x^2 = \frac{1}{N} \sum_n (\mathbf{B}^n \cdot \hat{x} - \langle \mathbf{B} \rangle \cdot \hat{x})^2, \quad \langle \hat{x} \rangle = 1$$

The minimisation procedure for this quantity is identical to solving the eigenvalue equation for the matrix \mathbf{M} , defined below, where the minimum eigenvalue defines the direction of minimum variance through its corresponding eigenvector. Thus, \mathbf{M} satisfies

$$\mathbf{M}\hat{x} = \lambda\hat{x}, \quad M_{ij} \equiv \frac{1}{N} \sum_n (B_i^n B_j^n - \langle B_i \rangle \langle B_j \rangle)$$

The diagonalisation of this variance matrix, which contains all information on the component variances, generally can always be performed and gives the principal values and axes which characterise the variance ellipsoid. The association of these principal values with the boundary normal of a discontinuity is a physical interpretation of the event being sampled, however. The quality of the final result, therefore, has a physical component as well as a component relating to the nature of the measurements. The former describes how accurate the model of the event is and how the presence of additional real structure may affect the calculation, while the latter is reflected in the statistical variations, which, nevertheless, arise also as a result of the detailed properties of the event. The way in which these contribute to the uncertainty in the determination of boundary normals, and the stability of the result, has therefore received much attention (see Chapter 8).

Although the common choice for normal determination is MVA in magnetic field based analysis, inconsistencies arising from particular properties of the boundary may require alternative methods for estimating the normals. For example, the procedure would respond differently when applied to both tangential and rotational discontinuities, particularly with regard to the planar ordering of the magnetic field. Various alternative methods of determining boundary normals for shocks and other discontinuities, often using multi-instrument data or the presence of additional particular properties, are given in Chapters 8 and 9.

The use of the deHoffmann-Teller frame, possible in all other boundaries except purely tangential discontinuities (for which Sonnerup MVA is particularly suited), forms the basis of other single spacecraft techniques. For shocks other conditions can be employed across the boundary, through the Rankine-Hugoniot relations and coplanarity (see Chapter 10). These are often employed preferentially since analysis through the shock is often confused by waves and other fluctuations either side of the shock which do not relate closely with the normal.

11.3.3 Planarity

In the event of magnetic structure taking the form of a 1-D boundary, which is crossed by a spacecraft, the solenoidality property of the magnetic field is equivalent to

$$\mathbf{B} \cdot \hat{\mathbf{n}} = \text{const}$$

where $\hat{\mathbf{n}}$ is the normal to the boundary. This is the physical starting point for the variance method outlined above (Section 11.3.2). Since we primarily have in mind tangential field structures (discontinuities) to demonstrate the DA technique, even for cases which deviate from planarity (modelled as shown in Section 11.A), it serves clarity to use the tangential form of the above condition,

$$\mathbf{B} \cdot \hat{\mathbf{n}} = 0 \Rightarrow \frac{\mathbf{B} \cdot \hat{\mathbf{n}}}{|\mathbf{B}|} = 0 \quad (11.1)$$

This equation has the strong geometric interpretation that the magnetic field vector always lies on the surface of the plane defined by $\hat{\mathbf{n}}$. A consequence of this is that the cross product of (average) field vectors on either side of the discontinuity surface is parallel to the normal to the surface, and thus this provides a simple way to find such normals (see equation 10.27 on page 263). However this estimate can be inaccurate, particularly in the case of nearly parallel field vectors.

Another method of normal determination which derives from equation 11.1 is as follows. For purely transverse variations ($|\mathbf{B}| = \text{const}$), equation 11.1 has an easy generalisation to that of conical order, when the non-zero constant is re-introduced. This geometric view forms the basis of a rugged test for magnetic order and is discussed by *Farrugia et al.* [1990] and *Dunlop et al.* [1995]. Constraining the field vector to lie in a plane (or the surface of a cone), has the effect that the Euler angles (θ, ϕ) , describing the orientation of the field vector (for *any* reference coordinates) are explicitly related through the above equation. Thus, for planar order

$$\tan \theta = \tan \alpha \sin (\xi - \phi)$$

where (α, ξ) are the polar and azimuth angles defining the orientation of $\hat{\mathbf{n}}$ in the same reference frame and θ is actually the latitude angle of the field vector. Conical order is governed by a slightly more complicated equation [*Farrugia et al.*, 1990].

In the above equation, the Euler angles for the sampled field are actually functions of time, and the equation relating them models the time series measurements. Therefore, they form a set of equations which can be statistically solved (by general least squares, or by singular value decomposition methods) for the angles (α, ξ) , which then define the normal and hence the orientation of the plane. Not only do singular value decomposition methods give the accuracy of the fit, but a plot of the elevation and azimuth of the field direction in some reference frame gives a strongly visual representation [see *Dunlop et al.*, 1995, for applications].

This simple, but powerful, geometric technique is used as a complementary test of event character and is part of the diagnostic test of normal stability employed during the analysis.

11.3.4 Stationarity

A simple test can be devised for the case of multiple spacecraft data which is described in *Chapman and Dunlop [1993]*. For planar, convecting structures, for example, it is clear that

$$B(\mathbf{r}_\beta, t) = B(\mathbf{r}_\alpha, t - \Delta t_{\alpha\beta})$$

where $\Delta t_{\alpha\beta}$ represents the convection time of the structure to move from spacecraft α to spacecraft β . A time Fourier transform of this signal (on the profile seen at each spacecraft) implies that (where ω is frequency).

$$B(\mathbf{r}_\beta, \omega) = B(\mathbf{r}_\alpha, \omega) e^{-i\omega \Delta t_{\alpha\beta}}$$

by using the familiar shift theorem in the Fourier domain. Thus, the ratio of the amplitude spectra between pairs of spacecraft will be approximately equal under these conditions. Further this ratio is also independent of frequency for non-dispersive structures. The phase spectra will differ, directly by $\omega \Delta t_{\alpha\beta}$.

For non-planar, but convecting, structures the amplitude ratio will still be constant, but not unity. The test therefore gives the necessary conditions for the sampled structure to be quasi-static (or non-dispersive) by inspection of the ratios in the signal between spacecraft in the frequency domain. Even if the signal is not fully static overall, there may be ranges in frequency (windows) for which it is, particularly if the time dependence is well separated in frequency. In these circumstances the signal can be filtered to remove the time dependent part, and re-analysed.

The test works best for planar structures, where the response is flat and unity, but can also be applied, approximately, with preferred sampling along principal directions of non-planar structures. The test has been successfully applied to a number of model structures. The effect of instrument and other noise has also been investigated. Normally the effect of instrumental response occurs only over some part of the frequency range. Most events will remain stationary over some range of ' ω ' (a window) where the model most strongly applies.

11.4 Discontinuity Analysis: Motional Properties

11.4.1 Planar-DA Technique

As described in Section 11.3, the planar-DA technique first attempts to fit a planar discontinuity, and then determine its motional properties. Initially, therefore, curvature is treated only as a test for non-planarity. Curvature introduces characteristic and systematic deviations from a planar fit (Section 11.4.5). Only in the case of very pure events can quantitative surface fitting be attempted, as we discuss in Section 11.5, under certain assumptions on the nature of the relative motion.

The core method (which is principally characterised by its use of independently determined normals) is therefore discussed from two viewpoints: firstly planar analysis; and secondly deviations from planarity. Figure 11.2 summarises the technical description of the analysis process. The conceptual flow in Figure 11.1 is not shown, but the form here shows the analysis choices which depend on the result of the individual boundary normal analysis. If the normals all agree (to within the implied uncertainty of the eigenvectors),

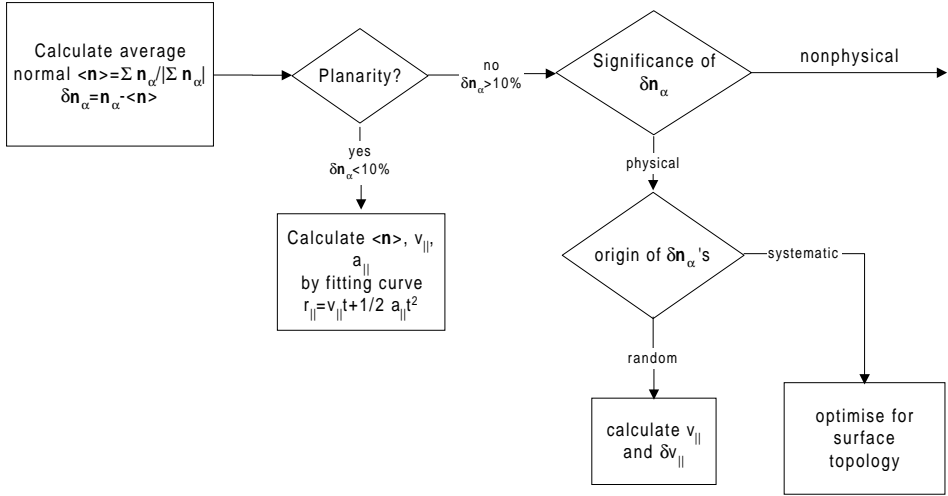


Figure 11.2: Technical flow for the Discontinuity Analyser.

the planar analysis can proceed to calculate the quantities indicated, as described later in Section 11.4.2.

Consider this planar route illustrated in Figure 11.3, where planar is arbitrarily defined as $\delta n_\alpha < 10\%$. The boundary is assumed to cross each spacecraft in the order shown, once, with the projected crossing distances r_\parallel along $\langle \hat{n} \rangle$ being known from spacecraft position data and the computed average normal. The individual boundary normals, obtained at each spacecraft, will provide an estimate of the mean normal, which will be a better estimate for the common boundary normal if the structure is planar. An error weighted mean is currently used for this (but shown as a sum in Figure 11.2, for simplicity). Although we can then estimate the velocity parallel to this normal for each pair of spacecraft individually using the crossing times, we choose to fit a polynomial to the projected spacecraft separation vectors, parallel to the average normal, of the form:

$$r_\parallel = v_\parallel^1 t + \frac{1}{2} a_\parallel t^2$$

Here t represents $t_{\alpha,\beta}$, the relative crossing times, and r_\parallel represents $\Delta \mathbf{r}_{\alpha,\beta} \cdot \langle \hat{n} \rangle$, the projected spatial separation for each spacecraft pair (note that this equation is similar to that quoted in Section 14.5.2). This allows a constant acceleration term only and some initial velocity at the first spacecraft crossing. Note that for planar boundaries, only the components of the motion along the normal have meaning. For non-planar boundaries, the transverse components immediately become significant, and then, of course, \mathbf{a} may not align with \mathbf{v} . There is a subtle problem of sorting crossing times against spacecraft number which is dealt with in the analysis technique but is not discussed here.

Of course, for planar boundaries having constant motion and a minimum of four spacecraft, v_\parallel and the common normal can be obtained directly from the relative crossings and knowledge of the spacecraft configuration, since there then exists a high degree of redundancy (this is the technique applied by *Russell et al.* [1983]). With less than four spacecraft

this analysis is ambiguous, unless the plane of the discontinuity is known in one direction, or unless other conditions along the normal can be employed (for example, in the case of shock jump conditions, covered in Chapter 10). With four spacecraft it is pointed out in Chapter 10 that the extra conditions can form an over-determined set of equations.

It is unlikely that any mission will be flown in the near future with an array of more than four spacecraft, so that this approach will continue to depend on fortuitous conjunctions of independently orbiting spacecraft (as in *Russell's* study). We feel the approach here is better able to distinguish deviations from such assumptions as rigorous planarity and constant velocity, together with the effect of errors and noise. Moreover, as we discuss in Section 11.5, the number of spacecraft required is not restricted and simply limits the information that can be reliably obtained.

The other route in Figure 11.2 (i.e. $\delta n_\alpha > 10\%$) implies possible non-planarity. The existence of stable, distinctly non-parallel normals is, however, not necessarily an indication of structure curvature. This is the question to be addressed in the box labelled “significance of n ” in the flowchart (Figure 11.2). For example, the differences between the normals may arise from effects other than simple boundary structure, such as surface waves. Further pre-filtering may be attempted, or there may exist no physical significance to the differences between the spacecraft normals and the analysis is ambiguous. By “physical” significance we mean the relevance to analysis by way of the discontinuity analyser technique. If, however, the origin of the normal differences is believed to be a curved thin magnetic structure, we can attempt to characterise the surface and its motion more fully. Systematic differences indicate the probable existence of curved field geometry and we can attempt a quantitative analysis of both the surface curvature and motion (see Section 11.5 for a more detailed discussion of curvature in which the planar-DA methodology is applied loosely, still within a discrete analysis).

Conversely, if the normals show differences that are inconsistent with simple curved field geometry but that are, nevertheless, believed to fundamentally derive from curved field geometry, then application of a planar analysis can still be performed, as described in Section 11.4.5, but then only linear fits to the velocity can be attempted. This provides a less quantitative estimate of the velocity (i.e. the estimate represents only a mean velocity during the measurements), and the non-planarity of the surface remains qualitative, since then the presence of acceleration in the motion is not explicitly treated. Any quadratic fit, attempting to obtain an acceleration estimate will be meaningless in this case because of the presence of curvature. Furthermore, curvature (and other effects), unknown at this stage of the analysis, can bias the estimate of velocity and even masquerade as an apparent acceleration.

11.4.2 Basic Application

Figure 11.3 demonstrates the technique for the case of planar analysis. The method has been tested, using suitable magnetic field models for the discontinuity, by flying the spacecraft through the model.

The top graph shows a typical result of flying the spacecraft with constant velocity. There may be a component of \mathbf{v} perpendicular to $\hat{\mathbf{n}}$, as shown, for which no information is revealed by a planar structure. The lower graph shows the result of an additional, constant acceleration. The relative times of the crossings are plotted against the relative, components of \mathbf{r} , parallel to $\hat{\mathbf{n}}$. Note that one point (for spacecraft 1, here) lies at the origin. For

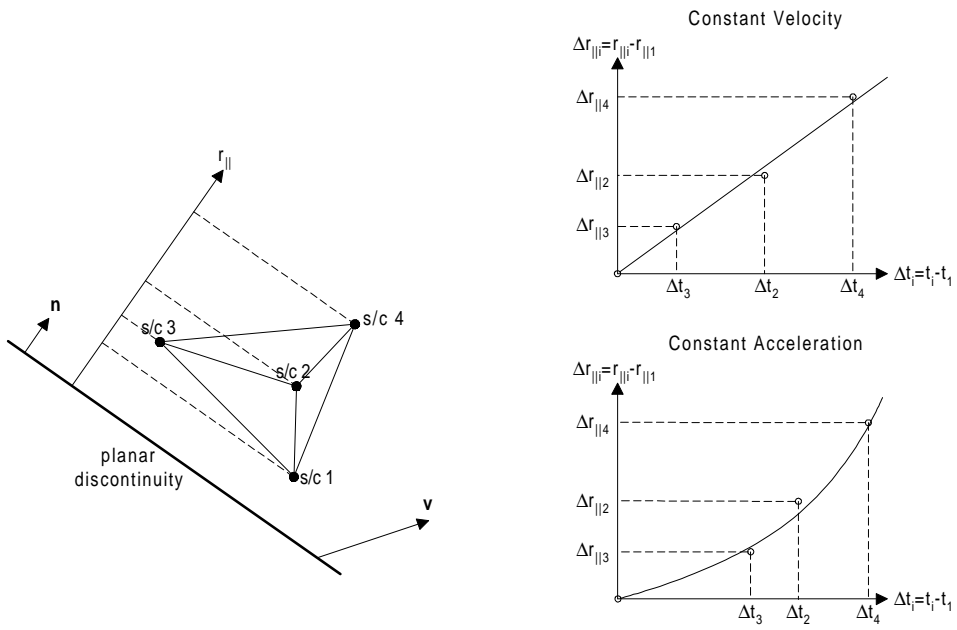


Figure 11.3: The basic discontinuity analyser method, described as the planar-DA in the text.

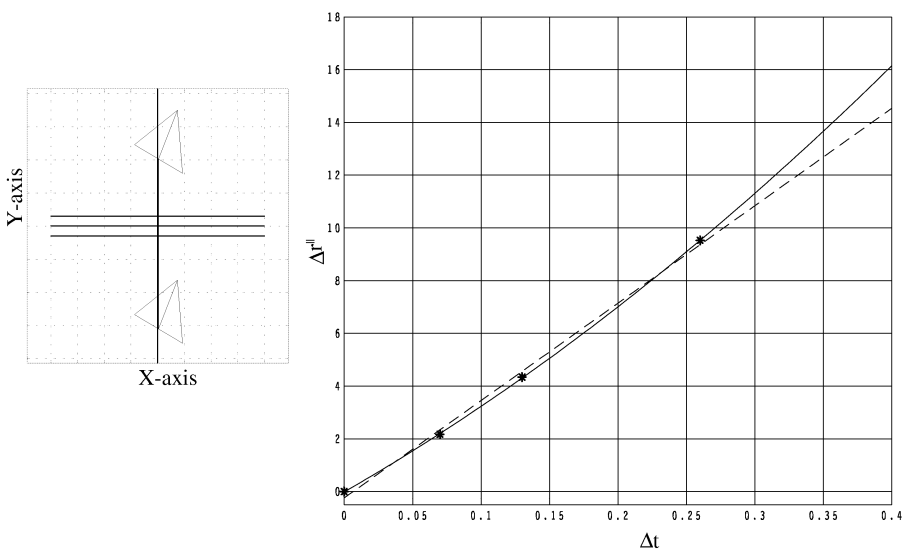


Figure 11.4: Sample run through the planar discontinuity using an accelerated fly-through. The two curves on the right-hand panel are linear and quadratic fits, respectively.

constant acceleration, a quadratic can easily be fitted to the points as shown.

An example output, showing a simulated accelerating trajectory is also given in Figure 11.4. In these tests both the actual trajectory flown and the quadratic fit to the crossing points can be superimposed (only the quadratic and a linear fit are shown in the Figure). Additionally, the effect of noise and other signal can be monitored by redoing the analysis for these cases (see below).

11.4.3 Noise: Error Combinations

The linear fit to the crossing points shown in Figure 11.3 determines the velocity from all independent pairs of spacecraft. Clearly,

$$v_{\parallel\alpha\beta} = \frac{\Delta \mathbf{r}_{\alpha\beta} \cdot \hat{\mathbf{n}}}{\Delta t_{\alpha\beta}}$$

so that the estimate of velocity depends upon errors in: the normal, the spacecraft separation vectors and the crossing times. In fact for application to a phased spacecraft array, timing errors usually remain well below 1% and are relatively unimportant in their contribution to the error in v_{\parallel} , δv_{\parallel} . We can therefore consider that the error in the fit lies predominantly in uncertainty in the values of Δr_{\parallel} . Relative errors in velocity then arise from $\delta \hat{\mathbf{n}}$ and $\delta r/|\Delta \mathbf{r}_{\alpha\beta}|$, where for convenience we have treated the separation error as common for all spacecraft pairs.

The presence of these errors, as a result of noise in the data signature or instrument uncertainty, will increase the scatter of the points about the linear fit depicted in Figure 11.3 and the corresponding uncertainty in \mathbf{v} . Noise will degrade the normal analysis through uncertainties in $\hat{\mathbf{n}}$. Clearly, also, the quality of the estimate of the acceleration term is degraded by the addition of noise. Since the mean normal, used for determining Δr_{\parallel} , is affected to a reduced degree compared to the individual normals, noise levels associated with normal uncertainty of up to 20% only affect the estimate of Δr_{\parallel} (and hence in \mathbf{v}), to below 10%. Of more practical concern are tracking errors, which are typically of order 10% in the projected Δr_{\parallel} (and hence in \mathbf{v}).

An acceleration term introduces a systematic change in the velocity of order $a_{\parallel} \Delta t$. The estimate of acceleration will remain quantitatively significant, therefore, if deviations from a linear fit (corresponding to implied acceleration, as in the quadratic fit in Figure 11.3) are greater than the random uncertainty in Δr_{\parallel} . In simulation tests of noise, which result in scatter of order the effect of the typical acceleration indicated in Figure 11.3 (although then not systematic), the fit often remains significant, although the estimate is quantitatively poor.

11.4.4 Noise: Data Examples

Before discussing the identification of non-planar effects, it is useful to first explore the visual effects of natural noise on the analysis. The visual effects of noise on the analysis can be simply shown in the format of the planar test, described in Section 11.3.3, where the Euler angles (θ, ϕ) of the magnetic field vectors are plotted against each other as a scatter plot. The equation for the model plane is superimposed as a curve in the plot of (θ, ϕ) . Figure 11.5, for example, shows a simulated data signature from the planar tangential

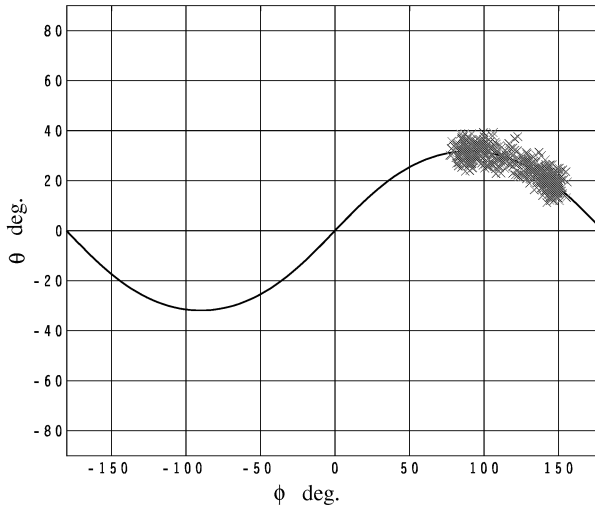


Figure 11.5: Scatter plot for a pure, planar discontinuity with the addition of instrumental noise.

model, with typical (in the sense of the argument in Section 11.4.3), high frequency scatter on the model data for one spacecraft. Note that this scatter is symmetric about the planar curve, drawn here to correspond to the exact model plane. Ultimately, this scatter will feed through as uncertainty in both the normals (represented by the plane here) and timings, affecting the positions of the crossing points, as in Figure 11.4.

The geometric technique is used both as a complementary test of event character (pre-selection) and as a diagnostic test of normal stability. The planar orientation may deviate from the MVA evaluation of the normals in some cases, particularly where real events have additional, complex properties which may confuse underlying boundary structure or may bias the estimated normals. The technique is applied in Figure 11.6, for instance, on a well studied magnetopause crossing; one of a number considered in *Paschmann et al.* [1986]. This example is chosen because it is quoted as presenting difficulties for MVA analysis of the boundary, due to variability in the implied normals. In fact it is often the case, for magnetopause boundary crossings, that MVA gives no better estimate of the normal than simple surface models.

Three intervals (thought to be rotational discontinuities) containing crossings of the magnetopause are chosen for the illustration. These are labelled 1, 2 and 3 at the top of the time series plots in Figure 11.6 and analysis is carried out for suitable intervals, centering on each. Scatter plots of each interval are shown on the right-hand side of Figure 11.6 in terms of the angles (θ, ϕ) of the magnetic field time series. The conical form of the planar equation is fitted for values of (α, ξ) , in the generalised least squares sense, to the scatter plot of the angles (θ, ϕ) for interval ‘1’, shown in the top panel. This plane corresponds closely, although not exactly, to the boundary normal, found via MVA. The other two panels then use this fit as a model curve for comparison to the other intervals. The curves are suggestive of the ordering seen by eye, even when the scatter here is large. Conversely, MVA then fails to produce any stable normals. The geometric technique is therefore useful

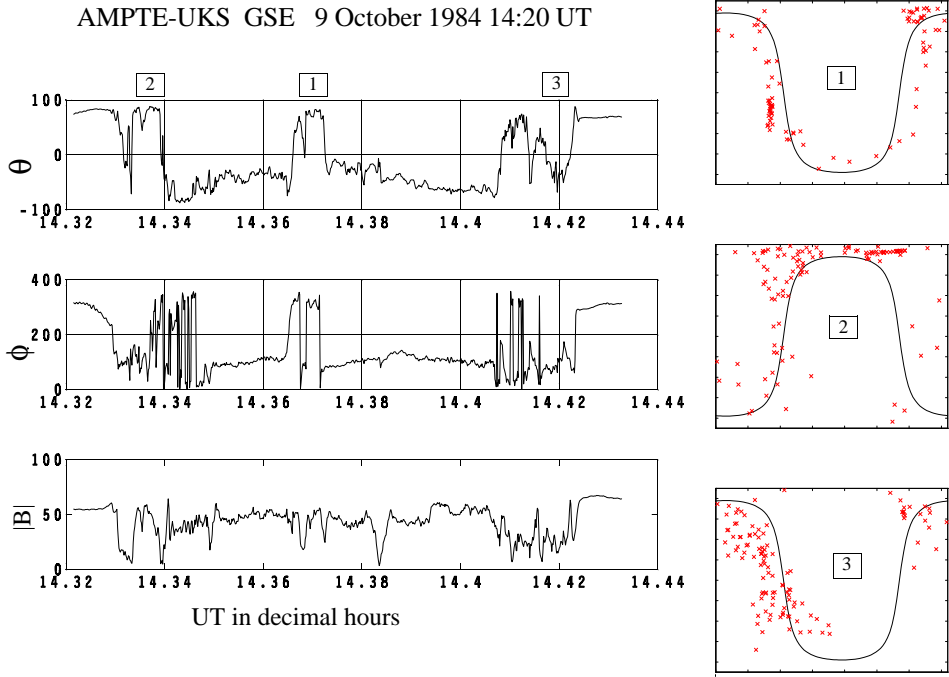


Figure 11.6: Demonstration of the planarity test as described in the text.

as an alternative method to MVA for these cases, where MVA gives confused results. Generally, the technique is applied best in situations of slow boundary crossings, giving a suitable number of data values through the boundary. The diagnostics for the normal determinations use similar (θ, ϕ) plots for visual representation. Clearly, the example here only represents a single spacecraft location.

11.4.5 Identification Of Non-Planar Structures

Consider an array of spacecraft encountering a curved discontinuity. The situation is drawn schematically in Figure 11.7, with the corresponding graph of crossing times in (a) and (b), plotted against the projected r for two relative orientations of the sense of the curvature. The curvature is assumed to be 2-D, and only three spacecraft are shown, for clarity. In this case, the dot-dashed line in the graphs are drawn to correspond to the constant velocity motion of a planar discontinuity. The situation shown, represents a typical result of simulated fly-throughs, using one of the simple 2-D magnetic boundary models (see Section 11.A).

For each model, the detail for the crossings will differ and hence the analysis will differ, but we can use the sketch in Figure 11.7a, which serves to indicate how the crossing times deviate asymmetrically from the planar line, to represent the key geometry. For instance, spacecraft 2 will cross the real boundary at t_2 but crosses an effective planar boundary at t'_2 . This shift $(t'_2 - t_1)$ is shown on the graph, and similarly for t_3 . The sense of this systematic deviation depends only on the relative orientation of the spacecraft configuration with

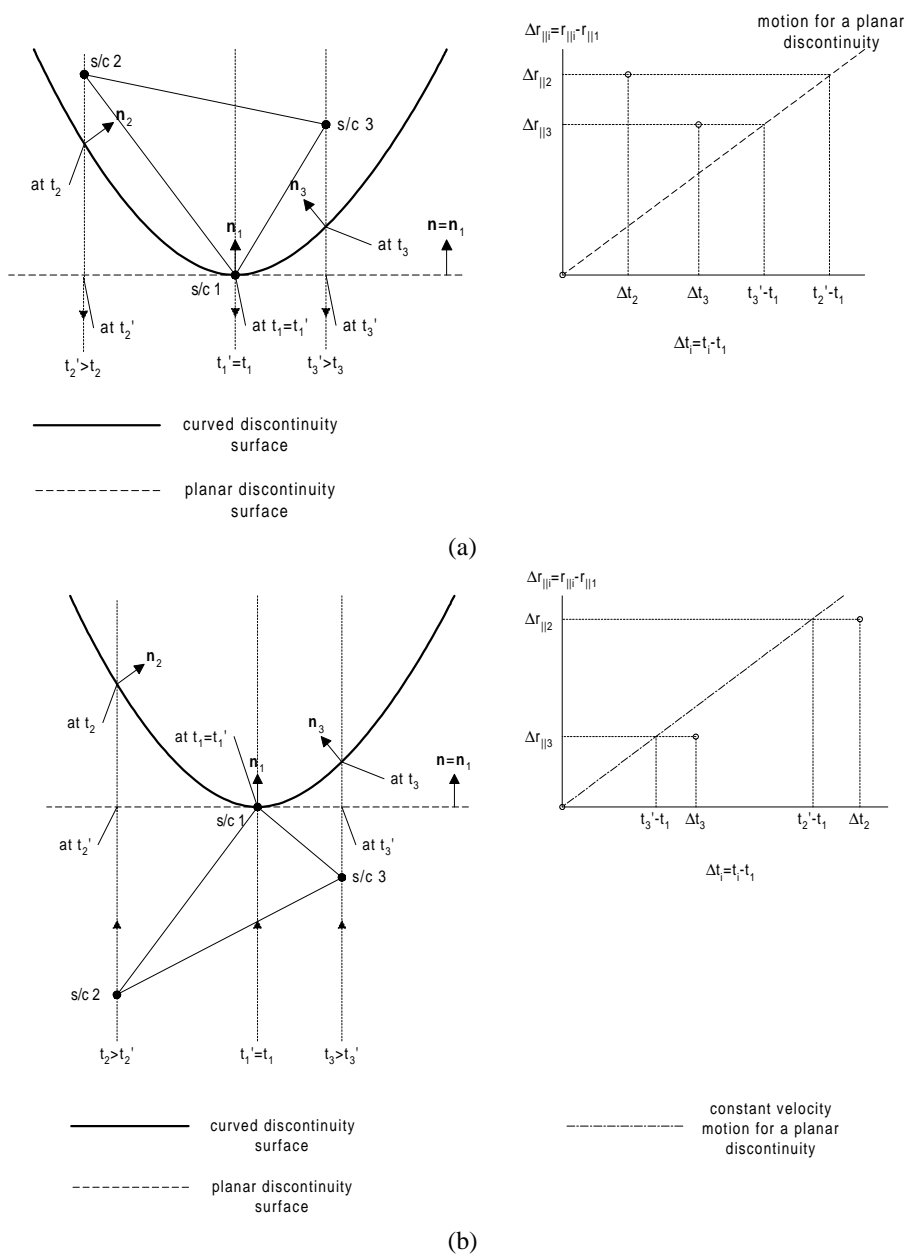


Figure 11.7: The analysis of a non-planar boundary structure.

respect to the curvature and not on the sense of motion (down or up along the line shown in Figure 11.7a), as is apparent by following a similar discussion of Figure 11.7b. For instance, spacecraft 2 again crosses the planar boundary at t'_2 , but now the geometry of the boundary is inverted with respect to the spacecraft configuration and this is reflected in the plot. We shall see later that the effect of noise more easily obscures the effect of acceleration in the presence of curvature; although the sense of the asymmetry may still be apparent (Section 11.5.5).

11.5 Discontinuity Analysis: Combined Motional and Curvature Analysis

Our intent in this chapter has been to deal conceptually with factors which can complicate the analysis of both curved discontinuities and situations with non-constant motion, particularly those factors relating to conflicts in event properties. Consequently, we have attempted to distinguish between conceptual issues and technique description at various points in the discussion. Now, we look at these issues in more detail, discussing required changes in methodology (from the planar-DA) when curvature analysis is introduced; for example, in the combination of the normal determination with the macroscopic analysis. We also look at the extent of self-consistency in each stage of the analysis, an issue which becomes deeper when curvature effects are considered. We first discuss these issues in general terms below, and then detail some specifics in the later sections.

11.5.1 Method Development

Use of multi-spacecraft

The amount of information that can be learned from different sized arrays of spacecraft is an important consideration in the development of the method. Intuitively we would expect that with more spacecraft available, more can be gleaned on the macroscopic properties of the physical event. With only a single spacecraft, for instance, planarity is implicitly assumed in the application of MVA and motional information can only be determined by combining multi-instrument data. With more than one spacecraft the multi-point data can be used and we summarise below the information that can be calculated from spacecraft arrays of increasing size (up to 4, the size of Cluster); firstly, in the case of planar discontinuities and secondly, in the presence of curved boundaries (it may help the visual interpretation to refer back to Figure 11.3).

1. Planar discontinuities

- 2 spacecraft:

Planarity may be checked to within the projection of the spacecraft separation by making use of computed boundary normals, \hat{n} (interpreted from MVA, for example). The boundary velocity, v_{\parallel} , relative to the spacecraft and parallel to \hat{n} may also be directly determined from the time delay, but with assumption of constant velocity (simple convection) and stationarity of the structure.

- *3 spacecraft:*

These give a better check on planarity (two spacecraft pairs are available, providing a mean, $\langle \hat{n} \rangle$). Although similar assumptions must be made as for two spacecraft, a more realistic velocity can be found since the two estimates may give some quantitative indication of acceleration. Again, the motion must be one dimensional (convective).

- *4 spacecraft:*

Here, a good planarity fit and velocity estimate can be obtained. In addition, a possible estimate of acceleration, a_{\parallel} , parallel to $\langle \hat{n} \rangle$, may be made if noise or other additional signal is sufficiently low and normal stability and colinearity is sufficiently high.

2. *Deviations from planarity (non-dispersive)*

Unless planarity is assumed, special, multiple boundary crossings are required with one spacecraft for even qualitative analysis, or the curvature must be small (nearly planar). The determination of quantitative information, in the presence of curved boundaries, generally requires limitations on either motion or curvature and more than one spacecraft. Explicitly, with four spacecraft, the limit could be constant velocity of a convected structure, which would allow some check on the level of curvature to be made, via surface fitting procedures. Alternatively, if some linear acceleration is known to be present, simple curvature properties must be assumed; for instance, a single radius of curvature (see Section 11.5.3). Generally, with less than four spacecraft, the number of unknowns sought, even at lowest order in curvature and acceleration, are too many for direct estimation. Either rigid, constant motion, or planarity must be assumed to proceed. Errors and noise, of course, severely limit the ability to disentangle motional from topological factors and it is possible that only qualitative knowledge can be obtained.

Macroscopic treatment

The basic assumption in the planar-DA, which requires justification, is that the spacecraft separation distances are much in excess of the scale lengths of the physical event. Sampling the characteristic spatial scales in this context allows the spatial content in the time series data to be interpreted individually; as compared to across spacecraft. This interpretation is consistent with the use of single spacecraft methods for boundary normal determination, since it limits the effect of non-planar structure on the local analysis of the individual spacecraft crossings even in the presence of some curvature, i.e. the integration of the solenoidality of the magnetic field remains, primarily, at the single spacecraft level.

Even under this “thin boundary” assumption, the full coordination of data from a number of spacecraft is sensitive to properties which generally take only secondary importance in the single spacecraft analogue. Effectively, the approach, termed *macroscopic* analysis, decouples the determination of macroscopic properties (through purely multipoint analysis) from the boundary normal identification (obtained through purely multichannel analysis, where the coordination in the data is primarily over an interval in each time series). This decoupling is not perfect, however, and is dependent on the event characteristics, particularly in the presence of curvature, which affects the normal determination to some degree, as we shall see.

Actual measurements result from *a priori* unknown encounters with the boundary, which must be checked self-consistently with any model assumptions. Moreover, the usability of the measurements is constrained by data issues such as, the sampling of the discontinuity, compatibility across data sets and data quality. A consequence is that developing the method for the combined use of multi-instrument data (having often lower time resolution and accuracy than that of the magnetic field alone) with multiple spacecraft data sets is not simple to design efficiently and to achieve overall enhancement of the performance of the technique. Special events may arise, of course, from particular sampling conditions or particular physical properties.

11.5.2 Non-Planar Analysis: Constant Motion

As discussed above, the planar-DA strictly applies to thin, nearly planar boundaries, but can provide a qualitative indication that curvature is present. However, by applying the methodology non-rigorously in this case, direct fitting of a surface to the individual normals is in principle possible; the assumption being that the normals remain determined independently from the surface fit at the individual spacecraft. Where planarity is demonstrated, the method gives a direct test of the motion (since the curvature is then known). There is no equivalent test, however, where curvature is unknown, since the individual crossing times will be affected by both curvature and motion. For instance, if a boundary moves with known, constant velocity, the crossings at each spacecraft can be mapped back to the positions at the time of the first crossing allowing the surface fit to proceed. The fit is then only mutually consistent with that velocity, however. Normally, all components of velocity need also to be treated as parameters in the fit.

There is therefore a need to ensure that the fit is achieved self-consistently by careful development of the method. Furthermore, although the normals might be significantly non-colinear, there will be some degree of uncertainty in their orientation. In fact it is not known, *a priori*, what weighting should be given to the errors in the individual normals in order to then apply some optimisation of the surface fit. One reason is that the analysis is further complicated by the fact that the normal determination is itself affected, or biased, at each crossing point, by the direction of motion through the boundary of each spacecraft, particularly if the curvature is high. This bias arises through the need to identify a finite data interval through the boundary (which maps to a distance along the motion) in order to perform the variance analysis. Unless this distance is small compared to the scale of the curvature, the field structure will not wholly represent the boundary orientation local to the crossing. This affects the resulting normal direction computed.

Figure 11.8 schematically emphasises this effect. In the illustration we have chosen a semicircular boundary (thick solid curve), across which the field reverses, and assume for argument that field lines (thin curves with arrow heads indicating field direction) in the vicinity of the boundary may be drawn as semicircles all with a common centre (O). Two spacecraft trajectories (t_1 and t_2) are drawn. The first one, t_1 , passes through the centre point, and normals to successive field lines along t_1 are all colinear, as illustrated in the figure.

The second and more realistic trajectory, t_2 , does not pass through the centre point, and the normals to successive field lines are not colinear. Single spacecraft boundary normal techniques determine a normal which is some average normal of the measured field over the interval taken. Clearly, the longer the interval over which the normal is calculated,

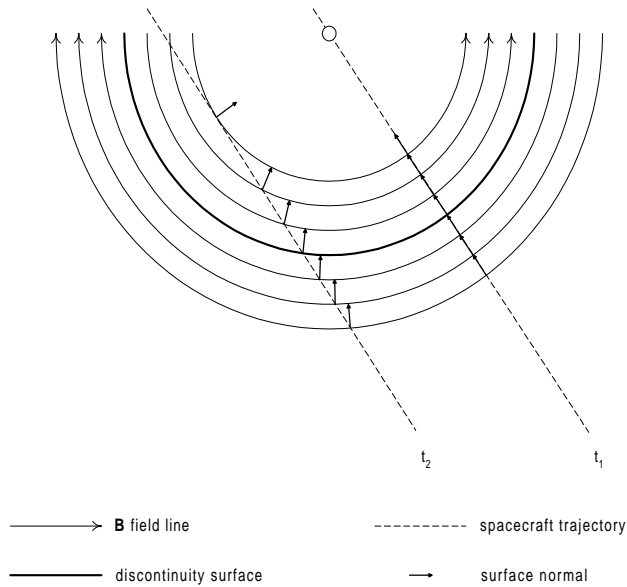


Figure 11.8: Diagrammatic representation of a circular discontinuity, showing the dependence of the field normal on trajectory.

the more it can vary from the real boundary surface normal. Other boundary models can be considered, but the conclusion here still remains the same: i.e. that trajectories not along certain specific (to the model) characteristic lines (i.e. realistic trajectories) will yield normals which vary according to the length of the data interval taken.

Although such a sensitivity of the computed normal to the data interval is a well known problem, its importance here is that any result of surface fitting cannot, therefore, be fully decoupled from the determination of normals, or curvature must not be significant on the scale of the boundary layer. Indeed, the effect of strong curvature can affect the normals quite dramatically and will introduce a dependence on both the time series interval used for analysis and the trajectory orientation through each crossing point. Investigation confirming this has been carried out in conjunction with the planar test on both the models and on a real data event (see Section 11.5.5) and suggests that a test for thin boundary events may be possible.

For completeness, a full inversion of the problem should be done, but this is not easy to set up correctly (see, for example Section 11.5.4). The above comments, together with those in Section 11.2, raise the issue of whether full curvature analysis can be done quantitatively at all. The two points raised for the planar-DA are essentially that independent determination of the boundary normals gives a direct test of planarity and that this then sets the context for the motional analysis. In other words, notwithstanding the above comments (that normal determination in curved geometry is open to possibly unknown error), the boundary normals have to be used in some way to provide information on the local surface orientation. Furthermore, there is no equivalent test for the nature of the motion as there is for establishing planarity. Events with constant velocity must be assumed, unless information is obtained from other instrument data, such as the plasma velocity.

Any actual acceleration will affect the surface fit to at least the same order as the curvature terms. Thus, deviations from assumptions of constant motion may artificially enhance, or mask, the curvature effect (see Section 11.5.4). The constraint to constant motion is used by *Mottez and Chanteur* [1994], who designed a surface fitting procedure based on a geodesic description of the surface and which requires the use of the Frenet curve relations. The analysis works to lowest order (given the motional constraint) and they successfully apply a reduced version to the case of a dual spacecraft magnetopause crossing. They do not explain, however, how, in general, events with constant motion are selected. Section 11.5.3 here also refers to their technique in terms of the extent of curvature analysis which can be attempted.

11.5.3 Non-Planar Analysis: Information Balance to Lowest Order

If the structure is non-planar, the foregoing discussion suggests that, in general, less information can be determined about the motion. In fact, topological and motional parameters compete for their representation in the data and in the nature of the analysis which can be performed. Use of the planar technique for qualitative analysis of a model, for instance, is possible since the planar analysis can be used to order the data so that deviations from planarity can yield unique information on the surface. Direct analysis of the surface topology, however, is limited by the need to ensure self-consistency with the normal analysis; currently a single spacecraft analysis in the manner discussed above. As mentioned, for example, the individual normals can be dependent on spacecraft trajectory, with the dependence being related to the form of the surface. Hence, any surface analysis will affect the interpretation of the normals.

Full curvature analysis, therefore, is generally possible only under conditions of constant motion or very limited curvature. This is indicated in Table 11.1 below, which results from the following equations governing the number of unknowns which are present in each assumed situation. In constructing this table, we assume that the surface normals determined at each spacecraft position are precise. This allows us to establish the lowest order determination of the unknown parameters characterising the surface. Firstly, we consider in detail the situation of a 2-D curved discontinuity. We assume that the surface of this physical event may be modelled by a curve of a single radius of curvature, and hence that the spacecraft array encounters the structure at its “nose”. For the corresponding 3-D case, we also take the view that the analysis attempts to find only a local expansion of the surface to the spacecraft configuration, which sets the orientation of the effective nose of the curved surface with respect to the spacecraft. This slight relaxation of mathematical rigour is, we believe, in the spirit of any realistic analysis using actual measurements.

Curvature with 2-D geometry

Consider a two-dimensional curved discontinuity (thick solid curves in Figure 11.9) moving linearly and without distortion or rotation. The discontinuity, which is indicated by the two thick solid curves in Figure 11.9, is shown at two positions in time. All quantities are assumed to be projected into the plane of the figure. This plane is known *a priori* for sampled events from the orientation of the measured normals, which would have to have a common component (be coplanar) by assumption. Referring to Figure 11.9, suppose that the discontinuity (curve passing through spacecraft 1) is detected at spacecraft 1 at a time

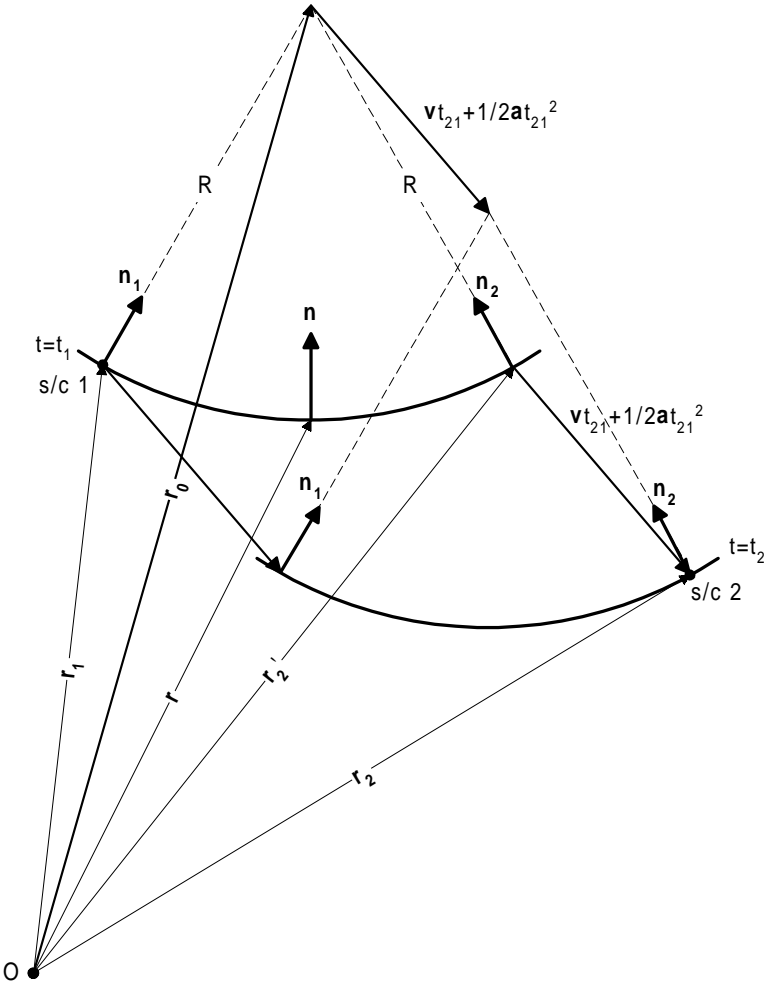


Figure 11.9: Geometrical construction for a curved discontinuity at lowest order curvature as described in the text.

t_1 (then at position \mathbf{r}_1 relative to some origin), and at spacecraft 2 at a time t_2 (then at position \mathbf{r}_2 and indicated by the curve passing through spacecraft 2).

If we assume that the discontinuity had a (relative) velocity \mathbf{v}_1 (with respect to the spacecraft) at time t_1 (assumed to be earlier than t_2) and was accelerating with constant acceleration \mathbf{a} , then, after a time $t_{21} \equiv t_2 - t_1$ when it was detected at spacecraft 2, it would have traversed a displacement $\mathbf{v}t_{21} + \frac{1}{2}\mathbf{a}t_{21}^2$. This motion is indicated in Figure 11.9 by the solid arrows connecting the two positions of the discontinuity at times t_1 and t_2 .

If we now assume that the discontinuity has constant curvature (i.e. has a constant radius of curvature, R), then it is possible to relate each of the discontinuity crossings by the two spacecraft. In principle, this may be done by using the assumed motional properties to locate the position on the surface at the earlier time t_1 (i.e. when the surface encountered

spacecraft 1) where spacecraft 2 detected it, and then using the assumed geometry to relate the two positions.

In particular, for an arbitrary point on the surface (see the upper thick curve passing through spacecraft 1), shown with position \mathbf{r} and normal vector $\hat{\mathbf{n}}$, the position \mathbf{r}_0 of the centre of curvature of the surface may be written as

$$\mathbf{r}_0 = \mathbf{r} + R\hat{\mathbf{n}} \quad (11.2)$$

Thus, considering the encounter with spacecraft 1, we may write

$$\mathbf{r}_0 = \mathbf{r}_1 + R\hat{\mathbf{n}}_1 \quad (11.3)$$

Since we have assumed a linear motion of the surface, we can relate the encounter with spacecraft 2 at time t_2 back to the position of the surface at time t_1 and the encounter with spacecraft 1. In particular, since the discontinuity is displaced by $\mathbf{v}t_{21} + \frac{1}{2}\mathbf{a}t_{21}^2$ during the period between its encounters with the two spacecraft, the position \mathbf{r}'_2 of the point on the surface detected by spacecraft 2 at the earlier time t_1 may be expressed as follows (see Figure 11.9):

$$\mathbf{r}'_2 = \mathbf{r}_2 - \mathbf{v}_1 t_{21} - \frac{1}{2}\mathbf{a}t_{21}^2 \quad (11.4)$$

Furthermore, the normal $\hat{\mathbf{n}}_2$ measured at spacecraft 2 at position \mathbf{r}_2 is parallel to the normal at position \mathbf{r}'_2 on the surface at the earlier time t_1 . Hence we may also write

$$\mathbf{r}_0 = \mathbf{r}'_2 + R\hat{\mathbf{n}}_2, \quad (11.5)$$

and combining equations 11.4 and 11.5 we derive

$$\mathbf{r}_0 = \mathbf{r}_2 + R\hat{\mathbf{n}}_2 - \mathbf{v}_1 t_{21} - \frac{1}{2}\mathbf{a}t_{21}^2 \quad (11.6)$$

Therefore, using equation 11.3 we derive a “motional” expression relating the (measured) displacements ($\mathbf{r}_{1,2}$), detection times ($t_{1,2}$) and normals ($\hat{\mathbf{n}}_{1,2}$) to the model parameters ($R, \mathbf{v}_1, \mathbf{a}$), as follows

$$\mathbf{r}_1 - \mathbf{r}_2 = R(\hat{\mathbf{n}}_2 - \hat{\mathbf{n}}_1) - \mathbf{v}_1 t_{21} - \frac{1}{2}\mathbf{a}t_{21}^2 \quad (11.7)$$

or

$$\mathbf{r}_{12} = R\hat{\mathbf{n}}_{21} - \mathbf{v}_1 t_{21} - \frac{1}{2}\mathbf{a}t_{21}^2 \quad (11.8)$$

where

$$\left. \begin{aligned} \mathbf{r}_{12} &\equiv \mathbf{r}_1 - \mathbf{r}_2 \\ \mathbf{n}_{21} &\equiv \hat{\mathbf{n}}_2 - \hat{\mathbf{n}}_1 \\ t_{21} &\equiv t_2 - t_1 \end{aligned} \right\} \quad (11.9)$$

If we have more than two spacecraft detecting the discontinuity, we may construct motional analogues to equations 11.8 for each spacecraft pair. We note, however, that by taking one reference spacecraft, e.g. spacecraft 1, to correspond to the earliest detection of the discontinuity by the spacecraft array, we have only one initial velocity \mathbf{v}_1 as a

model parameter to determine. Thus, we may write the set of equations for an array of 4 spacecraft as

$$\left. \begin{aligned} \mathbf{r}_{12} &= R\mathbf{n}_{21} - \mathbf{v}_1 t_{21} - \frac{1}{2} \mathbf{a} t_{21}^2 \\ \mathbf{r}_{13} &= R\mathbf{n}_{31} - \mathbf{v}_1 t_{31} - \frac{1}{2} \mathbf{a} t_{31}^2 \\ \mathbf{r}_{14} &= R\mathbf{n}_{41} - \mathbf{v}_1 t_{41} - \frac{1}{2} \mathbf{a} t_{41}^2 \end{aligned} \right\} \quad (11.10)$$

where definitions analogous to equations 11.9 apply.

For a two-dimensional model, as considered above, we can consider 2 component vectors in the plane of the normals. Thus, since the unknowns ($R, \mathbf{v}_1, \mathbf{a}$) amount to 5 scalar quantities altogether, we require 4 spacecraft (giving 6 equations) to determine them. If constant velocity can be assumed, only 3 spacecraft are required. Any fewer spacecraft will provide insufficient data for this type of model.

Curvature with 3-D geometry

Three-dimensional discontinuity surfaces present a more complicated problem. In particular, there are more geometrical parameters than in the case of two dimensions. Nevertheless, we outline here the basic (general) formalism which we only discuss by analogy to the above equations. The discussion below is similar to that presented by *Mottez and Chanteur* [1994], except that we relax the constant velocity assumption to one of constant acceleration.

We can assume a simple, general three-dimensional surface, characterised by two principal curvatures (with corresponding radii of curvature R_{p1} and R_{p2}). Thus, to describe the surface we require knowledge of these two scalar quantities, as well as their orientation in space. This latter factor is reduced to a single scalar quantity, namely an angle in the tangent plane to the surface (defined by the measured normal) at a reference point; taken as spacecraft 1 in our case. The motion of the surface is characterised by the velocity \mathbf{v}_1 and acceleration \mathbf{a} (i.e. 6 scalar quantities). Thus, 9 scalar quantities are required to describe the motion and geometry of this surface.

Consider, however, what information can be obtained from the surface by using an array of spacecraft. Building on the ideas discussed above for two-dimensional surfaces, but working with pairs of spacecraft in the three-dimensional case, we can derive motional equations, analogous to equations 11.10, which connect the local radius of curvature R between the spacecraft to the velocity \mathbf{v}_1 and acceleration \mathbf{a} of the surface as before.

Note that the local radius of curvature depends on the orientation of the section containing the two spacecraft and passing through the surface. Thus, we should label the local radius of curvature according to the two spacecraft between which it is measured (i.e. $R_{1\alpha}$; $\alpha = 2, 3, \dots$). In this way, each spacecraft pair, one of which being a reference spacecraft (e.g. spacecraft 1), contributes 3 equations. The unknown quantities are \mathbf{v}_1 and \mathbf{a} , as well as the local $R_{1\alpha}$.

This is not the whole story, however. Two (separated) points on a three-dimensional surface may be connected by a curve on the surface, and the shortest such curve is a geodesic. The important property of a geodesic curve is that the normal vector to it at any point coincides with the normal to the surface at the same point, and this normal is a quantity which we can determine from the magnetometer data using variance analysis,

for example. Thus, when attempting to parameterise the surface by measurements of location at a few points on the surface, we must construct normal sections, which contain the geodesic curves. These curves connect the points on the surface, through the local curvature between spacecraft.

The corresponding geodesic curves, however, will not, in general, be the curves with principal curvatures R_{p1}^{-1} and R_{p2}^{-1} of the surface, but will have different curvatures. Nevertheless, Euler's formula [Kreyszig, 1959] gives the curvature of a normal section in terms of the principal curvatures and the relative orientation of the normal section and the principal section with principal curvature R_{p1}^{-1} . The orientation of the normal section, relative to some coordinate system, is given by the direction of the tangent vector, which may be calculated using a discrete form of the Serret-Frenet curve formulas (as done by Mottez and Chanteur [1994]).

This leaves the orientation (θ_0 , say, relative to the same coordinate system) of the principal sections as unknown at this stage. Hence, each pair of spacecraft contributes 1 Euler equation with 4 unknowns: $R_{1\alpha}$, R_{p1} , R_{p2} and θ_0 . Thus, with 4 spacecraft (i.e. 3 pairs), we may construct 9 (scalar) motional equations in addition to 3 Euler equations, with a total of 12 unknown (scalar) quantities (v_1 , a , R_{p1} , R_{p2} , θ_0 , R_{12} , R_{13} , and R_{14}).

11.5.4 Consequences for Curvature Analysis

Within the framework of the above discussion, Table 11.1 below summarises the results. This table shows the (lowest order) quantitative information which is, in principle, obtainable from magnetic field analysis alone. To construct the table, it has been assumed that the normals are given and are not modified by the analysis (not always true). It is also assumed that the discontinuity is non-dispersive and convecting. The plus signs refer to situations where the number of equations, defined, as above, in terms of the motion through the structure, is more than the number of identifiable unknowns. The negative signs refer to the reverse situation where there are less equations than unknowns. The equals sign indicates situations where there are the same number of unknowns and equations (critically constrained).

For the cases which are under-constrained, so that not all parameters can be determined, qualitative indications may still be obtained, such as testing the existence of curvature, or acceleration, as discussed in the previous sections. Section 11.5.2, however, has discussed the problems associated with assuming the motion to have a constant velocity for the purpose of pursuing a curvature analysis. Clearly, from the table, this appears to allow lowest order determination of curvature to be made in more circumstances (up to 3-D or with less spacecraft). Analysis of the above equations in Section 11.5.3, however, reveals that the acceleration term in these equations enters at least at the same order as the curvature term. Thus, it is never known whether the curvature analysis is fundamentally in error (i.e. an artifact) as a result of an acceleration in the motion which had not been taken into account.

The converse situation occurs if planarity is assumed, anticipating a planar-DA analysis. The presence of actual curvature affects the motional fit at the same order as the acceleration term, and would infer an artificial acceleration (see, for example, the graphs in Figure 11.7). With a minimum of four spacecraft and to lowest order in curvature, as defined here, the motion can be determined without assumptions up to a constant acceleration term, given that the normals can be estimated by some independent method. In this

Table 11.1: Indication of the ability of the analysis to determine the unknowns in different physical situations and to lowest order in curvature.

Dimension	Number of spacecraft:	1 s/c	2 s/c	3 s/c	4 s/c
1-D	No acceleration	—	=	+	+
	Acceleration	—	—	=	+
2-D	No acceleration	—	—	+	+
	Acceleration	—	—	—	+
3-D	No acceleration	—	—	=	+
	Acceleration	—	—	—	=

case, therefore, the result of lowest order determination of the unknowns will have some uncertainties, but is in principle self-consistent: a more complex surface or motion is just represented approximately by the lowest order terms, not anomalously.

The use of other instrument data, of course, can potentially add information on structure, such as independent determination of velocity in the case of electric field measurements. This must be carefully assessed, however, since with the data restrictions (on resolution and quality), which are introduced through the incorporation of other data sets, it will require a different analysis methodology for their consistent integration. Except on a case by case basis, therefore, additional measurements are most easily added as independent information on the event, complementing the planarity or stationarity tests, for example, rather than as an extended, integrated multi-spacecraft method. Furthermore, the availability of more than 4 spacecraft can also potentially add information. However, again the methodology then adopted to fully integrate this information requires careful reconsideration.

Clearly, from the above discussion, if the normal directions show significant and systematic differences, then quantitative analysis of curvature can be attempted. If only a lowest order surface fit is attempted (for example, a locally spherical boundary is assumed), then, in principle, the motion can be determined up to a constant acceleration term, given that the normals can be estimated by some independent method. In practice, however, noise or other properties, present in real events precludes such a critical analysis mainly through their introduction of uncertainties in \hat{n} . Furthermore, there may be resulting bias from the determination of single spacecraft (MVA) normals in curved geometry (as demonstrated in the study below). Whilst other methods can be adopted for estimating the boundary normals their application and the associated errors depend on the nature of the event analysed. Inspection of the above equations, for example, in terms of an error analysis, indicates a weighting effect of curvature on any errors in \hat{n} . The effect may be more significant if the surface lies at an unfavourable orientation to the spacecraft configuration. It is clear from the plots in Figure 11.7, that, for actual encounters in the presence of curvature, where the relative orientations are not known, the effect of noise (or biasing, when significant) will more easily obscure the effect of acceleration; although the sense of the asymmetry may still be apparent. For instance, the study below shows that the asymmetry in the deviations shown in the trends in Figure 11.7 is also represented using the planar test on the individual data sets for each spacecraft.

⇒11.1

11.5.5 Data Analysis: Natural Noise

The problems discussed above in determining structure normals in curved geometries have fundamental impact on the methodology suitable for a “curved-DA” technique since the normals can no longer be assumed to be given. Here, we demonstrate the effect with data from a real event recorded by the AMPTE/UKS and IRM spacecraft pair. It is not surprising, of course, that MVA provides normal estimates in curved geometry which deviated from the geometric normals indicated by the models. The systematic effect of quantitative biasing on the normals, shown schematically in Figure 11.8, is important for their use within the methodology outlined above; i.e. as a single spacecraft analysis for \hat{n} , feeding into the equations. The significance of this systematic error with respect to the overall deviations of the normals between spacecraft depends upon the spacecraft trajectories through the structure, given the data sampling achieved. This sampling depends upon the combination of the minimum nesting interval applied and the data resolution with respect to the characteristic spatial scale of the structure. In practice, the sampling achieved depends upon the thin boundary properties of an event. The trajectories then set the spacecraft separation scale with respect to the characteristic scale of the model.

Different events will be sampled in different ways *in situ* so that for some the systematic error will be comparatively large and for others comparatively small. In the latter case, the methodology may be used, for such selected events, to estimate both lowest order curvature and acceleration terms (by fitting all terms in the above equations to the set of normals). As indicated by the flow in Figure 11.2, the rejection of qualitatively unsuitable events may be done if the normal analysis shows unstable results (with respect to nesting or shifted intervals) in the presence of significant deviations between the spacecraft normals. It is possible, however, that inspection of MVA normals, in comparison to the planar test at each spacecraft, will indicate suitable events for measured data; so providing a possible event selection stage in the procedure. The presence of stable MVA results, which are also consistent with the planar (geometric) test results, individually, is likely to be a good indicator that biasing is relatively small for the structure and that the individual normals can be interpreted geometrically.

Figure 11.10 shows a real event measured by the combined AMPTE/IRM (Figure 11.10a) and UKS (Figure 11.10b) spacecraft during a solar wind Barium release (see papers in *Nature*, **320**, 700–726, 1986). Both spacecraft were thought to have sampled an intense plasma cloud boundary at about the time indicated by the solid vertical line on the upper panels of each plot. Because of cloud expansion following the release, this boundary should be strongly curved between the two spacecraft and this appears to be confirmed by MVAB boundary normals obtained from each data set. The sense of the curvature implied by the normals, which were obtained using a standard run, is described in Dunlop *et al.* [1996] (equivalent curves are plotted in Figure 11.10), and is consistent with an expanding cloud where the effective tilt between the normal directions is $\sim 30\text{--}40^\circ$.

The curves in the lower panels of Figure 11.10 show individual fits of the planarity test at each spacecraft together with the corresponding planes defined by the individual MVA-normals. In this example, where there is significant curvature, the MVA and planar test give very different planes (on the scale of the deviations between spacecraft), even though the structure is strongly tangential. This, of course, is not unexpected in curved geometry unless the thin boundary description holds to a high degree. The planar test attempts the best fit, always, and hence finds a mean, geometrically, planar orientation for the scatter

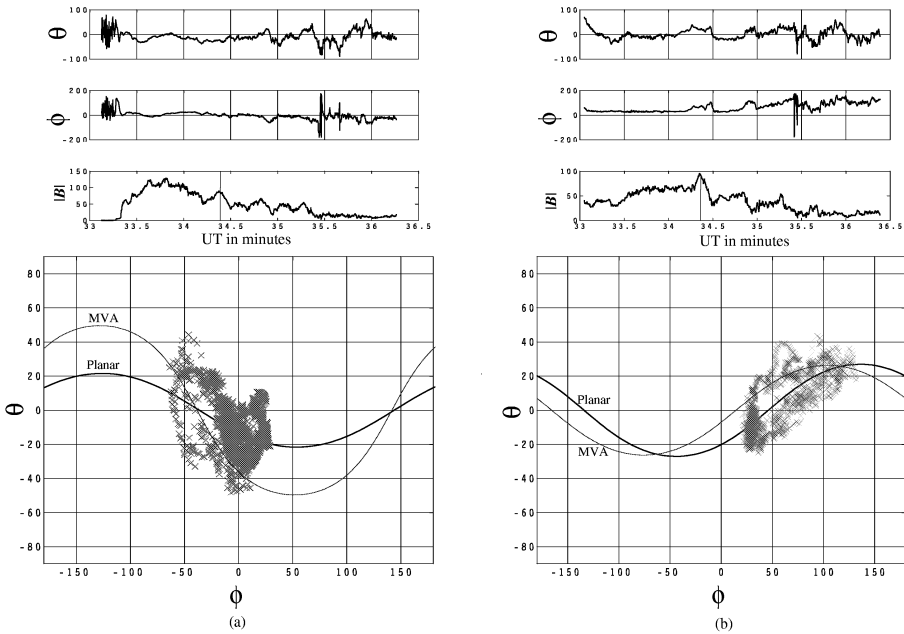


Figure 11.10: Data example of a curved discontinuity from the AMPTE/UKS and IRM solar wind Barium release, as discussed in the text.

plot (locally to each spacecraft). Either side of the boundary, the direction of the magnetic field will deviate systematically from the local plane inferred by the normals.

We therefore expect that both MVA (as well as the planar test) will return a biased normal estimate, which will be sensitive to the actual sampling path of the spacecraft, as discussed above. In fact, the scatter of data points shows a clear asymmetry about the MVA curves which is in the opposite sense for each spacecraft. This is consistent with the sense of the curvature implied by the boundary normals (following the arguments given in Section 11.4.5 for interpretation of non-planar structure). It shows that the boundary structure is effective over the analysis intervals and confirms the biasing of the normals is significant. Using the planarity test in conjunction with the MVA analysis obviously assists the geometric interpretation of the structure. Clearly, however, this event would be unsuitable for the discrete analysis in the manner discussed above.

11.5.6 Self-Consistent Curvature and Normal Analysis

Self-consistent methods are required which quantitatively disentangle the relative contributions arising from both non-constant motion and curvature; avoiding the need to make specific assumptions on properties. As was alluded to in Section 11.4, however, information sought on one of these aspects may only be done at the expense of details concerning the other if both are represented in the data. Even to lowest order, the information balance in the idealised model is still critical for four spacecraft. In the case of the presence of both these effects, the establishment of self-consistency will involve some interrelation

between the macroscopic and boundary normal analyses (checked iteratively). We can ask, for instance, how the determination of the individual boundary normals is relaxed to take account of a surface fit. This is an issue of weighting the normal values obtained during the optimisation of the surface fit.

A more general technique, than the motional analysis described in detail above, is in the process of being designed which combines both the multipoint and multichannel data in such an interrelated manner, within a slightly extended minimum variance scheme. Although this approach can potentially provide more rigorous estimates in the presence of both limited curvature and non-constant motion, it is not clear how fully a relaxation of the thin boundary assumption can be handled by it (although this is the hope). This will be the subject of future work. Effectively, the use of the variance definitions assumes that the data is characterised by the thin boundary regime, i.e. it assumes that data intervals used for the analysis (the sum over i , below) cover what can be considered to be a locally planar boundary.

The problem is one of inversion to find the coefficients describing the local surface shape, statistically optimised to remain consistent with the individual boundary normals. This can be defined in terms of the minimisation of a variance in the form of

$$\sigma^2 = \frac{1}{N} \sum_{i,\alpha} (\mathbf{B}_i^\alpha \cdot \hat{\mathbf{n}}^\alpha - \langle \mathbf{B}_i^\alpha \rangle \cdot \hat{\mathbf{n}}^\alpha)^2, \quad \alpha = 1, \dots, s.$$

under the particular constraints arising from the specification of the surface function, $U(x_\alpha, y_\alpha, z_\alpha) = C$. The index i runs over the data and the spacecraft are labelled by α . The \mathbf{B}_i^α are considered as (time shifted) time series vectors such that $\hat{\mathbf{n}}$ runs over a set of intervals centred on the crossings at each spacecraft. Singular value decomposition seems the best mathematical technique for solving the inversion. The definition of this variance is based on the construction of a statistical series of scalar products $\mathbf{B}_i^\alpha \cdot \hat{\mathbf{n}}^\alpha$, where \mathbf{B}_i^α is a single magnetometer measurement on spacecraft α and $\hat{\mathbf{n}}^\alpha$ is the (in principle, unknown) normal to the physical structure at that spacecraft.

In this form, the solenoidality of the magnetic field is integrated as part of the optimisation at the (nearly) single spacecraft level (rather than at a fully multi-spacecraft level). Therefore, the normals $\hat{\mathbf{n}}^\alpha$ are not related to each other through Gauss's law, but through the (unknown) form of the surface $U = C$. This surface is considered to contain the motional parameters in order to effect the time shifting of the \mathbf{B}_i^α 's. Thus, the above variance exhibits two characteristics. Firstly, it involves the variance at each spacecraft (i.e. the sum over i). Secondly, it entails the sum over the different spacecraft normals (i.e. the sum over α); the two characteristics being related through the form of the surface $U = C$. The minimisation of this variance determines the surface and motional parameters as well as, potentially, the individual boundary normals. Consider the case where the structure is actually planar. Then the form of the variance above decouples into the individual single spacecraft variances and a minimisation automatically should produce the MVA normals. In general, how closely the solution coincides with an independent boundary normal analysis at the individual spacecraft will depend upon how consistent, statistically, the normals are with the macroscopic nature of the sampled data (intervals selected, balance of motion and curvature effects).

Minimisation using the normals $\hat{\mathbf{n}}^\alpha$ implicitly would require a full inversion procedure, and represents a truly multi-spacecraft and multichannel methodology. In that case,

however, there is only an implicit weighting of each term: with too few spacecraft, or too complex definitions of the motion or topology, the information will be conflicting and optimisation with respect to all coefficients will not be possible. An explicit use of the normals, however, can be viewed as being halfway between the full-blown, combined methodology and the discrete methodology we have discussed in the preceding sections (where curvature actually exists). Such an explicit minimisation combines the normals \hat{n}^α , determined separately with each spacecraft data set, in the scheme of the variance defined above. The normals, considered to be independently calculated (and therefore part of the input data), must be weighted according to the confidence we have in them in the light of the uncertainties that they may possess as discussed in section 11.5.2 (random or systematic). In the case of a planar structure, for instance, in which systematic uncertainties would not be present, this reduces to the planar-DA method.

Clearly, one can also limit the procedure by primarily solving for the motion (given an assumption of fixed curvature) or for the curvature (given an assumption of fixed motion). Under the assumptions of lowest order curvature and constant motion, the minimisation is much simplified. The addition of more spacecraft, of other measurements (relating \mathbf{v} , $\hat{\mathbf{n}}$, \mathbf{B} , etc.), or other physical relations (for example, coplanarity, Rankine-Hugoniot relations for shocks), can also allow some statistical optimisation (by general least squares or other means) of motional or curvature parameters. For planar structure, for example, equations analogous to the type presented in equation (11.10) can be directly solved for \mathbf{v} and $\hat{\mathbf{n}}$ (as in the *Russell et al. [1983]* study). The issue of weighting is then not relevant, of course, since the normals are not independently calculated. Hence the inversion is a little artificial in these cases and only mutually consistent with the macroscopic assumptions, not necessarily the data characteristics.

11.6 Conclusions

This chapter has introduced a methodology for a magnetic field-based, discontinuity analyser technique which is suitable for multi-spacecraft analysis. It has discussed conceptually the issues arising for the study of different structures, particularly separating the approach required for (nearly) planar structures and significantly curved discontinuities. We also describe the application of the method to simulated and actual data (from 2 or more spacecraft). Such application of the discontinuity analyser (DA) technique requires more detailed examples in the future to fully test its behaviour. The success of the analysis depends on the degree of stationarity, and the planarity properties, of events and tests for these form part of the analysis. These pre-selection methods have been explicitly described here and have been implied in the description of procedure. The restrictions introduced by limitations in the number of spacecraft has also been addressed.

The planar technique, together with the associated motional analysis, has been demonstrated here, using example output from simulated data. The method relies on the independent determination of boundary normals to give a direct demonstration of planarity, which endows this method with an internal consistency between the (planar) model and the data characteristics. We have explored the natural progression from motional to topological parameters which can be extracted by considering, in degrees, the introduction of curvature. Thus, for planar structures one can potentially determine v_{\parallel} , a_{\parallel} , and a unique $\hat{\mathbf{n}}$ for the boundary. Uncertainties lead to a payoff between the determination of v_{\parallel} and a_{\parallel} , how-

ever, and events need to be particularly clean (in terms of noise, or the absence of complex properties) to analyse deviations from planarity. If the structure is non-planar, less information can be determined about the motion. The nature of the obtained parameters of course also depend upon the degree of stationarity found.

In general, therefore, topological and motional parameters compete for their representation in the data and typically we must attempt a lowest order determination of them, or know, or make assumptions, about their presence for particular events. For instance, two main issues regarding curvature analysis (surface fitting) are firstly, that the nature of the motion must be assumed (say, that of constant velocity) in order to attempt the analysis, and secondly, that boundary normals must be determined independently to give information on the surface geometry, where single spacecraft methods (MVA) for this determination are sensitive to the presence of curvature in the magnetic field. Data examples have been discussed to investigate this sensitivity as well as the general effect of natural noise, present in real (measured) events. A possible selection test, suitable for thin boundaries (for which the effect on the normals is minimised), has been proposed. It is apparent, then, that internal consistency of the method is weakened with the adaption to curved situations, except within the thin boundary data regime. We explored briefly the problem of constructing a self-consistent procedure for the curved-DA method.

We have had in mind an assumption that “multi-” actually means up to four spacecraft, this being the minimum required for adequate, three dimensional sampling. Although, in principle, the method can be generalised to any number of spacecraft, the method proposed here may not be the best which can be devised. Multi-instrument data, of course, may also extend the method. The use of other measured quantities than the magnetic field can add information on the properties of the structure sampled if used independently in the analysis. Its combined use, in a multi-spacecraft sense, must be treated with caution, however, because of the disparity in data resolution and quality of common quantities (see, for example, Chapters 8 and 9, for caveats on use of multi-instrument data in single spacecraft analysis, and Chapters 6 and 17, for a discussion of plasma moment determination and use).

In summary, if a structure is non-planar, it is difficult, if not impossible, to characterise its topology and motion fully. Direct analysis of surface topology is limited by the need to ensure internal consistency with the normal analysis, i.e. any surface analysis will affect the interpretation of the normals. Thus, curvature effects cannot be unambiguously determined in most situations. It is never known whether the curvature analysis is fundamentally in error as a result of an acceleration in the motion which has not been taken into account. Furthermore, for real events, determination of non-planar surface forms will depend upon the separation of noise and quality of each spacecraft normal from the parameters sought, as well as on the degree of stationarity found. For these reasons, we also, briefly, introduced the concept of a combined self-consistent methodology, which involves defining and solving a full inversion problem. The crux of this approach lies in the weighting of the various elements during the minimisation process, which endows the technique with its self-consistency. This will provide a fruitful area of development for the future.

Appendix

11.A Model Examples

A number of simple models have been employed to simulate a pure boundary structure (to which wave, noise and instrumental effects can be added). Some of these, and their properties, have been presented in *Dunlop et al. [1996]* and are therefore not reproduced here. The tangential field models are briefly defined below as key examples used in the tests of the technique summarised in Section 11.4.

11.A.1 Planar Model

The simplest model used to generate the data for a planar tangential discontinuity is defined as follows. In model cartesian coordinates B_z is zero, while B_x and B_y depend only on the coordinate z , thereby ensuring a divergence-free field. In the particular model chosen the magnetic field maintains a constant magnitude while rotating in the xy plane from parallel to the x axis for $z \ll 0$ to an angle θ_0 to the x axis for $z \gg 0$, the major change taking place between $z \pm d$, the thickness of the discontinuity.

Mathematically this is expressed as

$$\mathbf{B} = B (\cos \theta(z), \sin \theta(z), 0)$$

where B is the magnitude of the field and $\theta(z)$ rotates the field through an angle θ_0 across the discontinuity, thus:

$$\theta(z) = \frac{\theta_0}{2} \{1 + \tanh(z/d)\}$$

Figure 11.11a is a schematic diagram of the model.

11.A.2 Non-Planar Models

We shall define two basic models for a pure 2-D discontinuity, in both of which the strength of the field decreases to zero on approaching the boundary from one side, before flipping direction and increasing its strength on the other side. Mathematically, they are defined as follows:

$$\mathbf{b} = \nabla \varphi \times \hat{\mathbf{z}} ; \quad \varphi \equiv B_0 \log \left[2 \cosh \left(\frac{g(x, y)}{\sigma} \right) \right]$$

where 2σ is the boundary thickness and $g(x, y) = \text{const}$ represents a set of curves as described below.

There are two interesting classes of this type of model which have been used in the simulation tests of the planar-DA technique (see Section 11.4), namely: point-similar and curve-similar.

1. Point-similar

The essential features of this model are depicted in Figure 11.11b, where, for illustration, we have taken a circular geometry, with

$$g(x, y) \equiv R - \sqrt{x^2 + (y - R)^2} ; \quad y < R$$

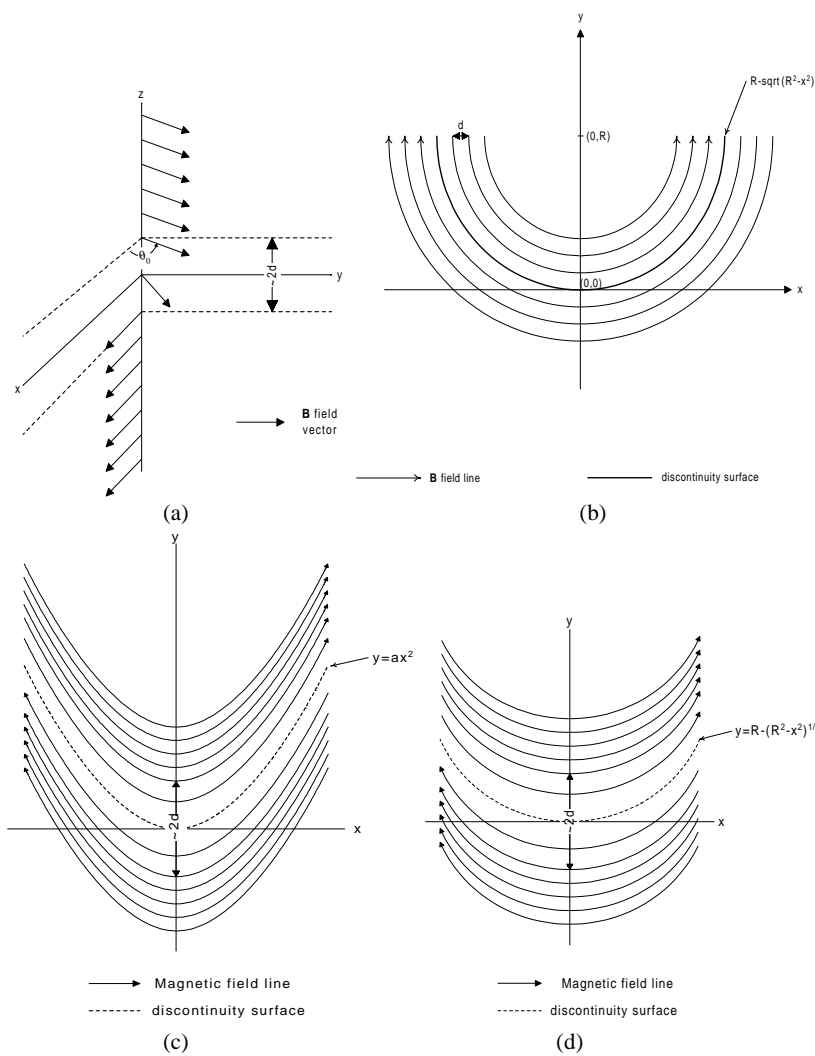


Figure 11.11: Schematics of the form of \mathbf{B} for the example models discussed in the text.

and R is the radius of curvature of the surface of the discontinuity in the xy plane. The point $(0, R)$ is its centre in this plane (the surface is of infinite extent in the z -direction).

2. Curve-similar

For this class

$$g(x, y) \equiv y - f(x)$$

where $y = f(x)$ represents the surface of the discontinuity in the xy plane. To illustrate this, consider two forms of surface.

- (a) A parabolic discontinuity, for which $f(x) \equiv ax^2$; $a = 1/2R$, with R being the radius of curvature of the parabola at its nose ($x = 0$). Figure 11.11c shows the characteristics of this model.
- (b) The circular discontinuity, for which $f(x) \equiv R - \sqrt{R^2 - x^2}$, where R is defined as before, depicted in Figure 11.11d.

The principal difference between the models lies in the position of the field lines (and hence the current distributions) on either side of the discontinuity surface, relative to the boundary itself. In the first case, point-similar, the field lines follow the shape of the surface at a constant distance perpendicular to the closest point on the surface. Thus, for a circular discontinuity, Figure 11.11b, the field lines are semicircles with the same centre point $(0, R)$, but with different radii as compared to the surface. The corresponding case of a parabolic geometry, for the point-similar class, is a little more subtle. On the other hand, the curve-similar, models have field lines which are a constant y -distance from the surface. For example, consider the case of a semicircular surface for this class, as in Figure 11.11d. It is seen that the field lines are semicircles of the same radius, but different centre points (displaced along the y axis) as compared to the surface. A similar geometry is apparent in the parabolic case, Figure 11.11c.

Curve similar models are easier to construct analytically (and computationally), but possess some undesirable features. These are that for a general surface, $y = f(x)$, the current density can increase monotonically away from the nose at $x = 0$, whereas for the point-similar model the current is naturally constant along the surface of the discontinuity. Furthermore, for the circular cases, the curve similar model does not describe the field outside of the region, $-R < x < R$, while the point-similar model does.

Finally, the point-similar model allows the field lines to evolve from straight ($y = -\infty$) to finite curvature geometry ($y > -\infty$), whereas this is not the case for curve-similar models. One drawback of the circular, point-similar model is that the current density can approach a delta function at the centre point $(0, R)$. The point similar model is used in the simulations described later in Section 11.4, being cautious not to approach the centre point.

Bibliography

- Bendat, J. S. and Piersol, A. G., *Random Data*, Wiley, New York, 1986.
 Chapman, S. C. and Dunlop, M. W., Some consequences of the shift theorem for multi-spacecraft measurements, *Geophys. Res. Lett.*, **20**, 2023–2026, 1993.

- Dunlop, M. W., Southwood, D. J., Glassmeier, K.-H., and Neubauer, F. M., Analysis of multipoint magnetometer data, *Adv. Space Res.*, **8**, (9)273–(9)277, 1988.
- Dunlop, M. W., Woodward, T. I., and Farrugia, C. J., Minimum variance analysis: Cluster themes, in *Proceedings of Cluster Workshops, Braunschweig, 28–30 Sep. 1994, Toulouse, 16–17 Nov. 1994*, ESA SP-371, pp. 33–42, European Space Agency, Paris, France, 1995.
- Dunlop, M. W., Woodward, T. I., Motschmann, U., Southwood, D. J., and Balogh, A., Analysis of non-planar structures with multipoint measurements, *Adv. Space Res.*, **18**, (8)309–(8)314, 1996.
- Farrugia, C. J., Dunlop, M. W., Geurts, F., Balogh, A., Southwood, D. J., Bryant, D. A., Neugebauer, M., and Etemadi, A., An interplanetary magnetic structure orientated at a large ($\sim 80^\circ$) angle to the Parker spiral, *Geophys. Res. Lett.*, **17**, 1025–1028, 1990.
- Kreyszig, E., *Differential Geometry*, Oxford University Press, Oxford, 1959.
- Mottez, F. and Chanteur, G., Surface crossing by a group of satellites: a theoretical study, *J. Geophys. Res.*, **99**, 13 499–13 507, 1994.
- Paschmann, G., Papamastorakis, I., Baumjohann, W., Sckopke, N., Carlson, C. W., Sonnerup, B. U. Ö., and Lühr, H., The magnetopause for large magnetic shear: AMPTE/IRM observations, *J. Geophys. Res.*, **91**, 11 099, 1986.
- Russell, C. T., Mellott, M. M., Smith, E. J., and King, J. H., Multiple spacecraft observations of interplanetary shocks: Four spacecraft determination of shock normals, *J. Geophys. Res.*, **88**, 4739–4748, 1983.
- Sonnerup, B. U. Ö. and Cahill, Jr., L. J., Magnetopause structure and attitude from Explorer 12 observations, *J. Geophys. Res.*, **72**, 171, 1967.

— 12 —

Spatial Gradients and the Volumetric Tensor

CHRISTOPHER C. HARVEY

*Observatoire de Paris-Meudon
Meudon, France*

12.1 Introduction

All physically observable field parameters, such as particle populations, electric and magnetic fields, vary in both space and time, and the understanding of the physical processes within the medium requires knowledge of both the temporal and the spatial variations. The ISEE-1 and -2 mission, launched in 1977, was the first attempt to separate systematically temporal and spatial gradients; two satellites allow the determination of the component of the spatial gradient in the direction of their separation vector.

The determination of all three components of a spatial gradient requires at least four spacecraft. These spacecraft define a polyhedron in space. Clearly the success with which any spatial gradient can be measured depends upon the size and shape of the polyhedron, and for the Cluster mission much thought has been given to the optimum geometry to meet specific scientific objectives.

In this chapter we examine from first principles the least squares determination of the spatial gradient using data acquired simultaneously from four or more spacecraft. It is found that the gradient is always expressed in terms of the inverse of a symmetric tensor formed from relative positions of the spacecraft. It is shown that this same tensor describes certain basic geometrical properties of the polyhedron defined by the spacecraft: its characteristic size (mean square thickness) in three mutually orthogonal directions, and the orientation of these directions in space. Conversely, these six geometrical parameters, three characteristic dimensions and three angles, define completely the symmetric tensor; therefore they contain the totality of the geometrical information required to determine the spatial gradient by the least squares method. These results are in agreement with what one intuitively expects: that the quality of the polyhedron for the determination of spatial gradients will involve its size, its anisotropy, and the orientation in space of that anisotropy.

This geometric tensor is shown to be closely related to the inertia tensor. In the special case of four spacecraft, the product of the three characteristic dimensions is exactly three times the volume of the tetrahedron; for this reason it is called the “volumetric” tensor.

The importance of the volumetric tensor for describing the geometry of a polyhedron was first noted by J. Schøenmækers of ESOC Flight Dynamics Division [private communication], but its fundamental importance lies in the key role it plays in the determination of spatial gradients, for which purpose it must be inverted. The magnitude and the direction of the smallest characteristic dimension define how well the spatial gradient can be determined or, indeed, whether it can be determined at all. The values of the other two

characteristic dimensions of the tetrahedron, and their directions, are required to estimate the precision of the gradient determination in both magnitude and direction.

Whatever the number of spacecraft used, each spacecraft will follow its own individual trajectory independently of the others. Therefore the size and shape of the polyhedron will vary continuously along the orbit. This ensures that most of the time the separation geometry will not be ideal. The volumetric tensor, through its eigenvalues and eigenvectors, provides a description of the geometry of the polyhedron which is intuitively simple and yet adequate to evaluate the geometric worth of the data for studying any particular scientific objective. This is explained further in Chapter 13. Of course, the final evaluation of the quality of the gradient determination is a combination of both orbital geometry and physical data quality considerations, as described in Chapter 17. And its overall scientific utility also depends upon the nature of the physical problem being investigated, as discussed in Chapter 13.

12.1.1 Timing

Multi-spacecraft observations exhibit a symmetry between space and time: the same physical observables are measured not only at different point in space, but also at different instants in time. Methods which may be used for resampling, so as to bring data from different instruments onto a common timeline, have been discussed in Chapter 2.

In all cases, it is necessary to use the spacecraft position at the time the physical measurement is made. In the case of a boundary, its motion is then obtained with respect to the orbital motion of the spacecraft.

For the determination of spatial gradients (Section 12.3), it is assumed that all physical measurements, including the spacecraft position coordinates, are brought to a common timeline using the methods described in Chapter 2. It is also assumed that the data have been smoothed (see Chapter 2) to eliminate all short period fluctuations associated with wave disturbances (see Chapters 3 and 4), so that only large-scale variations remain.

If the data sampling interval becomes too large, the resampling becomes convoluted with the spatial gradient which we are trying to measure. And if it becomes comparable with the tetrahedron transit time (the time required for the spacecraft to travel a distance comparable with the size of the tetrahedron), then the whole concept of a multi-spacecraft mission (to measure spatial gradients) fails.

12.1.2 A Simple Boundary Crossing

Let us assume that a discontinuity lies in a plane defined by the direction of its normal \hat{n} , and that this plane is moving in the direction \hat{n} with velocity V . In the simplest case (the one considered in Section 10.4.3) the discontinuity can be identified unambiguously on all four spacecraft. Let it be observed at time t_α by the spacecraft α , $1 \leq \alpha \leq 4$, which is located at position \mathbf{r}_α . These times t_α are assumed to be well determined, with no possibility of inconsistency. It is then easy to determine both \hat{n} and V . During the time $t_\alpha - t_4$ the plane of the discontinuity moves along the normal direction a distance $V(t_\alpha - t_4)$ which is equal to the projection of the separation distance $\mathbf{r}_\alpha - \mathbf{r}_4$ onto \hat{n} ,

$$(\mathbf{r}_\alpha - \mathbf{r}_4) \cdot \hat{n} = V(t_\alpha - t_4)$$

where spacecraft 4 has arbitrarily been taken as the reference.

Introducing the vector

$$\mathbf{m} = \frac{\hat{\mathbf{n}}}{V} \quad (12.1)$$

this may be written

$$\mathbf{D} \mathbf{m} = \mathbf{T} \quad (12.2)$$

where \mathbf{D} is the 3×3 matrix (not a tensor) defined by

$$\mathbf{D} = (\mathbf{r}_1 - \mathbf{r}_4, \mathbf{r}_2 - \mathbf{r}_4, \mathbf{r}_3 - \mathbf{r}_4)$$

and \mathbf{T} is the linear array (not a vector)

$$\mathbf{T} = \begin{pmatrix} t_1 - t_4 \\ t_2 - t_4 \\ t_3 - t_4 \end{pmatrix}$$

This set of equations (which are equivalent to equation 10.20 in Section 10.4.3) is solved by finding the inverse matrix \mathbf{D}^{-1} such that

$$\mathbf{D}^{-1} \mathbf{D} = \mathbf{I} = \text{the unit operator } (= \delta_{jk}) ;$$

hence \mathbf{m} is found,

$$\mathbf{m} = \mathbf{D}^{-1} \mathbf{T} \quad (12.3)$$

Note that a necessary condition for a solution (i.e., for the matrix \mathbf{D}^{-1} to exist) is that $|\mathbf{D}| \neq 0$. This condition is satisfied if, and only if, the four spacecraft are not coplanar.

Although equation 12.3 yields a solution, this solution is unsatisfactory.

1. The method is totally incapable of handling relative time differences determined independently between each of the six different pairs of spacecraft; these time differences may be subject to experimental errors and therefore, in the mathematical sense, they will be mutually inconsistent.
2. The method cannot be generalised to more than four spacecraft.
3. There is a lack of symmetry in the use of the satellite positions \mathbf{r}_α and the timing information t_α , despite the fact that the four satellites are identical. Of course, this consideration is purely aesthetic.

In the following section we present a treatment of the problem which overcomes these criticisms. It is a homogeneous least squares method applicable to four or more spacecraft, and which can be applied to either small or large-scale structures. *For small-scale structures it yields the normal and the speed of motion of a plane spatial structure; and for large-scale structures it yields a linear approximation to the spatial gradient.* The same results are obtained explicitly by a symmetric interpolation method described in Section 14.2; this method, which uses barycentric coordinates, has also been extended to higher order interpolation.

In the rest of this chapter, the dyadic vector and tensor notation used in most of this book is replaced by the summation convention whenever this simplifies the mathematics.

That is, a tensor \mathbf{X} is written in terms of its representation X_{ij} in some arbitrary orthogonal coordinate system; and when in any expression a term contains a product of two subscripted array elements with the same Latin index, then that term is understood to be summed over the values 1, 2 and 3 of the index. Summations over Greek indices, used to indicate the spacecraft, are written explicitly.

12.2 Orientation and Motion of a Plane Discontinuity

Consider a cluster of N spacecraft, where $N \geq 4$. Since they are to be treated symmetrically, it is convenient to identify their mean position, which we call the “mesocentre”; this coincides with the centre of mass if they have identical mass. Let \mathbf{r}_α be the position of spacecraft α relative to this mesocentre. Then,

$$\sum_{\alpha=1}^N \mathbf{r}_\alpha = \mathbf{0} \quad (12.4)$$

The relative positions \mathbf{r}_α , and the position the mesocentre, are all calculable from equation 12.4 and the known orbital positions of the N spacecraft.

