

TIME SERIES DATA ANALYSES IN SPACE PHYSICS

P. SONG^{1,2} and C. T. RUSSELL²

¹ *Space Physics Research Laboratory, The University of Michigan, Ann Arbor, MI 48109, U.S.A.;*

E-mail: pson@engin.umich.edu

² *Institute of Geophysics and Planetary Physics, University of California, Los Angeles, CA 90095, U.S.A.;*

E-mail: ctrussell@igpp.ucla.edu

(Received 24 July, 1997; accepted 6 February, 1998)

Abstract. Techniques of time series data analyses developed over the past decades are reviewed. We discuss the theoretical principles and mathematical descriptions of these analytical techniques that have been developed by scientists with different backgrounds and perspectives. These principles not only provide the guidelines to evaluate each particular technique but also point to directions for the development of new methods. Most time series analyses can be divided into three categories: discontinuity analysis, wave analysis and correlation analysis. Techniques for analyzing one-dimensional discontinuities have been well-developed and tested. The errors and ambiguities of discontinuity analyses are reasonably well, but not as widely, understood. Techniques for wave analyses have been developed for certain wave properties and are still under further active development. Problems in using these techniques are recognized to a certain extent. Because of the complicity of the waves in space and the limitation of probing, there are significant needs for the development of new methods. Although simple techniques for two-satellite correlation analyses have been developed and tested for some time, techniques for multiple satellites are in an embryonic stage. We expect to see significant advances in the development of new techniques and new concepts. We believe, however, that the problems in this area have not been fully appreciated.

Table of Contents

1. Introduction
 - 1.1. Frame of Reference
 - 1.2. Coordinate Systems
 - 1.3. Measurements and Their Uncertainties
 - 1.4. Principal Axis Analysis
2. Discontinuity Analyses
 - 2.1. Background
 - 2.2. Minimum Variance Analysis
 - 2.3. Tangential Discontinuity Analysis
 - 2.4. Coplanarity Analysis
 - 2.5. Maximum Variance Analysis
 - 2.6. DeHoffmann–Teller Frame and Walen Relation Test
 - 2.7. Suggested Procedures for Discontinuity Analysis
 - 2.8. Rankine–Hugoniot Relations Test
3. Wave Analyses
 - 3.1. Background



Space Science Reviews **87**: 387–463, 1999.

© 1999 Kluwer Academic Publishers. Printed in the Netherlands.

- 3.2. Routine Wave Analysis
- 3.3. Mode Identification
- 3.4. Frequency and Phase Velocity
- 3.5. Nonlinear Effects
- 3.6. Plasma Wave Analysis
- 4. Spatial Correlation Analyses
 - 4.1. Background
 - 4.2. One-Dimensional Discontinuity Analyses
 - 4.3. Two-Dimensional Structures
 - 4.4. Determination of Electric Currents
 - 4.5. Wave Analyses
- 5. Discussion and Summary

1. Introduction

Many of the observational data in space physics are obtained as time series. These time series contain measured physical quantities, that may be scalars, vectors, tensors, or multidimensional images. Examples of such quantities are temperatures and densities of the plasma, magnetic or electric field vectors, pressure tensors, and auroral images. These quantities can be measured either in space on a moving platform or on a fixed platform such as the surface of the Earth. Thus the variations in the time series may represent true temporal changes in the system or motion through spatial gradients or some combination of the two. Since most data gathered in space physics are initially in the form of time series, their use is widespread. To analyze these time series data, many data processing methods, analysis techniques and computer algorithms have been developed. In this review, we outline a set of principles for data analysis methods, describe a number of well-established data analysis techniques, discuss the uncertainties and limitations of each technique, and suggest procedures and criteria which may reduce the uncertainties of the results for some analyses. Many of the principles presented are derived for the first time.

Most time-series analyses can be divided into three categories: discontinuity analysis, wave analysis, and correlation analysis. Discontinuity analysis determines the orientation, thickness and motion of the interface between two different plasma regions or regimes. The methods for discontinuity analysis are well developed. However, due to the lack of wide recognition of the underlying assumptions, uncertainties and validity of each method, there are still many problems in this area. We present a comprehensive discussion of these problems. Wave analysis determines the properties and characteristics of a wave that identify which of the several possible wave modes allowed in a plasma a particular observed fluctuation might be. Techniques for wave analysis are under active development. We discuss the principles of these methods. Since discontinuities can be considered as steepened waves, some of the methods for wave analysis can be used in discontinuity analysis. Correlation analysis determines the relationship between observations at two or more spatially separated locations. Correlation studies can be performed in the

time domain, for example to determine the time lags between observers. Correlation analysis can also be performed in the frequency domain. We will give a brief introduction to correlation analysis and discuss how to apply it to the analysis of the data from a cluster of satellites.

From the beginning of the space age, scientists have sought means for quick qualitative visual examination of large amounts of data. With plotted traces, one can easily spot a discontinuity, estimate a wave frequency and correlate two traces by overlaying them. Even today these qualitative visual analyses can be powerful ways to check the results of an analysis, i.e., a quantitative result is suspicious if one cannot verify the consistency of the result by visual inspection of the data. As will be stressed in this paper, relying solely on automated computer analysis may lead to completely wrong results. As a data analyst, one has to frequently return to the examination of the original observations to perform what is often referred to as a 'sanity check' or a 'reality check'.

The first quantitative data analysis method for discontinuities was the minimum variance analysis for discontinuity analysis proposed by Sonnerup and Cahill (1967). In the 1970s, the wave analysis techniques using the magnetic field became mature. Along with the improvements of the plasma measurements, in the 1980s and 1990s, techniques incorporating plasma measurements have been under active development. The correlation analysis of signals made at two or more observation sites has become important since the launches of the ISEE satellites and it will play the key role in analyzing the multiple satellites data in the International Solar Terrestrial Physics (ISTP) program. While in theory the quantitative correlation analysis is most straightforward compared with other analyses, in practice, there are many problems associated with the coherence length of a phenomenon, compared with the spacecraft separation, for example. We suspect that many problems have not yet even been recognized. In the next few years, we expect to see major progress in understanding of correlation analyses as a result of multiple-satellite data analyses.

1.1. FRAME OF REFERENCE

Observational data are gathered in a frame of reference at rest with the observer. Physical laws are often stated in a frame of reference that is moving with respect to the observer. In particular, the Rankine–Hugoniot relations are given in the frame of reference at rest with the discontinuity, and the wave dispersion relations usually expressed in the plasma rest frame. In the non-relativistic limit, plasma density, pressure, magnetic field, and wavelength are not dependent on the frame of reference, but velocity, electric field, and frequency are. While it is always important to work in the appropriate frame of reference when dealing with a flowing plasma, this is critically true in the supersonic solar wind.

The total time derivative of an observed quantity Q , which can be either a scalar or a vector, is

$$\frac{DQ}{Dt} = \frac{\partial Q}{\partial t} + (\mathbf{v} \cdot \nabla)Q, \quad (1.1)$$

where \mathbf{v} is the velocity of the observer relative to the frame of reference in which the phenomena is described. The first term on the right is the temporal variation in the frame of reference of the plasma (say) and the second term on the right is the apparent temporal variation caused by motion through spatial gradients. Both are observed by a moving instrument as temporal variations. For example, when discussing observations of a wave in the plasma frame, \mathbf{v} is the measured plasma flow velocity, and the $\partial/\partial t$ term is nonzero. For a steady phenomenon, the $\partial/\partial t$ term is zero (e.g., in the plasma frame) and the observed time variation is due to the motion through stationary gradients. A time-domain correlation analysis assumes that, in a particular frame, for example, a wave frame or a shock frame, $\partial/\partial t = 0$. Therefore the velocity derived from the timing difference is the sum of the flow (convection) velocity and the propagation (phase) velocity. We recall that the group velocity measures the velocity of the envelope of the variations whereas the variations themselves move at the phase velocity. The energy of the wave packet flows with the group velocity and it is the group velocity that is restricted to velocities equal to or less than the speed of light, not the phase velocity.

For magnetic field measurements, using the frozen-in Faraday's law and divergence free condition, Equation (1.1) becomes,

$$\frac{D\mathbf{B}}{Dt} = (\mathbf{B} \cdot \nabla)\mathbf{v} - (\nabla \cdot \mathbf{v})\mathbf{B}. \quad (1.2)$$

Assume a coordinate system ℓ, m, n with variations only along the normal or a direction under one-dimensional assumption, for either a discontinuity or a wave, in which $\partial/\partial \ell$ and $\partial/\partial m$ are zero. It is easy to show that the measured time variation of the field is zero in the direction along n . No stationarity assumption is made here and this is true independent of the frame of reference. This particular feature of the magnetic field is actually the basis of most data analysis techniques. Note that in discontinuity analysis, one cannot always assume that the shock frame is known, nor in wave analysis, that the wave is in steady state.

For other quantities, temporal variations in the frame in which a theory is applied, may affect the results of an analysis. For example, the observed electric field variation is

$$\frac{D\mathbf{E}}{Dt} = \frac{\partial \mathbf{E}'}{\partial t} + \nabla(\mathbf{v} \cdot \mathbf{E}') - \frac{\partial \mathbf{v}}{\partial t} \times \mathbf{B}, \quad (1.3)$$

where $\mathbf{E}' = \mathbf{E} + \mathbf{v} \times \mathbf{B}$ is the electric field in the frame of reference in which the phenomenon is described. Note here that $\nabla \times \mathbf{E} = -\partial \mathbf{B}/\partial t$ and we have assumed \mathbf{v} to be uniform in space. When \mathbf{v} is not spatially uniform, more terms appear. Under a 1-D assumption, neither component is zero in general. But for 1-D steady state, the tangential components of the variation vanish whereas the normal component does not. This property of the electric field variation provides

the foundation of the maximum variance analysis to be discussed in Section 2.5. Note that this method requires the steady state assumption and thus is difficult to apply to wave analyses. For discontinuity analysis, not only unsteadiness of the discontinuity but also changes in the motion of the discontinuity relative to the observer affect the results. Since in space, boundaries are often in oscillation and not in simple motion, only on a few occasions is the latter effect unimportant. When comparing measurements from two spacecraft at different times, more effects will occur due to the relative motion between the spacecraft frame and the plasma frame (see more discussion in Section 4.1.3).

1.2. COORDINATE SYSTEMS

1.2.1. *Global Coordinate Systems*

Three global coordinate systems are often used in space physics. A global coordinate system is useful for studying phenomena of global effects.

1.2.1.1. *GSE coordinate system.* In the Geocentric-Solar-Ecliptic coordinate system, the x direction is from the Earth to the Sun. The z direction is along the normal to the ecliptic plane and pointing to the north. The y direction completes the right handed coordinate system and points to the east, opposite planetary motion. The statistical aberration of the solar wind by the Earth's orbital motion is most readily removed in this system. It is useful for problems in which the orientation of the Earth's dipole axis is not important, such as the bow shock and magnetosheath phenomena.

1.2.1.2. *GSM coordinate system.* In the Geocentric-Solar-Magnetospheric coordinate system, the x direction is also from the Earth to the Sun. The z direction lies in the plane containing the Sun–Earth line and the geomagnetic dipole, and is perpendicular to the Sun–Earth line and positive north of the Sun–Earth line. The y direction completes the right-handed coordinate system. The GSM y and z directions lie at an acute angle (around the x direction) to the GSE counterparts. The GSM coordinates are most useful for studies of solar wind-magnetosphere interaction and where the solar wind determines the geometry of the magnetosphere. These include phenomena near and at the magnetopause/boundary layer, in the outer magnetosphere and near earth magnetotail.

1.2.1.3. *SM coordinate system.* In the Solar Magnetic coordinate system, the z direction is along the Earth's magnetic dipole pointing north. The $x - z$ plane contains the solar direction with roughly toward the Sun. Its y direction is on the dawn-dusk meridian and points to dusk in the same direction as the GSM y direction. Thus the SM x and z directions differ from their GSM counterparts by an acute angle of rotation related to the tilt of the dipole. The SM coordinates are usually used in studies of the near earth phenomena, where the geomagnetic field

**BOUNDARY NORMAL
COORDINATES**

\hat{N} is normal to local
magnetopause surface.

$$\hat{M} = \frac{\hat{N} \times \hat{z}_{GSM}}{|\hat{N} \times \hat{z}_{GSM}|}$$

$$\hat{L} = \hat{M} \times \hat{N}$$

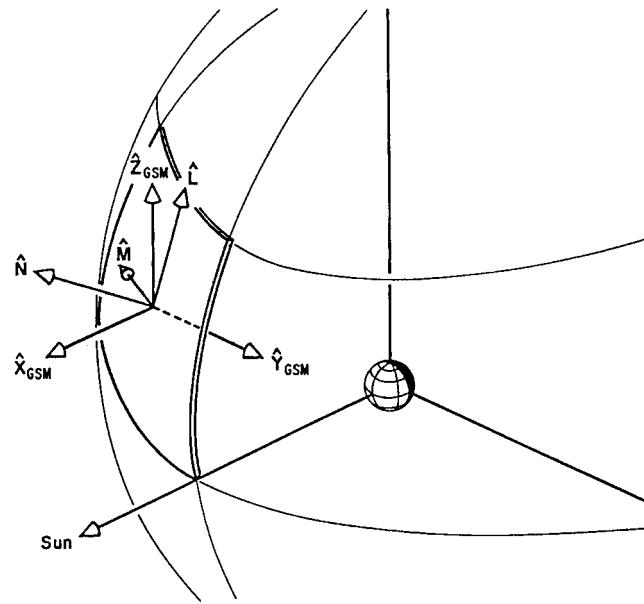


Figure 1.1. Boundary normal coordinate system (Russell and Elphic, 1978).

controls the processes, such as the ionosphere, inner magnetosphere, and ground observations.