12.2.1 A Symmetrical Treatment of the Simplest Case

Consider the example of Section 12.1.2, but with observations from N spacecraft, the time of observation on spacecraft α being t_α . We determine the “best” values of $\hat{\mathbf{n}}$ and V by minimising the expression

$$S = \sum_{\alpha=1}^N [\hat{\mathbf{n}} \cdot \mathbf{r}_\alpha - V(t_\alpha - t_0)]^2$$

where t_0 is some origin of time. Using equation 12.1, this equation may be written

$$S = \sum_{\alpha=1}^N [\mathbf{m} \cdot \mathbf{r}_\alpha - (t_\alpha - t_0)]^2 = \sum_{\alpha=1}^N [m_k r_{\alpha k} - (t_\alpha - t_0)]^2$$

Here the only undetermined quantities are t_0 and the three components m_k of the vector \mathbf{m} . Therefore we minimise S by putting $\partial S / \partial t_0 = 0$ and $\partial S / \partial m_k = 0$, to obtain

$$\sum_{\alpha=1}^N [m_j r_{\alpha j} - (t_\alpha - t_0)] = 0 \quad , \quad \sum_{\alpha=1}^N [m_j r_{\alpha j} - (t_\alpha - t_0)] r_{\alpha k} = 0$$

The first of these equations may be written, using equation 12.4,

$$t_0 = \frac{1}{N} \sum_{\alpha=1}^N [t_\alpha - m_j r_{\alpha j}] = \frac{1}{N} \sum_{\alpha=1}^N t_\alpha$$

t_0 is simply the time at which the mesocentre of the polyhedron crosses the discontinuity.

The second equation may be expressed

$$m_j R_{jk} = \frac{1}{N} \sum_{\alpha=1}^N t_{\alpha} r_{\alpha k} \quad (12.5)$$

where

$$R_{jk} = \frac{1}{N} \sum_{\alpha=1}^N r_{\alpha j} r_{\alpha k} \quad (12.6)$$

This tensor plays an essential role in the determination of spatial gradients. For reasons explained in Section 12.4.3 where its geometrical properties are studied in the special case of a tetrahedron, \mathbf{R} is called the “volumetric tensor”.

Provided that

$$|\mathbf{R}| \neq 0 \quad (12.7)$$

the inverse of the volumetric tensor exists and satisfies

$$R_{jk} R_{k\ell}^{-1} = \delta_{j\ell} \quad (12.8)$$

where $\delta_{j\ell}$ is the Kronecker delta. Then equation 12.5 may be solved, to give

$$m_{\ell} = \frac{1}{N} \left(\sum_{\alpha=1}^N t_{\alpha} r_{\alpha k} \right) R_{k\ell}^{-1} \quad (12.9)$$

A necessary and sufficient condition for equation 12.7 to be satisfied is that the N spacecraft are not coplanar. In reality, this precise mathematical condition must be replaced by a more physical condition which compares the tensor \mathbf{R} with the standard error in the relative spacecraft positions as determined from the orbit analysis, and the standard error of the experimental measurements themselves.

12.2.2 When Timing is only Relative

In general, geophysical events cannot be identified and unambiguously time-stamped on each of the spacecraft. The best that can be done is determine time delays of the observations made on one spacecraft with respect to those made on each of the other spacecraft, for example, by cross-correlation of the different data streams. With N spacecraft, this yields $N(N-1)/2$ time delays

$$t_{\alpha\beta} = t_{\alpha} - t_{\beta}$$

with $1 \leq \alpha \leq N$ and $1 \leq \beta < \alpha$.

To determine the boundary normal direction $\hat{\mathbf{n}}$ and the relative velocity V , we minimise the function

$$S = \sum_{\alpha=1}^N \sum_{\beta=1}^N [\hat{\mathbf{n}} \cdot (\mathbf{r}_{\alpha} - \mathbf{r}_{\beta}) - V t_{\alpha\beta}]^2$$

(noting that $t_{\beta\alpha} = -t_{\alpha\beta}$). Once again, we introduce \mathbf{m} defined by equation 12.1, to obtain

$$S = \sum_{\alpha=1}^N \sum_{\beta=1}^N [m_{\ell} (r_{\alpha\ell} - r_{\beta\ell}) - t_{\alpha\beta}]^2 \quad (12.10)$$

Putting $\partial S / \partial m_k = 0$, we obtain

$$\sum_{\alpha=1}^N \sum_{\beta=1}^N [m_\ell (r_{\alpha\ell} - r_{\beta\ell}) - t_{\alpha\beta}] [r_{\alpha k} - r_{\beta k}] = 0$$

which is

$$m_\ell \sum_{\alpha=1}^N \sum_{\beta=1}^N (r_{\alpha\ell} - r_{\beta\ell})(r_{\alpha k} - r_{\beta k}) = \sum_{\alpha=1}^N \sum_{\beta=1}^N t_{\alpha\beta} (r_{\alpha k} - r_{\beta k}) \quad (12.11)$$

The left-hand side of this equation may be expressed

$$m_\ell \sum_{\alpha=1}^N \sum_{\beta=1}^N (r_{\alpha\ell} r_{\alpha k} + r_{\beta\ell} r_{\beta k}) - m_\ell \sum_{\alpha=1}^N \sum_{\beta=1}^N (r_{\alpha\ell} r_{\beta k} + r_{\beta\ell} r_{\alpha k}) = 2 N^2 m_j R_{jk} \quad (12.12)$$

where equation 12.6 has been used to express the first two terms, and equation 12.4 to show that the last terms are zero. Thus equation 12.11 may be expressed in terms of the volumetric tensor,

$$2 N^2 m_j R_{jk} = \sum_{\alpha=1}^N \sum_{\beta=1}^N t_{\alpha\beta} (r_{\alpha k} - r_{\beta k})$$

which is analogous to equation 12.5. As before, provided that $|\mathbf{R}| \neq 0$, the inverse \mathbf{R}^{-1} exists and we obtain

$$m_\ell = \frac{1}{2 N^2} \left[\sum_{\alpha=1}^N \sum_{\beta=1}^N t_{\alpha\beta} (r_{\alpha k} - r_{\beta k}) \right] R_{k\ell}^{-1}$$

which may also be written

$$m_\ell = \frac{1}{N^2} \left[\sum_{\alpha \neq \beta} t_{\alpha\beta} (r_{\alpha k} - r_{\beta k}) \right] R_{k\ell}^{-1} \quad (12.13)$$

Here and elsewhere in this paper the symbol $\sum_{\alpha \neq \beta}$ indicates summation over the $N(N - 1)/2$ (and not $N(N - 1)$) terms with $\alpha \neq \beta$, all of which have all been used to determine the three components m_ℓ .

Caveat

The $N(N - 1)/2$ time delays $t_{\alpha\beta}$ may not all be independent. In other words, the condition

$$t_{\alpha\beta} + t_{\beta\gamma} + t_{\gamma\alpha} = 0 \quad (12.14)$$

may or may not be satisfied exactly for $\alpha \neq \beta \neq \gamma \neq \alpha$. This depends upon how the delays are determined. There are many methods, of which the two employed most frequently are probably:

- determine the delay corresponding to the maximum value of the cross-correlation function, and

- place two waveform plots, one over the other, and slide them around until “it looks about right”.

It is not clear whether equation 12.14 is satisfied when the first method is used between instruments which do not have identical transfer functions and internal noise characteristics, and which do not sample at precisely the same rate; this latter disparity requires that at least one of the data sets be resampled before cross-correlation is possible (see Chapter 2). But equation 12.14 is certainly not *exactly* true for the second method. And there are probably other ways to determine relative time delays, and it is impossible to predict whether they will satisfy equation 12.14.

If, however, condition 12.14 is satisfied, the $N(N - 1)/2$ equations are redundant. If it is satisfied for all values of α , β and γ there are only $N - 1$ independent time offsets $t_{\alpha\beta}$. In particular, for four spacecraft, the sum 12.10 of the residuals is zero, and the least squares method is formally identical to that of Section 12.1.2.

12.3 Spatial Gradients

The cluster of $N \geq 4$ spacecraft can be used to determine the large-scale local spatial gradient, as described below. It is assumed that all measurements are made simultaneously, if necessary by resampling using the methods described in Chapter 2. This includes, of course, the determination of the spacecraft positions. High frequency fluctuations will also have been removed from the data.

Let x be some parameter whose gradient \mathbf{k}

$$k_\ell = \frac{\partial x}{\partial r_\ell} \quad (12.15)$$

we wish to determine from the values x_α measured simultaneously on each of the spacecraft α , $1 \leq \alpha \leq N$. Our criterion for optimising the value of \mathbf{k} is that

$$S = \sum_{\alpha=1}^N \sum_{\beta=1}^N [\mathbf{k} \cdot (\mathbf{r}_\alpha - \mathbf{r}_\beta) - (x_\alpha - x_\beta)]^2$$

be minimum. This expression is identical to equation 12.10, but with x_α replacing t_α and \mathbf{k} replacing \mathbf{m} . Therefore the least squares value of the gradient is obtained from an expression analogous to equation 12.13, namely

$$k_\ell = \frac{1}{N^2} \left[\sum_{\alpha \neq \beta} (x_\alpha - x_\beta)(r_{\alpha\ell} - r_{\beta\ell}) \right] R_{k\ell}^{-1} \quad (12.16)$$

where the symbol $\sum_{\alpha \neq \beta}$ indicates summation over all $N(N - 1)/2$ independent terms with $\alpha \neq \beta$.

12.3.1 The Gradient of a Vector

The gradient $x_{n\ell} = \partial x_n / \partial r_\ell$ of a vector quantity x_n may be determined minimising the function

$$S = \sum_{n=1}^3 \sum_{\alpha=1}^N \sum_{\beta=1}^N [x_{nm}(r_{\alpha m} - r_{\beta m}) - (x_{\alpha n} - x_{\beta n})]^2 \quad (12.17)$$

To determine $x_{n\ell}$, we perform partial differentiation of S with respect to x_{nm} to obtain (see the derivation of equation 12.13 or, for details, the derivation of equation 12.20 with $\lambda = 0$) an expression analogous to equation 12.16,

$$x_{n\ell} = \frac{1}{N^2} \left[\sum_{\alpha \neq \beta} (x_{\alpha n} - x_{\beta n})(r_{\alpha k} - r_{\beta k}) \right] R_{k\ell}^{-1} \quad (12.18)$$

In effect, the gradient $x_{n\ell}$ has been determined independently for each component x_n of the vector.

12.3.2 The Gradient of a Solenoidal Vector

Very often the vector field is solenoidal, that is, it is known to satisfy

$$\nabla \cdot \mathbf{x} = x_{nn} = \sum_{n=1}^3 \frac{\partial x_n}{\partial r_n} = 0 \quad (12.19)$$

which is the case, for example, of both the magnetic field and the mass flux. This condition may be used to attempt to improve the determination of $\partial x_n / \partial r_m$, or to estimate the error in the determination; but caution must be exercised for reasons to be explained in Section 15.3. We wish to determine the parameters $x'_{n\ell} = \partial x_n / \partial r_\ell$ by minimising the function S of equation 12.17 subject to the condition 12.19. We introduce $F = \frac{1}{N^2} S + 4\lambda x_{nn}$ where λ is a Lagrangian multiplier (and the factors $1/N^2$ and 4 are introduced for later convenience), and solve the equations $\partial F / \partial \lambda = 0$ and $\partial F / \partial x_{nk} = 0$ (for $n = 1, 2, 3$ and $k = 1, 2, 3$), to obtain equation 12.19 and

$$\frac{\partial F}{\partial x_{nk}} = \frac{2}{N^2} \sum_{\alpha=1}^N \sum_{\beta=1}^N [x'_{nm}(r_{\alpha m} - r_{\beta m}) - (x_{\alpha n} - x_{\beta n})](r_{\alpha k} - r_{\beta k}) + 2\lambda \delta_{nk} = 0$$

This equation reduces to

$$x'_{nm} \sum_{\alpha} \sum_{\beta} [(r_{\alpha m} - r_{\beta m})(r_{\alpha k} - r_{\beta k})] - \sum_{\alpha} \sum_{\beta} [(x_{\alpha n} - x_{\beta n})(r_{\alpha k} - r_{\beta k})] + 2N^2 \lambda \delta_{nk} = 0$$

The first term is purely geometric, and may be expressed (compare equation 12.12) in terms of the volumetric tensor, to give

$$2x'_{nm} R_{mk} - \frac{1}{N^2} \sum_{\alpha} \sum_{\beta} [(x_{\alpha n} - x_{\beta n})(r_{\alpha k} - r_{\beta k})] + 2\lambda \delta_{nk} = 0$$

which may be solved, thus

$$x'_{n\ell} = \left[\frac{1}{N^2} \sum_{\alpha \neq \beta} [(x_{\alpha n} - x_{\beta n})(r_{\alpha k} - r_{\beta k})] - \lambda \delta_{nk} \right] R_{k\ell}^{-1} = x_{n\ell} - \lambda R_{n\ell}^{-1} \quad (12.20)$$

where $x_{n\ell}$ is defined by equation 12.18. This expression satisfies equation 12.19 if

$$\sum_n x'_{nn} = \sum_n \left[x_{nn} - \lambda R_{nn}^{-1} \right] = 0$$

that is, if the Lagrangian multiplier λ is determined by

$$\lambda = \frac{\sum_n x_{nn}}{\sum_n R_{nn}^{-1}} \quad (12.21)$$

Together, equations 12.20 and 12.21 determine $x'_{n\ell}$ subject to condition 12.19.

12.4 The Volumetric Tensor \mathbf{R}

The expressions 12.9 and 12.13 for the direction and speed of a moving boundary, and the expressions 12.16, 12.18 and 12.20 for a spatial gradient, all contain the inverse of the tensor \mathbf{R} defined by equation 12.6. This tensor, and its inverse, are of fundamental importance. Note that in an arbitrary system of coordinates, the mesocentre

$$\mathbf{r}_b = \frac{1}{N} \sum_{\alpha=1}^N \mathbf{r}_\alpha \quad (12.22)$$

is no longer at the origin of coordinates (compare equation 12.4), and the definition (equation 12.6) of \mathbf{R} generalises to

$$\mathbf{R} = \frac{1}{N} \sum_{\alpha=1}^N (\mathbf{r}_\alpha - \mathbf{r}_b) (\mathbf{r}_\alpha - \mathbf{r}_b)^T = \frac{1}{N} \sum_{\alpha=1}^N \mathbf{r}_\alpha \mathbf{r}_\alpha^T - \mathbf{r}_b \mathbf{r}_b^T \quad (12.23)$$

\mathbf{R} is determined uniquely from the known orbital positions of the N spacecraft. Its fundamental importance comes from the ubiquitous way in which it appears whenever a boundary normal or spatial gradient is determined from multipoint observations, for which purpose it must be inverted. (We note that the reciprocal tensor \mathbf{K} of section 14.3.1 is related to \mathbf{R} , $\mathbf{K} = \frac{1}{4}(\mathbf{R})^{-1}$, as will be proved in section 15.2). Mathematically, \mathbf{R} cannot be inverted if it is singular, that is, if one of its eigenvalues is zero, and the spacecraft are coplanar. Physically, the equivalent condition is that one of its eigenvalues is less than the uncertainty in the spacecraft position: the spacecraft are too close to being coplanar for the spatial gradient to be determined completely. (The determination of spatial gradients using the methods of this chapter and of Chapter 14 is the subject of Chapter 15.)

As will be shown in the following section, the tensor \mathbf{R} is closely related to the size and shape of the polyhedron. For this and further geometrical reasons explained in section 12.4.3, \mathbf{R} is called the “volumetric tensor” of the tetrahedron.

12.4.1 A Simple Geometric Interpretation

Consider a plane defined by its normal $\hat{\mathbf{n}}$. The mean square distance $\sigma^2(\hat{\mathbf{n}})$ of the N spacecraft from this plane is given by,

$$\sigma^2(\hat{\mathbf{n}}) = \frac{1}{N} \sum_{\alpha=1}^N [\hat{\mathbf{n}} \cdot (\mathbf{r}_\alpha - \mathbf{r}_0)]^2 = \frac{1}{N} \sum_{\alpha=1}^N (n_j r_{\alpha j} - a)(n_k r_{\alpha k} - a) \quad (12.24)$$

$$= \frac{1}{N} \sum_{\alpha=1}^N r_{\alpha j} r_{\alpha k} n_j n_k - 2a \frac{1}{N} \sum_{\alpha=1}^N \hat{\mathbf{n}} \cdot \mathbf{r}_{\alpha} + a^2$$

where \mathbf{r}_0 is any point in the plane, and $a = \hat{\mathbf{n}} \cdot \mathbf{r}_0$. It is easy to show that, for fixed \mathbf{r}_{α} and $\hat{\mathbf{n}}$, the minimum value of σ^2 occurs when $a = \frac{1}{N} \sum_{\alpha=1}^N \hat{\mathbf{n}} \cdot \mathbf{r}_{\alpha}$. This is true whatever the direction $\hat{\mathbf{n}}$; therefore

$$\mathbf{r}_0 = \frac{1}{N} \sum_{\alpha=1}^N \mathbf{r}_{\alpha} = \mathbf{r}_b$$

The plane of minimum mean squares contains the mesocentre \mathbf{r}_b of the polyhedron.

Introducing the tensor \mathbf{R} of equation 12.23 into equation 12.24, the root mean square “deviation” of the spacecraft from their mesocentre in the $\hat{\mathbf{n}}$ direction may be expressed

$$\text{RMS deviation} = \sigma(\hat{\mathbf{n}}) = \sqrt{R_{jk} n_j n_k} \quad (12.25)$$

The dimension, or thickness, of the polyhedron is twice this value:

$$\text{dimension} = 2 \times \sigma(\hat{\mathbf{n}}) = 2 \sqrt{R_{jk} n_j n_k}$$

As the direction of the unit vector $\hat{\mathbf{n}}$ varies over 4π steradians, so the value of $\sigma(\hat{\mathbf{n}})$ describes the surface of an quasi-ellipsoid. A line in any direction $\hat{\mathbf{n}}$ intersects this surface at two points whose separation is equal to the RMS “thickness” of the polyhedron in that direction. There are three directions in which $\sigma(\hat{\mathbf{n}})$ is stationary: these are the principle axes of the quasi-ellipsoid, defined by the eigenvectors of \mathbf{R} . We denote the corresponding values of $\sigma(\hat{\mathbf{n}})$ by a , b and c , with

$$a \geq b \geq c$$

a^2 , b^2 and c^2 are the eigenvalues of \mathbf{R} , and a , b and c represent respectively the major, middle and minor semiaxes of the quasi-ellipsoid.

12.4.2 Relation to the Inertia Tensor

The volumetric tensor is closely related to the inertia tensor.

If the N spacecraft were located rigidly with respect to one another and rotating at angular velocity $\boldsymbol{\omega}$ around an axis passing through the mesocentre, their angular momentum would be

$$\mathbf{h} = \sum_{\alpha=1}^N \left\{ \mathbf{r}_{\alpha} \times m \frac{\partial \mathbf{r}_{\alpha}}{\partial t} \right\} = m \sum_{\alpha=1}^N \{ \mathbf{r}_{\alpha} \times (\boldsymbol{\omega} \times \mathbf{r}_{\alpha}) \}$$

where m is the mass of each spacecraft. This may be written

$$h_j = m \sum_{\alpha=1}^N \epsilon_{jkl} r_{\alpha k} \epsilon_{lmn} \omega_m r_{\alpha n} = m \sum_{\alpha=1}^N (\delta_{jm} \delta_{kn} - \delta_{jn} \delta_{km}) r_{\alpha k} \omega_m r_{\alpha n}$$

which is

$$h_j = m \sum_{\alpha=1}^N \left(|r_{\alpha}|^2 \delta_{jk} - r_{\alpha j} r_{\alpha k} \right) \omega_k = I_{jk} \omega_k$$

where I_{jk} are the components of the inertia tensor \mathbf{I} . Thus,

$$\frac{1}{Nm} I_{jk} + R_{jk} = S^2 \delta_{jk} \quad \text{where} \quad S^2 = \frac{1}{N} \sum_{\alpha=1}^N |r_{\alpha}|^2 \quad (12.26)$$

is a parameter which characterises the size of the cluster of N spacecraft. The inertia tensor and the volumetric tensor have the same eigenvectors and their eigenvalues are related by

$$\frac{1}{Nm} I^{(n)} + R^{(4-n)} = S^2 \quad \text{for} \quad 1 \leq n \leq 3 \quad (12.27)$$

The eigenvector associated with the largest eigenvalue of the inertia tensor is associated with the smallest eigenvalue of the volumetric tensor, and *vice versa*.

12.4.3 The Special Case of a Tetrahedron

We have seen that twice the square roots of the eigenvalues, $2\sqrt{R^{(n)}}$ for $n = 1, 2$ or 3 , yield the characteristic dimensions $2a$, $2b$, and $2c$, of the tetrahedron, in three mutually orthogonal directions in space defined by the corresponding eigenvectors. It is proved below that these characteristic dimensions have the following properties in the special case of four spacecraft.

1. The volume V of the tetrahedron defined by the four spacecraft is one third of the product of the three characteristic dimensions. In other words,

$$\frac{V}{8abc} = \frac{1}{3} \quad \text{which is} \quad V = \frac{8}{3} \sqrt{|\mathbf{R}|} \quad (12.28)$$

The eigenvector \mathbf{R}^c associated with the smallest eigenvalue $R^{(c)}$ is the direction in which the tetrahedron is “flattened”. This is the component of the spatial gradient (or the boundary normal) which is least well defined, as explained further in Chapter 13.

2. If the spacecraft are coplanar, $c^2 = R^{(c)} = 0$ and the ratio of the area A of the quadrilateral defined by the four spacecraft to the product $4ab$ of the two non-zero characteristic dimensions lies in the range

$$\frac{1}{\sqrt{2}} \leq \frac{A}{4ab} \leq 1 \quad (12.29)$$

where:

- the upper limit applies when the spacecraft lie at the four corners of a parallelogram;
- the lower limit applies to the triangle defined by the spacecraft when any two of them are coincident.

The eigenvector \mathbf{R}^c corresponding to the null eigenvalue is the pole of the plane defined by the four spacecraft.

3. If the satellites are colinear, $b = c = 0$ and the overall distance L between the outer spacecraft is related to the characteristic dimension $2a$ by

$$1 \leq \frac{L}{2a} \leq \sqrt{2} \quad (12.30)$$

where:

- the upper limit applies when two of the spacecraft are coincident midway between the other two;
- the lower limit applies when there are two pairs of coincident spacecraft.

The eigenvector \mathbf{R}^a corresponding to the non-zero eigenvalue defines the direction of alignment of the spacecraft.

It is interesting to compare the expressions 12.28 through 12.30 with the definition (Section 13.3.3) of the characteristic size $L = 2a$ of the polyhedron.

Proof of the Expression for the Volume of a Tetrahedron

The expression 12.28 may be proved as follows. The volume of the tetrahedron with apexes defined by the four vectors \mathbf{r}_α for $1 \leq \alpha \leq 4$ is

$$\begin{aligned} V &= \pm \frac{1}{6} [\mathbf{r}_1 - \mathbf{r}_4] \cdot [(\mathbf{r}_2 - \mathbf{r}_4) \times (\mathbf{r}_3 - \mathbf{r}_4)] \\ &= \pm \frac{1}{6} \{ |\mathbf{r}_1, \mathbf{r}_2, \mathbf{r}_3| - |\mathbf{r}_2, \mathbf{r}_3, \mathbf{r}_4| + |\mathbf{r}_3, \mathbf{r}_4, \mathbf{r}_1| - |\mathbf{r}_4, \mathbf{r}_1, \mathbf{r}_2| \} \end{aligned} \quad (12.31)$$

where

$$|\mathbf{r}_\alpha, \mathbf{r}_\beta, \mathbf{r}_\gamma| = \mathbf{r}_\alpha \cdot (\mathbf{r}_\beta \times \mathbf{r}_\gamma)$$

The upper sign is used if $\mathbf{r}_1, \mathbf{r}_2, \mathbf{r}_3$ form a right-handed triad with respect to \mathbf{r}_4 , the lower sign otherwise. When the position vectors \mathbf{r}_α are measured from the position of their mean, $\mathbf{r}_b = 0$ and equation 12.22 reduces to equation 12.4 which may be used, together with the general properties of matrices, to eliminate \mathbf{r}_4 from equation 12.31. Thus,

$$|\mathbf{r}_2, \mathbf{r}_3, \mathbf{r}_4| = |\mathbf{r}_2, \mathbf{r}_3, -\mathbf{r}_1 - \mathbf{r}_2 - \mathbf{r}_3| = -|\mathbf{r}_2, \mathbf{r}_3, \mathbf{r}_1| = -|\mathbf{r}_1, \mathbf{r}_2, \mathbf{r}_3|$$

and similarly

$$\begin{aligned} |\mathbf{r}_3, \mathbf{r}_4, \mathbf{r}_1| &= -|\mathbf{r}_3, \mathbf{r}_2, \mathbf{r}_1| = +|\mathbf{r}_1, \mathbf{r}_2, \mathbf{r}_3| \\ |\mathbf{r}_4, \mathbf{r}_1, \mathbf{r}_2| &= -|\mathbf{r}_3, \mathbf{r}_1, \mathbf{r}_2| = -|\mathbf{r}_1, \mathbf{r}_2, \mathbf{r}_3| \end{aligned}$$

Thus, when coordinates with origin defined by equation 12.4 are used, equation 12.31 simplifies to

$$V = \pm \frac{2}{3} |\mathbf{r}_1, \mathbf{r}_2, \mathbf{r}_3| \quad (12.32)$$

To prove equation 12.28, we may use equations 12.4 and 12.6, to obtain

$$R_{jk} = \frac{1}{4} \left[\sum_{\alpha=1}^3 r_{\alpha j} r_{\alpha k} - (r_{1j} + r_{2j} + r_{3j}) r_{4k} \right] = \frac{1}{4} \sum_{\alpha=1}^3 r_{\alpha j} (r_{\alpha k} - r_{4k})$$

The right-hand side of the above is the product of two 3×3 matrices, with elements $r_{\alpha j}$ and $r_{\alpha k} - r_{4k}$, respectively, ($\alpha, j, k = 1, 2, 3$). Hence, using the rule for the determinant of a product of matrices,

$$|\mathbf{R}| = \frac{1}{4^3} |r_{\alpha j}| |r_{\alpha k} - r_{4k}|$$

Now, from equations 12.31 and 12.32 respectively

$$|r_{\alpha k} - r_{4k}| = \pm 6V \quad \text{and} \quad |r_{\alpha j}| = \pm \frac{3V}{2}$$

so that

$$|\mathbf{R}| = \frac{9V^2}{4^3} = \left(\frac{3V}{8}\right)^2$$

and the statement is proved.

An alternative proof of equation 12.28 is supplied in Appendix 12.A, where the inequalities 12.29 and 12.30 are also discussed.

12.5 Discussion

In this chapter we have derived expressions for the homogeneous least squares determination of a boundary normal or spatial gradient from measurements obtained on four or more spacecraft. The method allows the solenoidal condition (zero divergence) to be imposed when determining the gradient of a vector. The use of the method will be illustrated in Chapter 15.

All the expressions are derived in terms of the inverse of the symmetric tensor \mathbf{R} , which is defined in terms of the spacecraft spatial coordinates by equation 12.23. In the particular case of four spacecraft, \mathbf{R} yields the volume of the tetrahedron via equation 12.28; for this reason it is called the volumetric tensor. It is closely related to the inertia tensor (equation 12.26). We note that the reciprocal tensor introduced in Chapter 14 is shown in Section 15.2 to be equal to $\frac{1}{4}\mathbf{R}^{-1}$, thus proving the equivalence of the linear barycentric and the least squares methods of determining spatial gradients.

The eigenvalues of the volumetric tensor \mathbf{R} yield the characteristic dimension (mean square semi-thickness) of the polyhedron in three mutually orthogonal directions, plus the spatial orientation of these directions. The converse is also true; because a symmetric tensor is completely defined by its eigenvalues and eigenvectors, these characteristic dimensions and their orientation in space completely define the volumetric tensor. Therefore they, or parameters closely related to them, are to be recommended for describing the geometry of the polyhedron. In Section 13.3.3 of Chapter 13 the eigenvalues of \mathbf{R} are expressed in terms of more intuitively descriptive parameters: the “size”, “elongation” and “planarity” of the polyhedron, together with the corresponding (mutually orthogonal) axes of elongation and of planarity.

Appendix

12.A Spacecraft Non-Coplanar: Alternative Proof for the Tetrahedron Volume

A shorter proof of equation 12.28 uses the fact that a unimodular affine transformation leaves the value of the determinant (but not the matrix itself) unchanged. Therefore we can conveniently choose $\mathbf{r}_4 = \mathbf{0}$, so that

$$R_{jk} = \frac{1}{4} \sum_{\alpha=1}^3 r_{\alpha j} r_{\alpha k}$$

Then, using the rule for the determinant of a product and remembering that \mathbf{R} is a matrix of dimension 4,

$$|\mathbf{R}| = \frac{1}{4^4} |r_{\alpha j}| |r_{\alpha k}|$$

When $\mathbf{r}_4 = \mathbf{0}$ equation 12.4 is not satisfied, and we must use equation 12.31 instead of equation 12.32: thus

$$|r_{\alpha j}| = 6V$$

so that, finally,

$$|\mathbf{R}| = \frac{(6V)^2}{4^4} = \left(\frac{3V}{8}\right)^2$$

12.B Spacecraft Coplanar

When the four spacecraft are coplanar, one eigenvalue is zero and the corresponding eigenvector is normal to the plane of the spacecraft.

The area of the quadrilateral defined by the four spacecraft can be determined by consideration of small out-of-plane perturbations. These perturbations must be applied in such a way as

- to satisfy equation 12.4, and
- not to change the direction of the eigenvectors of \mathbf{R} .

There are two configurations for which perturbations satisfying these conditions can readily be found:

- the spacecraft lie at the four corners of a parallelogram, and
- two spacecraft are coincident, the four spacecraft thus defining a triangle.

Spacecraft at the Corners of a Parallelogram

Let the four spacecraft locations be

$$\begin{aligned} \mathbf{r}_1, & \quad \mathbf{r}_3 = -\mathbf{r}_1, \\ \mathbf{r}_2, & \quad \text{and} \quad \mathbf{r}_4 = -\mathbf{r}_2, \end{aligned} \quad (12.33)$$

where \mathbf{r}_1 and \mathbf{r}_2 are arbitrary. Consider the effect of perturbations $\delta\mathbf{r}$ applied in a way which preserves the validity of equation 12.4:

$$\begin{aligned} \mathbf{r}_1 &\rightarrow \mathbf{r}_1 + \delta\mathbf{r}, & \mathbf{r}_3 &\rightarrow \mathbf{r}_3 + \delta\mathbf{r} \\ \mathbf{r}_2 &\rightarrow \mathbf{r}_2 - \delta\mathbf{r}, & \mathbf{r}_4 &\rightarrow \mathbf{r}_4 - \delta\mathbf{r} \end{aligned} \quad (12.34)$$

To the first order in $\delta\mathbf{r}$,

$$R_{jk} \rightarrow R_{jk} + (r_{1j} - r_{2j} + r_{3j} - r_{4j})\delta r_k + \delta r_j(r_{1k} - r_{2k} + r_{3k} - r_{4k}) = R_{jk} \quad (12.35)$$

The first order perturbation of \mathbf{R} is zero, and neither the eigenvalues nor the eigenvectors are perturbed. We now derive the second order perturbation of the smallest eigenvalue.

Let the perturbation $\delta\mathbf{r}$ be parallel to the eigenvector $\mathbf{e}^{(c)}$, corresponding to the smallest eigenvalue and perpendicular to the plane of the spacecraft, say

$$\delta\mathbf{r} = \epsilon \mathbf{e}^{(c)}$$

Following application of the perturbation 12.34, the RMS thickness of the tetrahedron in the $\mathbf{e}^{(c)}$ direction is 2ϵ so that, from equation 12.25,

$$\lambda^{(1)} \rightarrow \lambda^{(1)} + 4\epsilon^2 = 4\epsilon^2$$

Simultaneously, spacecraft 4 is a distance 4ϵ above the plane of the (perturbed) triangle defined by spacecraft 1, 2 and 3. The volume of the tetrahedron is

$$\text{Volume} = \frac{1}{3} \times 4\epsilon \times \text{area of the triangle 123} = \frac{1}{3} \sqrt{\lambda^{(1)}\lambda^{(2)}\lambda^{(3)}} = \frac{1}{3} \sqrt{4\epsilon^2\lambda^{(2)}\lambda^{(3)}}$$

Therefore

$$\text{area of the triangle 123} = \frac{1}{2} \sqrt{\lambda^{(2)}\lambda^{(3)}}$$

The area of the parallelogram 1234 is twice the area of the triangle 123,

$$\text{area of the parallelogram 1234} = \sqrt{\lambda^{(2)}\lambda^{(3)}}$$

Two Spacecraft Coincident

Let the four spacecraft be located at

$$\mathbf{r}_1, \quad \mathbf{r}_2, \quad \mathbf{r}_3, \quad \text{and} \quad \mathbf{r}_4 = \mathbf{r}_3, \quad (12.36)$$

where \mathbf{r}_1 , \mathbf{r}_2 and \mathbf{r}_3 are arbitrary. Consider application of the perturbations

$$\mathbf{r}_3 \rightarrow \mathbf{r}_3 + \delta\mathbf{r} \quad \mathbf{r}_4 \rightarrow \mathbf{r}_4 - \delta\mathbf{r} \quad (12.37)$$

The first order perturbation of \mathbf{R} is

$$R_{jk} \rightarrow R_{jk} + (r_{3j} - r_{4j})\delta r_k + \delta r_j(r_{3k} - r_{4k}) = R_{jk} \quad (12.38)$$

Again, the first order perturbation of \mathbf{R} is zero, and we derive the second order perturbation of the smallest eigenvalue.

Let the perturbation $\delta \mathbf{r}$ be parallel to the eigenvector $\mathbf{e}^{(c)}$

$$\delta \mathbf{r} = \epsilon \mathbf{e}^{(1)} \quad (12.39)$$

The RMS thickness in the $\mathbf{e}^{(c)}$ direction is $\sqrt{2}\epsilon$ so that, from equation 12.25,

$$a^{(1)} \rightarrow a^{(1)} + 2\epsilon^2 = 2\epsilon^2$$

Following the perturbation 12.37, spacecraft 4 is at a distance 2ϵ above the plane of the triangle defined by spacecraft 1, 2 and 3. The volume of the tetrahedron is

$$\text{Volume} = \frac{1}{3} \times 2\epsilon \times \text{area of the triangle 123} = \frac{1}{3} \sqrt{\lambda^{(1)}\lambda^{(2)}\lambda^{(3)}} = \frac{1}{3} \sqrt{2\epsilon^2\lambda^{(2)}\lambda^{(3)}}$$

Therefore (spacecraft 4 being coincident with spacecraft 3)

$$\text{area of the triangle 123} = \sqrt{\frac{1}{2}\lambda^{(2)}\lambda^{(3)}}$$

Thus we have proved the two planar configurations considered in section 12.4.3 yield the limits of equation 12.29. But we have still not proved analytically that these two configurations do indeed correspond to the limiting (maximum and minimum) values of A in terms of a and b . This result has been inferred by numerical simulation.

12.C Spacecraft Colinear

A similar comment applies when the spacecraft are colinear. For the two colinear configurations considered in section 12.4.3, it is easy to establish the limits of equation 12.30.

Acknowledgements

The author wishes to thank the ISSI working group for stimulating discussion, and particularly Anders Eriksson and Joachim Vogt for simplification of the mathematics.

Bibliography

To the author's knowledge, there is none on the specific subject of this chapter. Nevertheless, the bibliography of Chapters 13 through 15 may usefully be consulted.

— 13 —

Tetrahedron Geometric Factors

PATRICK ROBERT AND ALAIN ROUX

*Centre National de la Recherche Scientifique
Vélizy, France*

CHRISTOPHER C. HARVEY

*Observatoire de Paris-Meudon
Meudon, France*

MALCOLM W. DUNLOP

*Imperial College of Science, Technology and Medicine
London, United Kingdom*

PATRICK W. DALY

*Max Planck Institut für Aeronomie
Katlenburg-Lindau, Germany*

KARL-HEINZ GLASSMEIER

*Technische Universität Braunschweig
Braunschweig, Germany*

13.1 Introduction

The geometrical shape of the tetrahedron formed by the spacecraft is an essential criterion in the choice of scientific investigation which can be performed with data from a multi-spacecraft mission, such as Cluster. The shape of this polyhedron evolves continuously along the orbital trajectory of the spacecraft, and this shape has a major impact on the accuracy of the determination of scientific parameters related to the spatial gradient, such as the current density, which is discussed in Chapter 16. The scientific importance of the shape, combined with its variability, inevitably led to many proposals for “quality factors” to attempt to describe the geometric shape of the tetrahedron, or for “performance indicators”, to indicate the likely error of a particular scientific parameter.

While these early geometric factors were all one-dimensional, 2-D parameters to characterise the geometrical shape of the tetrahedron have also been proposed, i.e., the “elongation” E and “planarity” P defined in terms of the eigenvalues of the volumetric tensor described in Chapter 12.

In this chapter, we use the E and P parameters to define five characteristic types of tetrahedra and we check the validity and the meaning of the 1-D geometric factors by a numerical simulation using an “homogeneous reservoir of tetrahedra” in the E - P configuration space. As a practical application, we present an example of the Cluster orbit, and the associated computation of the 1-D and 2-D geometric factors. We represent these quantities in the E - P diagram, which allows a better understanding of their meaning. Finally, we demonstrate the limits of the 1-D geometric factors and point out the advantages of a 2-D geometric factor.

13.2 Measurement Performance

Tetrahedral geometry is one of the principal factors affecting measurement performance; that is the precision of physical parameters derived by comparison of data acquired at four points in space. There are, in fact, three factors which affect this precision: the tetrahedral geometry, the structure (in time and space) of the phenomena sampled, and the inevitable experimental errors inherent to all physical measurements. Measurement accuracy includes not only instrumental accuracy, but also timing and location accuracy. As mentioned in the introduction, the treatment of errors is covered by Chapters 11, 16, 17, in terms of the determination of different sets of criteria for each physical parameter to be determined. Different analysis techniques are applied to four-point measurements to derive different physical parameters, such as the local current density (involving spatial field gradients), wave vector or mode, or global structure (boundaries). Each technique imposes different criteria on adequate sampling for measurement quality. Each of these criteria could be monitored separately, or given differing emphasis, depending on which particular physical property is of interest.

Measurement quality is therefore not determined only by the geometric “quality” of the tetrahedron (or polyhedron). Even for events which do not evolve structurally with time, the sampling achieved *of the physical event* depends upon the geometry (and scale) of the tetrahedron relative to that of the physical structure (anisotropy of the phenomena) present. For a highly anisotropic physical structure, a particular alignment of an anisotropic tetrahedral spacecraft configuration may be optimal, for example, to determine the spatial gradient. Different relative event scales, however, will result in different measurement performance for any given tetrahedral size and shape. Multipoint analysis typically involves the determination of gradients so that, for any given polyhedron overall size, derived quantities will, typically, be sensitive to the tetrahedral geometry when sampling similar physical structures.

Measurement quality depends also on the size of the tetrahedron, compared to the product of the measurement time resolution and the spacecraft relative velocity with respect to the physical structure, i.e. the interval the spacecraft travel into the structure within one data accumulation period. Note that for the particle experiments the data accumulation period is typically equal to the spacecraft spin, so as to sample a complete 3-D distribution function. In the case of Cluster this is 4 seconds, which determines the minimum size the tetrahedron should have for the various spacecraft/physical-structure relative velocities.

Consider Figure 13.1, for example, which indicates the evolution of the spacecraft configuration around a Cluster orbit for two proposed scenarios. Note how very different the evolution is and how the geometry varies widely in shape and size over the orbit. The insets show enlarged (by a factor of 50 with respect to the main figure) configurations, projected into the plane of view. The first group of insets, at positions 1, 2, and 3, show a highly elongated configuration at the southern magnetopause crossing (3). The orientation of this can be changed (not simply) by changing the orientation at 1 and, for instance, for some simulated mission phases has a more parallel alignment to the boundary. Faced with such a predicted tetrahedral geometry, the physical parameters which can be well determined depend upon the orientation with respect to the boundary. For such a nearly 1-D structure, techniques which determine those parameters depending on spatial structure (gradients, such as for $\nabla \times \mathbf{B}$ or $\nabla \times \mathbf{V}$) will typically require a configuration aligned with the boundary (i.e., matching the small and large gradients), but techniques which analyse

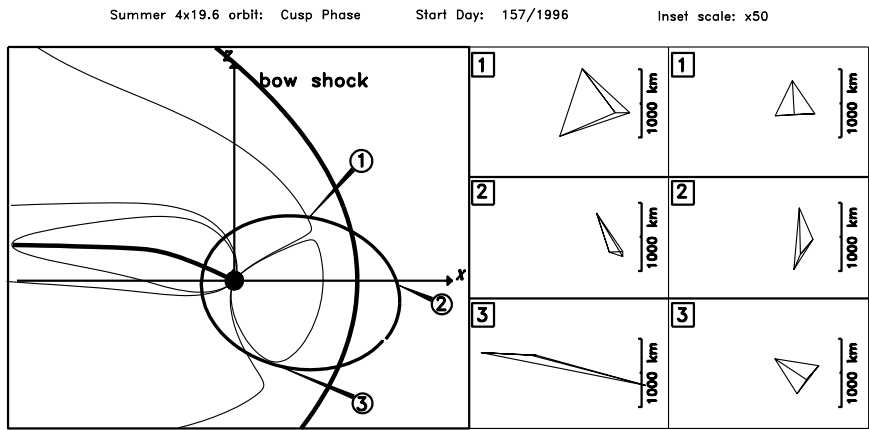


Figure 13.1: Evolution of the Cluster configuration around the nominal orbit for the day-side phase. The insets show two options which both target the northern cusp with a regular tetrahedron, but target the southern cusp with a regular tetrahedron only for the second case. The first option is optimised for fuel. [Reproduced from *Balogh et al., 1997.*]

macroscopic properties will benefit from an anisotropic configuration in different ways: motional properties will be best sampled by perpendicular alignment, whereas boundary shape (especially non-planar) is best sampled by parallel alignment.

A quality parameter that monitors only spacecraft configuration is particularly useful, however, when sampling of structure is not important. Such a parameter would best reflect performance relating to transient or fluctuating events, for instance, with no preferred orientation to the global structure. A large number of events are not predictable and therefore a regular tetrahedron is optimum in this situation. We call here a regular tetrahedron a particular tetrahedron where the separations between each pair of points are equal. For the second scenario in the figure, for instance, the target at 3 has been chosen to correspond to a second, regular tetrahedron in an attempt to regulate the evolution over the orbit. The effect of tetrahedral distortion in terms of geometric quality parameters is studied in detail in the next section.

Optimum configurations in terms of either physical sampling or measurement uncertainty, as discussed above, are only likely to be achieved over small segments of the orbit. It would seem sensible, therefore, to attempt to optimise for data quality over selected global regions, together with choice of spatial scale, as a primary constraint. For other regions, use may be made of the natural distortion of the configuration to achieve preferred orientations with respect to the sampled structure.

13.3 The Shape of the Tetrahedron

13.3.1 The 1-Dimensional Geometric Factors

Four points in space define a *tetrahedron*. If the separations between each pair of points are equal, then it is a *regular* tetrahedron. Four spacecraft will form a tetrahedron,

which in general will not be regular. How can we specify the degree to which regularity is achieved? A number of parameters have been proposed to accomplish this, which we present and compare below.

The Q_{GM} Parameter

The Q_{GM} parameter is defined as

$$Q_{GM} = \frac{\text{True Volume}}{\text{Ideal Volume}} + \frac{\text{True Surface}}{\text{Ideal Surface}} + 1 \quad (13.1)$$

The *ideal* volume and surface are calculated for a regular tetrahedron with a side length equal to the average of the 6 distances between the 4 points.

Q_{GM} takes values between 1 and 3, and attempts to describe the “fractional dimension” of the tetrahedron: a value of 1 indicates that the four spacecraft are in a line, while a value equal to 3 indicates that the tetrahedron is regular. There is nevertheless some difficulty with this interpretation: it is perfectly possible to deform a regular ($Q_{GM} = 3$) tetrahedron continuously until it resembles a straight line ($Q_{GM} = 1$) without it resembling a plane at any time; therefore $Q_{GM} = 2$ is not a sufficient condition for planarity.

The Q_{RR} Parameter

The Q_{RR} parameter is defined to be

$$Q_{RR} = \left(\frac{9\pi}{2\sqrt{3}} \cdot \frac{\text{True Volume}}{\text{Sphere Volume}} \right)^{\frac{1}{3}} \quad (13.2)$$

where the sphere is that circumscribing the tetrahedron (all four points on its surface). Q_{RR} is normalised to be equal to 1 for a regular tetrahedron; its minimum value is 0. This parameter was selected from many on the basis of its usefulness in estimating the error in the determination of the spatial gradient of the magnetic field. This is discussed in section 13.5.4.

The Q_{SR} Geometric Factor

Another of the 1-D parameters is known as the Q_{SR} geometric factor. This factor is simply defined by:

$$Q_{SR} = \frac{1}{2} \left(\frac{a + b + c}{a} - 1 \right) \quad (13.3)$$

where a, b, c are the lengths of the 3 axes of the pseudo-ellipsoid (see section 13.3.2).

The Q_{R8} Geometric Factor

Finally, another 1-D factor named Q_{R8} is defined by:

$$Q_{R8} = \frac{\text{True Volume}}{\text{Ideal Volume}} \quad (13.4)$$

These 1-D geometric factors are studied in Section 13.4.3 in order to establish a relationship between their values and the type of the tetrahedra defined in Section 13.4.1. To do this, we need to use a “reservoir of five types of tetrahedra” described in section 13.4.2.

13.3.2 A Geometric Representation of the Size, Shape, and Orientation of a Polyhedron

Since none of these 1-D parameters is sufficient to characterise both the shape of the tetrahedron and the accuracy of the \mathbf{J} determination, we now introduce two parameters to characterise the shape of the tetrahedron in a 2-D parameter space.

These two parameters are derived from the volumetric tensor introduced in Chapter 12 in connection with the determination of spatial gradients. It was shown in Chapter 15 that the linear barycentric and least squares methods of determining spatial gradients are equivalent; therefore the volumetric tensor must contain all the relevant geometrical information needed to determine the spatial gradient by either of these two methods. This suggests strongly that parameters which describe the volumetric tensor will be rather useful in practice. It may also be noted that the volumetric tensor, and parameters derived from it, are valid for a general polyhedron defined by four or more spacecraft.

The volumetric tensor is symmetric. A symmetric tensor describes a quadratic form which can be represented by an ellipsoid in space; this ellipsoid has three principal axes, each lying in the direction of one of the eigenvectors of the tensor, with semi-length determined by the corresponding eigenvalue.

We recall here the definition of the tensor \mathbf{R} , fully defined in Section 12.4 (page 315):

$$R_{jk} = \frac{1}{N} \sum_{\alpha=1}^N (r_{\alpha j} - r_{bj}) (r_{\alpha k} - r_{bk}) = \frac{1}{N} \sum_{\alpha=1}^N r_{\alpha j} r_{\alpha k} - r_{bj} r_{bk} \quad (13.5)$$

which is the component form of equation 12.23.

When N is the number of vertices (or spacecraft), $r_{\alpha j}$ is the j component of vertex α , and r_{bj} is the mean value, over all α , of $r_{\alpha j}$. If the origin of coordinates is chosen to be the mesocentre, then the tensor \mathbf{R} can be written

$$R_{jk} = \frac{1}{N} \sum_{\alpha=1}^N r_{\alpha j} r_{\alpha k} \quad (13.6)$$

\mathbf{R} is determined uniquely from the known orbital positions of the N spacecraft. It attempts to describe the size and the anisotropy of the polyhedron (see Chapter 12).

The principle axes of the pseudo-ellipsoid are given by the eigenvectors $\mathbf{R}^{(n)}$ of \mathbf{R} . If we order the eigenvalues as:

$$R^{(1)} \geq R^{(2)} \geq R^{(3)} \quad (13.7)$$

their square roots represent respectively the major, middle and minor semiaxes of the pseudo-ellipsoid:

$$\begin{aligned} a &= \sqrt{R^{(1)}} \\ b &= \sqrt{R^{(2)}} \\ c &= \sqrt{R^{(3)}} \end{aligned} \quad (13.8)$$

Thus, the volumetric tensor, and the associated ellipsoid, provide a simple way to visualise those features of the global shape of a polyhedron which are significant for the determination of gradients. For instance, an ellipsoid reduced to a sphere corresponds to a regular polyhedron, an ellipsoid reduced to a plane ellipse corresponds to the spacecraft being coplanar, and an ellipsoid reduced to a line corresponds, of course, to the alignment

of the spacecraft. The significance of the non-zero eigenvalues in the case of four spacecraft is explained in Section 12.4.3. Again considering only four spacecraft, it may be noted that, even if the volumetric tensor were to be renormalised so that the spacecraft of a regular tetrahedron actually lie on (the surface of) the sphere, for an arbitrary configuration the spacecraft would generally not lie on the corresponding ellipsoid.

13.3.3 Size, Elongation, and Planarity of a Polyhedron

The discussion of the preceding section, and of Chapter 12, clearly demonstrates the importance of the eigenvalues of the volumetric tensor with respect to both the description of the polyhedron geometry and the calculation of spatial gradients.

Three parameters are needed to describe the three eigenvalues. It is useful for these parameters to be “intuitively descriptive”. One parameter may be used to indicate the size of the polyhedron, and the other two, elongation and planarity, to describe its shape. Furthermore, in general (when it is anisotropic) two directions are required to define completely the orientation in space of the polyhedron. The reasoning behind this choice of parameters is as follows:

- When the polyhedron is isotropic, all three eigenvalues are equal.
- If it is stretched, a^2 becomes greater than the other two eigenvalues; if stretched (or rather, if squeezed in the two orthogonal directions) until $b = c = 0$, the spacecraft would lie on a straight line. We define the *elongation*, or prolateness, to be $E = 1 - (b/a)$. Furthermore, the eigenvector \mathbf{R}^a defines the direction of elongation.
- On the other hand, if the isotropic polyhedron is squashed in one direction, c^2 becomes smaller than the other two eigenvalues; if squashed until $c = 0$, the spacecraft would lie in a plane. We define the *planarity*, or oblateness, to be $P = 1 - (c/b)$. Furthermore, the eigenvector \mathbf{R}^c defines the normal (or pole) of planarity.
- In general the polyhedron is both stretched and squashed, in mutually orthogonal directions. Together, the elongation and planarity define completely (the ratios of) the eigenvalues, and thus the physically important characteristics of the shape of the polyhedron. It remains to define a parameter to describe the size; it is convenient to use the largest eigenvector, a^2 , which is always non-zero, and to define the characteristic size as $L = 2a$.

To summarise, the physically important characteristics of the polyhedron may be described completely by:

- **characteristic size** $L = 2a$ (in any convenient unit of length)
- **elongation** $E = 1 - (b/a)$
- **direction of elongation** \mathbf{R}^a
- **planarity** $P = 1 - (c/b)$
- **normal of planarity** \mathbf{R}^c .

The direction of elongation and the normal of planarity are (by definition) orthogonal, and so only three angles (e.g., the three Euler angles) are needed to describe completely the orientation of the quasi-ellipsoid in three dimensions. These three angles, plus the values of L , E and P , provide a complete description of the volumetric tensor. We may note that:

<div><div>P</div><div>E</div></div>	0	low	interme- diate	large	1
1	Circle	Ellipse of increasing eccentricity			Straight Line
large	Ellipsoid of increasing oblateness	Pancake	Elliptical Pancake	Knife Blade	
interme- diate		Thick Pancake	Potatoes	Flattened Cigar	
low		Pseudo- Sphere	Short Cigar	Cigar	
0	Sphere	Ellipsoid of increasing prolateness			

Figure 13.2: The shape of the polyhedron as a function of E and P .

- both E and P are dimensionless, and lie in the range $0 \leq E \leq 1$, $0 \leq P \leq 1$;
- when $E = 1$, P is undefined because $b = c = 0$.

The shape of the polyhedron as E and P vary over their permitted ranges is indicated in Figure 13.2. In Section 13.4.1, we will define a limited number of general shapes to characterise the tetrahedra during the investigation of this chapter.

Note that we have defined elongation and planarity to be $1 - (b/a)$ and $1 - (c/b)$, whereas the eccentricity of an ellipse is defined by $e = \sqrt{1 - (b/a)^2}$. Now e lies in the same range $0 \leq e \leq 1$ as E and P , and the question arises as to whether elongation and planarity would have been better defined as $\sqrt{1 - (b/a)^2}$ and $\sqrt{1 - (c/b)^2}$. Then the elongation and planarity would be simply the eccentricities $e_c = \sqrt{1 - (b/a)^2}$ and $e_a = \sqrt{1 - (c/b)^2}$ of the ellipsoid respectively in the plane of planarity (containing the middle and major axes), and in the plane perpendicular to the elongation (containing its minor and middle axes). Both definitions are acceptable, but the elongation and planarity as defined above yield a more uniform distribution of points in the E - P plane. This point is, of course, entirely subjective because there is no *a priori* uniform distribution; but the statement is certainly true for typical Cluster orbits, as explained in Section 13.5.

Note that if no single parameter can reproduce all the information contained in the volumetric tensor, the converse is also true: it is not possible to express analytically the various 1-D geometric parameters in terms of the volumetric tensor, because this tensor does not describe the tetrahedron completely. A complete description would require the position of three of the apexes with respect to the fourth apex, that is, nine independent quantities of which three describe orientation and six describe shape; the symmetric volumetric tensor has only six independent quantities, of which only three describe shape.

13.4 Study of the 1-D Geometric Factors over the Tetrahedron Reservoir

We now restrict our attention again to the special case of the tetrahedron, and study several 1-D geometric factors in terms of the parameters E and P .

13.4.1 The Five Types of Tetrahedra

It is useful to limit the number of characteristic tetrahedra given in Figure 13.2 and to define only 5 representative types by means of the E and P parameters, Figure 13.3 shows where, in the E - P plane, each type of tetrahedron would be. For low values of E and P we can define a “Pseudo-Sphere-shaped geometry” (bottom left corner of the E - P diagram) corresponding to the pseudo-regular tetrahedra. For a high value of P and a low value of E (top left corner of the E - P diagram) the ellipsoid is nearly a flat circle and we can define it as “Pancake-shaped”. At the opposite side (bottom right corner) we can find a long ellipsoid with a pseudo-circular section, that we can define as a “Cigar-shaped”. Finally, at the top right corner, we can find tetrahedra which are both elongated and flat, and we can call this type the “Knife-Blade-shaped”. Note that for elongated tetrahedron the flatness does not have much physical significance. Tetrahedra that do not belong to one of these categories or types, will be referred to “Potato type” and are located at the centre of the E - P diagram. The tetrahedra which correspond to these 5 types shown in Figure 13.3 are taken from a “five types reservoir” which is now defined.

13.4.2 Computation of a Reservoir of Five Types of Tetrahedra

Many tetrahedra corresponding to one or other of the five principal types shown in Figure 13.3 have been constructed as explained below, and placed in a “reservoir”. Such a reservoir is useful in simulations in order to study the consequences of each type of configuration on the derived parameters (see Section 13.4.3). All the tetrahedra have the same mean inter-spacecraft distance:

$$\langle D \rangle = \frac{1}{6} \sum_{\alpha=1}^6 d_{\alpha} \quad (13.9)$$

When computing the reservoir we start with $\langle D \rangle = 1$. The origin of coordinates of each tetrahedron is initially the mesocentre, with the axes being in accordance with Figure 13.13.

Pseudo-Sphere. There are two components of this population:

- “Regular”, for which $\langle D \rangle$ is equal to each of the 6 inter-spacecraft distances d_{α} .
- “Random”. The major part of the “Pseudo-Sphere” population is produced by perturbation of a regular tetrahedron, all three coordinates of each vertex suffering separately a random “displacement” uniformly distributed in the range $\pm \langle D \rangle \times 15\%$.

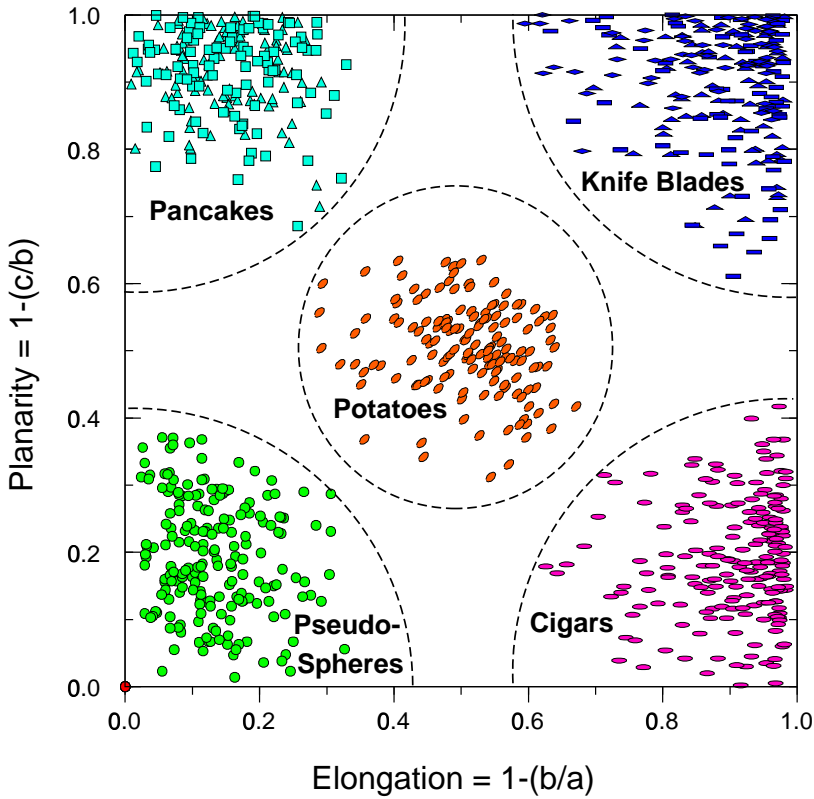


Figure 13.3: The five types of tetrahedra: Pseudo-Spheres, Pancakes, Cigars, Knife Blades, and Potatoes.

Cigar. This population is derived from the pseudo-spherical population by random elongation in the z direction, in such way as to obtain the distribution shown in Figure 13.3.

Pancake. There are two basic forms of pancake population:

- “Triangular”, derived from a regular triangle in the xy plane (with the 4th vertex taken at the mesocentre of the triangle);
- “Square”, derived from a square in the xy plane.

In both cases, the three coordinates of each vertex are perturbed by a random amount uniformly distributed in the range $\pm(D) \times 20\%$.

Knife Blade. This population contains three components:

- “Long Triangular” derived from the Triangular Pancake scaled in the x direction by a random factor in a such way to obtain the distribution shown in Figure 13.3;

- “Long Rectangular”, derived from the Square Pancake scaled in the x direction by a random factor.
- a “Long Diamond”, defined from a regular plane diamond in the xy plane, where the position of each vertex is perturbed by a random noise with an amplitude of $\langle D \rangle \times 15\%$ in a random direction for each cartesian component, and then scaled in the x direction by a random factor.

Potatoes. This population is derived from the Pseudo-Sphere type by elongation in both the x and z directions by different random factors in a such way to obtain the distribution shown in Figure 13.3.

After computation of the nine populations of tetrahedra defined above, the coordinates of each tetrahedron are computed with respect to its new (after perturbation of the vertices) mesocentre coordinate system. Then each tetrahedron is scaled so as to have the same mean inter-spacecraft distance $\langle D \rangle$; the value has been arbitrarily fixed at 1000 km. Finally, to randomise the spatial orientation of the tetrahedra, essential if we want to study the role of the tetrahedron direction, each tetrahedron is “shaken” in all directions, via three successive plane rotations, where the three rotation angles θ , ϕ , β are uniform random values.

To produce Figure 13.3, we have used a different number of tetrahedra for each type, as follows:

Regular	=	10
Pseudo-Sphere	=	200
Pancake (Triangular)	=	100
Pancake (Square)	=	100
Knife Blades (Long Triangle)	=	70
Knife Blades (Long Rectangle)	=	70
Knife Blades (Long Diamond)	=	70
Cigars	=	200
Potatoes	=	150

These numbers are chosen so that all five basic types contain about the same number of tetrahedra (200 for Pseudo-Sphere, 200 for Pancake, 210 for Knife Blades, 200 for Cigars, 150 for Potatoes), except the perfectly regular (10). It is worth noting that the cigar-type tetrahedra are largely over-represented; this must be taken into account in the simulations. In fact, there is no *a priori* “uniform” distribution for the shapes of the tetrahedra; any distribution which occurs in practice will be the result of a deliberate choice of orbital parameters for the spacecraft concerned.

13.4.3 The 1-D Geometric Factors and the Types of Tetrahedra

To study how the main 1-D geometric factors behave for each the five types of tetrahedra, we use the 5-types tetrahedra reservoir defined in Section 13.4.1. For each tetrahedron of the five types studied, we have computed the main geometric factors (Q_{GM} , Q_{RR} , Q_{SR} , and Q_{R8}), together with the E and P parameters. The results are given in Figure 13.4, where the x axis is the cumulative number of tetrahedra in each types. The total number of

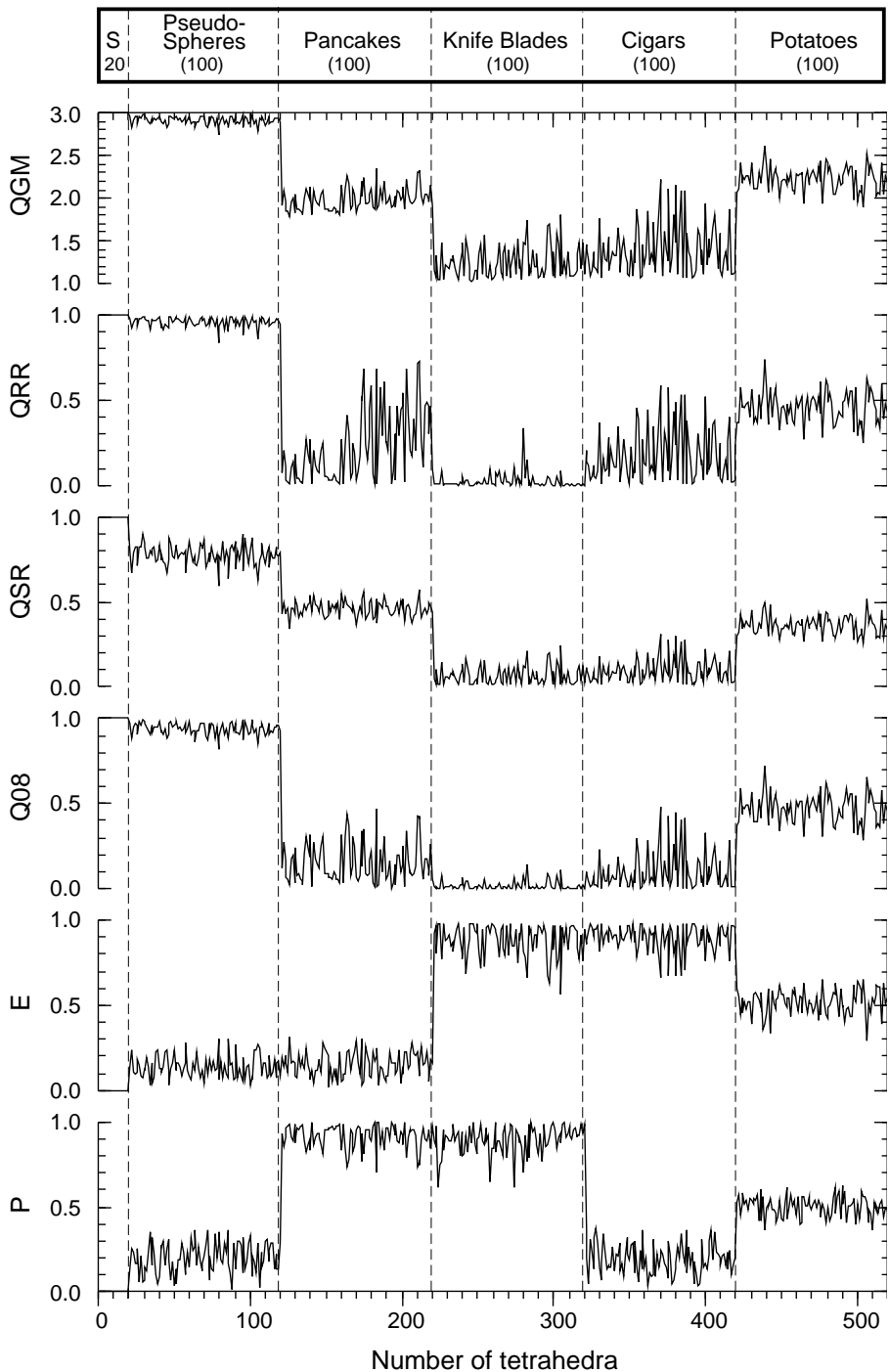


Figure 13.4: Plot of 4 1-D geometric factors and E , P parameters versus the 5 types of tetrahedron defined in Figure 13.3.

tetrahedra in each type is also indicated at the top of the figure. Note that the Q_{GM} factor varies in the (1–3) range, and the others parameters varies in the (0–1) range.

The small “S” category is the perfectly regular tetrahedron (the true sphere type). This type has been added, with a low number of tetrahedra, to verify that each 1-D parameter give the maximal value of its range (3 for Q_{GM} , 1 for the other parameters, except of course E and P).

The Pseudo-Spheres type tetrahedra gives the expected result: the Q_{GM} , Q_{RR} , and the Q_{R8} factors gives effectively a value very close to 3 for Q_{GM} and 1 for the others. Only the Q_{SR} parameter gives a value about 0.8, meaning that this kind of factor is very sensitive to the low change of a regular tetrahedron. The E and P parameters give the most real indication of the shape, in contrast to the 1-D parameters alone, and show, of course, that the elongation and planarity have low values already visible on Figure 13.3. As we will see, The E and P parameters will be obviously in accordance with the other types. Nevertheless, one should not forget that E and P alone are not sufficient to describe entirely the exact shape (see Chapter 12). They are only 2-D geometric parameters, and can indicate the main characteristics of a given tetrahedron; however, very much better that a single 1-D parameter.

The Pancake type tetrahedra gives a more complex result. The Q_{GM} factor gives the expected result, a value very close to 2, with a small variance. The Q_{SR} factor gives a value near 0.5. On the other hand, the Q_{RR} factor gives a value which varies from 0 to 0.7, the 0–0.4 part corresponding with the subtype of the triangular pancakes, and the 0.4–0.7 part corresponding to the subtype of square pancakes. The behaviour of the Q_{R8} factor is the same, but reaches only a value of 0.5. For these last two factors, it is not surprising, since these parameters are being computed from the circumscribing sphere, or for the volume of the tetrahedron, so a flat or a line tetrahedron leads to an infinite radius or a zero volume, and then a factor value near zero. Once again, one can separate the geometric factors giving information on the shape, such as Q_{GM} or Q_{SR} , from those such as Q_{RR} or Q_{R8} giving information on the accuracy of the measurement, as we will see in the next section and in Chapter 16.

The Knife Blades type gives also the broadly expected result. The Q_{RR} and the Q_{R8} geometric factors give a value very near zero, while the Q_{GM} factor, and, in a minor part, the Q_{SR} factor give a value near zero (1 for Q_{GM}) but with a rather high variance, these factors being probably more sensitive to the difference from an absolutely long and plane tetrahedron.

The Cigars type gives result which may be surprising, but can be easily understood. The Q_{GM} factor, in fact, does not make a large distinction between a cigar or a knife blade, because in the two cases the tetrahedron is long, and then give a “fractional dimension” in the range (1–2), with a high proportion close to 1. It is the same case for the Q_{SR} factor, which yields very similar result with the cigars type and the knife blade type, with a lower variance. This phenomena will be studied in details on the next section.

The “Potatoes type” gives the expected results, since the potato is an undefined shape, between the other well identified types, and gives values about 2.3 for Q_{GM} factor (0.65 if we normalise Q_{GM} in the 0–1 range) and about 0.5 for the others.

In conclusion, the study of the 1-D geometric factors with the 5-types of tetrahedra is limited, since the results are not surprising, although this kind of study allows us to be precise about the behaviour of these 1-D geometric factors with characteristic tetrahedra.

13.5 Study of the 1-D Geometric Factors with the E-P Parameters

The method used to study the meaning of the 1-D geometric factors is to plot the value of these parameters on a E - P diagram. To do that, we need a “homogeneous tetrahedron reservoir”, whose the E and P values must cover all the E - P plane. Then, for each of these tetrahedra, the values of the Q_{GM} , Q_{RR} , Q_{SR} , and Q_{R8} geometric factors are plotted on the E - P diagram. This highlights the significance and limitations of this kind of 1-D parameter.

13.5.1 Computing an Homogeneous Tetrahedra Reservoir

The method used to compute an homogeneous tetrahedra reservoir is explained below. Firstly, we take tetrahedra corresponding to the 9 basic forms used in Section 13.4.2. For each basic form, N_i tetrahedra are chosen in category i , we perturb (always in the mesocentre coordinate system) each vertex of the tetrahedra by a random noise with an amplitude of $\langle D \rangle \times 10\%$. Then, we define a grid of 0.1 steps in the E - P plane and decompose the E - P plane in 100 regular squares of 0.1 unit for each side. We also compute the E and P parameters for each tetrahedron, and determine the corresponding square in the E - P plane. Secondly, we begin again this process as many time as it is necessary (with a maximum of 10 times) so that each square of the E - P plane contains about 10 tetrahedra. Finally, to avoid a bias, all the tetrahedra have the same mean inter-spacecraft distance $\langle D \rangle$, arbitrarily fixed at 1000 km. The final result is a reservoir of about 1000 tetrahedra (ten per regular square of the 10×10 grid).

Figure 13.5 shows the result, and we can see that there is indeed an homogeneous distribution of representative points in the E - P plane. To make this reservoir, we used numbers of tetrahedra, deduced from each basic forms, as follows:

Regular	=	10
Pseudo-Spheres	=	300
Pancakes (Triangular)	=	150
Pancakes (Square)	=	150
Knife Blades (Long Triangle)	=	150
Knife Blades (Long Rectangle)	=	150
Knife Blades (Long Diamond)	=	150
Cigars	=	300
Potatoes	=	300

13.5.2 Cluster Orbit Tetrahedron in a Time Diagram

The more usual representation of the orbit of 4 spacecraft is a plot of the position of each spacecraft, and many other parameters, versus time. In Figure 13.6, we have plotted, for a typical Cluster orbit, and over one orbit, from top to bottom:

- the four geocentric distances,

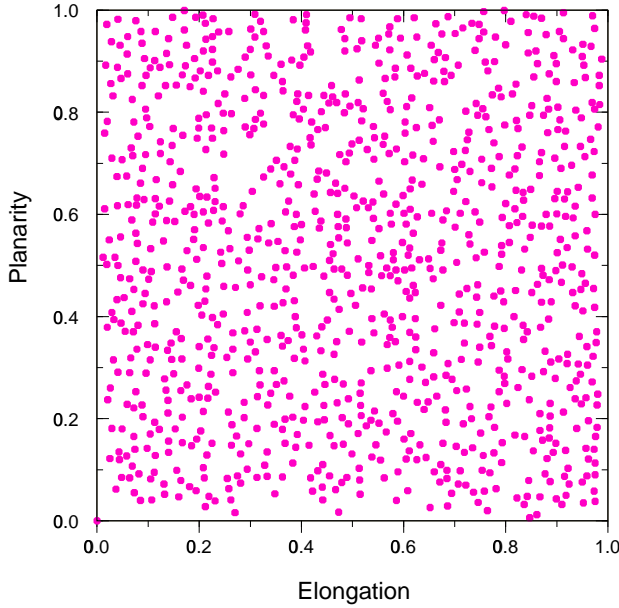


Figure 13.5: The tetrahedra reservoir used in the paper. Notice the homogeneous coverage of the E - P plane.

- the six inter-spacecraft distances,
- the volume of tetrahedron,
- the 3 geometric factors Q'_{GM} , Q_{RR} , Q_{SR} (note that Q'_{GM} factor is equal to $(Q_{GM} - 1)/2$ to have the same range (0–1) that the others parameters),
- the 3 semiaxes of the ellipsoid, a , b , c ,
- the E and P parameters.

The orbit has been given by ESA [Schænmackers, *private communication*], and was established initially on the basis of a launch in November 1995. Although the launch and the Cluster mission are delayed (failure launch of Ariane 501), the arguments remain the same. Regarding the geocentric distance, the spacecraft seem close to each other, and the four geocentric distances are superposed on the figure. One can see however that the inter-spacecraft distances vary in high proportions, and thus the shape of the tetrahedron has a strong variation along one orbit. In particular, the volume of the tetrahedron can reach a value very close to zero twice (at 26:30 UT and 33:10). This explains that, as we can see in Figure 13.6, the Q_{RR} geometric factor is very close to zero, and the Q'_{GM} and Q_{SR} reach a low value, as we have seen in preceding section. Since the minor semiaxis c of the ellipsoid is also equal to zero at these points and the middle semiaxis b has a non-zero value, the tetrahedra is fully flat (planarity parameter equal to 1, and elongation takes any value). Two other particular points can be observed, namely at 18:30 and 38:45, when $b = a$, and thus $E = 0$, corresponding to a regular sphere “flattened” in a single

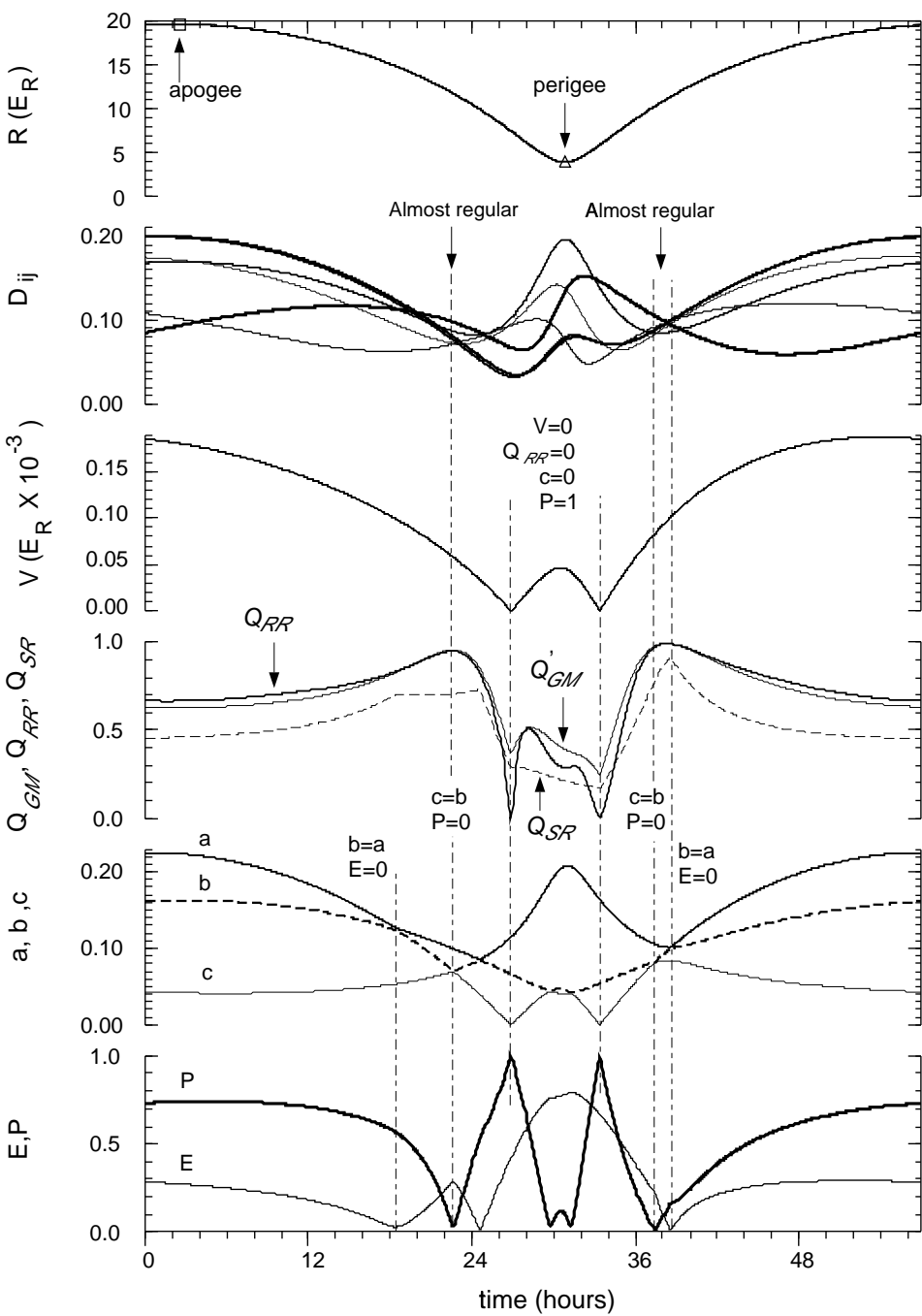


Figure 13.6: The main Cluster orbit parameters, and the 1-D and 2-D associated geometric factors, for a typical orbit of December 24, 1995 (data provided by ESA).

direction. Finally, Schoenmækers has shown that we can have two points in the orbit where the tetrahedron is regular. We find the first point at 22:30 where $P = 0$ and E have a low value. The second point is in fact a small duration in the interval [37:30–38:45] where $P = 0$ and E has a low value and then, after a short time, where P has a low value and $E = 0$. In both cases, these points can be rapidly found by examining Q'_{GM} and Q_{RR} which reach the maximum value of 1, corresponding to a regular tetrahedron. As we have seen, the Q_{SR} value has a maximum value less than 1, this parameter being very sensitive to the difference to a perfect tetrahedron.

This kind of figure can give a good indication on the shape of the tetrahedron during the orbit of the for spacecraft.

13.5.3 Cluster Orbit Tetrahedron in the E - P Diagram

In order to characterise quickly the shape of the Cluster tetrahedron along one orbit, rather than plotting the tetrahedron characteristics with time as done in previous section, another way is to plot an hodogram of the successive positions of the tetrahedron in a E - P diagram. In Figures 13.7 and 13.8, the shape of the Cluster tetrahedron is computed and plotted in the E - P diagram along a whole orbit. The time step is 6 minutes, and the arrow indicates the direction of the motion. The apogee corresponds to the portion of the figure where the different points are very close together, the velocity being low and the shape slowly varying. The perigee corresponds to the portion of the figure where the points are widely spaced, because the spacecraft velocity along the average trajectory is large.

In Figure 13.7 (December 24, 1995), as we have seen in the preceding section, the tetrahedron is regular at 2 points along the orbit, the first point being located near $(E, P) = (0.28, 0.01)$, and the second point is in fact a short period, from $(E, P) = (0.21, 0.01)$ to $(E, P) = (0.01, 0.16)$. These two points where the tetrahedron is regular are, of course, located in the region of the Pseudo-Spheres type (see Figure 13.4). During the rest of the curve, the E - P parameters can take extreme values. In particular, as we have seen before, the tetrahedron is absolutely flat ($P=0.99$) for 2 points along the orbit, but never completely linear (the maximum value of E is 0.8 near the perigee). For another example orbit shown in Figure 13.8 (June 24, 1996) the conclusions remain the same. During the course of the Cluster mission, all possible shapes of tetrahedra are expected, and thus, simulations must take into account any possible value in the E - P plane. Thus, the homogeneous reservoir will be used for the following 2-D simulations.

13.5.4 E - P Diagram for 1-D Geometric Factors

The idea is the same as that in the previous section on the shape of the tetrahedron along the orbit. In Section 13.4.3, we have studied the geometric factors among the 5 types of tetrahedra defined in 13.4.1. To have a more precise idea of what the different 1-D geometric factors studied mean, we have used the homogeneous reservoir of tetrahedron defined in Section 13.5.1 to compute the values of these 1-D parameters in the E - P plane. This presentation has an important advantage: by examination of the values of these 1-D geometrical parameters in the E - P diagram, we can directly correlate the value of the 1-D geometric factors to the shape of the tetrahedron which is very best defined by the E and P parameters, although, as we have already say, E and P are themselves an approximation of the exact shape.

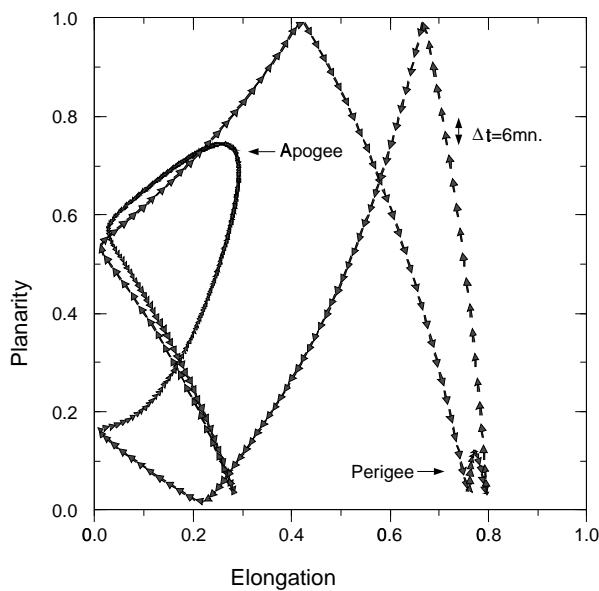


Figure 13.7: Evolution of the shape of the Cluster tetrahedron along its trajectory in a *E-P* diagram for December 24, 1995 (data provided by ESA).

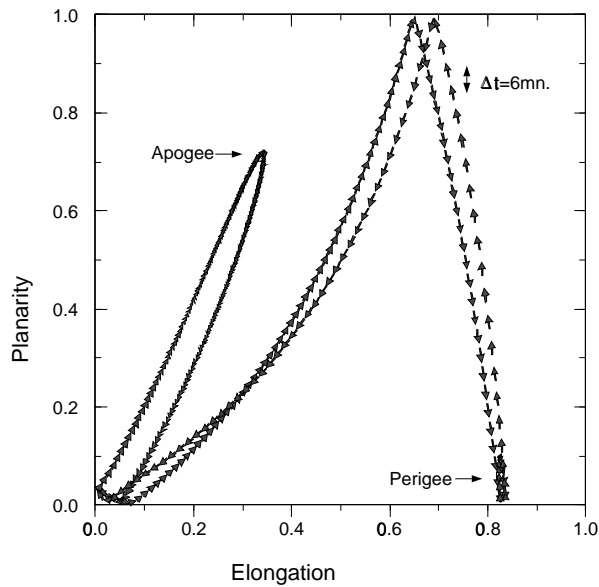


Figure 13.8: Evolution of the shape of the Cluster tetrahedron along its trajectory in a *E-P* diagram for June 24, 1996 (data provided by ESA).

The Q_{GM} and Q_{RR} Geometric Factors

The results are shown in Figures 13.9 and 13.10, where we have plotted in the E - P plane the values of the Q_{GM} geometric factor (Figure 13.9) and the Q_{RR} geometric factor (Figure 13.10). The size and the colour of the circles corresponds to the values of the geometric factors according to the legend given vertically on the right of each figure. At a first glance, there is an important difference between the distribution of the values of these factors in the E - P plane. Near the origin, for the low values of E and P ("Pseudo-Spheres type"), there is a similar behaviour of the two geometric factors, but, for high values of E and P , we have a difference. From the Q_{GM} factor, we can see an illustration of the fact that the P parameter becomes undefined when E is near 1 (see Section 13.3.3). In Section 13.4.3 we interpreted the meaning of Q_{GM} as the "fractional dimension" of the tetrahedron, but this kind of diagram reveals a fundamental question: does the fractional dimension exist? If we cover the sides of the E - P plane from the (0, 0) origin in the clockwise direction and having in mind the Figure 13.3 describing the five types of tetrahedron, from the "Pseudo-Spheres" type to the "Pancake" type, and then to the "Knife Blades type" and the "Cigars type", the Q_{GM} value varies from 3 to 2 and then to 1 for these 4 types, corresponding to the concept of a "fractional dimension". There is no difference between a "Knife Blade" and a "Cigar", both being considered as a line shape of dimension near 1. If we consider however the transition between the "Cigars type" ($D=1$) and the "Pseudo-Spheres type" ($D=3$) to finish the clockwise tour, we reach now a fundamental problem about the "fractional dimension". These two shapes are in fact very similar since the Cigars are deduced from the Pseudo-Spheres by a strong elongation in an arbitrary direction (13.4.2), and the transition between these two shapes from dimension $D=1$ to dimension $D=3$ has to pass by the value of $D=2$ which, in this case, does not correspond to a plane because the planarity P is near zero. In others words, a value of Q_{GM} equal to 2 does not imply a flat tetrahedron; it could also correspond to a rather long cigar with a rounded section. On the other hand, the fact that the Q_{GM} factor does not distinguish between Knife Blade and Cigars (both being considered as a long tetrahedron) cannot be essential, because this distinction becomes impossible near $E = 1$. In conclusion, the Q_{GM} geometric factor remains a good alternative to describe, in the strong limit of a single 1-D parameter, the geometrical shape of a tetrahedron, particularly in the extreme "pancake" region, although the concept of fractional dimension must be taken with care.

Concerning the Q_{RR} geometric factor (Figure 13.10), the result is fully different for the high values of E and P . The isovalues of this factor (not plotted here, but easily guessed) are roughly decreasing with the radius $r = \sqrt{E^2 + P^2}$. This factor is not directly connected to the geometric shape of the tetrahedron, because a Pancake type tetrahedron, a Knife Blade type, and a Cigar type lead approximately to the same value for Q_{RR} . Nevertheless, this kind of parameter is rather well connected to the relative error measurement of physical parameters such as $\nabla \times \mathbf{B}$ for which a regular tetrahedron is often the best shape to minimise the measurement errors (at least for isotropic signature, see also Chapter 16). This property is easily explained by examining Figure 13.10. In fact, this parameter has minimum values near $E = 1$ and $P = 1$, and particularly in the region where we have simultaneously E and P close to 1 (Knife Blades). Thus the Q_{RR} factor can be seen as an expression of the *degeneration* of the tetrahedron (i.e., when E or P are close to 1), and so can be used as a real geometric factor for the physical determination of scientific

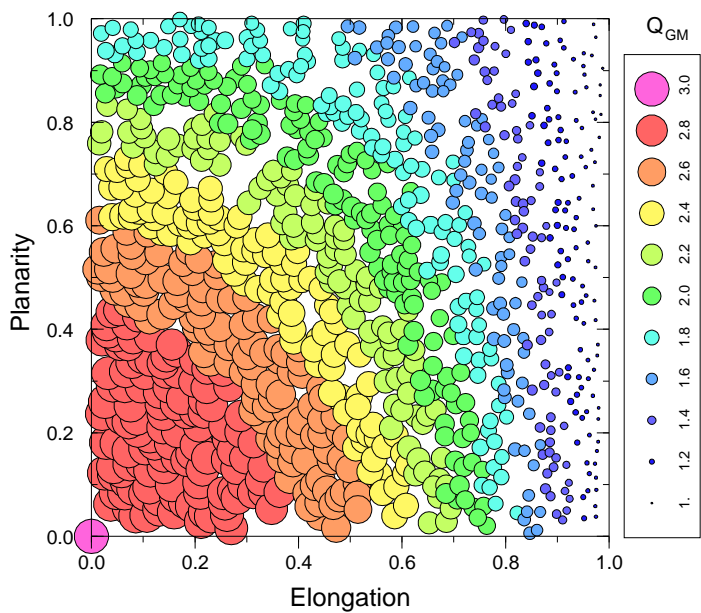


Figure 13.9: Plot in the E - P diagram of the Q_{GM} geometric factor.

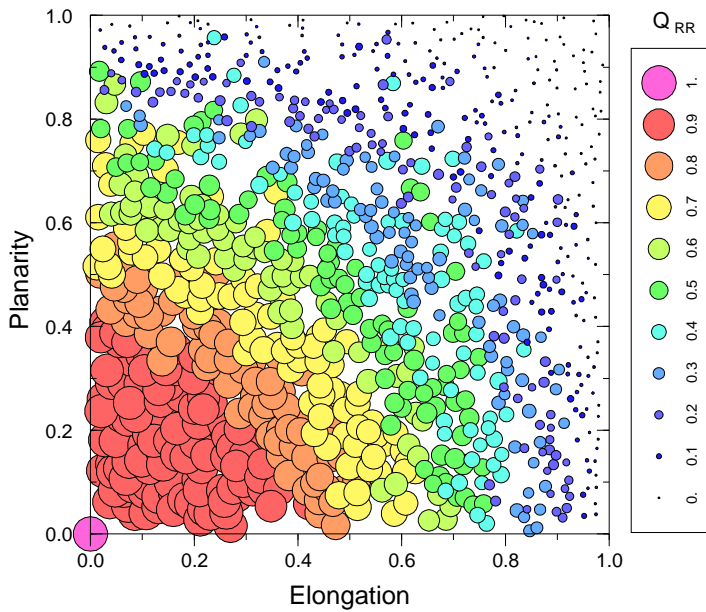


Figure 13.10: Plot in the E - P diagram of Q_{RR} geometric factor.

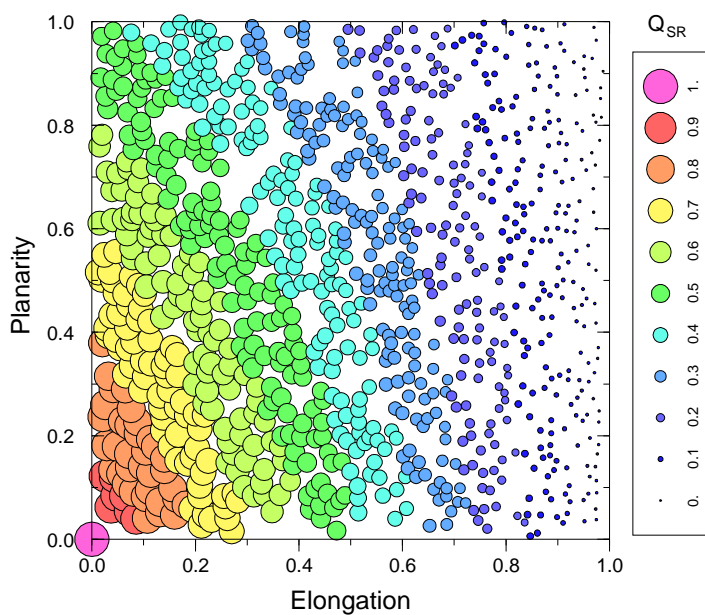


Figure 13.11: Plot in the E - P diagram of Q_{SR} geometric factor.

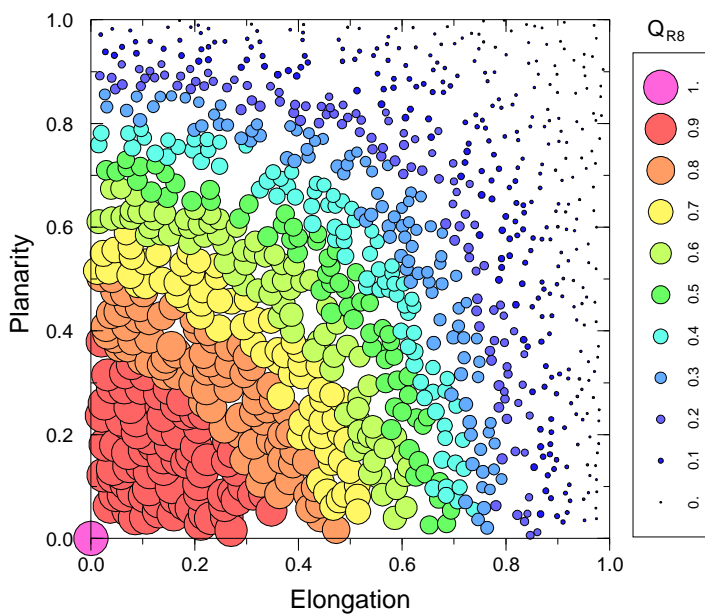


Figure 13.12: Plot in the E - P diagram of Q_{R8} geometric factor.

parameters which prefer a regular tetrahedron.

Nevertheless, for further studies, we can also define directly the degeneration of a tetrahedron as, for example, $d = \sqrt{E^2 + P^2}$, or something of the same kind.

The Q_{SR} and Q_{R8} Geometric Factors

Many 1-D geometric factors have been studied by examination of their values in the E - P diagram. In fact, we can find a lot of factors for which the E - P diagram looks the same as the two “preferred” Q_{GM} and Q_{RR} geometric factors. We present here only the Q_{SR} and the Q_{R8} geometric factors defined in Section 13.3.1 and already studied in Section 13.4.3.

The E - P diagram for the Q_{SR} factor is very similar to that of the Q_{GM} factor, as we can see in Figure 13.11. Nevertheless, there is a difference near the low values of E and P (see Section 13.4.3), where the Q_{SR} factor decrease very rapidly as soon as the E or P values are not close to zero, thus confirming the “sensitivity” of this geometric factor to a small deviation from a perfectly regular tetrahedron. Except for this difference, however, the main conclusion is the same as for the Q_{GM} geometric factor.

Concerning the Q_{R8} geometric factor (see Figure 13.12), apart from a much smoother transition, the E - P diagram of this geometric factor is very similar to the Q_{RR} one, confirming the fact that the normalised volume is a good indicator of the degree of “degeneration” of a tetrahedron.

13.6 Conclusions

The pseudo-ellipsoid, derived from the volumetric tensor of in Section 12.4.1, provides a useful and simple approach to characterise the shape of a tetrahedron, and its orientation in space. The E and P parameters allow an appropriate and easy-to-use description of this shape, and has been used to define 5 main types of tetrahedra: “Pseudo-Spheres”, “Pancakes”, “Knife Blades”, “Cigars”, and “Potatoes”. These E and P parameters are used to define a 2-D geometric factor, which is a very efficient way to describe the shape and the deviation to a regular tetrahedron rather than a single 1-D geometric factor, even if it is as best as possible.

The definition of the 5 types of tetrahedron, and the making of a corresponding “reservoir of five type”, has allowed us to study the response of the main 1-D geometric factors with respect to each type of tetrahedron.

On the other hand, the evolution of the shape of the tetrahedron along a typical Cluster orbit has been studied in a time diagram. By considering the different 1-D geometric factors, the length of the axes of the pseudo-ellipsoid, and the E and P parameters, we can obtain a good description of the evolution of this shape. But the introduction of the E - P diagram to plot, for instance, this orbit can give good information directly on the distortion of the tetrahedron, and its evolution. Notice that in a real case, such as the Cluster orbit, the E - P plane is well covered, and so all the values of E and P must be taken into account in any simulation.

The making of an homogeneous reservoir of tetrahedra in the E - P plane allows us to check the validity, meaning, and limits of the main 1-D geometric factors. Factors such as the Q_{GM} or Q_{SR} factors yield information on the geometrical shape, but, of course,

incompletely because of the strong limitation in single scalar values. On the other hand, the E - P diagram is not compatible with the notion of “fractional dimension” which remains an interesting concept but which has to be precisely defined. Other factors, such as the Q_{RR} factors and others, do not give real or direct information on the geometrical shape, but can be considered as the degree of degeneration of the tetrahedron and so are well related to the uncertainties in the determination of some physical parameters which prefer a regular tetrahedron.

To conclude, this study is based on the idea that a regular tetrahedron is the ideal form for good geometric measurements, but we do not forget that for special studies (for example, a boundary crossing) an alignment of the four points can be considered as the best form. Furthermore, for actual sampling of phenomena, we need to identify the relative scale and the orientation in space of the tetrahedron, which requires not only the knowledge of the length of the axes of the ellipsoid, but also their directions. When this information is unknown or is unimportant (as for an isotropic structure), since a single 1-D parameter is not sufficient to describe in a single scalar value the real shape of the tetrahedron, the use of a 2-D factor such as the E - P plane remains essential.

Appendix

13.A Calculation of Geometric Factors Q_{GM} and Q_{RR}

To calculate the geometric factors of equations 13.1 and 13.2, we need to study the geometrical properties of a tetrahedron. We consider the tetrahedron defined by four points in space numbered 1 to 4, with position vectors $\mathbf{r}_1, \mathbf{r}_2, \mathbf{r}_3, \mathbf{r}_4$. Without any loss of generality, we may consider only the differences $\mathbf{d}_\alpha = \mathbf{r}_\alpha - \mathbf{r}_4$ in describing the points.

Area of the Sides

The area of a parallelogram bounded by two vectors \mathbf{d}_1 and \mathbf{d}_2 is given by the magnitude of their cross product; any triangle is half of a parallelogram, so its area is

$$S = \frac{1}{2} |\mathbf{d}_1 \times \mathbf{d}_2|$$

where \mathbf{d}_1 and \mathbf{d}_2 are the vectors for any two sides of the triangle.

We specify side α of the tetrahedron to be the one opposite vertex α : that is, it does not contain the point α .

$$S_1 = \frac{1}{2} |\mathbf{d}_2 \times \mathbf{d}_3| \quad , \quad S_2 = \frac{1}{2} |\mathbf{d}_1 \times \mathbf{d}_3| \quad , \quad S_3 = \frac{1}{2} |\mathbf{d}_1 \times \mathbf{d}_2|$$

$$S_4 = \frac{1}{2} |(\mathbf{d}_2 - \mathbf{d}_1) \times (\mathbf{d}_3 - \mathbf{d}_1)| = \frac{1}{2} |\mathbf{d}_1 \times \mathbf{d}_2 + \mathbf{d}_2 \times \mathbf{d}_3 + \mathbf{d}_3 \times \mathbf{d}_1| \quad (13.10)$$

The total surface S is the sum $\sum_{\alpha=1}^4 S_\alpha$.

Volume

The volume of a parallelepiped defined by three vectors in space is the triple product of those vectors. Any tetrahedron is 1/6 of such a figure, hence

$$V = \frac{1}{6} |\mathbf{d}_1 \cdot \mathbf{d}_2 \times \mathbf{d}_3| = \frac{1}{6} \begin{vmatrix} d_{1x} & d_{1y} & d_{1z} \\ d_{2x} & d_{2y} & d_{2z} \\ d_{3x} & d_{3y} & d_{3z} \end{vmatrix} \quad (13.11)$$

Centre of the Circumscribing Sphere

To find the circumscribed sphere, we need the point that is equidistant from all four vertices, i.e., we want \mathbf{r} such that

$$(\mathbf{r} - \mathbf{r}_\alpha) \cdot (\mathbf{r} - \mathbf{r}_\alpha) = r^2 - 2\mathbf{r} \cdot \mathbf{r}_\alpha + |\mathbf{r}_\alpha|^2 = \rho^2; \quad \forall \alpha = 1, 4$$

If we take point 4 as the origin, that is, if we use the \mathbf{d}_α vectors in place of the \mathbf{r}_α , then $r^2 = \rho^2$, the sphere radius, and this equation reduces to

$$2\mathbf{r} \cdot \mathbf{d}_\alpha = |\mathbf{d}_\alpha|^2 \quad \forall \alpha = 1, 3$$

This matrix equation for the centre of the sphere can be solved for the vector \mathbf{r} and the radius of the sphere $\rho^2 = |\mathbf{r}|^2$. Note that the matrix $\{\mathbf{d}_\alpha\}$ in this equation is the same as the one whose determinant yields the volume of the tetrahedron (equation 13.11). The volume of the circumscribed sphere is then

$$V_o = \frac{4}{3}\pi\rho^3 \quad (13.12)$$

The Regular Tetrahedron

The regular tetrahedron of unit side is the ideal against which the true figure of the four spacecraft is to be measured. We may take (Figure 13.13)

$$\begin{aligned} \mathbf{d}_1 &= (1, 0, 0) \\ \mathbf{d}_2 &= \left(\frac{1}{2}, \frac{\sqrt{3}}{2}, 0\right) \\ \mathbf{d}_3 &= \left(\frac{1}{2}, \frac{\sqrt{3}}{6}, \frac{\sqrt{6}}{3}\right) \\ \mathbf{d}_4 &= (0, 0, 0) \end{aligned}$$

Values for the regular tetrahedron of unit side length are listed in Table 13.1.

The Geometric Factors Q_{GM} and Q_{RR}

From the above quantities, it is easy to calculate Q_{GM} and Q_{RR} .

For Q_{GM} , we average the 6 distances between the 4 points to get the side L of the “ideal” regular tetrahedron, with volume $V_{\text{ideal}} = L^3 \sqrt{2}/12$ and surface $S_{\text{ideal}} =$

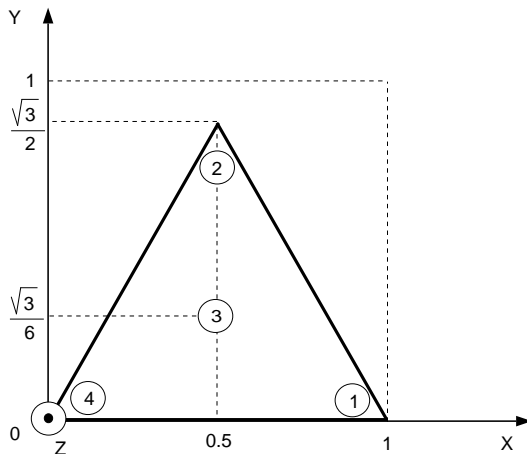


Figure 13.13: Conventions used to define a regular tetrahedron (1,2,3,4 correspond to the spacecraft position).

Table 13.1: Values for regular tetrahedron

Quantity	Value
S_{α}	$= \sqrt{3}/4$
S	$= \sqrt{3}$
V	$= \sqrt{2}/12$
ρ	$= \sqrt{6}/4$
V_{\circ}	$= \frac{4}{3}\pi \left(\frac{3}{8}\right)^{\frac{3}{2}}$

$L^2 \sqrt{3}$. The true volume V and true surface S are found from equations 13.11 and 13.10. Then we can express Q_{GM} as:

$$Q_{GM} = \frac{V}{V_{ideal}} + \frac{S}{S_{ideal}} + 1 \tag{13.13}$$

For Q_{RR} , the radius of the circumscribing sphere is calculated from equation 13.12. The actual volume of the sphere need not be calculated, for all the factors just go into the normalisation factor \mathcal{N} .

$$Q_{RR} = \left(\frac{9\sqrt{3}}{8} V\right)^{\frac{1}{3}} \cdot \rho^{-1} \tag{13.14}$$

Bibliography

The question of how to quantify the degree of tetrahedral regularity was first addressed in the context of the Cluster mission by:

vom Stein, R., Glassmeier, K.-H., and Dunlop, M., A Configuration Parameter for the Cluster Satellites, Tech. Rep. 2/1992, Institut für Geophysik und Meteorologie der Technischen Universität Braunschweig, 1992, who introduced the parameter Q_{GM} .

The parameter Q_{RR} was introduced and compared to Q_{GM} by:

Robert, P. and Roux, A., Influence of the shape of the tetrahedron on the accuracy of the estimation of the current density, in *Proc. International Conf. "Spatio-Temporal Analysis for Resolving plasma Turbulence (START)"*, Aussois, 31 Jan.–5 Feb. 1993, ESA WPP-047, pp. 289–293, European Space Agency, Paris, France, 1993.

Twenty-five geometric factors, among them Q_{SR} and Q_{R8} , were defined and compared with respect to their ability to provide a reliable index for the accuracy of the determination of \mathbf{J} by:

Robert, P., Roux, A., and Coeur-Joly, O., Validity of the estimate of the current density along Cluster orbit with simulated magnetic data, in *Proceedings of Cluster Workshops, Braunschweig, 28–30 Sep. 1994, Toulouse, 16–17 Nov. 1994*, ESA SP-371, pp. 229–233, European Space Agency, Paris, France, 1995.

The impact of the tetrahedral shape on the current determination has also been discussed by:

Dunlop, M. W., Southwood, D. J., Glassmeier, K.-H., and Neubauer, F. M., Analysis of multipoint magnetometer data, *Adv. Space Res.*, **8**, (9)273–(9)277, 1988.

Dunlop, M. W., Balogh, A., Southwood, D. J., Elphic, R. C., Glassmeier, K.-H., and Neubauer, F. M., Configurational sensitivity of multipoint magnetic field measurements, in *Proceedings of the International Workshop on "Space Plasma Physics Investigations by Cluster and Regatta"*, Graz, Feb. 20–22, 1990, ESA SP-306, pp. 23–28, European Space Agency, Paris, France, 1990.

Figure 13.1 has been reproduced from:

Balogh, A., Dunlop, M. W., Cowley, S. W. H., Southwood, D. J., Thomlinson, J. G., Glassmeier, K.-H., Musmann, G., Luhr, H., Buchert, S., Acuna, M., Fairfield, D. H., Slavin, J. A., Riedler, W., Schwingenschuh, K., and Kivelson, M. G., The Cluster magnetic field investigation, *Space Sci. Rev.*, **79**, 65–91, 1997.

The "elongation" and "planarity" factors (E , P) were introduced by:

Robert, P., Roux, A., and Chanteur, G., Accuracy of the determination of the current density via four satellites, in *Abstracts, International Union of Geodesy and Geophysics, XXI General Assembly, Boulder, Colorado, 1995*, presentation GAB51H-06.

An early description of the distortion of the spacecraft configuration was presented by:

Dunlop, M. W., Review of the Cluster orbit and separation strategy: Consequence for measurements, in *Proceedings of the International Workshop on "Space Plasma Physics Investigations by Cluster and Regatta"*, Graz, Feb. 20–22, 1990, ESA SP-306, pp. 17–

22, European Space Agency, Paris, France, 1990.

The ellipsoid corresponding to the volumetric tensor was first defined for a tetrahedron by J. Schænmækers of ESOC Flight Dynamics Division, who also provided the Cluster community with a code to compute the lengths and directions of the three axes of the ellipsoid defined by the four spacecraft.

The role of experimental errors in the Cluster context has been treated in some of the above references, and by:

Robert, P. and Roux, A., Accuracy of the estimate of J via multipoint measurements, in *Proceedings of the International Workshop on "Space Plasma Physics Investigations by Cluster and Regatta"*, Graz, Feb. 20–22, 1990, ESA SP-306, pp. 29–35, European Space Agency, Paris, France, 1990.

Dunlop, M. W. and Balogh, A., On the analysis and interpretation of four spacecraft magnetic field measurements in terms of small scale plasma processes, in *Proc. International Conf. "Spatio-Temporal Analysis for Resolving plasma Turbulence (START)"*, Aussois, 31 Jan.–5 Feb. 1993, ESA WPP-047, pp. 223–228, European Space Agency, Paris, France, 1993.

The effect of anisotropic spacecraft configurations in the presence of anisotropic physical structures has been studied for the case of the Earth's magnetosheath by:

Dunlop, M. W., Southwood, D. J., and Balogh, A., The Cluster configuration and the directional dependence of coherence lengths in the magnetosheath, in *Proc. International Conf. "Spatio-Temporal Analysis for Resolving plasma Turbulence (START)"*, Aussois, 31 Jan.–5 Feb. 1993, ESA WPP-047, pp. 295–299, European Space Agency, Paris, France, 1993.

The formulas for the geometric factors Q_{GM} and Q_{RR} given in the Appendix are taken from:

Daly, P. W., The tetrahedron quality factors of CSDS, Tech. Rep. MP Ae-W-100-94-27, Max-Planck-Institut für Aeronomie, D-37191 Katlenburg-Lindau, Germany, 1994, electronically available from <http://www.mpae.gwdg.de/~daly/tetra.html>.

— 14 —

Spatial Interpolation for Four Spacecraft: Theory

GÉRARD CHANTEUR
CETP-CNRS
Vélizy, France

14.1 Introduction

Spatial interpolation of data between the spacecraft and estimation of the effects of tetrahedron shape and attitude are key aspects of data analysis from four spacecraft. This chapter presents the fundamental concepts of barycentric coordinates and demonstrates their use for interpolating data within a tetrahedron. The reciprocal vectors, defined as the gradients of the barycentric coordinates, play a key role in the determination of gradients, the characterisation of discontinuities and the analysis of errors related to the uncertainties on the spacecraft positions. The first application of this formalism to a four spacecraft mission was proposed by *Chanteur and Mottez* [1993]. The method as developed here applies to only four spacecraft; it could likely be generalised to more than four spacecraft at the expense of a greater algebraic complexity, but this development has not been pursued.

Barycentric coordinates are well-known tools in applied mathematics but not in space physics. This is why a self-contained and detailed presentation is given in Section 14.2 together with the introduction of the reciprocal vectors and the demonstration of their fundamental properties. Linear and quadratic interpolation schemes are presented as well as the related estimators of gradients. The practical determination of the gradient of the magnetic field by these and others estimators along realistic orbits in a model magnetospheric field is the object of Chapter 15. Section 14.3 is devoted to the theoretical analysis of physical and geometrical errors. In a first step the expectation values and covariance matrix of the reciprocal vectors are explicitly derived from the uncertainties affecting the spacecraft positions for any configuration of the cluster. Then use is made of this result to predict the expectation values and covariances of the estimated components of the gradient of a vector field and to separate for errors of physical and geometrical origin. Section 14.4 gives a detailed analysis of the errors related to the truncation inherent to any interpolation scheme. To avoid unnecessary mathematical complication the analysis is conducted for the one-dimensional model of a thick and planar current sheet. Section 14.5 presents some other applications of the barycentric formalism, especially the derivation of the spatial aliasing condition for plane waves sampled by a tetrahedron and the characterisation of a planar discontinuity crossed by the cluster of spacecraft, i.e., the determination of the normal components of the velocity and acceleration of the discontinuity in the frame of the cluster. The conclusion section emphasises the usefulness of the barycentric formalism

and its possible applications to other topics not investigated in this chapter.

14.2 Interpolation within a Tetrahedron and Estimation of Gradients

In this and following sections involving the barycentric coordinates, dyadic notations will be used systematically. A column vector is denoted \mathbf{a} , its transpose \mathbf{a}^T being a line vector. The dyad $\mathbf{a}^T \mathbf{b}$ denotes the scalar product of two real vectors \mathbf{a} and \mathbf{b} , usually written $\mathbf{a} \cdot \mathbf{b}$, hence: ⇒14.1

$$\mathbf{a} \cdot \mathbf{b} = \mathbf{a}^T \mathbf{b} = \mathbf{b}^T \mathbf{a}$$

Meanwhile the dyad $\mathbf{a} \mathbf{b}^T$ is a tensor of rank two, the transpose of which is $\mathbf{b} \mathbf{a}^T$. The gradient of the scalar field $u(\mathbf{r})$ is the vector $\mathbf{G}[u]$ defined by the differential:

$$du = d\mathbf{r}^T \mathbf{G}[u] \quad (14.1)$$

Similarly the gradient of the vector field $\mathbf{v}(\mathbf{r})$ is the tensor of rank two $\mathbf{G}[\mathbf{v}]$ defined by the differential:

$$d\mathbf{v}^T = d\mathbf{r}^T \mathbf{G}[\mathbf{v}] \quad (14.2)$$

In cartesian coordinates this definition gives the usual definition for component kl of $\mathbf{G}[\mathbf{v}]$: $(\mathbf{G}[\mathbf{v}])_{kl} = \partial_k v_l$, partial derivative of component v_l with respect to coordinate x_k .

14.2.1 Linear Interpolation and Related Estimators

Definitions of the Barycentric Coordinates and Reciprocal Vectors

Let S_α ($\alpha = 1, 2, 3, 4$) be the vertices of an irregular tetrahedron, or cluster, and \mathbf{r}_α be their position vectors. A naturally occurring scalar field $u(\mathbf{r})$ or a vector field $\mathbf{v}(\mathbf{r})$ is known only by the values $u_\alpha = u(\mathbf{r}_\alpha)$ or $\mathbf{v}_\alpha = \mathbf{v}(\mathbf{r}_\alpha)$ measured at the vertices S_α . The linear interpolation of these values in the vicinity of the cluster requires four basic interpolating functions μ_α which are linear scalar functions of the position vector \mathbf{r} , and which satisfy the constraints $\mu_\alpha(\mathbf{r}_\beta) = \delta_{\alpha\beta}$ (Kronecker symbol). Let the linearly interpolated fields, denoted by $L[u]$ and $L[\mathbf{v}]$ be given by the following expressions:

$$L[u](\mathbf{r}) = \sum_{\alpha=1}^4 u_\alpha \mu_\alpha(\mathbf{r}) \quad (14.3)$$

$$L[\mathbf{v}](\mathbf{r}) = \sum_{\alpha=1}^4 \mathbf{v}_\alpha \mu_\alpha(\mathbf{r}) \quad (14.4)$$

The four values $\mu_\alpha(\mathbf{r})$, $1 \leq \alpha \leq 4$ are the barycentric coordinates of the point \mathbf{r} . They can be expressed $\mu_\alpha(\mathbf{r}) = v_\alpha + \mathbf{k}_\alpha \cdot \mathbf{r}$ where the v_α and \mathbf{k}_α are respectively scalar and vector constants to be determined. Using the constraints $\mu_\alpha(\mathbf{r}_\beta) = \delta_{\alpha\beta}$ we deduce that:

$$\mu_\alpha(\mathbf{r}) = 1 + \mathbf{k}_\alpha \cdot (\mathbf{r} - \mathbf{r}_\alpha) \quad (14.5)$$

$$\mathbf{k}_\alpha \cdot (\mathbf{r}_\beta - \mathbf{r}_\gamma) = \delta_{\alpha\beta} - \delta_{\alpha\gamma} \quad (14.6)$$

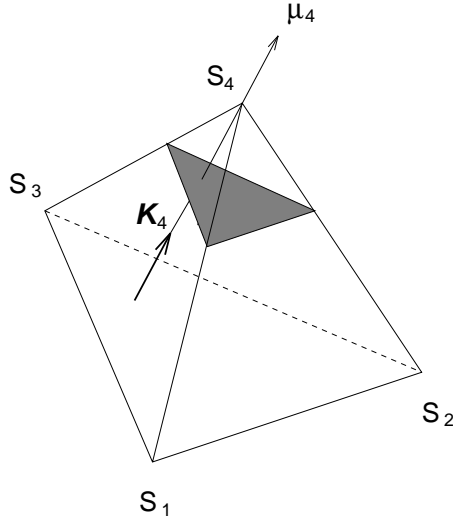


Figure 14.1: Barycentric vectors and coordinates. The shaded plane, perpendicular to the axis μ_α , is an isovalue surface of μ_α .

Equation 14.6 shows that \mathbf{k}_α is normal to Π_α , the face of the tetrahedron opposite to S_α . Hence, for example, by making use of $\mathbf{r}_{\beta\gamma} = \mathbf{r}_\gamma - \mathbf{r}_\beta$, \mathbf{k}_4 can be written:

$$\mathbf{k}_4 = \frac{\mathbf{r}_{12} \times \mathbf{r}_{13}}{\mathbf{r}_{14} \cdot (\mathbf{r}_{12} \times \mathbf{r}_{13})} \quad (14.7)$$

it appears to be proportional to the area of the face of the tetrahedron opposite to vertex S_4 and inversely proportional to the volume of the tetrahedron. Expressions for the other reciprocal vectors are obtained through cyclic permutations of the indices. We call \mathbf{k}_α the reciprocal vectors of the tetrahedron. From equation 14.5, they satisfy:

$$\mathbf{k}_\alpha = \mathbf{G}[\mu_\alpha] \quad (14.8)$$

Properties of the Barycentric Coordinates and Reciprocal Vectors

From a geometrical point of view, the barycentric coordinates have the following properties (see Figure 14.1):

1. $\mu_\alpha(\mathbf{r})$ is constant in a plane parallel to Π_α , the face of the tetrahedron opposite to S_α .
2. $\mu_\alpha(\mathbf{r}) < 0$ in the half space, relative to Π_α , not containing S_α .
3. $\mu_\alpha(\mathbf{r}) = 0$ for all points lying in Π_α .
4. $\mu_\alpha(\mathbf{r}) > 0$ in the half space, relative to Π_α , containing S_α .
5. $0 < \mu_\alpha(\mathbf{r}) < 1$ for all points lying inside the tetrahedron

It is worth mentioning the following properties of the barycentric coordinates and reciprocal vectors due to their importance in numerous topics that can be addressed by a four spacecraft mission. Firstly, from the definitions themselves it is obvious that $\mu_\alpha(\mathbf{r})$ and \mathbf{k}_α are independent of the origin of the position vectors. Secondly, the linear interpolation of a constant scalar field gives straightforwardly:

$$\sum_{\alpha=1}^4 \mu_\alpha(\mathbf{r}) = 1 \quad (14.9)$$

Taking the gradient of both sides of equation 14.9 leads to:

$$\sum_{\alpha=1}^4 \mathbf{k}_\alpha = \mathbf{0} \quad (14.10)$$

Thirdly, \mathbf{r} itself is a linear field, hence:

$$\mathbf{r} = L[\mathbf{r}] = \sum_{\alpha=1}^4 \mathbf{r}_\alpha \mu_\alpha(\mathbf{r})$$

According to definitions 14.2 and 14.8, taking the gradient of this equation gives:

$$\sum_{\alpha=1}^4 \mathbf{k}_\alpha \mathbf{r}_\alpha^T = \mathbf{I} = \sum_{\alpha=1}^4 \mathbf{r}_\alpha \mathbf{k}_\alpha^T \quad (14.11)$$

where \mathbf{I} is the unit tensor defined by $I_{ij} = \delta_{ij}$. Equation 14.11 contains the two following special relations:

$$\sum_{\alpha=1}^4 \mathbf{k}_\alpha \cdot \mathbf{r}_\alpha = 3 \quad \text{and} \quad \sum_{\alpha=1}^4 \mathbf{k}_\alpha \times \mathbf{r}_\alpha = \mathbf{0}$$

Another interesting property is obtained by taking the gradient of $\mathbf{r}^T \mathbf{A}$, where \mathbf{A} is a constant vector:

$$\mathbf{A} = \sum_{\alpha=1}^4 (\mathbf{r}_\alpha \cdot \mathbf{A}) \mathbf{k}_\alpha \quad (14.12)$$

Linear Estimators of Gradients

The gradients of the linearly interpolated fields $L[u]$ and $L[v]$ may be used as estimators of the gradients of the sampled physical fields u and v .

$$\mathbf{L}G[u] = \mathbf{G}[L[u]] \quad \text{and} \quad \mathbf{L}G[v] = \mathbf{G}[L[v]] \quad (14.13)$$

Hence, from the definitions 14.3 and 14.4 and making use of equation 14.8 the linear barycentric estimators are:

$$\mathbf{L}G[u] = \sum_{\alpha=1}^4 \mathbf{k}_\alpha u_\alpha \quad (14.14)$$

$$\mathbf{L}G[v] = \sum_{\alpha=1}^4 \mathbf{k}_\alpha v_\alpha^T \quad (14.15)$$

The linear estimators of the divergence and curl of field \mathbf{v} are thus:

$$LD[\mathbf{v}] = \sum_{\alpha=1}^4 \mathbf{k}_{\alpha} \cdot \mathbf{v}_{\alpha} \quad (14.16)$$

$$LC[\mathbf{v}] = \sum_{\alpha=1}^4 \mathbf{k}_{\alpha} \times \mathbf{v}_{\alpha} \quad (14.17)$$

These last expressions solve exactly the equations for $LD[\mathbf{v}]$ and $LC[\mathbf{v}]$ derived by *Dunlop et al.* [1990] from Ampère's and Gauss's integral theorems combined with linear interpolation of the fields.

The differences $\mathbf{LG}[u] - \mathbf{G}[u]$ or $\mathbf{LG}[\mathbf{v}] - \mathbf{G}[\mathbf{v}]$ are respectively, and by definition, the errors of truncation of $\mathbf{LG}[u]$ or $\mathbf{LG}[\mathbf{v}]$.

14.2.2 Quadratic Interpolation and Related Estimators

Quadratic Interpolation within a Tetrahedron

The information provided by four samples of a physical field u or \mathbf{v} , measured simultaneously at the vertices of the tetrahedron, allows linear interpolation of the field in the vicinity of the cluster, but not higher order interpolation. Quadratic interpolation requires six extra independent simultaneous measurements: for example, the values $u_{\alpha\beta}$ at the midpoints $S_{\alpha\beta}$ of the lines $S_{\alpha}S_{\beta}$.

The quadratic interpolation uses a base of ten quadratic functions, each of them being equal to unity on only one node, and zero on all other nodes. Because μ_{α} vanishes on Π_{α} , the face of the tetrahedron opposite to S_{α} , i.e. on the three other vertices and on the three midpoint nodes belonging to Π_{α} , it should be a factor of the basis function $q_{\alpha\alpha}$ related to vertex S_{α} . The second factor is $2\mu_{\alpha} - 1$ because $\mu_{\alpha} = \frac{1}{2}$ on the remaining three midpoint nodes between S_{α} and Π_{α} . Hence $\mu_{\alpha}(2\mu_{\alpha} - 1)$ is the desired function, being moreover equal to 1 on vertex S_{α} where $\mu_{\alpha} = 1$. The midpoint node $S_{\alpha\beta}$ is the only node which belongs neither to Π_{α} nor to Π_{β} , hence the associated basis function is proportional to $\mu_{\alpha}\mu_{\beta}$; a factor 4 is needed to obtain the right weight on node $S_{\alpha\beta}$. The basic quadratic interpolating functions are thus: $q_{\alpha\alpha} = \mu_{\alpha}(2\mu_{\alpha} - 1)$ which is equal to 1 at vertex S_{α} and 0 at all other nodes, and $q_{\alpha\beta} = 4\mu_{\alpha}\mu_{\beta}$ which is equal to 1 at midpoint $S_{\alpha\beta}$ and 0 at all other nodes. Relation 14.9 can be used to demonstrate $q_{\alpha\alpha} = \mu_{\alpha}\mu_{\alpha} - \sum' \mu_{\alpha}\mu_{\beta}$ where \sum' denotes a sum over β different from α . Thus quadratic interpolations may be done with homogeneous functions of the μ_{α} 's:

$$Q[u](\mathbf{r}) = \sum_{\alpha=1}^4 u_{\alpha} q_{\alpha\alpha} + \sum_{\alpha=1}^4 \sum_{\beta>\alpha} u_{\alpha\beta} q_{\alpha\beta} \quad (14.18)$$

$$Q[\mathbf{v}](\mathbf{r}) = \sum_{\alpha=1}^4 \mathbf{v}_{\alpha} q_{\alpha\alpha} + \sum_{\alpha=1}^4 \sum_{\beta>\alpha} \mathbf{v}_{\alpha\beta} q_{\alpha\beta} \quad (14.19)$$

These formulas reduce to the linear interpolation formulas when $u_{\alpha\beta} = \frac{1}{2}(u_{\alpha} + u_{\beta})$.

Quadratic Estimators of Gradients

From the definitions of the $q_{\alpha\beta}$'s it follows that:

$$G[q_{\alpha\alpha}] = (4\mu_\alpha - 1)\mathbf{k}_\alpha \quad (14.20)$$

$$G[q_{\alpha\beta}] = 4(\mu_\alpha\mathbf{k}_\beta + \mu_\beta\mathbf{k}_\alpha) \quad (14.21)$$

These formulas can be used to define quadratic barycentric estimators of differential operators through the application of these differential operators to the quadratically interpolated physical fields, spatially sampled by the spacecraft. The gradients of the quadratically interpolated fields are linear functions of the position vector, the gradient of the scalar field is defined by:

$$G[Q[u]] = \sum_{\alpha=1}^4 (4\mu_\alpha - 1)\mathbf{k}_\alpha u_\alpha + 4 \sum_{\alpha=1}^4 \sum_{\beta>\alpha} (\mu_\alpha\mathbf{k}_\beta + \mu_\beta\mathbf{k}_\alpha) u_{\alpha\beta}$$

and the gradient of the vector field is written:

$$G[Q[\mathbf{v}]] = \sum_{\alpha=1}^4 (4\mu_\alpha - 1)\mathbf{k}_\alpha \mathbf{v}_\alpha^T + 4 \sum_{\alpha=1}^4 \sum_{\beta>\alpha} (\mu_\alpha\mathbf{k}_\beta + \mu_\beta\mathbf{k}_\alpha) \mathbf{v}_{\alpha\beta}^T$$

The quadratic barycentric estimators of the gradients are chosen to be equal to the latter expressions evaluated at the barycentre of the tetrahedron, i.e.:

$$QG[u] = \sum_{\alpha=1}^4 \sum_{\beta>\alpha} (\mathbf{k}_\alpha + \mathbf{k}_\beta) u_{\alpha\beta} \quad (14.22)$$

$$QG[\mathbf{v}] = \sum_{\alpha=1}^4 \sum_{\beta>\alpha} (\mathbf{k}_\alpha + \mathbf{k}_\beta) \mathbf{v}_{\alpha\beta}^T \quad (14.23)$$

An Approximate Procedure Taking Advantage of the Orbital Motion

A midpoint sample can only be estimated at the expense of the time resolution by defining it as the mean value of its nearest vertical neighbours on a supplementary position of the cluster, but that would lead to midpoint and vertical samples defined at different times. A midpoint sample has to be defined by averaging this mean value on two positions of the cluster, one retarded and the other advanced with respect to the central position. One position of the cluster giving two independent midpoint samples on opposite edges of the tetrahedron, it appears that only three pairs of retarded/advanced clusters are sufficient to estimate the six midpoint samples. Figure 14.2 illustrate the following scheme:

$$u_{23}(t) = \frac{1}{4} (u_2(t - 2dt) + u_3(t - 2dt) + u_2(t + 2dt) + u_3(t + 2dt))$$

$$u_{14}(t) = \frac{1}{4} (u_1(t - 2dt) + u_4(t - 2dt) + u_1(t + 2dt) + u_4(t + 2dt))$$

The adequate interval of time dt will be specifically determined for each physical field; it should be small enough to preserve the accuracy of the estimation, but nevertheless large enough to guarantee the independence of the successive filtered samples of data.

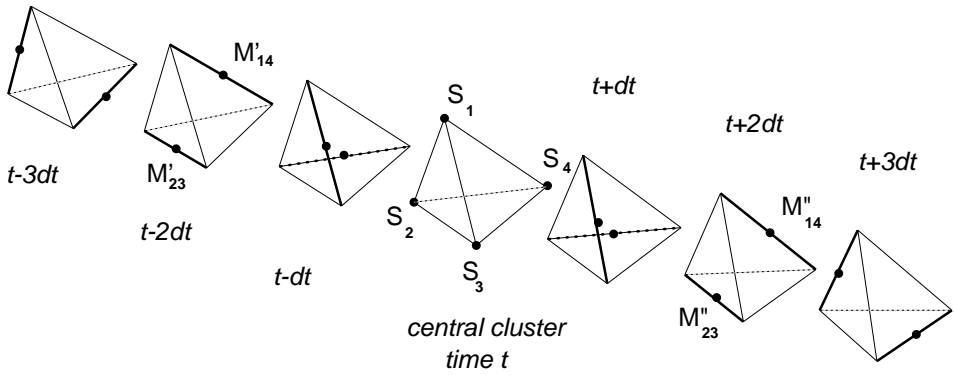


Figure 14.2: Sequence of seven tetrahedra necessary for the quadratic interpolation.

In fact due to the linear interpolations used to estimate the midpoint values this approximate quadratic estimator of the gradient is linearly degenerate and will be denoted $L_2G[B]$. It will be demonstrated in Section 14.4.3 that for a thick and planar current sheet: $L_2G[B] = LG[B]$. Nevertheless this identity relies upon power series expansions accurate to second order with respect to small parameters defined by the characteristic size of the cluster, or the distance between the successive clusters, divided by the scale length of the gradient. The two estimators are indeed slightly different, and $L_2G[B]$ could be used as an alternative to $LG[B]$, provided it is used with a smaller time interval between successive clusters.

14.3 Physical and Geometrical Errors

The investigation of the statistical properties of the barycentric estimators requires the knowledge of the statistical properties of the cluster configuration. We will only discuss the lowest order statistics, namely the expectation values and covariances.

14.3.1 Statistical Properties of the Reciprocal Vectors

The nominal position \mathbf{r}_α of spacecraft α differs from the true position $\mathbf{r}_\alpha + \delta\mathbf{r}_\alpha$. The difference between the true and nominal positions is supposed to be a random vector. The expectation value of the random variable X is noted $\langle X \rangle$ and the covariance of the random variables X and Y is $\langle (X - \langle X \rangle)(Y - \langle Y \rangle) \rangle$. Assuming that spacecraft positions are not affected by systematic errors, i.e. that $\langle \delta\mathbf{r}_\alpha \rangle = 0$ for any α , the covariance of the positions of spacecraft α and β is the tensor $\langle \delta\mathbf{r}_\alpha \delta\mathbf{r}_\beta^T \rangle$, which is not known *a priori* but should be part of the mission tracking program. Here, it will be supposed to be known. A unique base of reciprocal vectors can be computed from any configuration of the four spacecraft; the set $\{\mathbf{k}_\alpha, \alpha = 1 \text{ to } 4\}$ is the nominal reciprocal base computed from the nominal configuration $\{\mathbf{r}_\alpha, \alpha = 1 \text{ to } 4\}$, meanwhile $\{\mathbf{k}_\alpha + \delta\mathbf{k}_\alpha, \alpha = 1 \text{ to } 4\}$ is the true reciprocal base computed from the true cluster configuration $\{\mathbf{r}_\alpha + \delta\mathbf{r}_\alpha, \alpha = 1 \text{ to } 4\}$. The key issue consists in finding the relation between the deviations of the reciprocal and position vectors. This relation is

obtained by differentiating equation 14.11:

$$\sum_{\alpha=1}^4 \delta \mathbf{k}_{\alpha} \mathbf{r}_{\alpha}^T = - \sum_{\alpha=1}^4 \mathbf{k}_{\alpha} \delta \mathbf{r}_{\alpha}^T$$

Multiplying right by \mathbf{k}_{β} , then making use of equations 14.6 and 14.10 to evaluate the left-hand side of the above equation, we get:

$$\delta k_{\beta} = - \sum_{\alpha=1}^4 \left(\mathbf{k}_{\beta}^T \delta \mathbf{r}_{\alpha} \right) \mathbf{k}_{\alpha} = - \left(\sum_{\alpha=1}^4 \mathbf{k}_{\alpha} \delta \mathbf{r}_{\alpha}^T \right) \mathbf{k}_{\beta}, \text{ for any } \beta. \quad (14.24)$$

The first expression, involving the scalar products $\mathbf{k}_{\beta}^T \delta \mathbf{r}_{\alpha}$ will later appear as the most useful. The latter equation 14.24 ensures that, accordingly to the hypothesis of unbiased position measurements, the deviations of the reciprocal vectors have null expectation values:

$$\langle \delta \mathbf{k}_{\alpha} \rangle = \mathbf{0}, \text{ for any } \alpha. \quad (14.25)$$

Equation 14.24 allows for computing the covariances of the reciprocal vectors along the following lines:

$$\delta k_{\alpha i} \delta k_{\beta j} = \sum_{\gamma=1}^4 \sum_{v=1}^4 \left(\mathbf{k}_{\alpha}^T \delta \mathbf{r}_{\gamma} \right) \left(\mathbf{k}_{\beta}^T \delta \mathbf{r}_v \right) k_{\gamma i} k_{v j}$$

Changing $\mathbf{k}_{\beta}^T \delta \mathbf{r}_v$ into the equivalent expression $\delta \mathbf{r}_v^T \mathbf{k}_{\beta}$, the expectation value of the product $\left(\mathbf{k}_{\alpha}^T \delta \mathbf{r}_{\gamma} \right) \left(\mathbf{k}_{\beta}^T \delta \mathbf{r}_v \right)$ becomes $\mathbf{k}_{\alpha}^T \langle \delta \mathbf{r}_{\gamma} \delta \mathbf{r}_v^T \rangle \mathbf{k}_{\beta}$, thus revealing the role played by the covariances of the vector positions of the spacecraft. The general expression for the covariances of the components of the reciprocal vectors is thus:

$$\langle \delta k_{\alpha i} \delta k_{\beta j} \rangle = \sum_{\gamma=1}^4 \sum_{v=1}^4 \mathbf{k}_{\alpha}^T \langle \delta \mathbf{r}_{\gamma} \delta \mathbf{r}_v^T \rangle \mathbf{k}_{\beta} k_{\gamma i} k_{v j} \quad (14.26)$$

The covariance matrices of the spacecraft positions will be determined specifically for a given mission; they are not known *a priori*. Nevertheless, it is very likely that the position vectors of two spacecraft will be uncorrelated, furthermore it seems reasonable to assume that the covariance matrix of the position vector will be the same for all spacecraft of the cluster. This means that:

$$\langle \delta \mathbf{r}_{\gamma} \delta \mathbf{r}_v^T \rangle = \delta_{\gamma, v} \langle \delta \mathbf{r} \delta \mathbf{r}^T \rangle \quad (14.27)$$

Under these simplifying assumptions the general expression 14.26 is reduced to:

$$\langle \delta k_{\alpha i} \delta k_{\beta j} \rangle = \mathbf{k}_{\alpha}^T \langle \delta \mathbf{r} \delta \mathbf{r}^T \rangle \mathbf{k}_{\beta} K_{ij} \quad (14.28)$$

K_{ij} , the last factor, is the ij component of the reciprocal tensor defined by:

$$\mathbf{K} = \sum_{v=1}^4 \mathbf{k}_v \mathbf{k}_v^T \quad (14.29)$$

14.3.2 Statistical Properties of the Linear Estimator LG

This section investigates the statistical errors which affect the estimation of the gradient of a vector field sampled by a four spacecraft mission. Two kinds of statistical errors can be distinguished with regard to their physical or geometric origin. The first type originates in the measurement of the field itself and involves the results of the error analysis made for each relevant experiment. Of course, the resulting uncertainties on the gradient depend upon the geometry of the tetrahedron formed by the four spacecraft but only through the nominal configuration. The second type of statistical errors comes from the uncertainties on the spacecraft positions. Although the present theory is general and valid for any type of measured field, the discussion of the physical errors usually deserve specific considerations depending upon the nature of the measured field. Hence a detailed discussion of the accuracy of current determination via magnetic measurements is given in Chapter 16 and gradients of fields computed from the plasma measurements are discussed in Chapter 17. Let us proceed now with the general theory. First we derive expectation values and covariances of the components of the linear barycentric estimator of the gradient of an unspecified vector field \mathbf{v} , then specific considerations will be made in the case of the magnetic field.

Let \mathbf{v}_α be any vector field measured by spacecraft α at any given time, meanwhile the true field at the same time and location is $\mathbf{v}_\alpha + \delta\mathbf{v}_\alpha$. It is worth mentioning that short scale fluctuations should be filtered out from a physical field before trying to estimate its gradient. Thus the difference between the measured and true vectors is supposed to be a random vector having an expectation value $\langle\delta\mathbf{v}_\alpha\rangle$ that ideally should be equal to zero. Differentiating equation 14.15 we get the variation of the estimator, the expectation value of which is¹:

$$\langle\delta\mathbf{LG}\rangle = \sum_{\alpha=1}^4 \left(\mathbf{k}_\alpha \langle\delta\mathbf{v}_\alpha^T\rangle + \langle\delta\mathbf{k}_\alpha\rangle \mathbf{v}_\alpha^T \right)$$

When both the field and the position measurements are not spoiled by systematic errors, i.e. when $\langle\delta\mathbf{v}_\alpha\rangle = \mathbf{0}$ and $\langle\delta\mathbf{k}_\alpha\rangle = \mathbf{0}$, the linear barycentric estimator of the gradient of the vector field is not biased, i.e. $\langle\delta\mathbf{LG}\rangle = \mathbf{0}$.

The analysis of covariances of the components of $\langle\delta\mathbf{LG}\rangle$ takes advantage of the statistical independence of the geometrical and physical measurements and of the assumption of unbiased position measurements to eliminate the covariances of geometrical and physical variables:

$$\langle\delta k_{\alpha j} \delta v_{\beta m}\rangle = \langle\delta k_{\alpha j}\rangle \langle\delta v_{\beta m}\rangle = 0$$

Hence the covariance of the ij and mn components of \mathbf{LG} can be written:

$$\langle\delta LG_{ij} \delta LG_{mn}\rangle = \sum_{\alpha=1}^4 \sum_{\beta=1}^4 \left(\langle\delta v_{\alpha i} \delta v_{\beta m}\rangle k_{\alpha j} k_{\beta n} + \langle\delta k_{\alpha j} \delta k_{\beta n}\rangle v_{\alpha i} v_{\beta m} \right) \quad (14.30)$$

As far as the magnetic field is concerned the latter expression can be simplified by noticing that the magnetic measurements on two different spacecraft are statistically independent, nevertheless each spacecraft will have its own covariance matrix for the magnetic measurements, this means that:

$$\langle\delta B_{\alpha i} \delta B_{\beta m}\rangle = \delta_{\alpha\beta} \langle\delta B_i \delta B_m\rangle_\alpha$$

¹In order to simplify notations, functional $\mathbf{LG}[\mathbf{v}]$ is simply written \mathbf{LG} throughout this section.

Under these simplifying assumptions and those made to get the reduced form (14.28) of the covariances of the reciprocal vectors, the following simplified result holds:

$$\langle \delta L G_{ij} \delta L G_{mn} \rangle = \sum_{\alpha=1}^4 \langle \delta B_i \delta B_m \rangle_{\alpha} k_{\alpha j} k_{\alpha n} + \left(\sum_{\alpha=1}^4 \sum_{\beta=1}^4 k_{\alpha}^T \langle \delta \mathbf{r} \delta \mathbf{r}^T \rangle k_{\beta} B_{\alpha i} B_{\beta m} \right) K_{jn} \quad (14.31)$$

Usually the covariance matrices $\langle \delta B_i \delta B_m \rangle_{\alpha}$ will differ from one spacecraft to the other and will not be diagonal, consequently further simplifications will be impossible.

Both the general expression 14.30 and the one specialised to the magnetic field case 14.31 contain explicitly separated contributions originating in the physical and geometrical uncertainties.

14.3.3 Comparison of the Physical and Geometrical Errors for a Regular Tetrahedron

In order to compare the order of magnitudes of the different contributions to the errors affecting the estimation of the gradient of the magnetic field, let us consider the ideal case of a regular tetrahedron. It is convenient to work in a cartesian frame attached to the tetrahedron such that the four vertices $\{S_{\alpha}, \alpha = 1 \text{ to } 4\}$ have positions:

$$\begin{aligned} \mathbf{r}_1^T &= a(-1, -1, -1) \quad , \quad \mathbf{r}_2^T = a(+1, +1, -1) \\ \mathbf{r}_3^T &= a(+1, -1, +1) \quad , \quad \mathbf{r}_4^T = a(-1, +1, +1) \end{aligned} \quad (14.32)$$

The reciprocal vectors are:

$$\begin{aligned} \mathbf{k}_1^T &= \frac{1}{4a}(-1, -1, -1) \quad , \quad \mathbf{k}_2^T = \frac{1}{4a}(+1, +1, -1) \\ \mathbf{k}_3^T &= \frac{1}{4a}(+1, -1, +1) \quad , \quad \mathbf{k}_4^T = \frac{1}{4a}(-1, +1, +1) \end{aligned} \quad (14.33)$$

It is straightforward to show that the reciprocal tensor \mathbf{K} can be written:

$$\mathbf{K} = \frac{1}{4a^2} \mathbf{I} \quad (14.34)$$

For the sake of simplicity we make the further but, as already mentioned, unrealistic assumptions:

$$\langle \delta B_i \delta B_m \rangle_{\alpha} = (\Delta B)^2 \delta_{im} \quad \text{and} \quad \langle \delta \mathbf{r} \delta \mathbf{r}^T \rangle = (\Delta r)^2 \mathbf{I}$$

where ΔB and Δr denote respectively the uncertainties on the physical measures and on the coordinates of the spacecraft: $(\Delta B)^2$ is indeed the variance of the filtered data. It is then easy to obtain:

$$\langle \delta L G_{ij} \delta L G_{mn} \rangle = \left(\frac{\Delta B}{2a} \right)^2 \delta_{im} \delta_{jn} + \left(\frac{\Delta r}{2a} \right)^2 \frac{\delta_{jn}}{4a^2} \sum_{\alpha=1}^4 \sum_{\beta=1}^4 \left(\delta_{\alpha\beta} - \frac{1}{4} \right) B_{\alpha i} B_{\beta m}$$

where use has been made of the identity $\mathbf{k}_\alpha \cdot \mathbf{k}_\beta = \frac{1}{16a^2} (4\delta_{\alpha\beta} - 1)$. The double sum is of the order of B^2 and the final result holds for a regular tetrahedron:

$$\frac{\text{physical errors}}{\text{geometrical errors}} \approx \frac{\Delta B}{B} \frac{2a}{\Delta r}$$

In their investigation of errors affecting the estimation of $\nabla \times \mathbf{B}$ [Dunlop et al. \[1990\]](#) evaluate the maximum intrinsic error $\Delta B/B$ to be less or equal to 0.01; this value together with $\Delta r/2a \sim 0.01$, a typical value for the Cluster mission, gives a ratio of errors equal to 1, which means that in most cases the geometrical errors will compare to the physical ones. In case of degraded positional accuracy the geometrical errors will dominate as suggested by [Dunlop et al. \[1990\]](#) for the estimation of the current density; this contrasts with the plasma measurements discussed in Chapter 17. However the above argument just provides orders of magnitude, in practice the errors should be evaluated by formula 14.31 for the magnetic field, or by the more general formula 14.30, possibly combined with formula 14.28, for other fields, taking into account the effective covariances of the spacecraft positions and of the measured field components.

14.3.4 Effect of the Orbital Motion on LG

The estimation of the gradient of a physical field will usually require averaging of the field samples over some time interval in order to eliminate small scale fluctuations or to implement the approximate quadratic estimator, hence we have to investigate in which respect the estimation of the gradient will be affected by the deformation of the tetrahedron due to slight differences in the orbital motions of the four spacecraft. Let \mathbf{V}_α be the velocity of spacecraft α , and Δt the finite interval of time required to estimate the gradient; from definition 14.15 the variation of $\text{LG}[\mathbf{B}]$ induced by the orbital motion during Δt can be written:

$$\delta \text{LG}[\mathbf{B}] = \sum_{\beta=1}^4 \left(\delta \mathbf{k}_\beta \mathbf{B}_\beta^T + \Delta t \mathbf{k}_\beta \mathbf{V}_\beta^T \text{G}[\mathbf{B}] \right)$$

and the variation of the reciprocal vectors \mathbf{k}_β is given by equation 14.24 which can be written:

$$\delta \mathbf{k}_\beta = -\Delta t \left(\sum_{\alpha=1}^4 \mathbf{k}_\alpha \mathbf{V}_\alpha^T \right) \mathbf{k}_\beta$$

thus leading to:

$$\delta \text{LG}[\mathbf{B}] = \Delta t \left(\sum_{\alpha=1}^4 \mathbf{k}_\alpha \mathbf{V}_\alpha^T \right) (\text{G}[\mathbf{B}] - \text{LG}[\mathbf{B}])$$

The tensor $\sum_{\alpha=1}^4 \mathbf{k}_\alpha \mathbf{V}_\alpha^T$ is equal to $\sum_{\alpha=1}^4 \mathbf{k}_\alpha (\mathbf{V}_\alpha - \mathbf{V}_b)^T$, where \mathbf{V}_b is the velocity of the barycentre. The difference $\delta \mathbf{V}_\alpha = \mathbf{V}_\alpha - \mathbf{V}_b$ is supposed to be small compared to \mathbf{V}_b and is computed by using the angular momentum $\boldsymbol{\sigma}_\alpha = \mathbf{r}_\alpha \times m \mathbf{V}_\alpha$ and Runge's vector

$$\mathbf{R}_\alpha = \mathbf{V}_\alpha \times \boldsymbol{\sigma}_\alpha - \frac{GMm}{r_\alpha} \mathbf{r}_\alpha$$

which are approximate invariants of the orbital motion on a time scale Δt which is short compared to the orbital period. Moreover these invariants should be the same for the four

spacecraft because they have almost the same orbits, thus the velocity of the barycentre can be written:

$$\mathbf{V}_b = \frac{\boldsymbol{\sigma}}{\sigma^2} \times \left(\mathbf{R} + \frac{GMm}{r_b} \mathbf{r}_b \right)$$

where G is the gravitational constant, M the mass of the Earth, m the mass of one spacecraft and $\sigma^2 = \sigma_x^2 + \sigma_y^2 + \sigma_z^2$. Introducing $\delta \mathbf{r}_\alpha = \mathbf{r}_\alpha - \mathbf{r}_b$ and $\hat{\mathbf{e}} = \mathbf{r}_b / r_b$, the differentiation of \mathbf{V}_b gives:

$$\delta \mathbf{V}_\alpha = \frac{GMm}{\sigma^2 r_b} \bar{\boldsymbol{\sigma}} \left(\mathbf{I} - \hat{\mathbf{e}} \hat{\mathbf{e}}^T \right) \delta \mathbf{r}_\alpha \quad (14.35)$$

The cross product $\boldsymbol{\sigma} \times \mathbf{a}$ has been replaced by $\bar{\boldsymbol{\sigma}} \mathbf{a}$, with tensor $\bar{\boldsymbol{\sigma}}$ defined by:

$$\bar{\boldsymbol{\sigma}} = \begin{pmatrix} 0 & -\sigma_z & \sigma_y \\ \sigma_z & 0 & -\sigma_x \\ -\sigma_y & \sigma_x & 0 \end{pmatrix} \quad (14.36)$$

The final result for the variation of $\mathbf{LG}[\mathbf{B}]$ is:

$$\delta \mathbf{LG}[\mathbf{B}] = \Delta t \frac{GMm}{\sigma^2 r_b} \left(\mathbf{I} - \hat{\mathbf{e}} \hat{\mathbf{e}}^T \right) \bar{\boldsymbol{\sigma}} (\mathbf{LG}[\mathbf{B}] - \mathbf{G}[\mathbf{B}]) \quad (14.37)$$

For an elliptic orbit

$$\frac{GMm}{r_b} \sim m V_b^2 \quad \text{thus,} \quad \Delta t \frac{GMm}{\sigma r_b} \sim \frac{V_b \Delta t}{r_b}$$

For the Cluster mission, V_b will be of the order of 5 km/s near the perigee (geocentric distance of $4R_E$) and of 1 km/s near the perigee at $22R_E$. A typical Δt of 60 s gives:

$$(\text{apogee}) 10^{-4} \lesssim \Delta t \frac{GMm}{\sigma r_b} \lesssim 10^{-2} (\text{perigee})$$

In conclusion the errors induced by the orbital deformation of the cluster on $\mathbf{LG}[\mathbf{B}]$ are much smaller than the difference $\mathbf{LG}[\mathbf{B}] - \mathbf{G}[\mathbf{B}]$, named the truncation error of the linear estimator of the gradient. The next section is devoted to the analysis of this truncation error.

14.4 Truncation Errors for a One-dimensional Model: the Thick Plane Current Sheet

Let the magnetic field created by a one-dimensional current sheet be:

$$\mathbf{B} = f(\zeta) \hat{\mathbf{u}} + g(\zeta) \hat{\mathbf{v}} + B_n \hat{\mathbf{n}}$$

where $(\hat{\mathbf{u}}, \hat{\mathbf{v}}, \hat{\mathbf{n}})$ is a direct trihedron and $\zeta = \mathcal{K} \mathbf{r} \cdot \hat{\mathbf{n}}$; \mathcal{K}^{-1} being the thickness of the layer, meanwhile constant B_n is the component of the field normal to the sheet. Functions f and g are such that:

$$\lim_{\zeta \rightarrow -\infty} \mathbf{B} = \mathbf{B}_1 \quad \text{and} \quad \lim_{\zeta \rightarrow +\infty} \mathbf{B} = \mathbf{B}_2$$

According to definition 14.2, the gradient of this field is:

$$\mathbf{G}[\mathbf{B}] = \mathcal{K} \hat{\mathbf{n}} \left(\frac{df}{d\zeta} \hat{\mathbf{u}} + \frac{dg}{d\zeta} \hat{\mathbf{v}} \right)^T \quad (14.38)$$

The divergence of the field is equal to the trace of this tensor, i.e. zero owing to the orthogonality of the vectors $(\hat{\mathbf{u}}, \hat{\mathbf{v}}, \hat{\mathbf{n}})$. The errors of truncation of the linear, quadratic and linearly degenerate quadratic estimators of the gradient will now be evaluated for this simple model.

14.4.1 Truncation Errors for $\mathbf{LG}[\mathbf{B}]$

The position vector of spacecraft α is written $\mathbf{r}_\alpha = R(\hat{\mathbf{e}} + \boldsymbol{\varepsilon}_\alpha)$ where $R\hat{\mathbf{e}}$ is the position vector of the barycentre of the tetrahedron. The corresponding value of ζ is:

$$\zeta_\alpha = \mathcal{K} R \hat{\mathbf{e}} \cdot \hat{\mathbf{n}} + \mathcal{K} R \boldsymbol{\varepsilon}_\alpha \cdot \hat{\mathbf{n}} = \zeta_0 + \mathcal{K} R \varepsilon_\alpha$$

From definition 14.15 the linear barycentric estimator of the gradient can be written:

$$\mathbf{LG}[\mathbf{B}] = \sum_{\alpha=1}^4 \mathbf{k}_\alpha (f(\zeta_\alpha) \hat{\mathbf{u}} + g(\zeta_\alpha) \hat{\mathbf{v}})^T$$

Assuming that the four small parameters ε_α have magnitudes comparable to a/R , a being a length characterising the size of the cluster, the series expansion of $\mathbf{LG}[\mathbf{B}]$ accurate to first order in $\mathcal{K}a$ can be written:

$$\begin{aligned} \mathbf{LG}[\mathbf{B}] &= \sum_{\alpha=1}^4 \mathbf{k}_\alpha \mathbf{B}^T(\zeta_0) + \mathcal{K} \left(R \sum_{\alpha=1}^4 \mathbf{k}_\alpha (\boldsymbol{\varepsilon}_\alpha \cdot \hat{\mathbf{n}}) \right) \left(\frac{df}{d\zeta} \hat{\mathbf{u}} + \frac{dg}{d\zeta} \hat{\mathbf{v}} \right)^T_{\zeta=\zeta_0} \\ &\quad + \mathcal{K}a \mathcal{K} \left(\frac{R^2}{2a} \sum_{\alpha=1}^4 \mathbf{k}_\alpha \varepsilon_\alpha^2 \right) \left(\frac{d^2 f}{d\zeta^2} \hat{\mathbf{u}} + \frac{d^2 g}{d\zeta^2} \hat{\mathbf{v}} \right)^T_{\zeta=\zeta_0} + \mathcal{O}(\mathcal{K}^2 a^2) \end{aligned}$$

the first term is equal to zero due to property 14.10 and the second factor of the second term is equal to $\hat{\mathbf{n}}$ owing to property 14.12. Thus the linear estimator of the gradient is equal to:

$$\mathbf{LG}[\mathbf{B}] = \mathbf{G}_0[\mathbf{B}] + \mathcal{K}a \mathcal{K} \mathbf{C}_1 \left(\frac{d^2 f}{d\zeta^2} \hat{\mathbf{u}} + \frac{d^2 g}{d\zeta^2} \hat{\mathbf{v}} \right)^T + \mathcal{O}(\mathcal{K}^2 a^2)$$

where $\mathbf{G}_0[\mathbf{B}]$ is the true gradient of the field at the barycentre of the tetrahedron, i.e. tensor 14.38 evaluated at ζ_0 , and the second term involving \mathbf{C}_1 is the error of truncation. Vector \mathbf{C}_1 defined by:

$$\mathbf{C}_1 = \frac{R^2}{2a} \sum_{\alpha=1}^4 \mathbf{k}_\alpha (\boldsymbol{\varepsilon}_\alpha \cdot \hat{\mathbf{n}})^2 \quad (14.39)$$

depends upon the cluster configuration and the normal $\hat{\mathbf{n}}$ and is generally neither orthogonal to $\hat{\mathbf{u}}$ nor to $\hat{\mathbf{v}}$, hence the trace of $\mathbf{LG}[\mathbf{B}]$ is generally not equal to zero due to the

errors of truncation, which means that the linear estimation of $\nabla \cdot \mathbf{B}$ will differ from zero. The tensor of truncation errors being not symmetric will generally affect the estimation of the large scale current density. Vector \mathbf{C}_1 vanishes under exceptional circumstances, for example when ε_α^2 is independent of α .

The truncation error can be made more explicit in the ideal case of a regular tetrahedron. Considering the frame of reference attached to the tetrahedron, as defined by equations 14.32, the normal $\hat{\mathbf{n}}$ to the current sheet has components (n_x, n_y, n_z) and vector \mathbf{C}_1 can be written:

$$\mathbf{C}_1 = -(n_y n_z \hat{\mathbf{x}} + n_z n_x \hat{\mathbf{y}} + n_x n_y \hat{\mathbf{z}})$$

This vector vanishes only when the normal $\hat{\mathbf{n}}$ is parallel to one of the axes of the reference frame, a very unlikely situation. With the latter expression of vector \mathbf{C}_1 , the linear estimators of the divergence and curl of \mathbf{B} can be written:

$$\begin{aligned} LD[\mathbf{B}] &= -\mathcal{K}a \mathcal{K} \left(\frac{d^2 f}{d\zeta^2} (n_y n_z u_x + n_z n_x u_y + n_x n_y u_z) \right. \\ &\quad \left. + \frac{d^2 g}{d\zeta^2} (n_y n_z v_x + n_z n_x v_y + n_x n_y v_z) \right) \\ &\quad + \mathcal{O}(\mathcal{K}^2 a^2) \end{aligned}$$

$$(LC[\mathbf{B}])_z = -\mathcal{K}a \mathcal{K} \left(\frac{d^2 f}{d\zeta^2} n_z (n_y u_y - n_x u_x) + \frac{d^2 g}{d\zeta^2} n_z (n_y v_y - n_x v_x) \right) + \mathcal{O}(\mathcal{K}^2 a^2)$$

The other components of $LC[\mathbf{B}]$ are obtained through cyclic permutations of the indices x, y and z. The divergence of \mathbf{B} being equal to zero, its linear estimation is equal to its truncation error; this substantiates the intuitive idea of using this estimator to evaluate the errors of truncations which affect the estimation of the current density. Unfortunately the latter are not simply related to the former except under special circumstances, for example when the current density \mathbf{J} has a constant direction (e.g. along vector $\hat{\mathbf{v}}$ which means that function g vanishes identically); in this peculiar case the error of truncation affecting J_z is proportional to the linear estimation of the divergence:

$$\frac{(LC[\mathbf{B}])_z}{LD[\mathbf{B}]} = \frac{n_z (n_y u_y - n_x u_x)}{n_y n_z u_x + n_z n_x u_y + n_x n_y u_z}$$

This point is further investigated by simulations in Chapter 16: the numerical results corroborate that generally the linear estimation of $\nabla \cdot \mathbf{B}$ cannot be used to estimate the errors of truncation affecting the linear estimation of $\nabla \times \mathbf{B}$.

14.4.2 Truncation Errors for $\mathbf{QG}[\mathbf{B}]$

From definition 14.23 the quadratic barycentric estimator of the gradient can be written:

$$\mathbf{QG}[\mathbf{B}] = \sum_{\alpha=1}^4 \sum_{\beta>\alpha} (\mathbf{k}_\alpha + \mathbf{k}_\beta) (f(\zeta_{\alpha\beta}) \hat{\mathbf{u}} + g(\zeta_{\alpha\beta}) \hat{\mathbf{v}})^T$$

Making series expansions accurate to second order in $\mathcal{K}a$ of $f(\zeta_{\alpha\beta})$ and $g(\zeta_{\alpha\beta})$ with $\zeta_{\alpha\beta} = \zeta_0 + \frac{1}{2}\mathcal{K}R(\varepsilon_\alpha + \varepsilon_\beta)$, we get:

$$\begin{aligned} \text{QG}[\mathbf{B}] &= \sum_{\alpha=1}^4 \sum_{\beta>\alpha} (\mathbf{k}_\alpha + \mathbf{k}_\beta) \mathbf{B}^T(\zeta_0) \\ &+ \mathcal{K} \left(\frac{R}{2} \sum_{\alpha=1}^4 \sum_{\beta>\alpha} (\mathbf{k}_\alpha + \mathbf{k}_\beta)(\varepsilon_\alpha + \varepsilon_\beta) \right) \left(\frac{df}{d\zeta} \hat{\mathbf{u}} + \frac{dg}{d\zeta} \hat{\mathbf{v}} \right)_{\zeta=\zeta_0}^T \\ &+ \mathcal{K}a \mathcal{K} \left(\frac{R^2}{8a} \sum_{\alpha=1}^4 \sum_{\beta>\alpha} (\mathbf{k}_\alpha + \mathbf{k}_\beta)(\varepsilon_\alpha + \varepsilon_\beta)^2 \right) \left(\frac{d^2f}{d\zeta^2} \hat{\mathbf{u}} + \frac{d^2g}{d\zeta^2} \hat{\mathbf{v}} \right)_{\zeta=\zeta_0}^T \\ &+ \mathcal{O}(\mathcal{K}^2 a^2) \end{aligned}$$

Using properties 14.10 and 14.12 once more, it is easily demonstrated that the first and third terms of the right-hand side are equal to zero meanwhile the second term is equal to the exact gradient at the barycentre of the tetrahedron, thus the quadratic barycentric estimator of the gradient can be written:

$$\text{QG}[\mathbf{B}] = \mathbf{G}_0[\mathbf{B}] + \mathcal{O}(\mathcal{K}^2 a^2)$$

i.e. the first order errors of truncation vanish exactly whatever is the configuration of the cluster. This expected result demonstrates the improvement achieved by the quadratic interpolation. Unfortunately we do not have any possibility to measure the midpoint fields $\mathbf{B}_{\alpha\beta}$ and consequently this estimator will not be computable.

14.4.3 Truncation Errors for $\mathbf{L}_2\mathbf{G}[\mathbf{B}]$

We will now demonstrate that the approximate quadratic estimator defined in Section 14.2.2 is indeed linearly degenerate and identical to $\mathbf{LG}[\mathbf{B}]$ up to the first order in $\mathcal{K}a$.

Let us assume that the six auxiliary clusters we need are deduced from the central cluster (see Figure 14.2) by translations $(\pm q\mathbf{d}, \text{ for } q = 1, 2, 3)$; this is equivalent to the assumption that the four spacecraft have the same constant velocity. The estimated midpoint fields are defined by:

$$\mathbf{B}_{\alpha\beta} = \frac{1}{4} \left(\mathbf{B}_\alpha^+ + \mathbf{B}_\alpha^- + \mathbf{B}_\beta^+ + \mathbf{B}_\beta^- \right)$$

which involve functions f and g evaluated at $\zeta_\alpha^\pm = \zeta_0 + \mathcal{K}R(\varepsilon_\alpha \pm q\eta)$, where the small parameter $\eta = R^{-1}\mathbf{d} \cdot \hat{\mathbf{n}}$ is assumed to be of the same order as the ε_α with $q = 1, 2, 3$ for $(\alpha, \beta) = (1, 3)$ and $(2, 4)$, $(2, 3)$ and $(1, 4)$, and $(2, 1)$ and $(3, 4)$ respectively. The series expansion of 14.23 computed with the above approximate $\mathbf{B}_{\alpha\beta}$, accurate to the first order in $\mathcal{K}a$ is:

$$\mathbf{Q}_a\mathbf{G}[\mathbf{B}] = \mathbf{G}_0[\mathbf{B}]$$

$$\begin{aligned}
& + \mathcal{K}a \mathcal{K} \left(\frac{R^2}{4a} \sum_{\alpha=1}^4 \sum_{\beta>\alpha} (\mathbf{k}_\alpha + \mathbf{k}_\beta) (\varepsilon_\alpha^2 + \varepsilon_\beta^2) \right) \left(\frac{d^2 f}{d\zeta^2} \hat{\mathbf{u}} + \frac{d^2 g}{d\zeta^2} \hat{\mathbf{v}} \right)_{\zeta=\zeta_0}^T \\
& + \mathcal{O}(\mathcal{K}^2 a^2)
\end{aligned}$$

Terms proportional to η and η^2 vanish exactly and the second term is simplified by making use of property 14.12. It appears that this approximate quadratic estimator is affected by the same truncation error as the linear estimator, more precisely:

$$\mathbf{Q}_a \mathbf{G}[\mathbf{B}] = \mathbf{L} \mathbf{G}[\mathbf{B}] + \mathcal{O}(\mathcal{K}^2 a^2) \quad (14.40)$$

This result means that, for a uniform rectilinear motion of the cluster, the interpolation made to estimate the midpoint fields $\mathbf{B}_{\alpha\beta}$ causes the degeneracy of the quadratic estimator. This remark justifies the notation $\mathbf{L}_2 \mathbf{G}[\mathbf{B}]$ adopted for $\mathbf{Q}_a \mathbf{G}[\mathbf{B}]$.

14.5 Other Applications

14.5.1 Derivation of the Spatial Aliasing Condition

The base of reciprocal vectors allows a simple derivation of the spatial aliasing condition [Neubauer and Glassmeier, 1990]. Let us consider two monochromatic plane waves differing only by their wave vectors \mathcal{K}_1 and \mathcal{K}_2 :

$$\mathbf{B}_1 = \mathbf{B}_0 \exp i(\mathcal{K}_1 \cdot \mathbf{r} - \Omega t) \quad , \text{ and } \quad \mathbf{B}_2 = \mathbf{B}_0 \exp i(\mathcal{K}_2 \cdot \mathbf{r} - \Omega t)$$

The aliasing problem can be formulated by the question: “What relationship should exist between wave vectors \mathcal{K}_1 and \mathcal{K}_2 in order to make these two waves undistinguishable by the cluster?”. In other words the waveforms $\mathbf{B}(\mathbf{r}_\alpha, t)$, $\alpha = 1$ to 4, recorded by the four spacecraft should be identical, which means that

$$(\mathcal{K}_2 - \mathcal{K}_1) \cdot \mathbf{r}_\alpha = 2\pi n_\alpha \quad , \text{ for } \alpha = 1 \text{ to } 4, \text{ and where } n_\alpha \text{ are signed integers} \quad (14.41)$$

The solution for $\mathcal{K}_2 - \mathcal{K}_1$ is:

$$(\mathcal{K}_2 - \mathcal{K}_1) = 2\pi \sum_{\alpha=1}^4 n_\alpha \mathbf{k}_\alpha \quad (14.42)$$

where \mathbf{k}_α are the reciprocal vectors of the tetrahedron. This is proved at once as follows: any vector \mathcal{K} may be expressed in terms of its components κ_α on the reciprocal base,² thus

$$\mathcal{K} = \sum_{\alpha=1}^4 \kappa_\alpha \mathbf{k}_\alpha \quad (14.43)$$

²It is important to notice that the components of any vector on the four vectors of the reciprocal base are determined modulo a real constant due to equation 14.10.

hence, equations 14.41 become

$$\sum_{\beta=1}^4 (\kappa_{1,\beta} - \kappa_{2,\beta}) \mathbf{k}_\beta \cdot \mathbf{r}_\alpha = 2\pi n_\alpha, \text{ for } \alpha = 1 \text{ to } 4$$

or, using property 14.10 to eliminate, for example, \mathbf{k}_4 ,

$$\sum_{\beta=1}^3 [(\kappa_{1,\beta} - \kappa_{1,4}) - (\kappa_{2,\beta} - \kappa_{2,4})] \mathbf{k}_\beta \cdot \mathbf{r}_\alpha = 2\pi(n_\alpha - n_4), \text{ for } \alpha = 1 \text{ to } 4$$

then, subtracting the fourth equation from the first three equations

$$\sum_{\beta=1}^3 [(\kappa_{1,\beta} - \kappa_{1,4}) - (\kappa_{2,\beta} - \kappa_{2,4})] \mathbf{k}_\beta \cdot (\mathbf{r}_\alpha - \mathbf{r}_4) = 2\pi(n_\alpha - n_4), \text{ for } \alpha = 1 \text{ to } 3$$

equation 14.6 is eventually used to reduce the above set to the following

$$(\kappa_{1,\alpha} - \kappa_{1,4}) - (\kappa_{2,\alpha} - \kappa_{2,4}) = 2\pi(n_\alpha - n_4), \text{ for } \alpha = 1 \text{ to } 3$$

which leads to equation 14.42. This result is important for wave analyses taking advantage of the four spacecraft, especially for k -filtering technique (see Chapter 3).

14.5.2 Characterisation of a Planar Discontinuity

Although normals to shocks and discontinuities are considered in Chapters 10 and 11, it is worth here emphasising the usefulness of reciprocal vectors for this matter. Let \mathbf{D} be a planar surface of discontinuity crossed by the cluster and represented at some time t_0 on Figure 14.3. The position vector of spacecraft S_α is $\mathbf{r}_\alpha(t)$ at time t and this spacecraft crosses surface \mathbf{D} at time t_α through point P_α , the position of which is \mathbf{r}_α^0 at time t_0 . Let us suppose that \mathbf{D} is not deformable, does not rotate, has velocity \mathbf{V} at time t_0 and moves with constant acceleration $\mathbf{\Gamma}$ during the interval of time necessary for the cluster to cross it. The unknown vectors \mathbf{V} , $\mathbf{\Gamma}$ and \mathbf{r}_α^0 , for $\alpha = 1$ to 4, satisfy the four equations:

$$\mathbf{r}_\alpha(t_\alpha) = \mathbf{r}_\alpha^0 + \mathbf{V}(t_\alpha - t_0) + \frac{1}{2}\mathbf{\Gamma}(t_\alpha - t_0)^2, \text{ for } \alpha = 1 \text{ to } 4 \quad (14.44)$$

The positions of the four spacecraft, taken for each of them at the time when it crosses surface \mathbf{D} , form a tetrahedron which does not correspond to any instantaneous configuration of the cluster. It is nevertheless convenient to use this fictitious configuration together with its associated barycentric base of reciprocal vectors.

Let us first consider the simplest case of the *uniform* motion for which no *a priori* knowledge of the normal $\hat{\mathbf{N}}$ to the discontinuity is required. The scalar product of equation 14.44, with $\mathbf{\Gamma} = \mathbf{0}$, by normal $\hat{\mathbf{N}}$ gives:

$$\hat{\mathbf{N}} \cdot \mathbf{r}_\alpha(t_\alpha) = \hat{\mathbf{N}} \cdot \mathbf{r}_\alpha^0 + \hat{\mathbf{N}} \cdot \mathbf{V}(t_\alpha - t_0) \quad (14.45)$$

Owing to the planarity assumption all scalar products $\hat{\mathbf{N}} \cdot \mathbf{r}_\alpha^0$ are equal, thus multiplying the above equation by \mathbf{k}_α and summing over the four values of α we get:

$$\sum_{\alpha=1}^4 \mathbf{k}_\alpha \hat{\mathbf{N}} \cdot \mathbf{r}_\alpha(t_\alpha) = V_n \sum_{\alpha=1}^4 \mathbf{k}_\alpha t_\alpha$$

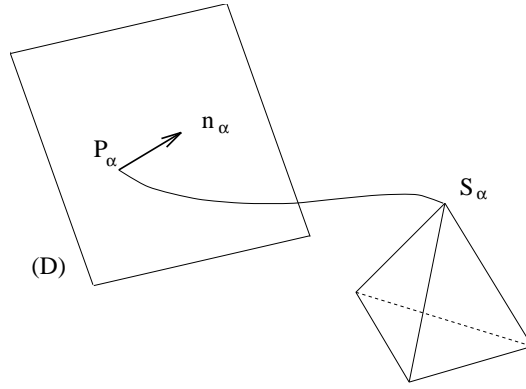


Figure 14.3: A surface of discontinuity \mathbf{D} that will be crossed by the cluster is represented at some time t_0 . Spacecraft S_α will cross it at time t_α through point P_α ; and $\hat{\mathbf{n}}_\alpha$ is the normal to the surface determined by spacecraft S_α . Each vertex of the tetrahedron is located at the position of the corresponding spacecraft when it crosses the surface, hence this tetrahedron does not correspond to any instantaneous configuration of the cluster.

where use has been made of equation 14.10. The reciprocal basis presently considered is the one associated to the fictitious tetrahedron that we have introduced, thus with the help of equation 14.12 the left-hand side of the above equation is just equal to $\hat{\mathbf{N}}$, which is eventually given by:

$$\hat{\mathbf{N}} = V_n \mathbf{Q}_1, \text{ where } \mathbf{Q}_1 = \sum_{\alpha=1}^4 \mathbf{k}_\alpha t_\alpha \quad (14.46)$$

The normal component of the relative velocity between the cluster of spacecraft and the discontinuity is given by the condition $\hat{\mathbf{N}} \cdot \hat{\mathbf{N}} = 1$, i.e.:

$$V_n = (\mathbf{Q}_1 \cdot \mathbf{Q}_1)^{-\frac{1}{2}} \quad (14.47)$$

This simple result is just the explicit solution of the problem posed by *Russell et al.* [1983] in their multi-spacecraft study of interplanetary shocks. Once V_n and $\hat{\mathbf{N}}$ have been computed with equations 14.46 and 14.47 it is necessary to check the consistency of the four equations 14.45 which mean that the four points, for $\alpha = 1$ to 4, having coordinates $(t_\alpha, \hat{\mathbf{N}} \cdot \mathbf{r}_\alpha(t_\alpha))$ should be aligned. A dramatic failure of this consistency test would mean that at least one of our assumptions, the planarity of the discontinuity or the uniformity of the motion, is not valid. A rotation or a deformation of the discontinuity, during the time interval needed by the cluster to cross it, are quite difficult to discuss. *Mottez and Chanteur* [1994] proposed a least square method, not related to the barycentric formalism, to determine the relative velocity of a non-planar discontinuity in the rest frame of the cluster, together with estimates of its principal directions and radii of curvature.

Let us now consider the case of the uniformly *accelerated* motion of a planar discontinuity. The normal $\hat{\mathbf{n}}_\alpha$ to the surface at the location and time of the crossing by spacecraft S_α is determined by measurements made onboard S_α only, for example by means of a minimum variance analysis of the magnetic fluctuations. When the mutual angles made by the

four normals are less than the experimental uncertainties on their directions it is reasonable to assume that the cluster crossed a planar discontinuity and that the relative motion did not involve a rotation. We will demonstrate at the end of this section that the estimations of \hat{N} and Γ are invariant in a time translation $t \rightarrow t + \tau$ and that $V \rightarrow V - \Gamma\tau$, as it should. Thus we choose $t_0 = 0$ for convenience in the following discussion which will be followed by a short discussion of the uniform motion case.

Derivation of the Estimations

Taking the scalar product of equation 14.44 with \hat{N} , for $t_0 = 0$, and using once more equation 14.10, we get the following system of equations:

$$\hat{N} \cdot \mathbf{r}_\alpha^0 + V_n t_\alpha + \frac{1}{2} \Gamma_n t_\alpha^2 = \hat{N} \cdot \mathbf{r}_\alpha(t_\alpha), \text{ for } \alpha = 1 \text{ to } 4$$

where V_n and Γ_n denote respectively the components of V and Γ along the normal. The first terms on the left-hand side of the above equations are equal to zero because of the planarity assumption, thus multiplying by \mathbf{k}_α and then summing over the four values of α gives, with the help of equation 14.12, the normal to the discontinuity in terms of vectors \mathbf{Q}_1 and \mathbf{Q}_2 :

$$\hat{N} = V_n \mathbf{Q}_1 + \Gamma_n \mathbf{Q}_2, \text{ where } \mathbf{Q}_1 = \sum_{\alpha=1}^4 t_\alpha \mathbf{k}_\alpha, \text{ and } \mathbf{Q}_2 = \frac{1}{2} \sum_{\alpha=1}^4 t_\alpha^2 \mathbf{k}_\alpha \quad (14.48)$$

The unknown parameters V_n and Γ_n are determined by a constrained least square procedure taking into account the fact that \hat{N} is a unit vector, i.e. by minimising the expression

$$\xi = \sum_{\alpha=1}^4 (\hat{n}_\alpha - \hat{N})^2 + 4(\Lambda - 1)(\hat{N}^2 - 1)$$

The Lagrange multiplier, written $4(\Lambda - 1)$ for later convenience, is determined by

$$\Lambda = \pm |v \mathbf{Q}_1 + \gamma \mathbf{Q}_2| \quad (14.49)$$

$$\text{where } v = \frac{Q_2^2 (N_{\text{mean}} \cdot \mathbf{Q}_1) - (\mathbf{Q}_1 \cdot \mathbf{Q}_2) (N_{\text{mean}} \cdot \mathbf{Q}_2)}{Q_1^2 Q_2^2 - (\mathbf{Q}_1 \cdot \mathbf{Q}_2)^2} \quad (14.50)$$

$$\gamma = \frac{Q_1^2 (N_{\text{mean}} \cdot \mathbf{Q}_2) - (\mathbf{Q}_1 \cdot \mathbf{Q}_2) (N_{\text{mean}} \cdot \mathbf{Q}_1)}{Q_1^2 Q_2^2 - (\mathbf{Q}_1 \cdot \mathbf{Q}_2)^2} \quad (14.51)$$

$$N_{\text{mean}} = \frac{1}{4} \sum_{\alpha=1}^4 \hat{n}_\alpha \quad (14.52)$$

The normal components of the velocity and the acceleration are

$$V_n = \frac{v}{\Lambda}, \text{ and } \Gamma_n = \frac{\gamma}{\Lambda} \quad (14.53)$$

Indeed looking for the extrema of ξ we have found two solutions for Λ , one of them corresponds to a minimum of ξ and the other one to a maximum; the right solution is selected by considering that the cosine of the angle between vectors \hat{N} and N_{mean} should be positive, i.e. that

$$\text{sign}(\Lambda) = \text{sign}[(v \mathbf{Q}_1 + \gamma \mathbf{Q}_2) \cdot N_{\text{mean}}] \quad (14.54)$$

The Uniform Motion Case

When vectors \mathbf{Q}_1 and N_{mean} are colinear, i.e. $N_{\text{mean}} = \lambda \mathbf{Q}_1$, it comes immediately from equations 14.49 and following that $\gamma = 0$ and $v = \lambda$. In that case

$$\hat{N} = \frac{N_{\text{mean}}}{|N_{\text{mean}}|}, \text{ and } V_n = \frac{1}{\mathbf{Q}_1 \cdot \hat{N}} \quad (14.55)$$

This result is consistent with our former discussion of the uniform motion case, especially with equations 14.46 and 14.47.

Changing the Origin of Time

Equation 14.10 is used to demonstrate that the time translation $t \rightarrow t + \tau$ leaves \mathbf{Q}_1 invariant and that $\mathbf{Q}_2 \rightarrow \mathbf{Q}_2 + \tau \mathbf{Q}_1$. Using these elementary properties of vectors \mathbf{Q}_1 and \mathbf{Q}_2 it is easily demonstrated that the following expressions are invariant when changing the origin of time

$$\begin{aligned} & Q_1^2 Q_2^2 - (\mathbf{Q}_1 \cdot \mathbf{Q}_2)^2 \\ & Q_1^2 (N_{\text{mean}} \cdot \mathbf{Q}_2) - (\mathbf{Q}_1 \cdot \mathbf{Q}_2) (N_{\text{mean}} \cdot \mathbf{Q}_1) \end{aligned}$$

It follows immediately that γ is an invariant of the time translation. Then it is readily demonstrated that $v \rightarrow v - \gamma \tau$ and that the vector $v \mathbf{Q}_1 + \gamma \mathbf{Q}_2$ is invariant. Being equal to plus or minus the norm of the latter vector the Lagrange multiplier Λ is also invariant. To summarise we have demonstrated the invariance of the estimations of \hat{N} and Γ_n given by equations 14.48 to 14.54 when $t \rightarrow t + \tau$, and that $V_n \rightarrow V_n - \Gamma_n \tau$ as expected.

14.6 Conclusions

The aim of this chapter was to introduce in a detailed and self-contained way the barycentric coordinates and the reciprocal vectors of a tetrahedron and to demonstrate the power of these geometrical concepts to analyse data provided by a four spacecraft mission. All the applications presented in this chapter, the estimation of gradients, the analysis of discontinuities or the spatial aliasing condition, demonstrate that the barycentric formalism allows an entirely symmetric handling of the data provided by the four spacecraft, leading to explicit results in which the four spacecraft play identical roles. Moreover the barycentric formalism allows a complete theoretical analysis of the various kinds of errors affecting the estimates of large scale gradients, for any shape of the tetrahedron formed by the spacecraft. Equation 14.31 shows that the knowledge of the covariance matrix $\langle \delta \mathbf{r} \delta \mathbf{r}^T \rangle$ of the uncertainties on the spacecraft positions is mandatory to estimate the geometrical errors affecting the evaluation of the gradients. The truncation errors of the barycentric

estimators of magnetic gradients will generally lead to errors in the estimations of the large scale electric currents; this has been demonstrated here for a thick and planar current sheet and this important point will be again investigated in Chapter 15 for a dipole field. Extensive simulations, making use of the linear barycentric estimators among others, are discussed in Chapter 16 with respect to the accuracy of the current determination for simple models of current tubes [see also *Robert and Roux, 1993*]. The analysis of the truncation errors has emphasised that the linear barycentric estimator of $\nabla \cdot \mathbf{B}$ is generally not equal to zero due to the linear interpolation of the magnetic field, but there is no general relationship between the truncation errors affecting $\nabla \cdot \mathbf{B}$ and \mathbf{J} . Although we have often focused on magnetic measurements, most of the general results derived here could be applied to other fields, for example the plasma velocity moments (see Chapter 17).

Bibliography

- Chanteur, G. and Mottez, F., Geometrical tools for Cluster data analysis, in *Proc. International Conf. "Spatio-Temporal Analysis for Resolving plasma Turbulence (START)"*, Aussois, 31 Jan.–5 Feb. 1993, ESA WPP-047, pp. 341–344, European Space Agency, Paris, France, 1993.
- Dunlop, M. W., Balogh, A., Southwood, D. J., Elphic, R. C., Glassmeier, K.-H., and Neubauer, F. M., Configurational sensitivity of multipoint magnetic field measurements, in *Proceedings of the International Workshop on "Space Plasma Physics Investigations by Cluster and Regatta"*, Graz, Feb. 20–22, 1990, ESA SP-306, pp. 23–28, European Space Agency, Paris, France, 1990.
- Mottez, F. and Chanteur, G., Surface crossing by a group of satellites: a theoretical study, *J. Geophys. Res.*, **99**, 13 499–13 507, 1994.
- Neubauer, F. M. and Glassmeier, K. H., Use of an array of satellites as a wave telescope, *J. Geophys. Res.*, **95**, 19 115–19 122, 1990.
- Robert, P. and Roux, A., Influence of the shape of the tetrahedron on the accuracy of the estimation of the current density, in *Proc. International Conf. "Spatio-Temporal Analysis for Resolving plasma Turbulence (START)"*, Aussois, 31 Jan.–5 Feb. 1993, ESA WPP-047, pp. 289–293, European Space Agency, Paris, France, 1993.
- Russell, C. T., Mellott, M. M., Smith, E. J., and King, J. H., Multiple spacecraft observations of interplanetary shocks: Four spacecraft determination of shock normals, *J. Geophys. Res.*, **88**, 4739–4748, 1983.

— 15 —

Spatial Interpolation for Four Spacecraft: Application to Magnetic Gradients

GÉRARD CHANTEUR
CETP-CNRS
Vélizy, France

CHRISTOPHER C. HARVEY
Observatoire de Paris-Meudon
Meudon, France

15.1 Introduction

The primary purpose of a multi-spacecraft mission such as Cluster is to distinguish between temporal and spatial variations by means of four or more spacecraft. An obvious question is how, precisely, the gradients of the spatial variations should be determined from such multipoint measurements. Some of the early studies related to the Cluster mission aimed to evaluate the components of the electric current density (see Chapter 16), i.e. of peculiar combinations of various components of the gradient of the magnetic field. But the first attempt to devise a method of determining the gradient of any field was due to *Chanteur and Mottez* [1993], who introduced barycentric coordinates for the linear interpolation of scalar and vector field quantities. This work has been considerably extended and a detailed and self-contained presentation has been made in Chapter 14; linear and quadratic estimators of the gradient of a field have been defined. Chapter 12 introduced two other linear estimators derived by least squares minimisation, constrained or not by the solenoidal condition.

The present chapter aims to evaluate the accuracy of these estimators by applying them to simulated data; it requires an understanding of Chapters 12 and 14. In Section 15.2 we prove the mathematical equivalence of the unconstrained least squares linear estimator with the linear barycentric estimator. In Section 15.3 a dipole field model is used to illustrate analytically the nature and magnitude of the errors caused by using linear interpolation to approximate a non-uniform field. In Section 15.4 we compare the linear barycentric estimator and the least squares estimator constrained by the solenoidality condition; the latter will be briefly referred to as the solenoidal estimator. The estimates of the gradient of the magnetic field are computed by both methods for simulated observations obtained by using a model of the Cluster orbit together with the *Tsyganenko* [1987] model of the geomagnetic field. The exact gradient of the simulated observations is computed directly from the model when preparing the simulated data sets. This comparison gives further insight into the difficulties related to the determination of gradients, and their likely variation as the tetrahedron geometry evolves along a typical spacecraft orbit. In the last section we discuss the smoothing of the data which is required before using it to determine the gradient. We also recall the result demonstrated in Chapter 14 (equation 14.31 on page 358) that the standard error in the determination of the gradient is intimately related

not only to the properties of the reciprocal tensor (or of its inverse, the volumetric tensor), but also to the properties of the covariance matrix of the error in the determination of the spacecraft position.

15.1.1 Statement of the Problem

Cluster can determine gradients in two different ways, depending upon the scale size of the spatial variation of the observed parameter.

1. Small-scale structures are spatial variations with scale size small compared with the inter-spacecraft separation distances. They are identifiable in the sense that the differences in their time of observation at the four spacecraft may be determined, for example, by cross-correlation. Such observations may be interpreted in terms of a one-dimensional spatial structure moving along the direction of its normal, and both the direction of the normal and the speed of the motion may be determined from the observations. This model is suitable for the study of shocks, discontinuities, and the like.
2. Large-scale structures have a spatial scale large compared with the spacecraft separation distances. Cross-correlation does not yield any meaningful result. But if there is a difference in the values of a parameter measured on the four spacecraft, it is possible to determine the corresponding large-scale gradient. In particular, an important objective of the Cluster project is to infer the mean current density from magnetic measurements made at different points.

These two cases are relatively clear. The situation is more complicated when the spatial scale is comparable with the inter-spacecraft separation, or when the tetrahedron is very anisotropic, and further study is required.

15.2 Relationship Between Homogeneous Least Squares and Barycentric Methods

We now show that the unconstrained homogeneous least squares method and the linear barycentric method described in Chapter 14 are equivalent. We do this by first showing that the reciprocal tensor defined by equation 14.29 (page 356)

$$\mathbf{K} = \sum_{\alpha=1}^4 \mathbf{k}_{\alpha} \mathbf{k}_{\alpha}^T$$

is equal to $\frac{1}{4}$ of the inverse of the volumetric tensor.