1.2.2. Local Coordinate Systems

To study a local phenomenon or a global phenomenon in a local context, a local coordinate system is most convenient. Three local coordinate systems are often used in the literature.

1.2.2.1. Boundary normal coordinate system (LMN) (Russell and Elphic, 1978).

The \hat{N} direction is along the normal of the boundary, which can be the bow shock or the magnetopause, or a current sheet. The direction is usually chosen to be outward from the Earth for both the bow shock and the magnetopause. The \hat{L} direction is along the geomagnetic field direction for the magnetopause and is along the projection of the upstream magnetic field on the boundary for the bow shock. The \hat{M} direction completes the right-handed coordinate system and points to dawn for the magnetopause (see Figure 1.1). At the subsolar magnetopause, neglecting the aberration caused by the finite velocity (29.5 km s^{-1}) of the Earth through the solar wind, the LMN coordinates are coincident with the GSM coordinates with $\hat{N} \rightarrow \hat{x}$, $\hat{L} \rightarrow \hat{z}$ and $\hat{M} \rightarrow -\hat{y}$.

The LMN coordinate system is useful only locally unless the boundary is planar. In some cases, a satellite may move along a curved boundary, e.g., the magnetopause, for a long period of time and cross it many times as the boundary rocks back and forth. The normal of the boundary at each crossing may be different.

When presenting the data in a fixed LMN coordinate system, one needs to be extremely careful in interpreting the results because the direction of the normal of the boundary may be varying. Similarly one should be cautious in using an LMN coordinate system derived from either the magnetopause or bow shock well away from that boundary such as in presenting the data from the bow shock to the magnetopause. One should not interpret, for example, the N component along a direction determined at the magnetopause as the normal component throughout the magnetosheath.

1.2.2.2. *Field-aligned coordinate system.* The average magnetic field direction (taken from the *in situ* measurements) is defined as a preferential direction. A second direction is usually defined according to the symmetry of the system. For example, for magnetospheric wave studies, the azimuthal direction parallel to the direction obtained from the cross product of the magnetic field and the radially outward direction is usually used as the second direction. This direction is eastward in the Earth's magnetosphere in the direction of electron drift. This direction is useful for displaying the field perturbations associated with sheets of field aligned currents that are roughly along shells of constant L-value or L-shells. Similarly, for magnetosheath studies, one can define a surface in the sheath which belongs to the family curves that are used for empirical models of the magnetopause and bow shock. The second direction can then be defined perpendicular to the field and along the surface.

1.2.2.3. *Geomagnetic dipole coordinate system.* The Geomagnetic Dipole coordinate system has its z direction along the geomagnetic dipole axis. The other two directions are defined in terms of geomagnetic latitude and longitude by analogy with geographic coordinates. Therefore, this coordinate system is useful in studying phenomena observed in ionosphere and ground stations. In order to study the global effects of these phenomena, an observation is often presented in geomagnetic local time, which is equivalent to that in SM coordinates, and is the angle between the meridian plane containing the Sun and that containing the point of observation converted to hours and increased by 12 hours.

1.2.3. *Spacecraft Coordinate System*

For some vector quantities only two components are measured by our spacecraft. Examples are the electric field, plasma waves and the plasma moments from certain instruments. The two components are in the plane perpendicular to the spacecraft spin axis. Frequently, the spin axis is nearly along one of the GSE axes (for magnetospheric missions) or perpendicular to the orbit plane (cartwheel mode for ionospheric missions). The fraction and the direction of the background magnetic field projected onto this plane are important to observations which deal with the anisotropies with respect to the magnetic field. The spin modulation of the high

time resolution signals can provide useful information. Often artifacts in measurements are most easily detected in spacecraft coordinates.

1.3. MEASUREMENTS AND THEIR UNCERTAINTIES

A measured physical quantity is derived from measurable quantities by a instrument. Therefore, its accuracy is limited by the instrument capability and the data reduction procedures. As a data analyst, it is very important to understand the principles of the instruments from which the data are measured and the schemes of the data reduction that are used. In this section, we will discuss several general issues concerning the measurements and data reduction. We will assume that the measurements are accurate in other sections in the paper. The word ‘fluctuation’ used in this paper refers to deviations from the average of the measured quantity. The word ‘noise’ refers to the unwanted signals that are not described by the idealized variations of a particular analysis method. It is important to point out that the ‘noise’ in one analysis can be a physical phenomenon for another analysis. Fluctuations include both wanted signals and noise. If one chooses a wrong method, the useful signals will be treated as noise.

1.3.1. Resolutions

1.3.1.1. *Time resolution.* Temporal resolution, the time between samples, is a major limit in many data analysis methods. For example, to use the Minimum Variance method (to be discussed in Section 2.2) requires the resolution to be at least half of the crossing duration of a discontinuity, i.e., one needs a sample in the middle of the crossing not just at either end. Temporal resolution also sets an upper cutoff frequency, the so-called Nyquist frequency, which is half of the sampling frequency, for Fourier analysis. Stated differently, in order to resolve a waveform, the sampling frequency should be at least twice the wave frequency. If frequencies above the Nyquist frequency are present when a signal is digitized it appears to oscillate at the frequency below the Nyquist frequency. Such a signal is called aliased. The timing difference derived from correlation analysis is limited by the accuracy of the clocks controlling the measurements. Synchronization of clocks of observers becomes very important when the timing differences are small. Sometimes components of vectors are not sampled simultaneously. Sometimes the time separation of sequential measurements is not uniform. Techniques exist for accurately analyzing signals under these circumstances but add complexity to the analysis.

If the bandwidth of the measurements is large compared to the frequency range of the phenomenon of interest, i.e., if the temporal resolution is too high, time domain analysis (which accept signals of any frequency) will be affected by ‘unwanted signals.’ For example, if within a discontinuity, there are waves which are polarized in the normal direction of the discontinuity, they will affect the result of a Minimum Variance analysis. Although these high frequency phenomena have

their own physical significance they may be treated as 'noise' in analysis of low-frequency phenomena. Therefore, in time domain analysis proper preparation of the data by either running average or filtering is important to limit the band of information to the time scale of the phenomenon of interest.

1.3.1.2. *Amplitude resolution.* Digitized measurements have finite amplitude resolution as well as finite temporal resolution. The result of the process is to add a square wave, or finite steps to the data. When a Fourier analysis is performed by digitized data, there is a minimum noise level set by the digitization of $D^2/12F_N$, where F_N is the Nyquist frequency and D is the digital window. This digital noise is spread uniformly over the bandwidth of the signal, i.e., 0 to F_N and when decreasing the bandwidth of the digitized signal by subsampling or 'decimating' as it is sometime called, it is important to low-pass filter the signal, or else the entire digital noise of the original time series will be added to the new narrower bandwidth. If the digital noise is comparable to the measured signal, the noise is most readily reduced by digitizing more finely than increasing bandwidth (Russell, 1972).

1.3.2. *Field Measurements*

One of the most accurately determined quantities in space is the magnetic field, if attention is paid to considerations of linearity, magnetic cleanliness and bandwidth relative to the Nyquist frequency. For an elliptical Earth orbiter its magnitude can be calibrated each time at perigee where the field is known and the zero levels of each sensor in the spin plane can be calibrated as the spacecraft rotates (e.g., Kepko et al., 1996). The intercalibration among different spacecraft can be done in current-free regions (e.g., Khurana et al., 1996).

The electric field measurements are affected by spacecraft charge and Debye length. The understanding of the calibration problem has been significantly improved and the validity of the measurements has been tested (Mozer et al., 1979, 1983). Nevertheless, special caution should be taken in the regions with significant plasma gradients and low plasma density.

High frequency (>10 Hz) electric field fluctuations are measured with electric dipole antennae. Often only the components in the spin plane are measured. The reliability of the data is generally good. The major limitation is caused by the finite data rate which results in competition between the frequency and temporal resolutions. For example, to infer the plasma density from waves at the plasma frequency, one needs a fine frequency resolution, but to determine the polarization of a wave based on spin modulation, one needs a temporal resolution a few times the spacecraft spin period. The corresponding oscillating magnetic field is usually measured with wire coil antennae that are sensitive to the time derivative of the magnetic field. A search coil magnetometer is a coil antenna with a highly permeable core and is used at ELF and VLF frequencies, say 10 to 10^4 Hz.

1.3.3. Plasma Moments

A plasma detector measures the number of particles ΔN in an energy range between U and $U + \Delta U$ that tranverse in a time Δt an area ΔA within an element of solid angle $\Delta\Omega$ around the normal to A . The differential direction (flux) intensity is defined as

$$J = \Delta N / (\Delta A \Delta \Omega \Delta U \Delta t) \quad (1.4)$$

In principle, ΔA is the cross-section of the detector, $\Delta\Omega$ and ΔU are the angular and energy resolutions of the instrument, and Δt is the sampling time which equals the data rate for continuous measurements. To obtain a complete distribution function, the instrument has to scan all look-directions and energy ranges. Often the angular and energy resolutions (and the signal-noise ratio) compete for the limited telemetry.

The plasma distribution function f is related to J by ($\Delta N = f \Delta \mathbf{v} \Delta \mathbf{r}$, $U = \frac{1}{2} m v^2$, $\Delta \mathbf{r} = v \Delta A \Delta t$, $\Delta \mathbf{v} = v^2 \Delta \Omega \Delta v$),

$$f = \frac{m}{v^2} J, \quad (1.5)$$

where m and v are the mass and velocity of a particle being detected. With the distribution function, the plasma moments, density, velocity, pressure and heat flux (for a particular species), can be derived,

$$N = \sum_{v_{LC}}^{v_{HC}} \sum_{\Omega} f v^2 \Delta \Omega \Delta v, \quad (1.6)$$

$$\mathbf{V} = \frac{1}{N} \sum_{v_{LC}}^{v_{HC}} \sum_{\Omega} \mathbf{v} f v^2 \Delta \Omega \Delta v, \quad (1.7)$$

$$P = m \sum_{v_{LC}}^{v_{HC}} \sum_{\Omega} (\mathbf{v} - \mathbf{V})^2 f v^2 \Delta \Omega \Delta v, \quad (1.8)$$

$$T = P / N, \quad (1.9)$$

where v_{LC} and v_{HC} are the lower and higher cutoff velocities that are determined by the lower and higher cutoff energies, E_{LC} and E_{HC} , and the spacecraft potential Φ , and are

$$v_{LC,HC} = \frac{2}{m} \sqrt{E_{LC,HC} + q\Phi}, \quad (1.10)$$

where q is the electric charge of the particle. For a 2-D detector, an assumption is needed in order to extrapolate the values of fluxes in undetected directions. This assumption becomes important when the anisotropy of the plasma is high. Note

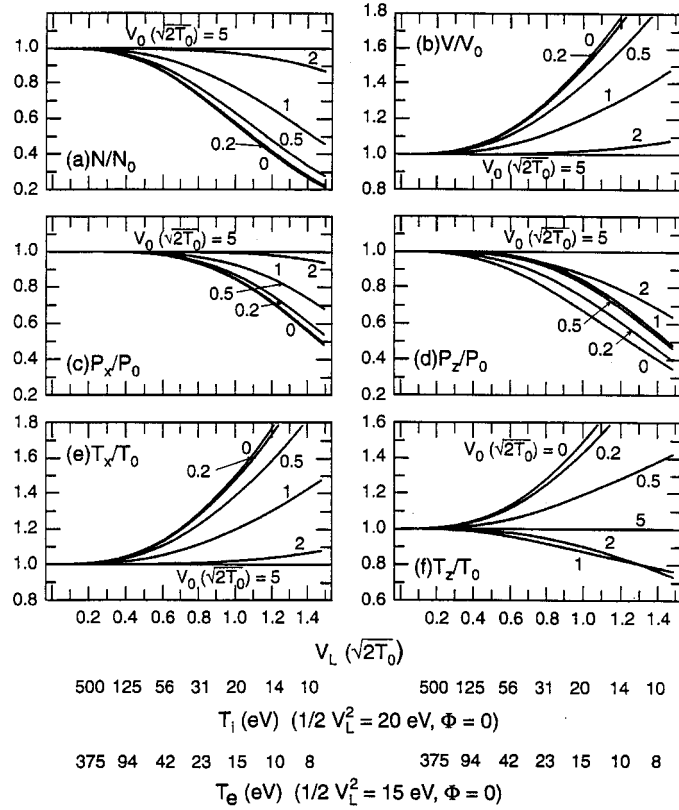


Figure 1.2. Ratios of the measured to real moments (Song et al., 1997). The lower cutoff velocity should be evaluated from the lower cutoff energy of the detector and the spacecraft potential at the time. The temperatures, in eV, for $\frac{1}{2}v_L^2 = 20$ eV ion detector and $\frac{1}{2}v_L^2 = 15$ eV electron detector, when the spacecraft is uncharged, are given for corresponding values in the x axis. Different lines are for different bulk velocities v_0 normalized by $(2T_0)^{1/2}$, which are close to the Mach numbers.

that the effects of the spacecraft charge depend on the species and that the cutoff velocities are very large for electrons compared with ions because of their small mass.

The effects of a finite sampling energy range on the moment measurements depend on the temperature and velocity of the plasma being detected (Song et al., 1997). If the temperature is many times lower than the higher cutoff energy, the higher cutoff has only minor effects. The lower cutoff has more extended effects. In general, see Figure 1.2, the density and pressure are underestimated, and the velocity and temperature overestimated. In order to reduce the errors, some data analysts interpolate the points below the lower cutoff or fill the hole with best-fit to a convective Maxwellian distribution. The energy resolution becomes an important issue in order to accurately derive the moments of cold plasmas. Without a good

energy resolution within the bulk of the distribution, the density cannot be derived accurately.

In summary, it is very important for a data analyst to understand the scheme of the algorithm with which plasma moments are derived. Caution should be taken in particular for electrons, cold plasmas, plasmas with large anisotropy using 2-D measurements, and plasmas of multiple populations.