15.2.1 Relationship Between the Reciprocal and Volumetric Tensors

The volumetric tensor is defined by equation 12.23 (page 315), where \mathbf{r}_b is the position vector of the barycentre of the tetrahedron. The reciprocal and volumetric tensors are

symmetric. A short lemma is needed concerning the scalar product

$$\mathbf{k}_\alpha \cdot (\mathbf{r}_\beta - \mathbf{r}_b) = \frac{1}{4} \sum_{\gamma} \mathbf{k}_\alpha \cdot (\mathbf{r}_\beta - \mathbf{r}_\gamma)$$

with the help of equation 14.6 (page 350), we obtain

$$\mathbf{k}_\alpha \cdot (\mathbf{r}_\beta - \mathbf{r}_b) = (\mathbf{r}_\beta - \mathbf{r}_b) \cdot \mathbf{k}_\alpha = \delta_{\alpha\beta} - \frac{1}{4} \quad (15.1)$$

The dyadic notation allows us to compute the product of the two tensors, thus:

$$4 \mathbf{K} \mathbf{R} = \sum_{\alpha} \sum_{\beta} \mathbf{k}_\alpha \mathbf{k}_\alpha^T (\mathbf{r}_\beta - \mathbf{r}_b) (\mathbf{r}_\beta - \mathbf{r}_b)^T$$

The inner dyad $\mathbf{k}_\alpha^T (\mathbf{r}_\beta - \mathbf{r}_b)$ is the scalar product of equation 15.1, so that

$$4 \mathbf{K} \mathbf{R} = \sum_{\alpha} \mathbf{k}_\alpha (\mathbf{r}_\alpha - \mathbf{r}_b)^T$$

Then equations 14.11 and 14.10 (page 352), and the symmetry of both tensors, allow to conclude that:

$$\mathbf{K} \mathbf{R} = \frac{1}{4} \mathbf{I} = \mathbf{R} \mathbf{K} \quad (15.2)$$

15.2.2 Identity of the HLS and Linear Barycentric Estimators of the Gradient

When the homogeneous least squares method is applied without imposing the condition of solenoidality (of Section 12.3.2), the resulting estimator $\text{HLSG}[\mathbf{v}]$ of the vector field \mathbf{v} minimises the scalar quantity (see equation 14.2 on page 350 for the definition of $\mathbf{G}^T[\mathbf{v}]$)

$$S = \sum_{\alpha} \sum_{\beta} \mathbf{w}^T \mathbf{w}, \text{ where } \mathbf{w} = \mathbf{G}^T[\mathbf{v}] (\mathbf{r}_\alpha - \mathbf{r}_\beta) - (\mathbf{v}_\alpha - \mathbf{v}_\beta)$$

It is obvious from definition 14.13 (page 352) that the linear barycentric estimator $\text{LG}^T[\mathbf{v}]$ satisfies:

$$\text{LG}^T[\mathbf{v}] (\mathbf{r}_\alpha - \mathbf{r}_\beta) = (\mathbf{v}_\alpha - \mathbf{v}_\beta)$$

hence $\text{LG}^T[\mathbf{v}]$ exactly cancels \mathbf{w} and S and

$$\text{HLSG}[\mathbf{v}] = \text{LG}[\mathbf{v}]$$

15.3 Truncation Errors of LG for a Magnetic Dipole

In this section we illustrate the inherent limitations of any linear (homogeneous least squares, barycentric, or other) estimation of the gradient. We do this by evaluating analytically the linear estimator obtained from a simple dipole magnetic field sampled at four

locations. Linear interpolation between these four field samples generally leads to an estimation of the gradient tensor which is asymmetric, with a trace different from zero; i.e., to the appearance of non-existent current and local violation of the conservation of the magnetic flux. Detailed expressions will be given for a regular tetrahedron.

These results explain, for example, why forcing the divergence of the estimated magnetic field to be zero (the method of Section 12.3.2) does not necessarily yield the best result, as will be discovered in Section 15.4.

15.3.1 The Dipole Field and its Gradient

The magnetic field created at location $\mathbf{R} = R\hat{\mathbf{e}}$ by the magnetic dipole $\mathcal{M} = \mathcal{M}\hat{\mathbf{u}}$ writes:

$$\mathbf{B}_{\text{dipole}}(\mathbf{R}) = -\nabla \left(R^{-3} \mathcal{M} \cdot \mathbf{R} \right) = \mathcal{M} R^{-3} \left(3\hat{\mathbf{e}} \hat{\mathbf{u}}^T \hat{\mathbf{e}} - \hat{\mathbf{u}} \right) \quad (15.3)$$

where $\hat{\mathbf{e}}$ and $\hat{\mathbf{u}}$ are unit vectors. This is obtained from definition 14.1 (page 350) of the gradient of a scalar field and with the help of the identity $R \, dR = \mathbf{R} \cdot d\mathbf{R}$. From definition 14.2 (page 350) of the gradient of a vector field we obtain:

$$\frac{R^4}{3\mathcal{M}} \mathbf{G}_{\text{dipole}}^T = (\hat{\mathbf{e}} \cdot \hat{\mathbf{u}}) (\mathbf{I} - 5\hat{\mathbf{e}} \hat{\mathbf{e}}^T) + \hat{\mathbf{e}} \hat{\mathbf{u}}^T + \hat{\mathbf{u}} \hat{\mathbf{e}}^T \quad (15.4)$$

where \mathbf{I} is the unit tensor. The gradient of the dipole field is obviously symmetric and its trace is equal to zero as it should.

15.3.2 Linear Estimation of the Gradient of the Dipole Field

The linear estimation of the gradient of the dipole field involves the magnetic field vectors measured by the four spacecraft. We assume that the characteristic size a of the tetrahedron is much less than R , the distance from the barycentre to the dipole, and we develop measured (dipole) vector magnetic field in the vicinity of the barycentre up to terms of second order in $(a/R)^2$. The vector position of spacecraft α is written $\mathbf{r}_\alpha = R(\hat{\mathbf{e}} + \boldsymbol{\varepsilon}_\alpha)$, which implies that the magnitude of $\boldsymbol{\varepsilon}_\alpha$ is of order a/R and $\Sigma \boldsymbol{\varepsilon}_\alpha = \mathbf{0}$. The linear estimator of the magnetic field gradient writes up to second order in a/R :

$$\begin{aligned} \text{LG}^T[\mathbf{B}] &= \mathbf{G}_{\text{dipole}}^T + \frac{a}{R} \frac{3\mathcal{M}}{R^4} \left\{ (1 - 5\hat{\mathbf{e}} \cdot \hat{\mathbf{u}}) \mathbf{T} \right. \\ &\quad \left. + (\hat{\mathbf{u}} - 5\hat{\mathbf{e}} \hat{\mathbf{u}}^T \hat{\mathbf{e}}) \mathbf{C}_0^T - 5(\hat{\mathbf{u}} - 7\hat{\mathbf{e}} \hat{\mathbf{u}}^T \hat{\mathbf{e}}) \mathbf{C}_1^T - 5\hat{\mathbf{e}} \mathbf{C}_2^T \right\} \end{aligned} \quad (15.5)$$

with the following definitions:

$$\mathbf{T} = \frac{R^2}{a} \sum_{\alpha=1}^4 (\hat{\mathbf{e}} \cdot \boldsymbol{\varepsilon}_\alpha) \boldsymbol{\varepsilon}_\alpha \mathbf{k}_\alpha^T \quad (15.6)$$

$$\mathbf{C}_0 = \frac{R^2}{2a} \sum_{\alpha=1}^4 (\boldsymbol{\varepsilon}_\alpha)^2 \mathbf{k}_\alpha \quad (15.7)$$

$$\mathbf{C}_1 = \frac{R^2}{2a} \sum_{\alpha=1}^4 (\hat{\mathbf{e}} \cdot \boldsymbol{\varepsilon}_\alpha)^2 \mathbf{k}_\alpha \quad (15.8)$$

$$\mathbf{C}_2 = \frac{R^2}{a} \sum_{\alpha=1}^4 (\hat{\mathbf{e}} \cdot \boldsymbol{\varepsilon}_\alpha)(\hat{\mathbf{u}} \cdot \boldsymbol{\varepsilon}_\alpha) \mathbf{k}_\alpha \quad (15.9)$$

As expected the first order contribution to $\text{LG}^T[\mathbf{B}]$ is equal to the exact gradient of the dipole field at the barycentre, but the general second order contribution is not symmetric and its trace is not equal to zero.

15.3.3 Truncation Errors for a Regular Tetrahedron

Consider the regular tetrahedron defined in Section 14.3.3 and let (u_x, u_y, u_z) and (e_x, e_y, e_z) be respectively the cartesian components of $\hat{\mathbf{u}}$ and $\hat{\mathbf{e}}$ in the reference frame of the tetrahedron. With these notations, the auxiliary vectors defined by equations 15.7 to 15.9 write:

$$\mathbf{C}_0 = \mathbf{0} \quad (15.10)$$

$$\mathbf{C}_1 = -\{e_y e_z \hat{\mathbf{x}} + e_z e_x \hat{\mathbf{y}} + e_x e_y \hat{\mathbf{z}}\} \quad (15.11)$$

$$\mathbf{C}_2 = -\{(e_y u_z + u_y e_z) \hat{\mathbf{x}} + (e_z u_x + u_z e_x) \hat{\mathbf{y}} + (e_x u_y + u_x e_y) \hat{\mathbf{z}}\} \quad (15.12)$$

where $\hat{\mathbf{x}}, \hat{\mathbf{y}}, \hat{\mathbf{z}}$ are the unit vectors of the reference frame. The tensor \mathbf{T} defined by equation 15.6 is equal to:

$$\mathbf{T} = -16 \begin{pmatrix} 0 & e_z & e_y \\ e_z & 0 & e_x \\ e_y & e_x & 0 \end{pmatrix} \quad (15.13)$$

From all the second order contributions to the errors of truncation, this term is the only one which does not contribute to false currents and to the non-conservation of the magnetic flux. Even in the simplest case of a regular cluster the truncation errors given by equation 15.5 to 15.9 are too much intricate to be really useful for analytical considerations. Nevertheless it is possible to obtain a few simple results, for example when the cluster lies in the equatorial plane of the dipole (in that case $\hat{\mathbf{e}} \cdot \hat{\mathbf{u}} = 0$), the linear estimator of $\nabla \cdot \mathbf{B}$ is, accordingly to equation 14.16 (page 353):

$$\frac{R}{B} \text{LD}[\mathbf{B}] = \frac{45a}{R} (u_x e_y e_z + u_y e_z e_x + u_z e_x e_y) \quad (15.14)$$

Thus the linear estimation of the divergence of \mathbf{B} can be equal to zero, for example for $u_z = e_x = 1$, but that is usually not the case.

15.4 Comparison of the Methods Using Simulated Data

In this section we test the barycentric and solenoidal least-squares estimators of the gradient of the magnetic field on simulated Cluster data obtained along predicted Cluster orbits embedded in the Tsyganenko-87 magnetic field model [Tsyganenko, 1987]. The orbits have been provided by the European Space Operations Centre (ESOC, Darmstadt, Germany) in the geocentric equatorial inertial frame of reference. Taking into account the Earth's rotation and orbital motion [Coeur-Joly et al., 1995], the magnetic data were computed in geocentric solar ecliptic (GSE) coordinates. An exhaustive investigation is

beyond the scope of this article, so we have selected a single Cluster orbit, with apogee in the geotail and hence the whole orbit lying inside the magnetosphere for the moderate index $Kp = 2$ which we have chosen. For such quiet conditions the gradient of the magnetic field in the outer magnetosphere has a characteristic scale size which is larger than the tetrahedron itself, and we may expect our determination to be a good approximation to the true gradient; the situation is much less favourable near perigee, where steeper gradients are seen with a tetrahedron considerably elongated in the direction of the spacecraft motion due to large orbital velocity. The precision of the gradient determination depends also upon the quality of intercalibration of the relevant experiments, and for some Cluster experiments, for example the magnetic field and the bulk flow velocity of the low energy plasma, this is critical. This point has been investigated by *Khurana et al. [1996]* for the dc magnetic field and will not be addressed here.

15.4.1 Evolution of the Tetrahedron along the Orbit

The geometry of the cluster of spacecraft at a given time can be characterised by various geometrical factors which have been introduced and discussed in Chapter 13. For this study we have selected six of these geometrical factors: the elongation E and planarity P defined in Section 13.3.3, and the four parameters Q_{RR} , Q_{SR} , Q_{R8} , Q_{GM} which are defined in Section 13.3.1.

Figure 15.1 shows the variations of these geometrical factors of the tetrahedron along one orbit between two successive apogees. It is worth noticing that the data are not periodic owing to the drift of the orbits in the GSE frame of reference. Time is normalised to the orbital period, and perigee has been chosen to occur at $t = 0.5$ so that apogees correspond to $t = 0$ and 1. This particular satellite configuration is isotropic at $t = 0.975$, shortly before apogee; at that time E and P are equal to zero and the four Q factors are equal to one. The top panel shows the elongation E (dotted curve) and the planarity P (solid curve). The centre panel shows the variations of Q_{SR} (dotted curve) and Q_{RR} (solid curve), and the bottom panel displays Q_{GM} (dotted curve) and Q_{R8} (solid curve). The factors Q_{RR} and Q_{R8} have values close to zero, meanwhile the planarity P is close to one, at times $t = 0.14, 0.42, 0.53$ and 0.67 ; this is indicative of a degenerate configuration of the spacecraft which are coplanar. The variations of Q_{SR} and Q_{GM} are smoother. The low values of Q_{RR} and Q_{R8} , as well as the high value of the elongation in the time interval extending from 0.5 to .7 indicate an unfavourable configuration of the cluster (either flattened when P is close to one, or elongated when P significantly differs from one). The shaded bars emphasise the time intervals during which the numerical accuracy is degraded due to the flattening of the cluster; this point will be made more precise at the end of the next section, especially by equation 15.16 (page 379) and the following lines.

15.4.2 Estimations of $\mathbf{G}[\mathbf{B}]$

In this section we compare the different estimations of the gradient $\mathbf{G}[\mathbf{B}]$ along the selected orbit with the “true” gradient of the Tsyganenko magnetic field model. This latter gradient was calculated at the same time as the field itself. We compared two finite difference methods, of respectively second and fourth order with respect to the spatial increment, which was 200 km. The two methods gave results which are indistinguishable

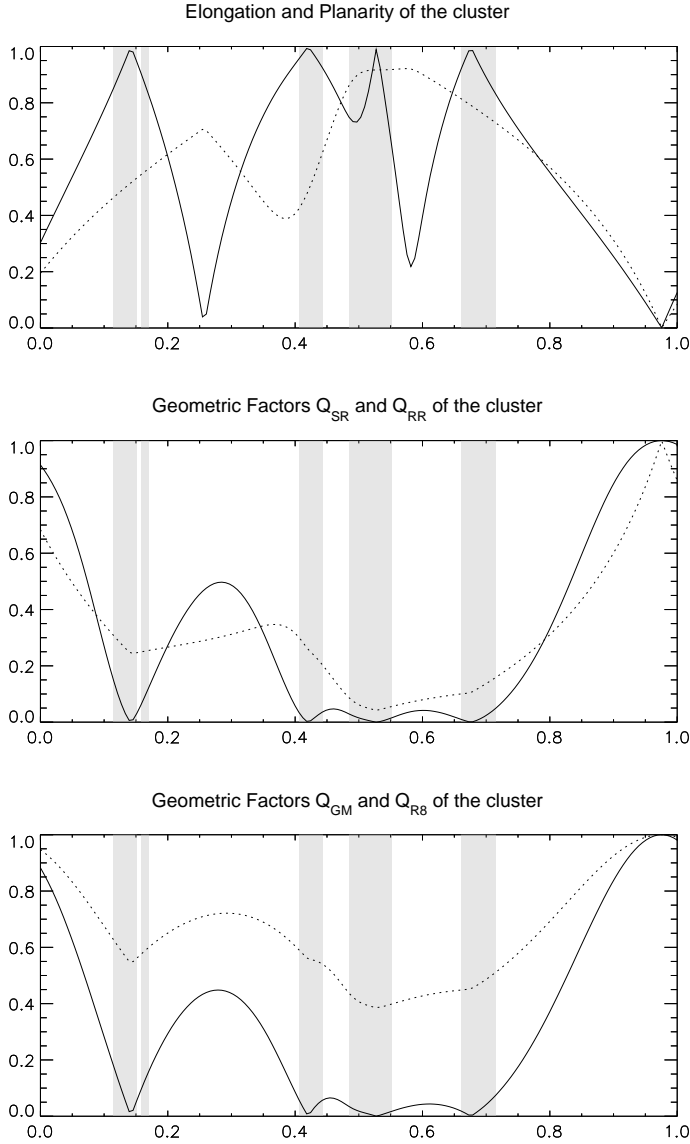


Figure 15.1: Characterisation of the tetrahedron along the orbit. The upper panel shows the variations with normalised time of the elongation E (dotted curve) and planarity P (solid curve) defined in Section 13.3.3. The centre panel shows the variations of Q_{SR} (dotted curve) and Q_{RR} (solid curve). The lower panel shows the variations of Q_{GM} (dotted curve) and Q_{R8} (solid curve). See Section 13.3.1 for the definitions of these geometrical factors. The shaded bars delimit intervals of degraded numerical accuracy (see Section 15.4.3).

for the present purposes. The fourth order method was used to compute what we will call the “true” gradient.

In order to compare the different estimations of $\mathbf{G}[\mathbf{B}]$ with the true gradient, a series of plots (Figures 15.2–15.10), present the variations of the components of $\mathbf{G}[\mathbf{B}]$, or of various estimates of these components, versus the normalised time in the GSE coordinate system. The various components of $\mathbf{G}[\mathbf{B}]$ have large variations, therefore, instead of the components $\partial_p B_q$ themselves (with $p, q = x, y$ or z), we have plotted the quantities

$$\frac{R_{\text{mean}}}{B_{\text{mean}}} (\partial_p B_q)$$

where R_{mean} and B_{mean} are the respective magnitudes of

$$R_{\text{mean}} = \frac{1}{4} \sum_{\alpha=1}^4 r_{\alpha}, \text{ and } B_{\text{mean}} = \frac{1}{4} \sum_{\alpha=1}^4 B_{\alpha}$$

Figures 15.2–15.4 compare the linear estimates, with and without the condition $\nabla \cdot \mathbf{B} = 0$, with each other and with the “true” components of $\mathbf{G}[\mathbf{B}]$. The “true” gradient is plotted as a solid grey curve, while the solenoidal least-squares and the barycentric linear estimates are respectively represented by the dotted and solid black curves. The unconstrained least-squares and barycentric linear estimations being identical, as shown in Section 15.2.2, the solid black lines represent estimates computed from both methods. It can be seen that large deviations of the estimated quantities occur during time intervals marked by the shaded bars and that the enforcement of the solenoidality condition generally does not improve the estimation of the gradient of the magnetic field. The deviations from the exact components are due to the errors of truncation as demonstrated by computing the quadratic barycentric estimator of the gradient. The exact quadratic estimator defined in Section 14.2.2 will not be computable with a four spacecraft mission but the simulation of the data allows one to compute the supplementary magnetic field vectors required to build the approximate estimator discussed in the same section. Figures 15.5–15.7 display the results provided by quadratic interpolation, which show a significant improvement of the estimates even throughout most of the singular time intervals marked by shaded bars. The light grey curves represent the exact components, and the black curves their quadratic estimates. The predicted components of $\mathbf{G}[\mathbf{B}]$ are quite accurate along the whole orbit except near perigee for components involving partial derivatives with respect to coordinate y (Figure 15.6).

The large deviations from the exact components around times of planarity of the cluster are due to the amplification of the truncation errors by singular configurations of the spacecraft. When the tetrahedron flattens, the largest of the reciprocal vectors is normal to the singular plane configuration and diverges as the inverse of the thickness of the tetrahedron. In order to demonstrate this amplification we subtract from the reciprocal vectors their singular components (parallel to the largest reciprocal vector) in the vicinity of the planar configurations according to the following formulas:

$$\mathbf{k}_{\alpha, \text{corrected}} = \mathbf{k}_{\alpha} - \gamma (\mathbf{k}_{\alpha} \cdot \hat{\mathbf{k}}_{\text{max}}) \hat{\mathbf{k}}_{\text{max}} \quad (15.15)$$

where $\hat{\mathbf{k}}_{\text{max}}$ is the unit vector parallel to the largest reciprocal vector and γ is a numerical factor which should be equal to zero almost everywhere except in the vicinity of the singular configurations where it should be equal to one. For any tetrahedron the sum of the

reciprocal vectors is equal to zero, but as will be shown below the numerical inaccuracy increases in the vicinity of a planar cluster and the magnitude of this sum is peaked around the singular times, although remaining small. Thus a practical choice for the function γ is:

$$\gamma = \left[\tanh \left((\mathbf{k}_1 + \mathbf{k}_2 + \mathbf{k}_3 + \mathbf{k}_4)^2 / S_0 \right) \right]^2 \quad (15.16)$$

where S_0 is a reference value chosen after inspection of the numerical values of $(\mathbf{k}_1 + \mathbf{k}_2 + \mathbf{k}_3 + \mathbf{k}_4)^2$; in the present case $S_0 = 10^{-18}$, when measuring the components of the reciprocal vectors in km^{-1} . The time intervals during which $\gamma \geq 0.1$ are marked by vertical shaded bars in all figures of this section. This corrected reciprocal base has been used to compute a corrected linear barycentric estimator of the gradient of the magnetic field. Figures 15.8–15.10 are to be compared with Figure 15.2–15.4; the large errors affecting the ∂_x and ∂_y components around $t = 0.53$ have been drastically reduced by this correction, as well as errors affecting ∂_x and ∂_z components near $t = 0.67$. This correction greatly reduces the amplification of the truncation errors by singular reciprocal bases, but not enough to provide reliable estimates during the singular intervals marked by shaded bars.

15.4.3 Quality of the Estimations

Numerical Precision

Although we do not know any *a priori* criterion of quality for the linear estimations, it is possible to define useful indicators of accuracy for both methods. In the barycentric method, the sum of the four reciprocal vectors of the tetrahedron formed by the spacecraft is theoretically equal to zero, but as this property is not used by the computational algorithm, it can be used as a numerical check. Hence an indicator of numerical accuracy is defined by:

$$\mathcal{Q}_{bary} = \log_{10} \left((\mathbf{k}_1 + \mathbf{k}_2 + \mathbf{k}_3 + \mathbf{k}_4)^2 \right) \quad (15.17)$$

and is represented by the solid black curve in the centre and bottom panels of Figure 15.11 versus time. Measuring components of the reciprocal vectors in km^{-1} and using simple precision IEEE arithmetic with words of 32 bits the background value of this indicator is slightly above -20 , but near the critical times the indicator increases and the peak values are greater than -15 , up to -9 , at the critical times for which the spacecraft are coplanar.

The method of least-squares requires the inversion of the volumetric tensor, it is thus natural to check how close to the unit tensor is the product of the volumetric tensor by its numerically computed inverse. Hence an empirical indicator of the numerical accuracy is defined by:

$$\mathcal{Q}_{LS} = \sum_{i \neq j} \left| \left(\mathbf{R} \mathbf{R}^{-1} \right)_{i,j} \right| \quad (15.18)$$

the base 10 logarithm of the sum of the absolute values of the non-diagonal elements of this product. The variations of this indicator along the orbit are represented by the solid black curve on the upper panel of Figure 15.11; the minimal value is of the order of -16 and peak values reach -11 exactly at the same critical times found by the former indicator.

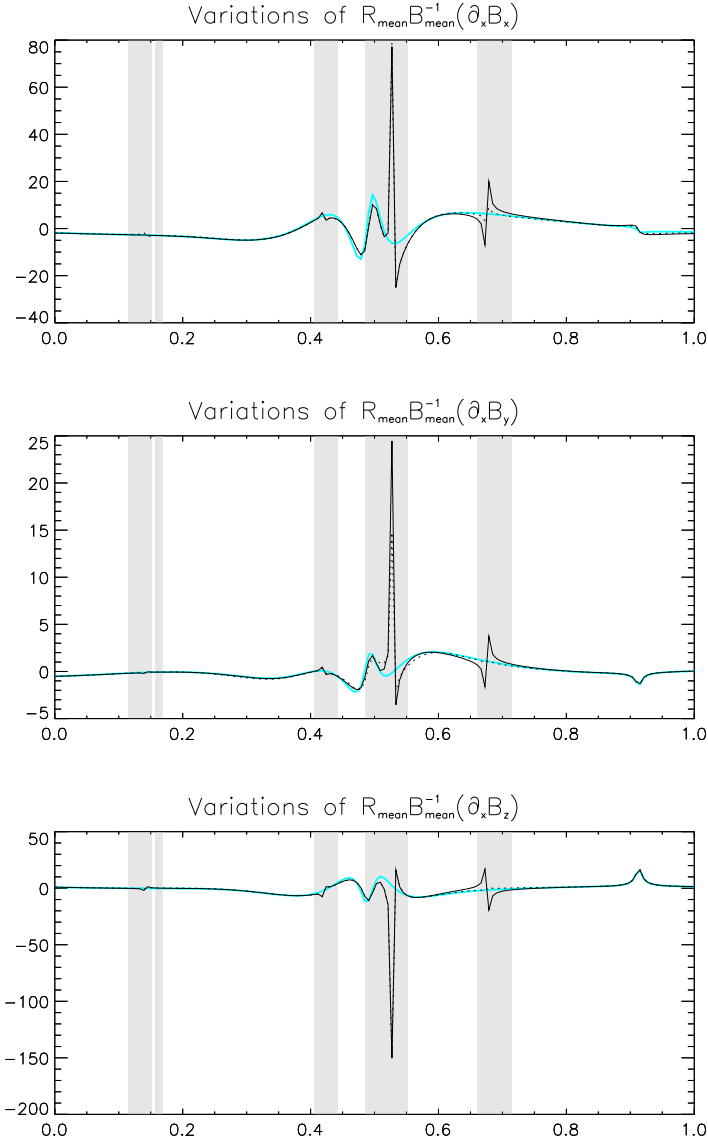


Figure 15.2: Variations of the linear estimations of the components $R_{\text{mean}} B_{\text{mean}}^{-1} \partial_x \mathbf{B}$ along a “tail orbit” of Cluster versus time normalised to the orbital period. The least-squares estimates, taking into account the solenoidal constraint $\nabla \cdot \mathbf{B} = 0$, are represented by the dotted black curves, while the barycentric, identical to the least-squares estimate without the solenoidal condition, are displayed by solid black curves. The solid grey curves are the true components of $\mathbf{G}[\mathbf{B}]$ for the Tsyganenko-87 model. The time resolution is 20 minutes.

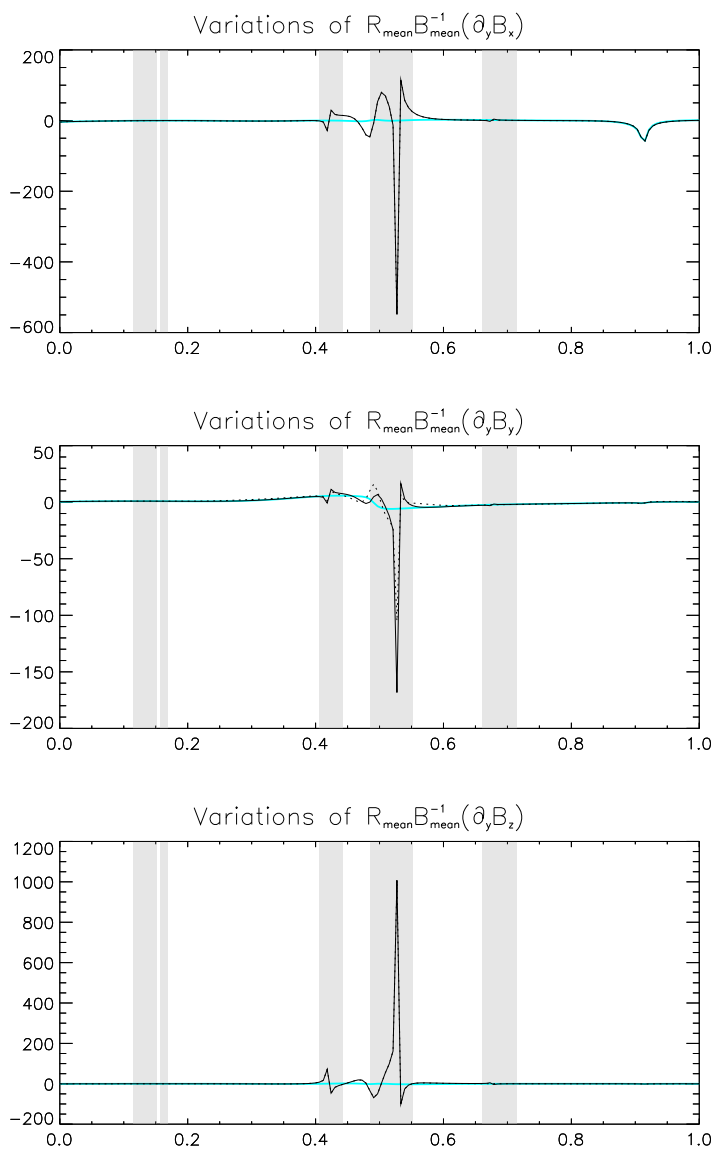


Figure 15.3: As Figure 15.2 but for the components $R_{\text{mean}} B_{\text{mean}}^{-1} \partial_y \mathbf{B}$.

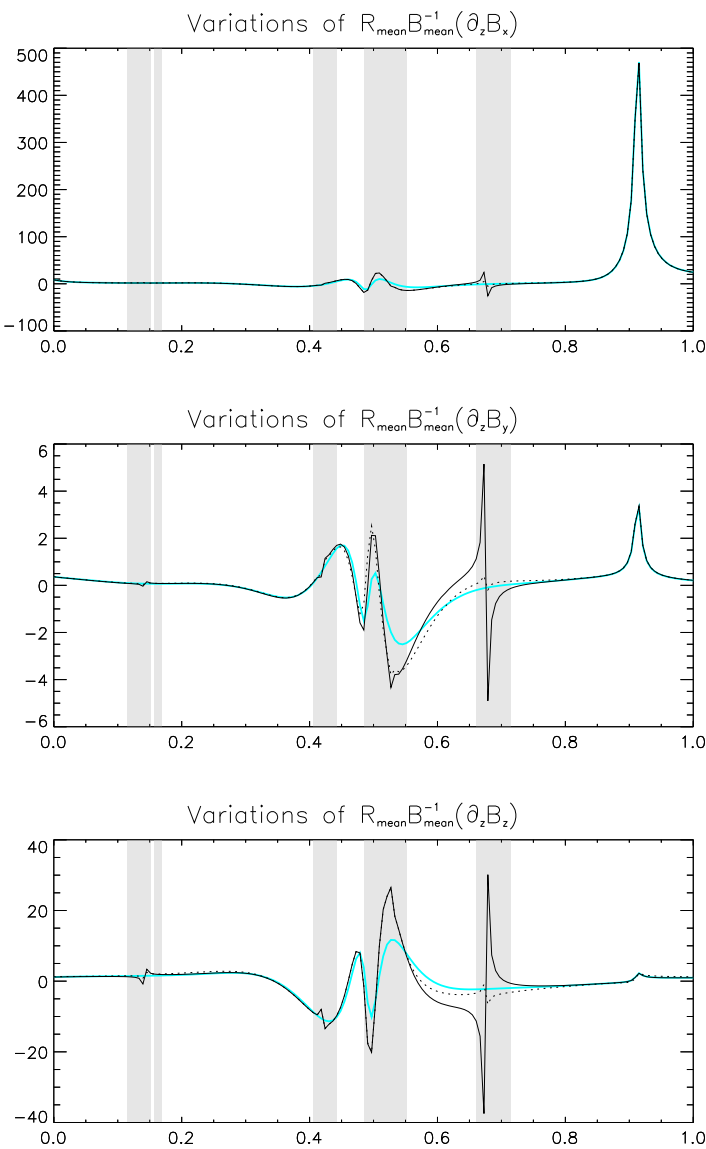


Figure 15.4: As Figure 15.2 but for the components $R_{\text{mean}} B_{\text{mean}}^{-1} \partial_z \mathbf{B}$.

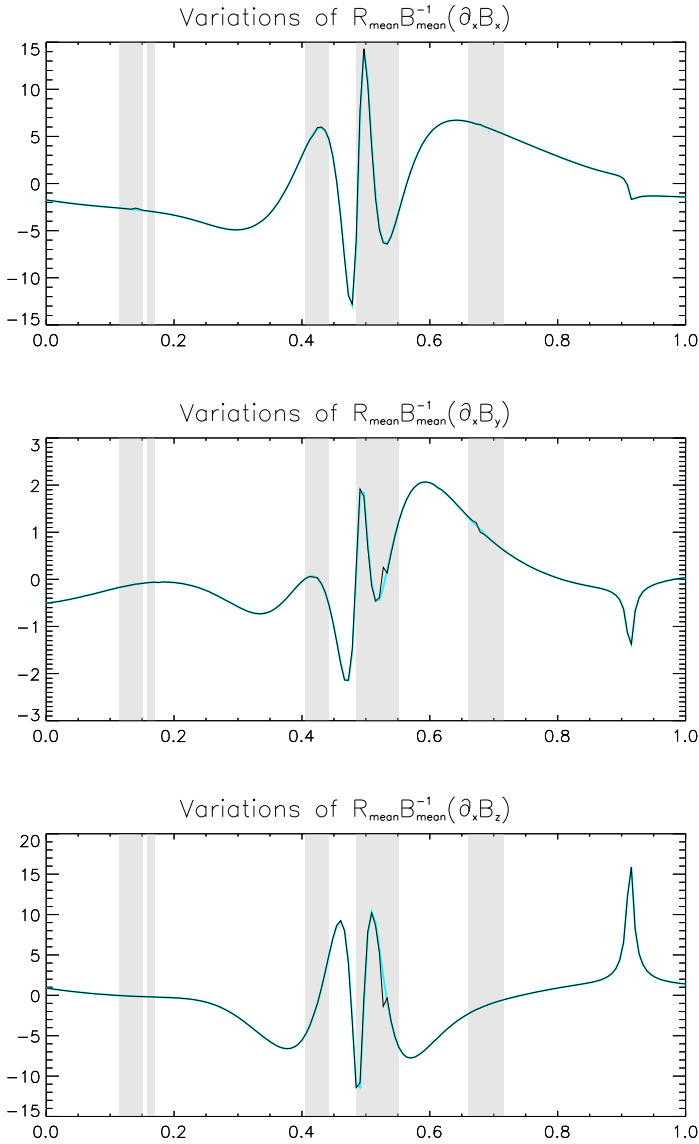


Figure 15.5: Components $R_{\text{mean}} B_{\text{mean}}^{-1} \partial_x \mathbf{B}$ estimated through a quadratic interpolation of \mathbf{B} for a time resolution of 20 minutes. The quadratic barycentric estimations are displayed as black curves, while the true components for the Tsyganenko-87 model, represented by the grey curves, are most of the time hidden by the estimated ones.

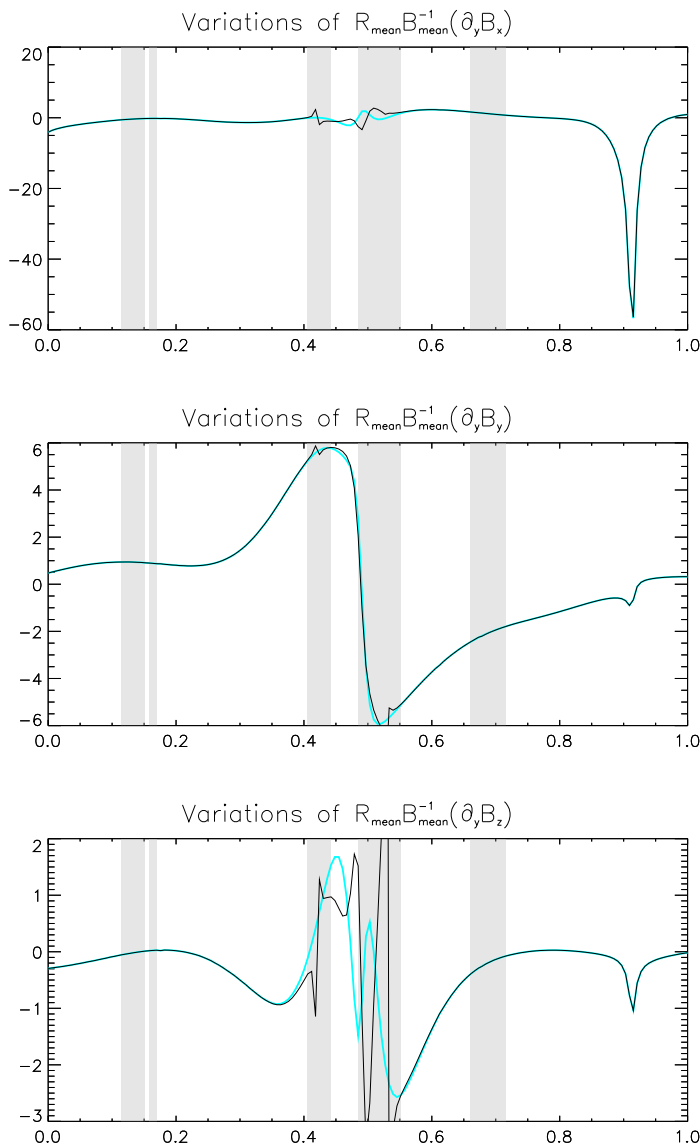


Figure 15.6: Components $R_{\text{mean}} B_{\text{mean}}^{-1} \partial_y \mathbf{B}$ estimated by quadratic interpolation of \mathbf{B} . Otherwise similar to Figure 15.5.

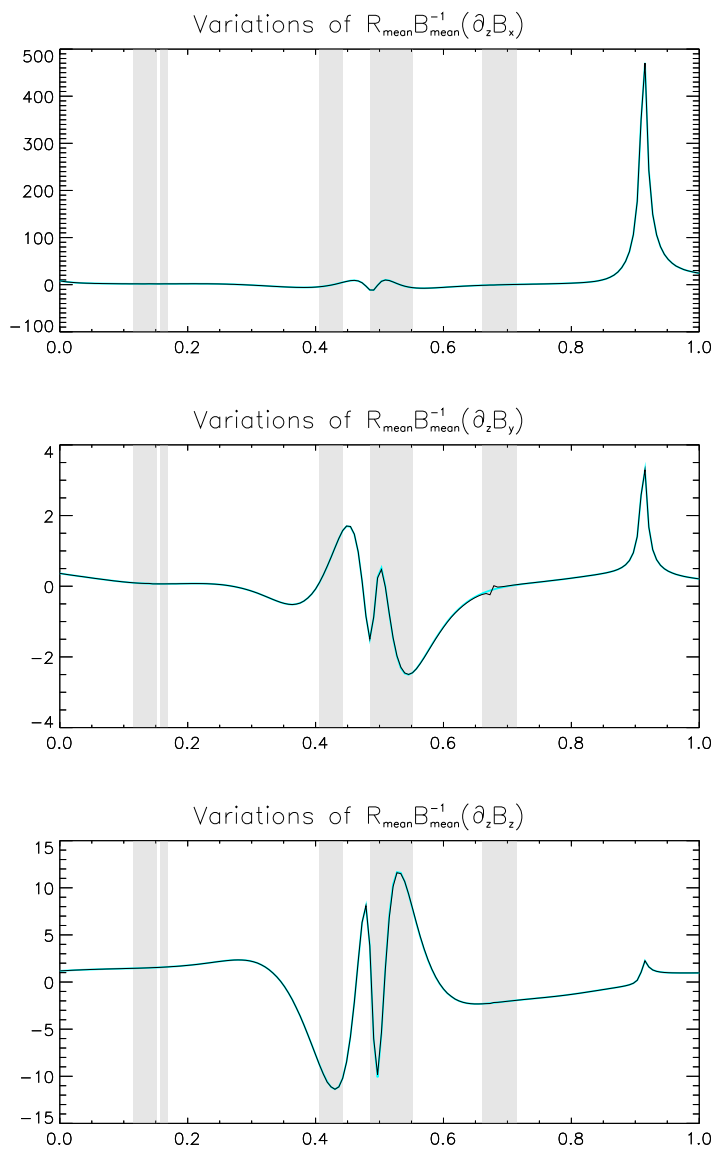


Figure 15.7: Components $R_{\text{mean}} B_{\text{mean}}^{-1} \partial_z \mathbf{B}$ estimated by quadratic interpolation of \mathbf{B} . Otherwise similar to Figure 15.5.

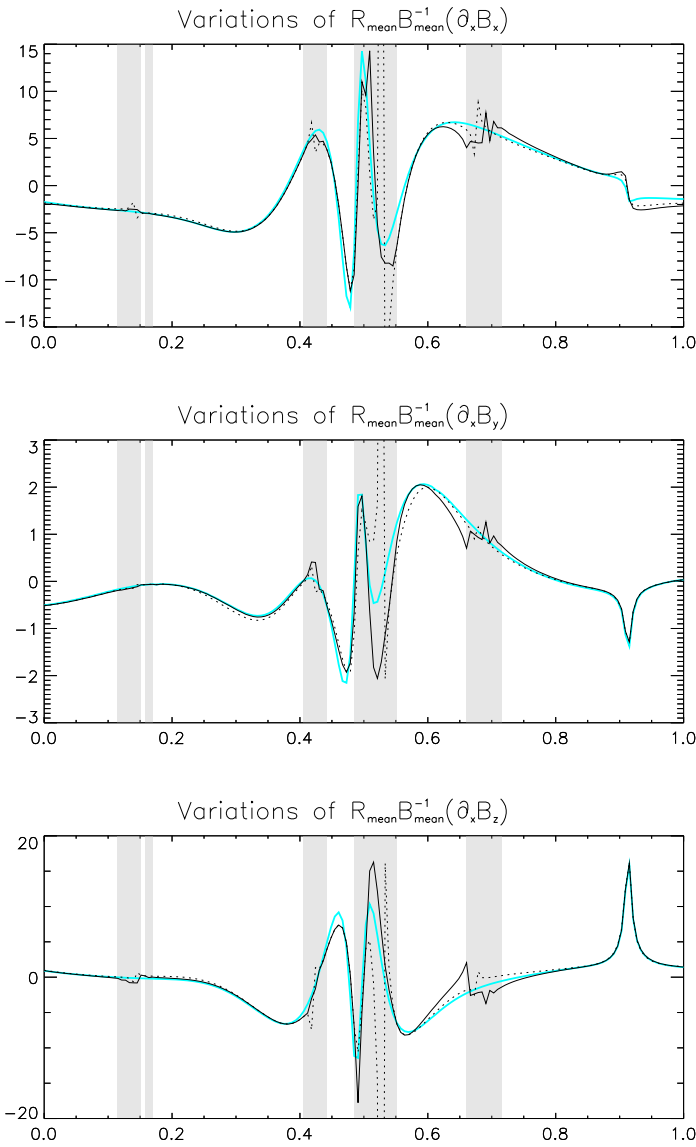


Figure 15.8: Components $R_{\text{mean}} B_{\text{mean}}^{-1} \partial_x \mathbf{B}$ estimated by the corrected linear estimator in order to reduce the amplification of the truncation errors by singular reciprocal bases. The exact components are plotted as solid grey curves, and the estimated components as dotted black curves for the solenoidal estimator and solid black curves for the corrected linear barycentric estimator.

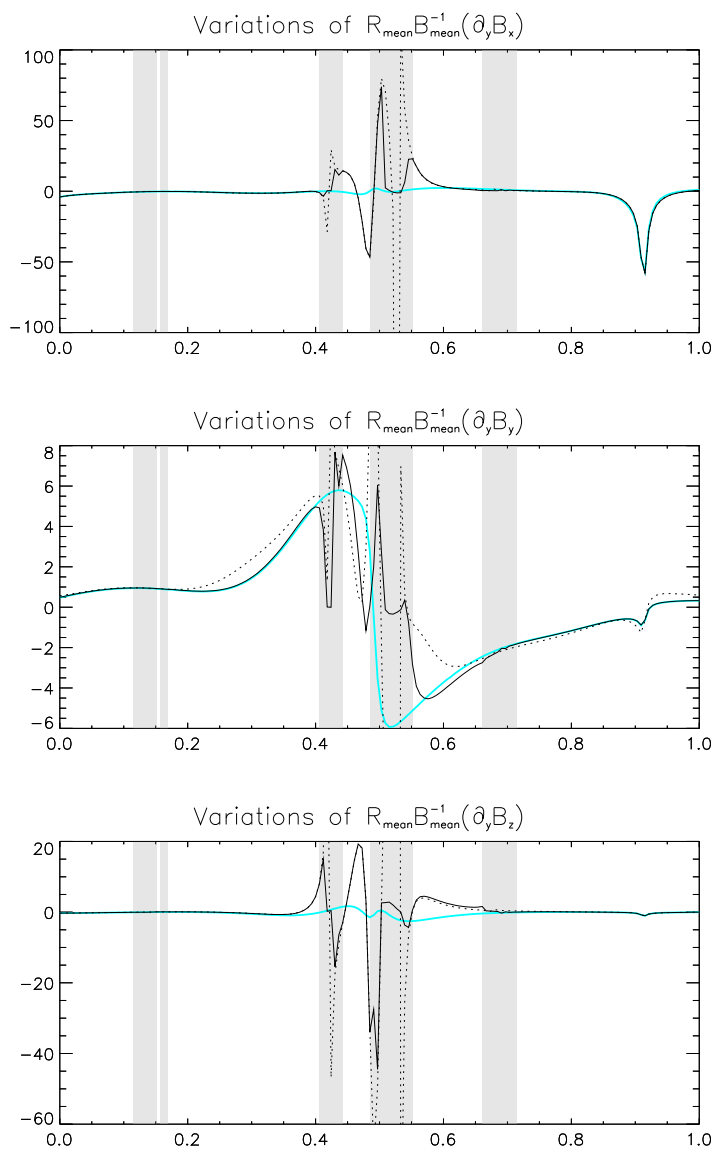


Figure 15.9: As Figure 15.8 but for the components $R_{\text{mean}} B_{\text{mean}}^{-1} \partial_y \mathbf{B}$.

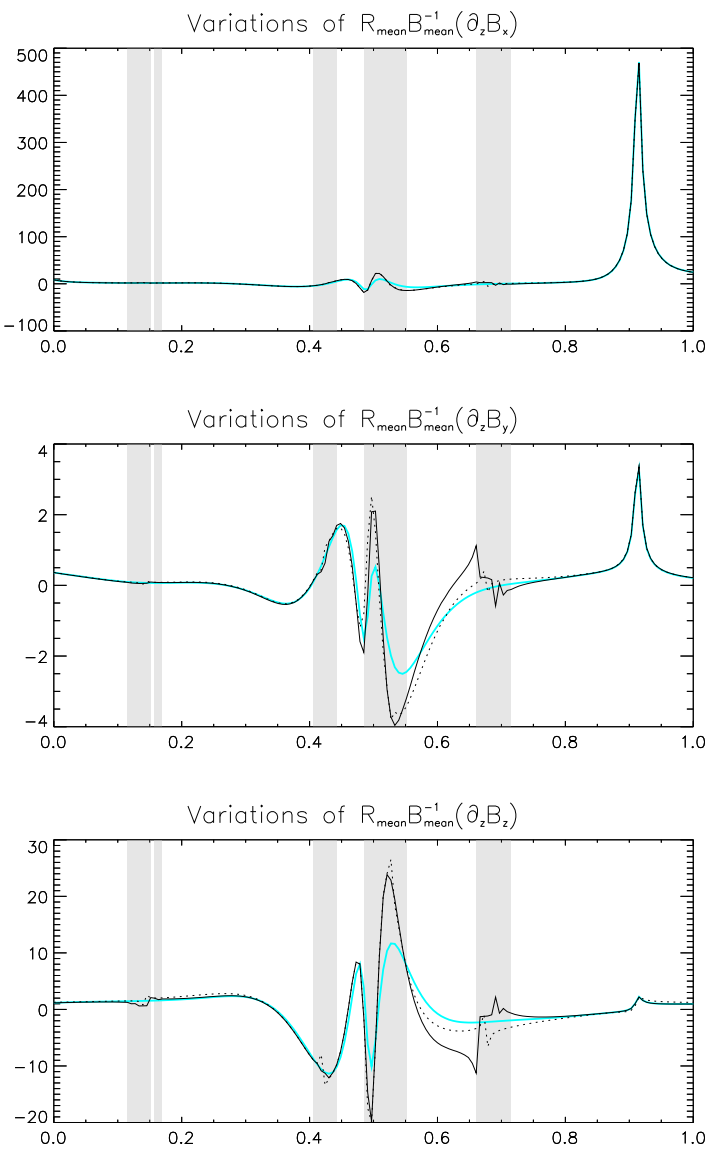


Figure 15.10: As Figure 15.8 but for the components $R_{\text{mean}} B_{\text{mean}}^{-1} \partial_z \mathbf{B}$.

Neither indicator requires diagonalisation of the volumetric tensor, both predict equally well the critical times, but neither describes what occurs to the tetrahedron.

Physical Precision

At the beginning of these investigations it was somewhat naturally thought that the estimation of $\nabla \cdot \mathbf{B}$, the trace of the estimated $\mathbf{G}[\mathbf{B}]$, should be indicative of the quality of the linear interpolation for a solenoidal field, such as \mathbf{B} . This quantity ignores the off-diagonal components, but nevertheless the plots of $\log_{10}(|\nabla \cdot \mathbf{B}|)$ for the Tsyganenko field model and the three estimates of the gradient deserve some comments.

1. The Tsyganenko-87 model itself is not completely divergence free, as indicated by the light-grey curves in the three panels of Figure 15.11. Except when the sign changes, the magnitude of $\nabla \cdot \mathbf{B}$ is larger than $10^{-5} \text{ nT km}^{-1}$ everywhere, and is greater than $10^{-3} \text{ nT km}^{-1}$ close to the Earth; in fact $|\nabla \cdot \mathbf{B}|/|\mathbf{B}|$ is almost constant along the orbit.
2. $\log_{10}(|\nabla \cdot \mathbf{B}|)$ computed for the solenoidal least squares estimator is plotted as the dotted curve of the upper panel of Figure 15.11; it varies between -20 and -16 , very small values indicating good respect of the imposed constraint. But, as already mentioned (Section 15.3), this does not guarantee that the resulting linear estimate is better than the one obtained without this constraint.
3. The linear barycentric estimation (dotted curve of the central panel) is greater than the divergence of the field model (light-grey curve) by one to three orders of magnitude. The quadratic estimation represented by the dotted line in the bottom panel follows more closely the divergence of the model, but nevertheless deviates from the divergence of the model at the third critical time, very near the perigee.

These results emphasise that the crucial point is the applicability of linear interpolation within the tetrahedron, but unfortunately there is presently no criterion to test this point. A pseudo-quadratic estimator as been proposed in Chapter 14 in an attempt to reduce the truncation errors by taking advantage of the orbital motion of the cluster; but, as shown for a planar current sheet model, this estimator is linearly degenerate and is affected by truncation errors which are of the same order as the truncation errors affecting the linear estimator.

It is worth noticing once more that truncation errors affecting $\nabla \cdot \mathbf{B}$ and $\nabla \times \mathbf{B}$ are independent: the simulations presented in this section illustrate this point especially during the time interval between $t = 0.85$ and 1.0 for which, according to Figures 15.8–15.10, all components of $\mathbf{G}[\mathbf{B}]$ involved in $\nabla \times \mathbf{B}$ are accurately estimated meanwhile the components involved in $\nabla \cdot \mathbf{B}$ obviously deviate from the “true” components.

15.5 Future Developments

15.5.1 Filtering of the Data

Real data will require the use of low pass filtering. It is assumed, of course, that the experimental data is already free from any effects of temporal aliasing, as explained in

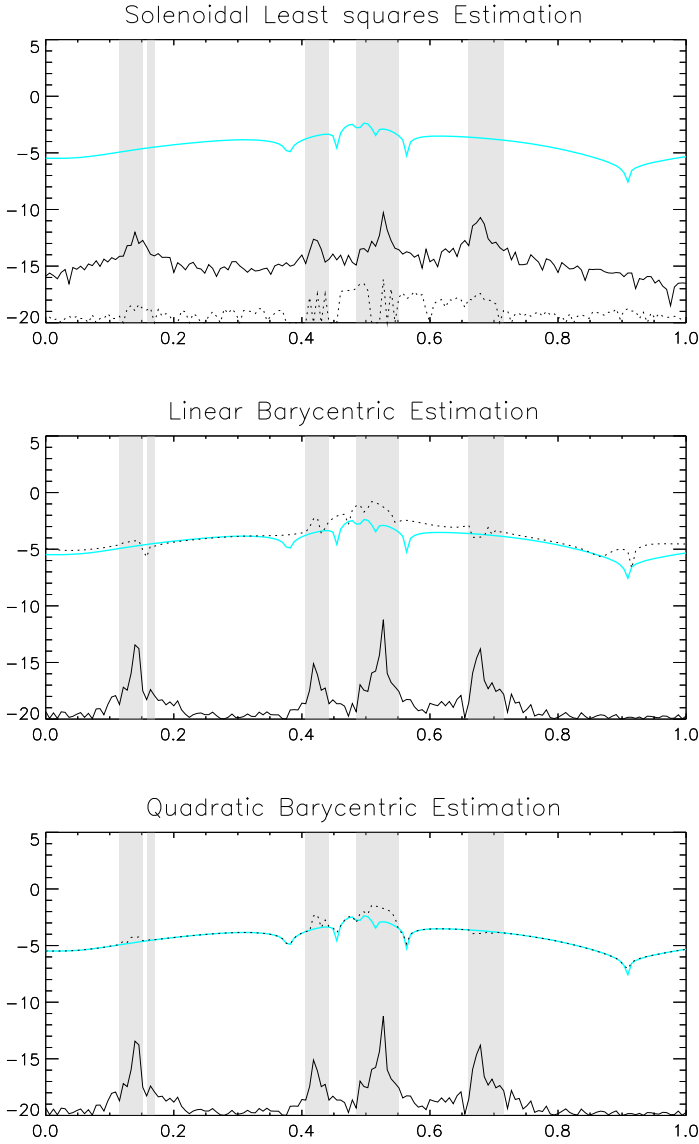


Figure 15.11: Estimated divergences and numerical accuracy indicators Q_{LS} , for the solenoidal least-squares estimator (upper panel), and Q_{bary} for the linear (central panel) and quadratic (bottom panel) barycentric estimators. In all frames the black solid curves represent these accuracy indicators defined by equations 15.18 (page 379) and 15.17 (page 379), respectively, and the light-grey curves represent the variation of $\log_{10}(|\nabla \cdot \mathbf{B}|)$ of the Tsyanenko-87 model along the orbit of the tetrahedron (where $\nabla \cdot \mathbf{B}$ is measured in nT km^{-1}); near the perigee $|\nabla \cdot \mathbf{B}|$ is greater than $10^{-3} \text{ nT km}^{-1}$. The dotted curves represent the variations of $\log_{10}(|\nabla \cdot \mathbf{B}|)$ of the estimators. The time resolution is 20 minutes.

Chapter 2. The effects we are talking about now are spatial aliasing: it is necessary to remove short wavelength (i.e., high k) fluctuations which would otherwise degrade the determination of the gradient on the scale of the size of the tetrahedron. The analysis of these high spatial frequency fluctuations is the subject of Chapter 3. The condition of spatial aliasing (Section 14.5.1) is easily derived with the barycentric formalism of Chapter 14. Real data will require averaging over a characteristic time such that the displacement of the tetrahedron during this time is comparable to, but less than, its size. Averaging over a longer interval would degrade the resolution.

15.5.2 The Spacecraft Position

The discussion of Section 14.3 shows that the precision of the determination of the satellite position is important. The probable error in the position of the spacecraft is generally not isotropic in space. This fact must be taken into account in the determination of the probable errors affecting the experimentally determined gradients. Furthermore, it is clear that for almost coplanar spacecraft the most important information is the precision of the spacecraft position in the direction of planarity; for this component of the gradient estimate problems are most likely to be encountered. It is therefore essential that the covariance matrix of the position error be computed along the orbit.

The three components of the error in the spacecraft position will vary differently with position along the orbit. To a first approximation, the component (of the error) parallel to the spacecraft velocity will have a minimum at apogee (where the orbital velocity is small) and maximum at perigee; this component depends mainly upon a timing offset. The components which depend upon a difference of orbital eccentricity or orbital plane will have the opposite behaviour, being larger near apogee than near perigee. These variations are similar to the variations of the geometry itself of the Cluster tetrahedron. Therefore the determination of gradients near apogee, when the tetrahedron is considerably flattened, may be rather better than predicted by the simple rule that the position accuracy is within a certain limit, e.g. ± 5 km. Note that for the numerical demonstration of Section 15.4 the orbit was chosen to be isotropic close to apogee, and consequently this flattening was not obvious. But in general the scientific objectives will require isotropy to be targetted at lower altitudes, leading to flattening of the tetrahedron near apogee.

It is essential to give more thought to this matter, because the Cluster experimenters will probably often want to push to the limit their knowledge of the orbital position, especially when the spacecraft are far from that part of the orbit concerned by a primary science objective.

15.6 Conclusions

The determination of gradients from measurements by four spacecraft is just the spatial interpolation of data within the tetrahedron formed by the spacecraft. The homogeneous least squares method presented in Section 12.2 complements the theoretical framework of Chapter 14. The unconstrained least square method was proved in Section 15.2 to be mathematically identical to the linear barycentric method described in Section 14.2. In principle the least squares approach allows account to be taken of the solenoidal constraint, but this does not necessarily improve the estimation of the gradient because, as demonstrated in

Section 15.3, the linearly interpolated gradient of a solenoidal field is generally not divergence free. The simulations presented in Section 15.4 show that in practice the two methods may yield different results, especially in the vicinity of singular configurations of the tetrahedron, with no indication as to which determination is better.

The errors of truncation may lead to completely erroneous estimates, especially for nearly planar clusters. Presently, no criterion is available which could give a hint about the validity of the linear interpolation of the field in between the spacecraft; this linear interpolation is the common basis of all methods designed to estimate the gradient of a given field. Such a criterion is mandatory to give credit or not to these estimates; more investigations are required of this crucial point. Nevertheless, the accuracy indicators discussed in section 15.4.3 are reliable warnings of inaccuracy in the vicinity of the planar clusters. In some cases, a regularisation of the reciprocal base can be attempted as demonstrated in the same section. The approximate quadratic estimator defined in Section 14.2 does not improve the estimation of the gradient along realistic orbits in a Tsyanenko model of the geomagnetic field. Note that the least squares method as described in Section 12.2 is directly applicable to more than four spacecraft.

The Tsyanenko-87 model we have used does not include a magnetopause, hence the simulations have been restricted to orbits completely within the magnetosphere, but from the theoretical analysis of Section 14.4 we conclude that a correct determination of the Chapman-Ferraro currents at the magnetopause will require a small size tetrahedron, with an inter-spacecraft distance of the order of 100 km; otherwise, the truncation errors will lead to meaningless estimations of the magnetic gradients and electric currents.

When the truncation errors are acceptably small, the accuracy of the estimated gradient is determined by the geometrical errors related to the shape of the tetrahedron and to the uncertainties on the spacecraft positions; to determine these errors it is mandatory to know the reciprocal vectors and the covariance matrix of the errors of the spacecraft positions. Lastly, it can be seen that the size, elongation and planarity of the tetrahedron, together with the three Euler angles describing the orientation of the principle axes (i.e., the direction of the axes of elongation and of planarity), are essential parameters for all multi-spacecraft science.

Acknowledgements

The authors are indebted to P. Robert for the files of simulated magnetic data in GSE coordinates and for the computation of the geometrical factors. They are also indebted to J. Vogt for having drawn their attention to the relationship 15.2 between the volumetric and reciprocal tensors.

Bibliography

- Chanteur, G. and Mottez, F., Geometrical tools for Cluster data analysis, in *Proc. International Conf. "Spatio-Temporal Analysis for Resolving plasma Turbulence (START)"*, Aussois, 31 Jan.–5 Feb. 1993, ESA WPP-047, pp. 341–344, European Space Agency, Paris, France, 1993.
- Coeur-Joly, O., Robert, P., Chanteur, G., and Roux, A., Simulated daily summaries of Cluster four-point magnetic field measurements, in *Proceedings of Cluster Workshops*,

- Braunschweig, 28–30 Sep. 1994, Toulouse, 16–17 Nov. 1994*, ESA SP-371, pp. 223–227, European Space Agency, Paris, France, 1995.
- Khurana, K. K., Kepko, E. L., Kivelson, M. G., and Elphic, R. C., Accurate determination of magnetic field gradients from four point vector measurements: II. use of natural constraints on vector data obtained from four spinning spacecraft, *IEEE Trans. Magn.*, **32**, 5193, 1996.
- Tsyganenko, N. A., Global quantitative models of the geomagnetic field in the cis-lunar magnetosphere for different disturbance levels, *Planet. Space Sci.*, **35**, 1347–1359, 1987.

— 16 —

Accuracy of Current Density Determination

PATRICK ROBERT

*Centre National de la Recherche Scientifique,
Vélizy, France*

MALCOLM W. DUNLOP

*Imperial College of Science, Technology and Medicine
London, United Kingdom*

ALAIN ROUX AND GERARD CHANTEUR

*Centre National de la Recherche Scientifique,
Vélizy, France*

16.1 Introduction

The four Cluster spacecraft will enable simultaneous measurements of the vector magnetic field at the vertices of a tetrahedron. Methods for analysing such data include: a contour integral method for determining the electric current density (from $\nabla \times \mathbf{B}$) in fields varying on scales relatively large compared to the spacecraft separations, and which was coined the “curlometer” technique; a wave telescope technique which is most reliable where the field scale variations are comparable to the spacecraft separations; and a discontinuity analysis technique for cases where the field scale variations are much shorter than the spacecraft separations. Note that the Curlometer also provides an estimate of $\nabla \cdot \mathbf{B}$ via Gauss’s Theorem. The finite difference equations at the core of the curlometer technique can also be derived and efficiently solved by making use of barycentric coordinates. A collection of methods based on barycentric coordinates are fully described in Chapter 14. In the linear approximation, the barycentric estimates for $\nabla \times \mathbf{B}$ and $\nabla \cdot \mathbf{B}$ are identical to those defined by contour integrals (as they should be for mathematical consistency).

For this reason, both forms are used here as a matter of mathematical convenience: to employ the computational efficiency of the barycentric equations for the statistical analysis, for instance. The resulting estimates of \mathbf{J} and the divergence of \mathbf{B} are subject to errors of which there are basically three types. The first relates to measurement uncertainties in \mathbf{B} , and in the spatial configuration of the four spacecraft. The second relates to the linear interpolation which is made between the various measurement points. The third relates to the simultaneity of the measurements. Sections 14.3 and 14.4 present theoretical investigations of the first and second types of errors respectively.

The influence of these errors on the accuracy of the estimate of \mathbf{J} or $\nabla \cdot \mathbf{B}$ is strongly related to the shape of the tetrahedron but also to the magnetic structure present. This, of course, is not known (at least explicitly) for measured data. Since the shape of this tetrahedron evolves along the mean trajectory of the 4 spacecraft, it is particularly important to

study the influence of the shape of the tetrahedron on the accuracy of the estimate of the current density.

First of all, we identify the general basis for the curlometer method. We briefly explore the effect of magnetic structure on the quality of the estimate of \mathbf{J} , particularly highlighting the possible effect of anisotropic structure on sampling, and set this in the Cluster context.

Then, using randomly simulated configurations for the case of a particular isotropic magnetic structure, we discuss the following quantitative questions, in some detail: (i) what parameters we have to use to characterise the shape of a given tetrahedron? (ii) what is the relation between the geometrical shape of the tetrahedron and the accuracy of the determination of \mathbf{J} via the estimate of $\nabla \times \mathbf{B}$? (iii) can $\nabla \cdot \mathbf{B}$ be used as an estimate of the error ΔJ , as opposed to its use as a general indicator of physical coverage¹? and (iv) when the tetrahedron is relatively flat, is there a relation between the accuracy of the determination of \mathbf{J} and the orientation of the current with respect to the orientation of the tetrahedron? These questions are studied with the help of a numerical simulation based on a large number of possible tetrahedra and a model for the current structure.

16.2 The Curlometer Technique

16.2.1 Background

Use of the magnetic field alone requires the electric current density to be estimated from Ampère's law, which is physically valid over the dynamic range of dc magnetometer measurements. By treating this current as constant over the tetrahedral volume formed by the four spacecraft, a difference estimate of $\nabla \times \mathbf{B}$ can be made. This estimate forms the basis of the curlometer analysis technique. In reality, the current will always vary to some degree over the tetrahedron and the best (*a priori*) knowledge of this lies in estimating $\nabla \cdot \mathbf{B}$ under the same assumptions. Because of the solenoidality of the magnetic field ($\nabla \cdot \mathbf{B} = 0$), any non-zero result arising from this estimate of $\nabla \cdot \mathbf{B}$ arises from the neglected nonlinear gradients in \mathbf{B} (assuming that the error measurement is weak by respect to the error due to the nonlinear gradients). These are of the same order as the second order terms in $\nabla \times \mathbf{B}$, dropped by differencing (see below). This quantity only partly reflects the physical error, arising from the coverage of the magnetic structure achieved by the spacecraft tetrahedron, in the context of the gradients contributing to $\nabla \cdot \mathbf{B}$. The gradient of \mathbf{B} and the nonlinear contributions for $\nabla \times \mathbf{B}$, in particular, are not checked explicitly. The use of the estimate of $\nabla \cdot \mathbf{B}$ serves in this method as a quality indicator only.

During the development of the method the suitability of $\nabla \cdot \mathbf{B}$ as a measure of the physical uncertainty was extensively tested. It was concluded that for some structures $\nabla \cdot \mathbf{B}$ does not necessarily form a particularly good estimate of the error in \mathbf{J} . In particular, for the case of a current tube, if statistically $\nabla \cdot \mathbf{B}$ and ΔJ show the same variation, there is no point to point correspondence (see Section 16.4.3). This has been also suggested by comparisons of the \mathbf{J} vector of the Tsyanenko-87 magnetic field model deduced from the Cluster tetrahedron by the barycentric method with the \mathbf{J} vector deduced by a finite difference method on an arbitrary small scale.

¹What is called here "physical coverage" and "nonlinear contributions" is related to the errors of truncation (see section 14.4).

For *in situ* analysis, the significance of $\nabla \cdot \mathbf{B}$ should therefore be quantified in terms of the interpretation of the data being made (see Section 16.2.3), i.e., in terms of the properties of different model structures being implied. Nevertheless, in many instances $\nabla \cdot \mathbf{B}/|\nabla \times \mathbf{B}|$ can be used to discuss the quality of the current estimate.

The difference approximation leads via the integral form of Ampère's law to:

$$\mu_0 \mathbf{J}_{\text{av}} \cdot (\mathbf{r}_{1\alpha} \times \mathbf{r}_{1\beta}) = \Delta \mathbf{B}_{1\alpha} \cdot \mathbf{r}_{1\beta} - \Delta \mathbf{B}_{1\beta} \cdot \mathbf{r}_{1\alpha} \quad (16.1)$$

$$\langle \nabla \cdot \mathbf{B} \rangle_{\text{av}} |\mathbf{r}_{1\alpha} \cdot (\mathbf{r}_{1\beta} \times \mathbf{r}_{1\gamma})| = \left| \sum_{\text{cyclic}} \Delta \mathbf{B}_{1\alpha} \cdot (\mathbf{r}_{1\beta} \times \mathbf{r}_{1\gamma}) \right| \quad (16.2)$$

where we have chosen to define differences between spacecraft to be relative to spacecraft 1; $\alpha, \beta, \gamma = 2, 3$ and 4 are indices referring to the apexes of the tetrahedron. \mathbf{J}_{av} represents the measured mean current over the tetrahedron volume ($= \mathbf{J}^{\text{estimated}}$ in Section 16.3), and $\langle \nabla \cdot \mathbf{B} \rangle_{\text{av}}$ is the differential estimate of $\nabla \cdot \mathbf{B}$ for the tetrahedron. $\mathbf{r}_{1\alpha}$ and $\Delta \mathbf{B}_{1\alpha}$ represent the separation vectors and field differences between respective spacecraft. These equations give the current normal to each face of the tetrahedron and therefore represent a coordinate independent, natural expression of the current relative to the tetrahedral geometry and orientation. As such, they immediately give a reflection of the physical coverage of the magnetic structure sampled by the spacecraft array (note that the fourth face, not involving spacecraft 1, forms a redundant estimate, but can be obtained by changing the reference spacecraft). If independent reference components of \mathbf{J}_{av} are required, these can, of course, be extracted by a further transformation, or by direct calculation, using the cartesian, differential form of the equations, for example, and also with an equivalent coordinate-independent formalism, employing barycentric coordinates. In Chapter 14 the use of the barycentric method is completely described, including the extraction of the difference estimates for $\nabla \times \mathbf{B}$ and $\nabla \cdot \mathbf{B}$.

Apart from the physical error (lack of accurate coverage by the spacecraft configuration), the change in the current estimate, δJ , due to errors in the spacecraft separations and in the magnetic field values measured at each spacecraft, can be estimated through an error analysis of equation 16.1. Other sources of error, such as timing errors, are usually relatively unimportant or are considered part of the field error. The relative measurement error $\delta J/|\mathbf{J}_{\text{av}}|$ is used as the prime monitor of measurement performance in \mathbf{J}_{av} . It is useful for understanding the variability found for δJ to express this conceptually as follows:

$$\frac{\delta J}{|\mathbf{J}_{\text{av}}|} = F_B \frac{\delta B}{\Delta B} + F_S \frac{\delta r}{\Delta r} \quad (16.3)$$

where the field error and the separation error are taken as component independent for convenience of expression here.

In general, the form for δJ is not separable, and is nonlinear. The above form, however, highlights the fact that the relative current error depends fundamentally on the relative measurement errors (in position and field), scaled by some factors, F_B and F_S , which depend on the field differences and the separation vectors (i.e., magnetic structure between the satellites and spacecraft configuration). Measurement and separation errors also contribute to $\nabla \cdot \mathbf{B}$ itself and may be a large part of the estimate obtained *in situ*. Although for actual data this raises a further issue of interpretation of $\nabla \cdot \mathbf{B}$, for this work the use

of model (simulated) data and known uncertainties allows the effect to be considered separately.

It should be noted that the method used in this chapter is not trivially extendable to the case of more than 4 spacecraft. The basis for calculation of spatial gradients changes in that case since the nonlinear terms can then be checked directly (second order terms in the field differences can be determined), in principle, or the added redundancy with more than 4 points can be used to improve the physical estimates. Consideration of the significance of the measurement accuracy would guide how this information is best used.

16.2.2 Application: Cluster Context

The factors F_B and F_S above are of order unity in the “best case” situation of a regular tetrahedral configuration, sampling structures with spatial scales which are well matched by the overall scale of the tetrahedron. In general, these increase as the tetrahedron distorts, adversely affecting the contribution of these error sources, and typically, F_B and F_S are ~ 2 – 10 , but this varies with magnetic structures also. Because of the critical nature of the error functions (for \mathbf{J}_{av} but also for other combined parameters) and their sensitivity to spacecraft formation, the scientific performance of each measurement (not restricted to the magnetic field) will be highly constrained by orbital evolution. Typical values of the magnetic field measurement errors onboard Cluster are ~ 0.1 nT for typical magnetospheric magnetic fields along the Cluster orbit (up to ~ 1000 nT). For the Cluster mission, a sequence of manoeuvres is anticipated which would modify the natural orbital evolution of the spacecraft configuration. Scientific analysis of a particular plasma environment is limited by the resulting scale size and shape of the spacecraft tetrahedron. A particular orbit phase will set both the configurational evolution over an orbit and the overall configuration scale in each magnetospheric region and this is depicted in Figure 16.1.

Two orbit phases are shown which have typically, 1000 km separations (dayside) and $1 R_E$ separations (nightside), respectively. The insets in the dayside plot are enlargements of the projected configurations ($\times 10$). Very different evolution of the configuration is apparent which samples the model magnetic structure, taken from the Tsyanenko-87 field model, in a highly varied way. The curlometer method above allows the deformation over an orbit to be monitored in terms of its affect on the error functions for \mathbf{J}_{av} . This deformation is to the extent that such differential measurements would be prohibited for significant fractions of the orbit. Similarly, the effect on $\nabla \cdot \mathbf{B}/|\nabla \times \mathbf{B}|$ can be monitored over the orbit. Each quantity can be used independently as an indicator of measurement performance: quality associated with measurement accuracy in the first case and quality of the physical coverage of gradients in the second case. Other quantities may have different regions of bad coverage.

In fact (see Figures 16.3 and 16.4), inspection of $\nabla \cdot \mathbf{B}/|\nabla \times \mathbf{B}|$ and $\delta J/|\mathbf{J}_{av}|$ reveals two facts. Firstly, small separations clearly achieve crudely better linearity estimates of spatial gradients, for some physical structure, since then the linear approximation is more accurate. Secondly, however, accuracy is limited by the relative measurement errors which become large at small separations because the absolute errors δB and δr are roughly constant. Thus, for a fixed orbital evolution, the overall quality represented by each parameter needs to be balanced: good linearity must still allow good relative measurement accuracy. This is represented in the Figure 16.2.

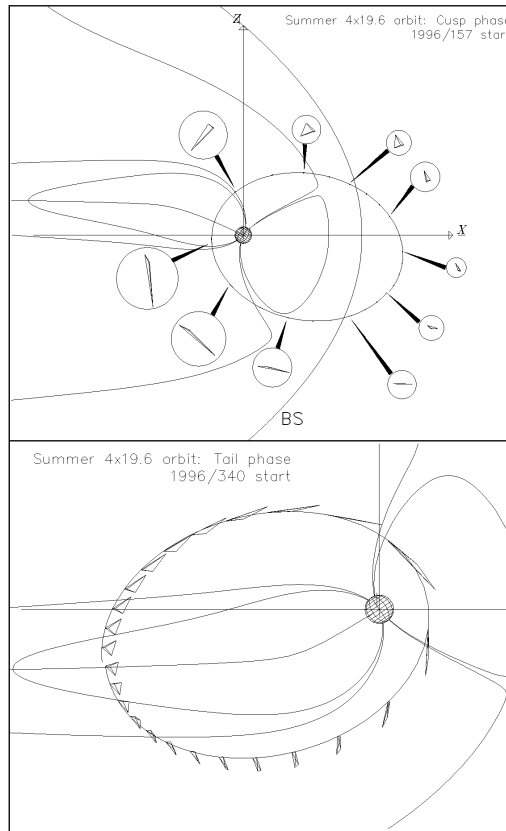


Figure 16.1: Dayside and nightside phases proposed for the first year of Cluster, showing the evolution of the spacecraft configuration.

Note that the curves shown are conceptual and, of course, change relative value depending upon the structure being sampled. The question is: what is achieved for a particular combined quantity (such as J) and for particular phenomena or regions of the orbit? For instance, the curves may cross at small or large separations and the quality of both the physical coverage and measurement may be high over a large range of scales, simultaneously. Not all magnetic structures are suitable for gradient analysis, of course, and in some instances other analysis techniques will be required for adequate performance.

A number of simple model structures have, in fact, been studied with simulated trajectories, having a range of spacecraft configurations, using the analysis tool described briefly in Section 16.2.3, below. It is clear from these studies that the characteristics contained in the magnetic structures (spatial variation), plays a critical role in the balance between the two conceptual curves shown above. The anisotropy and symmetry properties of the phenomena represented, changes the effect of the orientation of the spacecraft configuration relative to the structure, alters the contribution of the nonlinear variations to $\nabla \cdot \mathbf{B}/|\nabla \times \mathbf{B}|$, and, of course, affects the error functions defining $\delta J/|J_{av}|$. The characteristic spatial scales of a phenomenon set the appropriate relative scales for the spacecraft

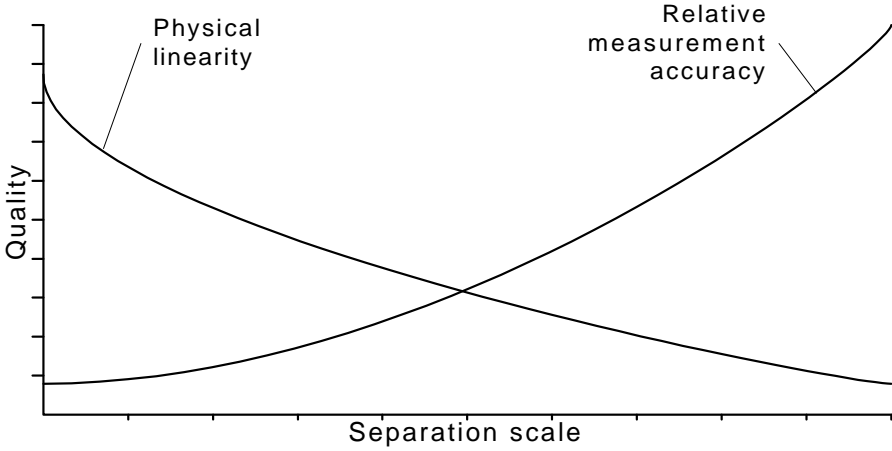


Figure 16.2: Comparison of quality trends with the overall scale of the spacecraft configuration.

configuration (a key factor in defining the Cluster orbital phases). This sensitivity to magnetic structure is often complex since sampled events may exhibit a number of properties (waves on boundaries, flux ropes, etc.) and is beyond the scope of the present chapter. In the second part of this chapter, rather, a detailed study of configuration dependence is discussed for a particular, important magnetic structure.

Over an orbit containing some configurational evolution, the two quality indicators ($\nabla \cdot \mathbf{B}/|\nabla \times \mathbf{B}|$ and $\delta J/|J_{av}|$) will exhibit individual error profiles, highly sensitive to both the characteristic anisotropy and spatial scales of the magnetic structure sampled and the geometry, relative orientation and relative size of the spacecraft configuration. Figures 16.3 and 16.4 show examples of these sensitivities for the case of anticipated Cluster situations in order to give a brief indication of the key results.

Referring back to the dayside orientation for the orbit shown in Figure 16.3, it should be noted that many other evolutions will result for different starting configurations (here chosen to be a regular tetrahedron at the northern cusp), but that the overall scale, within this evolution, is set by the scale of the starting configuration.

Profiles for the two parameters are shown for two, similar choices for dayside phases in the Figure 16.3. This plot shows variation against the true anomaly around the orbit, with the evolution outside the magnetopause not shown and where the field is sampled at the times over the orbit for each tetrahedral position. A nominal separation error of 5 km and a measurement error of 0.1 nT in the field have been assumed (no component error dependence is monitored for simplicity). The parameters monitor poor quality (large values here) when error contributions are high. Features in the trends arise from a combination of the effect of local magnetic structure, the relative separation scale and the degree of tetrahedral distortion. The effect of spacecraft configuration will be more apparent when sampled structure does not vary too widely or is unimportant for the calculation of $\nabla \cdot \mathbf{B}/|\nabla \times \mathbf{B}|$ and $\delta J/|J_{av}|$, the latter being true when the spacecraft configuration scale is large relative to the model structure for instance (here Tsyanenko-87). Then the value of $\nabla \cdot \mathbf{B}/|\nabla \times \mathbf{B}|$ is large, however, and the estimate of $|J_{av}|$ is unphysical ($\delta J/|J_{av}|$ can

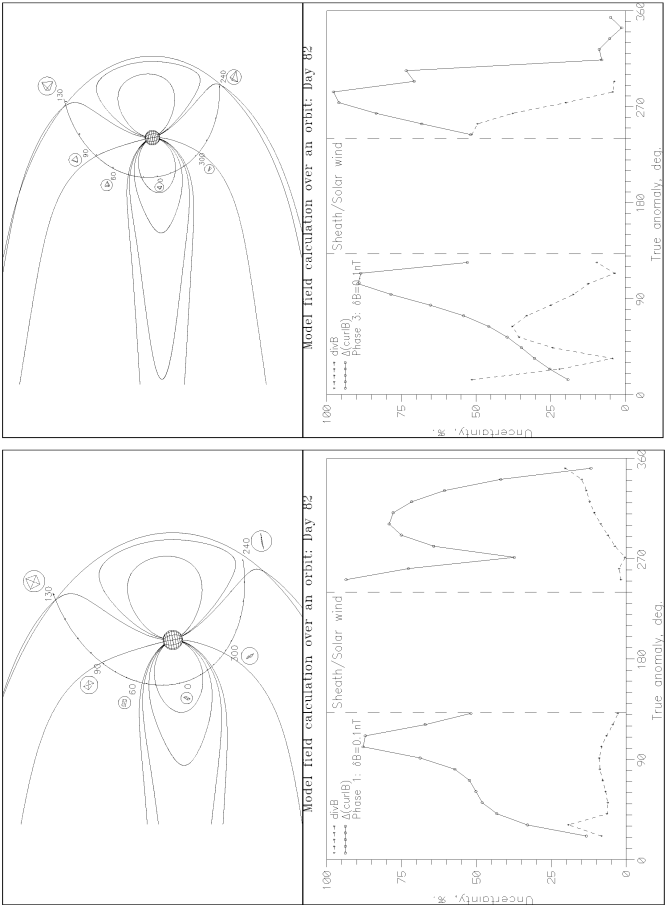


Figure 16.3: Error profiles for $\nabla \cdot \mathbf{B}/|\nabla \times \mathbf{B}|$ and $\delta J/|J_{av}|$ for two cases of dayside evolution, both with starting configurations at the northern cusp. The Tsyganenko model is used as model of magnetic structure.

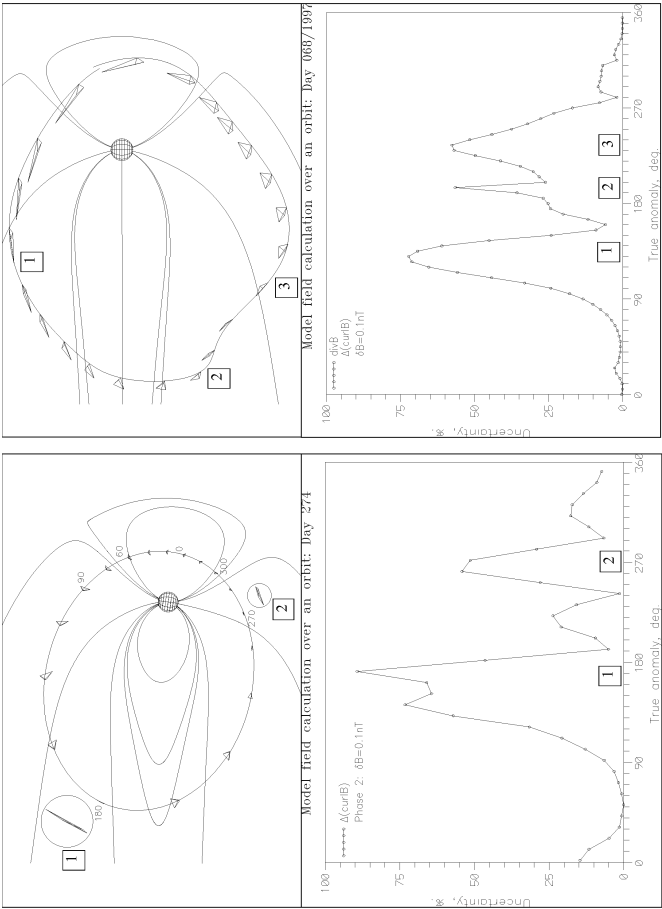


Figure 16.4: Error profiles for the nightside evolution, starting configuration at the magnetotail crossing (T87 model).

still be used as a quality estimate in this case).

In fact, in the context of the Cluster orbital configurations, it is only for the small separations shown in Figure 16.3 (<1000 km) that $\nabla \cdot \mathbf{B}/|\nabla \times \mathbf{B}|$ remains small. Note that, because of the different evolution in each dayside case, the details of profile are different, but that, overall, the trends are similar, particularly the large deviation between $\nabla \cdot \mathbf{B}/|\nabla \times \mathbf{B}|$ and $\delta J/|J_{av}|$: the physical coverage remains good, but measurement error is bad as the spacecraft move away from the strong field region.

Key factors influencing the trends are the overall tetrahedral volume, the shape of the configuration, the magnitude of \mathbf{B} and the nature of the field curvature. These factors combine to produce complicated dependencies and thus tetrahedral geometry alone is not a good indicator for predicting the errors in these cases. The consistent trend to large error in the plots for instance can be explained by the fact that the field magnitude varies considerably; falling dramatically as the magnetopause is approached. Generally, for small field intensities (and hence small differences between the spacecraft) the significance of the error is enhanced. For strong curvature in \mathbf{B} the value of $\nabla \cdot \mathbf{B}/|\nabla \times \mathbf{B}|$ may also grow. The peak at 30° in the left-hand plot, however, can be associated with extreme distortion in the tetrahedral shape. Similarly, the profiles after 240° are very different between the plots because the evolution is dramatically different (note that for the right-hand case the configuration is elongated in the direction perpendicular to the plane of view).

Figure 16.4 shows two cases for the nightside phase, for which a regular tetrahedron is set up close to the tail current sheet crossing. For these larger separations, while error quality is higher overall, reflecting the less critical nature of the separation error, $\nabla \cdot \mathbf{B}$ is now large (not shown for clarity). The evolution for the left-hand plot shows two clear regions of severe distortion (as indicated by the insets, scaled by $\times 10$): one between 120° – 180° and one centred on 270° . These positions correspond to large errors in the current estimate. Moreover, the current density peaks through the current sheet so that $\delta J/|J_{av}|$ is suppressed. The right-hand plot shows the orbit plotted in GSM coordinates. While, clearly, the trends differ, in particular with regard to the positions of the peaks, the striking feature is perhaps the similarity in profiles. In the right-hand plot the current sheet is more clearly identified and the variation over the inner dayside magnetopause is smoother since the cusp structure is not well sampled. The error peaks are still clearly identified with extreme tetrahedral distortion, however; now with a double peak corresponding to the positions ‘2’ and ‘3’.

Thus, for these large configurations, the effect of magnetic structure in $\delta J/|J_{av}|$ has been reduced, so that F_S and F_B depend mainly on tetrahedral geometry. This is in contrast to the dayside phases where the effect of the sampled magnetic structure masks the effect of the tetrahedral geometry. The dependence of $\nabla \cdot \mathbf{B}$ and $\delta J/|J_{av}|$ on tetrahedral geometry is covered quantitatively in detail for specific model structure in the next section.

16.2.3 Analysis Technique

The preceding discussion has concentrated on issues most relating to the analysis of *in situ* data (for which there is no *a priori* knowledge of magnetic structure) and results in a conceptual method arising from the combined use of $\nabla \cdot \mathbf{B}$ and $\delta J/|J_{av}|$ as quality monitors relating to the curlometer technique. The use of $\nabla \cdot \mathbf{B}$ makes no distinction between those spatial gradients which contribute differently to $\nabla \cdot \mathbf{B}$ and $\nabla \times \mathbf{B}$: there is an assumption that the terms in $\nabla \cdot \mathbf{B}$, as sampled, are as well represented, in combina-

tion, as the gradient terms in $\nabla \times \mathbf{B}$, or as poorly represented. The conceptual flow may contain further interpretation (or assumption) of event properties, however, following an initial analysis and combination of the data. This, of course, guides the application of the technique. The evidence obtained from simple models (e.g. where the direction of $\nabla \times \mathbf{B}$ does not change) is that $\nabla \cdot \mathbf{B}$ works well for indicating poor coverage for \mathbf{J} , at least in statistical sense. Such a technique has been designed and tested with a variety of simulated situations and is briefly described below.

Although an estimate for the current density can always be made, the above discussion suggests that two key areas of quality control can be pursued: the physical error, represented (crudely) by $\nabla \cdot \mathbf{B}/|\nabla \times \mathbf{B}|$, and the measurement error, represented by $\delta J/|J_{av}|$. The estimate of $\nabla \cdot \mathbf{B}$ itself separates the nature of the analysis which can be performed on actual events. Even if $\nabla \cdot \mathbf{B}/|\nabla \times \mathbf{B}|$ is well behaved, measurement quality may not be high for all components of \mathbf{J} , and this must be monitored as part of the analysis. A poor estimate of $\nabla \cdot \mathbf{B}/|\nabla \times \mathbf{B}|$ indicates a possibly poor physical estimate of some components of \mathbf{J} , requiring further (independent) interpretation of the sampled structure (for example, the size of the components of the gradient or stationarity properties). If $\nabla \cdot \mathbf{B}/|\nabla \times \mathbf{B}|$ is not well behaved, therefore, gradient analysis can still be performed on individual terms in the dyadic $\nabla \mathbf{B}$ and the time dependence of the event can be checked.

The technique has been used to investigate a number of simulated events using a variety of magnetic structures. It is clear from this study that for strongly anisotropic structures, having spatial scales at least of the order of the spatial scales of the spacecraft configuration, the orientation of the configuration with respect to model is crucial as suggested above. For some orientations the estimates show better coverage with distorted configurations than for regular configurations when comparing the estimated \mathbf{J} to the mean current for the model.

16.3 Accuracy of Current Density Determination

16.3.1 Parameters Used to Define the Shape of a Tetrahedron

Chapter 13 is dedicated to the study of “quality factors” which could describe the shape of a tetrahedron. In the past, the relationship between the accuracy of the current density determination and the shape of the tetrahedron has been studied with various 1-D parameters. The main result was that there exist two categories of 1-D parameter: one which attempts to describe the geometrical shape, and a second which is directly connected to the relative accuracy of the measurement. This was a conclusion deduced from observation and simulation, but not explained. Chapter 13 studies this observation, and has checked the validity of the 1-D parameters. The main conclusion of that chapter is that the best way to describe the shape of the tetrahedron, and understand what happens, is to use a 2-D geometric factor, made up of two parameters, the *elongation* E and *planarity* P parameters, which are deduced from the volumetric tensor (see Chapters 12 and 13).

Here we take again this method, and all parameters such as the total difference between the estimated and the theoretical, mean $|\mathbf{J}|$ defined by the model, $\Delta J/J$ (see precise definition in Section 16.3.5), and parameters such as $\nabla \cdot \mathbf{B}$ and $\nabla \times \mathbf{B}$ will be studied in an elongation-planarity (E - P) diagram, which allows us to characterise more precisely the relationship between the shape of the tetrahedron and the accuracy of the estimate of the

current density, and furthermore to understand why, in certain cases, the estimate of the measurement is good, and why, in other cases, this estimate is bad.

⇒ 16.1

16.3.2 Simulation Method

The following method has been used to check the possible relation between the shape of the tetrahedron and the accuracy of the current density determination:

- firstly, we use a high number of tetrahedra, corresponding to several possible geometric configurations, taken from a “configurations reservoir”;
- secondly, we use a model of current density structure, with a characteristic scale larger than the size of the tetrahedron, such that all the vertices of the tetrahedron are contained inside the current structure;
- thirdly, we add independent noise on the \mathbf{B} magnetic components at the 4 spacecraft positions (the four vertices of the tetrahedron) and we compute vector parameters such as $\nabla \cdot \mathbf{B}$, $\nabla \times \mathbf{B}$, and the relative error $\Delta J/J$ on the determination of J .

16.3.3 The Tetrahedron Reservoir

The shape of the tetrahedron is characterised by the E - P parameter. We try to identify a possible relationship between the value of these parameters and the accuracy of the estimate of $\nabla \times \mathbf{B}$ and $\nabla \cdot \mathbf{B}$. Firstly, to be sure that various kinds of tetrahedra are taken into account, we use the “homogeneous tetrahedra reservoir” defined in 13.5.1 of Chapter 13. This reservoir contains about 1000 tetrahedra and offers an homogeneous coverage of the E - P plane. So, this reservoir contains a wide variety of configurations. To avoid any bias, the tetrahedra taken from the reservoir are reprocessed as follows: (i) they are computed in the barycentric coordinates, (ii) all the tetrahedra are normalised to the same mean inter-spacecraft distance $\langle D \rangle$, (iii) their orientations are perturbed so as to have a random orientation in space, and (iv) the position of the centre of each tetrahedron is translated to -2500 km, in order to avoid the centre of the current structure which could be a particular case.

16.3.4 The Current Structure Models

The goal is to simulate the crossing of a current structure by the Cluster constellation. We have therefore to define a current structure model. The chosen model is described in Figures 16.5 and 16.6: it consists of a cylindrical current tube, with an homogeneous current density (Figure 16.5), or a Gaussian current density profile (Figure 16.6). In all cases, we assume that the size of the Cluster tetrahedron is smaller than the size of the current density structure, so that all the spacecraft are simultaneously located inside the current density structure. Typical values are $\langle D \rangle = 1000$ km, R or $\sigma = 5000$ km, $J_o = 10^{-8}$ amp/m².

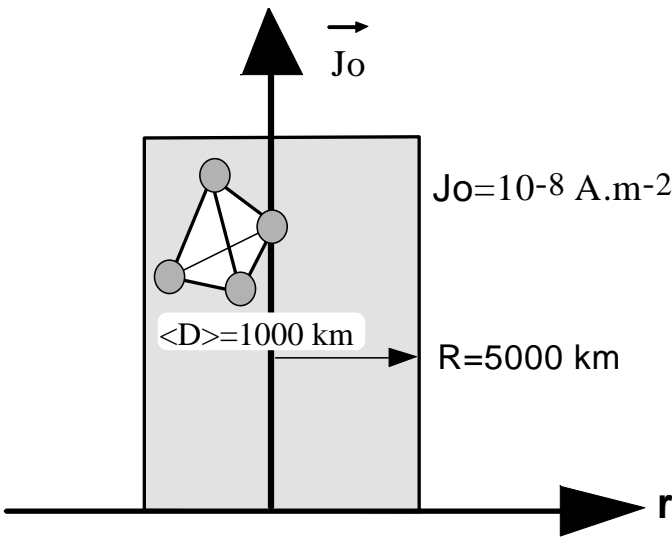


Figure 16.5: Current tube with homogeneous current density profile.

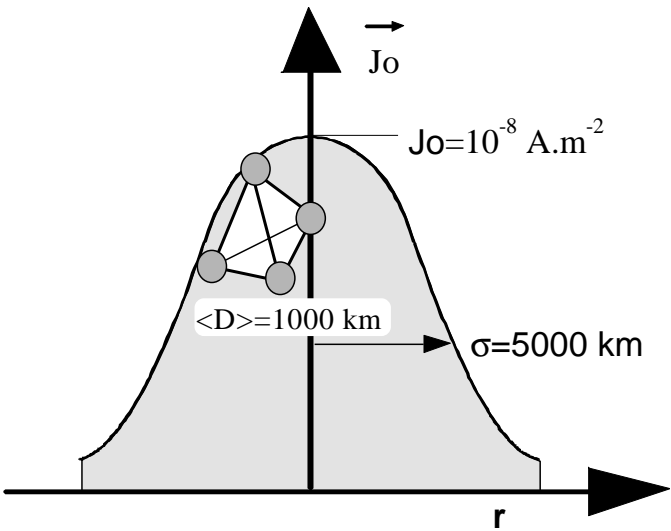


Figure 16.6: Current tube with Gaussian current density profile.

16.3.5 The Computation of \mathbf{J} and $\nabla \cdot \mathbf{B}$

The uncertainty in the measurements of \mathbf{B} (or the uncertainty in the restitution of the spacecraft attitude) is simulated by adding a random noise on the 3 components of the 4 \mathbf{B}_α vectors. Similarly a random noise is added to the 4 \mathbf{r}_α vectors describing the positions of the 4 spacecraft. The amplitude ΔB of the noise added on the 4 magnetic field vectors is independent of the components, and proportional to $|\mathbf{B}|$. The uncertainty Δr in the knowledge of the spacecraft position is taken to be proportional to $\langle D \rangle$, the average inter-spacecraft distance. $\Delta B/B$ or $\Delta r/r$ represent the relative accuracy in the determination of \mathbf{B}_α and \mathbf{r}_α . The relative errors $\Delta B/B$ and $\Delta r/r$ are defined as:

$$\Delta B/B = \Delta B / \frac{1}{N} \sum_{\alpha=1}^4 |\mathbf{B}_\alpha|$$

$$\Delta r/r = \Delta r / \frac{1}{N} \sum_{\alpha=1}^4 |\mathbf{r}_\alpha|$$

Typical values for $\Delta r/r$ are 1%, which correspond to the nominal values given by the Cluster project. $\nabla \cdot \mathbf{B}$ and $\mathbf{J} = \nabla \times \mathbf{B} / \mu_0$ are computed from the perturbed simulated data by the barycentric coordinates method, and we obtain an estimate of the \mathbf{J} or $\nabla \times \mathbf{B} / \mu_0$ vector which is compared to the real value of the average of \mathbf{J} at each vertex given by the model. Although the uncertainty in the modulus and the direction of \mathbf{J} have been studied, we present here only the results corresponding to the modulus of \mathbf{J} , that we note hereafter J . Thus the relative accuracy of the estimate of J , $\Delta J/J$, can be estimated. The definition of $\Delta J/J$ is:

$$\Delta J/J = \frac{\mu_0 \Delta J}{|\nabla \times \mathbf{B}|^{\text{model}}} = \frac{|\mathbf{J}^{\text{estimated}}| - |\mathbf{J}^{\text{model}}|}{|\mathbf{J}^{\text{model}}|} \quad (16.4)$$

where $|\mathbf{J}^{\text{model}}|$ is the mean value of the 4 $|\mathbf{J}_i|$ values of the model at the 4 vertices of the tetrahedron, and $|\mathbf{J}^{\text{estimated}}|$ is the estimate of $|\mathbf{J}|$ by the barycentric method, which is equivalent with $|\mathbf{J}_{av}|$ used in Section 16.2. Of course, from Ampère's law, and taking into account the different kinds of errors, we note $\mu_0 |\mathbf{J}^{\text{estimated}}| = |\nabla \times \mathbf{B}|^{\text{estimated}}$. It has been shown that the errors $\Delta B/B$ or $\Delta r/r$ have the same effect on the accuracy $\Delta J/J$. Therefore, for the sake of simplicity, we will only consider the perturbation $\Delta r/r$. This computation is made for all the tetrahedra taken from the reservoir defined in Chapter 13, thus all the E - P plane is covered.

16.4 Results

16.4.1 Influence of the Shape of the Tetrahedron on the Relative Accuracy $\Delta J/J$

First we consider an homogeneous current density profile, and therefore there is no error associated with the linear interpolation between the measurement made at the 4 spacecraft locations; only the uncertainties on the positions of the measurement points are taken into account ($\Delta r/r = 1\%$, $\Delta B/B = 0$).

The main results are shown in Figure 16.7, where we have plotted the relative accuracy $\Delta J/J$ in an E - P diagram. The size and the colours of the circles indicate the values of $\Delta J/J$: for $\Delta J/J = 0$, the radius of the circle is just an invisible point, the largest circles correspond to $\Delta J/J \geq 10\%$. Note that there is no small circle hidden behind a large one, $\Rightarrow 16.2$ the circles being sorted by size before plotting.

Adopting the five types of tetrahedra defined in Chapter 13 as description of shape, the main conclusion is that for a large fraction of the diagram, corresponding to the “Pseudo-Spheres type” and to a large fraction of the “Potatoes type”, the relative error $\Delta J/J$ remains below 2%. For values of the elongation or planarity parameter larger than about 0.6, $\Rightarrow 16.3$ the errors reach 3% or more. For E or $P > 0.9$, which corresponds to a very long or a $\Rightarrow 16.2$ very flat tetrahedron, the error can reach 10% and more, especially of course when both $E \Rightarrow 16.2$ and P get of the order of unity. As a matter of fact, the error increases roughly with the radius $r = \sqrt{(E^2 + P^2)}$, but this variation is not linear.

16.4.2 Influence of the Shape of the Tetrahedron on the Estimate of $\nabla \cdot \mathbf{B}/|\nabla \times \mathbf{B}|$

Similar results are shown in the Figure 16.8, for $\nabla \cdot \mathbf{B}$ and $\nabla \times \mathbf{B}$ estimated from the simulated measurement values. Notice that the current density profile being homogeneous, the theoretical value of $|\nabla \times \mathbf{B}|$ is the same for all the points, and the $|\nabla \times \mathbf{B}|$ estimated values differs from the theoretical ones according the results of Figure 16.7 where $\Delta J/J$ is defined by equation 16.4 (see Section 16.3.5).

The theoretical value of $\nabla \cdot \mathbf{B}$ is obviously equal to zero. Since $\nabla \cdot \mathbf{B}$ is not a normalised quantity, we have chosen to display the estimate of the ratio $\nabla \cdot \mathbf{B}/|\nabla \times \mathbf{B}|$ rather than the value of $\nabla \cdot \mathbf{B}$. The colour code is the same as for $\Delta J/J$. Roughly speaking, the diagram looks the same as for $\Delta J/J$, the “Pseudo-Spheres” and “Potatoes” types gives the lower values of the divergence, and a large value of E or P leads to a large value of the estimated divergence.

16.4.3 Relationship between $\Delta J/J$ and $\nabla \cdot \mathbf{B}/|\nabla \times \mathbf{B}|$

Since the $\nabla \cdot \mathbf{B}/|\nabla \times \mathbf{B}|$ diagram looks the same as the $\Delta J/J$ diagram, one would expect that the estimated ratio $\nabla \cdot \mathbf{B}/|\nabla \times \mathbf{B}|$ is an estimate of the error $\Delta J/J$. This is statistically true (see previous section), but a more careful investigation shows that there is no one to one correspondence between the two diagrams; a large value of $\Delta J/J$ can correspond to a small value of $\nabla \cdot \mathbf{B}/|\nabla \times \mathbf{B}|$ ratio, and *vice versa*: good estimates of \mathbf{J} can correspond to large value of the divergence (large $\nabla \cdot \mathbf{B}/|\nabla \times \mathbf{B}|$). This is particularly true for large values of E or P .

In order to reveal a possible relationship between $\Delta J/J$ and $\nabla \cdot \mathbf{B}/|\nabla \times \mathbf{B}|$, we have plotted these quantities in together in Figure 16.9. The colour code and the symbols correspond to the family of the tetrahedra defined in Chapter 13 (circle for Pseudo-Spheres, horizontal ellipsoid for Cigars, triangle for Pancakes, diamond for Knife Blade, and oblique ellipsoid for Potatoes).

A possible relationship between these two parameter would result in the alignment of the representative points. This is not observed; the distribution of the points has no preferred direction. Of course the area of the “Pseudo-Spheres” (round symbols) is restricted to the central part of the diagram, close to zero, while the other types cover all the diagram.

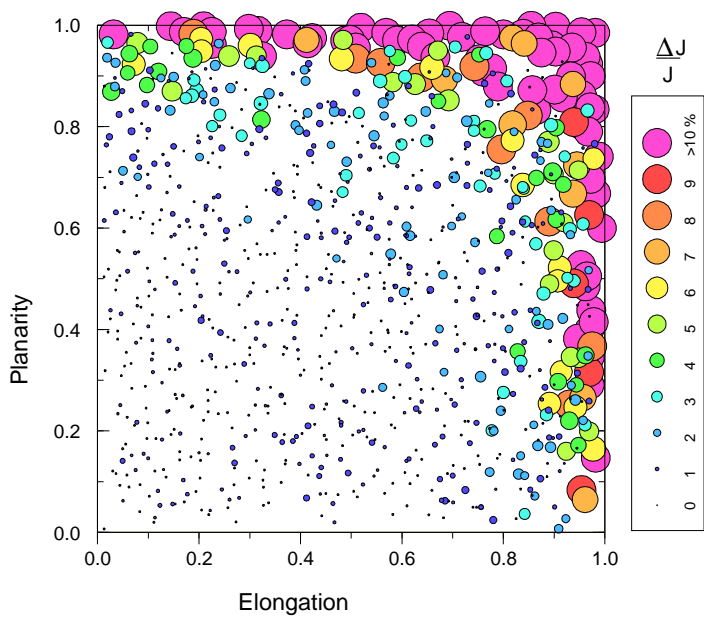


Figure 16.7: Influence of the shape of the tetrahedron on the estimate of $|J|$.

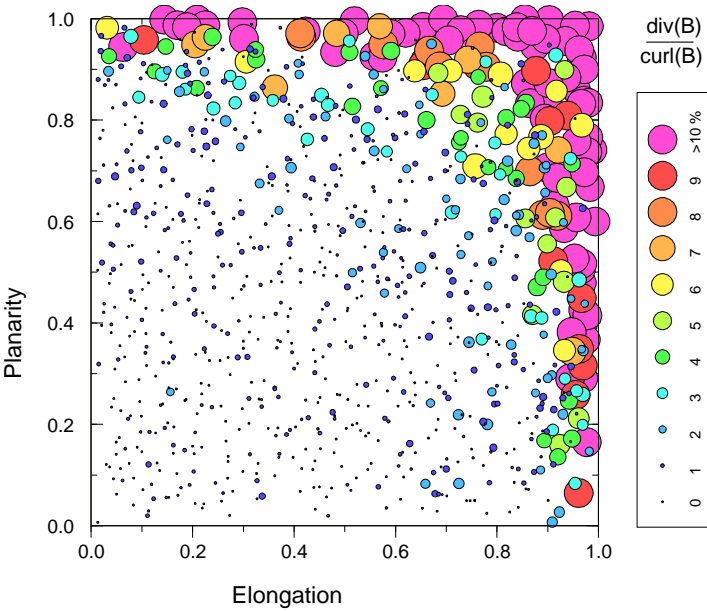


Figure 16.8: Influence of the shape of the tetrahedron on the estimate of the ratio $\nabla \cdot \mathbf{B} / |\nabla \times \mathbf{B}|$.

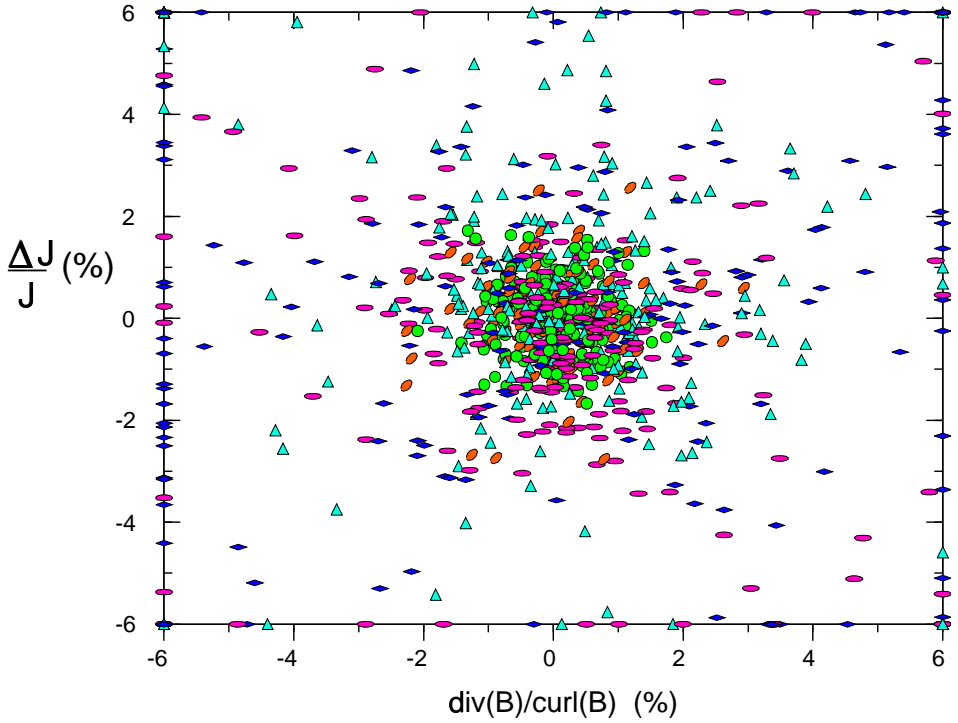


Figure 16.9: Relationship between the error $\Delta J/J$ and the ratio $\nabla \cdot \mathbf{B}/|\nabla \times \mathbf{B}|$.

The non-existence of a correlation between $\nabla \cdot \mathbf{B}$ and ΔJ can be explained by the fact that the computation of $\nabla \cdot \mathbf{B}$ and $\nabla \times \mathbf{B}$ involve different combinations of the components of the dyadic $\nabla \mathbf{B}$. The divergence is obtained from the diagonal terms, while the curl is built from the off-diagonal terms. These terms being perturbed by the addition of an independent random noise on the 12 components defining the 4 spacecraft positions to simulate the uncertainties on the measurement of these positions, the corresponding errors on the gradient tensor are not dependent. Practically, if the errors on the various components are effectively independent, it means that we cannot use the value of the estimated divergence to estimate the error ΔJ due to the measurement uncertainties. Thus, for sampling “blind events”, the uncertainty on \mathbf{J} can be calculated by equation 16.3 (see Section 16.2). Moreover, this result means that a particular configuration shape (classified by E - P) does not alone absolutely order the measurement error, except statistically (all orientations); the relative orientation to the magnetic structure also influences this (see Section 13.2 on “Measurement Performance”, page 324).

16.4.4 Influence of the Current Direction on the Error $\Delta J/J$

In the present section, we investigate the possible influence of the direction of the current with respect to the largest face of the tetrahedron on the errors. Figure 16.10 shows the result: the relative error $\Delta J/J$ is plotted versus the angle θ between the direction of the current and the normal to the main plane of the tetrahedron. The main plane of the

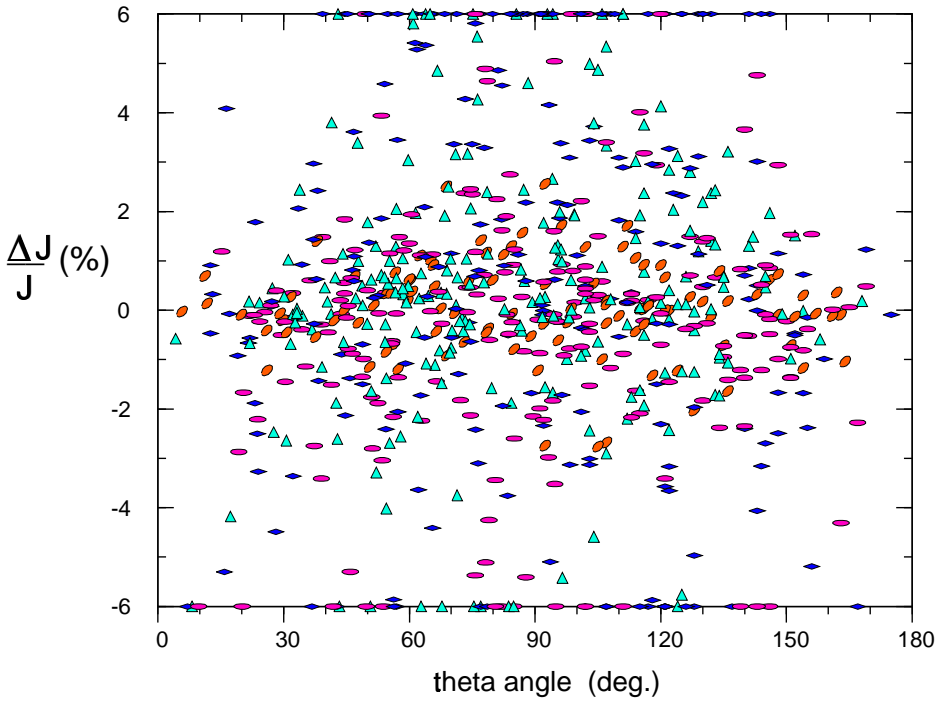


Figure 16.10: Influence of the current density direction on the error $\Delta J/J$.

tetrahedron is defined as the plane containing the major and middle semiaxes of the inertial ellipsoid, thus the normal to this plane is the minor semiaxis.

As defined in Section 16.4.3, a colour code and a symbol are used to separate the different families of tetrahedra defined in Chapter 13. The “Pseudo-Spheres” have been removed from the figure because the main plane has no meaning for a sphere. One could expect to find in this diagram a relationship between $\Delta J/J$ and θ , showing for instance that the quality of the estimate would be better when the current is orthogonal to the main plane. Examination of Figure 16.10 show that the (expected) minimum of $\Delta J/J$ for $\theta = 0^\circ$ or $\theta = 180^\circ$ seems roughly clear. One can observe a slight tendency in the case of the “Pancakes”, for which the main plane has of course the clearest meaning, but this relationship is not very obvious because the distribution of the point in this diagram is not homogeneous along the θ axis, and there is not enough point near $\theta = 0^\circ$ or $\theta = 180^\circ$ to get a clear-cut conclusion. Therefore, practically, it seems that the angle θ plays a role in the organisation of the diagram, and error $\Delta J/J$ can be related to this angle, at least for the particular model of an homogeneous current tube.

16.4.5 Heterogeneous Current Profile

The estimate of the current density inside the volume defined by the tetrahedron relies on the assumption that the magnetic field varies linearly between two spacecraft. If the current density profile is not homogeneous in space (as it is the case here), higher orders

derivatives introduce a key source of errors in the estimate of $\nabla \times \mathbf{B}$ and $\nabla \cdot \mathbf{B}$. To study its effect, we use a Gaussian shape for the current density profile, such as the one shown in Figure 16.6, and define an heterogeneity factor $h = \langle D \rangle / \sigma$ where $\langle D \rangle$ is the mean inter-spacecraft distance and σ is the root mean square deviation of the Gaussian. In order to better illustrate the effect of heterogeneity of the profile, we hereafter neglect the uncertainties $\Delta B/B$ and $\Delta r/r$ which are set to zero.

Figure 16.11 shows the relationship between the shape of the tetrahedron, again defined by the E and P parameters, and the accuracy $\Delta J/J$ of the estimate, for a low value of the heterogeneity factor.

The chosen value: $h=0.1$, being low, the profile of the current structure is not very heterogeneous at the scale of the Cluster tetrahedron, thus the relative error $\Delta J/J$ is very low, except for large values of E or P . For large values of E and P we find a result similar to the homogeneous case (but then with a relative error $\Delta r/r$ of 1%, see Figure 16.7 before). As before, the error grows up rapidly as soon as the tetrahedron degenerates to a very flat or a very elongated configuration. Increasing the heterogeneity factor, for instance $h=0.2$ (Figure 16.12), leads to a rapid growth of the error $\Delta J/J$, but the conclusion about the shape remains similar.

Nevertheless, it seems that the heterogeneous case is more sensitive to a flat or a linear tetrahedron than the homogeneous case; in other words it seems that the errors due to the linear interpolation are more sensitive for a non-regular tetrahedron than the errors associated with uncertainty on $\Delta r/r$ or $\Delta B/B$. This is particularly true for the $\nabla \cdot \mathbf{B}/|\nabla \times \mathbf{B}|$ ratio. Figures 16.13 and 16.14 show the ratio $\nabla \cdot \mathbf{B}/|\nabla \times \mathbf{B}|$ for $h=0.1$ and $h=0.2$. Comparisons between Figures 16.13 and 16.14, and between Figures 16.11 and 16.12 show that $\nabla \cdot \mathbf{B}/|\nabla \times \mathbf{B}|$ is more sensitive to the configuration than $\Delta J/J$. Since the errors on $\nabla \cdot \mathbf{B}$ and $|\nabla \times \mathbf{B}|$ are unrelated, the total error on the $\nabla \cdot \mathbf{B}/|\nabla \times \mathbf{B}|$ ratio is larger (no closely quantitative correspondence is anyway implied).

We have checked that the other conclusions, obtained in the homogeneous case, remain the same in the heterogeneous case. In particular, Figures 16.9 and 16.10 look the same with a finite low value of h . Thus, even in the heterogeneous case, there is no point to point correlation between $\Delta J/J$ and $\nabla \cdot \mathbf{B}/|\nabla \times \mathbf{B}|$. The relationship between $\Delta J/J$ and the angle θ is apparent, particularly near the limiting values $\theta = 0^\circ$ or $\theta = 180^\circ$, with the same restriction than the one for the homogeneous case.

16.5 Conclusions

The curlometer technique, for *in situ* measurements, uses two parameters to monitor different quality aspects: $\nabla \cdot \mathbf{B}/|\nabla \times \mathbf{B}|$ and $\delta J/|J_{av}|$. A large value of $\nabla \cdot \mathbf{B}$, due to non-linear dependence in the gradients contributing to $\nabla \cdot \mathbf{B}$ (assuming that the measurement errors are weak regarding the physical error), may indicate that \mathbf{J} is badly represented by the data, unless the sampling is favourable. The sampling achieved arises from both the configuration parameters (including relative orientation and scale) and the particular magnetic structure encountered (which defines the character of terms in the gradients matrix of \mathbf{B}). But, of course, when handling real data, we have to assume that \mathbf{J} will be badly measured, as indicated by $\nabla \cdot \mathbf{B}$, until we know otherwise. In this regard, other analysis techniques or model assumptions, which need to be checked for consistency, may be used iteratively with the curlometer to improve understanding of an event. When $\nabla \cdot \mathbf{B}$ is small,

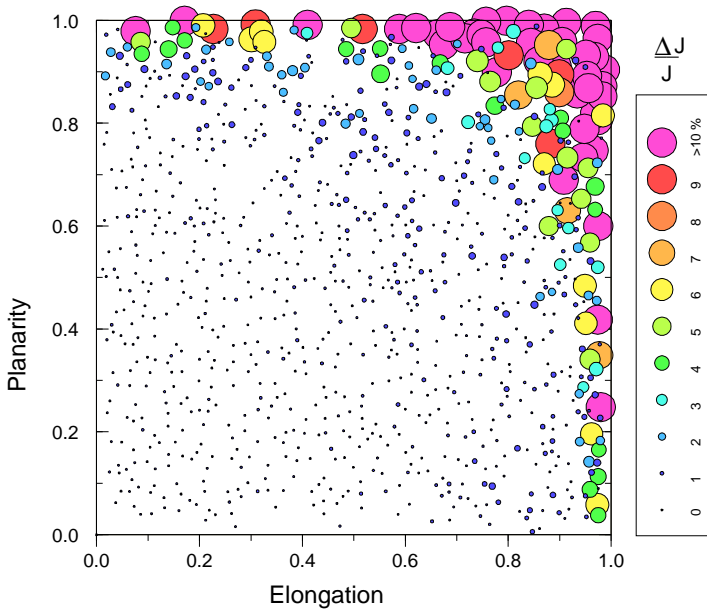


Figure 16.11: Influence of the shape of the tetrahedron on the estimate of J for a low degree of heterogeneity ($h = 0.1$).

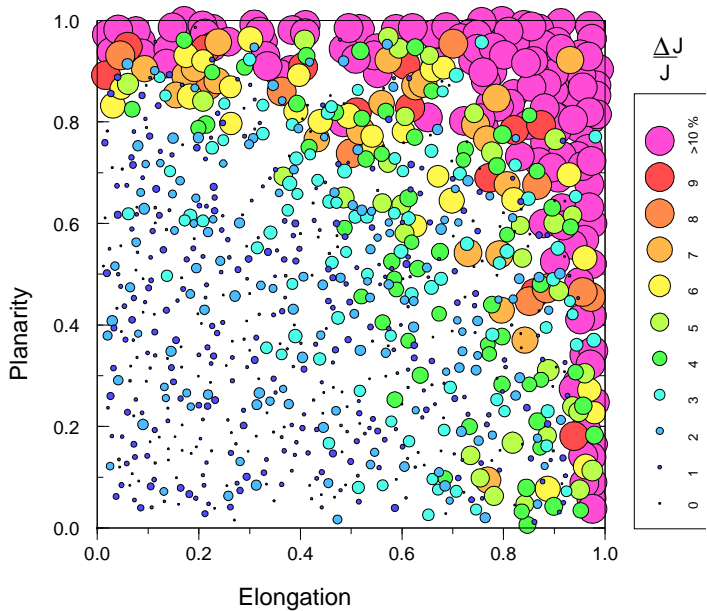


Figure 16.12: Same as Figure 16.11, but for a higher degree of heterogeneity ($h = 0.2$).

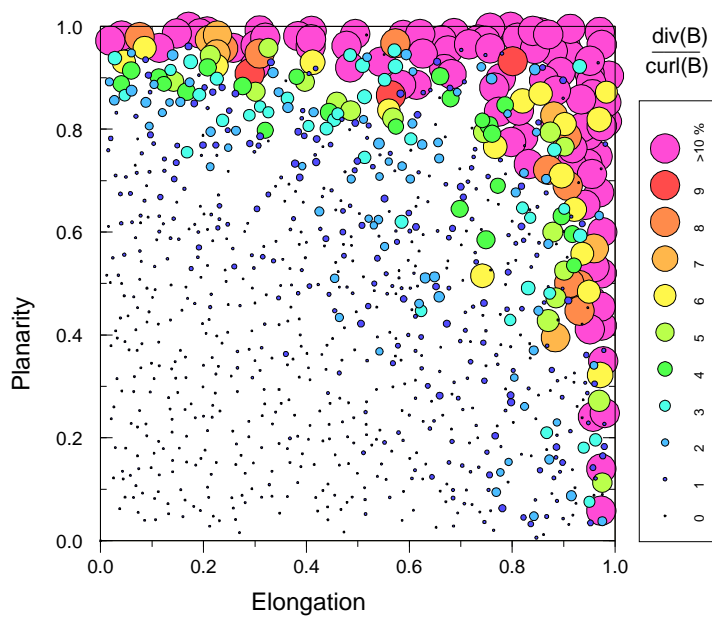


Figure 16.13: Influence of the shape of the tetrahedron on the estimated ratio $\nabla \cdot \mathbf{B} / |\nabla \times \mathbf{B}|$, for a low degree of heterogeneity ($h=0.1$, $\Delta B / B = \Delta r / r = 0$).

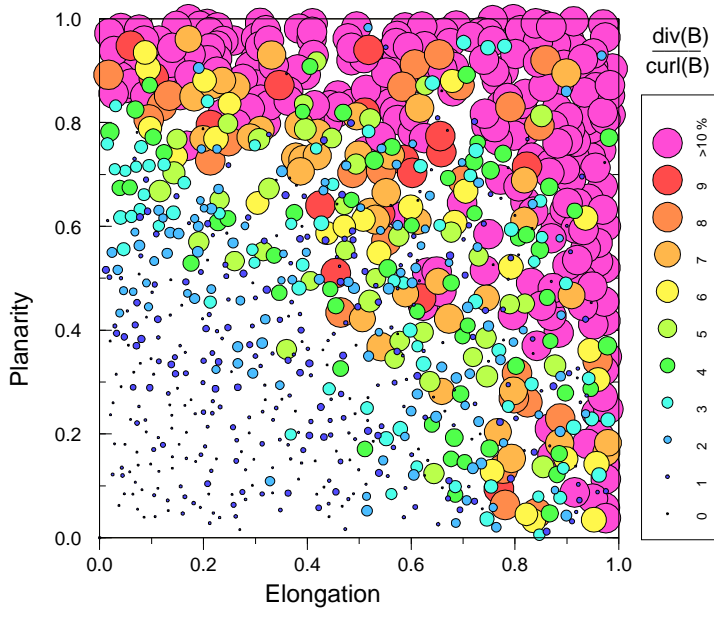


Figure 16.14: Same as Figure 16.13, but for a higher degree of heterogeneity ($h=0.2$).

we can assume that \mathbf{J} remains statistically (and statistically only, see below the conclusion about the relationship between $\nabla \cdot \mathbf{B}$ and δJ) well represented by the data unless the measurement error contribution $\delta J/|\mathbf{J}_{av}|$ (coming from δB and δr), to the error in \mathbf{J} , is large.

We therefore have two nearly independent parameters for the *in situ* monitoring: $\nabla \cdot \mathbf{B}$ which may indicate a physical sampling error when the nonlinear terms dominate, and $\delta J/|\mathbf{J}_{av}|$ which identifies measurement accuracy. It should be understood that not all situations are conducive to the technique and full analysis is performed by hierarchical application of a number of different techniques.

A significant result from Section 16.2 is that $\nabla \cdot \mathbf{B}$ is large for the Cluster orbit phases given, where the Cluster tetrahedron is moving into the Tsyganenko-87 magnetic field model. The parameter $\delta j/|\mathbf{J}_{av}|$ is smaller, overall, for the larger separations. This is more or less as expected: \mathbf{J} is physically better represented by the data when the tetrahedral dimensions are small, but $\delta j/|\mathbf{J}_{av}|$ is smaller for large configurations, due to the combined effects of larger field differences between the satellites and a small ($\sim 1\%$) relative error in the separations. For the mission phases considered here, an optimum range of separation scales, where both the physical estimate and the measurement accuracy remain good, does not exist. The parameters show several instances where the physical coverage (roughly represented by $\nabla \cdot \mathbf{B}$) is insufficient to identify overall quality for the measurement.

To characterise the shape of the tetrahedron, as we have seen in Chapter 12 and 13, the best way is to use the elongation and planarity parameters (E , P) defined and used in Chapter 13 for the description of the shape. Its orientation in space is defined by the three orthogonal vectors corresponding to the directions of the three axes of the pseudo-ellipsoid computed from the volumetric tensor defined in Chapter 12. Vector \mathbf{c} corresponds to the minor axis and is used for description of the orientation of the main face. By introducing the E - P diagram to plot quantities such as the relative accuracy $\Delta J/J$ or other interesting quantities, we have a good tool to analyse the relationship between a scientific parameter and the shape of the tetrahedron, for a given event model. In the same manner, by using the θ angle between the \mathbf{c} vector and the direction of the current, we can analyse the effect of the orientation of the tetrahedron on the accuracy of the measurement.

The relationship between the shape of the tetrahedron and $\Delta J/J$ has been shown by a simulation of the crossing, by 4 spacecraft, of a particular current density structure. This simulation allows an independent estimate of the effects of the various error sources, such as uncertainties on the position of the spacecraft, on the magnetometer measurements, and errors due to the spatial interpolation used for the computation of the various vector gradients. The simulation gives the relative accuracy $\Delta J/J$ of the measurement for all variables. In particular, the accuracy $\Delta J/J$ has been plotted in an (E - P) diagram, which organises the results well and relates easily the shape of the tetrahedron to the accuracy of the estimate of \mathbf{J} or $\nabla \cdot \mathbf{B}$. Furthermore, the E - P diagram is useful to get a quantitative estimate of this influence, and to check whether a regular tetrahedron leads to a more accurate estimate of \mathbf{J} and $\nabla \cdot \mathbf{B}$ than the corresponding estimates for a distorted tetrahedron. Roughly speaking, the distance between the representative point and the origin is proportional to the error on the accuracy of the estimate of both these quantities which, statistically only, show the same trend (see below).

The conclusions are roughly the same for homogeneous and for heterogeneous cases, but it seems that the error due to linear interpolation, in the heterogeneous case, is more

sensitive to a non-regular tetrahedron than the errors due to a relative uncertainty $\Delta r/r$ on the positions. This is particularly true for the $\nabla \cdot \mathbf{B}/|\nabla \times \mathbf{B}|$ ratio. In a future work, it would be interesting to quantify the amplitude of the linearity errors versus the heterogeneity parameter.

A potential relationship between $\nabla \cdot \mathbf{B}$ and ΔJ has been studied in detail. Statistically, the ratio $\nabla \cdot \mathbf{B}/|\nabla \times \mathbf{B}|$ has the same behaviour as $\Delta J/J$, but on close inspection there is no one to one correspondence. There are at least two reasons for this. Firstly, one reason is obvious: since $\nabla \cdot \mathbf{B}$ and $\nabla \times \mathbf{B}$ are computed from different terms of the dyadic $\nabla \mathbf{B}$ used by the barycentric method (see Chapters 14 and 15), and unless the errors are dependent (this can be nevertheless the case for special structures), there is no point to point correlation between the estimated $\nabla \cdot \mathbf{B}$ and the estimated $\nabla \times \mathbf{B}$ and therefore no correlation between $\nabla \cdot \mathbf{B}/|\nabla \times \mathbf{B}|$ and $\Delta J/J$. A second additional reason is that the orientation of the tetrahedron in particular cases, associated with the particular current density structure used, plays a controlling role in this lack of correspondence which has to be more explained. Thus, if the various contributions to measurement errors are independent or if the nonlinear contributions to $\nabla \cdot \mathbf{B}$ and $\nabla \times \mathbf{B}$ are independent, then $\nabla \cdot \mathbf{B}$ should not be used as a reliable estimate of the overall accuracy of the measurement of J . Nevertheless, for straight current tubes, the statistical behaviour of $\nabla \cdot \mathbf{B}/|\nabla \times \mathbf{B}|$ looks the same as $\Delta J/J$, which gives justification of the use of $\nabla \cdot \mathbf{B}$ as an indication of the quality of the physical coverage if the measurement errors can be neglected, or of the measurement error if the current structure is homogeneous.

The influence of the direction of the current with respect to the orientation of the tetrahedron has been examined. Indeed, for a nearly flat tetrahedron, one would expect a relationship between the angle θ (defined as the angle between the direction of the current and the normal to the main plane of the tetrahedron) and the $\Delta J/J$ error. The simulation shows that a relation exists, but the low number of points near θ close to 0° or 180° leads to a reserved conclusion. Other simulations are necessary to better clarify this relation, and in particular to identify the behaviour in this context of the different type of tetrahedra such as Pancakes, Knife Blades, Cigars, etc.

All these conclusions, of course, could be modified for another type of magnetic structure. The effect of different characteristic magnetic properties (events sampled) on measurement performance needs to be investigated more thoroughly, particularly with regard to anisotropic structure. Tools and simulations used in this chapter are well suited for such an event study and should be performed in conjunction with other possible methods.

Bibliography

The three methods for combining multipoint magnetometer data and their associated errors were discussed by:

- Dunlop, M. W., Southwood, D. J., Glassmeier, K.-H., and Neubauer, F. M., Analysis of multipoint magnetometer data, *Adv. Space Res.*, **8**, (9)273–(9)277, 1988.
- Dunlop, M. W., Balogh, A., Southwood, D. J., Elphic, R. C., Glassmeier, K.-H., and Neubauer, F. M., Configurational sensitivity of multipoint magnetic field measurements, in *Proceedings of the International Workshop on "Space Plasma Physics Investigations*

by Cluster and Regatta”, Graz, Feb. 20–22, 1990, ESA SP–306, pp. 23–28, European Space Agency, Paris, France, 1990.

Robert, P. and Roux, A., Accuracy of the estimate of J via multipoint measurements, in *Proceedings of the International Workshop on “Space Plasma Physics Investigations by Cluster and Regatta”*, Graz, Feb. 20–22, 1990, ESA SP–306, pp. 29–35, European Space Agency, Paris, France, 1990.

Barycentric coordinates were first applied to derivation of $\nabla \times \mathbf{B}$ by:

Chanteur, G. and Mottez, F., Geometrical tools for Cluster data analysis, in *Proc. International Conf. “Spatio-Temporal Analysis for Resolving plasma Turbulence (START)”*, Aussois, 31 Jan.–5 Feb. 1993, ESA WPP–047, pp. 341–344, European Space Agency, Paris, France, 1993.

The dependence of the accuracy of J or $\nabla \times \mathbf{B}$ was evaluated by:

Coeur-Joly, O., Robert, P., Chanteur, G., and Roux, A., Simulated daily summaries of Cluster four-point magnetic field measurements, in *Proceedings of Cluster Workshops, Braunschweig, 28–30 Sep. 1994, Toulouse, 16–17 Nov. 1994*, ESA SP–371, pp. 223–227, European Space Agency, Paris, France, 1995.

Dunlop, M. W. and Balogh, A., On the analysis and interpretation of four spacecraft magnetic field measurements in terms of small scale plasma processes, in *Proc. International Conf. “Spatio-Temporal Analysis for Resolving plasma Turbulence (START)”*, Aussois, 31 Jan.–5 Feb. 1993, ESA WPP–047, pp. 223–228, European Space Agency, Paris, France, 1993.

Dunlop, M. W., Southwood, D. J., and Balogh, A., The Cluster configuration and the directional dependence of coherence lengths in the magnetosheath, in *Proc. International Conf. “Spatio-Temporal Analysis for Resolving plasma Turbulence (START)”*, Aussois, 31 Jan.–5 Feb. 1993, ESA WPP–047, pp. 295–299, European Space Agency, Paris, France, 1993.

Robert, P. and Roux, A., Influence of the shape of the tetrahedron on the accuracy of the estimation of the current density, in *Proc. International Conf. “Spatio-Temporal Analysis for Resolving plasma Turbulence (START)”*, Aussois, 31 Jan.–5 Feb. 1993, ESA WPP–047, pp. 289–293, European Space Agency, Paris, France, 1993.

Robert, P., Roux, A., and Coeur-Joly, O., Validity of the estimate of the current density along Cluster orbit with simulated magnetic data, in *Proceedings of Cluster Workshops, Braunschweig, 28–30 Sep. 1994, Toulouse, 16–17 Nov. 1994*, ESA SP–371, pp. 229–233, European Space Agency, Paris, France, 1995.

Robert, P., Roux, A., and Chanteur, G., Accuracy of the determination of the current density via four satellites, in *Abstracts, International Union of Geodesy and Geophysics, XXI General Assembly, Boulder, Colorado, 1995*, presentation GAB51H-06.

The Tsyganenko-87 magnetic field model has been used to test the applicability of the proposed techniques:

Tsyganenko, N. A., Global quantitative models of the geomagnetic field in the cis-lunar magnetosphere for different disturbance levels, *Planet. Space Sci.*, **35**, 1347–1359, 1987.

The FGM magnetometer instrument on board Cluster is described by:

Balogh, A., Dunlop, M. W., Cowley, S. W. H., Southwood, D. J., Thomlinson, J. G., Glassmeier, K.-H., Musmann, G., Luhr, H., Buchert, S., Acuna, M., Fairfield, D. H., Slavin, J. A., Riedler, W., Schwingenschuh, K., and Kivelson, M. G., The Cluster magnetic field investigation, *Space Sci. Rev.*, **79**, 65–91, 1997.

— 17 —

Accuracy of Plasma Moment Derivatives

JOACHIM VOGT AND GÖTZ PASCHMANN
*Max-Planck-Institut für extraterrestrische Physik
Garching, Germany*

Introduction

Fundamental conservation equations in plasma fluid theories involve spatial and temporal derivatives of plasma moments like density, bulk velocity and pressure (see Chapter 6). Multi-satellite missions offer the opportunity to estimate such derivatives. The purpose of this chapter is to investigate the plasma moment derivative accuracy by algebraic means, i.e. formulas for the mean square errors are derived. We consider two sources of error: those of measurement and those of spacecraft position. Partial and directional derivatives are investigated first. Then the application to divergence and curl estimates is discussed.

We start from a simplified satellite configuration to clarify some basic aspects of derivative estimation accuracy. Error formulas for general cluster geometries are deduced thereafter. Applications of the obtained error formulas to the case of the Earth's magnetosphere are discussed.

The estimation of spatial gradients with clusters of spacecraft is also the subject of Chapter 15, where the principles of the analysis techniques are described, with special emphasis is on magnetic field measurements. The investigations of the accuracy of (electric) current density estimations in Chapter 16 (see also references therein) are examples for statistical studies, where measurements of large sets of satellite clusters in a given model field are used to determine the deviations of the measured values from the given ones. This approach is in some sense complementary to the one chosen in the present chapter, where only algebraic means are used and the emphasis is on plasma moment derivatives.

It is important to note that the methods of analysis investigated in this chapter are only applicable to spatial variations *where the scale lengths are not smaller than the spacecraft separation distance*. The differences of the measured values at the distinct spacecraft can then be used to estimate the derivatives. Other methods of data analysis, for example cross-correlation techniques, are required if the spatial scale of the structure of interest is small compared with the inter-spacecraft distance. Discontinuities and shocks are examples for such small-scale structures. They are discussed in Chapters 10 and 11.

Two types of indices are used. Greek letters (α, β, \dots) take values from the set $\{0, 1, 2, 3\}$. These numbers are used to distinguish different spacecraft. Latin characters (i, j, \dots) indicate the range $\{1, 2, 3\}$ and are usually used for the cartesian components of a vector.

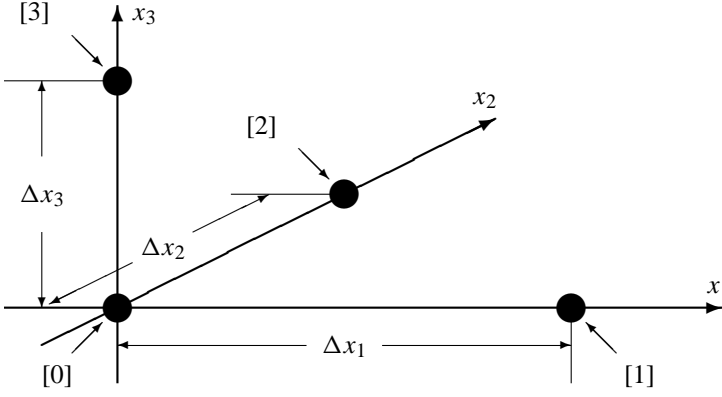


Figure 17.1: Sketch of the three-dimensional model satellite configuration used in Section 17.1.

17.1 Error Estimates from a Simplified Model

In the following we estimate the accuracy of derivative determination on the basis of the simplified satellite configuration sketched in Figure 17.1: The position vectors are $(0, 0, 0)$, $(\Delta x_1, 0, 0)$, $(0, \Delta x_2, 0)$, $(0, 0, \Delta x_3)$. The special cluster geometry together with the choice of the coordinate axes offer a convenient possibility to determine partial derivatives: Let g denote an arbitrary physical quantity, then $\partial g / \partial x_i$ can be estimated by the satellite pair $(0, i)$. In this situation it is inconvenient to use different types of characters for component indices and spacecraft numbers. In order to distinguish the latter from the former in conflicting cases, we write spacecraft numbers in square brackets: [0], [1], [2], [3].

A difference operator $\Delta_{[i0]}$ is introduced by the relation:

$$\Delta_{[i0]}g = g_{[i]} - g_{[0]} \quad (17.1)$$

where $g_{[i]}$ and $g_{[0]}$ denotes the measured value of the physical quantity g at spacecraft [i] and [0], respectively.

17.1.1 Partial Derivative Estimation Accuracy

Partial derivatives are key quantities in the derivative estimation process: gradient, divergence and curl are linear combinations of such expressions. Therefore, partial derivatives are the starting point of our investigation. $\partial g / \partial x_i$ can be approximated in the following way:

$$\frac{\partial g}{\partial x_i} \simeq \frac{\Delta_{[i0]}g}{\Delta_{[i0]}x_i} \equiv \frac{g_{[i]} - g_{[0]}}{x_{i[i]} - x_{i[0]}} \quad (17.2)$$

The accuracy of this estimator is studied in detail in Appendix 17.A.1. Expressions for linear errors are derived there which are squared and ensemble averaged (denoted by a

horizontal line over the averaged quantities) to obtain mean square errors. Cross terms, i.e. averages of products of different derivatives, are treated in the appendix as well. We summarise the resulting equations:

$$\overline{\left(\delta \left[\frac{\partial g}{\partial x_i} \right] \right)^2} = \left\{ \frac{2}{(\Delta x_i)^2} \right\} \left((\delta g)^2 + |\nabla g|^2 (\delta x)^2 \right) \quad (17.3)$$

$$\overline{\left(\delta \left[\frac{\partial g}{\partial x_i} \right] \right) \left(\delta \left[\frac{\partial g}{\partial x_j} \right] \right)} = \left\{ \frac{1}{\Delta x_i \Delta x_j} \right\} \left((\delta g)^2 + |\nabla g|^2 (\delta x)^2 \right) \quad (17.4)$$

$$\overline{\left(\delta \left[\frac{\partial g}{\partial x_i} \right] \right) \left(\delta \left[\frac{\partial h}{\partial x_j} \right] \right)} = \left\{ \frac{1}{\Delta x_i \Delta x_j} \right\} \left([(\nabla g) \cdot (\nabla h)] (\delta x)^2 \right) \quad (17.5)$$

$$\overline{\left(\delta \left[\frac{\partial g}{\partial x_i} \right] \right) \left(\delta \left[\frac{\partial h}{\partial x_i} \right] \right)} = \left\{ \frac{2}{(\Delta x_i)^2} \right\} \left([(\nabla g) \cdot (\nabla h)] (\delta x)^2 \right) \quad (17.6)$$

The formulas concerning the cross terms are valid if g and h are different physical quantities (two scalar functions or two components of the same vector field) with uncorrelated errors and $i \neq j$. Here $(\delta x)^2$ and $(\delta g)^2$ should be interpreted as mean square errors of the spacecraft position determination and the measurement of the quantity g , respectively. Often relative accuracies are more important than absolute ones, i.e. the residual errors after inter-spacecraft calibration should be taken as a measure for the quantities $(\delta x)^2$ and $(\delta g)^2$. More details can be found in Appendix 17.A.1.

17.1.2 Discussion of Competing Error Terms

In the error formulas given above there are two factors which can contribute to the partial derivative inaccuracies. The curly brackets $\{ \}$ contain the influence of the satellite cluster geometry. The other factor comprises the contribution of the two error sources. In order to compare the competing error terms $(\delta g)^2$ and $|\nabla g|^2 (\delta x)^2$ we denote by Λ a typical inter-spacecraft distance and introduce the quantity:

$$L_{\nabla g}^2 = \frac{|g|^2}{|\nabla g|^2}$$

$L_{\nabla g}$ is the gradient scale length of g .

Since small-scale structures like discontinuities and shocks cannot be treated with the finite difference method investigated here, we can restrict ourselves to those cases where the length scale $L_{\nabla g}$ is comparable to Λ or larger. Thus, the quantity

$$\frac{|\nabla g| \delta x}{|g|} \simeq \frac{\delta x}{L_{\nabla g}} = \frac{\Lambda}{L_{\nabla g}} \frac{\delta x}{\Lambda}$$

should be smaller or at least not larger than the relative position error $\delta x / \Lambda$, which we expect to be around 1%. On the other hand, typical values for the relative uncertainties $\delta g / |g|$ for various plasma velocity moments g are of the order of 10%, maybe 5% after inter-spacecraft calibration (see Chapter 6). Such differences are amplified in the error

formulas because square values are added. Thus, with regard to plasma moment derivative estimation it is reasonable to assume

$$\begin{aligned} |\nabla g|^2 (\delta x)^2 &\ll (\delta g)^2 \\ [(\nabla g) \cdot (\nabla h)] (\delta x)^2 &\ll (\delta g)^2, (\delta h)^2 \end{aligned}$$

i.e. to neglect the contribution of spacecraft position inaccuracies against g -measurement errors. Consequently, the cross terms $\overline{(\delta[\partial g/\partial x_i])(\delta[\partial h/\partial x_j])}$ and $\overline{(\delta[\partial g/\partial x_i])(\delta[\partial h/\partial x_i])}$ can be ignored and the other error formulas simplify:

$$\overline{\left(\delta \left[\frac{\partial g}{\partial x_i} \right] \right)^2} \simeq \frac{2(\delta g)^2}{(\Delta x_i)^2} \quad (17.7)$$

$$\overline{\left(\delta \left[\frac{\partial g}{\partial x_i} \right] \right) \left(\delta \left[\frac{\partial g}{\partial x_j} \right] \right)} \simeq \frac{(\delta g)^2}{\Delta x_i \Delta x_j} \quad (17.8)$$

We make use of this approximation in those cases (divergence and curl estimation) where considerable simplifications can be achieved this way.

Errors of magnetic field measurements are usually much smaller than plasma moment inaccuracies. Thus, if e.g. the electric current is determined by the spacecraft cluster via $\nabla \times \mathbf{B}$, then the spacecraft position errors may contribute significantly and thus the cross terms cannot be safely neglected. Error formulas which are deduced on the basis of the above approximation can be understood as a lower limit for the total error in such a case.

17.1.3 Directional Derivative Estimation Accuracy

Directional derivatives, also known as (unit) vector gradients, may serve as a measure for variations of physical quantities along a predefined unit vector $\hat{\mathbf{e}}$, which is given e.g. by the magnetic field direction or a discontinuity normal vector. They also play an important role in time derivative estimation.

The directional derivative of a quantity g can be approximated by:

$$\hat{\mathbf{e}} \cdot \nabla g \equiv (\hat{\mathbf{e}} \cdot \nabla)g = \sum_{i=1}^3 e_i \frac{\partial g}{\partial x_i} \simeq \sum_{i=1}^3 e_i \frac{\Delta_{[i0]}g}{\Delta_{[i0]}x_i}$$

The mean square error of this estimator is given by the following expression if inaccuracies from imperfect knowledge of $\hat{\mathbf{e}}$ are neglected:

$$\begin{aligned} \overline{(\delta [\hat{\mathbf{e}} \cdot \nabla g])^2} &= \overline{\left(\sum_{i=1}^3 e_i \delta \left[\frac{\partial g}{\partial x_i} \right] \right) \left(\sum_{j=1}^3 e_j \delta \left[\frac{\partial g}{\partial x_j} \right] \right)} \\ &= \sum_{i=1}^3 \sum_{j=1}^3 e_i e_j \overline{\left(\delta \left[\frac{\partial g}{\partial x_i} \right] \right) \left(\delta \left[\frac{\partial g}{\partial x_j} \right] \right)} \\ &= \left\{ \sum_{i=1}^3 \sum_{j=1}^3 (1 + \delta_{ij}) \frac{e_i e_j}{\Delta x_i \Delta x_j} \right\} ((\delta g)^2 + |\nabla g|^2 (\delta x)^2) \\ &= \left\{ \left(\sum_{i=1}^3 \frac{e_i}{\Delta x_i} \right)^2 + \sum_{i=1}^3 \left(\frac{e_i}{\Delta x_i} \right)^2 \right\} ((\delta g)^2 + |\nabla g|^2 (\delta x)^2) \quad (17.9) \end{aligned}$$

δ_{ij} denotes the Kronecker symbol ($= 1$ if $i = j$, otherwise $= 0$). The expression in curly brackets is a measure for the sensitivity of the spacecraft cluster in direction \hat{e} . It corresponds to a positive definite quadratic form and thus to an ellipsoidal surface. The directions of highest and lowest sensitivity correspond to principle axes of the ellipsoid.

Uncertainties in \hat{e} introduce additional errors: If the cartesian component errors δe_i are uncorrelated and their mean square values are of equal size $\overline{(\delta e_i)^2} = (\delta e)^2$, the additional terms can be written as $|\nabla g|^2 (\delta e)^2$. We do not pursue this source of error further because its contribution to the total mean square error depends very much on the accuracy of the \hat{e} determination.

In some cases it might be interesting to have a measure for the omnidirectional sensitivity of the spacecraft cluster, i.e. a mean value independent of direction. In order to obtain such a mean value we average over all possible directions, i.e. over all possible unit vectors and denote the result by $\langle \dots \rangle_{\hat{e}}$. This procedure leads to:

$$\left\langle \overline{(\delta [\hat{e} \cdot \nabla g])^2} \right\rangle_{\hat{e}} = \frac{1}{3} \left\{ 2 \sum_{i=1}^3 \frac{1}{(\Delta x_i)^2} \right\} \left((\delta g)^2 + |\nabla g|^2 (\delta x)^2 \right) \quad (17.10)$$

This expression may also serve as a measure for the magnitude error $\overline{(\delta |\nabla g|)^2}$.

17.1.4 Time Derivative Estimation Accuracy

Conservation equations in fluid theories relate spatial derivatives to time derivatives $\partial g / \partial t$ at a fixed location in an inertial coordinate frame. The choice of the inertial system depends on the physical processes of interest. It may be appropriate, for example, to choose a coordinate frame where the magnetospheric boundaries are at rest. On the other hand, satellite measurements lead to time series $g_\alpha(t)$ in a coordinate frame which is *moving* with the spacecraft velocity \mathbf{u}_α relative to the chosen coordinate frame. The corresponding time derivative Dg/Dt can be obtained directly from the measurements and is related to the time derivative at a fixed location by:

$$\left(\frac{Dg}{Dt} \right)_\alpha = \left. \frac{\partial g}{\partial t} \right|_{\mathbf{r}_\alpha} + \mathbf{u}_\alpha \cdot \nabla|_{\mathbf{r}_\alpha} g$$

\mathbf{r}_α denotes the position vector of spacecraft α . Summing up these equations and dividing by 4 yields an expression where average values of various quantities Y appear: $\langle Y_\alpha \rangle_\alpha \equiv (1/4) \sum_{\alpha=0}^3 Y_\alpha$. We denote the deviations from the average values by $\Delta Y_\alpha = Y_\alpha - \langle Y_\alpha \rangle_\alpha$. With the help of a Taylor expansion it can be shown that

$$\begin{aligned} \left. \frac{\partial g}{\partial t} \right|_{\mathbf{r}_{MC}} &\simeq \left\langle \left. \frac{\partial g}{\partial t} \right|_{\mathbf{r}_\alpha} \right\rangle_\alpha \simeq \left\langle \left(\frac{Dg}{Dt} \right)_\alpha \right\rangle_\alpha - \langle \mathbf{u}_\alpha \rangle_\alpha \cdot \nabla|_{\langle \mathbf{r}_\alpha \rangle_\alpha} g \\ &\simeq \frac{d\langle g_\alpha \rangle_\alpha}{dt} - \mathbf{u}_{MC} \cdot \nabla|_{\mathbf{r}_{MC}} g \end{aligned} \quad (17.11)$$

when terms of second order in $\Delta \mathbf{r}_\alpha$ and $\Delta \mathbf{u}_\alpha$ are disregarded. $\mathbf{r}_{MC} \equiv \langle \mathbf{r}_\alpha \rangle_\alpha$ is the mesocentre of the satellite configuration and $\mathbf{u}_{MC} \equiv \langle \mathbf{u}_\alpha \rangle_\alpha$ its velocity.

The directional derivative on the right-hand side of the averaged equation and its accuracy can be estimated with the formulas given in the last section (the spacecraft velocities

can be assumed to be very accurately known). Since the temporal resolution is usually better than the spatial resolution, we expect the moving time derivative Dg/Dt to be more accurate than the directional derivative. Thus, the error of the time derivative $\partial g/\partial t$ should be mainly determined by the error of the directional derivative $\mathbf{u}_{MC} \cdot \nabla g$.

17.1.5 Divergence Estimation Accuracy

The divergence of a vector field \mathbf{V} (such as the bulk velocity, the heat flux vector or the energy flux density vector) can be estimated by the model satellite configuration in Figure 17.1 in the following way:

$$\nabla \cdot \mathbf{V} \simeq \sum_{i=1}^3 \frac{\Delta_{i[0]} V_i}{\Delta x_i} \equiv \sum_{i=1}^3 \frac{V_{i[i]} - V_{i[0]}}{x_{i[i]} - x_{i[0]}} \quad (17.12)$$

This estimator has a mean square error of:

$$\begin{aligned} \overline{(\delta [\nabla \cdot \mathbf{V}])^2} &= \overline{\left(\sum_{i=1}^3 \delta \left[\frac{\partial V_i}{\partial x_i} \right] \right) \left(\sum_{j=1}^3 \delta \left[\frac{\partial V_j}{\partial x_j} \right] \right)} \\ &= \sum_{i=1}^3 \sum_{j=1}^3 \overline{\left(\delta \left[\frac{\partial V_i}{\partial x_i} \right] \right) \left(\delta \left[\frac{\partial V_j}{\partial x_j} \right] \right)} \\ &= \left\{ 2 \sum_{i=1}^3 \frac{1}{(\Delta x_i)^2} \right\} (\delta V)^2 \\ &\quad + \sum_{i=1}^3 \sum_{j=1}^3 (1 + \delta_{ij}) \frac{[(\nabla V_i) \cdot (\nabla V_j)] (\delta x)^2}{\Delta x_i \Delta x_j} \end{aligned} \quad (17.13)$$

It has been assumed that the errors of the different components of \mathbf{V} are uncorrelated and that their mean square values are equal: $(\delta V_i)^2 = (\delta V)^2$.

Considerable simplification is achieved when the approximation $|\nabla V_i|^2 (\delta x)^2 \ll (\delta V)^2$ is used that was discussed earlier:

$$\overline{(\delta [\nabla \cdot \mathbf{V}])^2} \simeq \left\{ 2 \sum_{i=1}^3 \frac{1}{(\Delta x_i)^2} \right\} (\delta V)^2 \quad (17.14)$$

As an example we consider a vector field of the following form

$$\mathbf{V} = \begin{pmatrix} V_1 \\ V_2 \\ V_3 \end{pmatrix} = \frac{\xi}{3} \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} = \frac{\xi}{3} \mathbf{r}$$

where ξ is an arbitrary scalar. It is easy to show that $\nabla \cdot \mathbf{V} = \xi$. The general error formula reduces to:

$$\overline{(\delta [\nabla \cdot \mathbf{V}])^2} = \left\{ 2 \sum_{i=1}^3 \frac{1}{(\Delta x_i)^2} \right\} \left((\delta V)^2 + \frac{\xi^2}{9} (\delta x)^2 \right)$$

17.1.6 Curl Estimation Accuracy: The General Case

The curl of the bulk velocity \mathbf{V} is called the vorticity: $\boldsymbol{\Omega} = \nabla \times \mathbf{V}$. In Chapter 6 the physical importance of this quantity is discussed. In the following we investigate the accuracy of a vorticity estimator based on the model satellite configuration sketched in Figure 17.1. The results can be easily transferred to other vector fields such as the heat flux vector or the energy flux density vector.

Cartesian Components of the Vorticity Vector

The i -th component of the vorticity is given by the formula:

$$\Omega_i = (\nabla \times \mathbf{V})_i = \frac{\partial V_k}{\partial x_j} - \frac{\partial V_j}{\partial x_k} \simeq \frac{\Delta_{[j0]} V_k}{\Delta x_j} - \frac{\Delta_{[k0]} V_j}{\Delta x_k} \quad (17.15)$$

where (i, j, k) must be a cyclic permutation of $(1, 2, 3)$. The procedure to compute the mean square error is similar to the divergence estimation case. The expression

$$\delta \Omega_i = \delta \left[\frac{\partial V_k}{\partial x_j} \right] - \delta \left[\frac{\partial V_j}{\partial x_k} \right] \quad (17.16)$$

is squared and ensemble averaged. We finally obtain:

$$\begin{aligned} \overline{(\delta \Omega_i)^2} &= \overline{\left(\delta \left[\frac{\partial V_k}{\partial x_j} \right] \right)^2} + \overline{\left(\delta \left[\frac{\partial V_j}{\partial x_k} \right] \right)^2} - 2 \overline{\left(\delta \left[\frac{\partial V_k}{\partial x_j} \right] \right) \left(\delta \left[\frac{\partial V_j}{\partial x_k} \right] \right)} \\ &= \left\{ \frac{2}{(\Delta x_j)^2} + \frac{2}{(\Delta x_k)^2} \right\} (\delta V)^2 \\ &\quad + \left\{ \frac{2|\nabla V_k|^2}{(\Delta x_j)^2} + \frac{2|\nabla V_j|^2}{(\Delta x_k)^2} - \frac{2(\nabla V_k) \cdot (\nabla V_j)}{\Delta x_j \Delta x_k} \right\} (\delta x)^2 \end{aligned} \quad (17.17)$$

As in the divergence estimation case, using $|\nabla V_{j,k}|^2 (\delta x)^2 \ll (\delta V)^2$ leads to a much simpler expression:

$$\overline{(\delta \Omega_i)^2} \simeq \left\{ \frac{2}{(\Delta x_j)^2} + \frac{2}{(\Delta x_k)^2} \right\} (\delta V)^2 \quad (17.18)$$

The following shear field can serve as an example:

$$\mathbf{V} = \frac{1}{2} \boldsymbol{\Omega} \times \mathbf{r} = \frac{1}{2} \begin{pmatrix} \Omega_1 \\ \Omega_2 \\ \Omega_3 \end{pmatrix} \times \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix}$$

In this case the mean square error is given by:

$$\overline{(\delta \Omega_i)^2} = \left\{ \frac{2}{(\Delta x_j)^2} + \frac{2}{(\Delta x_k)^2} \right\} (\delta V)^2 + \frac{1}{2} \left\{ \frac{\Omega_i^2 + \Omega_j^2}{(\Delta x_j)^2} + \frac{\Omega_i^2 + \Omega_k^2}{(\Delta x_k)^2} + \frac{\Omega_j \Omega_k}{\Delta x_j \Delta x_k} \right\} (\delta x)^2$$

It should be noted that the curl of a vector field depends only on the anti-symmetric part $\mathbf{J}_A = (\mathbf{J} - \mathbf{J}^T)/2$ of the (jacobian) matrix $\mathbf{J} = (\nabla \mathbf{V})^T$, i.e. \mathbf{J} is the transpose of the

dyadic tensor $\nabla \mathbf{V}$: $J_{ij} = (\nabla \mathbf{V})_{ji} = \partial V_i / \partial x_j$. The example vector field is special with regard to curl estimation because it can be shown that its jacobian matrix is already anti-symmetric. A non-zero symmetric part $\mathbf{J}_S = (\mathbf{J} + \mathbf{J}^T)/2$ contributes further error terms which are proportional to the spacecraft position uncertainties $(\delta x)^2$ and thus leads to even more complicated expressions. Therefore, in the remaining part of this section we make use of the approximation $|\nabla V_{j,k}|^2 (\delta x)^2 \ll (\delta V)^2$ which is well justified in the Earth's magnetosphere for typical values of the errors $(\delta V)^2$ and $(\delta x)^2$ as we pointed out in a previous section.

Magnitude of the Vorticity Vector

The mean square error of the vorticity magnitude $\Omega \equiv |\Omega|$ is computed in the appendix (Chapter 17.A.2). The result is:

$$\begin{aligned} \Omega^2 \overline{(\delta \Omega)^2} &= 2 \left\{ \frac{\Omega_2^2 + \Omega_3^2}{(\Delta x_1)^2} + \frac{\Omega_3^2 + \Omega_1^2}{(\Delta x_2)^2} + \frac{\Omega_1^2 + \Omega_2^2}{(\Delta x_3)^2} \right\} (\delta V)^2 \\ &\quad - 2 \left\{ \frac{\Omega_1 \Omega_2}{\Delta x_1 \Delta x_2} + \frac{\Omega_2 \Omega_3}{\Delta x_2 \Delta x_3} + \frac{\Omega_3 \Omega_1}{\Delta x_3 \Delta x_1} \right\} (\delta V)^2 \end{aligned} \quad (17.19)$$

If we are interested mainly in the influence of the shape of the satellite configuration on the magnitude error we can get rid of the vorticity direction dependence by performing an average over the sphere $\Omega = \text{constant}$. This procedure leads to the following expression:

$$\left\langle \overline{(\delta \Omega)^2} \right\rangle_{\Omega=\text{const}} = \frac{2}{3} \left\{ 2 \sum_{i=1}^3 \frac{1}{(\Delta x_i)^2} \right\} (\delta V)^2 \quad (17.20)$$

The influence of the satellite configuration on the averaged magnitude error is contained in the expression in $\{\}$ brackets. We may refer to it as a geometric factor of the spacecraft tetrahedron. It is dominated by the shortest separation length of the configuration. This expression appeared in the equations for the divergence estimation error and the averaged directional derivative error as well. It is a key quantity with regard to the accuracy of derivative estimation by the spacecraft cluster.

17.1.7 Curl Estimation Accuracy: The Special Case $\Delta x_1 = \Delta x_2$

In this section we investigate the special case where two separation distances Δx_i are equal. In order to distinguish the special from the general case we introduce the subscripts ' \perp ' and ' \parallel ' by $\Delta x_\perp = \Delta x_1 = \Delta x_2$ and $\Delta x_\parallel = \Delta x_3$ (this notation should not be confused with magnetic-field-aligned coordinate systems or the like). It is now appropriate to express the vorticity as:

$$\mathbf{\Omega} = \begin{pmatrix} \Omega_1 \\ \Omega_2 \\ \Omega_3 \end{pmatrix} = \begin{pmatrix} \Omega_\perp \cos \lambda \\ \Omega_\perp \sin \lambda \\ \Omega_\parallel \end{pmatrix} = \begin{pmatrix} \Omega \cos \lambda \sin \phi \\ \Omega \sin \lambda \sin \phi \\ \Omega \cos \phi \end{pmatrix}$$

λ is the azimuth measured from the x_1 axis and ϕ the polar angle of the vector $\mathbf{\Omega}$ in a polar coordinate system where the polar axis is along $x_3 \equiv x_\parallel$. The situation is sketched in Figure 17.2 for the case $\lambda = 0^\circ$.

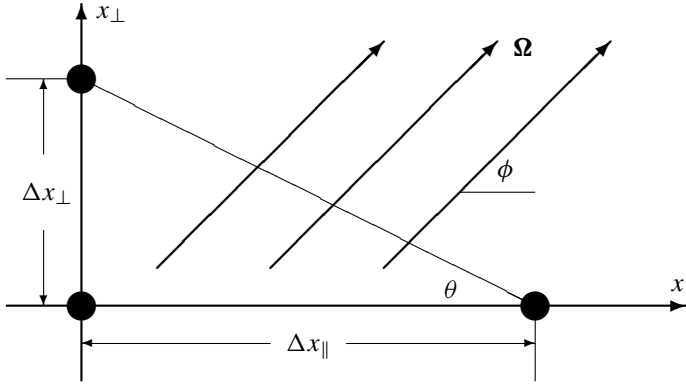


Figure 17.2: The special case $\Delta x_1 = \Delta x_2 = \Delta x_\perp$ (and $\Delta x_3 = \Delta x_\parallel$): Sketch of the spacecraft positions, the vorticity field and the angles ϕ and θ for the case $\lambda = 0^\circ$. The fourth spacecraft is at distance Δx_\perp out of the plane of the figure.

The Errors $\delta\Omega_\parallel$, $\delta\Omega_\perp$, and $\delta\lambda$

Since the parallel component is cartesian, the mean square error is given by:

$$\overline{(\delta\Omega_\parallel)^2} \equiv \overline{(\delta\Omega_3)^2} = \frac{4(\delta V)^2}{(\Delta x_\perp)^2} \quad (17.21)$$

This expression does not depend on the angle λ .

The mean square errors of Ω_\perp and λ are calculated in Appendix 17.A.3. The results are:

$$\overline{(\delta\Omega_\perp)^2} = \left\{ \frac{2 - \sin 2\lambda}{(\Delta x_\perp)^2} + \frac{2}{(\Delta x_\parallel)^2} \right\} (\delta V)^2 \quad (17.22)$$

$$\Omega_\perp^2 \overline{(\delta\lambda)^2} = \left\{ \frac{2 + \sin 2\lambda}{(\Delta x_\perp)^2} + \frac{2}{(\Delta x_\parallel)^2} \right\} (\delta V)^2 \quad (17.23)$$

In Appendix 17.A.3 it is pointed out that the dependence of the errors on the angle λ is of minor importance. Thus, in most cases it should be sufficient to use the mean values (i.e. averaged over the angle λ):

$$\left\langle \Omega_\perp^2 \overline{(\delta\lambda)^2} \right\rangle_\lambda = \left\langle \overline{(\delta\Omega_\perp)^2} \right\rangle_\lambda = \left\{ \frac{2}{(\Delta x_\perp)^2} + \frac{2}{(\Delta x_\parallel)^2} \right\} (\delta V)^2 \quad (17.24)$$

The mean values coincide with the actual values if $\Omega_1 = 0$ or $\Omega_2 = 0$.

The Influence of $\phi = \tan^{-1}(\Omega_\perp / \Omega_\parallel)$

The polar angle ϕ is related to the vorticity components and magnitude:

$$\begin{aligned} \Omega_\parallel &= \Omega \cos \phi \\ \Omega_\perp &= \Omega \sin \phi \end{aligned}$$

In the following we concentrate on the influence of the polar angle ϕ on the errors $\overline{(\delta\Omega)^2}$ and $\Omega^2\overline{(\delta\phi)^2}$. To eliminate the azimuthal dependence we perform averages over the angle λ .

Three configurations are of special interest:

$$\begin{aligned} \Delta x_{\parallel} &\gg \Delta x_{\perp} & : & \quad \text{“long” tetrahedron} \\ \Delta x_{\parallel} &\simeq \Delta x_{\perp} & : & \quad \text{“regular” tetrahedron} \\ \Delta x_{\parallel} &\ll \Delta x_{\perp} & : & \quad \text{“flat” tetrahedron} \end{aligned}$$

The angle θ , defined through:

$$\begin{aligned} \Delta x_{\parallel} &= \Delta r \cos \theta \\ \Delta x_{\perp} &= \Delta r \sin \theta \end{aligned}$$

where $\Delta r = \sqrt{(\Delta x_{\parallel})^2 + (\Delta x_{\perp})^2}$, may serve as a measure for the shape of the satellite configuration: $\tan \theta = \Delta x_{\perp} / \Delta x_{\parallel}$ (see also Figure 17.2).

The mean square magnitude error $\overline{(\delta\Omega)^2}$ and the mean square directional error $\overline{(\delta\phi)^2}$ are calculated in Appendix 17.A.3. The results are:

$$\begin{aligned} \frac{\langle \overline{(\delta\Omega)^2} \rangle_{\lambda}}{\Omega^2} &= \left\{ \frac{2}{(\Omega\Delta x_{\perp})^2} + \frac{2}{(\Omega\Delta x_{\parallel})^2} \right\} \sin^2 \phi (\delta V)^2 + \frac{4 \cos^2 \phi (\delta V)^2}{(\Omega\Delta x_{\perp})^2} \\ &= \left\{ \frac{2}{\sin^2 \theta} \right\} \frac{(\delta V)^2}{(\Omega\Delta r)^2} + 2 \left\{ \frac{\cos^2 \phi}{\sin^2 \theta} + \frac{\sin^2 \phi}{\cos^2 \theta} \right\} \frac{(\delta V)^2}{(\Omega\Delta r)^2} \quad (17.25) \end{aligned}$$

$$\begin{aligned} \overline{(\delta\phi)^2} &= \left\{ \frac{2}{(\Omega\Delta x_{\perp})^2} + \frac{2}{(\Omega\Delta x_{\parallel})^2} \right\} \cos^2 \phi (\delta V)^2 + \frac{4 \sin^2 \phi (\delta V)^2}{(\Omega\Delta x_{\perp})^2} \\ &= \left\{ \frac{2}{\sin^2 \theta} \right\} \frac{(\delta V)^2}{(\Omega\Delta r)^2} + 2 \left\{ \frac{\sin^2 \phi}{\sin^2 \theta} + \frac{\cos^2 \phi}{\cos^2 \theta} \right\} \frac{(\delta V)^2}{(\Omega\Delta r)^2} \quad (17.26) \end{aligned}$$

The scaling quantity $\Omega\Delta r$ is a measure for the typical bulk velocity jump over the distance Δr .

The relative magnitude error $\delta\Omega/\Omega \equiv \left(\langle \overline{(\delta\Omega)^2} \rangle_{\lambda} / \Omega^2 \right)^{1/2}$ (i.e. the normalised standard deviation of Ω) as a function of the angles θ and ϕ is shown in Figure 17.3. A value of 1% for the scaled error $\delta V / (\Omega\Delta r)$ has been chosen for convenience. Since the standard deviation $\delta\Omega/\Omega$ is proportional to the scaled error $\delta V / (\Omega\Delta r)$, the quality of the results does not depend on it. A value of $p\%$ yields errors which are simply p times higher. Realistic error values in the Earth’s magnetosphere are discussed in Section 17.3.

In Figure 17.3 there is a strong divergence for $\theta \rightarrow 0^\circ$ and $\theta \rightarrow 90^\circ$. For θ near 0° (long tetrahedra) the error is very large for all vorticity directions, i.e. there is no particular direction which allows a much more accurate determination of the magnitude than other directions. For moderate values of θ (more or less regular tetrahedra) the variability of the relative error $\delta\Omega/\Omega$ with the angle ϕ becomes rather weak. This picture changes for very flat tetrahedra, $\theta \approx 90^\circ$. In general the error becomes rather large, but if the vorticity vector is parallel to the x_{\parallel} -direction, $\phi \approx 0^\circ$, the magnitude can be determined quite accurately at the same time.

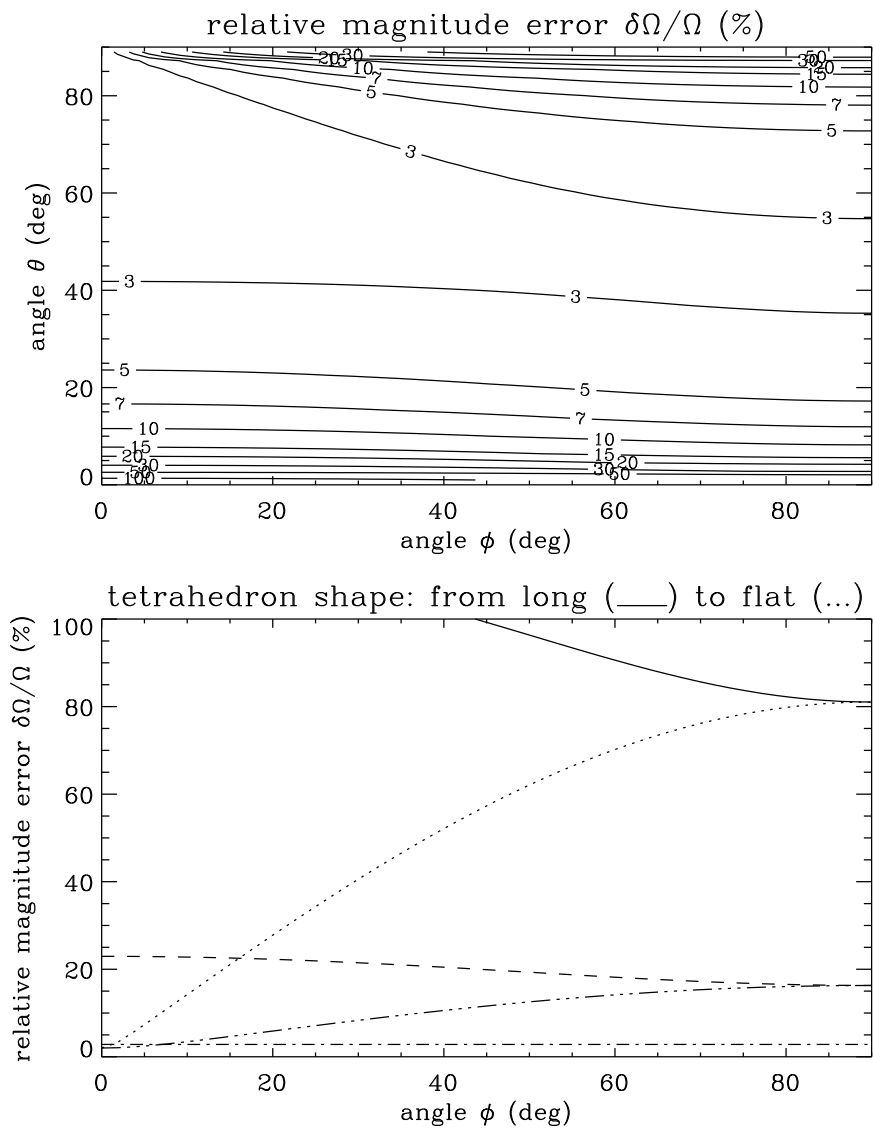


Figure 17.3: Relative mean error of the vorticity magnitude as a function of angles θ and ϕ . A scaled error $\delta V/(\Omega\Delta r)$ of 1% has been assumed. Top: contour plot. Bottom: magnitude error as a function of the angle ϕ for five distinct values of θ : 1° (—), 5° (---), 45° (- · -), 85° (- - - -), 89° (···).

Figure 17.4 shows the directional error $\delta\phi \equiv \left(\langle (\delta\phi)^2 \rangle_\lambda \right)^{1/2}$ for the same value of the scaled error as above. Once again a strong divergence of the error is found for $\theta \rightarrow 0^\circ$ and $\theta \rightarrow 90^\circ$. Very long tetrahedra lead to large $\delta\phi$ for all vorticity directions. The error becomes much smaller for moderate values of θ , i.e. regular tetrahedra. The behaviour of $\delta\phi$ for very flat satellite configuration, $\theta \approx 90^\circ$, is in contrast to the magnitude error: the vorticity direction may be determined with an acceptable accuracy only when $\phi \approx 90^\circ$, i.e. when the vorticity vector is perpendicular to the x_\parallel -direction. We conclude that for a degenerate configuration there may be special cases where either the vorticity direction or the magnitude can be determined with some accuracy, but never both quantities.

17.2 Arbitrary Satellite Configurations

The previous section referred to the special satellite configuration sketched in Figure 17.1. In this section we turn to the general case, i.e. four-point derivatives from arbitrary cluster tetrahedra. They can be obtained with the help of barycentric coordinates which are briefly reviewed below. The error formulas derived with the help of this method allow to define a cluster geometric factor for the general case

17.2.1 Derivatives from Barycentric Coordinates

Barycentric coordinates are discussed extensively in Chapter 14. Briefly, they provide a convenient means to linearly interpolate a physical quantity g inside a satellite cluster tetrahedron by using the measured values g_α at the four spacecraft positions \mathbf{r}_α :

$$\tilde{g}(\mathbf{r}) = \sum_{\alpha=0}^3 \mu_\alpha(\mathbf{r}) g_\alpha$$

where:

$$\mu_\alpha(\mathbf{r}) = 1 + \mathbf{k}_\alpha \cdot (\mathbf{r} - \mathbf{r}_\alpha)$$

\tilde{g} denotes the linear function that interpolates between the measurements. The vectors \mathbf{k}_α are given by the formula:

$$\mathbf{k}_\alpha = \frac{\mathbf{r}_{\beta\gamma} \times \mathbf{r}_{\beta\lambda}}{\mathbf{r}_{\beta\alpha} \cdot (\mathbf{r}_{\beta\gamma} \times \mathbf{r}_{\beta\lambda})}$$

$(\alpha, \beta, \gamma, \lambda)$ must be a permutation of $(0, 1, 2, 3)$. Relative position vectors are denoted by $\mathbf{r}_{\alpha\beta} = \mathbf{r}_\beta - \mathbf{r}_\alpha$. The set $\{\mathbf{k}_\alpha\}$ is called the reciprocal base of the tetrahedron.

Vector functions, such as \mathbf{V} , can be handled in a similar way by applying the above formulas to the cartesian components. Since \tilde{g} and $\tilde{\mathbf{V}}$ are linear functions, the calculation of the derivatives can be done quite easily. The results are:

$$\nabla g \simeq \nabla \tilde{g} = \sum_{\alpha=0}^3 \mathbf{k}_\alpha g_\alpha \quad (17.27)$$

$$\hat{\mathbf{e}} \cdot \nabla g \simeq \hat{\mathbf{e}} \cdot \nabla \tilde{g} = \sum_{\alpha=0}^3 (\hat{\mathbf{e}} \cdot \mathbf{k}_\alpha) g_\alpha \quad (17.28)$$

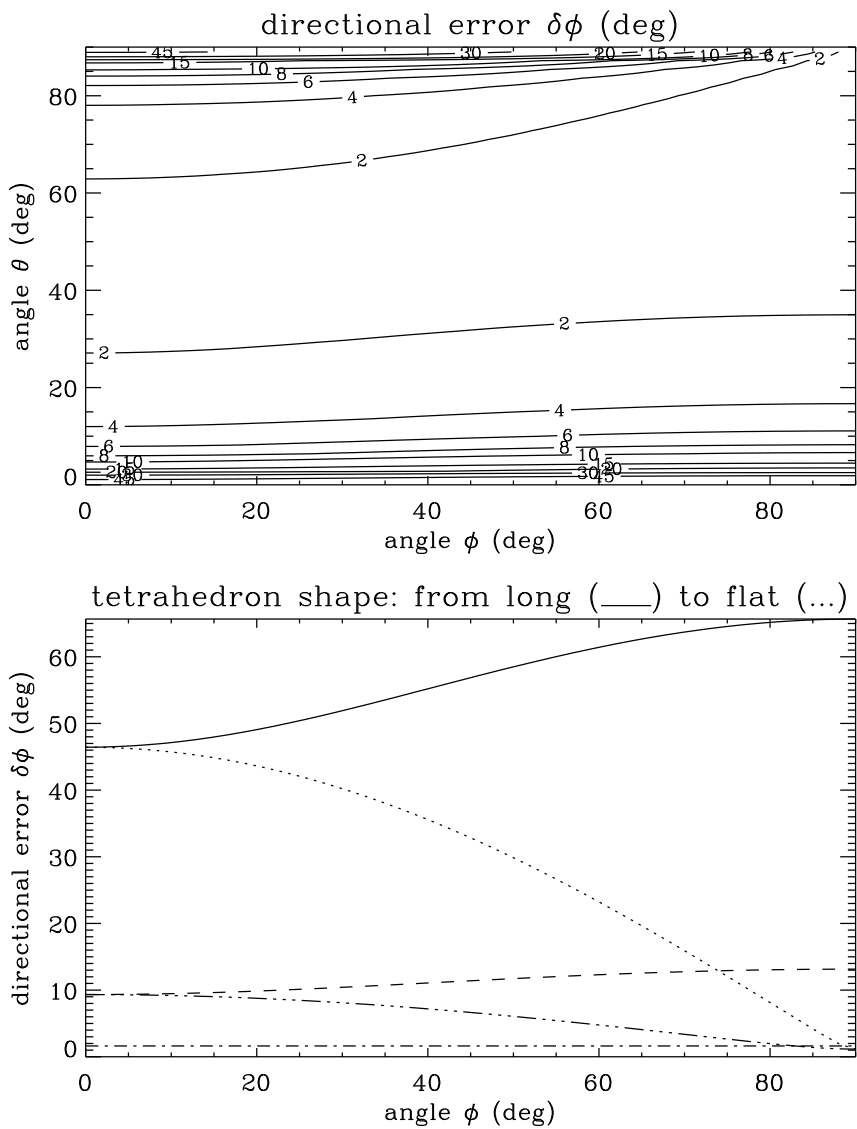


Figure 17.4: Mean error of the vorticity direction as a function of the angles θ and ϕ . A scaled error $\delta V/(\Omega\Delta r)$ of 1% has been assumed. Top: contour plots. Bottom: directional error as a function of the angle ϕ for five distinct values of θ : 1° (—), 5° (---), 45° (- · -), 85° (- - - -), 89° (···).

$$\nabla \cdot \mathbf{V} \simeq \nabla \cdot \tilde{\mathbf{V}} = \sum_{\alpha=0}^3 \mathbf{k}_{\alpha} \cdot \mathbf{V}_{\alpha} \quad (17.29)$$

$$\nabla \times \mathbf{V} \simeq \nabla \times \tilde{\mathbf{V}} = \sum_{\alpha=0}^3 \mathbf{k}_{\alpha} \times \mathbf{V}_{\alpha} \quad (17.30)$$

The element (i, j) of the matrix $\nabla \mathbf{V}$ is given by:

$$\frac{\partial V_j}{\partial x_i} \equiv (\nabla \mathbf{V})_{ij} \simeq \sum_{\alpha=0}^3 (\mathbf{k}_{\alpha} \mathbf{V}_{\alpha})_{ij} \equiv \sum_{\alpha=0}^3 k_{\alpha i} V_{\alpha j} \quad (17.31)$$

With regard to error estimation it is important to notice that $\nabla \times \mathbf{V}$ and $\nabla \cdot \mathbf{V}$ are just linear combinations of various $(\nabla \mathbf{V})_{ij}$'s and thus of terms like $k_{\alpha i} V_{\alpha j}$, with $i = j$ or $i \neq j$.

Local time derivatives such as $\partial g / \partial t$ can be estimated with the help of moving time derivatives Dg/Dt and the directional derivatives $\mathbf{u}_{MC} \cdot \nabla g$ in the same way as discussed in section 17.1.4 (\mathbf{u}_{MC} is the velocity of the mesocentre of the spacecraft cluster).

It should be noted that there are other methods to compute estimates from multi-satellite measurements for the above derivatives, e.g. contour integrals or a combination of the measured finite differences. All these methods have to interpolate the measurements in a certain way, at least implicitly. As long as the underlying interpolation is linear—and that is all one can get from a single set of four point measurements—the various methods must lead to identical results because of the uniqueness of the linear interpolation function. Higher-order interpolations are possible only when measurements at more than one time are used.

17.2.2 Error Estimation

Mean square error formulas for various derivative operators in the general case are derived in Appendix 17.B. As discussed in a previous section, contributions from the spacecraft position uncertainties should be of minor importance to these errors. Thus, only measurement errors such as $(\delta g)^2$ and $(\delta V)^2$ are considered. Furthermore, correlations between measurement errors of different spacecraft are neglected. Different components of vector field measurement errors are treated as uncorrelated quantities. We summarise the results from Appendix 17.B. ⇒17.1

The mean square error of a directional derivative can be written as: ⇒17.2

$$\overline{(\delta[\hat{\mathbf{e}} \cdot \nabla g])^2} = \left\{ \sum_{\alpha=0}^3 (\hat{\mathbf{e}} \cdot \mathbf{k}_{\alpha})^2 \right\} (\delta g)^2 \quad (17.32)$$

In particular, the mean square errors of partial derivatives such as $\partial g / \partial x_i$ and $\partial V_j / \partial x_i$ are given by:

$$\overline{\left(\delta \left[\frac{\partial g}{\partial x_i} \right] \right)^2} = \left\{ \sum_{\alpha=0}^3 k_{\alpha i}^2 \right\} (\delta g)^2 \quad (17.33)$$

$$\overline{\left(\delta \left(\frac{\partial V_j}{\partial x_i} \right) \right)^2} = \left\{ \sum_{\alpha=0}^3 k_{\alpha i}^2 \right\} (\delta V)^2 \quad (17.34)$$

Averaging the directional derivative error over all possible unit vectors $\hat{\mathbf{e}}$ leads to a mean value which is independent of direction:

$$\left\langle \overline{(\delta [\hat{\mathbf{e}} \cdot \nabla g])^2} \right\rangle_{\hat{\mathbf{e}}} = \frac{1}{3} \left\{ \sum_{\alpha} |\mathbf{k}_{\alpha}|^2 \right\} (\delta g)^2 \quad (17.35)$$

The divergence estimation error is given by:

$$\overline{(\delta [\nabla \cdot \mathbf{V}])^2} = \left\{ \sum_{\alpha=0}^3 |\mathbf{k}_{\alpha}|^2 \right\} (\delta V)^2 \quad (17.36)$$

For the i -th component of the vorticity we find:

$$\overline{(\delta \Omega_i)^2} = \left\{ \sum_{\alpha=0}^3 \sum_{l=1}^3 (\hat{\mathbf{e}}_i \times \mathbf{k}_{\alpha})_l^2 \right\} (\delta V)^2 \equiv \left\{ \sum_{\alpha=0}^3 \sum_{j \neq i} k_{\alpha j}^2 \right\} (\delta V)^2 \quad (17.37)$$

where $\hat{\mathbf{e}}_i$ denotes the unit vector in the x_i -direction.

The mean square error of the vorticity magnitude $\Omega \equiv |\boldsymbol{\Omega}| \equiv |\nabla \times \mathbf{V}|$ can be written as:

$$\Omega^2 \overline{(\delta \Omega)^2} = \left\{ \Omega^2 \sum_{\alpha} |\mathbf{k}_{\alpha}|^2 - \sum_{i, \alpha, m} \Omega_i \Omega_m k_{\alpha m} k_{\alpha i} \right\} (\delta V)^2 \quad (17.38)$$

If the influence of the vorticity direction on the error is of minor interest, an average formula over the sphere $\Omega = \text{constant}$ could be useful:

$$\left\langle (\delta \Omega)^2 \right\rangle_{\Omega=\text{const}} = \frac{2}{3} \left\{ \sum_{\alpha=0}^3 |\mathbf{k}_{\alpha}|^2 \right\} (\delta V)^2 \quad (17.39)$$

The terms in curly brackets contain the influence of the satellite configuration on the above mentioned square errors. We may refer to them as various geometric factors of the spacecraft cluster (CGF).

It is easy to check the following relation for the special satellite configuration sketched in figure 17.1:

$$\left\{ \sum_{\alpha=0}^3 (k_{\alpha i})^2 \right\} = \frac{2}{(\Delta x_i)^2}$$

which leads to:

$$\left\{ \sum_{\alpha=0}^3 |\mathbf{k}_{\alpha}|^2 \right\} = 2 \sum_{i=1}^3 \frac{1}{(\Delta x_i)^2}$$

With the help of these expression the formulas derived in the previous sections can be regained.