1.3.4. *Intercalibration*

For quantitative data analyses, the calibration of a measurement becomes essential. A calibration factor is often a function of time and plasma conditions. For example, degradation of an instrument with time requires the measurements be calibrated and recalibrated, and the calibration factor may change significantly across a shock as the plasma condition differs (Sckopke et al., 1990; Song et al., 1997).

When data analysis involves more than one instrument, intercalibration among these instruments becomes important. For example, the sonic Mach number is the ratio of two moments measured by the same detector. Even though the absolute value of each moment may not be accurate, one may suspect their ratio to be reasonably accurate. The Alfvén Mach number, on the other hand, involves three measured quantities from two different instruments. The chance for error is much greater. Another interesting example is that the nonlinear response of two instrument could lead to significant differences in the same physical quantity (Petrinec and Russell, 1993) and these differences may vary with the measured parameters. One way to intercalibrate the magnetic field and plasma moments measurements is to use the force balance requirement under some known conditions, such as near a stagnation region (Song et al., 1993), see Figure 1.3 and its caption.

Quantitative comparison among measurements from different satellites requires knowledge about the above issues for all satellites involved. Without careful intercalibration, one could draw a wrong conclusion. For example, the difference in calibration for two satellites could make normal fluctuations of two physical quantities into two clusters which lead to a linear relationship. This relationship could be mistakenly interpreted as a dependence between the two physical quantities, see Figure 1.4 for example.

Advances in technology and accumulation of experience have made multiple-instrument multiple-satellite studies possible. Plasma density can be intercalibrated by comparing particle measurements with plasma frequency measured by plasma wave experiments or wave propagation experiment (e.g. Harvey et al., 1978). Plasma velocity can be intercalibrated by comparing particle measurements with the field convection velocity, $\mathbf{E} \times \mathbf{B}$ (e.g., Mozer et al., 1983). However, the latter velocity has components only perpendicular to the magnetic field.

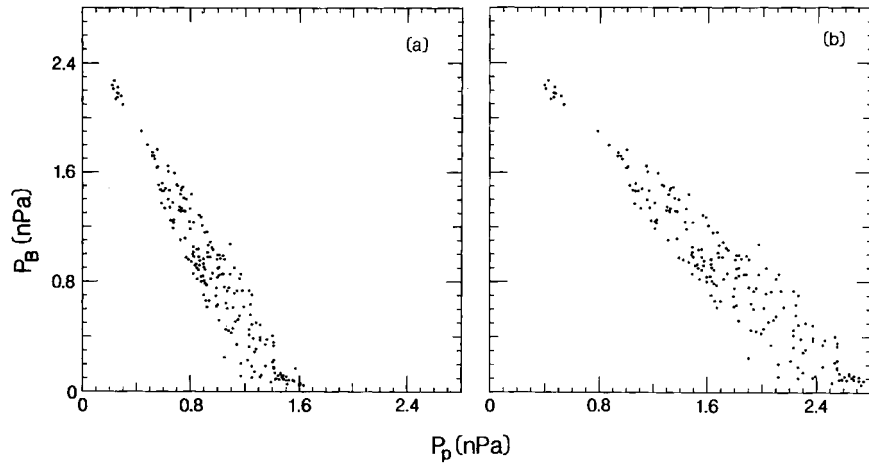


Figure 1.3. An example of intercalibration between the magnetic field and plasma measurements. At a stagnation region, the sum of the plasma pressure and magnetic pressure should be constant, or the variations in the two should be anticorrelated with a factor of -1 . The raw data panel (a) shows that the two pressures are anticorrelated but with a factor differing from -1 . A calibration factor is introduced to the plasma pressure to make the slope -1 . The intercalibrated plasma density is compared with the densities measured by other instruments to validate the method (Song et al., 1993).

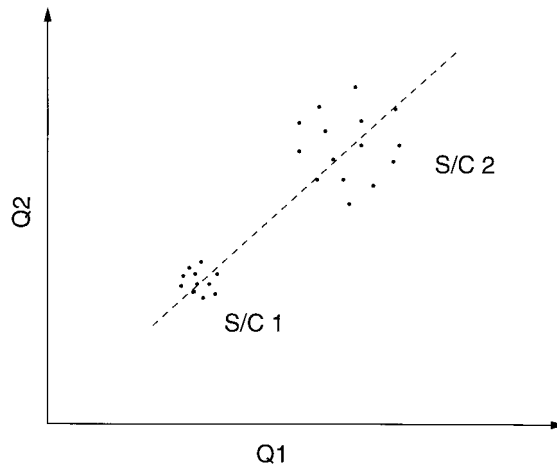


Figure 1.4. Random uncorrelated fluctuations in quantities Q_1 and Q_2 measured from spacecraft A and B which are not intercalibrated can sometimes be misinterpreted as a linear relation between Q_1 and Q_2 .

1.4. PRINCIPAL AXIS ANALYSIS

Principal Axis Analysis provides the mathematical basis for the Minimum Variance Analysis of discontinuity analyses in Section 2.2 and for the covariant matrix analyses of wave analyses in Section 3.1. For more introductory readings, one is referred to textbooks of multivariate analysis (e.g., Anderson, 1958). In space physics data analyses, the multivariates are often the three components of the magnetic field, $\mathbf{B}(t_i)$. We define a so-called covariance matrix

$$M_{\alpha\beta} = \overline{B_\alpha B_\beta} - \overline{B_\alpha} \overline{B_\beta}, \quad \alpha, \beta = 1, 2, 3, \quad (1.11)$$

where $\overline{B_\alpha B_\beta}$, $\overline{B_\alpha}$ and $\overline{B_\beta}$ are averages of $B_\alpha(t)B_\beta(t)$, $B_\alpha(t)$ and $B_\beta(t)$, respectively. Similarly, the covariance matrix can also be defined in the frequency domain, or $B_{\alpha,\beta}(t)$ are replaced by $B_{\alpha,\beta}(\omega)$. When $\alpha \neq \beta$, $M_{\alpha\beta}$ gives the cross-correlation between the two involved components of the field, and $M_{\alpha\alpha}$ is the auto-correlation. Principal Axis Analysis provides a tool for a coordinate transformation. In the new coordinate system, the cross-correlation, $M'_{\alpha\beta}$, between two components vanishes, or

$$\mathbf{M}' = \mathbf{T}^{-1}\mathbf{M}\mathbf{T}, \quad (1.12)$$

where \mathbf{T} and \mathbf{T}^{-1} are the transformation matrix and its inverse, and \mathbf{M}' is a diagonal matrix. The magnetic field in the new coordinate system is

$$\mathbf{B}' = \mathbf{T}\mathbf{B}. \quad (1.13)$$

Mathematically, to find such a transformation is to find the eigenvector ξ and eigenvalue λ of \mathbf{M} , or solve for

$$\mathbf{M}\xi = \lambda\xi. \quad (1.14)$$

Because \mathbf{M} is a 3×3 matrix, there are three solutions, ξ_1 , ξ_2 , and ξ_3 with $\lambda_1 \geq \lambda_2 \geq \lambda_3$. The three eigenvectors referred to as the principal axes (in rows) form the transformation matrix \mathbf{T} and the three eigenvalues referred to as the maximum, medium, and minimum eigenvalues, respectively, are the diagonal elements of \mathbf{M}' . As will be discussed below, Minimum Variance Analysis (Sections 2.2 and 3.1) assumes ξ_3 to be the normal direction of a discontinuity or the propagation direction of a wave and Maximum Variance Analysis (Section 2.5) assumes ξ_1 (for a different variable) to be the normal direction.

Loosely speaking, Principal Axis Analysis can be visualized as follows. The tip of the measured (magnetic field) vector draws points around the average field in three-dimensional space due to variations. A best-fit ellipsoidal surface centered at the tip of the average field that approximates these points is then obtained. The three axes of the ellipsoidal surface are the three principal axes. The lengths of the principal axes represent the standard deviation of the field fluctuations about the average field in the three directions, and their squares are the eigenvalues. The

above picture describes very well the perturbations associated with a wave. In the case of a discontinuity the field rotation across it is usually far less than 360° . For example, if the field rotates 180° , the field vector will vary on only one side of the maximum variation. In this case, the direction of the maximum eigenvector usually, depending on the distribution of the variations, remains parallel to the direction of the maximum variation, but the maximum eigenvalue will be different from the maximum variation. It is worth mentioning that for linearly polarized perturbations (in contrast to rotational perturbations) only one principal axis is determined and the other two have no definitive direction. This behavior causes uncertainty in data analyses and will be discussed in the corresponding subsections.

In general, in the frequency domain, the covariance matrix is complex. The Principal Axis Analysis is concerned only with the real part of the matrix. The meaning of the imaginary part of the matrix will be discussed in Section 3.1.

2. Discontinuity Analyses

There are many discontinuities in space. These can be classified by where they are and what function they play, e.g., the bow shock, the magnetopause, interplanetary shocks, solar wind discontinuities, the neutral sheet, and the heliospheric current sheet. They can also be classified by the physical nature of the boundary, fast shock, slow shock, rotational discontinuity or tangential discontinuity for example. The field and plasma properties usually change significantly across a discontinuity. In most theoretical studies a discontinuity is treated, for simplicity, as a one dimensional problem, namely, the physical quantities change only along the normal of the discontinuity. Observationally, while a spacecraft moves relative to a discontinuity, it measures the upstream and downstream conditions in time series. For a scalar quantity, the time series can be easily converted into a function of distance relative to the discontinuity if the motion of the discontinuity relative to the spacecraft is known assuming stationarity of the upstream conditions. For a vector quantity, to understand the physical behavior of a discontinuity and the processes near it, it will be convenient if the measurements are presented in a boundary normal coordinate system as has been discussed in Section 1.2.2. Such a system can often result in variations only in two dimensions and allow easier visualization and understanding of the behavior of the plasma. To find such a coordinate system, the normal direction of the discontinuity has to be determined. Different methods of discontinuity analysis have been developed to allow this determination. In Section 2.1, we briefly introduce the background and principles for discontinuity analysis. In Sections 2.2 to 2.4, we describe and discuss the three most useful methods in discontinuity analysis. In Section 2.5, we describe a method which has been proposed most recently.

2.1. BACKGROUND

2.1.1. Rankine–Hugoniot Relations

The plasma conditions on the two sides of a discontinuity are linked by the Magnetohydrodynamic (MHD) equations which describe the requirements for macroscopic continuity, pressure balance, and energy budget. If the discontinuity is planar and stationary, the MHD equations can be simplified to the form known as the Rankine–Hugoniot (R–H) relations. In isotropic plasmas, the R–H relations are

$$[\rho u_n] = 0, \quad (2.1)$$

$$[\mathbf{E}_T] = 0, \quad (2.2a)$$

$$[B_n] = 0, \quad (2.3)$$

$$[\rho u_n^2 + P + B_T^2/2\mu_0] = 0, \quad (2.4)$$

$$[\rho u_n \mathbf{u}_T - B_n \mathbf{B}_T/\mu_0] = 0, \quad (2.5)$$

$$\left[\left(\frac{\rho u^2}{2} + \frac{P}{\gamma - 1} + P \right) u_n + S_n \right] = 0, \quad (2.6a)$$

where ρ , \mathbf{u} , P , \mathbf{E} , γ , μ_0 , and $\mathbf{S} = \mathbf{E} \times \mathbf{B}/\mu_0$ are the density, velocity, pressure, electric field, ratio of specific heats, magnetic permeability in vacuum and Poynting vector, the square brackets denote the changes across the discontinuity, and the subscripts n and T denote the normal and tangential components to the discontinuity. For most of the problems in space physics, the frozen-in condition is applicable, or $\mathbf{E} = -\mathbf{u} \times \mathbf{B}$. Thus, Equations (2.2a) and (2.6a) can be written as

$$[(\mathbf{u} \times \mathbf{B})_T] = 0, \quad (2.2b)$$

$$\left[\left(\frac{\rho u^2}{2} + \frac{P}{\gamma - 1} + P \right) u_n + \mathbf{B}_T \cdot (\mathbf{B}_T u_n - B_n \mathbf{u}_T) \right] = 0, \quad (2.6b)$$

Equation (2.2b) holds upstream and downstream from a discontinuity even if the frozen-in condition is broken within the thin layer of the discontinuity. The R–H relations as written above hold in the shock frame rather than in the spacecraft frame, so that

$$\mathbf{u} = \mathbf{V} - \mathbf{V}_{\text{disc}}, \quad (2.7)$$

where \mathbf{V}_{disc} and \mathbf{V} are the velocity of the spacecraft relative to the discontinuity and the velocity of the flow measured in the spacecraft frame.

The R–H relations contain 8 equations and 19 parameters including the velocity of the discontinuity \mathbf{V}_{disc} . It is important to point out that the goal of data analysis is often not to simply apply the R–H relations but to verify them, or to

determine how well these relations hold in the situation being studied since several approximations have been made in applying the R–H relations. Ideally, one should substitute measured parameters in the left-hand side of Equations (2.1)–(2.6). The difference between the upstream value and downstream value of the quantity in each equation should be much smaller than either the upstream or the downstream value if the R–H relations are verified, or

$$\frac{[Q]}{|Q|} \sim 0, \quad (2.8)$$

where Q is the quantity in each of the R–H relations. The ratio on the left-hand side of Equation (2.8) gives the uncertainty of the analysis and Equation (2.8) is used as the basis of the discussions of the uncertainty of each method in the following subsections. If the R–H relations are not verified, one or more of the approximations made may not be valid. There may be temporal variations and/or curvature of the discontinuity, the anisotropy of the plasma, and/or significant presence of heat flow. Perhaps the identification of the nature of the discontinuity is incorrect. In these cases, conclusions should be drawn carefully from the analysis. However, at the present time, even in the best situation, with two spacecraft and full three dimensional measurements, observations provide only 18 parameters (16 plasma and field parameters, one timing difference and one distance measurement). Thus the R–H relations cannot be verified completely from observations and some additional assumptions must be made. Usually, one may assume a subset of the R–H relations as given, and then, use the remaining relations as confirmation. It is however not appropriate to assume all the R–H relations as given and then to determine remaining unmeasured parameters using, for example, optimization because different equations in the R–H relations have different uncertainties and because the results of such fittings are often found not to be a solution of the R–H relations (Chao, 1995). A cluster of four closely spaced satellites will enable us to verify the R–H relations independently. We will discuss this issue later in Section 4.2.

If some assumptions must be made, choosing the right subset of R–H relations can minimize the uncertainty of the results. Among the R–H relations, the continuity of the normal magnetic field, Equation (2.3), has the least uncertainty since it is not affected by time variations (see Equation (1.2)) and usually the magnetic field is the most accurately measured quantity with relatively high time resolution as discussed in Section 1.3. Almost all the present methods of discontinuity analysis are in fact based on this assumption. However, as will be discussed later in this section, there are ambiguities in some instances. An alternative is to use the continuity of the tangential electric field, Equation (1.3) or (2.2), if either the electric field or the plasma velocity can be measured accurately in three dimensions with a relatively high time resolution. We will discuss this method briefly in Section 2.5. Here we emphasize that comparing Equation (1.2) with (1.3), the required assumptions for the continuity of the tangential electric field are more than that for normal magnetic field conservation. Unless the velocity can be measured accurately (see discussion

in Section 1.3.3), Equation (2.2b) should not be used since it involves the cross product of two vectors and the uncertainties in the measurements will be amplified in the calculations. In this case, the conservation of mass, Equation (2.1), may provide a relatively smaller uncertainty than other relations except Equation (2.3). However, as has been discussed in Section 1.3.3, the calibration factors of plasma moments may change significantly across the bow shock or the magnetopause. One has to be extremely careful when using these moments. At the present time, it is suggested not to assume Equations (2.4)–(2.6) as given, rather to use them as confirmation, since they involve more complicated calculations and the effects of the temperature anisotropy and the intercalibration between the magnetic field and plasma measurements may become important.

Ideally, application of the Rankine–Hugoniot relations would involve simultaneous measurements both upstream and downstream of a discontinuity using two independent measuring platforms. In practice such application is usually performed using a single observatory moving across the discontinuity under the assumption that the external conditions do not change. Often the discontinuity is encountered because there has been a temporal change in these conditions, so caution must always be exercised and the time stationarity assumption verified when using the R–H relations.

2.1.2. *Types of Discontinuities*

There are several simplified types of discontinuities which have been commonly used to characterize and classify discontinuities (see the recent review by Lin and Lee, 1994). A discontinuity is called a tangential discontinuity (TD) if there is neither magnetic flux nor mass flux across it, or $u_n = B_n = 0$ in Equations (2.1) and (2.3). A TD is a current sheet separating two different plasmas.

A discontinuity is called a rotational discontinuity (RD) if there is magnetic flux across it but the density and the field strength (in an isotropic plasma) are same on the two sides of it, or $B_n \neq 0, u_{1n} = u_{2n} \neq 0, [B] = [\rho] = 0$, where subscripts 1 and 2 denote the values upstream and downstream of the discontinuity. The field strength and density may change within an RD. An RD is a propagating, usually non-linear, Alfvén wave front and satisfies Equation (2.2) in the form of

$$[\mathbf{u}_T] = \frac{u_n}{B_n} [\mathbf{B}_T] \quad (2.9)$$

and $u_n = B_n / \sqrt{\rho \mu_0}$. Equation (2.9) is the so-called Walén relation for isotropic plasmas and will be further discussed in Section 2.7. As B_n, u_{1n} and u_{2n} go to zero, an RD may degenerate into a TD. However, this TD is different from a general TD because its velocity change has to be parallel to its field change but a general TD has not.

A discontinuity is called a shock if there are both magnetic flux and mass flux across it and if there is a change in the density, or $B_n \neq 0, u_{1n} \neq u_{2n} \neq 0$, and $\rho_1 < \rho_2$. A shock is associated with a dissipation process and usually with heating

of the plasma. Across a shock, the flow velocity decreases from above to below a characteristic speed, such as the fast mode speed, intermediate mode speed or slow mode speed (for more discussion on the modes, see Section 3.3.1), in the frame at rest to the discontinuity. Similar to a shock but with $\rho_1 > \rho_2$, a discontinuity is called a rarefaction wave. A rarefaction wave may occur in an expansion fan such as formed when flow moves across a ledge and expands into a vacuum. In theory, an expansion fan cannot steepen but in reality because of the rapid motion between an expansion fan and the spacecraft, it can appear sharp in the time series data. An observed discontinuity may be a superposition of these elementary discontinuities and also may not be in steady state, but in the regions far from where a discontinuity is generated, these elementary discontinuities are expected to separate because of their difference in speed.

2.2. THE MINIMUM VARIANCE ANALYSIS

This method is based on the Principal Axis Analysis (Section 1.4) and the fact that the magnetic field is divergence-free, $\nabla \cdot \mathbf{B} = 0$, the derivative form of Equation (2.3) (see also the discussion of Equation (1.2)). For an infinitesimally thin discontinuity, Equation (2.3) should hold across it if it is planar. If a structure consists of many such discontinuities and they are parallel to each other, the sum of the fluctuations normal to the discontinuity should be zero. In reality, the fluctuations within the structure are equivalent to distortions of these thin discontinuities. Thus locally, the normal direction of a discontinuity may not be the same as of the overall structure, and thus there can be magnetic fluctuations along the direction of the average normal. The minimum variance method assumes that the distortions of these thin discontinuities from the overall structure are small compared to the changes in the magnetic field in the plane of the boundary. Thus the field fluctuations are smallest in the direction normal to the overall structure (Sonnerup and Cahill, 1967). Therefore, to determine the normal direction of a discontinuity with internal structure is equivalent to finding the minimum eigenvector direction of the principal axis analysis (see Section 1.4).

Note that the minimum variance analysis assumes only that the variations are smallest along the normal but the normal component of the steady field is not necessarily smallest. As will be discussed next and in Sections 2.3 and 2.4, in theory, the minimum variance method has a large uncertainty for tangential discontinuities and shocks (see Figure 2.1 (Lepping and Behanon, 1980)) since both theoretically consist of linearly polarized variations of the field and the normal may lie anywhere perpendicular to this direction of the maximum change (see further discussion in Section 2.7). In one situation minimum variance will give an accurate shock normal, when there is a standing whistler mode precursor propagating upstream along the shock normal. This is usually seen for subcritical shocks (Mellott and Greenstadt, 1984) and it assumes the upstream whistler wave propagates in the

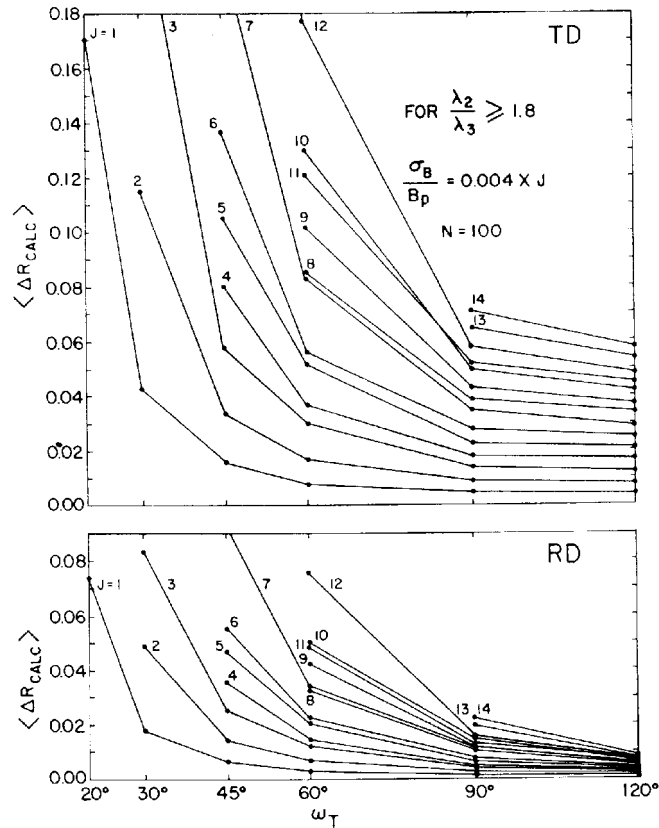


Figure 2.1. The errors of the minimum variance analysis from a numerical experiment (Lepping and Behannon, 1980). The upper (lower) panel shows the errors for TDs (RDs). ω_T is the shear angle of the field across a discontinuity. The errors for TDs are much greater than for RDs.

same direction as the shock wave does. The minimum variance method is most useful for rotational discontinuities and other more complicated situations.

In principle, one may also apply the minimum variance method to the mass flux, $\rho \mathbf{V}$, to determine the normal direction (Sonnerup et al., 1987), since in steady state we have $\nabla \cdot \rho \mathbf{u} = 0$. However, in practice, due to temporal variations, the most important cause of which comes from the relative motion between the discontinuity and the spacecraft (see Equation (1.1)), and relatively large uncertainties in the plasma measurements, for example, due to changes in composition and/or calibration factor (see Section 1.3.3) across the discontinuity, this method has a much larger uncertainty than the magnetic field minimum variance in addition to the uncertainties discussed below.

What limits the accuracy of the minimum variance method? Following the steps of the description discussed above, we know that the minimum variance method

can be limited by the data resolution and wave activity within the discontinuity. As the determination of the principal axes is equivalent to a three-free-parameter fit, a small number of data points will lead to a large uncertainty in the fit. In an extreme case, if there is no measurement within the discontinuity, this method should be used with caution since the minimum variance direction would then be determined by the wave activity upstream and downstream. Therefore, high resolution measurements and slow motion of the discontinuity relative to the spacecraft will minimize the uncertainty. On the other hand, as the resolution increases, one may be able to resolve the wave activity within the discontinuity. These waves may cause uncertainty in the determination of the normal direction as well. As discussed earlier, the assumption made in this method is that the infinitesimally thin surfaces within the overall structure have only small perturbations. Waves and small structures within the discontinuity may destroy the validity of this assumption. For example, if the magnetic perturbations within the structure due to structure and waves are mainly along the normal, the minimum variance direction will not be the normal direction for a discontinuity with a small field change across it. Filtering the data to pass only frequencies consistent with the thickness of the structure will help in reducing the uncertainty of the normal determination. The uncertainty of the minimum variance analysis was first discussed quantitatively by Sonnerup (1971) (in using Equation (13) of Sonnerup (1971) note that there is a typographical error that $\langle B^2 \rangle$ should be $\langle B \rangle^2$) and then investigated comprehensively and numerically by Lepping and Behanon (1980). Recently, Kawano and Higuchi (1995) used the bootstrap method to estimate the errors in the minimum variance analysis.

In principle, one may select many different time intervals for the minimum variance analysis and each of them provides a different normal. How to evaluate a result of the minimum variance analysis? Here are the several key issues to check.

(1) Check the normal component of the field. After rotating the field into the minimum variance coordinates, the average fields in the minimum variance direction on the two sides of the discontinuity should be the same at least in the regions close to the discontinuity. Often a visual inspection can quickly determine whether the rotation is good. Since the minimum variance analysis provides the direction of smallest field fluctuations only within the selected time interval, if the interval does not include all the major field changes, one may find that the average fields in the minimum variance direction are different on the two sides of the discontinuity. In the case of the magnetopause, the field in the minimum variance direction may increase or decrease continuously on the magnetospheric side due to the curvature of the magnetospheric field.

(2) Check the ratios of the eigenvalues. The square root of an eigenvalue is the standard deviation of the field along that direction. Ideally the minimum eigenvalue should be zero. In practice, if the minimum eigenvalue is much smaller than the other two eigenvalues, the minimum variance direction is well determined. Usually, a normal direction is considered as to be well determined if the minimum eigenvalue is one order smaller than the intermediate eigenvalue, or the amplitudes

of the perturbations along the normal are less than one third of the smaller one of the two components perpendicular to the normal.

(3) Check the minimum eigenvalue. Select a different time interval across a discontinuity to provide a set of normal directions. Since a smaller minimum eigenvalue indicates a better determination of the normal direction, one may choose the normal direction with a smaller minimum eigenvalue but also with a smaller ratio of the minimum and intermediate eigenvalues. However, usually, a shorter interval, or a smaller number of data points provides a smaller minimum eigenvalue. In the extreme case, if only three data points are selected, the minimum eigenvalue may go to zero since the ellipsoidal surface degenerates into a plane. In this case, the smaller minimum eigenvalue is obtained with some sacrifice in statistics. In practice one should include in the analysis only the variations associated with the discontinuity being analyzed. In principle, one should choose the normal direction with a smaller minimum eigenvalue and a longer time interval. The ratio of the minimum variation of the field and the strength of the average field should be very small, less than few percent, or $\sqrt{\lambda_3}/|B| \sim 0$, where λ_3 is the minimum eigenvalue.

(4) Check the ratio of the minimum variation of the field to the average field in the minimum variance direction, or $\sqrt{\lambda_3}/(j-2)/B_{\min}$ where B_{\min} is the average field in the minimum variance direction and j is the number of data points. A large value of this ratio indicates a large uncertainty in the analysis as will be discussed in the section of tangential discontinuity analysis. A large number of data points will reduce the uncertainty. When B_{\min} is extremely small, the discontinuity could be a tangential discontinuity which needs to take additional caution when using the minimum variance analysis.