Of special interest is the sum $\left\{ \sum_{\alpha=0}^3 |\mathbf{k}_{\alpha}|^2 \right\}$ which appears in both the curl and divergence estimation case. If the factor $2/3$ in the curl estimation error formula is disregarded, an estimate for the accuracy of both the curl and the divergence determination can be written as:

$$\left\langle \overline{(\delta \Omega)^2} \right\rangle_{\Omega=\text{const}} \simeq \overline{(\delta [\nabla \cdot \mathbf{V}])^2} \simeq \left\{ \sum_{\alpha=0}^3 |\mathbf{k}_{\alpha}|^2 \right\} (\delta V)^2 \quad (17.40)$$

This should be a useful expression because it is applicable to arbitrary satellite configurations and very easy to handle. ⇒17.3

The Error Formulas and the Volumetric Tensor

In Chapter 12, section 12.4, the volumetric tensor \mathbf{R} is introduced. In the mesocentre frame of reference, i.e. $\sum_{\alpha=0}^3 \mathbf{r}_{\alpha} = 0$, it can be written as $\mathbf{R} = \frac{1}{4} \sum_{\alpha=0}^3 \mathbf{r}_{\alpha} \mathbf{r}_{\alpha}^T$ (be aware that \mathbf{r}_{α} denotes a column vector and \mathbf{r}_{α}^T a row vector). Since the inverse of the volumetric tensor is related to the reciprocal vectors by $\mathbf{R}^{-1} = 4 \sum_{\alpha=0}^3 \mathbf{k}_{\alpha} \mathbf{k}_{\alpha}^T$, the formula for the mean square error of a directional derivative (see above) can be rearranged: ⇒17.4

$$\overline{(\delta[\hat{\mathbf{e}} \cdot \nabla g])^2} / (\delta g)^2 = \left\{ \sum_{\alpha=0}^3 (\hat{\mathbf{e}} \cdot \mathbf{k}_{\alpha})^2 \right\} = \hat{\mathbf{e}}^T \left(\sum_{\alpha=0}^3 \mathbf{k}_{\alpha} \mathbf{k}_{\alpha}^T \right) \hat{\mathbf{e}} = \hat{\mathbf{e}}^T \mathbf{R}^{-1} \hat{\mathbf{e}} / 4 \quad (17.41)$$

This is the quadratic form associated with the matrix \mathbf{R}^{-1} . If $\hat{\mathbf{e}}$ is a unit eigenvector, we obtain the inverse of the corresponding characteristic dimension squared. Therefore, the characteristic directions minimise/maximise the quartic form and thus the directional error.

Furthermore, the quantity $\left\{ \sum_{\alpha=0}^3 |\mathbf{k}_{\alpha}|^2 \right\}$, which can be interpreted as a cluster geometric factor, is 0.25 times the trace of the matrix \mathbf{R}^{-1} . ⇒17.4

17.3 Application to the Earth's Magnetosphere

In the following we discuss the application of the obtained results to the Earth's magnetosphere. The error formulas are evaluated by inserting typical values. A simple procedure to estimate plasma moment derivatives and their accuracies is suggested. We start with a short discussion of a time averaging process which can improve the accuracy of plasma moment derivative estimation significantly.

Time Averaging

The error formulas derived in this study are valid for a set of four-point measurements at a single time. A time average leads to mean values for the derivatives of interest which usually have an improved statistical significance compared to the corresponding single measurements (in time). Such a time average process can only be done at the cost of time resolution. But in our case this should not matter too much because we expect the time resolution (times the cluster velocity u) to be much better than the resolution in space. Furthermore, to be consistent with the assumption of linear varying velocity fields, the derivative of interest should not deviate much from its mean value on a length scale comparable to a typical inter-spacecraft distance Λ (or some fraction of it). It may be prescribed or computed from the relative spacecraft positions $\mathbf{r}_{\alpha\beta}$ with the help of the formula:

$$\Lambda^2 = \frac{1}{6} \sum_{\alpha < \beta} |\mathbf{r}_{\alpha\beta}|^2$$

Therefore, as long as the time interval is smaller than Λ divided by the cluster velocity u , averaging may improve the derivative estimation procedure: if N points in time are used, standard deviations are expected to decrease by a factor of \sqrt{N} .

Nonlinear Variations

Since the barycentric coordinate method provides basically a linear interpolation between the measured values of the cluster tetrahedron, the derivatives computed with this method are meaningful only as long as the underlying physical quantity is predominantly a linear function of the space coordinates. Nonlinear variations imply non-constant derivatives. Thus, as a measure of their influence we suggest the square deviation of an individual partial derivative from its mean value:

$$\left(\delta \left[\frac{\partial g}{\partial x_i} \right] \right)_{\text{nonlin}}^2 \equiv \left\langle \left(\frac{\partial g}{\partial x_i} - \left\langle \frac{\partial g}{\partial x_i} \right\rangle_{\Lambda} \right)^2 \right\rangle_{\Lambda} \quad (17.42)$$

$\langle \dots \rangle_{\Lambda}$ denotes a (time) average over some distance along the cluster orbit whose length should be of the order of a typical inter-spacecraft distance Λ (see above). This formula can be easily transferred to the various components of the jacobian matrix of a vector field. With regard to curl and divergence estimation we suggest also the consideration of the quantities:

$$(\delta \Omega)_{\text{nonlin}}^2 \equiv \left\langle (\Omega - \langle \Omega \rangle_{\Lambda})^2 \right\rangle_{\Lambda} \quad (17.43)$$

and

$$(\delta [\nabla \cdot \mathbf{V}])_{\text{nonlin}}^2 \equiv \left\langle (\nabla \cdot \mathbf{V} - \langle \nabla \cdot \mathbf{V} \rangle_{\Lambda})^2 \right\rangle_{\Lambda} \quad (17.44)$$

These quantities should be compared with a threshold value to judge whether the linear assumption is justified or not. The square mean value of the corresponding derivative may be useful in such a comparison. However, this should be tested with real (or simulated) data.

Approximate Evaluation of the Error Formulas

Now we try to get an idea of the accuracies that can be achieved in the plasma moment derivative estimation process. The error formulas derived earlier are evaluated with the help of typical instrument and spacecraft position inaccuracies and expected values for velocity inhomogeneities in the Earth's magnetosphere.

We look at the particle pressure gradient first and examine the accuracy of the i -th partial derivative with the help of the error formulas based on the simplified cluster geometry sketched in Figure 17.1:

$$\begin{aligned} \overline{\left(\delta \left[\frac{\partial p}{\partial x_i} \right] \right)^2} &= \left\{ \frac{2}{(\Delta x_i)^2} \right\} \left((\delta p)^2 + |\nabla p|^2 (\delta x)^2 \right) \\ &= \left\{ \frac{2|p|^2}{(\Delta x_i)^2} \right\} \left(\frac{(\delta p)^2}{|p|^2} + \frac{|\nabla p|^2}{|p|^2} (\delta x)^2 \right) \\ &= \left\{ \frac{2|p|^2}{(\Delta x_i)^2} \right\} \left(\frac{(\delta p)^2}{|p|^2} + \frac{(\delta x)^2}{L_{\nabla p}^2} \right) \end{aligned}$$

The gradient length $L_{\nabla p}$ is implicitly defined through the last relation (see also Section 17.1.2). For an arbitrary spacecraft configuration the factor $\{2/(\Delta x_i)^2\}$ must be replaced by $\{\sum_{\alpha=0}^3 (k_{\alpha i})^2\}$. As pointed out in a previous section, we expect the relative error

$(\delta p)^2/|p|^2$ to be larger than the scaled error $(\delta x)^2/L_{\nabla p}^2$. Thus, the normalised standard deviation is given by:

$$\frac{\delta[\partial p/\partial x_i]}{|\partial p/\partial x_i|} \simeq \frac{\sqrt{2} \delta p}{\Delta x_i |\partial p/\partial x_i|} = \frac{\sqrt{2} L_{\partial_i p}}{\Delta x_i} \frac{\delta p}{|p|} \quad (17.45)$$

with the length scale $L_{\partial_i p} = |p|/|\partial p/\partial x_i|$. If this quantity is of the order of the inter-spacecraft distance Δx_i , the (normalised) error of the i -th partial derivative is of the order of the relative error $\delta p/|p|$ (around 10%, for details see Chapter 6). The error of the i -th partial derivative increases with larger length scales.

As a second example we use the $\nabla \times \mathbf{V}$ error formulas of the symmetric case (equations 17.25 and 17.26) with $\phi \simeq 45^\circ$ and $\theta \simeq 45^\circ$ and get for the standard deviations:

$$\frac{\delta \Omega}{\Omega} \simeq \delta \phi \simeq \sqrt{8} \frac{\delta V}{\Omega \Delta r}$$

Usually the quantity $\delta V/V$ is better known than the scaled error $\delta V/(\Omega \Delta r)$. They are related in the following way:

$$\frac{\delta V}{\Omega \Delta r} = \left(\frac{\delta V}{V} \right) \frac{(V/\Omega)}{\Delta r} \equiv \left(\frac{\delta V}{V} \right) \frac{L_{\text{shear}}}{\Delta r}$$

where the shear length has been implicitly defined as $L_{\text{shear}} = V/\Omega$. If absolute calibration accuracy is considered, the error $\delta V/V$ should be around 10% (see Chapter 6 for details). More important for derivative estimation is the relative calibration accuracy of the plasma instruments on the different spacecraft: on such a basis the error $\delta V/V$ can be assumed to be 5%.

If the shear length is of the order of the inter-spacecraft distance $\Delta r \simeq \Lambda$, the scaled error $\delta V/(\Omega \Delta r)$ will be around 10%. Inserting this value into the above mentioned error formulas yields $\delta \Omega/\Omega \simeq \delta \phi \simeq 30\%$. As noted above, time averaging can improve this accuracy. The time resolution of the measured bulk velocity \mathbf{V} can be assumed to be about 4 seconds, and the relative motion of the spacecraft cluster with respect to the magnetospheric boundaries is of the order of 10 km/s. Thus, along an orbit trajectory of length $\Lambda = 1000$ km there are about 25 measurements of the derivatives, which improves the accuracy by a factor of $\sqrt{25} = 5$. Thus, the mean error of the time averaged vorticity is of the order: $\delta \langle \Omega \rangle_\Lambda / \langle \Omega \rangle_\Lambda \simeq \delta \langle \phi \rangle_\Lambda \simeq 5\%$. The situation gets worse for degenerate satellite configurations and weaker inhomogeneities, i.e. larger shear lengths.

The shear length introduced above is the relevant length scale if $\nabla \times \mathbf{V}$ is calculated. A more general inhomogeneity length scale can be defined with the help of the norm of the jacobian matrix: $L_{\nabla \mathbf{V}} = V/|\nabla \mathbf{V}|$.

Relation to Previous Work

Since magnetic field measurements provide an accuracy which is comparable to or even better than the spacecraft position determination accuracy, the error formulas of this chapter cannot be compared directly with the results of the statistical studies concerning current density determination (see Section 17.1.2 for a more detailed discussion). But we can look at the work of *Martz, Sauvaud, and Rème* [1993] (*MSR93*), where the simplified spacecraft configuration sketched in Figure 17.1 was used in the case

$\Delta x_1 = \Delta x_2 = \Delta x_3 \equiv D$ to estimate the diamagnetic current, i.e. the electric current stemming from the pressure gradient term in the MHD momentum equation. In Section 6.3 of *MSR93* the influence of the ratio $D/L_{\nabla p}$ (denoted $\nabla p/p$ by *MSR93*) and the relative experiment inaccuracy $\delta A/A$ (called intercalibration error by *MSR93*) is investigated. The last quantity corresponds to the relative error in pressure estimation $\delta p/|p|$. Possible errors in spacecraft position determination are ignored. Thus, we rearrange equation 17.45 and get:

$$\frac{\delta[\partial p/\partial x_i]}{|\partial p/\partial x_i|} \simeq \frac{\sqrt{2} L_{\partial_i p}}{\Delta x_i} \frac{\delta p}{|p|} \simeq \left(\frac{\delta A/A}{\nabla p/p} \right)_{MSR93}$$

This formula fits the variation of the relative error with $\delta A/A$ and $\nabla p/p$, which was computed on the basis of a statistical study and is sketched in Figure 15 of *MSR93*.

A Road Map to Plasma Moment Derivative Estimation

A procedure to estimate plasma moment derivatives and their errors should at least consider the following three steps:

1. Take the plasma moments (density ρ , bulk velocity \mathbf{V} , pressure p) and compute the derivatives of interest with the help of the reciprocal base of the cluster tetrahedron $\{\mathbf{k}_\alpha\}$ (Eqs. 17.27-17.31).
2. Compute the cluster geometric factor and the square error of the interesting derivative with the help of the errors formulas 17.32-17.40. If the resulting numbers are too large, the spacecraft cluster is not sensitive enough to determine such derivatives, at least from a set of four-point measurements at a single time. Averaging may improve the situation.
3. Check the influence of nonlinear variations: Compute the square deviation of the interesting derivative from its mean value and check if it is small enough (Eqs. 17.42-17.44). If not, a discontinuity analysis technique may be appropriate, i.e. a special nonlinear model has to be assumed and a set of parameters has to be fitted to the data.

Summary

In order to illustrate some basic aspects of plasma moment derivative accuracy, the first part of this report deals with a simplified cluster geometry. The main contribution to the derivative inaccuracies stems from measurement errors. The lack of knowledge of the exact spacecraft positions should be of minor importance.

With regard to curl and divergence estimation accuracy, large deviations from a regular tetrahedron, i.e. degenerate spacecraft configurations, cause in general large errors. Omnidirectional sensitivity is required in particular if both magnitude and direction of the vorticity vector or the divergence of the velocity field are to be determined. There are only very few exceptions where degenerate configurations, i.e. flat or even long tetrahedra, are of any use at all.

Arbitrary spacecraft configurations can be treated with the help of barycentric coordinates. The reciprocal base of the tetrahedron can be used to define a cluster geometric

factor. We expect it to be key quantity with regard to the determination of all plasma moment derivatives.

In the last section some applications of the results and the error formulas to the Earth's magnetosphere are discussed. Time averaging can improve the accuracy significantly. Furthermore, it can be used to compute a measure of the importance of nonlinear variations. The important steps in the process of plasma moment derivative estimation are summarised.

Appendix

17.A Details of Error Estimation: Simplified Geometry

17.A.1 Error Estimates for Partial Derivatives

On the basis of the model satellite configuration sketched in Figure 17.1 the partial derivative $\partial g / \partial x_i$ can be approximated in the following way:

$$\frac{\partial g}{\partial x_i} \simeq \frac{\Delta_{[i0]} g}{\Delta_{[i0]} x_i} \equiv \frac{g_{[i]} - g_{[0]}}{x_{i[i]} - x_{i[0]}}$$

where $\mathbf{r}_{[\alpha]} \equiv (x_1, x_2, x_3)_{[\alpha]} \equiv (x_{1[\alpha]}, x_{2[\alpha]}, x_{3[\alpha]})$ denotes the measured value of the position of satellite $[\alpha]$. g can be a (scalar) function or a single component of a vector field. Here $g_{[\alpha]}$ denotes the measured value of g at satellite $[\alpha]$.

Sources of Error

Two sources of error are considered:

1. *g-measurement errors*: $\delta g_{[\alpha]}$. They stem e.g. from imperfect calibration of the instruments which measure the quantity g .
2. *Spacecraft position errors*: $\delta \mathbf{r}_{[\alpha]}$. The true spacecraft positions are not the measured values $\mathbf{r}_{[\alpha]}$, but $\mathbf{r}_{[\alpha]} + \delta \mathbf{r}_{[\alpha]}$. The measurements of g take place at the latter points in space.

We name them *primary* errors. The measured values of g can be written in the following way:

$$g_{[\alpha]} = g(\mathbf{r}_{[\alpha]} + \delta \mathbf{r}_{[\alpha]}) + \delta g_{[\alpha]} \simeq g(\mathbf{r}_{[\alpha]}) + (\nabla g) \cdot \delta \mathbf{r}_{[\alpha]} + \delta g_{[\alpha]}$$

This applies if nonlinear variations of g are neglected (they are discussed in a separate section). Since a linear relationship $g(\mathbf{r})$ implies homogeneous partial derivatives, it is not necessary to specify the location where ∇g is taken.

With the help of the last expression the formula for the $\partial g / \partial x_i$ estimator given above can be easily rearranged:

$$\frac{g_{[i]} - g_{[0]}}{x_{i[i]} - x_{i[0]}} = \frac{g(\mathbf{r}_{[i]}) - g(\mathbf{r}_{[0]})}{x_{i[i]} - x_{i[0]}} + \frac{(\nabla g) \cdot (\delta \mathbf{r}_{[i]} - \delta \mathbf{r}_{[0]}) + (\delta g_{[i]} - \delta g_{[0]})}{x_{i[i]} - x_{i[0]}}$$

The first term on the right-hand side equals $\partial g / \partial x_i$ because g is assumed to vary linear with x_i . Thus, the second term is the error of the estimator:

$$\begin{aligned} \delta \left[\frac{\partial g}{\partial x_i} \right] &= \frac{1}{\Delta x_i} [(\delta g_{[i]} - \delta g_{[0]}) + (\nabla g) \cdot (\delta \mathbf{r}_{[i]} - \delta \mathbf{r}_{[0]})] \\ &= \frac{1}{\Delta x_i} \left[(\delta g_{[i]} - \delta g_{[0]}) + \sum_{j=1}^3 \frac{\partial g}{\partial x_j} (\delta x_{j[i]} - \delta x_{j[0]}) \right] \end{aligned} \quad (17.46)$$

where $\Delta x_i \equiv \Delta_{[i0]} x_i \equiv x_{i[i]} - x_{i[0]}$.

Mean Square Errors and Correlations of Primary Errors

In order to compute the mean square error of the partial derivative estimator we square $\delta[\partial g / \partial x_i]$ and perform an ensemble average. We obtain:

$$\begin{aligned} \overline{\left(\delta \left[\frac{\partial g}{\partial x_i} \right] \right)^2} &= \\ &= \frac{1}{(\Delta x_i)^2} \left[\overline{(\delta g_{[i]} - \delta g_{[0]})^2} + \sum_{j=1}^3 \overline{\left(\frac{\partial g}{\partial x_j} (\delta x_{j[i]} - \delta x_{j[0]}) \right)^2} \right] \end{aligned} \quad (17.47)$$

It has been assumed that the g -measurement errors are statistically independent from the errors of the spacecraft position determination.

The last formula simplifies when correlations of *different components* of the spacecraft position errors can be neglected:

$$\begin{aligned} \overline{\left(\delta \left[\frac{\partial g}{\partial x_i} \right] \right)^2} &= \\ &= \frac{1}{(\Delta x_i)^2} \left[\overline{(\delta g_{[i]} - \delta g_{[0]})^2} + \sum_{j=1}^3 \left(\frac{\partial g}{\partial x_j} \right)^2 \overline{(\delta x_{j[i]} - \delta x_{j[0]})^2} \right] \end{aligned} \quad (17.48)$$

$(\delta x_{j[i]} - \delta x_{j[0]})$ denotes the j -th component of the *relative* position error. It depends e.g. on the geocentric distance and on the orientation of the spacecraft cluster.

Further simplifications can be achieved if measurement errors (of the satellite position and the quantity g) at *different spacecraft* are assumed to be uncorrelated:

$$\begin{aligned} \overline{\left(\delta \left[\frac{\partial g}{\partial x_i} \right] \right)^2} &= \\ &= \frac{1}{(\Delta x_i)^2} \left[\overline{(\delta g_{[i]})^2} + \overline{(\delta g_{[0]})^2} + \sum_{j=1}^3 \left(\frac{\partial g}{\partial x_j} \right)^2 \left(\overline{(\delta x_{j[i]})^2} + \overline{(\delta x_{j[0]})^2} \right) \right] \end{aligned} \quad (17.49)$$

But: since every inter-spacecraft calibration process introduces correlations of the above neglected type, the last mentioned formula and the associated assumption require further comments (see below).

Convenient Error Formulas

The purpose of this report is to find some simplified error formulas for derivatives such as gradient, divergence and curl. The expressions given above are still too complicated to produce simple formulas, although all correlations of the primary errors $\delta g_{[i]}$, $\delta g_{[0]}$, $\delta x_{j[i]}$ and $\delta x_{j[0]}$ have already been disregarded. Therefore, we have to make further assumptions:

- The mean square errors of the different spacecraft positions are assumed to be equal and independent of orientation, i.e.

$$\overline{(\delta x_{j[\alpha]})^2} = (\delta_a x)^2$$

for all $j, [\alpha]$.

- The mean square g -measurement errors of the different instruments are assumed to have the same size:

$$\overline{(\delta g_{[\alpha]})^2} = (\delta_a g)^2$$

for all $[\alpha]$.

The index a in $(\delta_a x)^2$ or $(\delta_a g)^2$ indicates that these quantities are *absolute* calibration errors rather than *relative* ones which are discussed below. We finally obtain:

$$\overline{\left(\delta \left[\frac{\partial g}{\partial x_i} \right] \right)^2} = \frac{2}{(\Delta x_i)^2} \left((\delta_a g)^2 + |\nabla g|^2 (\delta_a x)^2 \right) \quad (17.50)$$

By the same procedure one can derive a formula for the mean error of the product of different partial derivative errors ($i \neq j$):

$$\overline{\left(\delta \left[\frac{\partial g}{\partial x_i} \right] \right) \left(\delta \left[\frac{\partial g}{\partial x_j} \right] \right)} = \frac{1}{\Delta x_i \Delta x_j} \left((\delta_a g)^2 + |\nabla g|^2 (\delta_a x)^2 \right) \quad (17.51)$$

If h is different from g (two functions or two components of the same vector field) and the errors δg and δh are uncorrelated, we obtain ($i \neq j$):

$$\overline{\left(\delta \left[\frac{\partial g}{\partial x_i} \right] \right) \left(\delta \left[\frac{\partial h}{\partial x_j} \right] \right)} = \frac{1}{\Delta x_i \Delta x_j} \left([(\nabla g) \cdot (\nabla h)] (\delta_a x)^2 \right) \quad (17.52)$$

$$\overline{\left(\delta \left[\frac{\partial g}{\partial x_i} \right] \right) \left(\delta \left[\frac{\partial h}{\partial x_i} \right] \right)} = \frac{2}{(\Delta x_i)^2} \left([(\nabla g) \cdot (\nabla h)] (\delta_a x)^2 \right) \quad (17.53)$$

Such *cross terms* appear in the error formulas of divergence and curl estimates. The last two formulas can be used if correlations of different components of the vector field under investigation can be neglected.

Inter-Spacecraft Calibration

The aim of inter-spacecraft calibration processes should be the reduction of *relative errors*, e.g. $\overline{(\delta g_{[i]} - \delta g_{[0]})^2}$. This can be achieved by the introduction of a positive correlation between $\delta g_{[i]}$ and $\delta g_{[0]}$, i.e. between errors at different spacecraft:

$$\overline{(\delta g_{[i]} - \delta g_{[0]})^2} = \overline{(\delta g_{[i]})^2} + \overline{(\delta g_{[0]})^2} - 2\overline{(\delta g_{[i]})(\delta g_{[0]})}$$

As before we assume equal absolute errors $(\delta_a g)^2 = \overline{(\delta g_{[\alpha]})^2}$ for different instruments $[\alpha]$. To investigate the basic consequences of the correlation of errors, we use a correlation coefficient $C_{[\alpha][\beta]}^g \equiv C^g (\in [0, 1])$ which is the same for all possible binary combinations of the various instruments, i.e.

$$\overline{(\delta g_{[\alpha]})(\delta g_{[\beta]})} = C_{[\alpha][\beta]}^g (\delta_a g)^2 = C^g (\delta_a g)^2$$

for all $[\alpha] \neq [\beta]$. With the help of this assumption one can compute the mean square relative errors:

$$\overline{(\delta g_{[i]} - \delta g_{[0]})^2} = 2(1 - C^g)(\delta_a g)^2$$

The correlation of relative errors can be written in the following form ($[j] \neq [i]$):

$$\overline{(\delta g_{[i]} - \delta g_{[0]})(\delta g_{[j]} - \delta g_{[0]})} = (1 - C^g)(\delta_a g)^2 = \frac{1}{2} \overline{(\delta g_{[i]} - \delta g_{[0]})^2}$$

Similar formulas can be written down for the relative errors of the spacecraft positions. In addition to the absolute errors $(\delta_a g)^2$ and $(\delta_a x)^2$ we define mean square relative errors by:

$$(\delta_r g)^2 = (1 - C^g)(\delta_a g)^2 = \frac{1}{2} \overline{(\delta g_{[i]} - \delta g_{[0]})^2} \quad (17.54)$$

$$(\delta_r x)^2 = (1 - C^x)(\delta_a x)^2 = \frac{1}{2} \overline{(\delta x_{j[i]} - \delta x_{j[0]})^2} \quad (17.55)$$

It is easy to show that the structure of the equations derived above doesn't change if $C^{g,x} \neq 0$, but the errors must be replaced: $(\delta_a \{g, x\})^2 \rightarrow (\delta_r \{g, x\})^2$. Thus, we finally obtain ($i \neq j$):

$$\overline{\left(\delta \left[\frac{\partial g}{\partial x_i} \right] \right)^2} = \frac{2}{(\Delta x_i)^2} \left((\delta_r g)^2 + |\nabla g|^2 (\delta_r x)^2 \right) \quad (17.56)$$

$$\overline{\left(\delta \left[\frac{\partial g}{\partial x_i} \right] \right) \left(\delta \left[\frac{\partial g}{\partial x_j} \right] \right)} = \frac{1}{\Delta x_i \Delta x_j} \left((\delta_r g)^2 + |\nabla g|^2 (\delta_r x)^2 \right) \quad (17.57)$$

$$\overline{\left(\delta \left[\frac{\partial g}{\partial x_i} \right] \right) \left(\delta \left[\frac{\partial h}{\partial x_j} \right] \right)} = \frac{1}{\Delta x_i \Delta x_j} \left([(\nabla g) \cdot (\nabla h)] (\delta_r x)^2 \right) \quad (17.58)$$

$$\overline{\left(\delta \left[\frac{\partial g}{\partial x_i} \right] \right) \left(\delta \left[\frac{\partial h}{\partial x_i} \right] \right)} = \frac{2}{(\Delta x_i)^2} \left([(\nabla g) \cdot (\nabla h)] (\delta_r x)^2 \right) \quad (17.59)$$

We summarise the assumptions made during the calculation of these expressions: Linear varying functions or vector field components have been considered in a simplified satellite configuration. Correlations between different components of the spacecraft position errors $\delta \mathbf{r}_{[\alpha]}$ and the measurement errors $\delta g_{[\alpha]}$ have been neglected. Furthermore, we have assumed that measurements taken at different satellites are equally accurate. The correlations introduced by inter-spacecraft calibrations have been modelled by a fixed correlation coefficient.

17.A.2 Curl Estimation Accuracy: The General Case

The relation $\Omega^2 = \sum_{i=1}^3 \Omega_i^2$ leads to the following expression for the error $\delta\Omega$ of the vorticity magnitude:

$$\Omega \delta\Omega = \sum_{i=1}^3 \Omega_i \delta\Omega_i$$

The errors of the cartesian components of the vorticity are related to the errors of the partial derivatives by:

$$\delta\Omega_i = \delta \left[\frac{\partial V_k}{\partial x_j} \right] - \delta \left[\frac{\partial V_j}{\partial x_k} \right]$$

where (i, j, k) must be a cyclic permutation of $(1, 2, 3)$. In order to compute the mean square error of the vorticity magnitude we have to insert this expression into $\sum_{i=1}^3 \Omega_i \delta\Omega_i$, then square and average the result. We assume (as in the divergence error estimation case) that different components of the bulk velocity have uncorrelated measurement errors and make use of the approximation $|\nabla V_{j,k}|^2 (\delta x)^2 \ll (\delta V)^2$. This leads to:

$$\begin{aligned} \Omega^2 \overline{(\delta\Omega)^2} &= \left(\Omega_1^2 \left[\overline{\left(\delta \left[\frac{\partial V_3}{\partial x_2} \right] \right)^2} + \overline{\left(\delta \left[\frac{\partial V_2}{\partial x_3} \right] \right)^2} \right] + \text{cyclic permutations} \right) \\ &\quad - \left(2\Omega_1\Omega_2 \overline{\left(\delta \left[\frac{\partial V_3}{\partial x_2} \right] \right) \left(\delta \left[\frac{\partial V_3}{\partial x_1} \right] \right)} + \text{cyclic permutations} \right) \\ &= 2 \left\{ \frac{\Omega_2^2 + \Omega_3^2}{(\Delta x_1)^2} + \frac{\Omega_3^2 + \Omega_1^2}{(\Delta x_2)^2} + \frac{\Omega_1^2 + \Omega_2^2}{(\Delta x_3)^2} \right\} (\delta V)^2 \\ &\quad - 2 \left\{ \frac{\Omega_1\Omega_2}{\Delta x_1 \Delta x_2} + \frac{\Omega_2\Omega_3}{\Delta x_2 \Delta x_3} + \frac{\Omega_3\Omega_1}{\Delta x_3 \Delta x_1} \right\} (\delta V)^2 \end{aligned} \quad (17.60)$$

This is still a rather complicated expression for the quantity $\overline{(\delta\Omega)^2}$. It depends on the shape of the satellite configuration, on the magnitude Ω and on the direction of the vorticity with respect to the spacecraft tetrahedron.

The easiest way to isolate the dependence of the magnitude error on the shape of the satellite configuration is to get rid of the vorticity direction dependence in the above formula by performing an average over the sphere $\Omega = \text{constant}$:

$$\langle g \rangle_{\Omega=\text{const}} = \frac{1}{4\pi} \int_0^{2\pi} d\lambda \int_0^\pi d\phi \sin \phi g(\lambda, \phi)$$

(ϕ : polar angle, λ : azimuth). With the help of the relations:

$$\begin{aligned} \frac{\langle \Omega_1^2 \rangle_{\Omega=\text{const}}}{\Omega^2} &= \frac{\langle \Omega_2^2 \rangle_{\Omega=\text{const}}}{\Omega^2} = \frac{\langle \Omega_3^2 \rangle_{\Omega=\text{const}}}{\Omega^2} = \frac{1}{3} \\ \frac{\langle \Omega_1\Omega_2 \rangle_{\Omega=\text{const}}}{\Omega^2} &= \frac{\langle \Omega_2\Omega_3 \rangle_{\Omega=\text{const}}}{\Omega^2} = \frac{\langle \Omega_3\Omega_1 \rangle_{\Omega=\text{const}}}{\Omega^2} = 0 \end{aligned}$$

we obtain:

$$\overline{(\delta\Omega)^2}_{\Omega=\text{const}} = \frac{2}{3} \left\{ 2 \sum_{i=1}^3 \frac{1}{(\Delta x_i)^2} \right\} (\delta V)^2 \quad (17.61)$$

17.A.3 Curl Estimation Accuracy: The Special Case $\Delta x_1 = \Delta x_2$

The Error $\delta\Omega_\perp$

The error $\delta\Omega_\perp$ of the vorticity component in the x_1x_2 plane obeys the relation:

$$\Omega_\perp \delta\Omega_\perp = \Omega_1 \delta\Omega_1 + \Omega_2 \delta\Omega_2$$

which can be deduced from $\Omega_\perp^2 = \Omega_1^2 + \Omega_2^2$. Squaring and averaging yields:

$$\Omega_\perp^2 \overline{(\delta\Omega_\perp)^2} = \Omega_1^2 \overline{(\delta\Omega_1)^2} + \Omega_2^2 \overline{(\delta\Omega_2)^2} + 2\Omega_1\Omega_2 \overline{(\delta\Omega_1)(\delta\Omega_2)}$$

To evaluate these expressions we make use of the same assumptions as in the vorticity magnitude case (see above). We obtain:

$$\begin{aligned} \overline{(\delta\Omega_1)^2} = \overline{(\delta\Omega_2)^2} &= \left\{ \frac{2}{(\Delta x_\perp)^2} + \frac{2}{(\Delta x_\parallel)^2} \right\} (\delta V)^2 \\ \overline{(\delta\Omega_1)(\delta\Omega_2)} &= \frac{-(\delta V)^2}{(\Delta x_\perp)^2} \end{aligned}$$

and finally:

$$\begin{aligned} \overline{(\delta\Omega_\perp)^2} &= \left\{ \frac{2}{(\Delta x_\perp)^2} + \frac{2}{(\Delta x_\parallel)^2} \right\} (\delta V)^2 - \frac{2 \cos \lambda \sin \lambda (\delta V)^2}{(\Delta x_\perp)^2} \\ &= \left\{ \frac{2 - \sin 2\lambda}{(\Delta x_\perp)^2} + \frac{2}{(\Delta x_\parallel)^2} \right\} (\delta V)^2 \end{aligned} \quad (17.62)$$

The Error $\delta\lambda$

The angle λ is given by $\lambda = \tan^{-1}(\Omega_2/\Omega_1)$. Elementary calculus leads to the following relation:

$$\Omega_\perp \delta\lambda = \cos \lambda \delta\Omega_2 - \sin \lambda \delta\Omega_1$$

The mean square error $\overline{(\delta\lambda)^2}$ is derived exactly in the same way as in the $\overline{(\delta\Omega_\perp)^2}$ case. The result is:

$$\begin{aligned} \Omega_\perp^2 \overline{(\delta\lambda)^2} &= \sin^2 \lambda \overline{(\delta\Omega_1)^2} + \cos^2 \lambda \overline{(\delta\Omega_2)^2} - 2 \sin \lambda \cos \lambda \overline{(\delta\Omega_1)(\delta\Omega_2)} \\ &= \left\{ \frac{2}{(\Delta x_\perp)^2} + \frac{2}{(\Delta x_\parallel)^2} \right\} (\delta V)^2 + \frac{2 \cos \lambda \sin \lambda (\delta V)^2}{(\Delta x_\perp)^2} \\ &= \left\{ \frac{2 + \sin 2\lambda}{(\Delta x_\perp)^2} + \frac{2}{(\Delta x_\parallel)^2} \right\} (\delta V)^2 \end{aligned} \quad (17.63)$$

Dependence of $\overline{(\delta\Omega_\perp)^2}$ and $\overline{(\delta\lambda)^2}$ on the Angle λ

Both errors $\overline{(\delta\Omega_\perp)^2}$ and $\overline{(\delta\lambda)^2}$ vary with the angle λ . But if accuracy is gained in the determination of one quantity it is lost in the determination of the other one because the sum of both mean square errors does not depend on the angle λ :

$$\Omega_\perp^2 \overline{(\delta\lambda)^2} + \overline{(\delta\Omega_\perp)^2} = \text{constant}$$

Even the variations of the individual errors with the angle λ are rather moderate compared to their dependence on the separation lengths Δx_{\parallel} and Δx_{\perp} or on the measurement error δV : From the formulas given above we can deduce maximum and minimum values of the mean square errors which cannot deviate more than 50% from the mean value which is given by:

$$\left\langle \Omega_{\perp}^2 (\delta \lambda)^2 \right\rangle_{\lambda} = \left\langle (\delta \Omega_{\perp})^2 \right\rangle_{\lambda} = \left\{ \frac{2}{(\delta x_{\perp})^2} + \frac{2}{(\delta x_{\parallel})^2} \right\} (\delta V)^2 \quad (17.64)$$

Averaging over the angle λ is denoted by $\langle \dots \rangle_{\lambda}$. The maximum uncertainty of $\pm 50\%$ is reached only if $(\Delta x_{\parallel})^2 \gg (\Delta x_{\perp})^2$, for more regular tetrahedra $((\Delta x_{\parallel})^2 \approx (\Delta x_{\perp})^2)$ it should be around $\pm 25\%$ which corresponds to a standard deviation of less than $\pm 15\%$. Thus, in most cases one does not need to take care of the dependence on the angle λ and it should be sufficient to use the formula for the mean values of $\Omega_{\perp}^2 (\delta \lambda)^2$ and $(\delta \Omega_{\perp})^2$.

The Error $\delta \Omega$

Since $\Omega^2 = \Omega_{\perp}^2 + \Omega_{\parallel}^2$ the vorticity magnitude error $\delta \Omega$ obeys the relation:

$$\Omega \delta \Omega = \Omega_{\perp} \delta \Omega_{\perp} + \Omega_{\parallel} \delta \Omega_{\parallel}$$

Squaring and averaging leads to:

$$\Omega^2 \overline{(\delta \Omega)^2} = \Omega_{\perp}^2 \overline{(\delta \Omega_{\perp})^2} + \Omega_{\parallel}^2 \overline{(\delta \Omega_{\parallel})^2} + 2 \Omega_{\perp} \Omega_{\parallel} \overline{(\delta \Omega_{\perp})(\delta \Omega_{\parallel})}$$

The cross term is evaluated in same way as before, the result is:

$$\overline{(\delta \Omega_{\perp})(\delta \Omega_{\parallel})} = -(\cos \lambda + \sin \lambda) \frac{(\delta V)^2}{\Delta x_{\perp} \Delta x_{\parallel}}$$

Since we are mainly interested in the influence of the polar angle $\phi = \tan^{-1}(\Omega_{\perp}/\Omega_{\parallel})$ on the error $\overline{(\delta \Omega)^2}$, we perform an average over the azimuth λ . This kills the cross term. We finally obtain after some rearrangements:

$$\begin{aligned} \frac{\left\langle (\delta \Omega)^2 \right\rangle_{\lambda}}{\Omega^2} &= \frac{\Omega_{\perp}^2}{\Omega^2} \frac{\left\langle (\delta \Omega_{\perp})^2 \right\rangle_{\lambda}}{\Omega^2} + \frac{\Omega_{\parallel}^2}{\Omega^2} \frac{\left\langle (\delta \Omega_{\parallel})^2 \right\rangle_{\lambda}}{\Omega^2} \\ &= \left\{ \frac{2}{(\Omega \Delta x_{\perp})^2} + \frac{2}{(\Omega \Delta x_{\parallel})^2} \right\} \sin^2 \phi (\delta V)^2 + \frac{4 \cos^2 \phi (\delta V)^2}{(\Omega \Delta x_{\perp})} \\ &= \left\{ \frac{2}{\sin^2 \theta} \right\} \frac{(\delta V)^2}{(\Omega \Delta r)^2} + 2 \left\{ \frac{\cos^2 \phi}{\sin^2 \theta} + \frac{\sin^2 \phi}{\cos^2 \theta} \right\} \frac{(\delta V)^2}{(\Omega \Delta r)^2} \end{aligned} \quad (17.65)$$

with $\Delta r = \sqrt{(\Delta x_{\perp})^2 + (\Delta x_{\parallel})^2}$ and $\tan \theta = \Delta x_{\perp}/\Delta x_{\parallel}$.

The Error $\delta \phi$

The relation $\tan \phi = \Omega_{\perp}/\Omega_{\parallel}$ implies:

$$\Omega \delta \phi = \cos \phi \delta \Omega_{\perp} - \sin \phi \delta \Omega_{\parallel}$$

We proceed in the same manner as in the $\delta\Omega$ case to obtain the mean square error of the polar angle ϕ . We finally get:

$$\begin{aligned} \overline{(\delta\phi)^2} &= \left\{ \frac{2}{(\Omega\Delta x_\perp)^2} + \frac{2}{(\Omega\Delta x_\parallel)^2} \right\} \cos^2 \phi (\delta V)^2 + \frac{4 \sin^2 \phi (\delta V)^2}{(\Omega\Delta x_\perp)} \\ &= \left\{ \frac{2}{\sin^2 \theta} \right\} \frac{(\delta V)^2}{(\Omega\Delta r)^2} + 2 \left\{ \frac{\sin^2 \phi}{\sin^2 \theta} + \frac{\cos^2 \phi}{\cos^2 \theta} \right\} \frac{(\delta V)^2}{(\Omega\Delta r)^2} \end{aligned} \quad (17.66)$$

17.B Details of Error Estimation: General Geometry

In the following we estimate the derivative errors for arbitrary satellite configurations with the help of barycentric coordinates. Contributions of spacecraft position uncertainties are neglected, i.e. the approximation $|\nabla g|^2 (\delta x)^2 \ll (\delta g)^2$ is used. With regard to vector fields \mathbf{V} , we treat different components of $\delta \mathbf{V}$ as uncorrelated quantities. Furthermore, we assume uncorrelated measurement errors at different spacecraft.

For convenience the summation limits are omitted. We keep in mind that Greek indices (α, β, \dots) assume values 0,1,2,3 whereas latin indices (i, j, \dots) assume values 1,2,3.

17.B.1 Directional Derivatives and Partial Derivatives

The derivative of a quantity g in the direction of vector $\hat{\mathbf{e}}$ is estimated by:

$$\hat{\mathbf{e}} \cdot \nabla g \simeq \sum_{\alpha} (\hat{\mathbf{e}} \cdot \mathbf{k}_{\alpha}) g_{\alpha}$$

Thus, the error of the directional derivative can be written as:

$$\delta[\hat{\mathbf{e}} \cdot \nabla g] = \sum_{\alpha} (\hat{\mathbf{e}} \cdot \mathbf{k}_{\alpha}) \delta g_{\alpha}$$

Squaring and averaging yields:

$$\overline{(\delta[\hat{\mathbf{e}} \cdot \nabla g])^2} = \sum_{\alpha, \beta} (\hat{\mathbf{e}} \cdot \mathbf{k}_{\alpha}) (\hat{\mathbf{e}} \cdot \mathbf{k}_{\beta}) \overline{(\delta g_{\alpha})(\delta g_{\beta})}$$

The assumptions mentioned above imply:

$$\overline{(\delta g_{\alpha})(\delta g_{\beta})} = \delta_{\alpha\beta} (\delta g)^2$$

where $\delta_{\alpha\beta}$ denotes the Kronecker symbol ($= 1$ if $\alpha = \beta$, otherwise $= 0$). We finally get:

$$\overline{(\delta[\hat{\mathbf{e}} \cdot \nabla g])^2} = \left\{ \sum_{\alpha} (\hat{\mathbf{e}} \cdot \mathbf{k}_{\alpha})^2 \right\} (\delta g)^2 \quad (17.67)$$

The mean square error of a partial derivative is found by inserting $\hat{\mathbf{e}} = \hat{\mathbf{e}}_i$ (unit vector in direction x_i):

$$\left(\delta \left[\frac{\partial g}{\partial x_i} \right] \right)^2 = \left\{ \sum_{\alpha} k_{\alpha i}^2 \right\} (\delta g)^2 \quad (17.68)$$

The result can be transferred to the elements of the jacobian matrix of a vector field V :

$$\left(\overline{\delta \left[\frac{\partial V_j}{\partial x_i} \right]} \right)^2 = \left\{ \sum_{\alpha} k_{\alpha i}^2 \right\} (\delta V)^2 \quad (17.69)$$

A mean value for the directional derivative error can be obtained by averaging over the set of unit vectors \hat{e} , i.e. over all possible directions:

$$\left\langle \overline{(\delta [\hat{e} \cdot \nabla g])^2} \right\rangle_{\hat{e}} = \frac{1}{3} \left\{ \sum_{\alpha} |\mathbf{k}_{\alpha}|^2 \right\} (\delta g)^2 \quad (17.70)$$

17.B.2 Divergence Estimation Accuracy

From the estimator for the divergence of a vector field V

$$\nabla \cdot V \simeq \sum_{\alpha} \mathbf{k}_{\alpha} \cdot V_{\alpha}$$

we can deduce the error:

$$\delta[\nabla \cdot V] \simeq \sum_{\alpha} \mathbf{k}_{\alpha} \cdot \delta V_{\alpha} = \sum_{\alpha, i} k_{\alpha i} \delta V_{\alpha i}$$

Squaring and averaging leads to:

$$\overline{(\delta[\nabla \cdot V])^2} = \sum_{\alpha, \beta, i, j} k_{\alpha i} k_{\beta j} \overline{(\delta V_{\alpha i})(\delta V_{\beta j})}$$

We make use of the assumptions concerning the uncorrelated measurement errors:

$$\overline{(\delta V_{\alpha i})(\delta V_{\beta j})} = \delta_{\alpha\beta} \delta_{ij} (\delta V)^2$$

and finally obtain:

$$\overline{(\delta[\nabla \cdot V])^2} = \left\{ \sum_{\alpha, i} k_{\alpha i}^2 \right\} (\delta V)^2 = \left\{ \sum_{\alpha} |\mathbf{k}_{\alpha}|^2 \right\} (\delta V)^2 \quad (17.71)$$

17.B.3 Curl Estimation Accuracy

The estimator for the i -th cartesian component of the vorticity can be written in the following form:

$$\begin{aligned} \Omega_i &= (\nabla \times V) \cdot \hat{e}_i \simeq \sum_{\alpha} (\mathbf{k}_{\alpha} \times V_{\alpha}) \cdot \hat{e}_i = \sum_{\alpha} (\hat{e}_i \times \mathbf{k}_{\alpha}) \cdot V_{\alpha} \\ &= \sum_{\alpha, j} (\hat{e}_i \times \mathbf{k}_{\alpha})_j V_{\alpha j} \end{aligned}$$

We proceed in the same way as above to calculate the mean square errors and get:

$$\overline{(\delta \Omega_i)^2} = \left\{ \sum_{\alpha, j} (\hat{e}_i \times \mathbf{k}_{\alpha})_j^2 \right\} (\delta V)^2 \quad (17.72)$$

The vorticity magnitude error is related to the component errors by:

$$\Omega \delta \Omega = \sum_i \Omega_i \delta \Omega_i = \sum_{i,\alpha,j} \Omega_i (\hat{\mathbf{e}}_i \times \mathbf{k}_\alpha)_j \delta V_{\alpha j} = \sum_{i,\alpha,j,l,m} \Omega_i \epsilon_{jlm} (\hat{\mathbf{e}}_i)_l k_{\alpha m} \delta V_{\alpha j}$$

The symbol ϵ_{jlm} denotes the total anti-symmetric tensor ($= 1$, if (j, l, m) cyclic permutation of $(1, 2, 3)$; $= -1$, if anti-cyclic; otherwise zero). $(\hat{\mathbf{e}}_i)_l$ denotes the l -th component of unit vector $\hat{\mathbf{e}}_i$, thus $(\hat{\mathbf{e}}_i)_l = \delta_{il}$ and

$$\Omega \delta \Omega = \sum_{i,\alpha,j,m} \Omega_i \epsilon_{jim} k_{\alpha m} \delta V_{\alpha j}$$

Squaring and averaging yields:

$$\begin{aligned} \Omega^2 \overline{(\delta \Omega)^2} &= \sum_{i,\alpha,j,m} \sum_{n,\beta,p,q} \Omega_i \Omega_n \epsilon_{jim} \epsilon_{pnq} k_{\alpha m} k_{\beta q} \underbrace{\overline{\delta V_{\alpha j} \delta V_{\beta p}}}_{=\delta_{\alpha\beta} \delta_{jp} (\delta V)^2} \\ &= (\delta V)^2 \sum_{i,\alpha,m,n,q} \Omega_i \Omega_n k_{\alpha m} k_{\alpha q} \underbrace{\sum_j \epsilon_{jim} \epsilon_{jnq}}_{=\delta_{in} \delta_{mq} - \delta_{iq} \delta_{mn}} \end{aligned}$$

We finally obtain:

$$\begin{aligned} \Omega^2 \overline{(\delta \Omega)^2} &= (\delta V)^2 \sum_{i,\alpha,m} \Omega_i \Omega_i k_{\alpha m} k_{\alpha m} - (\delta V)^2 \sum_{i,\alpha,m} \Omega_i \Omega_m k_{\alpha m} k_{\alpha i} \\ &= \left\{ \Omega^2 \sum_{\alpha} |\mathbf{k}_\alpha|^2 - \sum_{i,\alpha,m} \Omega_i \Omega_m k_{\alpha m} k_{\alpha i} \right\} (\delta V)^2 \end{aligned} \quad (17.73)$$

The last expression simplifies when averaged over the sphere $\Omega = \text{constant}$, i.e. over all possible vorticity directions. The result is:

$$\langle \overline{(\delta \Omega)^2} \rangle_{\Omega=\text{const}} = \frac{2}{3} \left\{ \sum_{\alpha} |\mathbf{k}_\alpha|^2 \right\} (\delta V)^2 \quad (17.74)$$

Acknowledgements

We thank Steven J. Schwartz for helpful comments and valuable suggestions.

Bibliography

Martz, C., Sauvaud, J. A., and Rème, H., Accuracy of ion distribution measurements and related parameters using the Cluster CIS experiment, in *Proc. International Conf. "Spatio-Temporal Analysis for Resolving plasma Turbulence (START)"*, Aussois, 31 Jan.–5 Feb. 1993, ESA WPP-047, pp. 229–239, European Space Agency, Paris, France, 1993.

— 18 —

Numerical Modelling and Simulation for Multi-Spacecraft Data Analysis: Approaches and Examples

JÖRG BÜCHNER, JENS-PEER KUSKA, HEINZ WIECHEN

*Max-Planck-Institut für Aeronomie
Katlenburg-Lindau, Germany*

18.1 Introduction

The primary advantage of multi-spacecraft *in situ* measurements is the principal opportunity to discriminate spatial structures from temporal evolutions. These two aspects of space processes are undistinguishably intermixed in single satellite measurements. Further, multi-spacecraft measurements allow the determination of divergences, gradients, curls, currents, wave vectors and other non-scalar parameters—in principle. The interpretation of multi-spacecraft measurements often depends, however, on the underlying models.

Numerical modelling and simulation is useful for multi-spacecraft investigations first to develop and test specific analysis methods. Numerical simulations are able to provide test data before the satellites are launched.

The second purpose of numerical modelling and simulations for multi-spacecraft investigations is to support the interpretation. The systematic errors of single satellite observations were usually large enough to suffice with more or less qualitative models for data interpretation. The potentially higher information content of future multi-spacecraft investigations, however, can be deduced only by comparison of the observations with appropriate numerical modelling and simulation results.

The third purpose of numerical modelling and simulations for multi-spacecraft data analysis results from the very aim of a multi-spacecraft mission: to better understand the essentials of the plasma dynamics in space and time from just a few measurable parameters of plasma, particles and fields obtained just along the satellite orbits. In this, multi-spacecraft missions have still to cope with the principal limits of *in situ* satellite measurements: they have to solve inverse problems with only an insufficient amount of information available, although they are more discriminative in principle. The limited number of measurements and the insufficient spatial coverage can be dealt with only by means of models.

Many different numerical modelling and simulation approaches to space plasma problems exist. Before choosing a concrete approach one has to realise that no code can answer all questions. Despite the amazing growth of computer resources, numerical simulations

never will duplicate nature. Their purpose is, instead, to verify hypotheses about the essential links and laws in nature. The way they do this is to predict consequences of model assumptions which then can be compared with observations. The choice of the most appropriate code is governed, therefore, by the specifics of the problem one has in mind.

How can one decide what numerical model and simulation approach is most appropriate for a given multi-spacecraft measurement? The criterion is to simplify the space plasma phenomena as much as possible, while still describing the most essential interactions. Indeed, one has to take into account that most space plasmas are collisionless. As a result interactions by binary collisions are negligible while weak interactions via collective electromagnetic forces prevail. Fortunately, depending on the problem at hand, parts of the interactions decouple. Since the derivation of most appropriate numerical models is still an on-going research topic, a coverage of all existing approaches and problems in a handbook chapter is impossible. Hints to the extended literature are given in the short but incomplete bibliography at the end of this chapter.

The aim of this chapter is to demonstrate by means of typical examples the utility and limitations of the three main numerical modelling and simulation approaches for multi-spacecraft measurements. In Section 18.2 we illustrate the possibilities of test particle calculations, in Section 18.3 of a magnetohydrodynamic modelling of fields and macroscopic plasma parameters, neglecting particle interactions and kinetic effects while in Section 18.4 the use of self-consistent kinetic plasma simulations is treated.

Test particle calculations provide an appropriate treatment mainly of acceleration and propagation of energetic particles (Section 18.2). They help to utilise non-adiabatic acceleration and propagation effects for the remote sensing of particle energisation processes (Section 18.2.1) as well as the mass-dependent spectra formation (Section 18.2.2).

Conversely, if one is interested in the macroscopic dynamics of fields and plasmas rather than in particle effects, it is better to take a fluid approach. In Section 18.3 we demonstrate the applicability of fluid approaches by several examples of applying single-fluid resistive magnetohydrodynamic (MHD) simulations in order to discriminate physical hypotheses by multi-satellite plasma and field observations. We give examples of how even the simplest MHD approach still can help to identify new wave modes (Section 18.3.1), to discriminate magnetopause entry mechanisms (Section 18.3.2) or mechanisms of the magnetotail energy release (Section 18.3.3).

On the other hand, it has become obvious over the past decades of exploring space plasmas that the most important processes (transport through boundaries, the sudden releases of energy, the acceleration of particles to high energies, collisionless shock waves and reconnection) are essentially controlled by the interactions of particles and their collective fields. Kinetic processes, however, cannot be described by fluid models. However, only the simplest kinetic plasma models, like one-dimensional electromagnetic or two-dimensional electrostatic models, are analytically solvable. More realistic models have necessarily to be treated numerically. In Section 18.4 we demonstrate utilities and limitations of currently available kinetic simulation approaches for multi-spacecraft observations. Although modern computer technique makes the numerical simulation of more and more complicated models possible, they easily reach any technical limit. One way to cope with this is to simplify the approach by neglecting some less important interactions. In Sections 18.4.1, 18.4.2, and 18.4.3 we give examples of such restricted kinetic approaches. We focus especially on those, which were developed in the framework of the European Simulation Network preparing the first attempt of a Cluster mission [Burgess,

1994; *Chanteur and Roux, 1997*]. A fully adequate kinetic description, however, must be self-consistent, considering all particle species, their full electromagnetic field effect and the complete feedback of the fields to the particles. Only at this level can simulations reveal all cross-scale dynamic space plasma phenomena. Fully three-dimensional self-consistent kinetic simulations have now been started. In Section 18.4.4 we demonstrate their opportunities for multi-spacecraft observations.

18.2 Test Particle Calculations

Multi-spacecraft missions provide for the first time simultaneously measured spatially distributed observations of the three-dimensional velocity space spectra of particles covering a broad energy range, including mass separation. These measurements contain a vast amount of information about the mechanisms of particle energisation in space. Although, strictly speaking, only self-consistent kinetic simulations can describe the underlying plasma processes adequately (see Section 18.4.4), important features of the distribution can be modelled in the much simpler test particle approach, at least for energetic particles. The test particle approach neglects all particle interactions, considering only the action of external forces. The trajectories of particles are, therefore, calculated in prescribed electric and magnetic fields neglecting any feedback reactions of particle flows. This usually restricts the applicability of the test particle approach to energetic particles, their energy and angular spectra, their propagation and spatial distribution and calls for best possible field models.

Mathematically the test particle approach leads to a system of nonlinear but ordinary differential equations. In a simply connected phase space, the Liouville theorem allows the reconstruction of phase space distributions by mapping along test particle orbits. If, in addition, perturbation methods are applicable, as in slowly and smoothly varying fields, test particle orbits can easily be found using adiabatic invariants as approximate integrals of motion. Typical applications of this approach are described in Chapter 7. The methods presented there are suitable as long as the phase space structure is regular and simply connected. The phase space is regular if the equations of motion are integrable or almost integrable. The latter applies, for example, if the particles are magnetised and the magnetic moment μ is conserved. In this situation a straightforward application of the Liouville theorem allows predictions of distribution functions using just a few typical particle orbits. Important predictions can be made in this limit, like the D -shape of the distribution function with a lower energy cutoff, the spatial dispersion by velocity filtering of particles, accelerated at different moments of time to different energies for the diagnostics of time-dependent reconnection, for example, at the dayside magnetopause (see Chapter 7).

In reality, however, the phase space can be much more complex. Over the past ten years it became obvious that the phase space of energetic particles in space is not always regular. Deviations from the regularity of motion may cause the formation of structures in configurational and velocity space. In Section 18.2.1 we demonstrate results of test particle calculations leading beyond the limits of adiabaticity and a straightforward applicability of the Liouville theorem. We show how the consequent structuring of energetic particle flows in space and velocity space, resulting from non-adiabatic effects, can be used for remote sensing in multi-satellite studies. In Section 18.2.2 we then demonstrate the use of the test particle approach in interpreting multi-spacecraft mass-spectrometer results.

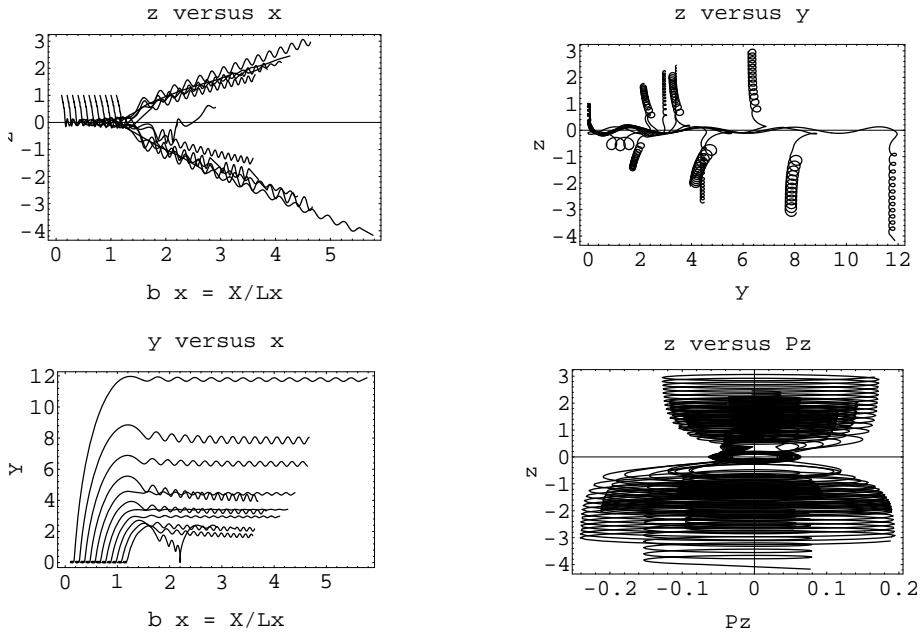


Figure 18.1: H^+ orbits illustrating the non-adiabatic scattering and acceleration near a reconnection X-line.

18.2.1 Non-adiabatic Particle Motion

The regularity of the phase space and adiabaticity of charged particle orbits breaks down in inhomogeneous as well as in rapidly changing fields. Both effects can cause a violation of the conservation of the magnetic moment $\mu = m v_{\perp}^2 / |B|$, for example in boundaries with their strongly inhomogeneous fields. As a result, the equations of motion can become non-integrable, the particle motion is non-gyrotropic, non-adiabatic scattering and deterministic chaos can occur. If this happens then trajectories started nearby each other diverge considerably over finite distances. In this case the application of the Liouville theorem must be handled with care. On the other hand the properties of non-adiabatic scattering can be used for multi-spacecraft remote diagnostics. Observable features caused by non-adiabatic motion are non-gyrotropic distribution functions, non-monotonic and multiple-peaked spectra. Let us illustrate the formation of multiple-peaked spectra by acceleration in a reconnection field

$$\mathbf{B} = B_o \tanh(Z) \hat{\mathbf{e}}_x + B_n \tanh(X \cdot B_n / B_o) \hat{\mathbf{e}}_z; \quad \hat{\mathbf{e}} = \mathbf{E} = E_o \hat{\mathbf{e}}_y. \quad (18.1)$$

Figure 18.1 depicts three projections of test H^+ orbits and one orbit projection in mixed space-velocity space (lower right) in a field given by equation 18.1. The parameters are chosen, in accordance to the geomagnetic tail, $B_n / B_o = 0.1$ and $E_o = 0.1 \cdot v_A B_o$, where v_A is the Alfvén velocity in the ambient plasma. The test particle orbits are obtained launching initially cold protons, i.e. with at a thermal velocity $V_{th} \ll E_o / B_n$, at $Z = 1$ from different positions in the X direction.

Figure 18.1 shows how the protons first gyrate and drift toward the equatorial plane $Z = 0$. Near the equator they become accelerated while meandering across the equatorial plane. The upper left panel of the figure depicts how particles, started near each other, can reach quite different locations in space. As one can see they even switch easily between the upper and lower half space! It is this behaviour that makes the phase space irregular. It lets any application of the Liouville theorem sensitively depend on the choice of test particle orbits. On the other hand, non-adiabatic motion and also chaotic scattering are obviously generated by solving deterministic equations. They control the formation of process-specific new structures. The small ellipses in the centre of the lower right panel of Figure 18.1 indicate for example the temporary existence of a quasi-integral of the motion during the acceleration phase. The upper right panel of Figure 18.1, on the other hand, demonstrates that the ejection points of the accelerated ions are bunched. Since the particle energy gain is directly proportional to the shift of the particles in the y -direction (see equations 18.1), the resulting particle energy is modulated accordingly. This effect forms structures in velocity space and energy spectrum.

Reconnection accelerates particles to highest energies along and inside the magnetic field separatrix as shown in the upper left panel of Figure 18.1. There the selective non-adiabatic scattering splits the ion flow into beamlets which continue to diverge in space due to velocity filtering outside the acceleration region. The histogram in Figure 18.2 depicts the spatial distribution of the count rates of energetic particles formed by reconnection in a field model according to equation 18.1 versus the distance from the equatorial plane. Figure 18.3 additionally resolves the velocity spectrum for different distances from the equatorial plane. Altogether this shows that the chaotic scattering of particles accelerated by reconnection does not at all mean a loss of information. Instead, non-adiabatic and chaotic scattering cause the formation of structures in space and velocity space which can be used for remote diagnostics. The dispersion between the beamlets is even enhanced by the velocity filter effect. As a result the whole magnetosphere can act as a mass spectrometer. The modulation of the energy spectrum detected by several satellites gives additional indications of the acceleration process. It allows the conclusion about whether the ions were accelerated by reconnection, which causes a modulation with velocity peaks at $v_N \propto (N + \frac{1}{2})^{2/3} (E_o/B_n)^{2/3}$ with $N = \text{integers}$, for current sheet acceleration $v_N \propto N^\alpha$ with $\alpha = 1 \dots 2$ in dependence on the tail magnetic field variation.

The simultaneous measurement of the energetic particle flow structure, observed from different satellite positions in multi-spacecraft studies, can act as a dispersiometer as demonstrated by Figure 18.4. The figure illustrates the spatial scaling of the dispersion of protons, accelerated $20 R_E$ away from the spacecraft. While the beamlets of energetic protons (50 keV and higher) are well resolved by the four Cluster-type satellites, a fifth spacecraft, like Interball or Geotail located further away, will see the lower energy particles stemming from the same acceleration region.

18.2.2 Mass-dependent Energetic Particle Spectra

There are more utilities to interpret multi-spacecraft observations of energetic ions in the framework of the test particle approach. The RAPID time-of-flight spectrometer onboard the Cluster spacecraft, for example, will resolve, in addition to the energy spectra, the mass composition of energetic particle from dozens of keV up to MeV energies. Let us consider the acceleration of O^+ ions, 16 times heavier than protons, in the recon-

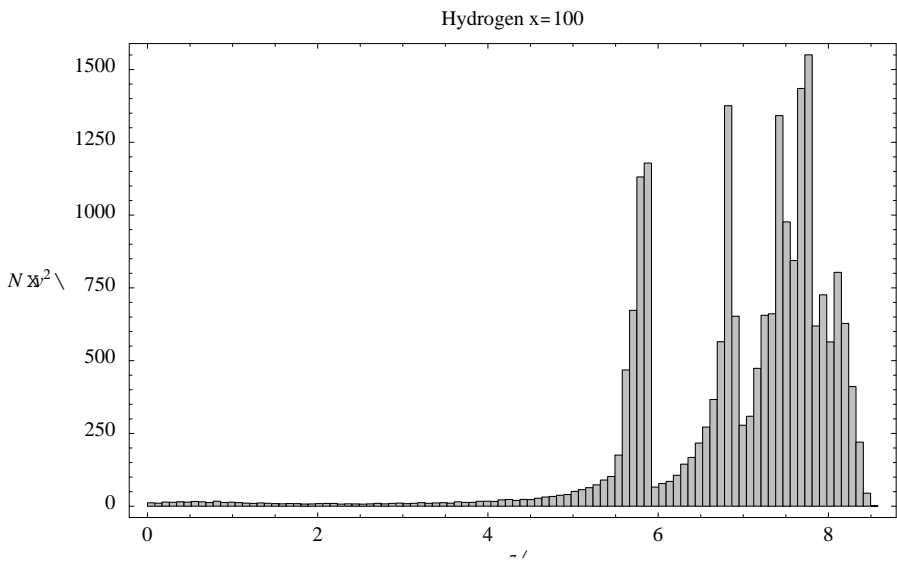


Figure 18.2: Histogram of the particle count rate due to beamlet formation by scattering of reconnection-accelerated protons.

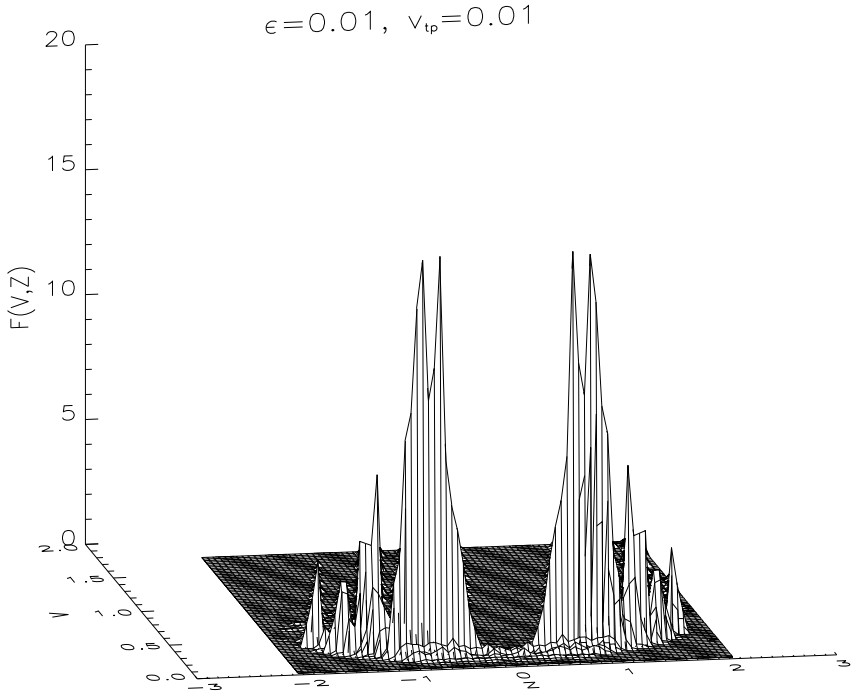


Figure 18.3: Spatio-velocity space distribution function due to non-adiabatic scattering versus the distance z from the equatorial plane.

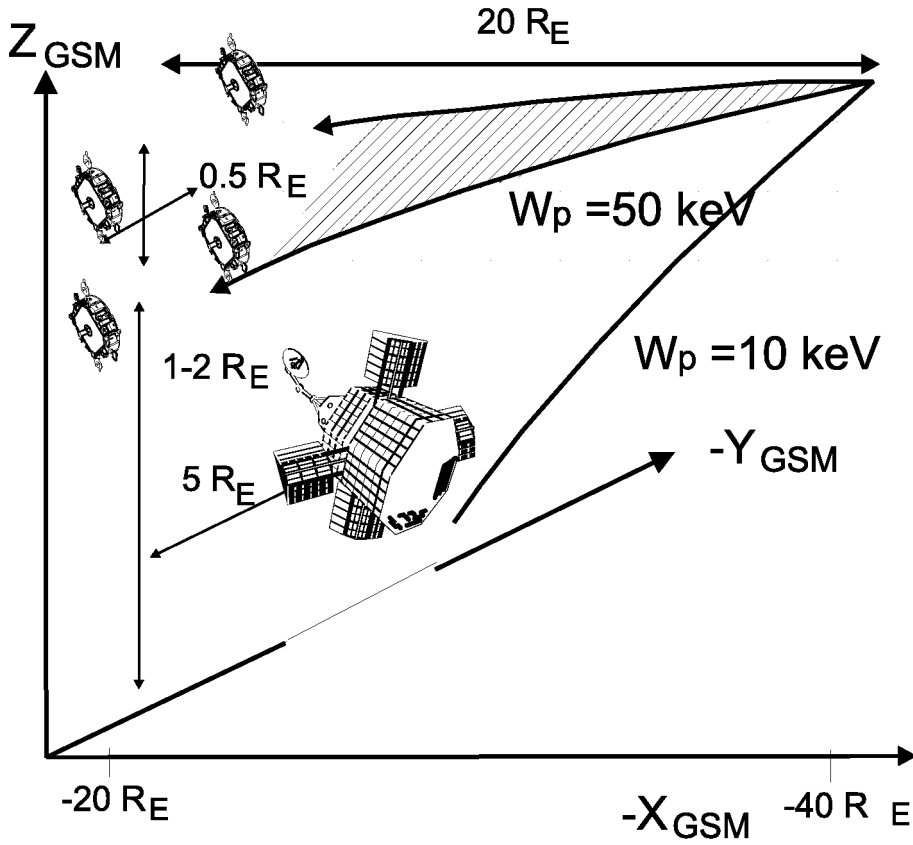


Figure 18.4: Four Cluster satellites and one additional spacecraft forming a dispersiometer using the Earth's magnetic field as a energy spectrometer [from Büchner, 1990].

nection field given by equation 18.1. One finds important similarities but also significant differences compared to the protons. Figure 18.5 illustrates the acceleration of O^+ ions by the same orbit projections used in Figure 18.1. Although the principal features of the non-adiabatic acceleration process of H^+ and O^+ are the same, one sees that the energies gained are much higher for the latter. Due to the higher ion mass, however, the final velocities of the accelerated oxygen ions are even smaller than the proton velocities. A comparison of the velocity spectra (Figure 18.6) shows clearly how the rigidity of the velocity spectra increases with ion mass. A recently developed theory allows the fitting of the spectra. According to Büchner [1995] the spectral function is

$$f(v) \propto \frac{M^{1/2}}{v^{3/2}} \frac{B_n \cdot E_{\text{norm}}}{B_o \cdot E_o} \exp \left\{ -\frac{2B_n \cdot E_{\text{norm}}}{3B_o \cdot E_o} M^{1/2} v^{3/2} \right\} \quad (18.2)$$

where E_{norm} is a normalisation parameter.

The theoretical spectra predicted by equation 18.2 are also shown in Figure 18.6 as solid lines together with the results of test particle calculations, indicated by solid circles,

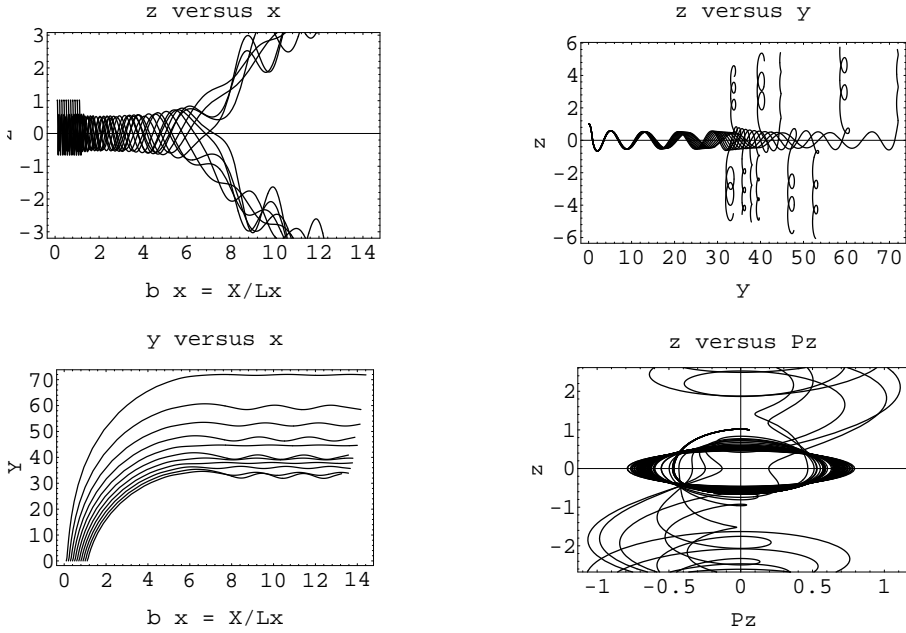


Figure 18.5: O^+ orbits illustrating chaotic scattering near a reconnection X-line.

squares and triangles for H^+ , He^{++} and O^+ , respectively. Due to the characteristic ϵ/b dependence of the spectra this important parameter of the accelerating field can be directly obtained from fitting the observed spectra. Note that the three spectra of H^+ , He^+ and O^+ shown in Figure 18.6, belong to different beamlets and appear at spatially dislocated positions. The observation of different ion species by different Cluster spacecraft will, therefore, provide insights into the structure of the acceleration region just by remote sensing.

So far we have illustrated consequences of the non-adiabatic particle dynamics relevant to multi-spacecraft observations by using analytically prescribed electromagnetic fields. The use of analytically prescribed fields allows the principal theoretical understanding and the prediction of observable non-adiabatic effects. In order to get closer to the concrete situation of a multi-spacecraft observation, one can also combine the test particle approach with empirical models, like those of Fairfield [Delcourt et al., 1990] or Tsyganenko [Ashour-Abdalla et al., 1993]. A dynamic evolution of the fields can be taken from analytical models [Chapman and Watkins, 1993], from local MHD simulations [Scholer and Jamnitzki, 1987], from two-dimensional turbulent reconnection [Ambrosiano et al., 1988], from three-dimensional MHD reconnection [Birn and Hesse, 1994], or even from global MHD simulations (see Section 18.3) [Joyce et al., 1995]. These modifications enhance the predictive power of test particle calculations, they are also appropriate to describe the transport and propagation of energetic particles over larger distances. As for all test particle approaches, however, they still neglect the feedback of the particles to the field configuration. Hence, they too cannot correctly describe the action of the self-consistent

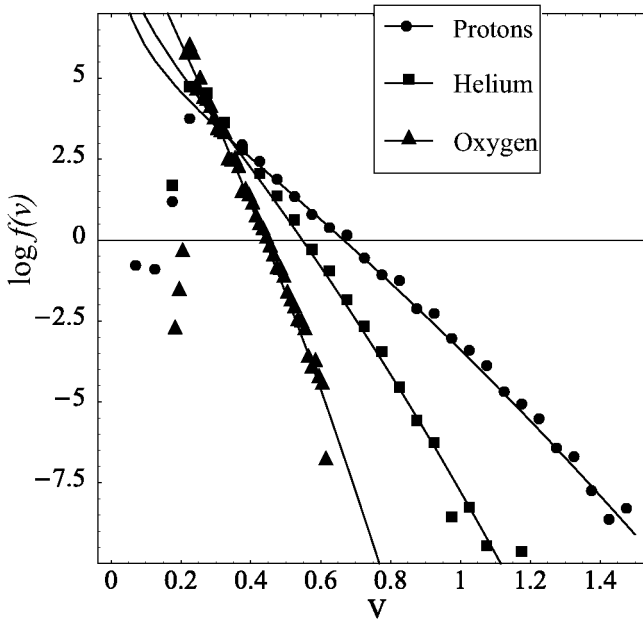


Figure 18.6: Calculated O^+ , He^{++} and H^+ test particle spectra $f(v)$ versus v and fitted curves corresponding to equation 18.2.

electric fields which are so important in shocks, double layers, reconnection or other important energisation sites in space plasmas. They are not applicable, therefore, to describe the thermal component of the plasma. For the latter, one has to apply self-consistent kinetic plasma simulations of the plasma and the fields. Before turning to the most difficult case of self-consistent kinetic plasma simulations in Section 18.4 we will discuss in Section 18.3 the advantages and limitations of MHD models for multi-spacecraft investigations.

18.3 Magnetohydrodynamic Modelling

If the multi-spacecraft investigations are aimed at the study of plasma flows and macroscopic, slowly varying electromagnetic fields and not at the self-consistent formation of distribution functions nor at the investigation of wave-particle resonances or other essentially kinetic effects, then magnetofluid approaches are most appropriate. The applicability of magnetohydrodynamic (MHD) or more sophisticated fluid approaches is based on the fact that correlated electromagnetic forces let space plasma behave, under certain circumstances, like a conducting or even ideally conducting magnetised fluid. The technical advantage of fluid approaches is that they do not have to follow the details of particle motion. They, instead, consider the dynamics of fluid elements which represent the average and undifferentiated behaviour of many particles at once. In combination with appropriate boundary conditions and assumed transport coefficients, the partial differential equations of fluid mechanics describe, therefore, the slow evolution of observable macroscopic parameters well, as there are density, bulk flows, temperature, pressure and magnetic fields.

Since fluid models of typically inhomogeneous and highly variable in time, space plasmas cannot be solved analytically, numerical simulations have to be carried out. In the form of so-called global models, numerical fluid models allow the simulation of large regions including the whole magnetosphere and the local solar wind [e.g. *Fedder and Lyon, 1995; Ogino et al., 1994; Raeder, 1994*]. For multi-spacecraft investigations, global fluid models can be best used on top of empirical models in order to describe the general situation of the magnetosphere. Due to there large grid distances global fluid models, however, do not resolve the typical Cluster distance scales. But these are typically the scales at which physical effects, other than kinetic effects, start to work. While global fluid simulations provide background information for Cluster-type multi-spacecraft measurements, more focused and higher resolution fluid simulations are necessary to accompany the investigation concentrating on key parts of the magnetosphere, such as boundaries, the inner edge of the plasma sheet, the cusp or the magnetosheath. We mention, again, that the smaller scale limits of magnetofluid approaches excluding kinetic effects are the ion gyroradii and time intervals which compare with the characteristic particle time scales.

Fluid models exist ranging from the simplest single fluid magnetohydrodynamic (MHD) approach to fields, temperature, density, bulk flow velocities and plasma pressure, to models considering temperature anisotropies, several ion components, heat fluxes etc. For example, the Chew-Goldberger-Low approach to collisionless plasmas, considers the temperatures parallel and perpendicular to the magnetic field separately. In combination with test particle calculations (Section 18.2), fluid simulations can also help to estimate spectra and distribution functions of energetic particles in evolving macroscopic fields but only in a non-self-consistent way.

Still, non-global numerical MHD models will play an important role for interpreting multi-spacecraft observations. In this section we give examples of how MHD simulations can reveal temporal variations of field parameters and plasma parameters at satellite positions. Numerical MHD simulations allow one to derive the parameter variations along satellite orbits due to the motion of plasma and field structures relative to the spacecraft. This way they generate data for developing and testing, e.g. multi-spacecraft wave analysis methods (see Section 18.3.1), they allow the recognition of the three-dimensional structure and the temporal evolution of space plasmas by multi-spacecraft measurements. We demonstrate the latter for multi-spacecraft investigations of the magnetopause in Section 18.3.2 and for the investigation of substorm related energy release processes in Section 18.3.3.

18.3.1 Identification of Wave Modes

In this section we give an example of the heuristic power of numerical MHD modelling which supports the development and testing of multi-spacecraft data analysis methods. The examples deal with a method of estimating spatial and temporal power spectra from multi-spacecraft measurements of electromagnetic fields, as described in Chapters 3 and 4. The method is basically a minimum variance analysis modified for wavelength and direction filtering. It has been tested to identify a nonlinear wave coupling process. Independent checks of the selection rules for frequencies and wave vectors have been carried out without making assumptions about the dispersion relation of the modes. The waves were generated by a two-dimensional half-compressible MHD code. Initial conditions of the simulation were chosen in a way that slow, fast, and shear Alfvén waves should be

expected. At assumed satellite positions the values of the electromagnetic fields and of the plasma density were diagnosed. Processing of these simulated data in accordance with the generalised minimum variance analysis recovered the three expected wave modes. In addition, however, a new mode was found, which was not present when initialising the simulation. It was practically impossible to identify the new mode by classical methods. Further theoretical investigations have then shown that such a mode arises from coupling between one of the shear waves and the first harmonic of the slow wave. This example is a good demonstration of the power of combining multipoint wave diagnostics with simulations. In the following section we aim at the multi-spacecraft recognition of typical signatures of large scale processes at the magnetopause and in the magnetotail.

18.3.2 Magnetopause Entry Mechanisms

The magnetopause plays an important role in solar-terrestrial relations, being the interface between the solar wind and magnetosphere. Its first theoretical description was given by Chapman and Ferraro in 1931. Since then the mechanism of the interaction between the magnetised solar wind and the Earth's magnetic field and magnetospheric plasma are still not properly understood. Multi-spacecraft investigations are expected to contribute to the solution of the puzzles concerned.

One important magnetopause transport mechanism which has to be considered is magnetic reconnection. Reconnection magnetically opens the magnetosphere through the magnetopause. *Otto et al. [1993]* have first used an MHD model of three-dimensional magnetopause reconnection to provide data from simulations for multi-spacecraft measurements. For this purpose the authors chose a simple model equilibrium for the dayside magnetopause. They allowed reconnection by imposing a localised finite resistivity. When comparing the simulated data with single satellite observations of flux transfer events (FTE), the simulation revealed a plausible match to the corresponding traces emerging from the simulation. Nevertheless, the interpretation of single spacecraft measurements remained ambiguous: an important alternative transport mechanism through the boundary, magnetopause oscillations, could create similar signatures! In order to discriminate between the different mechanisms multi-spacecraft measurements combined with numerical MHD modelling would be of use, as we will demonstrate here.

The starting point of the following simulations is the fact that the bulk velocity of the slowed solar wind in the magnetosheath still exceeds the magnetospheric plasma velocity at the inner side of the magnetopause by far. As a result, a Kelvin-Helmholtz instability (KHI) may be excited and the resulting surface waves may cause transport through the magnetopause, at least of energy and momentum, perhaps also of plasma particles if going beyond the ideal MHD approach. It is, however, difficult to verify a KHI-generated wave experimentally since it causes complex three-dimensional vortex structures corresponding to extraordinary large amplitude oscillations. Their signatures are, unfortunately, similar to those of FTEs. Single satellite observations, in the past interpreted as FTEs, are ambiguous, even when considering ideal MHD beyond accelerated and intermixed plasma. It is, therefore, an open question, whether the characteristic bipolar magnetic field signatures are due to magnetic reconnection or due to magnetopause oscillations. In addition to reconnection and KHI, similar signatures may also be caused by solar wind pressure pulses. Only multi-spacecraft measurements can allow a distinction between the potential candidates if combined with numerical MHD modelling of the magnetic fields, plasma

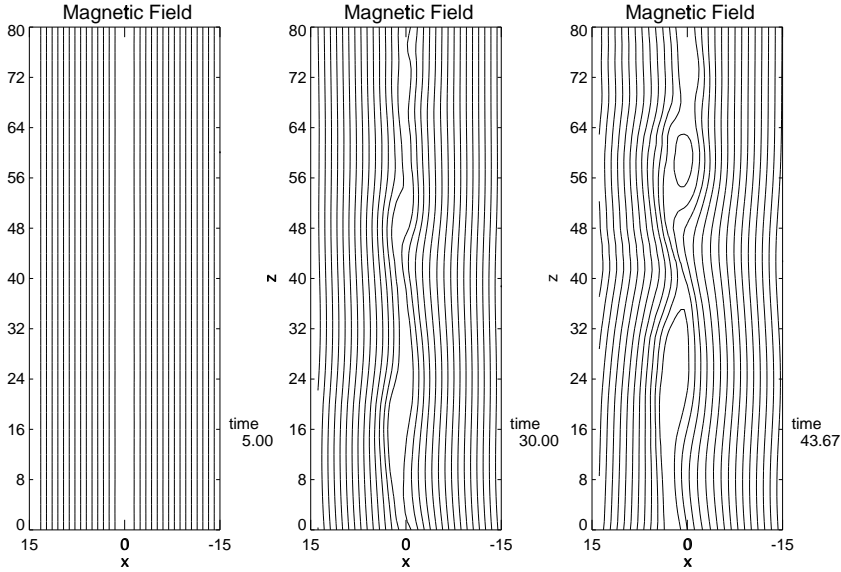
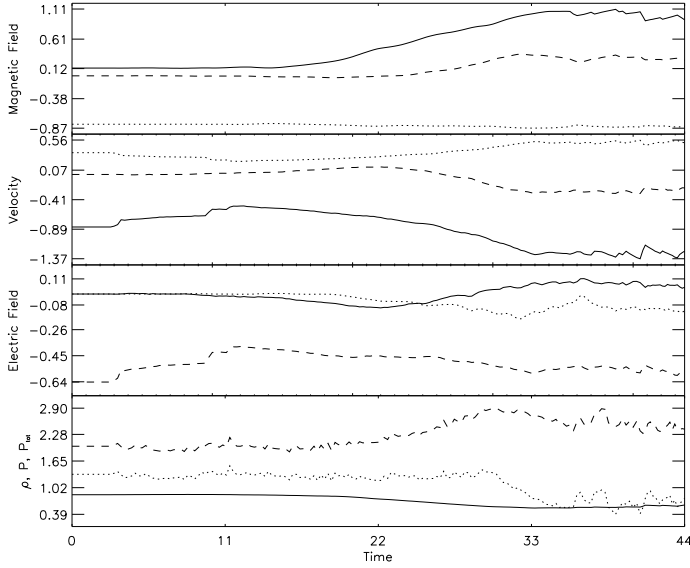


Figure 18.7: Magnetic field lines, projected into a plane $Y=\text{const.}$ after 5, 30 and 44 τ_A . The half-space $X < 0$ correspond to the magnetosphere, while $X > 0$ corresponds to the magnetosheath.

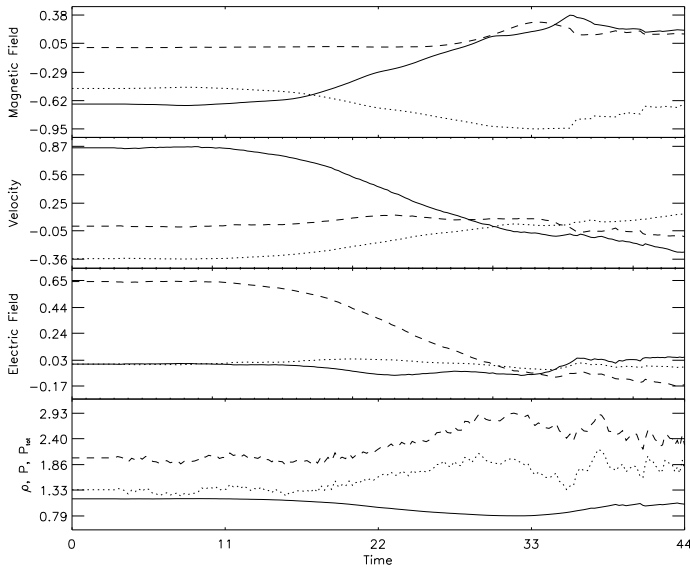
flow and pressure distribution. Magnetopause MHD simulations can be carried out for different ambient magnetic shear and wave vectors. Thus, they may yield information about the circumstances under which observed strong temporal and spatial variations are due to a KHI, due to pressure pulses or due to reconnection.

To give an example, we have initialised a two-dimensional magnetopause model for an asymptotic magnetic shear angle of 135° between the ambient fields of the magnetosheath and magnetosphere. In order to cause a KHI, an asymptotic velocity shear of $\Delta v = 4v_A$ was chosen between the magnetospheric ($x < 0$) and the magnetosheath side ($x > 0$). This value of Δv is in accordance with the theoretically predicted threshold of KHI (about $2v_A$).

Figures 18.7 and 18.8 illustrate a simulated Kelvin-Helmholtz unstable magnetopause. Figure 18.7 depicts the instantaneous magnetic field configuration after 5, 30 and 44 τ_A , where τ_A denotes the Alfvén transient times of a wave propagating through the current sheet. If rescaled to typical magnetopause parameters, the τ_A unit is of the order of one second. Figure 18.8 shows time series of the magnetic field, velocity, electric field, mass density, thermal pressure and total pressure calculated from the simulation runs at two different satellite positions. The satellites are spaced about 1000 km apart (the normalising distance in Figure 18.7). This is a typical separation for the Cluster project as well as for the Interball-1 main probe and its sub-satellite Magion 4 in the first flight-phase. In our example, the satellite positions are assumed to be at rest with respect to the magnetopause,



Position $X = -0.5$, $Z = 39.5$ (inside the magnetosphere).



Position $X = 0.5$, $Z = 39.5$ (in the magnetosheath).

Figure 18.8: Simulated time series of the relevant macroscopic plasma and field parameters obtained at assumed satellite positions in the simulation box depicted in Figure 18.7. For the three upper panels (\mathbf{B} , \mathbf{V} , \mathbf{E}) the solid, dotted and dashed lines represent the l (z), m ($-y$) and n (x) components, respectively. For the bottom panel, the plots show the mass density ρ (solid), thermal pressure (dotted), total pressure (dashed).

one in the magnetosphere ($X = -0.5, Z = 39.5$) and the other in the magnetosheath ($X = 0.5, Z = 39.5$). Both satellites see large KHI-induced oscillations in the later part of the time series. Also, the satellites observe a large scale KHI-wave structure. Both satellites indicate an increase of the normal component of the magnetic field. The satellite on the magnetospheric side observes an increase in B_z and the satellite initially located in the magnetosheath even measures a change of sign of B_z . With single-spacecraft observations only, one would not be able to decide on the magnetosheath side, whether there is a large scale wave or whether the satellite simply crosses the central current sheet. A correlated analysis of both data sets, however, helps to interpret the observations correctly.

More information can be obtained from the magnetosheath satellite time series by determining the magnetopause normal using a minimum variance analysis of the magnetic field and a maximum variance analysis of the electric field, respectively (see Chapter 8). Figure 18.9 shows the results of both variance analysis techniques. The vectors \hat{i} , \hat{j} , \hat{k} denote the eigenvectors corresponding to maximum, intermediate and minimum variance, respectively. The component of the eigenvectors are given in $\hat{l} (\equiv z)$, $\hat{m} (\equiv -y)$ and $\hat{n} (\equiv x)$ in this order. From our simulations we know that the magnetopause normal should have a strong \hat{n} component and some \hat{l} component due to the wave structure. The maximum variance analysis of the electric field yields such a result, while that of the minimum variance analysis of the magnetic field is completely different. Again, with the data from the magnetosheath satellite alone it would be impossible to decide which result is correct. However, using the two satellites together, a correlated analysis indicates a KHI wave, as mentioned above. Thus, fluctuations of the magnetic field, especially of the \hat{n} component, may make the minimum variance analysis of the magnetic field fail, while the maximum variance analysis of the electric field becomes more credible. More systematic data analysis will include additional diagnostics like the search for a deHoffmann-Teller frame of reference (where the electric field vanishes and the magnetic field is at rest, see Chapter 9), a test of the Walén relation (Section 9.3.3 of Chapter 9), the investigation of the boundary orientation and motion (Chapter 11). They all can be tested by using the results of MHD magnetopause simulations as indicated above. Notice that extensions of the MHD modelling approach are necessary when considering the high-latitude magnetopause and the cusp region, where one has to consider the Earth's magnetic field explicitly. Such simulations can be used for systematic studies of three-dimensional reconnection and its dependence of the orientation of the interplanetary magnetic field to be compared with multi-spacecraft *in situ* measurements.