The suggested procedures are as follows:

(1) In the time interval selection, try to minimize the number of the data points on the two sides of the discontinuity but try to maximize the number of the data points within the discontinuity. Too many data points on the two sides of the discontinuity will put too much weight on the fields on the two sides. More data points within the discontinuity in general will increase the statistical significance of the determination.

(2) Examine the fluctuations in the field during the crossing. If there are strong waves present that are not part of the discontinuity structure being analyzed, low-pass filter the data before a minimum variance analysis.

(3) Perform the minimum variance analysis for several different selected time intervals and compare the results according to the discussion above. Experience indicates that if the results are essentially the same for several neighboring 'nested' data segments, they are perhaps believable (Sonnerup, private communication, 1992).

(4) Compare with the normal directions predicted by geometric models if there are any. Occasionally, one may find the normal direction determined by the minimum variance is orthogonal to the model prediction. Most likely, this is caused by the 90° ambiguity to be discussed in Section 2.7.

(5) Since the difference among the normals determined from different time intervals provides a measure of the uncertainty of the analysis, never draw qualitative conclusions which may not be true given the uncertainty.

2.3. TANGENTIAL DISCONTINUITY ANALYSIS

In theory tangential discontinuities are those with B_n and u_n zero. There is no magnetic field or mass flux across a tangential discontinuity. As discussed in Section 2.1, Equation (2.8), a good determination of the normal is indicated by a small ratio between the difference and the average of the quantity in the R–H relation across a discontinuity. Noting that the ratio of the standard deviation and the probable error of the mean is $j^{-1/2}$, the ratio $j^{-1/2} \Delta B_n / \langle B_n \rangle$, where ΔB_n and $\langle B_n \rangle$ are the minimum variation and the average of the field in the minimum variation direction, provides a measure of the uncertainty of the minimum variance method in verifying Equation (2.3). For a TD, since $\langle B_n \rangle$ is small, the ratio becomes very large and hence the minimum variance has a large uncertainty. The effect of such a large uncertainty can be seen when one selects different intervals and finds different normal directions but $\langle B_n \rangle$ remains similar.

Since the magnetic field is tangential to a TD surface, the normal direction of the discontinuity is perpendicular to the fields upstream and downstream, or

$$\mathbf{n} \parallel \mathbf{B}_1 \times \mathbf{B}_2, \quad (2.10)$$

where \mathbf{B}_1 and \mathbf{B}_2 are determined by selecting a relatively stable interval on each side of the discontinuity. Ideally, this is done using simultaneous data from two spacecraft. When one spacecraft is used care must be exercised to ensure that the changes observed are solely due to the spatial gradients across the discontinuity.

The uncertainty in the determination of the normal for a tangential discontinuity arises from the uncertainties in measurements of the two fields due to the fluctuations near the discontinuity. The uncertainties for the measurements of the two fields are the probable errors of mean, not the standard deviation. A large number of data points in measuring each of the fields may reduce the uncertainty. In reality, however, there may be temporal changes in the magnetic field near the discontinuity of interest. Some are oscillations and others may be either gradual or sudden changes. The effects of oscillations can be removed by averaging over many wave cycles. Again if there are temporal changes in conditions as the spatial discontinuity is crossed, the calculated normal will be affected. Finally, the uncertainty of the TD method becomes large when the fields on the two sides are nearly parallel to each other.

In summary, the tangential discontinuity analysis has a relatively small uncertainty in determining the normal of a tangential discontinuity if the fields on the two sides of the discontinuity are not parallel to each other (with a change only in magnitude). However, one has to verify carefully that a discontinuity is a tangential discontinuity before using the method. To minimize the uncertainty in the normal

direction of the discontinuity, one should try to select time intervals as long as possible on the two sides of the discontinuity to minimize the effect of the wave but without major changes in the field on the two sides to minimize the effects of changing external conditions.

2.4. COPLANARITY ANALYSIS

A discontinuity is called a shock if there are magnetic flux and mass flux through the discontinuity and the velocity changes from supersonic relative to the discontinuity to subsonic. This velocity change causes a change in the density across the discontinuity. A shock is called a fast (slow) shock if the density changes in (out of) phase with the magnetic field strength across the shock. Here we have ignored the intermediate shocks in which the density and the field strength may vary either in phase or out of phase but the rotation in the field tangential to the discontinuity must be exactly 180° . From the R–H relations, one can show that the magnetic fields on the two sides of the shock and the normal of the shock are coplanar, and that the normal is also perpendicular to the vector of $(\mathbf{B}_1 - \mathbf{B}_2)$ (Colburn and Sonett, 1966). The normal direction, thus, is

$$\mathbf{n} \parallel (\mathbf{B}_1 - \mathbf{B}_2) \times (\mathbf{B}_1 \times \mathbf{B}_2). \quad (2.11)$$

The fields are coplanar only in the region in which there is no electric field along the normal. In boundary normal coordinates, the non-coplanar component is almost zero on the two sides of the shock. The normal component remains constant but with a finite value through the shock. Consistent with this theorem, non-coplanar magnetic fields are frequently observed within the quasi-perpendicular subcritical shock associated with the dissipation, see Figure 2.2 for example. Under this circumstance the minimum variance analysis should be applicable as it is when there is a standing wave upstream of the shock along the normal. When there is a sizable non-coplanar component its magnitude can be used to derive the shock velocity from a single spacecraft (Newbury et al., 1997). Since under typical conditions the variations in both the normal and noncoplanar components are small, the minimum variance analysis generally has a large uncertainty at the shock.

Similar to the tangential discontinuity analysis, ideally the calculation is made with simultaneous measurements on two sides of the discontinuity. If as usual the calculation is made from the measurements on a single spacecraft, uncertainty in the coplanar analysis is mainly caused by temporal variations. Since usually there are strong fluctuations near a shock, especially for a high Mach number shock, the two fields should be measured in the regions which are relatively quiet. The trailing wavetrains downstream from shocks are usually not coplanar with the normal. One should avoid selecting these regions as the downstream condition. For a weak shock, the uncertainty for this method may become large since the fields on two sides of the shock may be similar and the two vectors in the brackets in Equation (2.11) are both close to zero (Russell et al., 1983).

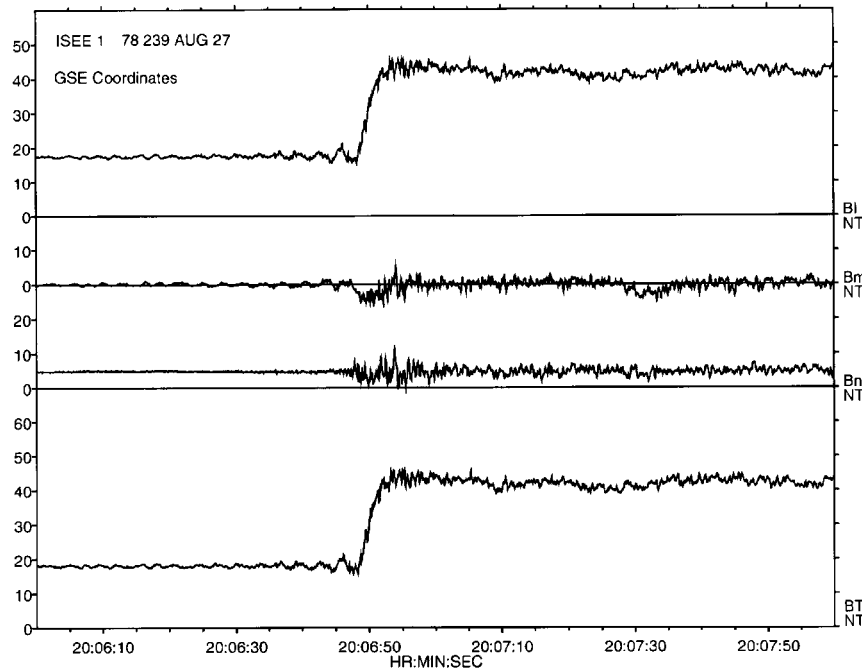


Figure 2.2. An example of bow shock crossings. The normal direction is determined using the coplanar analysis. Note that B_m , the non-coplanar component, is near zero on both sides of the shock. If one used the tangential discontinuity analysis, the normal direction would be the m direction. In this case, because of a significant non-coplanar field within the shock, the minimum variance analysis will provide correct normal direction. Without such a noncoplanar component, the minimum variance analysis will have a very large uncertainty.

For a quasi-parallel shock, the normal of which is nearly parallel to the upstream magnetic field, the average field change across the shock is small. The uncertainty in the normal determination is expected to be large. Furthermore and more importantly, there are usually large amplitude fluctuations present near such a shock, and the shock front is often not clearly defined. This makes the shock normal analysis extremely difficult. Computer simulations have shown that the shock front undergoes a continuous reformation process (Krauss-Varban and Omidi, 1993). Therefore the stationarity approximation based on which the R–H relations are derived may not be valid. How to analyze quasi-parallel shocks has not been systematically investigated.

2.5. THE MAXIMUM VARIANCE ANALYSIS

The determination of the normal of a discontinuity from the minimum variance analysis of the magnetic field has a very large uncertainty when the intermediate and minimum eigenvalues are close to each other. This is the usual situation when the field shear across the discontinuity is small (e.g., when the major field

change is in its strength) or when the observations cannot resolve the interior of the discontinuity either due to low time resolution of the measurements or fast motion of the discontinuity relative to the spacecraft. In these circumstances, the maximum variance analysis of the electric field may offer a better determination of the normal. The maximum variance analysis of the electric field is based on the fact that the electric field is curl free in steady state (see also discussions on Equation (1.3)) or

$$\nabla \times \mathbf{E} = -\frac{\partial \mathbf{B}}{\partial t} = 0. \quad (2.12a)$$

Thus,

$$\mathbf{n} \times \Delta \mathbf{E} = 0, \quad (2.12b)$$

namely, the normal is along the electric field change. Analogous to the minimum variance analysis of the magnetic field, the normal direction of the discontinuity is along the maximum variance direction of the electric field.

The electric field data for the analysis can be from either direct measurements of the electric field or the convective electric field derived from the magnetic field and plasma velocity measurements according to the frozen-in condition, or $\mathbf{E} = -\mathbf{v} \times \mathbf{B}$. If the electric field measurements are two dimensional, the third component of the electric field can be obtained from the frozen-in condition, or $\mathbf{E} \cdot \mathbf{B} = 0$.

The major uncertainty in this method comes from the temporal variation terms in Equation (1.3). Both unsteadiness of the discontinuity itself and changes of its motion relative to the observer will affect the results. In particular in the case of the magnetopause, the boundary usually oscillates instead of being in constant motion. Another important uncertainty comes from the relative motion between the spacecraft and the boundary, even if the motion is steady. From Equation (2.2), one obtains a tangential electric field change in the spacecraft frame, $\Delta \mathbf{E}_T = \mathbf{V}_{\text{disc}} \times (\mathbf{B}_2 - \mathbf{B}_1)$. To remove this effect, one has to transfer the electric field into a frame which is at rest in the discontinuity. However, since the motion of the discontinuity is in general unknown before the normal direction of the discontinuity is determined, it is difficult to completely remove this effect. Sonnerup et al. (1987) developed a method and gave a comprehensive discussions on how to minimize this effect. One way to reduce this effect is as follows.

- (1) Find the maximum variance direction of the electric field, \mathbf{n} .
- (2) Measure the average velocity and magnetic field along \mathbf{n} , v_n and B_n , within the discontinuity.
- (3) If the discontinuity is not a shock, calculate

$$u_n = \pm B_n / \sqrt{\mu_0 \rho}, \quad (2.13)$$

where u_n is similar to the flow velocity across the discontinuity, and we have assumed that the discontinuity is a rotational discontinuity.

- (4) The relative velocity of the discontinuity to the spacecraft is approximately

$$V_{\text{disc}} = v_n - u_n . \quad (2.14)$$

(5) Subtract the electric field due to the relative motion

$$\mathbf{E}' = \mathbf{E} - \mathbf{V}_{\text{disc}} \times \mathbf{B} . \quad (2.15)$$

(6) Using E' as corrected electric field, repeat the procedures above, until \mathbf{n} does not change.

Another source of uncertainty in the maximum variance method is due to the assumption of the frozen-in condition to calculate the electric field if it is not measured directly in three dimensions. Under the frozen-in approximation the effects due to the Hall term, the resistivity term, electron pressure gradient term and electron inertial term in Ohm's law have been ignored. These effects may be important in sharp changes within a discontinuity.

The principles to evaluate a result of the maximum variance analysis is similar to some of that for the minimum variance analysis discussed in Section 2.2. The success of the method requires a much larger maximum eigenvalue than the other two eigenvalues. The continuity of the normal component of the magnetic field and tangential electric field can be used as a check on the results. The maximum variance analysis does not require many data points within a discontinuity, a significant advantage over the minimum variance analysis. However, since the plasma velocity measurements have usually a lower time resolution than the magnetic field measurements, fewer data points are obtained within and near a discontinuity. If the fluctuations near the discontinuity are not small, the result may be very sensitive to the number of the data points used in the analysis.

In summary, the maximum variance analysis of the electric field can be used as an alternative in cases when the minimum variance analysis has a large uncertainty. It should be used with caution.