18.3.3 Magnetotail Energy Release

Another important aim of multi-spacecraft investigations, where the comparison with numerical MHD modelling and simulations is needed, is the investigation of substorms and the dynamics of the corresponding energy release processes in the Earth's magnetotail. While kinetic simulations are needed to understand the causes of substorms and geomagnetic activity (see Section 18.4), MHD modelling will be of use for considering the large scale consequences of the magnetospheric activity by means of multi-spacecraft observations. Reconnection, again, is the main mechanism to be considered in order to understand the magnetospheric activity. Since reconnection is a non-local, three-dimensional and non-stationary process, numerical simulations are necessary to derive the consequences

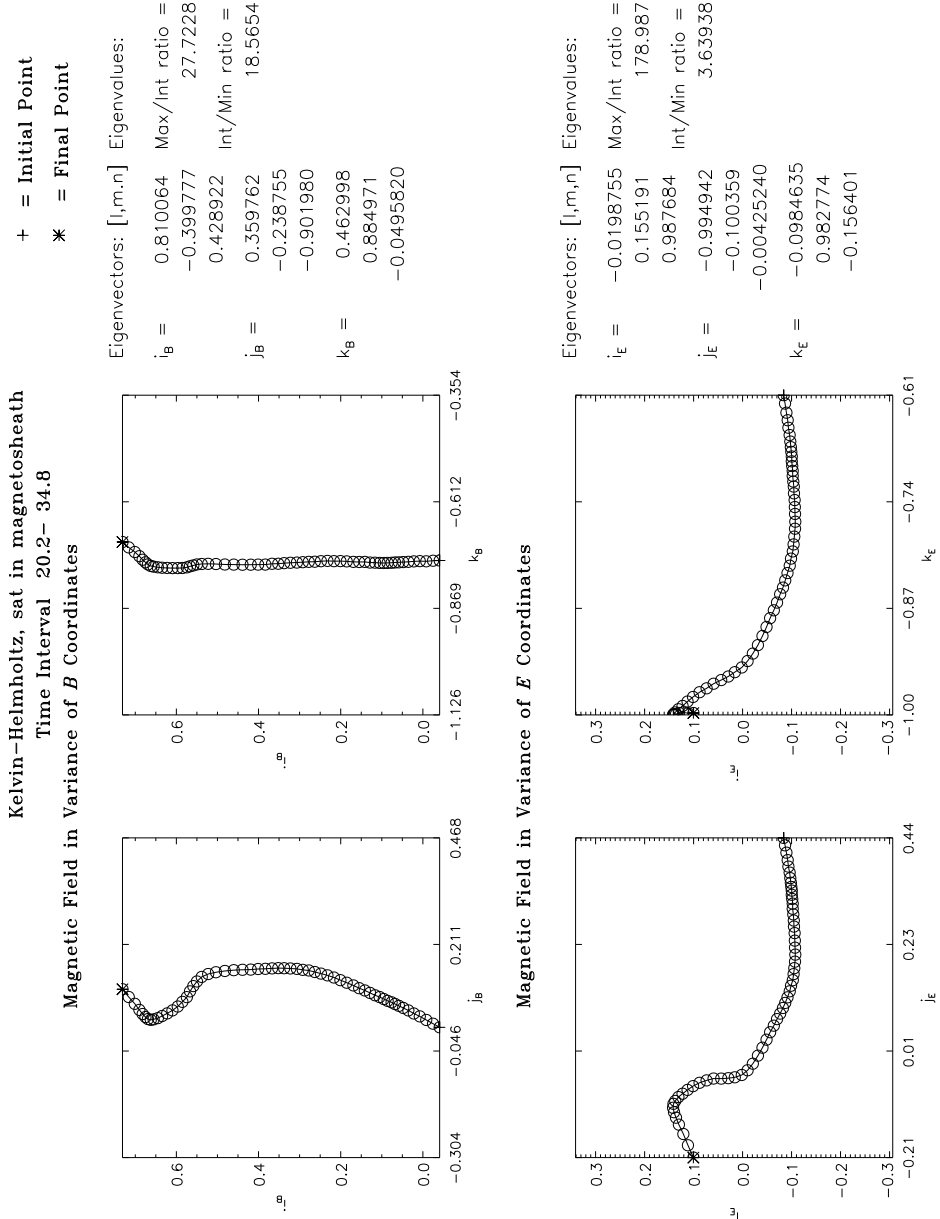


Figure 18.9: Results of a minimum variance analysis of the MHD simulated magnetic field data, and a maximum variance analysis of the simulated electric field data of the magnetosheath satellite crossing the central current sheet. The \hat{i} , \hat{j} , \hat{k} vectors denote the eigenvectors of maximum, intermediate and minimum variance, respectively, given in \hat{l} , \hat{m} , \hat{n} coordinates.

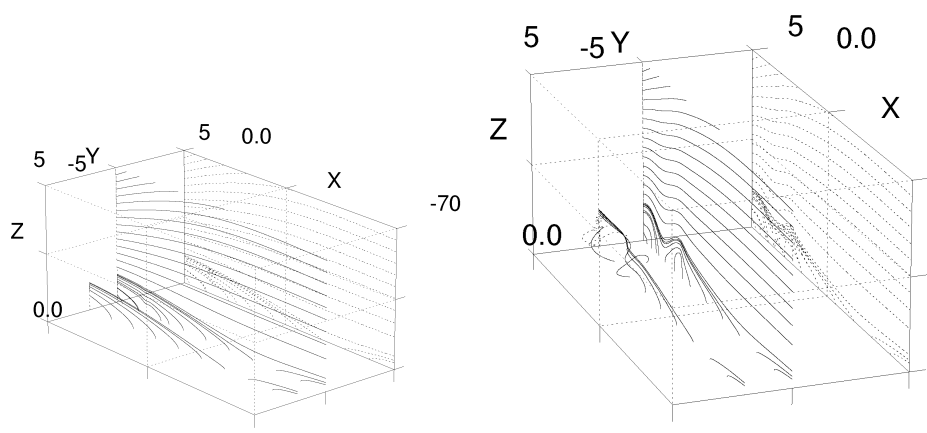


Figure 18.10: Reconnecting magnetic field at different moments of time.

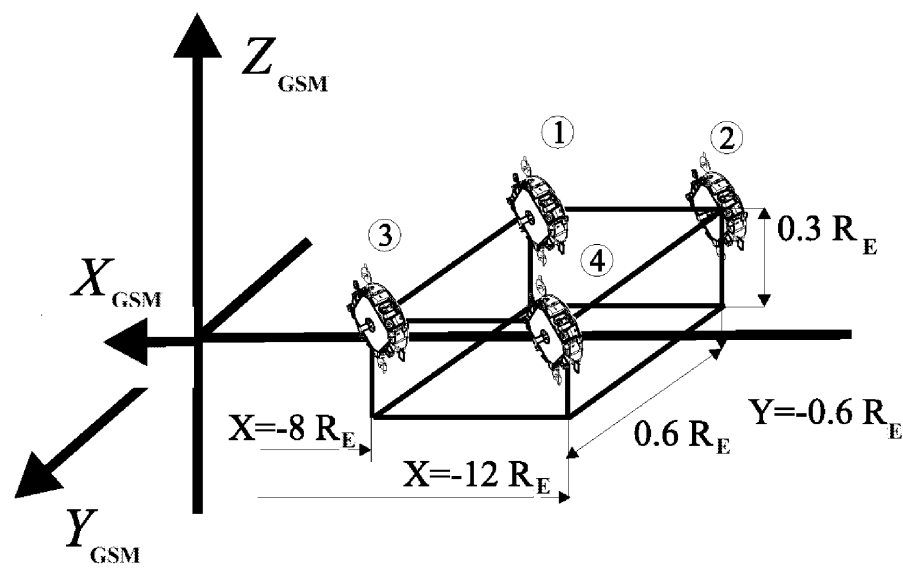


Figure 18.11: Satellite positions used for the multi-spacecraft diagnosis carried out in Section 18.3.3.

of different models. This concerns the evolution of flux tubes, the plasmoid dynamics, a detailed analysis of the plasma flows and systematic studies of the influence of different microscopic transport models on the nonlinear evolution of reconnection. In particular, studies of the evolution of the current density distribution are of interest with respect to multi-spacecraft data interpretation. Simulated multi-spacecraft data time series yield then the opportunity of a direct comparison with satellite data.

In order to demonstrate the strength of numerical MHD modelling for interpreting multi-spacecraft data, we have diagnosed a tail reconnection simulation from the points of view of four satellites, located around a reconnection region. The simulation is three-dimensional, starting with a two-dimensional tail equilibrium, invariant in the Y -direction. At $t = 0$ a Gaussian shaped three-dimensional resistivity profile is imposed, localised near $X = -10 R_E$ and $Y = Z = 0$ in the equatorial plane. The start configuration includes a far-Earth neutral line near $X = -50 R_E$. Figure 18.10 depicts magnetic field lines at two different moments of time during a MHD simulated three-dimensional magnetotail evolution. In the figure X, Y, Z correspond to GSM coordinates given in Earth radii (R_E). Following the global evolution of the configuration, three-dimensional reconnection evolves together with a developing plasmoid structure.

From the simulation data, time series can be deduced as they would be seen from four different satellite positions. The spacecraft are all positioned around the reconnection region located at $X = -10 R_E$, $X = 0$ and $Z = 0$ (Figure 18.11). For demonstration purposes a case is shown where all four satellites are at rest with respect to the reconnection region. The satellites are located northward of the equatorial plane at a distance of $Z = 0.3 R_E$. Two satellites are located earthward of the reconnection region. The corresponding data time series at these positions are shown in Figure 18.12 and Figure 18.14, respectively. Two more satellites are situated tailward of the reconnection region (Figure 18.13 and Figure 18.15, respectively).

Although the three-dimensional dynamics of the configuration is rather complicated, a correlated analysis of the simulated time series of data yields clear signatures of three-dimensional magnetic reconnection. Satellite 1 at $X = -8 R_E$, $Y = -0.6 R_E$, $Z = 0.3 R_E$ (Figure 18.12) measures a positive, i.e. earthward directed, velocity v_x , together with a positive B_z . Satellite 2 at $x = -12$, $y = -0.6$, $z = 0.3$ (Figure 18.13), located tailwards of the reconnection region, measures tailward-directed plasma flow (negative v_x) and a negative B_z component. The same feature is found in the data sets of satellites 3 and 4, located at the duskside at $y = 0.6$ (Figures 18.14 and 18.15). The time series of v_x and B_z of the earthward satellite (Figure 18.14) correspond to that found in Figure 18.12. The v_x, B_z data in Figure 18.15 (tailward located satellite) fit the feature in Figure 18.13. Inflow and outflow from the flanks of the reconnection region can be studied in detail from the data sets obtained for the satellite cluster. Even just concentrating on the observation of v_x and B_z , the correlated analysis of multi-spacecraft measurements would yield information about the consequences of three-dimensional MHD reconnection for macroscopic parameters, allowing a distinction between different model approaches of geomagnetic activity phenomena.

What MHD models do not allow in principle is the investigation of the microscopic nature of the energy release processes, based on wave-particle resonances or other kinetic effects. Particle effects have to be addressed by kinetic models, which, again, require numerical simulations as described in the following sections of this chapter.

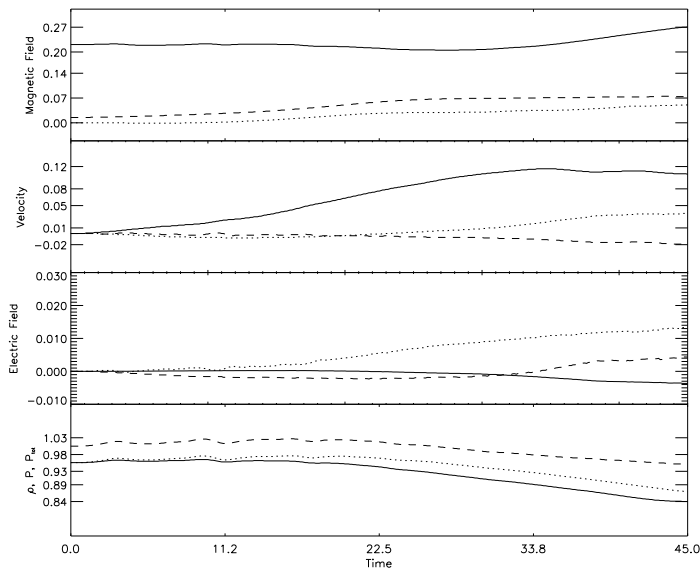


Figure 18.12: Simulated time series of plasma and field data measured at satellite position 1 ($X = -8 R_E$, $Y = -0.6 R_E$, $Z = 0.3 R_E$). For the three upper panels (\mathbf{B} , \mathbf{V} , \mathbf{E}) the solid, dotted, dashed lines represent the x , y , z components, respectively. For the bottom panel, the plots are mass density ρ (solid), thermal pressure (dotted), total pressure (dashed).

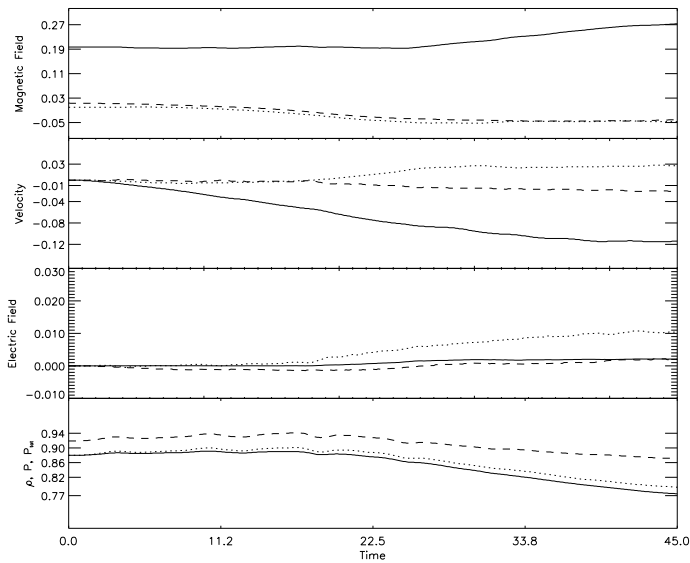


Figure 18.13: Simulated time series of plasma and field data measured at satellite position 2 ($x = -12$, $y = -0.6$, $z = 0.3$). The line types have the same meaning as in Figure 18.12.

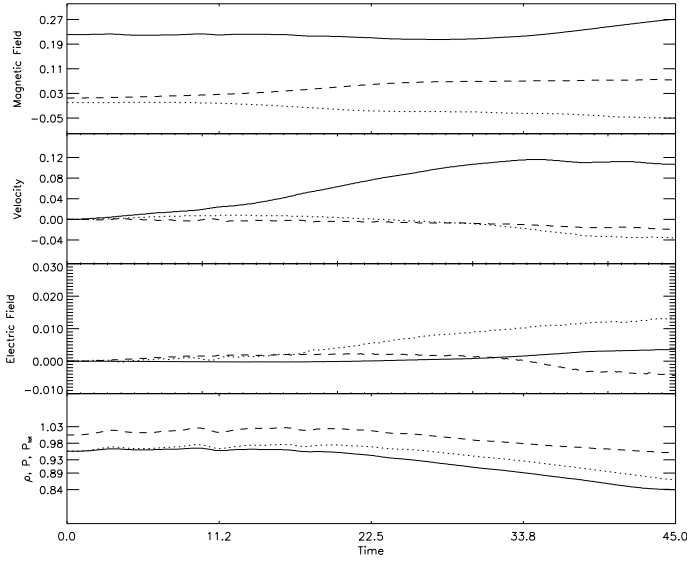


Figure 18.14: Simulated time series of plasma and field data measured at satellite position 3 ($X = -8 R_E$, $Y = 0.6 R_E$, $Z = 0.3 R_E$). For the three upper panels (\mathbf{B} , \mathbf{V} , \mathbf{E}) the solid, dotted and dashed lines represent the X , Y and Z components, respectively. For the bottom panel, the plots are mass density ρ (solid), thermal pressure (dotted), total pressure (dashed).

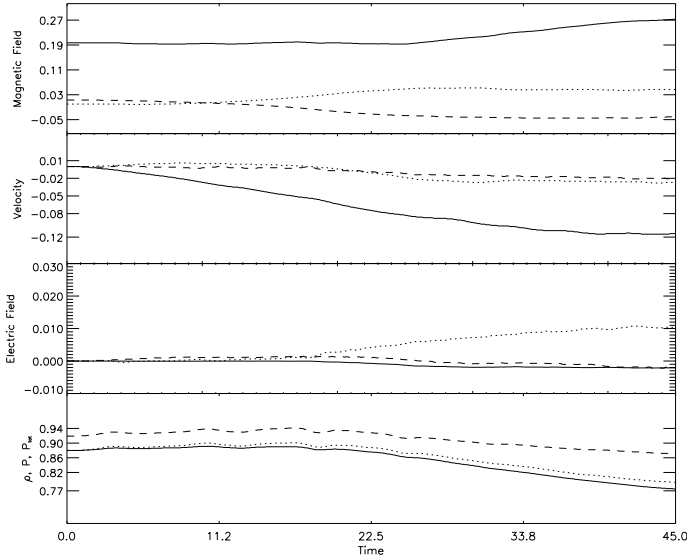


Figure 18.15: Simulated time series of plasma and field data measured at satellite position 4 ($X = -12 R_E$, $Y = 0.6 R_E$, $Z = 0.3 R_E$). The line types are same as in Figure 18.14.

18.4 Kinetic Plasma Simulations

Although test particle simulations, discussed in Section 18.2, can be used to interpret multi-spacecraft measurements of energetic particles, they are not self-consistent. Hence, they do not describe, e.g., plasma heating or feedback of particles on the fields. MHD simulations, discussed in Section 18.3 are self-consistent, but they still neglect particle effects. They are limited to spatial scales larger than ion gyroradii and to time intervals longer than gyro- and plasma periods. They do not reveal particle distribution functions and spectra self-consistently. These limits can be overcome only by a self-consistent kinetic plasma description. Unfortunately, only the simplest kinetic models, like one-dimensional electromagnetic or two-dimensional electrostatic approaches, can be treated analytically. They are useful for demonstration purposes, having helped to understand, e.g., the principles of local cyclotron resonances, of Landau damping, and unstable wave generation. For low-amplitude waves the corresponding nonlinear differential equations can be linearised and quasi-linear methods describe a weak feedback of waves on the particles. Even some essentially nonlinear equations are analytically solvable, describing, for example, low-dimensional collapse and solitons. Nevertheless, these models and equations over-simplify reality. They do not properly describe the situation met in multi-satellite studies. In slightly more realistic considerations of, e.g., wave-particle resonances in inhomogeneous plasmas, one has to deal with several coupled partial differential equations for the electromagnetic fields combined with nonlinear ordinary differential equations of the particle motion. For their solution analytical methods fail. Kinetic simulations help to describe the consequences of particle and field interactions self-consistently [e.g. *Winske and Omidi, 1996*].

In self-consistent kinetic plasma simulations, now available, fields are calculated on finite grids of limited size although gridless solutions are possible in principle. The field values are interpolated with a high degree of accuracy to any spatial position. In particle-in-cell (PIC) codes, particles are clustered in so-called macro-particles; they move in the fields interpolated between the grid-points. Macroscopic moments like density, pressure tensor elements, temperatures, heat flows as well as phase space distribution functions are calculated from the particles at any time step by relating the relative contribution of all particles to a grid point. Due to the box-type simulation approach, the choice of initial and boundary conditions is very important. As a matter of fact, their choice requires high sophistication. The most important limit of kinetic plasma simulations arises, however, from the finite number of particles, which is dictated by the limits of computer resources. For special applications, therefore, a restriction to not fully self-consistent kinetic simulations may be of use. In Section 18.4.1 (Fokker-Planck simulations), Section 18.4.2 (guiding-centre oriented gyrokinetic simulations), and Section 18.4.3 (hybrid simulations) we give examples of different restricted kinetic approaches, developed in the framework of the European network preparing the first Cluster mission. In Section 18.4.4 we finally turn to fully self-consistent kinetic electromagnetic plasma simulations which have now started to be carried out in three dimensions necessary for comparison with multi-spacecraft measurements. We will illustrate the opportunities of fully kinetic plasma simulations by results obtained with the three-dimensional, fully self-consistent particle code GISMO.

18.4.1 Fokker-Planck Simulations

Phase space diffusion processes can be investigated by numerically solving Fokker-Planck equations of the space density evolution. A group at AEA Culham, United Kingdom, has designed a two-dimensional and finite difference Fokker-Planck code that describes the diffusive acceleration of electrons by resonant lower hybrid waves. A simulation of processes producing the non-thermal velocity distributions of the auroral electrons has illustrated the opportunities of this approach. In the context of multi-spacecraft observations Fokker-Planck simulations can be used for testing theoretically derived transport coefficients and comparing their predictions of the phase space evolution (scattering processes, etc.) with observations.

18.4.2 Guiding Centre Simulations

Guiding centre codes follow just the guiding centres of one or more particle species (mostly electrons) instead of the particles themselves. Such an approach is applicable as long as the particles are magnetised, conserving the first adiabatic moment of motion μ (Section 18.2.1) and as long as the gyration phase is unimportant.

In a collaboration of three groups of the European Simulation Network for Cluster (CETP, École Polytechnique, and Culham) an implicit electromagnetic code designed at the École Polytechnique was extended to simulate the dynamics of the guiding centres of the electrons. This new scheme does not require projections on the local magnetic field, hence it should be more efficient than previous guiding centre codes. Starting from a guiding-centre approximation, one can reduce the calculation efforts of a kinetic code even further. A group at Rutherford Appleton Laboratory, for example, has designed an explicit electromagnetic code in the so-called Darwin approximation of Maxwell's equations which neglects inductive electric fields. They make use of a guiding-centre approach for electrons and considering ring-shaped ion distributions. Simulations of tearing modes and drift waves are among potential applications of this operational two-dimensional code.

18.4.3 Hybrid Simulations

Hybrid plasma simulations combine the kinetic treatment of one or several plasma species as particles with a fluid model of the remaining plasma components. They have been developed and broadly used over the past ten years, for example, for investigating the ion acceleration at shock waves or the role of waves in shock formation [Dubouloz and Scholer, 1993], or for investigations of certain aspects of the reconnection process [Hess et al., 1995]. In the framework of the European simulation network, a new numerical scheme for hybrid simulations has been designed [Matthews, 1994]. Both one- and two-dimensional implementations have been made and tested. The ability of this code to handle multiple ion species efficiently makes it a valuable tool to investigate the physics of collisionless shocks in the solar wind, low-frequency micro-instabilities at the magnetopause, and the associated diffusion of ions through the magnetopause. The new scheme, called Current Advance Method and Cyclic Leapfrog (CAM-CL), is distinct from other hybrid schemes. First, multiple ion species may be treated with only a single computational pass through the particle data. While the moment method advances the fluid velocity, CAM advances the ionic current density and therefore computes more efficiently the dynamics

of multiple ion species. This current advance is made easier by the collection of a free streaming ionic current density. CL is a leapfrog scheme for advancing the magnetic field, adapted from the modified midpoint method. It allows sub-stepping of the magnetic field. An application of this code which will be of use for multi-spacecraft diagnostics is the diagnosis of ion distribution functions formed by reconnection [*Scholer and Lottermoser, 1997*].

18.4.4 Fully Self-consistent Kinetic Plasma Simulations

In many cases, however, the simplifications of restricted kinetic plasma simulations, discussed in Sections 18.4.1–18.4.3, go too far. They over-simplify if, e.g., particle resonance effects cannot be neglected for neither the electrons nor the ions. Typical examples of resonance effects are plasma instabilities, the dissipation of waves as in collisionless shocks, reconnection problems, or other microscale-based energy releases. In such cases, fully self-consistent kinetic plasma simulations have to be carried out. Not only do they contain all the plasma physics self-consistently, they also provide the necessary resolution of spatial, temporal, angular and velocity information for the planned multi-spacecraft diagnostics.

Self-consistent kinetic plasma simulations are based on the solution of a huge number of differential equations, the Maxwell equations for the fields and the equations of motion of many particles. All this information has to be stored, at least temporarily. The PIC approach reduces the amount of equations and information to be stored because many particles together are represented by a smaller number of “macro-particles”. Details of the method can be found in the literature, [e.g. *Birdsall and Langdon, 1991*; *Matsumoto and Sato, 1985*; *Tajima, 1989*]. But the number of macro-particles must still be large enough to guarantee a sufficient statistical significance of the calculated plasma density and currents. These days several millions can still be handled by workstations. Tens of millions of macro-particles, have to be dealt with by parallel computers. Even with the best existing computers, however, problems like the simulation of time dependent, statistically significant distribution functions in three dimensions, the treatment of large boxes and realistic mass ratios (usually of the order ten or a hundred for M_i/m_e) are still unsolved.

Let us demonstrate opportunities and limitations of fully self-consistent particle simulations by presenting new results on the instability and decay dynamics of collisionless current sheets. The latter has to do with the very cause for reconnection, like at substorm onsets. It is also a typical example of the necessity of considering both ions and electrons kinetically. In the simplest case of a laminar Harris-type current-sheet the theoretically predicted decay is due to an absolutely growing tearing mode leading to two-dimensional reconnection. Unfortunately, the predicted mode of current sheet decay was observed neither at the Earth’s magnetopause nor in the magnetotail. Indeed, as it can be shown theoretically, sheets containing a finite cross-sheet normal magnetic field component are stable—at least in two dimensions. Two-dimensional fully self-consistent kinetic plasma simulations have confirmed this result.

The situation changes completely in three spatial dimensions. Since the consequences of three-dimensional non-local kinetic interactions cannot be considered analytically, one has to carry out kinetic plasma simulations to reveal the properties of the three-dimensional current sheet decay as well as typical signatures of this process for experimental multi-

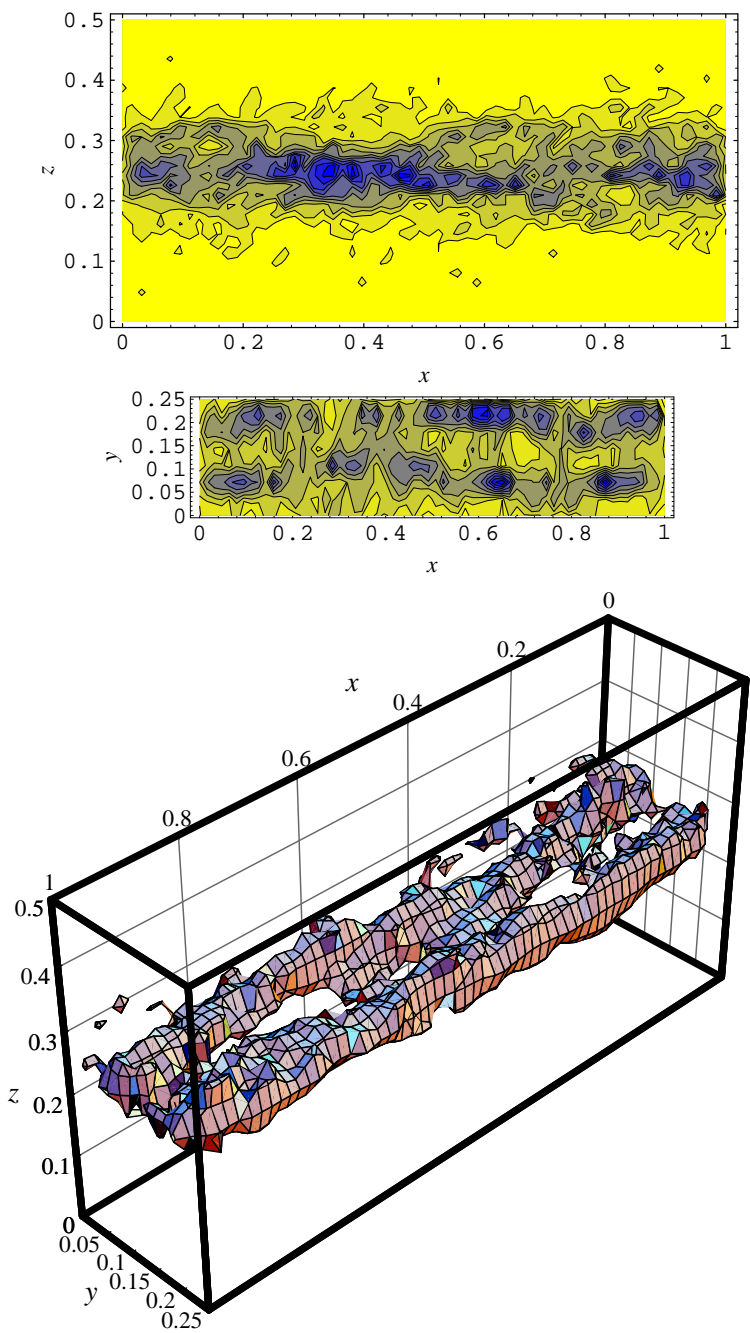


Figure 18.16: Three-dimensional Harris-type current sheet decay: perspective view at an isodensity surface (lower panel) and cuts through its meridian (upper panel— $y = 0.125$) and equatorial plane ($z = 0.25$) [after *Büchner and Kuska, 1996*].

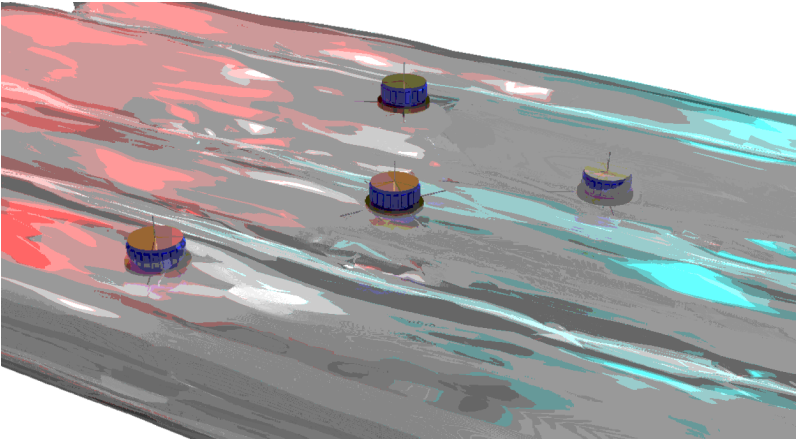


Figure 18.17: Current instability, observed by four satellites.

spacecraft investigations. A kinetic approach is needed since particle properties of both electrons and ions have to be considered. In our demonstration we use results obtained with the GISMO code, a three-dimensional, electromagnetic, fully kinetic code which we have developed especially for multi-spacecraft applications. As usual for PIC codes, since one has to resolve the full electron dynamics, an artificial mass ratio must be set, otherwise the time scales, to be resolved, would cover too many orders of magnitude. At the moment the mass ratios applied reach up to 1:256. This allows the consideration both of the electron-ion mass separation and of the role of heavy ions, e.g. of ionospheric O^+ ions, 16 times heavier than protons. Figure 18.16 summarises the resulting structure of a three-dimensional Harris current sheet decay. The upper two plots depict cross-sections of the three-dimensional isodensity surfaces through the meridian and equatorial planes, respectively, while the lower panel shows a three-dimensional perspective view of the iso-surface of $2/3$ of the maximum plasma density distribution. How does such structure arise? An investigation by the same code has shown that in the case of thin current sheets, reconnection in three dimensions is preceded by a bulk current instability. This instability forms density structures depicted in Figure 18.17. The figure shows an instantaneous isodensity surface zoomed out of the current sheet centre region. Before reconnection starts, the thin current sheet becomes modulated by a sausage-type bulk instability. The resulting density modulation, shown in the figure, propagates at about the ion drift speed in the current direction. Such structure formation can be observed, obviously, only by means of multi-satellite studies. We put four satellites, positioned at distances which correspond to those of four Cluster spacecraft, in the figure. This way the figure demonstrates that multi-spacecraft observations are able to recognise the moving current sheet density structure which should precede magnetotail reconnection. The instability illustrated by Figure 18.17 enables the growth of magnetic fields connecting the two half spaces above and below the sheet. As soon as these fields have reached a certain strength fast reconnection starts. The three-dimensional combination of sausage and reconnection instability develops at a much faster rate rather than the two-dimensional tearing instability. In loop-like fields (with a finite cross-sheet magnetic field component present from the very beginning), the creation

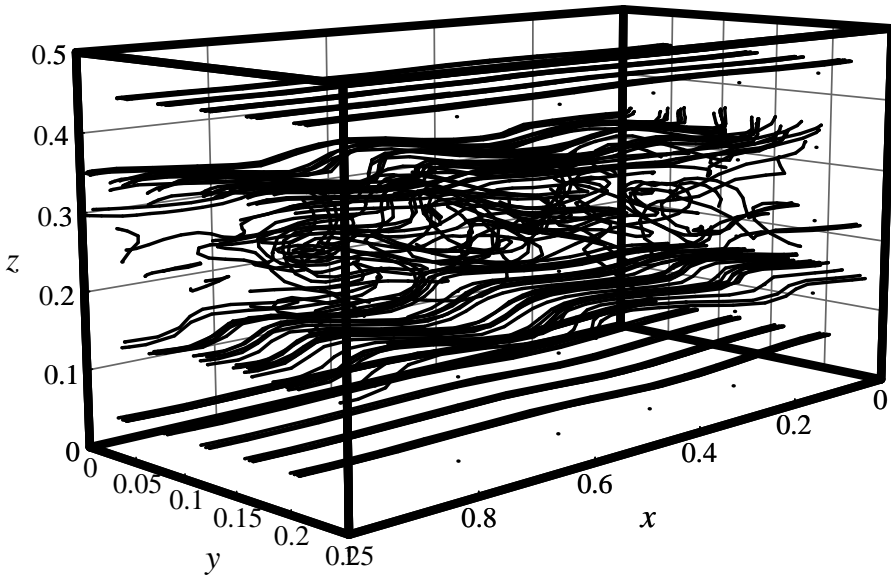


Figure 18.18: Three-dimensional reconnection magnetic field, sheared due to Hall currents.

of reconnection magnetic fields across the sheet allows reconnection while the classical tearing instability cannot develop at all.

However, the fully kinetic plasma simulations by GISMO have shown more features of interest for multi-spacecraft studies which cannot be seen in MHD models. As a consequence of the mass differences between electrons and ions they have revealed, e.g. that reconnection is surrounded by a three-dimensional Hall current system. The latter causes a sheared magnetic field structure shown in Figure 18.18 which can be used to identify three-dimensional reconnection in the magnetospheric tail by multi-spacecraft magnetic field measurements.

Other multi-spacecraft investigations which can be carried out by comparing with numerical PIC-code simulation results, are the consequences of the presence of heavy ions of ionospheric origin like O^+ in the magnetosphere. These are the possible causal roles of ionospheric ions in initiating at least subsequent substorms and storms and their heating as indications of specific energisation processes.

Another feature, which was not explained on the basis of single spacecraft observations alone and which requires self-consistent kinetic plasma simulations are so-called bursty bulk flows (BBF) in the magnetotail. Fully kinetic GISMO simulations reveal that BBFs may indicate just the transitional states between very localised and patchy acceleration and larger scale coherent reconnection. While BBF would correspond to small scale transient releases of energy from the overstressed magnetotail, the formation of plasmoids would be indicative of a large scale instability on the scale of the whole tail.

Although the limitations of computer resources still strain the performance of kinetic plasma simulations, kinetic codes will soon allow the self-consistent determination of

spatial-temporal variations of velocity space distributions and the simulation of larger boxes and mass ratios, realistic enough to be directly compared with multi-spacecraft measurements.

Summary and Conclusion

With the coming era of multi-spacecraft measurements the analysis of satellite data will become much more quantitative and discriminative as compared with single point observations.

In order to deduce the potentially rich information contained in multi-spacecraft observations, however, numerical modelling and simulations have to be applied. Numerical simulations are needed starting with the development and testing of specific analysis methods, all the more for the interpretation of real data. In the latter process, numerical models will have to fill the gap between hypotheses and real data.

Nowadays one can choose among many existing numerical modelling and simulation approaches in accordance with the physical problem one intends to solve. In this chapter we have discussed opportunities, applicability and limitations of numerical modelling and simulations for multi-spacecraft investigations giving examples based on the three main approaches.

The test particle approach, discussed in Section 18.2, is quite relevant for the multi-spacecraft diagnostics of remote acceleration processes using energetic particles as tracers. The non-adiabatic energetic particle dynamics in the inhomogeneous cosmic electromagnetic fields and plasma boundaries forms the basis for remote multi-spacecraft diagnostics—resulting space and velocity space distributions can be well described in the framework of a test particle approach (Section 18.2.1). Utilising the results of numerical test particle calculations, multi-spacecraft configurations can even form a dispersiometer using the Earth's magnetic field as a mass and energy spectrometer. Test particle calculations are a particularly valuable tool for the interpretation of mass-separating particle spectra (Section 18.2.2). A combination of the test particle approach with empirical or large scale magnetohydrodynamic field models, providing spatially distributed and time dependent fields, can be used to predict the propagation of energetic particles. The success of test particle calculations depends on the quality of the field models used since they are not calculated self-consistently. In most cases the test particle approach is not applicable to the thermal plasma component.

Fluid approaches like magnetohydrodynamics (MHD), discussed in Section 18.3, are self-consistent, if transport coefficients are postulated, i.e. as long as kinetic details and smaller scales can be neglected. Fluid modelling continues to be of worth to describe and analyse slowly varying fields, flows, density, pressure and temperatures. As demonstrated in Section 18.3, MHD simulations are well suited to predict and analyse the evolution of fields and macroscopic plasma parameters in space and time, for example, for low frequency plasma waves, magnetopause transport scenarios, and the magnetospheric energy release processes. In particular we have cited in Section 18.3.1 the successful identification of wave modes using numerical MHD modelling, in Section 18.3.2 the abilities of MHD models to predict discriminating signatures of surface waves generated by KHI, in contrast to magnetopause reconnection as the two main candidates for the transport between solar wind and magnetosphere through the magnetopause. The prediction of typical

multi-spacecraft signatures in plasma flows and magnetic field variations in a reconnecting magnetospheric tail have been discussed in Section 18.3.3.

The importance of kinetic plasma simulations for multi-spacecraft measurements has been argued for in Section 18.4. Only self-consistent kinetic plasma simulations can meet the sophistication of advanced multi-spacecraft space plasma and particle diagnostics. Only they contain all plasma phenomena and can provide complete information about velocity space distributions for any given point and related to any moment of time during the simulation run together with the fields, evolving with the particles and plasma. The combination of multi-spacecraft observations and kinetic plasma simulations allows the investigation of the real plasma dynamics below the spatio-temporal resolution of the MHD approximation, of wave-particle resonance effects as well as of the consequences of the mass difference between electrons and ions.

Since fully kinetic simulations consume considerable computer resources, one may apply restricted kinetic codes for special applications. We have demonstrated utilities and limitations of restricted kinetic simulations by citing codes, developed in the framework of a European network program supporting numerical simulations for multi-spacecraft missions in (Sections 18.4.1-18.4.3). A first kind of restricted simulation uses known numerical schemes. It targets operational codes to do physical simulations in the short term. Restricted kinetic simulations of a second kind are more exploratory. They aim at the design of new numerical schemes, especially tailored to investigate the physics of boundary layers where the large, typical fluid scales, and the small, typically kinetic scales interact (see Section 18.4.3). They can be of use for comparison with multi-spacecraft observations as well.

Finally, in Section 18.4.4 we have demonstrated how multi-spacecraft observations in combination with self-consistent kinetic plasma simulations enable the understanding of essentially kinetic plasma physical processes like, e.g., energy release in collisionless space plasmas. As an example we have treated the dynamic evolution of current sheets, a process, essentially governed by kinetic plasma effects. We have demonstrated the development of a three-dimensional kinetic current instability and the transition to reconnection. The mass separation causes a three-dimensional Hall current system around a reconnection region which can be used for multi-spacecraft diagnostics. Other effects which can be described only by kinetic simulations, are the influence of heavy ions, the transition from localised and patchy energy releases to larger scale coherent structures, the spatial and temporal variation of distribution functions as fine indicators of the physical processes in collisionless plasmas. In contrast to test particle calculations, kinetic simulations allow the derivation of the shape and spatial variation of the velocity space distribution of all particles, not only of the energetic ones, in a self-consistent way. The limiting factor of self-consistent kinetic plasma simulations is the necessity to resolve the Debye length and the plasma period as well as the electron cyclotron period, whichever is smaller. The propagation of electromagnetic waves between the grid points must be resolved as well. Further, the number of particles followed is always limited. If one intends to simulate realistic plasma regimes one has to limit the simulation box size. Global kinetic plasma simulations are, therefore, in contrast to global MHD simulations, a totally future effort.

Summing up, the rich information contained in multi-spacecraft measurements, can evidently be deduced only with the help of appropriate numerical modelling and simulations efforts. For this purpose the different approaches have to be further developed according to different aims, especially kinetic plasma simulations, to be available at the

moment when multi-spacecraft observations become reality.

Acknowledgements

The DARA contribution to H. Wiechen's work (contract 50 QN 9202) and the DFG contribution to the development of GISMO by J.-P. Kuska (contract Bu 777 2-1) are gratefully acknowledged as well as helpful contacts and discussions with the Cluster team, especially with A. Balogh, G. Paschmann, G. Chanteur, P. W. Daly, S. J. Schwartz and B. Wilken.

Bibliography

- Ambrosiano, J., Matthaeus, W. H., Goldstein, M. L., and Plarte, D., Test particle acceleration in turbulent reconnecting magnetic fields, *J. Geophys. Res.*, **93**, 14 383, 1988.
- Ashour-Abdalla, M., Berchem, J., Büchner, J., and Zelenyi, L., Shaping of the magnetotail from the mantle: Global and local structuring, *J. Geophys. Res.*, **98**, 5651, 1993.
- Birdsall, C. K. and Langdon, A. B., *Plasma Physics via Computer Simulation*, Plasma Physics Series, IOP Publishing, 1991.
- Birn, J. and Hesse, M., Particle acceleration in the dynamic magnetotail: Orbits in self-consistent three-dimensional MHD fields, *J. Geophys. Res.*, **99**, 109, 1994.
- Büchner, J., Energetic particles in the magnetotail: Cluster and Regatta as a dispersiometer, in *Proceedings of the International Workshop on "Space Plasma Physics Investigations by Cluster and Regatta"*, Graz, Feb. 20–22, 1990, ESA SP-306, p. 117, European Space Agency, Paris, France, 1990.
- Büchner, J., The nonlinear dynamics of strong particle acceleration by reconnection, in *Proceedings of the Cambridge Symposium/Workshop in Geoplasma Physics on "Multi-scale Phenomena in Space Plasmas"*, edited by T. Chang and J. R. Jasperse, no. 14 in *Physics of Space Plasmas*, pp. 57–77, MIT Center for Theoretical Geo/Cosmo Plasma Physics, Cambridge, MA, 1995.
- Büchner, J. and Kuska, J.-P., Three-dimensional collisionless reconnection through thin current sheets: Theory and self-consistent simulations, in *Proc. 3rd International Conference on Substorms*, ESA SP-389, p. 373, European Space Agency, Paris, France, 1996.
- Burgess, D., Mission oriented modelling in Europe, in *Solar-Terrestrial Energy Program*, edited by D. N. Baker, V. Papitashvili, and M. Teague, vol. 5 of *COSPAR Colloquia series*, p. 723, Pergamon Press, 1994.
- Chanteur, G. and Roux, A., European network for the numerical simulation of space plasmas, *Space Sci. Rev.*, **79**, 583–598, 1997.
- Chapman, S. and Watkins, N. W., Parametrization of chaotic particle dynamics in a simple time-dependent field reversal, *J. Geophys. Res.*, **98**, 165, 1993.
- Delcourt, D. C., Sauvaud, J. A., and Pedersen, A., Dynamics of single-particle orbits during substorm expansion phase, *J. Geophys. Res.*, **95**, 20 853, 1990.
- Dubouloz, N. and Scholer, M., On the origin of the short large-amplitude magnetic structures upstream of quasi-parallel collisionless shocks, *Geophys. Res. Lett.*, **20**, 547, 1993.
- Fedder, J. A. and Lyon, J. G., The Earth's magnetosphere is 165 R_E long: Self-consistent currents, convection, magnetospheric structure, and processes for northward interplanetary magnetic field, *J. Geophys. Res.*, **100**, 3623–3635, 1995.

- Hess, M., Winske, D., and Kuznetsova, M. M., Hybrid modeling of collisionless reconnection in two-dimensional current sheets: Simulations, *J. Geophys. Res.*, **100**, 21 815–21 825, 1995.
- Joyce, G., Chen, J., Slinker, S., and Holland, D. L., Particle energization near an X-line in the magnetotail based on global MHD fields, *J. Geophys. Res.*, **100**, 19 167, 1995.
- Matsumoto, H. and Sato, T., *Computer Simulation of Space plasma*, D. Reidel Publ. Comp., Dordrecht, Boston, Lancaster, 1985.
- Matthews, A., Current advance method and cyclic leapfrog for 2D multispecies hybrid simulations, *J. Comp. Phys.*, **112**, 102, 1994.
- Ogino, T., Walker, R. J., and Ashour-Abdalla, M., A global magnetohydrodynamic simulation of the response of the magnetosphere to a northward turning of the interplanetary magnetic field, *J. Geophys. Res.*, **99**, 11 027–11 042, 1994.
- Otto, A., Ziegler, H. J., and Birk, G., Plasma and magnetic signatures generated by three-dimensional MHD-simulations, in *Proc. International Conf. "Spatio-Temporal Analysis for Resolving plasma Turbulence (START)"*, Aussois, 31 Jan.–5 Feb. 1993, ESA WPP–047, p. 337, European Space Agency, Paris, France, 1993.
- Raeder, J., Global MHD simulations of the dynamics of the magnetosphere: Weak and strong solar wind forcing, in *Proc. of the 2nd Int. Conf. on Substorms*, edited by J. R. Kan, J. D. Craven, and S.-I. Akasofu, p. 561, Univ. of Alaska, Fairbanks, Alaska, 1994.
- Scholer, M. and Jamnitzki, F., Particle orbits during the development of plasmoids, *J. Geophys. Res.*, **92**, 12 181, 1987.
- Scholer, M. and Lottermoser, Distribution functions in reconnection—hybrid simulations, *J. Geophys. Res.*, 1997, submitted for publication.
- Tajima, T., *Computational Plasma Physics*, Addison-Wesley, 1989.
- Winske, D. and Omidi, N., A nonspecialist's guide to kinetic simulations of space plasmas, *J. Geophys. Res.*, **101**, 17 287–17 303, 1996.

Notes

The use of endnotes and their markers in the text is described in the [Preface](#) to the electronic edition.

Note 0.1 (page 1)

For the electronic version, a new first paragraph and expanded list of references have been added to the Introduction.

Note 0.2 (page 3)

In lieu of a chapter on turbulence, see the article by

Horbury, T. S., Cluster II analysis of turbulence using correlation functions, in *Proceedings of the Cluster-II Workshop Multiscale/Multipoint Plasma Measurements, Imperial College, London, 22–24 September 1999*, ESA SP-449, pp. 89–97, European Space Agency, Paris, France, 2000,

a [reprint](#) of which is attached to this book.

Note 1.1 (page 29)

For actual calculation of the Morlet wavelet transform [1.37](#) from a time series sampled at finite sampling frequency, the integral can simply be replaced by its Riemann sum approximation. The result is still classified as a continuous wavelet transform (CWT), which usually is evaluated at more discrete values of τ and f than actually allowed by the time-frequency resolution of Panel (d) in Figure [1.2](#). There is at least one obvious reason for this practice: plot routines for representing a function of two variables (τ and f) often require the function to be evaluated on a uniform rectangular grid (like Panels (a)–(c) rather than (d) in Figure [1.2](#)). The term discrete wavelet transform (DWT) is usually reserved for the coefficients describing the expansion of the time series in a set of discrete of functions obtained by stretching and translation of the mother wavelet by factors of two, thus directly corresponding to Panel (d) in Figure [1.2](#) (Najmi, A.-H. and Sadowsky, J., The continuous wavelet transform and variable resolution time-frequency analysis, *Johns Hopkins APL Technical Digest*, **18**, 134–139, 1997).

Note 2.1 (page 43)

Techniques for spectral and cross-correlation estimators based on irregularly sampled and/or unsynchronised data do exist, but we do not consider them here. The calculation of gradients, and also time series of mixed parameters (i.e., those calculated using data from more than one time series) necessitate resampling the original data.

Note 2.2 (page 48)

In the paper edition, the words “increased” and “reduced” were interchanged in points 1 and 2, since they were meant to refer to the time resolution. An increased sample interval corresponds to reduced time resolution, and vice versa. The words now correctly refer to the sample interval.

Note 6.1 (page 138)

For a more recent example, see Fazakerley, A. and Szita, S., Multi-point science with the

Cluster plasma electron data, in *Proceedings of the Cluster-II Workshop Multiscale/Multipoint Plasma Measurements, Imperial College, London, 22–24 September 1999*, ESA SP-449, pp. 155–161, European Space Agency, Paris, France, 2000, especially the figure on page 159 of that work.

Note 7.1 (page 169)

In the paper edition, equation 7.22, was erroneously given as:

$$f(\mathbf{r}, v) = f_o(\mathbf{r}, v) \left[1 + \hat{\mathbf{v}} \cdot \boldsymbol{\varepsilon}_o + \mathbf{R}_g \cdot \nabla \ln f_o(\mathbf{r}, v) - \mathbf{V} \cdot \hat{\mathbf{v}} \frac{\partial f_o(\mathbf{r}, v)}{\partial v} + \mathcal{O} \left(S, \frac{R_g}{L} \varepsilon_o, \frac{V}{v} \varepsilon_o, \left(\frac{R_g}{L} \varepsilon_o \right)^2, \frac{R_g}{L} \frac{V}{v}, \left(\frac{V}{v} \right)^2 \right) \right]$$

i.e., the last partial derivative on the first line was missing a \ln function.

Note 8.1 (page 186)

Because of the assumption of one-dimensionality, i.e., $\partial/\partial x = 0 = \partial/\partial y$, Faraday's law becomes:

$$\partial B_z / \partial t = -(\nabla \times \mathbf{E})_z = \partial E_x / \partial y - \partial E_y / \partial x = 0$$

Note 8.2 (page 188)

Note that the variance of the actual measured data set is used rather than the variance estimate for an ensemble, of which the measured set is a member. Such an estimate would be larger by the factor $M/(M-1)$ and so would the eigenvalues of the variance matrix.

Note 8.3 (page 198)

For MVAB with constraint $\langle \mathbf{B} \rangle \cdot \hat{\mathbf{n}} = 0$ (Section 8.2.6), the angular error (radians) for rotation of $\hat{\mathbf{n}} \equiv \mathbf{x}_{\min}$ in the plane perpendicular to $\langle \mathbf{B} \rangle$ is

$$\pm \left\{ \lambda_{\max} \lambda_{\min} / (M-1) (\lambda_{\max} - \lambda_{\min})^2 \right\}^{1/2}$$

while for rotation about \mathbf{x}_{\max} it is

$$\pm \left\{ \langle \mathbf{B} \cdot \mathbf{x}_{\min} \rangle^2 / (M-1) |\langle \mathbf{B} \rangle|^2 \right\}^{1/2}$$

\mathbf{x}_{\max} and \mathbf{x}_{\min} being the eigenvectors corresponding to λ_{\max} and λ_{\min} (in variance space, the major and minor axes of the shaded ellipse in Figure 8.5).

Note 8.4 (page 200)

Kawano and Higuchi [private communication] have pointed out that only the fluctuating part (the noise) superimposed on the underlying signal but not the signal itself, should be resampled in the bootstrap technique so that each simulated realisation of the data is a sum of the resampled noise and the deterministic signal. However, for $\lambda_3 \ll \lambda_2$ (as is the case for the event studied in this chapter), it can be shown that the bootstrap error estimate for $\hat{\mathbf{n}}$, but not for $\langle \mathbf{B} \rangle \cdot \mathbf{x}_3$, remains accurate even if the noise is not separated from the

signal. To properly obtain the uncertainty in $\langle \mathbf{B} \rangle \cdot \mathbf{x}_3$ under the condition $\lambda_3 \ll \lambda_2$, it is sufficient to resample only the fluctuations in B_n , because the (unknown) fluctuations in the systematically varying tangential field components produce a negligible effect. With the resampling performed in this manner, the numerical value of the bootstrap uncertainty in $\langle \mathbf{B} \rangle \cdot \mathbf{x}_3$ agrees with the analytical estimate in equation 8.24 (i.e., ± 1.63 nT for the event studied).

Note 8.5 (page 207)

A. V. Khrabrov [private communication] has shown that the MVA ρv determination of $\hat{\mathbf{n}}$ and u_n can be formulated as an eigenvalue problem in a manner similar to that described by *Khrabrov and Sonnerup* [1998] for the case of the electric field analysis, thereby avoiding a cumbersome iterative process.

Khrabrov, A. V. and Sonnerup, B. U. Ö., Orientation and motion of current layers: Minimization of the Faraday residue, *Geophys. Res. Lett.*, **25**, 2372–2376, 1998.

Note 8.6 (page 216)

The constraint $\langle B_3 \rangle = 0$ should be corrected to read $\langle \mathbf{B} \rangle \cdot \hat{\mathbf{n}} = 0$. Note also that use of this constraint fails to provide an improved $\hat{\mathbf{n}}$ vector in the special case of a TD such that $\lambda_2 \simeq \lambda_3$ and $\langle B_2 \rangle \simeq \langle B_3 \rangle \simeq 0$.

Note 9.1 (page 227)

The data used to obtain numerical results in this chapter in fact exhibit small deviations from the values in Table 8.4. The latter data yield $\mathbf{V}_{\text{HT}} = (-122.6, -223.6, 75.6)$ km/s instead. Other numerical results in this chapter also change slightly.

Note 9.2 (page 241)

For the data in Table 8.4, the results are slightly different:

$$\begin{aligned} \mathbf{V}_{\text{HT0}} &= (-102.2, -204.3, 22.1) \text{ km/s} \\ \mathbf{a}_{\text{HT}} &= (-0.62, -0.53, 0.87) \text{ km/s}^2 \end{aligned}$$

Numerical values, given in the remainder of Section 9.5, also change slightly.

Note 9.3 (page 246)

When the current sheet is close to being a tangential discontinuity, as is the case for the event studied in this chapter, separation of the noise may not be necessary in determining the bootstrap errors for \mathbf{V}_{HT} .

Note 11.1 (page 296)

This issue has been discussed in detail by

Dunlop, M. W. and Woodward, T. I., Analysis of thick, non-linear boundaries using the discontinuity analyser, *Ann. Geophys.*, **17**, 984–995, 1999.

An [extract](#) from this article is attached to this book.

Note 14.1 (page 350)

The equations presented here may also be used to explain the simulation results of Section 16.3. See for example

Chanteur, G., Accuracy of field gradient estimations by Cluster: Explanation of its dependency upon elongation and planarity of the tetrahedron, in *Proceedings of the Cluster-II Workshop Multiscale/Multipoint Plasma Measurements, Imperial College, London, 22–24 September 1999*, ESA SP-449, pp. 265–268, European Space Agency, Paris, France, 2000,

a [reprint](#) of which is attached to this book.

Note 16.1 (page 405)

For a theoretical explanation of the simulation results, see the [article](#) referred to in Note 14.1.

Note 16.2 (page 408)

In the paper edition, the percentages quoted on page 408 and in Figures 16.7, 16.8, 16.11–16.14 were 10 times too high; i.e., they incorrectly ranged 0–100%, instead of 0–10%.

Note 16.3 (page 408)

Note that this error is comparable to the relative uncertainty $\Delta r/r = 1\%$.

Note 17.1 (page 432)

Errors affecting magnetic gradients result from both physical and geometrical uncertainties by contrast with errors affecting plasma gradients. Section 14.3 presents a general discussion of these two contributions. Beyond this difference, it is worth underlining that angular averaging leads to the appearance of the trace of the reciprocal tensor in the following equation 17.35 or 17.70 and equation 14 in the [article](#) referred to in Note 14.1.

Note 17.2 (page 432)

Assuming that electric field measurement errors are also considerably larger than spacecraft position inaccuracies (see Section 17.1.2), the error formulas for plasma moment derivatives should be applicable to electric field spatial derivatives as well. This way, error estimates of $\nabla \cdot \mathbf{E}$ and $\nabla \times \mathbf{E}$ can be obtained.

Note 17.3 (page 434)

The “omni-directional” error equations of Section 17.2.2, in particular 17.35, 17.36, and 17.39, can be comprised into a single symbolic formula as

$$\left\langle (\delta D_f)^2 \right\rangle = \frac{f}{3} \left\{ \sum_{\alpha} |k_{\alpha}|^2 \right\} \equiv \frac{f}{3} \text{Tr}(\mathbf{K})$$

Here, D_f denotes a derivative operator with f “degrees of freedom”, and the left-hand side denotes the expectation value of the associated mean square error averaged over all possible directions. Since the directional derivative acts in one direction only, $f = 1$ which results in equation 17.35. The curl operator is effectively two-dimensional (recall Stokes’ theorem), hence $f = 2$ which yields equation 17.36. The div operator acts in all three spatial dimensions, thus $f = 3$ which leads to equation 17.39. $\mathbf{K} = \sum_{\alpha} \mathbf{k}_{\alpha} \mathbf{k}_{\alpha}^T$ denotes the reciprocal tensor.

Note 17.4 (page 434)

In the paper edition, the reciprocal volumetric tensor was erroneously given as

$$\mathbf{R}^{-1} = \sum_{\alpha=0}^3 \mathbf{k}_{\alpha} \mathbf{k}_{\alpha}^T$$

i.e., the factor of 4 was missing. This also affects equation 17.41 and the last statement of this section about the trace of \mathbf{R}^{-1} .

Authors' Addresses

Note: Authors without an email address are no longer working at the address given.

Jörg Büchner

Max-Planck-Institut für Aeronomie
Max-Planck-Str. 2
37191 Katlenburg-Lindau, Germany
Email: buechner@linmpi.mpg.de

Gérard Chanteur

CETP/CNRS
10–12, avenue de l'Europe
78140 Vélizy, France
Email: gerard.chanteur@cept.ipsl.fr

Patrick W. Daly

Max-Planck-Institut für Aeronomie
Max-Planck-Str. 2
37191 Katlenburg-Lindau, Germany
Email: daly@linmpi.mpg.de

Malcolm W. Dunlop

Space and Atmospheric Physics Group
Physics Department
Imperial College of Science, Technology and Medicine
London, SW7 2BZ, United Kingdom
Email: m.dunlop@ic.ac.uk

Anders I. Eriksson

Swedish Institute of Space Physics
Uppsala Division
S-755 91 Uppsala, Sweden
Email: aie@irfu.se

Andrew N. Fazakerley

Mullard Space Science Laboratory
Holmbury St. Mary
Dorking, United Kingdom
Email: anf@mssl.ucl.ac.uk

Karl-Heinz Glassmeier

Institut für Geophysik und Meteorologie
Technische Universität Braunschweig
38106 Braunschweig, Germany
Email: khg@tu-bs.de

Christopher C. Harvey

Centre d'Etude Spatiale des Rayonnements
B.P. 4346
31029 Toulouse CEDEX, France
Dept. Recherche Spatiale
Email: harvey@cesr.fr

Alexander V. Khrabrov

Dartmouth College
Hanover, NH 03755, U.S.A.
Email: sasha@northstar.dartmouth.edu

Jens-Peter Kuska

Max-Planck-Institut für Aeronomie
Max-Planck-Str. 2
37191 Katlenburg-Lindau, Germany

Uwe Motschmann

Institut für Geophysik und Meteorologie
Technische Universität Braunschweig
38106 Braunschweig, Germany
Email: uwe@tu-bs.de

Götz Paschmann

Max-Planck-Institut für extraterrestrische Physik
Postfach 1312
85741 Garching, Germany
alternative address:
ISSI
Hallerstr. 6
3012 Bern, Switzerland
Email: gep@mpe.mpg.de

Jean-Louis Pinçon

LPCE/CNRS
3A, avenue de la Recherche Scientifique
45071 Orléans CEDEX, France
Email: jlpincon@cnrs-orleans.fr

Patrick Robert

CETP/CNRS
10–12, avenue de l'Europe
78140 Vélizy, France
Email: patrick.robert@cept.ipsl.fr

Alain Roux
CETP/CNRS
10–12, avenue de l'Europe
78140 Vélizy, France
Email: alain.roux@cept.ipsl.fr

Maureen Scheible
Dartmouth College
Hanover, NH 03755, U.S.A.

Steven J. Schwartz
Astronomy Unit
Queen Mary and Westfield College
London, E1 4NS, United Kingdom
Email: s.j.schwartz@qmw.ac.uk

Bengt U. Ö. Sonnerup
Thayer School of Engineering
Dartmouth College
Hanover, NH 03755, U.S.A.
Email: sonnerup@dartmouth.edu

Joachim Vogt
Institut für Geophysik und Meteorologie
Technische Universität Braunschweig
38106 Braunschweig, Germany
Email: j.vogt@tu-bs.de

Heinz Wiechen
Max-Planck-Institut für Aeronomie
Max-Planck-Str. 2
37191 Katlenburg-Lindau, Germany

Timothy I. Woodward
Space and Atmospheric Physics Group
Physics Department
Imperial College of Science, Technology and
Medicine
London, SW7 2BZ, United Kingdom
Email: t.woodward@ic.ac.uk

Index

- AMPTE, 272, 297
 - Barium release, 297
 - electron data, 131
 - ion data, 132, 137
 - magnetopause, 189, 207, 212, 226, 284
 - benchmark data, 218
 - on-board moment computations, 135
 - plasma data, 129
- anisotropies
 - particle, 165
- anti-aliasing filter, 50
- autoregressive method, 37
- Bartlett window, 18
- barycentric coordinates, 349, 395, 397, 430, 437, 445
 - definition, 350
 - properties, 351–352
- boundary
 - curvature, 294
 - 2-D, 291
 - non-planar models, 302
 - normals
 - rotational discontinuity, 263
 - shocks, 253
 - tangential discontinuities, 263
 - variance analysis, 185–206, 276
 - planar model, 302
 - planarity, 278
 - remote sensing
 - absorbing boundary, 175
 - finite gyroradius, 165
 - source boundary, 175
 - velocity filter, 164
- bow shock
 - crossing, 133
 - electron mapping, 175
 - electrons, 164
 - foreshock, 165
 - heat flux, 150
 - ions, 151
 - kinetic temperature at, 149
 - models, 259, 260, 263, 267
 - particle entry, 105
 - remote sensing, 177
- boxcar averaging, 47, 53
- bulk velocity, 146
- Chapman-Ferraro currents, 392
- Cluster, 43, 135, 307, 359, 453
 - CIS instrument, 123, 133, 138
 - FGM instrument, 417
 - field data, 70
 - geometric factors, 335, 337
 - magnetic field errors, 398
 - objectives, 395
 - orbit, 325, 336, 375, 399
 - orbital deformation, 360
 - PEACE instrument, 123, 133
 - RAPID instrument, 123, 453
 - separations, 71
 - spin period, 324
 - STAFF experiment, 32
 - tetrahedron, 391, 396
 - tetrahedron evolution, 338, 339
 - WEC instruments, 44
- concatenation of data sets, 46
- configuration, *see* tetrahedron geometric factors
- cross-spectral density, 32
- curlometer, 272, 396
- current density, *see* density
- current structure models, 405
- curved plate analyser, 92, 99
- cusp ions, 178
- data gap, 45, 57
- degeneracy, in variance analysis, 192
- deHoffmann-Teller frame, 221, 251, 265, 277, 462
- density
 - charge, 150
 - current, 150
 - accuracy, 404–416
 - mass, 146

- number, 146
- discontinuity, 249
 - analysis, 271
 - normal velocity and acceleration, 365–368
 - planar, 273
 - rotational, 156, 206, 222, 229, 253, 263, 284
 - tangential, 156, 186, 190, 195, 206, 217, 222, 225, 263, 266
 - tangential stress balance, 151
 - variance analysis, 196, 310–312
- discontinuity analyser, 272
 - boundary properties, 275
 - crossing times, 275
 - data pre-selection, 274
 - hierarchically-based methodology, 274
 - motional properties, 279
 - multi-spacecraft, 287
 - non-planar, 285–287
 - normal stability, 275
 - planar, 273
 - self-consistent normal analysis, 298
- efficiency, particle detector, 116, 135
- electrostatic analyser, *see* plasma measurement
- elongation, *see* tetrahedron geometric factors
- energetic particle spectra
 - test particle calculations, 453
- Equator-S, 135
- errors
 - absolute, 440
 - estimates
 - analytical, 196, 230, 234
 - angular, 201
 - bootstrap, 198, 233
 - formulas
 - approximate evaluation, 435
 - curl estimation, 425, 426, 428, 433
 - directional derivatives, 423, 432
 - divergence estimation, 424, 433
 - partial derivatives, 421, 433
 - gradients
 - physical and geometrical, *see* statistical properties
 - truncation, *see* truncation errors
 - in boundary velocity, 283
 - in current density, 397
 - magnetic field measurements, 422
 - mean square, 421, 439
 - orbit determination, 391
 - primary, 438, 439
 - relative, 440
 - sources, 421
 - calibration, 438
 - data noise, 283
 - orbit determination, 391, 438
 - plasma moments, 130–135
 - relative importance, 422
 - tetrahedron shape, 407
 - variance analysis, 196
 - variance ellipsoid, 191
- European Simulation Network, 450, 469
- FFT, *see* Fourier analysis
- filtering technique, 65
 - filter bank, 67
 - for gradients, 389
 - k filtering, 83
 - variance analysis, 204
- finite gyroradius effects, 165
- flux transfer event, 178, 212, 223, 459
- Fokker-Planck simulations, 469
- foreshock, 152, 165
 - electron mapping, 176
- Fourier analysis, 15–25
 - continuous transform, 29
 - discrete transform, 8
 - fast transform (FFT), 15, 31
 - filtering, 66
 - integral, 6
 - Parseval's relation, 6
 - series, 6, 79
 - short-time transform, 10
 - wave decomposition, 80
 - Welch method, 24
 - windowed transform, 10
 - with wavelets, 27
- Freja satellite
 - electric field data, 29

magnetic field data, 19
fuzzy join, 47

Gaussian window, 18
geometric factors, *see* tetrahedron geometric factors
geometry factor, particle detector, 113, 135

Geotail, 453

gradient, 371
definition
scalar field, 350
vector field, 350
error of linear determination, 373
evaluation using simulated data, 375–389
identity of least squares and barycentric methods, 372
least squares method, 313
linear barycentric estimators
definitions, 352
of a solenoidal vector, 314
of a vector, 313
of dipole field, 374
or discontinuity, 372
least squares method, 310
pre-filtering of the data, 389
quadratic barycentric estimators
definitions, 354
gradient scale length, 421
guiding centre simulations, 469
gyromotion, 167
gyrotropic symmetry, 119
gyroviscosity, 148, 155

Hamming window, 18

Hann window, 18

heat flux vector, 149

heterogeneity factor, 412

hybrid simulations, 469

inter-spacecraft calibration, 440

Interball, 123, 142, 453, 460

ISEE, 272, 307

FTE crossing, 178

plasma experiment, 128

joining data sets, 46

jump relations, 156, 258

k filtering, 83

Kelvin-Helmholtz instability, 459

kinetic plasma simulations, 468

kinetic temperature, 148

linear estimators, 309, 319, 352, 357, 361, 364, 371, 373, 374

linear interpolation, 47, 56, 350–353, 389, 392

Liouville mapping
inaccessible regions, 162
known fields, 161
test particles, 451
unknown fields, 163

Liouville's Theorem, 160

magnetic moment, 160

magnetic reconnection, *see* magnetopause

magnetopause
AMPTE crossing, 189, 207, 212, 226, 284
crossing, 10
current density, 150
heat flux, 149
MHD methods, 459
models, 259, 260
motion, 243
normal, 210, 227, 241, 462
particle entry, 165, 450
reconnection, 156, 178, 229, 459, 474
remote sensing, 173, 175
surface waves, 212
tangential discontinuity, 196, 230
travelling bulge, 231

magnetosphere
open or closed, 156

maximum variance analysis, *see* variance analysis

merging data sets, 46

mesocentre, 310

metadata, 58

MHD

ideal fluid, 154

- modelling, 457
- tail dynamics, 462
- wave mode analysis, 458
- wave mode decomposition, 79
- minimum variance analysis, *see* variance analysis
- mode decomposition of MHD waves
 - method, 79
 - testing, 82
- modelling, 449–477
 - MHD, 457
 - test particle calculations, 451
- momentum flux, 147
- nonlinear variations, 435
- normals
 - boundary, *see* boundary
 - discontinuities, *see* discontinuities
 - shocks, *see* shocks
- notation conventions, 4
- numerical modelling and simulation, *see* modelling
- Nyquist frequency, 48
- orbit, error of determination, 391
- Parseval's relation, 6
- particle kinetics, 159
- phase spectrum, 7
- pitch angle distribution, 119
- planarity, *see* tetrahedron geometric factors
- plasma measurement
 - 3-D distribution, 120
 - calibration, 116
 - curved plate analyser, 92, 99
 - mass per charge, 105
 - pitch angle, 119
 - reduced 3-D distribution, 121
 - retarding potential analyser, 99
 - time of flight analysis, 106
 - “Top Hat” analyser, 92
 - velocity distribution, 92
 - velocity space filter, 92, 113
- plasma moments
 - accuracy, 91, 128
 - one-fluid, 126, 150
 - single-species, 125, 146
- power spectral density, 7
- pressure tensor, 147, 150
- principle axes, *see* tetrahedron geometric factors
- Q_{GM} parameter, 326, 340
 - calculation, 344
- Q_{R8} parameter, 326, 343
- Q_{RR} parameter, 326, 340
 - calculation, 344
- Q_{SR} parameter, 326, 343
- quadratic estimators, 354, 364, 378, 389
- quadratic interpolation, 353–355
- Rankine-Hugoniot relations, 156, 249, 277
- reciprocal tensor, 372
 - definition, 356
- reciprocal vectors
 - definition, 351
 - properties, 351–352
- reconnection, *see* magnetopause
- remote sensing, 164
 - boundaries, 172
 - bow shock, 177
 - finite gyroradius, 165
 - flux transfer event, 178
- resampling of data sets, 46
- retarding potential analyser, 99
- sample duration, 45
- sample interval, 45
- sampling factor, 45, 50
- shocks, 156, 186, 249
 - coplanarity plane, 252
 - deHoffmann-Teller frame, 251
 - geometry, 250
 - Mach number, 253
 - Normal Incidence Frame, 251
- normals
 - coplanarity methods, 254
 - empirical models, 259
 - jump condition method, 258
 - multi-spacecraft timings, 256
 - single spacecraft methods, 254
 - variance analyses, 254
- Rankine-Hugoniot relations, 249

-
- speed determination, 263
 - simulations
 - Fokker-Planck, 469
 - guiding centre, 469
 - hybrid, 469
 - kinetic plasma, 468
 - fully self-consistent, 470
 - solenoidal condition, 276, 319, 371, 380
 - spatial aliasing, 69, 364–365
 - spatial gradient, *see* gradient
 - spatial interpolation, 371–391
 - linear, 350–353
 - quadratic, 353–355
 - spectral density matrix, 82, 83, 86
 - spectral energy density, 66–68
 - spectral power density, 7, 85
 - spherical harmonic analysis, 152
 - spline interpolation, cubic, 47, 56
 - stationarity, 279
 - deHoffmann-Teller analysis, 236
 - variance analysis, 202
 - statistical properties
 - linear barycentric estimators, 357–358
 - reciprocal vectors, 355–356
 - status data, 58
 - surface wave detector, electromagnetic, 75
 - tail dynamics by MHD, 462
 - temperature tensor, 148, 150
 - test particle calculations, 451
 - energetic particle spectra, 453
 - non-adiabatic, 452
 - tetrahedron geometric factors, 317, 330, 433
 - E-P* diagram, 338
 - calculation, 344
 - Cluster orbit, 335, 337
 - elongation and planarity, 328, 404
 - inertia tensor, 316
 - principle axes, 316, 327, 328
 - Q_{GM} parameter, 326, 340
 - Q_{R8} parameter, 326, 343
 - Q_{RR} parameter, 326, 340
 - Q_{SR} parameter, 326, 343
 - reciprocal tensor, 372
 - reciprocal vectors, 351
 - volumetric tensor, 315, 327, 404
 - tetrahedron types, 330, 408
 - thin boundary, 271
 - time averaging, 434
 - time derivatives, 423, 432
 - time series analysis, 43
 - time tag, 44
 - time-frequency analysis, 10
 - time-of-flight analysis, 106
 - “Top Hat” analyser, 92
 - trends in data signals, 15
 - truncation errors
 - definition, 353
 - thick current sheet, 360–364
 - Tsyganenko-87 field model, 375, 380, 389, 392, 398, 415
 - variance analysis, 185–218
 - degeneracy, 192
 - electric field, 209
 - for pre-selection, 275
 - magnetic field, 187, 277
 - magnetopause normal, 462
 - mass flux, 206
 - mode decomposition, 83
 - stationarity, 202
 - variance ellipsoid, 191
 - velocity dispersion, 164
 - velocity filter, 164
 - volumetric tensor, *see* tetrahedron geometric factors
 - and error formulas, 434
 - and spatial gradients, 315
 - vorticity, 425
 - Walén relation, 229, 245, 462
 - wave mode analysis, 458
 - wavelet techniques, 25–32
 - continuous transform, 28
 - Morlet wavelet transform, 27
 - with Fourier methods, 27
 - weighted averaging, 48
 - Welch method, 24

Extract from:

Dunlop, M. W. and Woodward, T. I., Analysis of thick, non-linear boundaries using the discontinuity analyser, *Ann. Geophys.*, **17**, 984–995, 1999.

Analysis of thick, non-planar boundaries using the discontinuity analyser

M. W. Dunlop and T. I. Woodward

Space Physics Group, Physics Department, Imperial College of Science, Technology and Medicine, London, SW7 2BZ, UK.

ABSTRACT

The advent of missions comprised of phased arrays of spacecraft, with separation distances ranging down to at least meso-scales, provides the scientific community with an opportunity to accurately analyse the spatial and temporal dependencies of structures in space plasmas. Exploitation of the multi-point data sets, giving vastly more information than in previous missions, thereby allows unique study of their small scale physics. It remains an outstanding problem, however, to understand in what way comparative information across spacecraft is best built into any analysis of the combined data. Different investigations appear to demand different methods of data co-ordination. Of the various multi-spacecraft data analysis techniques developed to affect this exploitation, the so-called discontinuity analyser has been designed to investigate the macroscopic properties (topology and motion) of boundaries, revealed by multi-spacecraft magnetometer data, where the possibility of at least mesoscale structure is considered. It has been found that the analysis of planar structures is more straightforward than the analysis of non-planar boundaries, where the effects of topology and motion become interwoven in the data, and we argue here that it becomes necessary to customise the analysis for non-planar events to the type of structure at hand. One issue central to the discontinuity analyser, for instance, is the calculation of normal vectors to the structure. In the case of planar and ‘thin’ non-planar structures, the method of normal determination is well-defined, although subject to uncertainties arising from unwanted signatures. In the case of ‘thick’, non-planar structures, however, the method of determination becomes particularly sensitive to the type of physical sampling that is present. It is the purpose of this article to firstly review the discontinuity analyser technique and secondly, to discuss the analysis of the normals to thick non-planar structures detected in magnetometer data.

1. Introduction

Knowledge of the motion and size of structures in space plasmas provides vital clues as to their physics (such as composition, structure, relationship to the local and/or global plasma conditions). For example, the determination of motional parameters for a non-dispersive structure permits the separation of spatial and temporal variations, which *in situ* produce mixed effects in time series data signatures. Single spacecraft measurements are generally insufficient for the unambiguous separation of spatial and temporal dependencies, however, except under rather extraordinary circumstances [e.g. *Fairfield and Feldman*, 1975]. Measurements made by two or more satellites, which happen to be appropriately or fortuitously positioned, can help to unravel this mixed spatio-temporal behaviour, often through simple timing studies. Indeed several unplanned conjunctions of spacecraft, as well as planned, dual satellite missions (such as the ISEE and AMPTE missions) have resulted in numerous studies whose important conclusions hinge on establishing such knowledge of the motion of some underlying structure. For example, Greenstadt and co-workers used multi-spacecraft data sets to analyse the terrestrial bow shock and propose formation mechanisms based on the thicknesses they determined [e.g. *Greenstadt et al.*, 1975 and *Morse and Greenstadt*, 1976]. During the ISEE-1 and -2 era, dual spacecraft data sets were again used to separate both spatial and temporal behaviour [e.g. *Russell and Greenstadt*, 1979; *Bame et al.*, 1979; *Russell et al.*, 1982 and *Mellot and Greenstadt*, 1984]. Non-planar boundary surface topology, thickness of the boundary layer, and motion has been of interest to investigators in, for example, the solar wind where interplanetary shocks have been studied [e.g. *Chao and Lepping*, 1974; and *Lepping and Chao*, 1976]. Another topic where such macroscopic properties have been sought are magnetopause surface waves: for example, *Song et al.* [1988] used simultaneous data from the closely separated ISEE-1 and -2 spacecraft to determine the amplitude, wavelength and period of these waves.

A major objective of the European Space Agency's four spacecraft Cluster mission is to affect this spatio-temporal separation on a routine basis and on spatial scales down to, at least, mesoscale variations. The above studies used data measured during the fortuitous conjunctions of several satellites, separated by distances large compared to the typical Cluster separations [*Dunlop et al.*, 1990]. With the mission a reality again for launch in the year 2000, further investigation in this area is of paramount significance to the fruitful exploitation of what would be a unique data set. Clearly, such a phased array of at least four spacecraft is generally sufficient to resolve 3-D information. Nevertheless, properties of highly time dependant structures, or those which vary on spatial scales which are small compared to the array size, still present obvious problems for modeling, empirically, observed behaviour. Under certain assumptions, for example those of stationarity or comparative spatial scales, some particular analysis may be possible. But a fundamental problem always exists: how should the comparative information between spacecraft be built into any analysis? The determination of different properties demands different methods of co-ordination [*Dunlop et al.* 1988]; often having a large overlap between them, with parameters in common. In terms of boundary analysis, although there is an apparent distinction between macroscopic parameters (such as motion, orientation and form) and the boundary structure itself (as sampled individually by each spacecraft), it isn't always *a priori* obvious how spacecraft differences are to be combined with time series information, except where the boundary is planar or 'thin' [*Dunlop and Woodward*, 1998]. Both sources of information may in principle provide independent knowledge of the boundary normals, for instance, but depend on these key boundary properties. The effect of different physical structure in the boundary can therefore provide a context for the time series analysis and needs to be known at some level of detail, depending upon behaviour. The general problem posed in extending the analysis to multi-spacecraft data in curved geometry is how to combine the information self-consistently when the single spacecraft assumptions on the macroscopic behaviour are relaxed.

Dunlop and Woodward [1998] describe in detail a multi-spacecraft magnetometer data analysis technique, which they call the *Discontinuity Analyser* (DA), designed to determine the structure, orientation and motion of boundaries using calculated normal vectors at each spacecraft to the structure as well as the associated times of encounter with the structure. Their technique is focused on the use of magnetometer data only, primarily for data resolution considerations. The method is applicable to thin, non-dispersive, simple planar and non-planar structures detected by a small array of closely separated spacecraft. Their requirement of "thin" structures, where the thickness of the structure is far less than its local radius of curvature, is fundamental to the design of their method, since a unique normal direction at any point on a curved surface is only well-defined when the structure is thin. The requirement is not so critical where the boundary is planar, however. Clearly, not all structures detected in space may fit this requirement on their thickness. Nevertheless, this problem is generally present in other treatments of boundaries, and relates to the issue of how to deal with comparative information, raised above.

The purpose of this paper is twofold: firstly, to summarize, in a practical way, the discontinuity analyser; and secondly, to discuss the issues for such analysis of thick, non-planar boundaries. A pivotal concept of the non-planar technique is the degree of certainty one has in the normal vectors determined at each spacecraft. The subject of normal determination in the presence of thick boundaries is the topic of the next section, where we discuss the variation of the normal determined using minimum variance analysis [*Sonnerup and Cahill*, 1967] for a curved, thick, 2-D model boundary under various conditions of sampling, namely data interval length, symmetry about the centre of the event, data resolution and spacecraft trajectory. A similar theme of analysis, appropriate to real spacecraft data, is presented for a barium release recorded in the magnetometer data of the

AMPTE-IRM/UKS spacecraft pair in section 3, confirming the model analysis. In the last section, we conclude with a discussion of the DA and the ramifications on it of the thick boundary normal analysis.

2. Thick, non-planar boundaries

In the presence of curved field geometry, normal determination is a subtle issue, particularly when the boundary is not thin. Systematic uncertainties on the normals arise, and these disrupt the delicate balance of information indicated in Table I. This then questions the validity of performing a surface analysis, since uncertainties in the normals imply uncertainties in both the topology and the acceleration of the structure. In contrast to the discrete methodology, *Dunlop and Woodward* [1998] introduced the concept of the combined approach which recognises that independent determination of the normals can be compromised. This methodology is based on the fully self-consistent determination of the surface topology and motion of the structure past the spacecraft array, cast in the framework of an inversion problem. They also discussed a compromise methodology between that of the discrete analysis and that of the combined approach. In this they suggested the explicit use of single-spacecraft determined normals weighted by confidence factors. In the following section we discuss normal determination in the presence of thick, non-planar boundaries, first presenting the conceptual scenario and then illustrating our ideas with an analysis of a model 2-D structure. This is followed up in section 3 with the analysis of an event recorded in the dual spacecraft AMPTE-IRM/UKS mission.

2.1. Normal determination

In single spacecraft analysis, normal determination is affected by the direction of motion through the boundary of the spacecraft (particularly if the curvature is high) as well as the nature of the data interval used in the normal analysis (Figure 3). Strong curvature arises when the thickness of the boundary surface is not negligible in comparison to the radius of curvature of the surface. In this case normal determination becomes difficult, as a result of the need to identify a finite data interval (which maps to a distance along the motion) through the boundary in order to perform a boundary normal analysis. Unless this distance is small on the scale of the curvature, the field structure will not wholly represent the boundary orientation local to the crossing, thus affecting the implied normal direction (as indicated in Figure 3). Boundary normal estimates will produce some average normal which will be sensitive to the interval chosen.

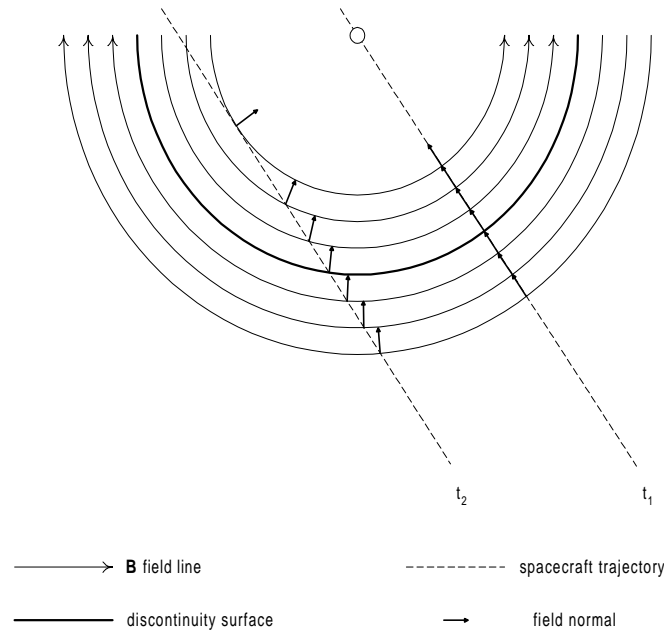


Figure 3: Diagrammatic illustration of the effect of thick, curved boundaries on normal determination. Shown is a circular discontinuity (thick curve), with two spacecraft trajectories (t_1 and t_2) and normal vectors (short thick arrows) to the field lines (thin curves with arrowheads indicating field direction) along each trajectory. Taken from *Dunlop and Woodward* [1998].

The individual normals are therefore dependent on the spacecraft trajectory, with the dependence being related to the *a priori* unknown form of the surface. Any result of surface fitting cannot therefore be fully de-coupled from the determination of normals. This raises the question of whether or not a full curvature analysis can be done quantitatively unless curvature is not significant on the scale of the boundary layer. Nevertheless, normals determined at two spacecraft, which deviate by a direction well in excess of that implied by the uncertainties, do indicate curved field geometry. For real events, however, it is not known,

a priori, what weighting should be given to the errors in the individual normals, in order to then apply some optimisation of the surface fit. One can either attempt to refine the uncertainties using the implied curvature (iterative process) and proceed with the discrete analysis, or feed the data into the combined methodology, knowing that real curvature exists. We explore the applicability of the former route below.

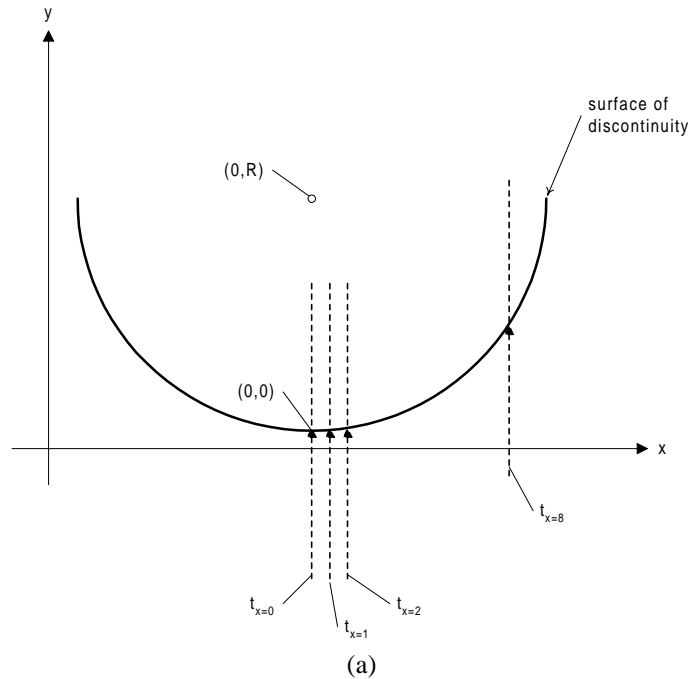
2.2. Data analysis

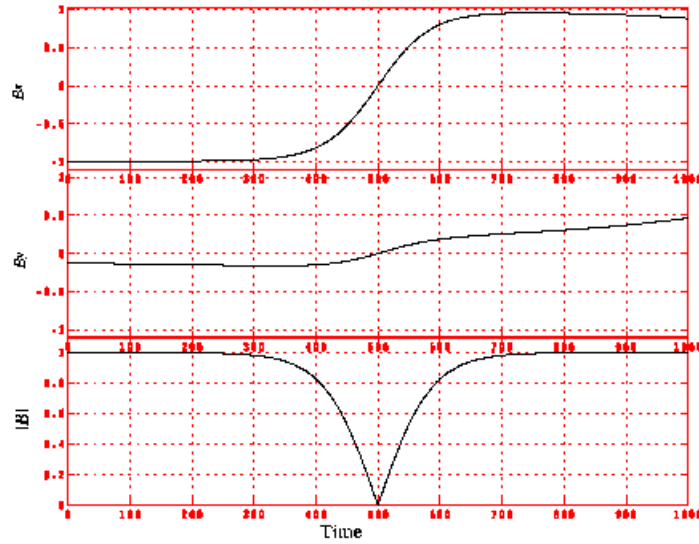
In order to quantitatively illustrate the variation of the normal along sampling trajectories through curved boundaries, we make use of a simple 2-D model, performing several fly-throughs or sampling trajectories and analyse the data thus recorded in various ways. The 2-D model we use is shown in the upper panel of Figure 4, together with four sampling trajectories labelled $t_{x=0,1,2,8}$. In the lower panel, typical magnetic field data signatures for the samplings are indicated.

We highlight the effects of data interval length, symmetry about the centre of the boundary, data resolution and spacecraft trajectory on the calculation of boundary normals using MVA (hereafter referred to as MVA normals). Since we have defined a simple model, theoretical normals are readily calculated in order to assess the accuracy of the MVA normals. Similar analyses have also been carried out with other curved 2-D models, obtaining qualitatively similar results.

Figure 4 illustrates the relative positions of 4 sampling trajectories across a 2-D circular discontinuity surface with a radius of curvature of 10 units (part a), as well as a typical data signature (part b). The first trajectory $t_{x=0}$ traverses along the y-axis, while the others, $t_{x=1,2,8}$, traverse parallel to the y-axis along the appropriate x-value. We shall compare and contrast the above-described effects on the MVA normals using these 4 trajectories. Firstly, consider the 4 trajectories in relation to the point symmetry of this particular model. Trajectory $t_{x=0}$ passes along a radial direction, while the other trajectories are increasingly inclined to this characteristic direction. Three characteristic runs were performed: one, where the data interval used is nested about the crossing, successively increasing its size; one, where the interval is shifted relative to the crossing time, successively changing the ratio of data intervals before and after the crossing, and another, where the effects of data resolution were explored.

Figure 5, which is a plot of the angular deviation of the MVA normals from the theoretical ones versus the relative length of the data interval over which the MVA normal is calculated, illustrates the effect of nesting. Note that we do not calculate the sense, clockwise or anti-clockwise, of the angular deviations. It is seen that the effect is more marked the greater the inclination of the trajectory is to the characteristic direction, is greater for longer data intervals, and can be of the order of several degrees. As the interval approaches a thin sheath around the centre of the discontinuity, or if $t_{x=0}$ is chosen, the effect is minimal. This is because for longer and more inclined trajectories the spacecraft samples magnetic fields of very different directions over the analysis interval, and hence of different normal directions. However, there is a slight anomaly from zero deviation for all trajectories, except $t_{x=0}$, at the smallest data interval, as well as a minimum deviation which is not at the shortest interval. This is explained below when we consider the asymmetry of the data interval about the discontinuity centre.





(b)

Figure 4: Schematic drawing of sampling trajectories through a circular discontinuity model and the associated data (time and magnetic field values in arbitrary units). The magnetic field model is defined as follows:

$$\mathbf{b} = \nabla \varphi \times \hat{\mathbf{z}} \quad ; \quad \varphi \equiv B_{0l} \log \left[2 \cosh \left(\frac{g(x, y)}{\sigma} \right) \right] \quad ; \quad g(x, y) \equiv R - \sqrt{(x^2 + (y - R)^2)} \quad ; \quad y < R$$

where 2σ is the boundary thickness, B_{0l} is a constant with units of length times magnetic field and R is the radius of curvature of the discontinuity surface. The field line topology is indicated in Figure 3. In this model the strength of the field decreases to zero on approaching the boundary from one side, before flipping direction and increasing its strength on the other side.

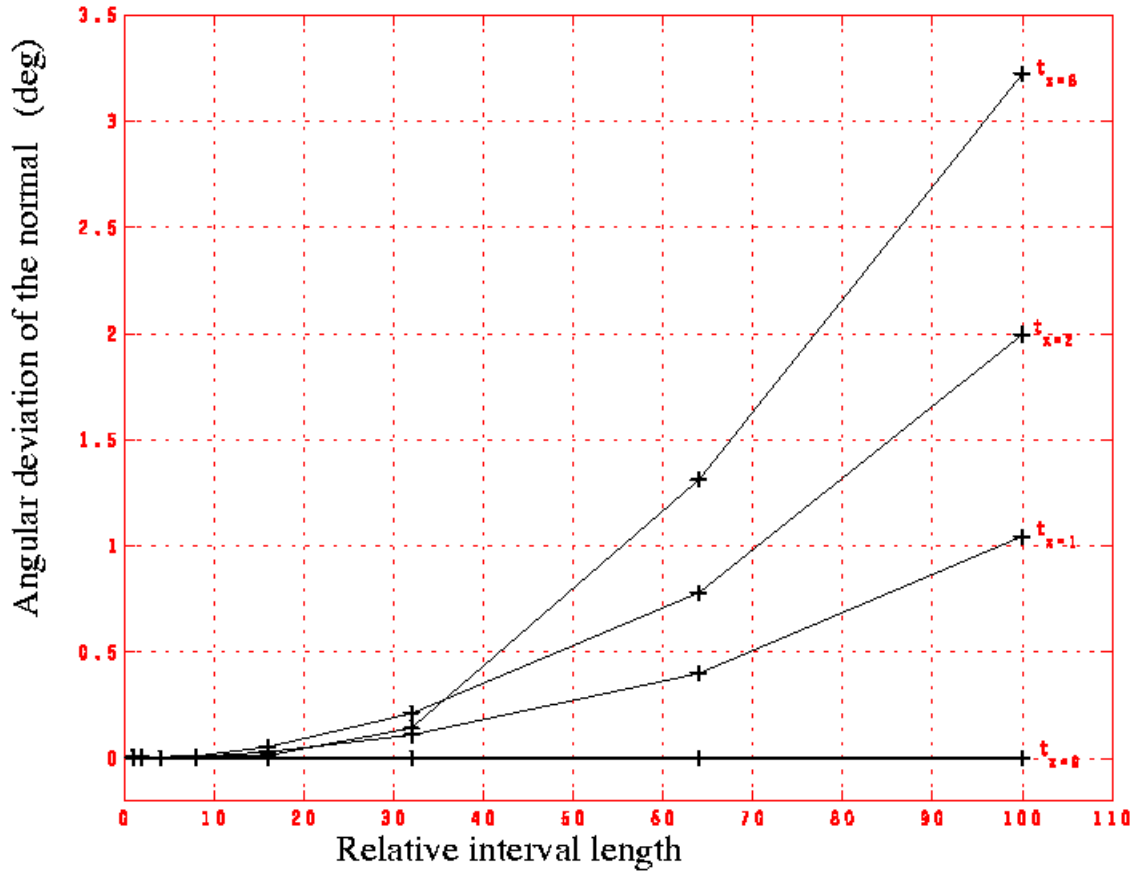


Figure 5: Sensitivity of MVA analysis to the model shown in Figure 5 in terms of interval nesting.