2.6. DEHOFFMANN–TELLER FRAME AND WALEN RELATION TEST

The deHoffmann–Teller (HT) frame is one of the shock frames, i.e., a frame at rest in the discontinuity. Here we recall that frames at rest in the discontinuity can have different tangential velocities. The HT frame moves along the shock front with a velocity such that the magnetic field and velocity are parallel, and hence the tangential electric field vanishes (see a brief review by Sonnerup et al., 1995). The HT frame moves at a speed

$$\mathbf{V}_{\text{HT}} = \mathbf{V}_i \pm \mathbf{B}_i \frac{u_{\text{in}}}{B_n} \quad (i = 1, 2) , \quad (2.16)$$

relative to the observer. The plus and minus signs correspond to the normal component of the velocity and magnetic field to be antiparallel and parallel, respectively. In the normal incidence case, because $\mathbf{V}_{1T} = 0$, the downstream tangential velocity $\mathbf{V}_{2T} = (\mathbf{B}_{2T}u_2 - \mathbf{B}_{1T}u_1)/B_n$ is in general nonzero. The flow is accelerated

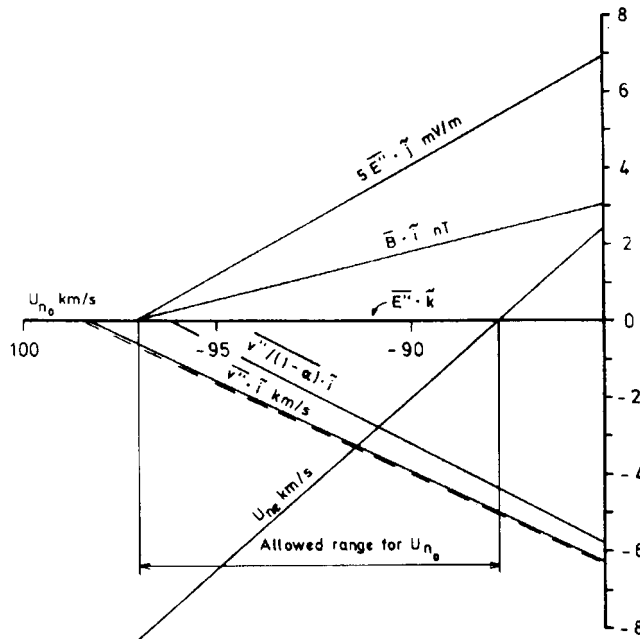


Figure 2.3. In the most advanced development (Sonnerup et al., 1987), the acceleration of the HT frame can be introduced to improve the fit and hence the temporal variations of the HT frame can be partially resolved.

tangentially in crossing the shock due to the kink force of the field. Since the electric field in the HT frame is zero, the electric field in the spacecraft frame is

$$\mathbf{E}_i = -\mathbf{V}_{HT} \times \mathbf{B}_i. \quad (2.17)$$

The proportionality between the electric and the magnetic field variations can be used to determine the velocity of the HT frame. Practice indicates that a well-determined HT frame can often be found in the magnetopause (Sonnerup et al., 1990; Walthour et al., 1993). Since the HT frame is a shock frame ($V_{HTn} = V_{disc}$), it can be used to solve the difficulty in determination of a shock frame discussed in Section 2.5. However, in general, since the normal direction is unknown, it needs iteration before a satisfactory result is reached. In its most developed form, using the information from the maximum variance analysis of the magnetic and electric fields and the minimum variance analysis of \mathbf{V}_{HT} , through iteration, one can derive not only the normal direction, but also the normal velocity and its acceleration, see Figure 2.3, for example. The effect of the acceleration is important as seen in the last term of Equation (1.3).

Equation (2.17) can also be written in the form of finite field perturbations. If the magnetic field change is in the \hat{L} direction, and when the discontinuity has no motion in the normal direction, the electric field change is in the \hat{N} direction, the

HT frame moves mainly along the \hat{M} direction. To derive \mathbf{V}_{HT} in Equation (2.17) is equivalent to a three-parameter fit or a minimum variance problem. Sonnerup et al. (1987) provided the expression for the covariance matrix. Here we should note that unless \mathbf{B}_i and \mathbf{V}_i outside the shock ramp are not coplanar, \mathbf{V}_{HT} and the magnetic fields are all in a plane orthogonal to the shock surface. The derived electric field change is along the shock surface and not along the normal direction.

For RDs, combination of Equations (2.5), (2.9), and (2.16) yields the Walen relation in the spacecraft frame,

$$\mathbf{V} = \mathbf{V}_{\text{HT}} \pm \mathbf{C}_A, \quad (2.18)$$

where $\mathbf{C}_A = \mathbf{B}/\sqrt{\mu_0\rho}$ for an isotropic plasma and $\mathbf{C}_A = \sqrt{\xi_0}\mathbf{B}/\sqrt{\mu_0\rho}$ for an anisotropic plasma, $\xi_0 = \mu_0(P_\perp - P_\parallel)/B^2$ is the anisotropy factor (Chao, 1970), and P_\perp and P_\parallel are the pressures perpendicular and parallel to the magnetic field, respectively. The plus and minus signs denote the RDs propagate antiparallel and parallel to the magnetic field, respectively. The subscript i has been neglected assuming that the relationship applies to every measurement.

The Walen relation is a more restrictive test for a discontinuity. It requires not only a well determined HT frame but also that the discontinuity propagates with the Alfvén speed relative to the flow. A positive result of the test verifies the discontinuity to be an RD. Here we should emphasize that a linear relationship between the plasma velocity and magnetic field (or the Alfvén velocity) variations does not necessarily imply the discontinuity to be an RD unless the offset between the two, \mathbf{V}_{HT} (see Equation (2.16)) equals the proportional factor between the electric and magnetic field variations (see Equation (2.17)).

2.7. SUGGESTED PROCEDURES FOR DISCONTINUITY ANALYSIS

As discussed in Sections 2.2 to 2.4, the three methods commonly used for discontinuity analyses are for different purposes. None is intrinsically better than others. However, different methods may provide different normal directions. For example, the noncoplanar direction in the coplanarity analysis is very close to the normal direction in the tangential discontinuity analysis for the same discontinuity (see Figure 2.2 for an example). The minimum variance direction could be the noncoplanar direction of a shock. Thus, the question is how to analyze a discontinuity without a presumption about its type. The following is a suggested approach.

(1) Collect as much information as possible for the interesting discontinuity in addition to the magnetic field, such as plasma measurements and the measurements from other close spacecraft if there are any. These measurements may help to constrain the results.

(2) Make an overview of the discontinuity to decide where are the upstream and downstream regions and which major change is most interesting. There may be more than one choice. Keep in mind that a single discontinuity, when its internal structure can be resolved, may appear to consist of more than one sharp

changes, that a discontinuity can oscillate back and forth, and that two distinct discontinuities could be very close to each other in time series records.

(3) Begin the analysis by using the minimum variance analysis with caution as discussed in Section 2.2. For low time resolution measurements or when the discontinuity being studied moves fast relative to the spacecraft, there may be no measurement within the discontinuity. In this case, the minimum variance method can provide only the direction of maximum variance and the normal of the discontinuity cannot be determined with only magnetic field measurements.

This step only provides hints to the nature of the discontinuity. One may only guess the nature of the discontinuity from the results of the minimum variance analysis. For example, if the field in the minimum variance direction is close to zero, the discontinuity may be a tangential discontinuity. If the field in the minimum variance direction is not small and the field strength is similar on the two sides of the discontinuity, the discontinuity may be a rotational discontinuity. If the major field change is in only one component and the field strength changes, the discontinuity may be a shock.

Further determinations of the properties of the discontinuity need plasma measurements or assumptions.

(4) Make assumptions of the nature of the discontinuity if the processes associated with the discontinuity are known. In many of discontinuity studies, the nature of the discontinuity has been carefully studied previously with particle measurements and hence known, but the plasma moments are not available. In these cases, assumptions of the nature of a discontinuity can be made, but keep in mind that the results are conditional depending on the accuracy of the assumptions. For most of bow shock studies, coplanarity is a good assumption. In fact, most of these studies skip step 3 and use the coplanarity analysis directly. In most circumstances, the magnetopause can be considered as either a tangential discontinuity or a rotational discontinuity. Thus one may try the tangential discontinuity analysis if the minimum variance component is close to zero. The neutral sheet can be considered as a rotational discontinuity. However, to analyze the neutral sheet using the minimum variance method is difficult because the variance in one of the tangential components (\hat{y}) can be very small. The field aligned current sheets in the low beta magnetosphere can often be treated as tangential discontinuities. For most interplanetary discontinuities, since their natures are unknown without plasma measurements, an analysis using only magnetic field measurements cannot be conclusive except if the field strength remains nearly constant across the discontinuity. In this latter case, the discontinuity can be either a tangential discontinuity or a rotational discontinuity and can be treated similarly to the magnetopause situation just discussed above. However, because often an interplanetary discontinuity moves with a speed similar to that of the solar wind, the flow velocity relative to the discontinuity is not easy to resolve. Caution should be taken when interpreting it as either a TD or an RD. The main difficulty in analyzing an interplanetary

discontinuity is the lower time resolution of the data caused by the fast passage of the discontinuity.

With an assumption of the type of a discontinuity, one can determine the normal direction of the discontinuity using the methods discussed in previous subsections.

(5) Use plasma measurements. With plasma measurements, if the normal direction of the discontinuity has been determined in the last step, the normal velocity of the discontinuity can be determined, according to Equations (2.1) and (2.7),

$$V_{\text{disc}} = \frac{\rho_1 V_{1n} - \rho_2 V_{2n}}{\rho_1 - \rho_2}. \quad (2.19)$$

If the plasma velocity measurements are not three dimensional, an additional assumption has to be made, hence more uncertainties are introduced, in the calculations. Although this relationship is widely used, we would like to point out that when V_{disc} is much less than both V_{1n} and V_{2n} , it could be in the range of the uncertainty of the velocity measurements. In this case, Equation (2.19) is not useful. The uncertainty in this calculation comes from these major sources. The first is in the direction of the normal. This error is discussed above. The second is the fluctuations upstream and downstream. In some cases, this source may not be important. The third is due to the change in the calibration factor of the instrument associated with the change of plasma state as discussed in Section 1.3.3. While velocity measurements, at least of the solar wind ions are usually quite accurate, the plasma density upon which Equation (2.19) depends are not so accurate. If the temperature and speed of the plasma remain similar across the discontinuity, this second source may not be important. Unfortunately under these circumstances, the density change is small, leading to a very large uncertainty in the denominator of Equation (2.19). At shocks, both the temperature and speed vary across the discontinuity in such a way that the density calibration in some instruments can be drastically affected (Petrinec and Russell, 1993). An example of where this calibration change may have seriously affected measurements across the bow shock can be found in Lepidi et al. (1996).

For interplanetary discontinuities, plasma measurements can help to determine the type of a discontinuity. As discussed in step 4, its type can be determined when there is no change in the field strength across the discontinuity. If there is a change in the field strength, in the simplest case, the discontinuity is either a tangential discontinuity or a shock. (If the plasma is anisotropic, even an RD can be accompanied by a change in field strength.) As discussed before, the minimum variance method has a large uncertainty in these two situations, and the tangential discontinuity analysis and the coplanarity analysis provide two orthogonal normal directions. Which of these two normals is correct? For a tangential discontinuity, since $u_{1n} = u_{2n} = 0$, or $V_{1n} = V_{2n} = V_{\text{disc}}$, the velocity difference $\mathbf{V}_1 - \mathbf{V}_2$ is tangential to the discontinuity. Thus, comparing the direction of the velocity difference with the two normals may eliminate one of the two. There is a possibility that the three vectors are orthogonal to each other. Another way to distinguish a tangential

discontinuity from a shock is to check the total pressure, the sum of the thermal pressure and the magnetic pressure $B^2/2\mu_0$. For a tangential discontinuity, since $u_n = B_n = 0$, from Equation (2.4), the total pressure remains constant across the discontinuity. However, this method relies on a good intercalibration between the magnetic field measurements and the plasma measurements. For slow shocks, since the change in the total pressure is expected to be small, it is difficult to eliminate the possibility of slow shocks solely based on the total pressure balance.

(6) Use the timing difference from two spacecraft. Measurements from two spatially separated spacecraft may be able to determine the velocity of a discontinuity. If the separation of the two spacecraft along the discontinuity is smaller than the curvature of the discontinuity and the time scale of the variations of the discontinuity is much larger than the time scale of the time delay between the encounters of the discontinuity by the two spacecraft (the determination of the time delay will be discussed in Section 4.1), the discontinuity can be considered as a stationary planar wave front. The curvature of the discontinuity can be determined by comparing the two normals from the two spacecraft if only a single crossing is observed by each of them. The velocity of the discontinuity in the plasma frame can be obtained by

$$V_{\text{disc}} = \frac{\mathbf{L} \cdot \mathbf{n}}{\Delta t} - V_{(1,2)} \cos \eta_{(1,2)}, \quad (2.20)$$

where \mathbf{L} , Δt , and η are the separation vector between the two spacecraft, the time delay between the observations of the discontinuity from the two spacecraft and the angle between the flow velocity, either upstream or downstream, and the normal of the discontinuity. Note that Equation (2.20) requires both 3-D plasma measurements and the normal direction. Therefore, this step is not independent of the last two steps. It provides additional information and can be used in combination with the last two steps or as a verification of the results from the last two steps.

2.8. R–H RELATIONS TEST

In the previous steps, we have used two of the R–H relations, the mass conservation and the continuity of the normal field. These two relations have the least uncertainty in practice. However, in steps 4 and 5, to determine the normal of a discontinuity, we have approximated it by a simplified discontinuity. In fact, a discontinuity may not be any of these simple discontinuities assumed rather a combination of them. Also, an analysis may provide many different parameter sets as the result of multiple interval selections. One should use these parameter sets to test other R–H relations. With this step, one may find the best parameter set for the study and estimate the uncertainty of the analysis. The uncertainty is the ratio of the difference to the average of the quantity, Equation (2.8), in each of the R–H relations. The conclusions of any analysis should be drawn consistent with this uncertainty.

In the procedure described above, the determination of the shock speed carries enormous weight in determining the nature of a discontinuity. As discussed in Section 1.3.3, the calibration factor of plasma measurements may change across a shock. To accurately determine the shock velocity is extremely difficult. To solve this problem, Chao et al. (1995) recently developed a method to derive the properties of a shock without predetermining the shock velocity and then derive the shock velocity afterward. In their method, the R–H relations are combined and written in normalized parameters, such as ratios of upstream and downstream values, angles, plasma betas, and Mach numbers. It is realized that to determine an isotropic shock requires only any three independent such parameters. Therefore, one can choose three observationally best determined parameters to substitute into the normalized R–H relations and then derive those remaining. We suggest using the ratio of the ratio of the magnetic field strengths, B_2/B_1 , the upstream shock angle and downstream plasma beta. All these three parameters are independent of frame reference. The ratio of the magnetic field strengths has little uncertainty. The shock angle depends only on the accuracy of the shock normal determination. As shown in Figure 1.2, among three plasma moments, the uncertainty in the pressure measurements is least and is even smaller for higher temperature plasmas. Thus the downstream plasma beta is the most accurately measured quantity although its absolute calibration needs to be verified.

An interesting finding by Chao et al. (1995) from a parametrical study is that the ratio of the upstream and downstream plasma densities is less sensitive than the ratio of plasma velocities. In other words, a small difference in the density ratio corresponds to a large range of shock parameters. A small uncertainty in the density measurements will hence cause drastically different conclusions about the shock properties. Such temperature and velocity dependences of the density calibration of several plasma instruments as found by Petrinec and Russell (1993) would cause such uncertainties across almost any shock crossing. Therefore, it is not recommended to use the density ratio as a primary parameter in shock normal determination unless the density calibration on both sides of the shock is well known.

Chao (1995) shows that in many cases, the parameters derived using the best fit to all R–H relations are actually far from possible solutions of the R–H relations. This result indicates that because of the complicated relationship among different variables, the values of the unmeasured quantities determined from a best fit actually do not represent well the real values of these quantities for each set of the parameters at a given point. It is not recommended to use a best fit of all R–H relations.

In summary, using only magnetic field measurements one can determine the normal direction of a discontinuity in simple cases. The minimum variance method applies only if there are measurements within a discontinuity. It has a large uncertainty for either a tangential discontinuity or a shock. The tangential discontinuity and coplanarity analyses impose strong assumptions on the type of a discontinuity.

They should be used carefully in combination with the minimum variance analysis and plasma measurements.

3. Wave Analyses

Wave analyses are used to determine the properties and characteristics of waves, such as propagation direction, wave frequency, wave amplitude, polarization, and mode. With measurements at multiple locations one can determine the phase speed, wave number, and etc. The waves of concern here are at frequencies near or lower than the ion gyrofrequency, or so-called ULF (Ultra Low Frequency) waves, measured in the spacecraft frame. Waves of higher frequencies will be discussed briefly in Section 3.6. We will focus on electromagnetic waves with emphasis on MHD waves. Methods for wave analysis based solely on the magnetic field measurements have been well developed and will be discussed as routine procedures in Section 3.2. From measurements at a single site these analyses can provide only the propagation direction (with an ambiguity in sign) and wave parameters in the spacecraft frame. To resolve the Doppler shift and to determine the direction of propagation, one needs to either use more than one satellite and plasma measurements or to make assumptions on the mode of the wave and use the dispersion relation calculated from theory as an accurate description of the wave. We note that dispersion relations developed for small amplitude linear waves may not be accurate for the large amplitude waves encountered in space. Also dispersion relations that ignore the interaction with the gyro and thermal motions, such as in the Hall-MHD treatment, may be inaccurate in the moderate and high beta collision-free plasmas in space. Related issues are discussed in Section 3.3. To identify the mode of a perturbation is crucial to understand the underlying physics of the processes that generate the wave. Several schemes to make the identification have been developed in recent years and will be discussed in Section 3.4.

3.1. BACKGROUND

The most powerful tool in general use for wave analyses is the Fourier analysis. Its principles can be found in most time-series data-analysis textbooks. However, in space physics, we deal with Fourier analysis of vectors. For a wave the perturbations in different components of a vector are correlated. This relationship defines a covariance matrix in the frequency domain, or the so-called spectral matrix. The Principal Axis Analysis can be used to analyze the covariance matrix in much the same way as it is used above in the time domain analysis (McPherron et al., 1972). Fourier analysis, the resultant spectral matrix, and the Principal Axis Analysis form the main ingredients of the most popular wave analysis technique for waves in space plasmas.

3.1.1. *Approximations*

Most of the existing wave analysis techniques in the time domain are based on the assumption that there is one dominant wave being analyzed. Most time series arise from a combination of many waves from various sources. If the coupling among different coexisting waves is weak, the analysis can be performed in the frequency domain. Spectral analysis transforms the time series into a simple superposition of waves of different frequencies. Usually when we analyze a wave we treat a finite band of frequencies, i.e., several successive Fourier amplitudes are grouped together. At each of the individual frequencies there is an amplitude and a phase, or a cosine (in phase) and a sine (quadrature phase) amplitudes. If there is a single source of all the wave amplitudes seen in this band, then the ratios of the amplitudes seen in two different components of the magnetic field over the band will be fixed (but not necessarily the amplitudes themselves) this quality is measured by the parameter called coherence which varies from zero to unity. In practice a random signal has a non-zero coherence whose value is determined by the number of Fourier estimates in the band analyzed.

Many magnetospheric waves appear with clear sinusoidal patterns, however, they may not be propagating but standing, in the sense they are the sum of two waves propagating in opposite directions with nodes in the ionosphere. For these waves, the Fourier analysis to be discussed in the following is applicable, but the matrix analysis is not. We note that there is another type of standing wave, one in which the wave propagation velocity is exactly balanced by the plasma bulk velocity so that the wave train stays fixed relative to its source. Such waves can be analyzed with the matrix method if the motion of the spacecraft or the source carries the observer through the wave train. We note further that the special property of the whistler mode that the group velocity exceeds the phase velocity allows energy to be pumped into a phase standing wave.

If the amplitude of the fluctuations is large, the harmonic generation associated with the nonlinear effects will affect the analysis. We discuss this issue in Section 3.5.

3.1.2. *An Ideal Wave*

The magnetic field vector can usually be measured relatively easily with adequate accuracy and time resolution. Moreover, it carries most of the Poynting flux in electromagnetic waves in plasmas of interest to us. (Here we recall that the ratio of the electric to magnetic energies for an electromagnetic wave equals the ratio of the phase velocity to the speed of light.) Thus it is often used to model wave fluctuations. The magnetic vector for a monochromatic plane wave propagating in the z direction is

$$\begin{aligned} B_x &= B_{x0} + \delta B_x \exp[-i(\omega t - \mathbf{k} \cdot \mathbf{r})] , \\ B_y &= B_{y0} + \delta B_y \exp[-i(\omega t - \mathbf{k} \cdot \mathbf{r} + \phi)] , \end{aligned} \quad (3.1)$$

$$B_z = B_{z0} ,$$

where subscript 0 and prefix δ denote the average and deviation from the average, respectively, and ω , \mathbf{k} , \mathbf{r} and ϕ are the frequency, wave number, spatial vector, and the phase, respectively. Since $\nabla \cdot \mathbf{B} = 0$, $\mathbf{k} \cdot \delta \mathbf{B} = 0$ for plane waves and therefore $\delta B_z = 0$. This is a most important feature in wave analyses. The power spectral matrix is

$$P(\omega) = \begin{pmatrix} \delta B_x^2 & \delta B_x \delta B_y e^{i\phi} & 0 \\ \delta B_x \delta B_y e^{-i\phi} & \delta B_y^2 & 0 \\ 0 & 0 & 0 \end{pmatrix} , \quad (3.2)$$

where $P_{ij}(\omega) = B_i(\omega)B_j^*(\omega)$; $i, j = x, y, z$; and $B_i(\omega)$ is the Fourier transform of $B_i(t)$. The asterisk denotes the corresponding complex conjugate.

In general, the off-diagonal elements can be written as a real part plus an imaginary part. The real (imaginary) part corresponds to the component of which δB_i and δB_j are in or 180° out of phase (of phase shift of $\pm 90^\circ$).

The intensity of the wave is defined as

$$I(\omega) = \text{Tr}P(\omega) = \delta B_x^2 + \delta B_y^2 \quad (3.3a)$$

and the ellipticity (Rankin and Kurtz, 1970)

$$\epsilon = \tan \psi , \quad (3.4a)$$

where

$$\sin 2\psi = \frac{2 \text{Im}(P_{xy})}{[(\text{Tr}P_{xy})^2 - 4|P_{xy}|]^1/2}$$

and P_{xy} is the 2×2 subtensor of P . For a linearly polarized wave, $\phi = 0$ or π , $\epsilon = 0$. For a circularly polarized wave, $\phi = \pm \frac{1}{2}\pi$, $\delta B_x = \delta B_y$ and $\epsilon = \pm 1$, where the plus sign is for right-hand polarization and the minus sign is for left-hand polarization.

3.1.3. Spectral Analysis

For a given interval, the spectral matrix $P(\omega)$ can be evaluated for a selected frequency range, $\Delta\omega$. In general the matrix can be written as

$$P = \text{Re}(P_{ij}) + i \text{Im}(P_{ij}) . \quad (3.5)$$

Its real part is symmetric, and its imaginary part is antisymmetric and consists of only off-diagonal elements because the covariance matrix is Hermitian. The real part can be diagonalized using the Principal Axis Analysis (see Section 1.4). In principal axis coordinates, the matrix is

$$\mathbf{P}' = \begin{pmatrix} \lambda_1 & & \\ & \lambda_2 & \\ & & \lambda_3 \end{pmatrix} + i \operatorname{Im} \begin{pmatrix} 0 & P'_{xy} & P'_{xz} \\ -P'_{xy} & 0 & P'_{yz} \\ -P'_{xz} & -P'_{yz} & 0 \end{pmatrix}, \quad (3.6)$$

where $\lambda_1 > \lambda_2 > \lambda_3$ and are real. The matrix for isotropic noise is

$$\mathbf{P}_{\text{noise}} = \alpha \begin{pmatrix} 1 & & \\ & 1 & \\ & & 1 \end{pmatrix}. \quad (3.7)$$

Comparing the above expressions with Equation (3.2), one finds that λ_3 corresponds to the noise in the \mathbf{k} direction. The wave intensity should then be defined, if the noise is isotropic, as,

$$I = \lambda_1 + \lambda_2 - 2\lambda_3. \quad (3.3b)$$

As we discussed in Section 3.1.2, each imaginary element gives the correlation of the fractions of the two corresponding components that have $\pm 90^\circ$ phase shift. $\operatorname{Im}(P'_{ij})$ should be equal to or less than $\sqrt{\lambda_i \lambda_j}$. If \mathbf{k} is along \hat{z}' , $\operatorname{Im} P'_{xz}$ and $\operatorname{Im} P'_{yz}$ should be very small. For a purely elliptically polarized wave with isotropic noise, $\operatorname{Im} P'_{xy} = \sqrt{(\lambda_1 - \lambda_3)(\lambda_2 - \lambda_3)}$, and $\sqrt{\lambda_1 - \lambda_3}$ and $\sqrt{\lambda_2 - \lambda_3}$ are the lengths of the major and minor axes of the polarization ellipse, respectively. For a nearly linearly polarized wave, $\lambda_2 \approx \lambda_3$. The ellipticity is

$$|\epsilon| = \sqrt{\frac{\lambda_2 - \lambda_3}{\lambda_1 - \lambda_3}} \quad (3.4b)$$

and its sign is the same as of $\operatorname{Im} P'_{xy}$. The sense of the polarization is determined by the sign of $\operatorname{Im} P'_{xy}$ (right hand for plus and left hand for minus). The amplitude of the wave can then be defined as

$$a = \sqrt{\frac{I}{1 + \epsilon^2}}. \quad (3.8a)$$

A linearly polarized wave can be decomposed as two circularly, but with opposite senses, polarized waves. An elliptical polarized wave can also be decomposed into two circularly polarized waves with opposite senses but different amplitudes. Therefore, every wave has left-hand and right-hand components. One can show that the amplitudes of the right- and left-hand components are (Kodera et al., 1977)

$$a_{\begin{matrix} R \\ L \end{matrix}} = \frac{(\lambda_1 - \lambda_3) \pm \operatorname{Im} P'_{xy}}{2\sqrt{\lambda_1 - \lambda_3}} = \frac{\sqrt{\lambda_1 - \lambda_3}}{2} (1 \pm \epsilon). \quad (3.8b)$$

3.2. ROUTINE WAVE ANALYSIS

In practice a time series of data usually contains some flags and gaps. Before any wave analysis, one has to remove these flags and fill the gaps. Sometimes, spikes in the data can be considered as flags if their time scale is much smaller than the period of the wave being studied. In general, deflagging reduces the wave power at higher frequencies and degapping adds wave power to low frequencies. Some instruments are operated periodically leaving periodic gaps in the data. The gaps will create false wave power in the corresponding periods. After the preparation of the data, one is ready for the routine wave analysis.

3.2.1. *Coordinate System*

A wave can be analyzed more easily by choosing a proper coordinate system. Since one way to characterize a wave is to see if the wave is compressional or transverse, we suggest the use of field-aligned coordinates as discussed in Section 1.2.2. In this coordinate system, if the perturbations are mainly along the background magnetic field direction, the wave is compressional, otherwise it is transverse.

Assume in magnetic field coordinates the field direction as \hat{z} . In general the spectral matrix is not diagonal, P_{zz} is the compressional power and $P_{xx} + P_{yy}$ is the power transverse to the field. The real part of an off-diagonal element $\text{Re } P_{ij} / \sqrt{P_{ii} P_{jj}}$ gives the portion of which the two components are in phase (when positive) or 180° out of phase (when negative). Note here ij satisfies the right-hand rule, otherwise a minus sign needs to be added. Similarly, $\text{Im } P_{ij} / \sqrt{P_{ii} P_{jj}}$ gives the portion of the signals that are $\pm 90^\circ$ out of phase. The properties of a wave can be decided by careful examination of each element of the spectral matrix as will be discussed in Sections 3.2.3 to 3.2.6.