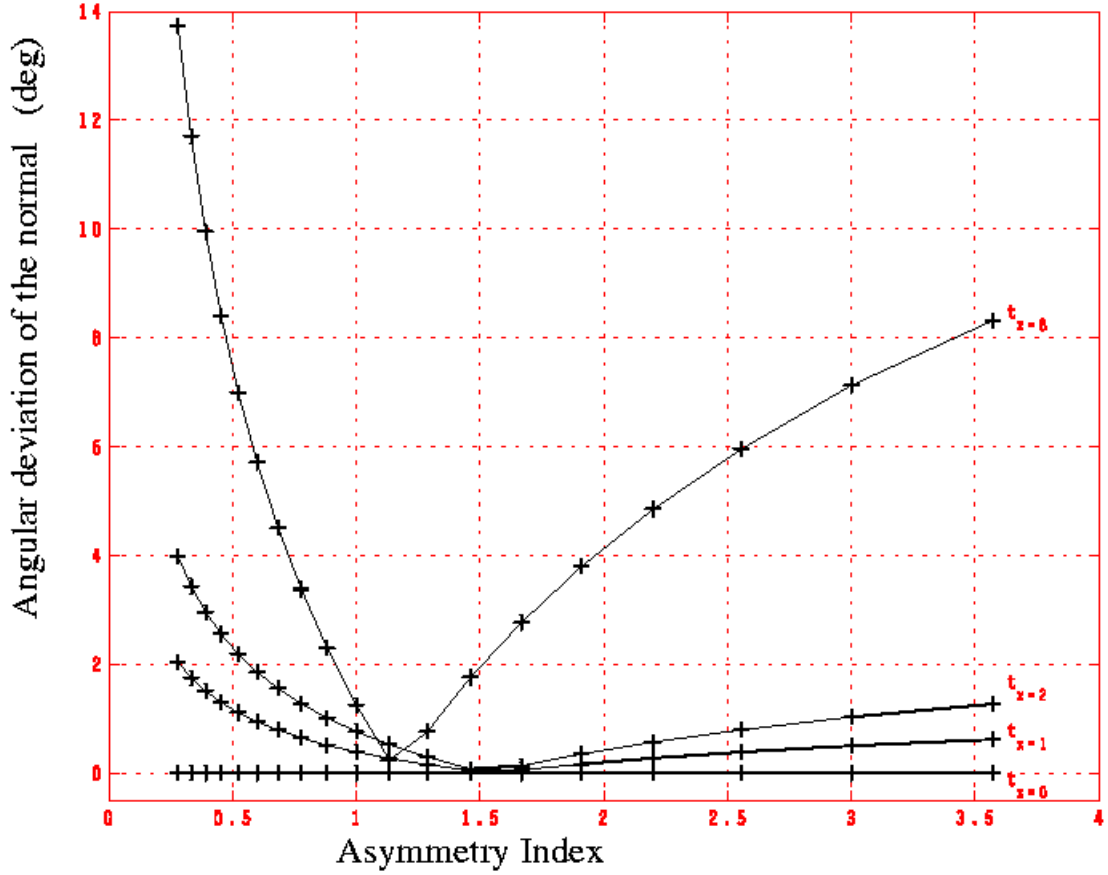


Figure 6: Sensitivity of MVA analysis to the model shown in Figure 5 in terms of interval asymmetry.

In order to study the effect of data interval asymmetry, we sample along all the trajectories, $t_{x=0,1,2,8}$, with some resolution (e.g. 1000 points for the whole data interval). Then taking a fixed length of data (e.g. 640 points) shorter than the whole interval sampled and containing the centre of the discontinuity somewhere inside, we determine the asymmetry of the chosen sub-interval and perform MVA on it. This is repeated for several asymmetries and for each trajectory. We define an *asymmetry index* of the selected data interval to be the ratio of the number of data points recorded before the centre of the discontinuity is encountered to that measured after encounter, with all these points within in the chosen interval length (e.g. 640 points). Thus symmetry indices can only be positive. Unit symmetry index implies that there are an equal number of points on either side of the discontinuity centre: i.e. a perfectly symmetric interval.

Figure 6 shows the dependence of the deviation of the MVA normal with data interval symmetry index. Again, the effect is greater the higher the inclination of the trajectory. For highly inclined trajectories, there is more than an order of magnitude change in the deviations of the MVA normals as the asymmetry of the data interval is varied. Furthermore, the minimum deviations occur for non-unitary symmetry indices (i.e. asymmetric data intervals) and this minimum varies according to the inclination of the trajectory. The minimum occurs at higher symmetry indices (i.e. more asymmetric data intervals) for lower trajectory inclinations. Since the curvature of the field lines is tighter after the discontinuity centre is encountered compared to the situation prior to encounter, the corresponding field normal directions are (monotonically) changing faster after rather than before encounter. Since MVA finds some average normal over the sampled data interval, you would not expect to see zero deviation for a perfectly symmetric interval, and moreover would expect minimal/zero deviation to occur for symmetry indices greater than unity.

The effects of data resolution on the calculation of discontinuity normals is illustrated in Figure 7, using one analysis. It shows that the deviation is of the order of a few degrees. Intuitively we expect to find that MVA normals are of higher accuracy the greater the data resolution, as in Figure 7. However, this is not always the case and the situation is not as clear-cut as the effects described above. We leave this to a more detailed later study. The main point we wish to make here is that data resolution does have an affect on MVA normal determination and is of the order of a few degrees.

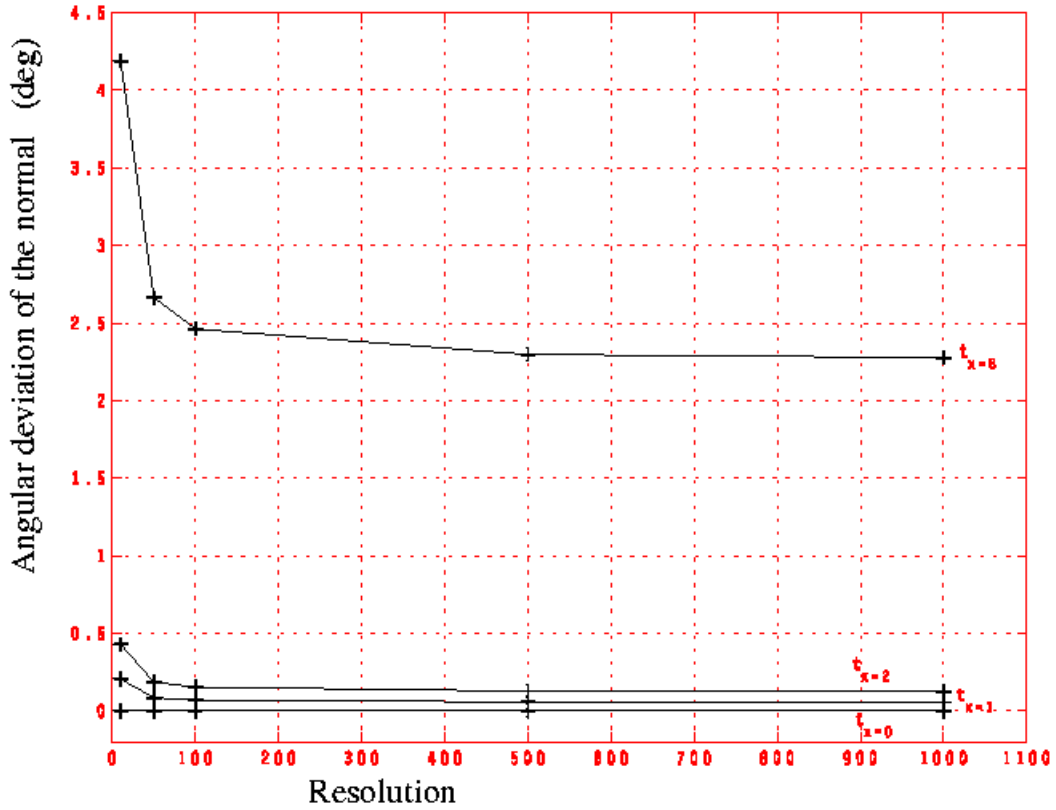


Figure 7: Sensitivity of MVA analysis to the model shown in Figure 5 in terms of data resolution.

It isn't surprising, of course, that MVA provides normal estimates in curved geometry which deviate from the geometric normals indicated by the models. We explore the effect here to investigate the quantitative biasing on the normals in the sense of their use within the discrete methodology above; i.e. as a single spacecraft analysis for the surface normal \mathbf{n} , feeding into the equation of motion. This systematic error depends on the model used, of course, but the trends show that the significance of this error with respect to the overall deviations of the normals between spacecraft can be optimised by choice of trajectory, given the data sampling achieved. The sampling depends upon the combination of the minimum nesting interval applied and the data resolution with respect to the characteristic spatial scale of the structure. In practice, the sampling achieved depends upon the thin boundary properties of an event. The trajectory selected then sets the spacecraft separation scale with respect to the characteristic scale of the model.

Clearly, different events will be sampled in different ways *in situ* so that for some the systematic error will be comparatively large and for others comparatively small. In the latter case, the discrete methodology may be used for such selected events to estimate both lowest order curvature and acceleration terms. As indicated by the flow in Figure 2, the rejection of qualitatively unsuitable events may be done if the normal analysis shows unstable results (with respect to nesting or shifted intervals) in the presence of significant deviations between the spacecraft normals.

3. Analysis of a real event

Figure 8 shows a real event measured by the combined AMPTE IRM (Figure 8(a)) and UKS (Figure 8(b)) spacecraft during a solar wind Barium release (see papers in *Nature*, **320**, 700-726, 1986), plotted in GSE co-ordinates. Both spacecraft were thought to have sampled an intense plasma cloud boundary at about the time (within a few seconds) indicated by the solid vertical lines on the $|B|$ plots in the upper panels. Because of cloud expansion following the release, this boundary should be strongly curved between the two spacecraft and this appears to be confirmed by boundary normals obtained from each data set. The sense of the curvature implied by the normals, which were obtained using a standard run, is described in *Dunlop et al.* [1996] and is consistent with an expanding cloud where the effective tilt between the normal directions is ~ 30 -40 (deg). Clearly, this event would be unsuitable for the discrete analysis in the manner discussed above in section 2.2.

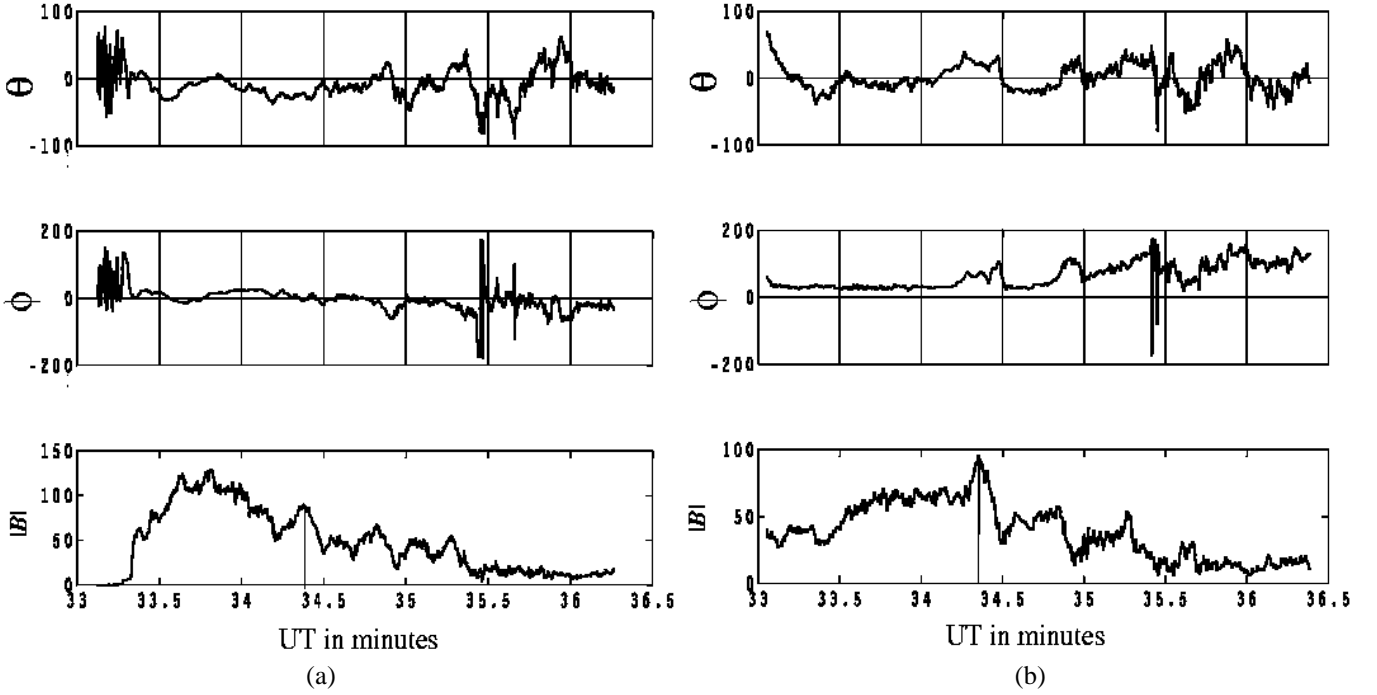
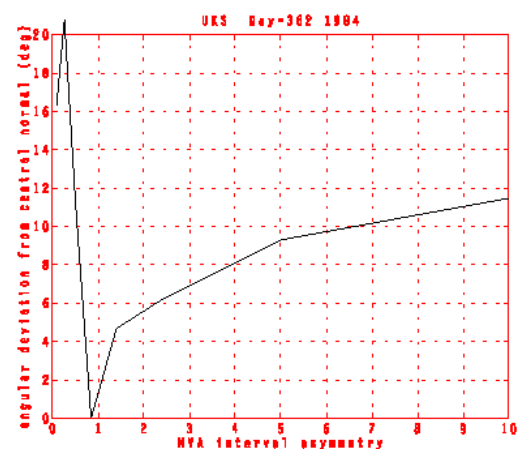
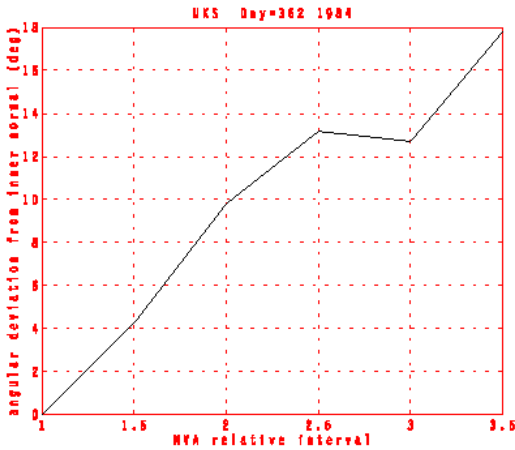
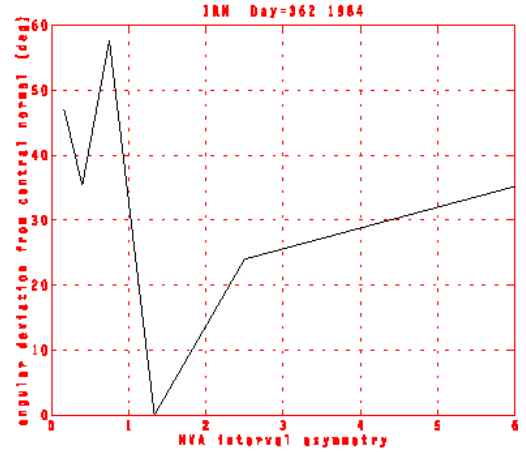
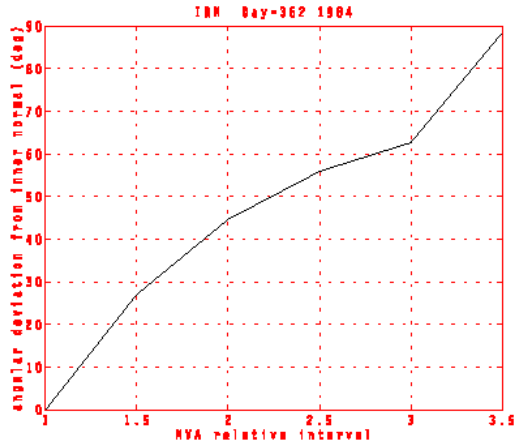


Figure 8: Example of a curved discontinuity from the AMPTE-UKS/IRM solar wind Barium release, as discussed in the text.

We therefore expect that MVA will return a biased normal estimate, which will be sensitive to the actual sampling path of the spacecraft, as discussed above. To investigate the sensitivity of the MVA normals, a similar set of runs has been performed, successively adjusting the data intervals used to determine them. For this real event, covered only by two spacecraft, we clearly only have two fixed trajectories, and have fixed data resolution. We do not have the equivalent of Figure 7, therefore, but can analyse in the spirit of Figures 5 and 6. Effectively, the IRM spacecraft corresponds to the model sampling trajectory, $t_{x=0}$, being in the centre of the cloud. The UKS spacecraft corresponds to an intermediate trajectory, inclined at ~ 30 - 40 (deg) to the boundary, traversing the outer cloud region before exiting. The data signature, however, is complicated by properties other than the simple model described above. In particular, the magnetic signature changes character over the extent of the cloud, and between the two spacecraft (since they traverse different regions). Also, upstream conditions (times following the exit time) are obviously not static and contain clear wave power at characteristic frequencies. Thus, we cannot test the quantitative comparison between the trajectories (spacecraft), but Figure 9 does show the sensitivity to changes in the MVA interval extremely well.

The two vertical panels in Figure 9(a) show the equivalent trends to those in Figure 5 for the two spacecraft separately, where the analysis interval has been increased in steps, from 20 seconds up to one and a half minutes, centred on 12:34:20 UT. Both panels show initially, the steady increase in deviation of the boundary normal, here defined as the angle of successive estimates with respect to the inner interval. At some point the trend breaks down, since other properties are being sampled well within or upstream of the cloud and its structure is likely to be still changing in time. The panels in Figure 9(b) are perhaps more significant, since they show the equivalent trends to those in Figure 6, where an analysis interval of 2 minutes (for UKS), and 1 minute (for IRM), has been shifted in steps of 10 seconds about 12:34:20 UT. The interval asymmetry has been defined as the ratio of inner to outer portions of the time interval, as for Figure 6. Clearly, both spacecraft, but particularly UKS, show deviations (defined with respect to the symmetric interval) which have closely similar character to those in Figure 6.



(a)

(b)

Figure 9: Sensitivity of MVA analysis to the event shown in Figure 9, (a) in terms of nesting and (b) in terms of asymmetry.

4. Discussion and conclusion

In this paper we have presented a summary of the DA technique, based on magnetometer data. This seeks to characterise the topological structure and motion of thin, non-dispersive structures present in space magneto-plasmas. It assumes *in situ* detection by a small array (up to four) of closely separated spacecraft. The methodology underlying the technique here, namely the discrete, sequential approach, first determines topological parameters, defined with respect to boundary normals, independently before these are fed into the motional equations. The planar form of the DA, which is based on the discrete analysis framework, relies on the independent determination of boundary normals to give a direct demonstration of planarity, and this then sets the context for the subsequent motional analysis. We point out that in the case of non-planar structures, topological and motional properties compete for: (i) their representation in the data, and (ii) the nature of the analysis that can be performed. Techniques such as those presented by *Chao and Lepping* [1974] and *Russell et al.* [1983] rely on an assumption of planarity, rather than its demonstration. *Mottez and Chanteur* [1994] assume constant velocity and relax the planarity assumption, relying on accurate single spacecraft-determined normal directions, in their analysis. Thus, these techniques are only mutually consistent with their assumptions; but these assumptions may not be consistent with the data, leading to the determination of an anomalous motion, on the one hand, or an anomalous curvature on the other.

The non-planar form of the DA therefore represents the more intricate analysis problem, with assumptions on topology affecting the motional determination and vice versa. The discrete methodology orders the analysis sequentially, but under the requirement that boundary normals can be accurately obtained (from individual, time series analysis). We can ask, however, how the individual normal directions are adjusted to take account of a surface fitting procedure: the above implies that equal weight is given to each. Uncertainties arising from normal determinations in particular, can therefore be a major contributor to the procedural dilemma for establishing all macroscopic parameters, since it isn't obvious that normal directions, so determined, are unbiased in curved geometry. Errors are likely to be systematic, especially in the context of boundaries that are not thin. A combined approach has been outlined by *Dunlop and Woodward* [1998], which attempts to deal with this dilemma, in which all surface and motional parameters are determined simultaneously: some weighted adjustment of the surface away from the strict orientations of the boundary normals is allowed in order to optimise for all parameters. Nevertheless, a data regime is applicable for the DA technique, as conceived, in which boundary structure can be described as approximately thin.

In view of the importance of this issue, we have presented an analysis of the variation of normals (determined using minimum variance analysis) to a curved 2-D model boundary from actual model normals under various scenarios of sampling. The model topologically orders the field far from its centre and is hence ‘thick’. We found that the greatest deviations arose for asymmetry of the analysis data interval about the centre of the event, with higher deviations occurring for trajectories highly inclined to the characteristic directions of the model. Smaller deviations were observed for the size of the data interval and data resolution, and again the greater the inclination of the sampling trajectory to the characteristic directions of the model the greater the deviations were in both cases. Analysis of barium release data from the AMPTE-IRM/UKS spacecraft provided a real event demonstration of the trends found with the model event, indicating significant deviations of the normals from the average normal of up to 90 deg., and confirming the model analysis. Thus, we conclude that normal determination in curved geometries is open to error, and furthermore cannot be completely de-coupled from the determination of the surface topology. One could attempt to refine the uncertainties using the implied curvature (iterative process) and proceed with the discrete analysis methodology. Conversely the combined methodology may be adopted, using confidence factors implied by analysis similar to that described here. Furthermore, a possible event selection test, to identify ‘thin’ boundaries by combining the individual time series analyses, has been conjectured by *Dunlop and Woodward* [1998].

In any multi-spacecraft analysis, it is generally important to understand in what way comparative information across spacecraft is best built into any analysis of the combined data. Although for the DA, there is at first sight an apparently clear separation of the handling of multi-point and time series information, in general, different investigations appear to demand different methods of data co-ordination. The use of other instrument data, of course, can potentially add information on structure, such as independent determination of boundary motion in the case of electric field measurements, or via DeHoffmann-Teller analysis using plasma data. This must be carefully assessed, however, since with data restrictions on resolution and quality of these other data sets, a different analysis methodology may be required and such additional measurements are most easily added as independent information on the event. The same is true of analysis using more than four spacecraft (the minimum for direct 3-D spatial information), where there will be at least a play off between use of the redundant information to improve spatial accuracy or establish temporal evolution.

Acknowledgements

We would like to thank Professor Karl-Heinz Glassmeier and Professor Peter Weidelt of the Institut für Geophysik und Meteorologie, Technische Universität Braunschweig, Germany for useful discussions on elements of this work. We acknowledge financial support from PPARC in the U.K.

References

- Bame, S. J., J. R. Asbridge, J. T. Gosling, M. Halbig, G. Paschmann, N. Sckopke, and H. Rosenbauer, High temporal resolution observations of electron heating at the bow shock, *Space Sci. Rev.*, **23**, 75, 1979.
- Chao, J. K., and R. P. Lepping, A correlative study of ssc's, interplanetary shocks, and solar activity, *J. Geophys. Res.*, **79**, 1799-1807, 1974.
- Dunlop, M. W., D. J. Southwood, K.-H. Glassmeier, and F. M. Neubauer, Analysis of multipoint magnetometer data, *Adv. Space Res.*, **8**, (9)273-(9)277, 1988.
- Dunlop M. W., Review of the Cluster orbit and separation strategy: consequence for measurements, *SP-306*, p17, ESA Publications, Noordwijk, The Netherlands, February, 1990.
- Dunlop, M. W., T. I. Woodward, and C. J. Farrugia, Minimum variance analysis: CLUSTER themes, *Proc. Cluster workshop on Physical Measurements and Mission Orientated Theory*, **ESA SP-371**, 33-42, 1995.
- Dunlop, M. W., T. I. Woodward, U. Motschmann, D. J. Southwood, and A. Balogh, Analysis of non-planar structures with multipoint measurements, *Adv. Space Res.*, **18**, (8)309, 1996.
- Dunlop, M. W., and T. I. Woodward, Discontinuity analysis: orientation and motion, in *Analysis methods for multispacecraft data*, ISSI Scientific Report **SR-001**, p271, ESA Publications, Noordwijk, The Netherlands, 1998.
- Fairfield, D. H., and W. C. Feldman, Standing waves at low Mach number laminar bow shocks, *J. Geophys. Res.*, **80**, 515-522, 1975.
- Greenstadt, E. W., C. T. Russell, F. L. Scarf, V. Formisano, and M. Neugebauer, Structure of the quasi-perpendicular laminar bow shock, *J. Geophys. Res.*, **80**, 502-514, 1975.
- Khrabrov, A. V., and B. U. Ö. Sonnerup, DeHoffmann-Teller analysis, in *Analysis methods for multispacecraft data*, ISSI Scientific Report **SR-001**, p221, ESA Publications, Noordwijk, The Netherlands, 1998.
- Lepping, R. P., and J. K. Chao, A shock surface geometry: The February 15-16, 1967, event, *J. Geophys. Res.*, **81**, 60-64, 1976.
- Mellott, M. M., and E. W. Greenstadt, The structure of oblique subcritical bow shocks: ISEE-1 and -2 observations, *J. Geophys. Res.*, **89**, 2151, 1984.
- Morse D. L., and E. W. Greenstadt, Thickness of magnetic structures associated with the Earth's bow shock, *J. Geophys. Res.*, **81**, 1791, 1976.
- Mottez, F., and G. Chanteur, Surface crossing by a group of satellites; a theoretical study, *J. Geophys. Res.*, **99**, 13,499-13,507, 1994.
- Russell, C. T., and E. W. Greenstadt, Initial ISEE Magnetometer Results: Shock Observations, *Space Science Rev.*, 3-37, 1979
- Russell, C. T., M. M. Hoppe, W. A. Livesey, J. T. Gosling, and S. J. Bame, ISEE-1 and -2 observations of laminar bow shocks: Velocity and thickness, *Geophys. Res. Lett.*, **9**, 1171, 1982.
- Russell, C. T., M. M. Mellott, E. J. Smith, and J. H. King, Multiple observations of interplanetary shocks: four spacecraft determination of shock normals, *J. Geophys. Res.*, **88**, 4739-4748, 1983.
- Song, P., R. C. Elphic, and C. T. Russell, Multi-spacecraft observations of magnetopause surface waves: ISEE-1 and -2 determinations of amplitude, wavelength and period, *Adv. Space Res.*, **8**, (9)245-(9)248, 1988.
- Sonnerup, B. U. O., and L. J. Cahill, Magnetopause structure and attitude from Explorer 12 observations, *J. Geophys. Res.*, **72**, 171-183, 1967.

Reprint from:

Chanteur, G., Accuracy of field gradient estimations by Cluster: Explanation of its dependency upon elongation and planarity of the tetrahedron, in *Proceedings of the Cluster-II Workshop Multiscale/Multipoint Plasma Measurements, Imperial College, London, 22–24 September 1999*, ESA SP-449, pp. 265–268, European Space Agency, Paris, France, 2000

ACCURACY OF FIELDS GRADIENTS ESTIMATIONS BY CLUSTER: EXPLANATION OF ITS DEPENDENCY UPON ELONGATION AND PLANARITY OF THE TETRAHEDRON

G. Chanteur

Centre d'étude des Environnements Terrestre et Planétaires, CNRS, Vélizy, France

Reprinted from:

Proceedings of the Cluster-II Workshop Multiscale/Multipoint Plasma Measurements
ESA SP-449, European Space Agency, Paris
265–268, 2000

ABSTRACT

The accuracy of estimations of fields gradients by CLUSTER will be influenced by the aspect of the tetrahedron formed by the four spacecraft. This key problem for the mission has been investigated by means of numerical simulations and theoretical results have been derived within the framework of the barycentric formalism. The present study provides a link between theoretical and simulation approaches under simplifying assumptions about the covariance matrices of the position vectors of spacecraft and the covariance matrices of the measured field components.

1. INTRODUCTION

The CLUSTER II mission with its four identical sets of scientific instruments will for the first time allow the unambiguous estimation of fields gradients through the simultaneous measurement of physical fields in four distinct points. Since the beginning of the CLUSTER project different methods have been designed for estimating fields gradients: the early used methods of finite differences, poorly adapted to irregular tetrahedra, were soon replaced by integral theorems (for example Ampere's and Gauss theorems) to estimate $\nabla \times \mathbf{B}$ and $\nabla \cdot \mathbf{B}$ (Dunlop et al. 1990; Robert & Roux 1990). Later Chanteur & Mottez (1993) introduced the method of barycentric coordinates and gave an estimator of the gradient of a vector field. A detailed presentation of this approach and its applications is given by Chanteur (1998). Harvey (1998) introduced a method of least-squares with the possibility to enforce the solenoidal constraint in the case of the magnetic field. Least-squares and barycentric estimations of magnetic gradients along simulated realistic CLUSTER orbits inside a Tsyanenko magnetospheric model have been presented in Chanteur & Harvey (1998). All these approaches, except for the constrained least-squares, rely upon a linear interpolation of fields through the cluster of spacecraft, nevertheless the

method of barycentric coordinates is the only one that provides simple analytical formulas and allows a theoretical investigation of errors affecting the estimations of fields gradients for any configuration of the cluster.

Vogt & Paschmann (1998) have examined the accuracy of the derivatives of plasma moments making use of finite differences when the four spacecraft form an orthogonal frame and barycentric coordinates in the general case. A general discussion of the different types of errors affecting the estimations of fields gradients is given by Chanteur (1998). Truncature errors originating from the linear interpolation of the field inbetween spacecraft essentially depend upon the field configuration sampled by the cluster: they are discussed in Chanteur (1998) for a thick and plane current sheet, and in Chanteur & Harvey (1998) for a dipolar field. Physical errors originate in the uncertainty affecting the field measurements meanwhile geometrical errors are due the uncertainty on spacecraft positions in space. Numerical simulations (Robert et al. 1998) demonstrated that the accuracy of current density estimations made from clusters located inside simple current tubes (*i.e.* for which \mathbf{J} is monodirectional) is a function of the elongation and planarity of the cluster. The aim of the present contribution is to demonstrate that simulation results can be explained on the basis of theoretical formula given by Chanteur (1998). The main results concerning errors and gradients estimations are briefly reviewed in the next section. Section 3 is devoted to a simplification of the theoretical formulas which leads to the explicit dependency of physical and geometrical errors upon size, elongation and planarity of the cluster. Section 4 presents an application to a force-free current tube with twisted field lines.

2. ESTIMATIONS OF GRADIENTS AND THEIR ERRORS

For the sake of completeness let us begin with a short account of some of the results demonstrated in sections 14.2 and 14.3 of Chanteur (1998) that will be hereafter

used. The key objects in the barycentric approach are the reciprocal vectors of the tetrahedron formed by the four spacecraft. Figure 1 shows the cluster of spacecraft and emphasizes the geometrical meaning of the reciprocal vectors. The reciprocal vector \mathbf{k}_4 is given by:

$$\mathbf{k}_4 = \frac{\mathbf{r}_{12} \times \mathbf{r}_{13}}{\mathbf{r}_{14} \cdot (\mathbf{r}_{12} \times \mathbf{r}_{13})} \quad (1)$$

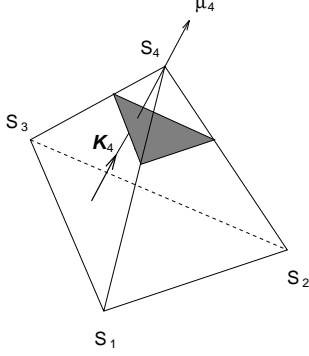


Figure 1. Tetrahedron formed by four spacecraft. The reciprocal vector \mathbf{k}_4 , defined by formula 1, points towards vertex S_4 and is the gradient of the barycentric coordinate μ_4 which is constant on any plane parallel to the face $S_1S_2S_3$.

The linear estimator of the gradient of a vector field \mathbf{A} is the following tensor built from the reciprocal vectors and the field vectors measured by the four spacecraft:

$$\mathbf{G} = \sum_{\alpha=1}^4 \mathbf{k}_{\alpha} \mathbf{A}_{\alpha}^t \quad (2)$$

The covariance of the (i, j) and (m, n) components of \mathbf{G} writes, accordingly to formula (14.30) in Chanteur (1998):

$$\langle \delta G_{ij} \delta G_{mn} \rangle = \sum_{\alpha=1}^4 \sum_{\beta=1}^4 (\langle \delta A_{\alpha i} \delta A_{\beta m} \rangle k_{\alpha j} k_{\beta n} + \langle \delta k_{\alpha j} \delta k_{\beta n} \rangle A_{\alpha i} A_{\beta m}) \quad (3)$$

The general expression for the covariance of components i and j of the reciprocal vectors \mathbf{k}_{α} and \mathbf{k}_{β} is:

$$\langle \delta k_{\alpha i} \delta k_{\beta j} \rangle = \sum_{\gamma=1}^4 \sum_{v=1}^4 \mathbf{k}_{\alpha}^t \langle \delta \mathbf{r}_{\gamma} \delta \mathbf{r}_v^t \rangle \mathbf{k}_{\beta} k_{\gamma i} k_{v j} \quad (4)$$

When the position vectors of the four spacecraft are uncorrelated and when covariance matrices of the four position vectors are identical we can write by introducing Kronecker symbols $\delta_{\gamma, v}$:

$$\langle \delta \mathbf{r}_{\gamma} \delta \mathbf{r}_v^t \rangle = \delta_{\gamma, v} \langle \delta \mathbf{r} \delta \mathbf{r}^t \rangle \quad (5)$$

which leads to a simplified form of equation 4:

$$\langle \delta k_{\alpha i} \delta k_{\beta j} \rangle = \mathbf{k}_{\alpha}^t \langle \delta \mathbf{r} \delta \mathbf{r}^t \rangle \mathbf{k}_{\beta} K_{ij} \quad (6)$$

where the K_{ij} 's are the components of the reciprocal tensor defined by:

$$\mathbf{K} = \sum_{v=1}^4 \mathbf{k}_v \mathbf{k}_v^t \quad (7)$$

3. DEPENDENCY OF PHYSICAL AND GEOMETRICAL ERRORS UPON ELONGATION AND PLANARITY

3.1. Simplified formulas for covariances

In default of information about the covariance tensor of position vectors we assume that:

$$\langle \delta \mathbf{r} \delta \mathbf{r}^t \rangle = (\Delta r)^2 \mathbf{I} \quad (8)$$

where \mathbf{I} is the unit tensor and Δr is the uncertainty on spacecraft positions and will be chosen equal to 1% of the size of the tetrahedron for the numerical applications. Terms corresponding to geometrical errors then factorize and equation 3 can be written:

$$\langle \delta G_{ij} \delta G_{mn} \rangle = \sum_{\alpha=1}^4 \sum_{\beta=1}^4 \langle \delta A_{\alpha i} \delta A_{\beta m} \rangle k_{\alpha j} k_{\beta n} + (\Delta r)^2 \nabla A_i \cdot \nabla A_m K_{jn} \quad (9)$$

where equation 2 has been used to notice that for negligible truncation errors:

$$\nabla A_i = \sum_{\alpha=1}^4 \mathbf{k}_{\alpha} A_{\alpha i}$$

For the sake of simplicity we reduce the physical errors in the same way we did for the geometrical errors, *i.e.* we assume that:

$$\langle \delta A_{\alpha i} \delta A_{\beta m} \rangle = (\Delta A)^2 \delta_{\alpha, \beta} \delta_{i, m} \quad (10)$$

where ΔA is the uncertainty of the field measurement. This latter hypothesis on the covariance matrix of the measured vectors should be questioned for a given physical field by taking into account the details of the experiment designed for this measure. Our simple hypothesis leads to the following final expression for the covariances of the components of \mathbf{G} , the barycentric estimator of tensor $\nabla \mathbf{A}$:

$$\langle \delta G_{ij} \delta G_{mn} \rangle = \left((\Delta A)^2 \delta_{i,m} + (\Delta r)^2 \nabla A_i \cdot \nabla A_m \right) K_{jn} \quad (11)$$

It is worth noticing that, under the simplifying assumptions we have made, the physical and geometrical contributions to the covariances of the gradient depend upon the geometry of the cluster only through the components of the reciprocal tensor. A theoretical explanation of the simulation results of Robert et al. (1998) will be derived from this simple result.

3.2. Averaging over equiprobable eigenbases

A tensor \mathbf{K} with eigenvalues $(\lambda_1, \lambda_2, \lambda_3)$ and orthonormal eigenvectors $(\mathbf{e}_1, \mathbf{e}_2, \mathbf{e}_3)$ can be represented by the matrix:

$$\mathbf{K} = \mathbf{M} \text{Diag}(\lambda_1, \lambda_2, \lambda_3) \mathbf{M}^t \quad (12)$$

where \mathbf{M} is the 3x3 matrix of which the k -th column is equal to \mathbf{e}_k and $\text{Diag}(\lambda_1, \lambda_2, \lambda_3)$ is the diagonal matrix with elements $(\lambda_1, \lambda_2, \lambda_3)$. Specifying the orientation of the eigenbasis in space by its Euler angles (ψ, θ, ϕ) allows to compute the average of \mathbf{K} for given probability distributions of the Euler angles: details will be published elsewhere. When all orientations of the eigenbasis are equiprobable the expectation $\overline{\mathbf{K}}$ of \mathbf{K} over the orientations is given by:

$$\overline{\mathbf{K}} = \frac{1}{3} \text{Trace}(\mathbf{K}) \text{Diag}(1, 1, 1) \quad (13)$$

For given size a , elongation e and planarity p , the eigenvalues of the reciprocal tensor \mathbf{K} are equal to:

$$a^{-2}, \quad a^{-2}(1-e)^{-2}, \quad a^{-2}(1-e)^{-2}(1-p)^{-2}$$

and the average of equation 11 over equiprobable orientations gives:

$$\langle \delta G_{ij} \delta G_{mn} \rangle = \left((\Delta A)^2 \delta_{i,m} + (\Delta r)^2 \nabla A_i \cdot \nabla A_m \right) \frac{g(e, p)}{a^2} \delta_{j,n} \quad (14)$$

Figure 2 shows the contour levels of the function $g(e, p) = \frac{a^2}{3} \text{Trace}(\mathbf{K})$ which reproduces the behaviour of $\Delta J(e, p)/J$ illustrated by Figures 16.7, 16.11 and 16.12 of Robert et al. (1998). This similarity will be explicated in the next section.

A factor equal to the trace of the reciprocal tensor also appears in the angular averaged error affecting the directional derivative of a scalar quantity related to plasma measurements; see equation 17.35 or 17.70 of Vogt & Paschmann (1998).

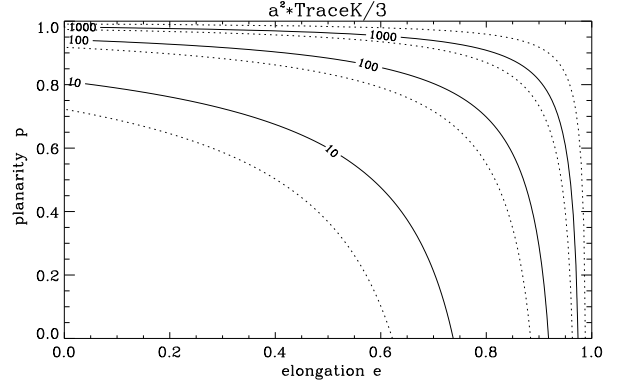


Figure 2. Isovalues of $g(e, p) = \frac{a^2}{3} \text{Trace}(\mathbf{K})$ versus elongation e and planarity p of the tetrahedron.

4. RESULTS FOR CURRENT TUBES

4.1. Model of a force-free current tube

A force-free magnetic field with a structure function $\alpha(\mathbf{r})$ is a solution of the following equations:

$$\nabla \times \mathbf{B} = \alpha(\mathbf{r}) \mathbf{B} \quad \text{and} \quad \nabla \cdot \mathbf{B} = 0$$

In cylindrical coordinates (r, ϕ, z) axisymmetric and axially invariant solutions have a zero radial component meanwhile the angular and axial components B_ϕ and B_z are solutions of the coupled equations with the axial conditions $B_\phi(0) = 0$ and $B_z(0) = B_0$:

$$\frac{dB_\phi}{dr} = -\frac{B_\phi}{r} + \alpha(r)B_z \quad \text{and} \quad \frac{dB_z}{dr} = -\alpha(r)B_\phi$$

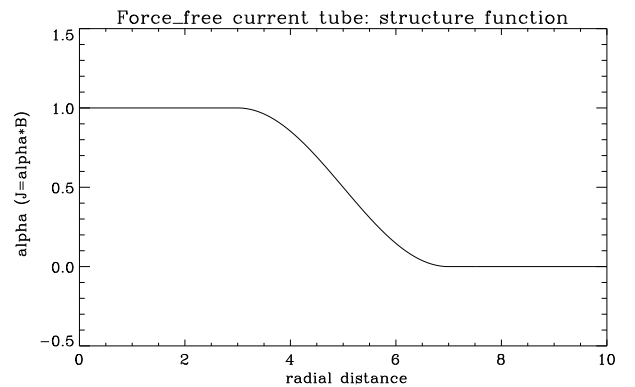


Figure 3. Chosen structure function which determines the force-free field used in this study: $r_1 = 3$. and $r_2 = 7$.

If for $r < r_1$, $\alpha(r) = \alpha_0$ (constant) then for $r < r_1$ the solution is $(B_\phi, B_z) = B_0 (J_1(\alpha_0 r), J_0(\alpha_0 r))$, where J_0 and J_1 are Bessel functions of order zero and one. If for $r > r_2$, $\alpha(r) = 0$ then for $r > r_2$ the solution is determined by the equations $B_z(r) = B_z(r_2)$ and $r B_\phi(r) = r_2 B_\phi(r_2)$. A numerical integration is usually required to determine the solution for a given structure function $\alpha(r)$. Figure 4 presents the radial profiles of the

magnetic components for the structure function specified by Fig. 3. The field lines of the present model have both curvature and twist in contrast with previously used models of current tubes.

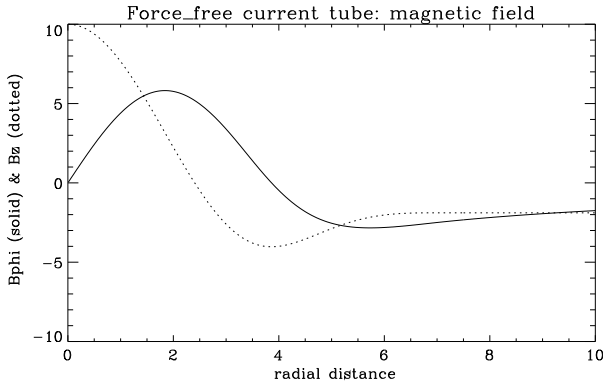


Figure 4. Angular and axial components of the force-free magnetic field determined by the structure function illustrated by Fig. 3.

4.2. Covariances of the components of \mathbf{J}

Making use of equation 12 we construct a statistical ensemble of reciprocal tensors having specified (a, e, p) and equiprobable orientations (16000 in this study) to determine the probability distribution function of the axial component of the current density estimated by a cluster located on the edge of the force-free current tube presented in Section 4.1: the result is displayed on Figure 5. The maximal value of $\Delta J_z / |J_z|$ (see definition below) is equal to 0.15 and the most probable value is equal to 0.08 for $(a, e, p) = (1., 0.8, 0.5)$. A detailed study of these probability distributions will be published later. The angular averaged covariance of the estimate of J_z is expressed as follows:

$$\overline{\langle \delta J_z \delta J_z \rangle} = \left(2(\Delta B)^2 + (\Delta r)^2 [(\nabla B_x)^2 + (\nabla B_y)^2] \right) \frac{g(e, p)}{a^2} \quad (15)$$

Thus it appears that the variations of $\Delta J_z / |J_z| = \sqrt{\langle \delta J_z \delta J_z \rangle} / J_z^2$ with elongation e and planarity p have their origin in the function $g(e, p)$ introduced in Section 3.2 (see Figure 2).

5. CONCLUSIONS

By making simplifying assumptions we have demonstrated that the theoretical investigation of the accuracy of fields gradients estimations by a cluster of four spacecraft (Chanteur 1998) can explain the simulation results of Robert et al. (1998). The assumptions explicated by equations 5, 8 and 10 lead to the simple form 14 for the average of equation 3 over equiprobable orientations of

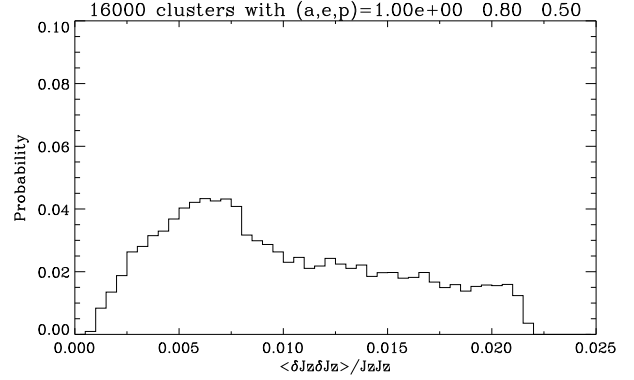


Figure 5. Simulated probability distribution function of the reduced variance of J_z for elongated tetrahedra, $(a, e, p) = (1.0, 0.8, 0.5)$, located on the edge, $r = 5.0$, of the force-free current tube described in Section 4.1.

the tetrahedron. This theoretical result explains the variations of the accuracy of current estimations with the elongation and planarity of the cluster formerly demonstrated by numerical simulations (Robert et al. 1998). This theoretical approach also allows to evaluate the probability distribution functions of the covariances of the estimations of fields gradients.

Acknowledgements

The author thanks Dr P.W. Daly for his help with \LaTeX and \BIBTeX softwares and for providing style files.

REFERENCES

- Chanteur G., 1998, In: Paschmann G., Daly P.W. (eds.) Analysis Methods for Multi-Spacecraft Data, ISSI Scientific Report SR-001, 349–369, International Space Science Institute, Bern, Switzerland
- Chanteur G., Harvey C.C., 1998, In: Paschmann G., Daly P.W. (eds.) Analysis Methods for Multi-Spacecraft Data, ISSI Scientific Report SR-001, 371–393, International Space Science Institute, Bern, Switzerland
- Chanteur G., Mottez F., 1993, In: Spatio-Temporal Analysis for Resolving Plasma Turbulence, ESA WPP-047, 341–344, ESA, Paris, France
- Dunlop M.W., Balogh A., Southwood D.J., et al., 1990, In: Space Plasma Physics Investigations by CLUSTER and Regatta, ESA SP-306, 23–28, ESA, Paris, France
- Harvey C.C., 1998, In: Paschmann G., Daly P.W. (eds.) Analysis Methods for Multi-Spacecraft Data, ISSI Scientific Report SR-001, 307–322, International Space Science Institute, Bern, Switzerland
- Robert P., Roux A., 1990, In: Space Plasma Physics Investigations by CLUSTER and Regatta, ESA SP-306, 29–35, ESA, Paris, France
- Robert P., Dunlop M.W., Roux A., Chanteur G., 1998, In: Paschmann G., Daly P.W. (eds.) Analysis Methods for Multi-Spacecraft Data, ISSI Scientific Report SR-001, 395–418, International Space Science Institute, Bern, Switzerland

Vogt J., Paschmann G., 1998, In: Paschmann G., Daly P.W. (eds.) *Analysis Methods for Multi-Spacecraft*

Data, ISSI Scientific Report SR-001, 419–447, International Space Science Institute, Bern, Switzerland

Reprint from:

Horbury, T. S., Cluster II analysis of turbulence using correlation functions, in *Proceedings of the Cluster-II Workshop Multiscale/Multipoint Plasma Measurements, Imperial College, London, 22–24 September 1999*, ESA SP-449, pp. 89–97, European Space Agency, Paris, France, 2000

CLUSTER II ANALYSIS OF TURBULENCE USING CORRELATION FUNCTIONS

T. S. Horbury

Astronomy Unit
Queen Mary and Westfield College
Mile End Road
London
E1 4NS

Email: t.horbury@qmw.ac.uk

Abstract

The Cluster II mission offers the prospect of studying, for the first time, small scale magnetospheric and solar wind turbulence in three dimensions and with a wide-ranging set of instrumentation. We discuss possible Cluster II applications to this topic, concentrating on the correlation function as a generic measure of turbulence structure. We stress what is possible, and not possible, with four spacecraft data and discuss the effects of varying spacecraft separation lengths and orientations. Finally, we consider limitations of the Cluster II spacecraft instrumentation which may impact on turbulence measurements.

Introduction

A major aim of the Cluster II mission is to study small scale structures and dynamics in the near-Earth environment. Some of the key target areas for the mission (e.g. Escoubet *et al.*, 1997) are the cusp, bowshock, magnetopause, plasmashet and solar wind. All these regions are highly structured and often turbulent. As such, the study of turbulence in these environments is of considerable interest. Plasma turbulence, on both magnetohydrodynamic (MHD) and kinetic scales, is important for energetic particle propagation; shock dynamics; energy and momentum transfer; and of course is a fundamental topic of interest in its own right. The Cluster II formation offers the prospect of significant advances in the understanding of several of these topics.

In this paper, we describe some methods of analysing Cluster II data to study turbulence. A considerable literature already exists on turbulence; multi-point measurements; and Cluster II, so this work will not provide details of any of these subjects. Many key concepts regarding analysis of data from multiple spacecraft are covered by Paschmann and Daly (1998) and papers therein and we refer the reader to these papers for more detail on these topics than is presented here.

The state of the art instruments onboard the four Cluster II spacecraft are likely to provide significant advances in our understanding of plasma processes in the magnetosphere and solar wind, even with single spacecraft data. However, in this paper we concentrate on techniques that can be used by combining data from the four spacecraft formation. We discuss magnetic field and plasma moment data in particular, since these measurements are most often used in the analysis of MHD scale turbulence in the heliosphere and they can be used to illustrate generic multi-point analysis concepts, although it is likely that other data products from a range of instruments will be useful in the study of these processes. In addition, many of the concepts presented here are applicable to the combination of data from multiple spacecraft in contexts other than turbulence analysis.

Turbulence science topics

Extensive introductions to heliospheric turbulence, particularly in the MHD regime, have been presented by Tu and Marsch (1995), Goldstein and Roberts (1995) and Matthaeus *et al.* (1995). Frisch (1995) is a good introduction to turbulence in general, although it does not discuss MHD turbulence. Marsch (1995) and Goldstein (1995) reviewed the state of MHD turbulence research from the perspective of Cluster and we refer the reader to these works for a more detailed introduction to the subject.

It is well established that the solar wind is pervaded by waves and turbulence on MHD scales. Changes in the shape of the power spectrum with solar distance, first shown by Bavassano *et al.* (1982), demonstrate that there is active energy transfer between scales and that the fluctuations are participating in an ongoing turbulent cascade, at least on relatively small scales. However, many aspects of this cascade are poorly understood. In particular, the three dimensional structure of MHD turbulence is not well known. Single spacecraft measurements cannot be used to determine this structure in anything other than a statistical sense.

Turbulent structure is important for a number of reasons. Firstly, it is produced by, and influences, energy transfer and the turbulent cascade. Understanding of neutral fluid turbulence has increased with the discovery of the presence of discrete structures (vortex sheets, “mushrooms,” etc.) within the turbulence and the fact that much of the turbulent energy is located around them. These structures were discovered when measurements were taken at multiple points within turbulent flows. In MHD turbulence, discontinuities may play a similar role but single spacecraft measurements cannot provide sufficient information about the medium in three dimensions.

A second motivation for studying turbulent structure is the role it plays in energetic particle transport. The complex magnetic field structure within MHD turbulence greatly affects the motion of energetic particles such as cosmic rays (see, e.g. Potgieter, 1998). Recent work (e.g. Bieber *et al.*, 1996) has shown that anisotropy in the turbulence, a manifestation of structures within the fluid, can greatly alter cosmic ray diffusion coefficients and hence particle transport throughout the heliosphere. The scale invariance of MHD means that qualitatively similar structures are likely to exist in magnetofluid turbulence throughout the Universe, with consequences for energetic particle transport in many environments and on many scales.

The Cluster II formation, by sampling in four locations simultaneously, will provide additional information on the structure of the turbulence than is possible with one spacecraft. Some of the techniques developed for analysing multiple spacecraft data for other Cluster II applications, most obviously the “wave telescope” (e.g. Pinçon and Motschmann, 1998) and “discontinuity analyser” (e.g. Dunlop and Woodward, 1998), may have useful applications to turbulence measurements but we do not discuss these further here. In the next section, we describe some of the simple methods of analysing single spacecraft data using the two point correlation function and later show how to extend this to multiple spacecraft.

The correlation function

In this paper, we concentrate on methods for calculating the two point correlation function of a temporally invariant three dimensional vector or scalar field – for example, the magnetic field; bulk plasma velocity or density; or energetic particle flux. The correlation function is perhaps the simplest method for combining data from multiple sampling points and it is inevitable that considerably more sophisticated techniques (cross-wavelets, bispectra and structure functions, for example) will be employed in the analysis of Cluster II data. However, the correlation function illustrates the concepts of combining multiple spacecraft data, which are still relevant to more advanced methods.

The two point correlation tensor, R_{ij} , of a vector field $\mathbf{f}(\mathbf{x})$ is defined as

$$R_{ij}(\mathbf{r}) = \langle f_i(\mathbf{x}+\mathbf{r})f_j(\mathbf{x}) \rangle$$

where $\langle \rangle$ denote a spatial average over \mathbf{x} . For a scalar field, one can only define a correlation function, R ,

$$R(\mathbf{r}) = \langle f(\mathbf{x}+\mathbf{r})f(\mathbf{x}) \rangle.$$

In this paper, we are concerned with the coverage of vector separation \mathbf{r} , and as a result we need not distinguish between the correlation tensor and correlation function. We use “correlation function” throughout the paper on the understanding that, depending on the measured parameter of interest (density, velocity or field magnitude, for example) $R_{ij}(\mathbf{r})$ may be appropriate.

As discussed in the previous section, the three dimensional structure of plasma turbulence is of interest for a variety of reasons: the correlation function is one way of measuring and describing this structure. As we discuss later, the Cluster II formation offers the prospect of a significant advance in the measurement of the three dimensional correlation function of plasma turbulence.

Taylor’s hypothesis

A fundamental problem when analysing spacecraft data is that of separating spatial and temporal variations at the spacecraft – this is, of course, one of the topics which Cluster II is designed to address. In a fast-flowing medium such as the solar wind, however, it is possible to show that a time series recorded at a spacecraft can be considered as a spatial sample through the medium, using the so-called “Taylor’s hypothesis,” first discussed in the context of turbulence in wind tunnels by Taylor (1938).

A spatial variation in a fluid over a scale λ can be described by a wavenumber $k=2\pi/\lambda$. In the fluid frame, this variation changes on a characteristic time scale $t_\lambda=\lambda/V_{\text{wave}}$ where V_{wave} is the speed of the wave mode, typically the Alfvén mode in the solar wind. A spacecraft typically travels much slower (a few km/s) than the solar wind (several hundred km/s relative to the Sun or Earth) and can be assumed to be stationary, with a fast-moving wind passing over it. The spacecraft measures a variation on a plasma spatial scale λ in the flow direction on a time $t_{\text{sc}}=\lambda/V_{\text{flow}}$ where V_{flow} is the relative flow speed of the fluid. If $V_{\text{flow}} \gg V_{\text{wave}}$ then the ratio of the sampling time to the variation timescale, which is given by $t_{\text{sc}}/t_\lambda = V_{\text{wave}}/V_{\text{flow}}$, is much less than 1, so variations on a scale λ pass over the spacecraft much faster than they vary temporally and therefore the time series recorded by the spacecraft is effectively a spatial “snapshot.” In this case, one can convert a spacecraft recorded time t to a spatial position in the plasma \mathbf{x} as $\mathbf{x}=\mathbf{x}_0 + \mathbf{V}_{\text{flow}}t$ where \mathbf{V}_{flow} is approximately the solar wind velocity, which is nearly anti-sunward. In the solar wind near 1 AU, the Alfvén speed is a few tens of km/s, much less than the solar wind speed, so Taylor’s hypothesis is valid and one can therefore treat a time series as a spatial snapshot. This greatly simplifies the analysis of spacecraft data.

While Taylor's hypothesis is almost always valid in the solar wind at 1 AU, it is often not valid in the magnetosphere, where the magnetic field is often stronger (so wave speeds are higher) and flows are slower. However, some regions, such as parts of the magnetosheath, are usually sufficiently quickly flowing to make the approximation valid. Regions such as the magnetotail occasionally support rapid plasma flows (for example, so-called "bursty bulk flows") and it may prove possible to use Taylor's hypothesis at these times. Later in this paper we discuss the validity of Taylor's hypothesis for multiple spacecraft observations and show that the situation is rather more complex than the single spacecraft case.

Measuring the correlation function with one spacecraft

Using a single spacecraft in a sufficiently fast flow, one can measure the spatial correlation function in one direction, that of the flow. This is sufficient to measure the scaling properties of the fluctuations, but is obviously insufficient to measure their three dimensional structure. Various methods have been used to help overcome this limitation while analysing heliospheric turbulence, but these all involve assumptions about the structure of the fluctuations. The ability to measure the correlation function in a number of directions simultaneously, which Cluster II will provide, will obviously help to improve our understanding of anisotropy in heliospheric and magnetospheric turbulence.

Limitations of the correlation function

It has become increasingly clear in solar wind turbulence studies that the turbulence is not homogeneous – that is, that turbulent energy is distributed unevenly throughout the medium. This is related to the presence of structures in the fluid, as discussed earlier. A key assumption of the calculation of a correlation function is that the fluctuations are spatially homogeneous, but this is often not the case in MHD turbulence and care must be used to select intervals without large scale structures. However, the methods of combining data from multiple spacecraft discussed in this paper can be applied to more complex techniques, such as higher order correlation functions or structure functions (see, for example, Marsch and Tu, 1997 for a recent review of higher order observations), which can be used to quantify spatial inhomogeneity.

Cluster II measurements of the correlation function

The use of multiple measuring points to determine properties of three dimensional vector and scalar fields is clearly a general one. In this section we review some of the most important concepts and emphasise applications to Cluster II.

Earth-orbiting and interplanetary spacecraft already provide multi-point measurements in the sense that the movement of either the spacecraft or the underlying medium effectively turn measured time series into spatial samples using Taylor's hypothesis, discussed earlier. However, the obvious limitation of a one

dimensional sample is that variations in only one direction can be studied.

Size and shape of the formation

The four Cluster II spacecraft are expected to operate at separations ranging from a few hundred to 18000 km, but typically around 1000 to 2000 km apart. Four points form a tetrahedron, but the shape and size of the tetrahedron will change along the Cluster II orbit and as a result of trajectory manoeuvres. The effects of changes in the shape of the formation have been extensively considered by a variety of authors: see, for example, Robert *et al.* (1998).

In general, for reasonably regularly shaped tetrahedra, the distances between all the spacecraft are comparable, to within a factor of 2. The formation as a whole is therefore sensitive to variations on these scales. Variations on much larger scales are inaccessible, unless carried across the formation by a flow or propagating across the formation. Even then, only variations in the flow or propagation direction can be measured, as for a single spacecraft.

Similarly, variations on scales much smaller than the formation size cannot be accurately measured, and such variations can appear to some analysis tools (for example, discontinuity normal measurements) as noise.

This strong scale selection effect limits the range of phenomena which Cluster II can measure. For example, the separation scale will often be comparable to the proton gyroscale, so the medium on these scales cannot be well described as a magnetofluid. Of course, the transition from magnetohydrodynamic (MHD) to kinetic behaviour, which occurs near the proton gyroscale, is an important one and one that is important to turbulent processes, but a consequence of this separation scale is that some assumptions of solar wind turbulence analysis, which is often undertaken at MHD scales, may not be valid.

Constructing a correlation function in three dimensions

As we have seen, with a single spacecraft one can construct a correlation function in only one direction, aligned with the flow. With more than one spacecraft it is possible to sample other directions. As a fast moving plasma flows over the spacecraft formation, each spacecraft will sample a line through the medium, producing a linear spatial sample. With four spacecraft, a long time series is effectively a sampled "tube" through the flow (Figure 1).

By comparing data points from different spacecraft, it is possible to calculate differences in the measured parameter (for example, the magnetic field vector) in directions other than the flow vector. Figure 2 shows two formation configurations projected onto the plane perpendicular to the flow direction: we consider variations perpendicular to this plane in the next section but for now assume that all the spacecraft lie in the flow-perpendicular plane. The lines joining the spacecraft are the vectors between them, along which variations in parameters can be measured. These vectors

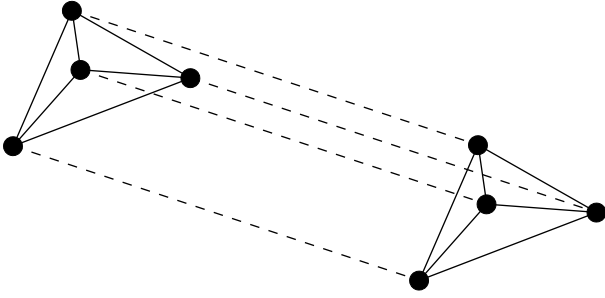


Figure 1. Flow across the Cluster II formation produces a “tube” swept through the plasma.

are shown in the right hand panels of Figure 2. Their lengths are separations between the spacecraft.

One would like as comprehensive coverage in angle as possible to best measure the correlation function. It is clear that the top configuration in Figure 2 has less regular angle coverage than the bottom formation. This is despite the fact that they are projections of almost the same configuration, simply at different orientations. The lower configuration, while it has better angular coverage, has large variations in the lengths of the vectors between the spacecraft. Therefore different scales are measured in different directions. The upper configuration is close to a square, so although all the separation vectors are of similar lengths, two pairs of vectors point in approximately the same direction, producing the uneven angle coverage.

Figure 2 shows that there are important consequences for angle and scale coverage of not only the size and shape of the formation, but also the orientation. There is no “ideal” shape: better angle coverage results in worse scale consistency and vice versa. As discussed in the next section, scale variations may be removed by estimating power scalings, so it may be better to maximise angle coverage by choosing a formation projection similar to the lower example in Figure 2 and compensating for scale changes.

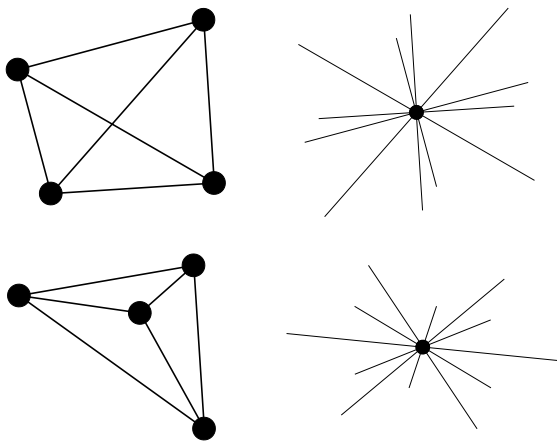


Figure 2. Two possible spacecraft formations projected perpendicular to the flow direction (left); corresponding comparison angles and minimum separations perpendicular to the flow (right).

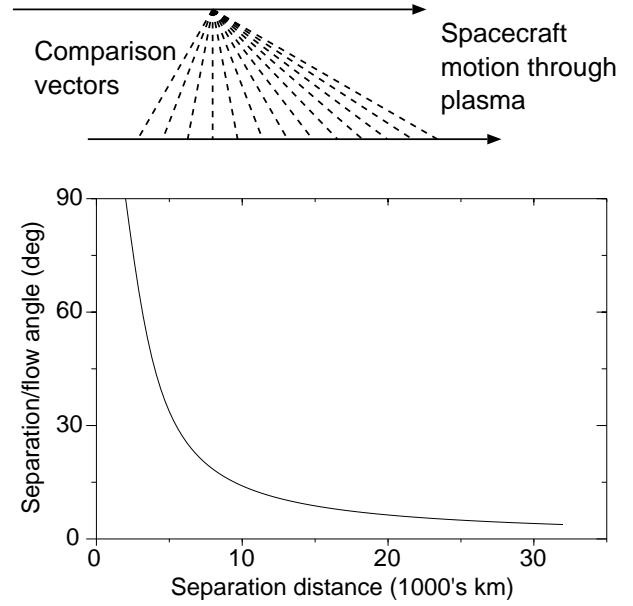


Figure 3. Top panel: schematic of angle/separation relation when comparing data from two spacecraft. Bottom panel: Angle/separation dependence for a flow-perpendicular separation of 2000km.

Variations perpendicular to the flow

We now consider the effect of comparing data points taken at different times. In the previous section, we assumed that the formation was planar perpendicular to the flow. Of course, this will not be the case in practice. However, since using Taylor’s hypothesis one can convert time series into spatial samples, this is not a restriction in general. With four spacecraft, we have four linear samples through the flow. Any two data points on the same or different spacecraft can be compared. Each pair of points is joined by a vector separation *in the plasma frame*.

Figure 3 illustrates comparisons between two spacecraft. The top panel shows the motion of two spacecraft relative to the plasma: their separation is in the plane of the Figure. Dashed lines show comparisons between one data point from one spacecraft and several samples from the other. It is clear that there is a dependence of the angle of the separation vector on its length, as shown in the bottom panel. The separation length is the scale on which spatial variations are sampled, so the unfortunate result of this angle/scale dependence is that one cannot measure any angle at more than one scale, or any scale at more than (at most) two angles.

While the angle/scale dependence is a significant problem, it is likely that it can be largely overcome. Figure 3 shows that around 60° of the separation/flow angle can be sampled with around a factor of 2 change in separation distance. With a single spacecraft one can measure the scale dependence of the correlation function in the flow direction. In general, this is a power law function of scale, at least in the MHD regime for

inertial range turbulence – although this is not the case near the proton cyclotron scale. Given the scale dependence of the correlation function in one direction, and assuming that this is the same in every direction (not necessarily the case), one can “normalise” each separation vector to the same separation length and thereby measure the angle variation in the correlation function.

The comparison vectors between spacecraft shown in Figure 3 all lie in a plane which joins the two spacecraft motion vectors through the plasma. Each of these planes forms one of the lines shown in the right panels of Figure 2. Therefore, the spacecraft separations define 6 planes over which comparisons can be made and the minimum separation distance in the direction along that plane perpendicular to the flow. This fully defines the set of possible angles over which correlation function estimates can be made.

The most convenient coordinate system in which to consider the set of sampled angles is one defined by two angles θ and ϕ and the separation distance r . θ is the flow/separation angle (as shown in Figure 3) and ϕ is the “clock angle” perpendicular to the flow (as shown in Figure 2).

The symmetry of the correlation function means that $R(\mathbf{r})=R(-\mathbf{r})$, simply because two sampled points can be compared in either direction. This symmetry of the correlation function means that each pair of data points contributes an estimate of the function in two opposing directions. This is the reason that there are 12, not 6, vectors in the right hand panels of Figure 2. However, because the two opposing vectors are in reality calculated from the same data, they are effectively identical. This effect is similar to that in a Fourier transform and in practice, one can display or calculate only half the correlation function with no loss of information.

Typical angle coverage

In practice, to convert many individual estimates of the correlation function at many separate angles into an approximation of the function over all angles can be most easily achieved by quantising the angles θ and ϕ into a set of “bins” and averaging all the correlation function estimates that lie within each bin. This procedure implicitly assumes that the correlation function varies smoothly on scales of the bin size, which may not be the case for realistic bin widths of around 10° . Bin size is a compromise between statistical uncertainties (sufficient data points per bin) and angular resolution. However, with only 4 spacecraft many angle bins will have no correlation function estimates within them. Each two spacecraft – there are 6 pair comparisons – results in samples within one plane, which corresponds to one ϕ value. Comparing different points in the two spacecraft time series varies θ , but not ϕ .

Figure 4 shows the angle coverage for a simulated spacecraft formation, for 50 seconds of $\frac{1}{2}$ s sampled data in a flow of 300 km/s. The Cluster II formation for

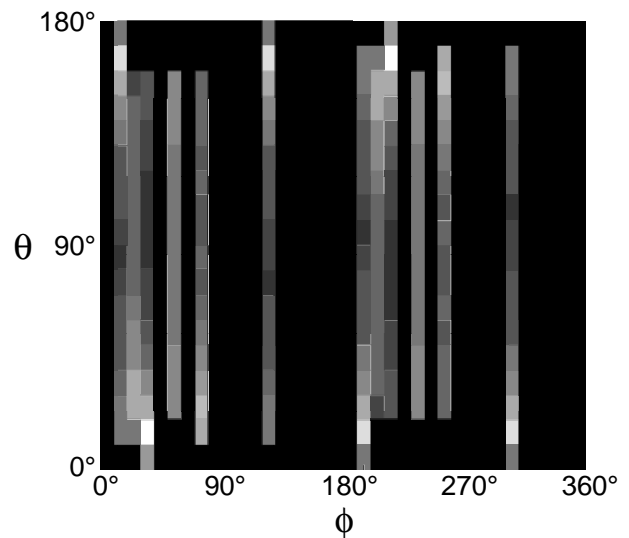


Figure 4. Typical angle coverage using two spacecraft, 300 km/s solar wind and 50s of 2 samples/s magnetic field data, with a typical flow-perpendicular separation of 2000km. Black regions have no coverage; white regions have most coverage.

this example was randomly generated, with typical inter-spacecraft distances around 2000 km. The number of pairs of data points corresponding to a given angle is shown on a grey scale. Black areas of the figure have no coverage, while white areas have most. The angles have been quantised into 18 bins in θ and 36 in ϕ . Each pair of spacecraft contributes a vertical line in the figure: this is one plane of comparisons, with varying θ values. There are 12 vertical strips because of the symmetry of the correlation process. The lowest number of points in any populated bin is 10 in this example; the highest is 3770.

The angle coverage in Figure 4 is sparse and there are large regions with no data points. However, the coverage is probably sufficient to produce a sufficiently accurate estimate of the total correlation function for several applications. One could, for example assume that the function varies as a sinusoidal function of the angles.

The angle coverage achieved is a strong function of the Cluster II formation shape and orientation, as discussed above. Randomly generated formations generally result in rather poor angle coverage: Figure 4 shows a more regular example than average. However, it is likely that the formation will be rather regular in practice, due to active management of the spacings for sampling targets of choice in the magnetosphere.

Data set length

Comparing data points from two spacecraft taken at very different times will result in a large separation between their sample locations in the plasma in the flow direction. In this case, the separation vector will be nearly in the flow direction. This is essentially the single spacecraft case, since the flow-perpendicular separation is much smaller than the flow-parallel spacing.

In general, points separated by a distance in the plasma frame that is much greater than the typical inter-spacecraft distance perpendicular to the flow will be nearly flow-aligned. This is not a useful case in the sense that such comparisons can already be made with a single spacecraft and they do not increase angle coverage. One can reasonably restrict data sets to spatial scales of a few times the formation scale, therefore. For typical Cluster II separation distances of a few thousand km, one needs only about 10^5 km of data in the flow direction. In the solar wind, with a typical flow speed of around 300 km/s, this corresponds to around 30s of data, a very small sample compared to typical interval lengths of single spacecraft data used for turbulence analysis. This short interval length, over which one can build up good angle coverage, means that it should prove possible to calculate “snapshot” correlation functions in different regions, for example immediately upstream and downstream of the bow shock, or at several points in the magnetosheath to measure turbulent development. The ability to estimate, albeit statistically, the three dimensional structure of the plasma on such short time scales is likely to greatly help the analysis of several solar wind and magnetospheric phenomena.

An implicit assumption of the methods described here is that the fluctuations are homogeneous over the length of the data interval. The short duration of the necessary data set means that this is often likely to be the case, even in highly structured environments.

Avoiding oversampling

The comparison of multiple data points as discussed above leads to oversampling of the data: individual points are used repeatedly. Oversampling should be avoided, but this can only be achieved by using any given data point at most once. It is not clear at this time how best this can be achieved, although a simple random method may be sufficient. Using this technique, random data points would be selected from the total available data from all spacecraft, and a correlation estimate and separation calculated, then placed in the correct (θ, ϕ) bin. The two points used would then be discarded and another two selected, this procedure being repeated until no more pairs of data points were available.

More complex methods are certainly possible. It may be that selecting pairs on the basis of gaps in the angle coverage will prove more useful. More advanced techniques are likely to be developed after studying data intervals, rather than before launch.

Taylor’s hypothesis and Cluster II

We have previously discussed Taylor’s hypothesis in the context of a single spacecraft. In this case, we simply require $V_{flow} \gg V_{wave}$. For multiple spacecraft, the principle is the same but the condition is more complex to calculate and rather more restrictive. Indeed, we show in this section that for realistic formation configurations and plasma conditions, Taylor’s hypothesis will not be satisfied for some angle ranges.

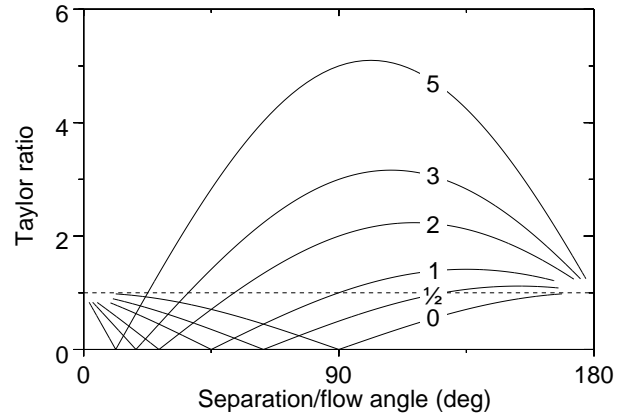


Figure 5. The “Taylor ratio” for several flow parallel/perpendicular spacecraft separations, as functions of the separation/flow angle θ . A Taylor ratio of 1 corresponds to the single spacecraft case. Taylor’s hypothesis is only valid for angles where the ratio is much less than the ratio of the plasma flow speed to the local wave speed.

For multiple spacecraft as for just one, we still require that the time between samples being taken be much less than the period of a wave with a wavelength of the separation between the two samples *in the plasma frame*.

Consider the case of two spacecraft, separated by a distance x in the plasma flow direction and y perpendicular to this. If we compare data samples taken at the same moment by both spacecraft, they are a distance $r=(x^2+y^2)^{1/2}$ apart in the fluid. Since they are taken at the same time, $t_{SC}=0$, while $t_\lambda=r/V_{wave}$. The ratio of these times, t_{SC}/t_λ , is zero, and so Taylor’s hypothesis of, of course, well satisfied.

If, however, we take a data point from the upstream spacecraft at a time $\delta t=x/V_{flow}$ before that in the downstream spacecraft, the separation in the flow direction between these two points *in the plasma frame* is zero, so the separation distance $r=y$. In this case, $t_{SC}=\delta t$ and $t_\lambda=r/V_{wave}$, so $t_{SC}/t_\lambda=(x/r)\cdot(V_{wave}/V_{flow})$, or, $t_{SC}/t_\lambda=(x/y)\cdot(V_{wave}/V_{flow})$. Compare this to the single spacecraft condition, which is simply given by the ratio $t_{SC}/t_\lambda=(V_{wave}/V_{flow})$. In both cases, we require this ratio to be much less than 1 for Taylor’s hypothesis to be satisfied. The additional factor x/y can be much greater than 1, invalidating Taylor’s hypothesis.

The two cases above are extreme examples. In general, when comparing two data points taken at different times on two spacecraft, the sampling points in the plasma frame will be separated by a flow-parallel distance x and a flow-perpendicular distance y and hence a total distance $r=(x^2+y^2)^{1/2}$. The flow-parallel distance will not be the flow-parallel spacecraft separation in general.

We refer to the ratio x/r , the ratio of flow-parallel and total plasma frame separations, as “Taylor’s ratio” in this paper. For Taylor’s hypothesis to be valid, we require this ratio to be much less than the ratio V_{wave}/V_{flow} , the single spacecraft condition for Taylor’s

hypothesis. For a given spacecraft separation, one can change the angle between the plasma frame separation and the flow, θ , and hence this ratio, by selecting data point pairs with different time delays t_{SC} .

Figure 5 shows values of Taylor's ratio for a number of ratios of the flow parallel to flow perpendicular spacecraft separation distances (equivalently, the flow-spacecraft separation angle), for values of the plasma separation/flow angle, θ . Variations in θ correspond to changing the time lag between the compared data points. For each separation angle, there is one value of θ where the ratio is zero. This occurs when θ is the same as the spacecraft separation angle: one can measure variations along this angle with no time lag between them.

It is clear from Figure 5 that, for formations that are anything other than exactly planar and perpendicular to the flow vector, there will be a range of angles where Taylor's hypothesis is less well satisfied than the single spacecraft case, although it will be better satisfied for some other angles. For flow parallel to perpendicular spacecraft separation ratios above 2, Taylor's ratio can greatly exceed 1 for some angles. In the solar wind, the ratio of flow to wave speeds is typically 5 to 10. In many regions of the magnetosphere it is well below 1, but in some regions, such as the magnetosheath, it is above 1 but less than 5. In these situations, care must be taken to ensure that Taylor's hypothesis is satisfied for each angle considered and it may be necessary not to consider some ranges of angles in some situations. In contrast, there are some angles (around the spacecraft separation angle) where Taylor's hypothesis will be satisfied for two spacecraft when it is not for one, making at least some spatial sampling of the correlation function possible in slow-flowing regions such as the cusp.

Field-aligned anisotropy

We have discussed measurement of the correlation function in terms of a flow-aligned coordinate system. However, it is well known that in general, on MHD scales turbulent magnetic field fluctuations tend to be perpendicular to the local background magnetic field direction (e.g. Horbury *et al.*, 1995). For the specific application of studying MHD turbulence, therefore, it may prove useful to consider variations in a coordinate system that is aligned with the background magnetic field. If the background field is constant over the interval of interest, this is a simple coordinate transformation. If the background field varies, however, it is possible to use a wavelet method to construct a dynamic coordinate system that tracks changes in the local background magnetic field. This is useful because small scale turbulent fluctuations "ride" on larger scale variations in the magnetic field. Changes in the background field can therefore allow the Cluster II formation to measure fluctuations in different orientations relative to the field, without changing the spacecraft positions. Effectively, this "blurs" the coverage shown in Figure 4 to cover a wider range of angles. Oughton *et al.* (1997) showed that the form of

the two point correlation function of MHD turbulence is restricted and it may prove possible to fit Cluster II observations to this form to produce an estimate of the full correlation function in a form that is usable for other applications, for example cosmic ray diffusion.

We do not describe the technique in detail here, but merely note that such methods are likely to prove useful in the analysis of Cluster II data.

Additional spacecraft

There are six vectors linking four spacecraft, while for five spacecraft there are 10 such vectors. For good angle coverage, it is preferable to have as many comparison vectors as possible. As a result, one would like to use data from additional Earth-orbiting spacecraft to increase the number of comparison vectors. However, such spacecraft are likely to pass considerably further from the formation than the inter-spacecraft distance within the cluster. If this separation is largely perpendicular to the local flow direction, then the minimum separation distance of the spacecraft to any of the Cluster II spacecraft will be much larger than the inter-Cluster distances. Since the separation vectors from all the Cluster II spacecraft to the additional craft will be in approximately the same direction, such an arrangement will not provide significant additional information: in effect, this scenario treats the Cluster II formation as one spacecraft.

If the separation is largely in the flow direction, then one can use time-delayed comparisons: effectively, waiting for plasma passing over the single spacecraft to propagate towards the Cluster II formation, or vice versa. When the flow-perpendicular separation is comparable to the tetrahedron scale, the additional spacecraft can act like a member of the Cluster II formation and provide extra information. However, if the spacecraft is far upstream, Taylor's hypothesis will be invalid because the time for the plasma to flow from the formation to the additional craft will be larger than the time for the fluctuations to change significantly, making the interpretation of the resulting correlation function considerably harder.

A possible use for additional spacecraft data can occur if the position of the formation is such that magnetic field lines that pass through the formation also pass over, or close to, the extra craft. In this case, shown in Figure 6, the four separation vectors between the Cluster II formation and the additional craft are all in approximately the same direction, and using data taken earlier or later at the formation one can find vectors that are approximately parallel or anti-parallel to the mean magnetic field direction. Then, the four separation vectors from the formation can be used to study variations in, for example, the correlation function at angles close to the mean field direction. The mean field direction is often poorly sampled in single or multiple spacecraft correlations. However, this is the case because angles within, for example 10° of a given direction cover a small solid angle, making them

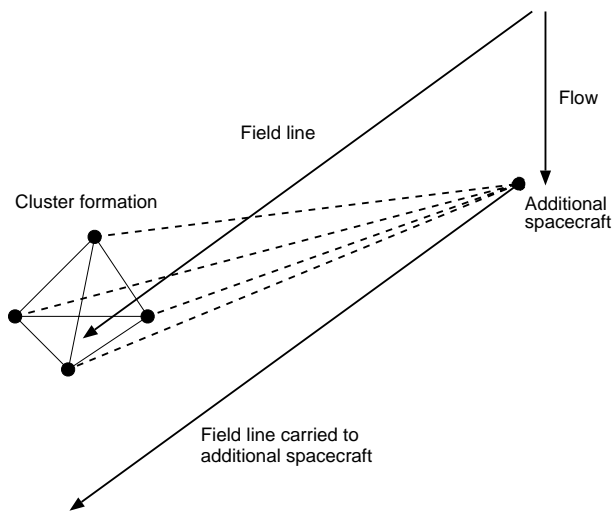


Figure 6. Using an additional spacecraft to study anisotropy close to the mean magnetic field direction. With correct placement of the extra spacecraft, field lines passing over the Cluster II formation also pass over the spacecraft at an earlier or later time. Comparison vectors from the extra spacecraft to the formation allow measurements of the variation in fluctuations in 4 directions close to the mean field.

unlikely to be sampled: separation vectors rarely point in the desired direction. Of course, this effect will also affect the scenario described above and it is not obvious if usable intervals of data with spacecraft encounters and a correctly pointed mean magnetic field will occur during the lifetime of the Cluster II mission. However, the fact that only a minute of data is needed, and that the additional spacecraft is likely to move slowly over a large angular range relative to the Cluster II formation, make the scenario possible.

Limitations of Cluster II

While it is clear that the Cluster II mission will greatly improve our understanding of a wide variety of microscale processes throughout the magnetosphere and the solar wind, it is important to stress that many questions will remain unanswered.

For a complete understanding of magnetospheric processes, one would ideally sample the entire magnetosphere and immediately upstream solar wind at a sufficiently fine scale, with high time resolution and high precision. For some applications, a smaller volume is sufficient, but it is clear that the reality of the available data will be much worse: the medium will be grossly undersampled. The way forward is obviously to make assumptions of some sort, which will vary with the problem of interest. However, it is important to stress that any analysis of spacecraft data involves assumptions of some kind and Cluster II will be no exception.

As we have seen, four spacecraft provide only 6 planes in which to make comparisons in a supersonic flow. It is not clear whether one can extrapolate usefully from

these 6 planes to the entire unit sphere, for example when estimating the correlation function. A field-aligned analysis method, as described in this paper, can help, but once again this introduces an assumption (that the field direction is an axis of symmetry of the fluctuations) that may not be justified.

An obvious problem that may be difficult to overcome is that of inter-spacecraft calibration. Taking the magnetic field as an example, on typical Cluster II spacecraft separations of a few thousand km, variations in the magnetic field can be around 100pT, but sometimes even less. To reliably compare measurements between instruments therefore requires an accuracy of around 10 pT in each instrument. This is near the limit of what can be achieved in practice, although these problems can be alleviated with techniques such as background removal. For bulk plasma measurements, the precision may be even more demanding relative to what is usually achieved. As a result we expect that inter-spacecraft offsets will be a significant issue in Cluster II data analysis. In addition, combining Cluster II data with additional spacecraft, with different instrumentation, will be particularly challenging.

An issue related to inter-spacecraft calibration is that of instrument precision: for example, the Cluster II magnetometer resolution is 8pT (Balogh *et al.*, 1995), about 10% of typical variations on the scales of interest. Typically, bulk plasma instruments can measure velocities to within a few percent, and if one is measuring variations on small scales, where the fluctuations are well correlated so that spatial changes are small, these changes will be close to the instrument precision and any variations seen will be either quantised or strongly affected by noise. While this is a problem with even single spacecraft, the combination of such noisy signals will make the problem more acute. Again, such problems will have to be carefully considered when analysing Cluster II data, although with careful data analysis it is likely that they can be largely overcome.

A further instrumental limitation is that of time resolution. To construct a correlation function using data points from several spacecraft, at least when Taylor's hypothesis can be used, the time between samples needs to be comparable to the time taken for the plasma to travel across the formation. For example, with a separation of 3000km and a flow speed of 300 km/s, this travel time is 10s. The data sampling time must be a small fraction of this time to achieve good angle resolution when comparing data between spacecraft. Plasma moment data produced at spin time resolution, approximately 4s, may not be sufficient for this application, and although it may be possible to produce plasma data at a higher time resolution this is unlikely to be a routine procedure. Elsässer variables, which require combined velocity, density and magnetic field data, can only be produced at the time resolution of the plasma data and are therefore unlikely to be produced at resolutions above 4s. Consequently, multi-spacecraft Elsässer variable comparisons are unlikely to be achieved with Cluster II data without a significant

analysis effort. The importance of Elsässer variables in the study of plasma turbulence means that this effort is likely to be made.

An additional restriction of plasma moment data is that, in contrast to magnetic field data, they are produced over an extended time as detectors scan over look directions and energies. It is therefore impossible to give a precise time stamp to each plasma moment. When these moments are generated over times that are significant fractions of the time for plasma to flow over the formation, it is not possible to spatially localise each measurement with sufficient precision to make accurate comparisons between spacecraft. With four spacecraft, each spinning at slightly different rates and hence with different spin phases at any given time, the comparison of high time resolution plasma moments from one spacecraft to another will be extremely difficult. It may prove possible to recover consistent plasma differences between the spacecraft, but this is likely to be a slow and complex process which will be achieved only with considerable effort and using instrument expertise which will only be gained during in-orbit operations.

Summary

The Cluster II mission offers the prospect of greatly increasing our understanding of small scale plasma processes. With four spacecraft, analysis of the three dimensional structure of both structured and turbulent plasmas can be performed rigorously for the first time. However, considerable challenges remain. Instrumental effects are likely to be important at such small scales and considerable care will be needed to ensure that reliable inter-spacecraft comparisons can be made at such small scales.

Acknowledgements

The author is grateful to the organisers of the Cluster II workshop, A. Balogh and C. P. Escoubet, for the opportunity to present this paper and to D. Burgess and S. J. Schwartz for useful discussions. This work was supported by PPARC(UK) grant GR/L29347.

References

Bavassano B., M. Dobrowolny, F. Mariani and N. F. Ness, Radial evolution of power spectra of interplanetary Alfvénic turbulence, *J. Geophys. Res.*, **87**, 3617–3622, 1982

Bieber J. W., W. Wanner and W. H. Matthaeus, Dominant two-dimensional solar wind turbulence with implications for cosmic ray transport, *J. Geophys. Res.*, **101**, 2511–2522, 1996

Balogh A., M. W. Dunlop, S. W. H. Cowley, D. J. Southwood, J. G. Thomlinson, K. H. Glassmeier, G. Musmann, H. Luhr, S. Buchert, M. H. Acuna, D. H. Fairfield, J. A. Slavin, W. Riedler, K. Schwingenschuh and M. G. Kivelson, The Cluster magnetic field investigation, *Space Sci. Rev.*, **79**, 65–91, 1997

Dunlop M. W. and T. I. Woodward, Multi-spacecraft discontinuity analysis: orientation and motion, in *Analysis methods for multi-spacecraft data*, ed. G. Paschmann and P. W. Daly, International Space Science Institute, Bern, 1998

Escoubet C. P., R. Schmidt and M. L. Goldstein, Cluster - Science and mission overview, *Space Sci. Rev.*, **79**, 11–32, 1997

Frisch U., *Turbulence*, Cambridge University Press, Cambridge, 1995

Goldstein M. L., Magnetohydrodynamic turbulence in the solar wind, 137–148, *ESA SP-371: Proceedings of the Cluster Workshops*, ESA Publications Division, Noordwijk, 1995

Goldstein M. L. and D. A. Roberts, Magnetohydrodynamic turbulence in the solar wind, *Ann. Rev. Astron. Astrophys.*, **33**, 283–325, 1995

Horbury T. S., A. Balogh, R. J. Forsyth and E. J. Smith, Anisotropy of inertial range turbulence in the polar heliosphere, *Geophys. Res. Lett.*, **22**, 3405–3408, 1995

Marsch E., Analysis of MHD turbulence: spectra of invariants, structure functions and intermittency scalings, 107–118 *ESA SP-371: Proceedings of the Cluster Workshops*, ESA Publications Division, Noordwijk, 1995

Marsch E. and C.-Y. Tu, Intermittency, non-Gaussian statistics and fractal scaling of MHD fluctuations in the solar wind, *Nonlinear Processes in Geophysics*, **4**, 101–124, 1997

Matthaeus W. H., J. W. Bieber and G. P. Zank, Unquiet on any front - anisotropic turbulence in the solar wind, *Rev. Geophys.*, **33**, 609–614, 1995

Oughton S., K.-H. Radler and W. H. Matthaeus, General second-rank correlation tensors for homogeneous magnetohydrodynamic turbulence, *Phys. Rev. E*, **56**, 2875–2888, 1997

Paschmann G. and P. W. Daly, *Analysis methods for multi-spacecraft data*, International Space Science Institute, Bern, 1998

Pinçon J.-L. and U. Motschmann, Multi-spacecraft filtering: general framework, in *Analysis methods for multi-spacecraft data*, ed. G. Paschmann and P. W. Daly, International Space Science Institute, Bern, 1998

Potgieter M. S., The modulation of galactic cosmic rays in the heliosphere: theory and models, *Space Sci. Rev.*, **83**, 147–158, 1998

Robert P., A. Roux, C. C. Harvey, M. W. Dunlop, P. W. Daly and K.-H. Glassmeier, Tetrahedron geometric factors, in *Analysis methods for multi-spacecraft data*, ed. G. Paschmann and P. W. Daly, International Space Science Institute, Bern, 1998

Taylor G. I., The spectrum of turbulence, *Proc. R. Soc. Lond.*, **A164**, 476–490, 1938, reproduced in *Turbulence - classic papers on statistical theory*, ed. S. K. Friedlander and L. Topper, Interscience, New York, 1961

Tu C.-Y. and E. Marsch, MHD structures, waves and turbulence in the solar wind - observations and theories, *Space Sci. Rev.*, **73**, 1–210, 1995