3.2.2. *Detrending*

Most wave analysis methods essentially perform Fourier analysis on a segment of a time series data. A critical but implicit assumption of Fourier analysis is that the time series is periodic. Thus a slow trend in the signal is transformed into a sawtooth variation by the Fourier analysis technique. Figure 3.1 shows the effects of various detrending methods on subsequent Fourier analysis. The top line shows the Fourier spectra of linear, quadratic or cubic trends of the field increasing from zero to one nanotesla. The spectra for the three trends are almost identical. The slope is about -2 . This slope will have a profound effect on studying the slopes of spectra as discussed in Section 3.2.3. The corresponding amplitude can be as large as 10% in the low frequency range and decreases one order for every two-order increase in frequency. A non-zero average of a constant signal has only minimal effects on the spectral analysis as shown in the lowest line in Figure 3.1. A linear detrending can remove a linear trend completely and lower the effects down to the level of the non-zero average trace shown. Similarly quadratic detrending will remove the power due to the quadratic trend. A linear detrending is effective to reduce the error

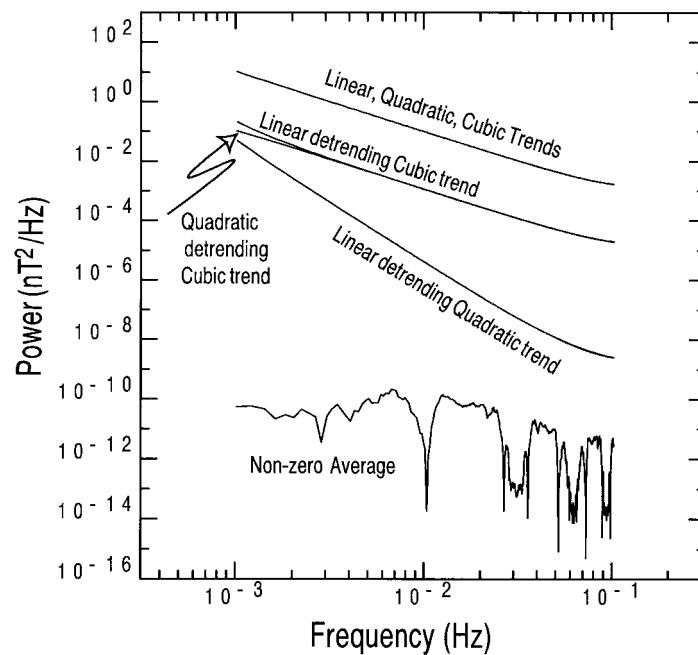


Figure 3.1. Effects of the trends in the background field on the Fourier analysis. The magnitude of the trend is one nanotesla over an interval of 2000 s. The lowest line show the spectrum of a constant background. The noise (10 orders smaller than the trend) is associated with the finite digitization of the calculation. The top line shows the spectra of linear, quadratic and cubic trends. The middle three lines show the spectra of each specified trend after detrending.

in higher frequencies for a quadratic trend, but not as effective at lower frequencies as shown in the second line from the bottom. The effects of a cubic trend are difficult to remove using either linear or quadratic detrending. The effects of higher order trends are especially important for magnetospheric wave studies since the measured background magnetospheric field compared with wave signals consists of strong higher order trends when a spacecraft moves radially through the Earth's field. A popular method, in addition to linear and quadratic detrending, is so-called prewhitening techniques. An example of such is the differencing method that analyzes the difference between two neighboring measurements. In principle, the differencing is similar to linear detrending between every two data points. Trends are removed unevenly throughout the interval being studied. It works very well in regions without discontinuities. In the differencing process, the power spectrum has been increased by a factor of ω^2 . The additional factor is slightly smaller than ω^2 at the low frequency end (Takahashi et al., 1990).

3.2.3. Power Spectrum

After detrending the data, one can examine the power spectra of the waves. To make the spectra, one selects a time interval which includes the wave activity of interest.

Observed waves often are not perfectly sinusoidal for a long period. How many wave cycles should be included for an analysis? The longest period in a Fourier spectrum is determined by the length of the selected interval. If only one wave cycle is selected, one would not obtain a meaningful peak in the spectrum. Furthermore, any residual trend in the background will strongly affect the analysis. It is common practice to select an interval containing at least five to six wave cycles in order to derive a statistically significant result. The more wave cycles are included, the better are its statistics, but the more chance has one to include waves of different sources or effects of source variation.

The spectrum can be spiky. To enhance the robustness of the result, one can average several Fourier estimates. The number of the total Fourier estimates in the average is called 'bandwidth'. Increasing the bandwidth improves statistical significance of the spectrum, reduces the resolution of frequency, and increases the lower-cutoff frequency of the spectrum.

To examine power spectra, firstly one may compare the power in the frequency range of interest for the different components and the strength of the field. Usually, the power for one or two components is much higher than that for the others. To determine whether a wave is compressional or transverse one can compare the power in the strength of the field P_t (the compressional component) with $P_x + P_y + P_z - P_t$ (the transverse component) where P_x , P_y , and P_z are powers in B_x , B_y , and B_z , respectively.

Secondly, one may look for peaks in the spectra. A sharp peak indicates that the wave is nearly monochromatic and a resonance of some sort may occur in the process being studied. However, if a sharp peak occurs exactly at the frequency of the spin of the spacecraft or at some multiple, it may be caused by an imperfect despinning process. Usually the spin tone occurs only in the two transverse components to the spin axis and usually little in the field strength. A broad peak may indicate that many wave modes contribute to the wave. If the plasma is not in a rapid motion, or the Doppler shift is small (see detailed discussions in Section 3.3), one may compare the peak frequency with the characteristic frequencies of the plasma, such as the gyrofrequencies. If the wave frequency is much smaller than the ion gyrofrequency, the wave is usually referred to, rightly or wrongly, as an MHD wave. If the wave frequency is near the ion gyrofrequency, the wave may be associated with ion gyromotion. Sometimes the trough between two peaks may be also of interest since it may be associated with resonant absorption by particles (Young et al., 1981; Anderson et al., 1991). Thus a trough could occur at a local characteristic frequency for certain species.

Thirdly, one may examine the slope of the spectrum (LaBelle and Treumann, 1988). The slope may provide some information about cascading processes in the frequency domain. Cascading processes describe the evolution of wave power in wave number. A wave cycle may break into two wave cycles with smaller wavelengths or two wave cycles may coalesce to form one wave cycle with a longer wavelength. If this process continues, a single wave mode may evolve into a spec-

trum with many wave numbers. However, we should point out that to compare the measured slopes with those given by theory is not as straight forward as many people thought. The cascading processes are described in the wave number domain in almost all theoretical work. They may appear differently in the frequency domain. In interplanetary space, since the solar wind velocity is much greater than the phase velocity for MHD waves, the measured spectrum is essentially due to the Doppler shift and hence is proportional to that of the wave number. Thus, a power spectrum in the frequency domain can be converted linearly into the wave number domain if the waves at different frequencies propagate in the same direction (which needs to be demonstrated), and is easily used in theoretical investigations. In other cases, the conversion from the frequency domain to the wave number domain may not be linear due to the dispersion. The slope of the power spectrum in the frequency domain may not be purely due to cascading. We should point out that an imperfect detrending process will strongly affect the analysis of the slope of a spectrum. One way to check this is to examine whether the amplitude in the lower frequency end is significantly smaller than a few percent of the trend in the background field.

3.2.4. *Coherence*

In any region in space several wave modes may coexist. Each mode has a particular polarization at a particular frequency. When we observe these waves we measure the sum of the wave modes and the noise of the observing system. One way to separate these wave modes is to check the cross correlations between different components as a function of frequency. Coherence analysis is a particular way of using the cross-correlations, i.e., the off-diagonal elements in Equation (3.5). The coherence is defined as the square-root of $|P_{ij}|^2 / (P_{ii} P_{jj})$. (See also Section 4.2). In the coordinate system of the waves, for a purely compressional wave, the field aligned component should have little coherence with the two transverse components. On the other hand, for a transverse wave, the two transverse components may have high coherence. Note, however, if one is not performing the analysis in the coordinate system of the waves so that any wave might excite all three sensors or directions of analysis, then high coherence could result in high coherence in all these possible pairs of signals. Combining the information gained from the coherence analysis with that from the power spectra, one may find sometimes that different peaks in the power spectra correspond to different wave modes. In coherence analysis, the phase difference between two components may also be calculated, $\phi_{ij} = \text{Im } P_{ij} / |\text{Im } P_{ij}| \tan^{-1}(\text{Im } P_{ij} / \text{Re } P_{ij})$. From the phase difference between the two transverse components, one may be able to determine the polarization of a transverse wave. The phase difference can also be used for mode identification as discussed in Section 3.4.

3.2.5. *Propagation Direction*

In principle, there are two methods to determine the propagation direction \mathbf{k} . Comparing Equation (3.6) with (3.2), in analogy of the minimum variance analysis

in the discontinuity analysis, one can take \mathbf{k} either parallel or antiparallel to the minimum eigenvector. This method is referred to as Born-Wolf method or principal axis analysis (Born and Wolf, 1965). This method is based on the information contained in the real part of the covariance matrix, sometimes called the cospectrum. Its applicability is limited by the ratio of λ_2/λ_3 in the same manner as the minimum variance analysis. An alternative method is based on the information contained in the imaginary part of the covariance matrix, sometimes called the quaspectrum. Means (1972) shows that the unit vector \mathbf{k}^0 can be given by

$$\begin{aligned} k_x^0 &= \pm \text{Im } P'_{yz}/IP, \\ k_y^0 &= \mp \text{Im } P'_{xz}/IP, \\ k_z^0 &= \pm \text{Im } P'_{xy}/IP, \end{aligned} \tag{3.9}$$

where $IP = (\text{Im } P_{xy}'^2 + \text{Im } P_{xz}'^2 + \text{Im } P_{yz}'^2)^{1/2}$, the plus (minus) sign is for right (left)-handed polarization.

The differences between the two methods were studied by Arthur et al. (1976), who suggested that the Born-Wolf method may be better for linear polarization and the Means method for circular polarization. One advantage of the Means method is that it provides the sense of the polarization.

For a linearly polarized wave, the propagation direction cannot be well determined from the minimum variance analysis. If the waves are mainly compressional, their propagation direction can be determined according to the coplanarity theorem assuming that they are either fast or slow modes (see more discussion on wave modes in Section 3.3). The coplanarity theorem requires that the background field, the perturbed field and the wave vector be coplanar for the fast and slow mode waves. Since the perturbed field is perpendicular to the wave vector, the propagation direction is (Russell et al., 1987)

$$\mathbf{k} \parallel (\Delta \mathbf{B} \times \mathbf{B}_0) \times \Delta \mathbf{B}, \tag{3.10}$$

where the direction of the perturbed field, $\Delta \mathbf{B}$, is in the direction of the maximum variance.

The propagation direction cannot be well determined for nearly linearly polarized intermediate mode waves.

3.2.6. Wave Properties

From principal axis analysis of the covariance matrix, one can derive properties of the wave such as its amplitude, ellipticity and compressibility using the definitions given in Section 3.1. For a linearly polarized wave the intermediate eigenvalue and minimum eigenvalue may be close to each other, and hence the propagation direction not well determined. In this case, the direction of the perturbed field which is the direction of maximum eigenvector is well determined. If the linearly

polarized wave is compressional (fast or slow mode), the propagation direction can be determined according to the coplanarity theorem. On the other hand, for a nearly circularly polarized wave, the maximum eigenvalue is close to the intermediate eigenvalue. In this case, the propagation direction is well determined. If all three eigenvalues are similar, say the differences are less than an order of magnitude, the fluctuations are turbulent. Neither the propagation direction nor polarization is well determined. Note that the wave properties are derived for only a selected frequency band. When one integrates over the entire frequency range, the results give the average properties over the whole spectrum and should be identical to the results of principal axis time series analysis. Wave analysis needs the background magnetic field for referencing the wave polarization. Thus if the wave properties are obtained after detrending or filtering, one must make sure that the background field be made available in the analysis.

3.2.7. *Filtering*

Filtering the time series data can be used to examine the behavior, amplitude for example, of the wave changing with time or location in space. The interesting issues here are generation or damping of the waves, the nodes of the field line resonance and wave packets. Filtering may be performed in the time domain or the frequency domain. In the time domain filters may be symmetric about the point of interest with identical weights multiplying the data on either side of the central point. Such filters have no phase lag, but have the disadvantage for real time use that they look forward in time. Filters can also be recursive using only data previous to the point in question. Such filters have phase variations with frequency but can be used in real time situations. Filtering can also be done with weighting the powers in the frequency domain. Such a filtering process is accomplished by multiplying a filter function by the Fourier spectrum to be filtered, and then inverting the resultant spectrum back to the time domain. An ideal filter function removes the variations in the unwanted frequency range while keeping the rest unchanged. However, in reality, this process may either introduce artificial fluctuations into the resulting time series or leave some residual power in the frequencies to be filtered.

Figure 3.2 shows examples of commonly used filter functions using time domain weighting. One selects a filter according to the requirements of his/her analysis. There are three major concerns in choosing a filter: how close to unity its pass-band is, how close to zero its stop band is and how sharp the cutoff is. In general, each filter is good in one or two of the three aspects. For example, the rectangular window, Figure 3.2(a), has a sharp cutoff, but its stop band is not clean and it creates artificial waves in the pass band. The Kaiser window, Figure 3.2(c), has the broadest cutoff transition but with the lowest stop band response. The Hanning window, Figure 3.2(b), is a compromise between the rectangular and Kaiser windows and is widely used in data analysis. Hamming window, not shown, is similar to the Hanning window and is also popularly used.

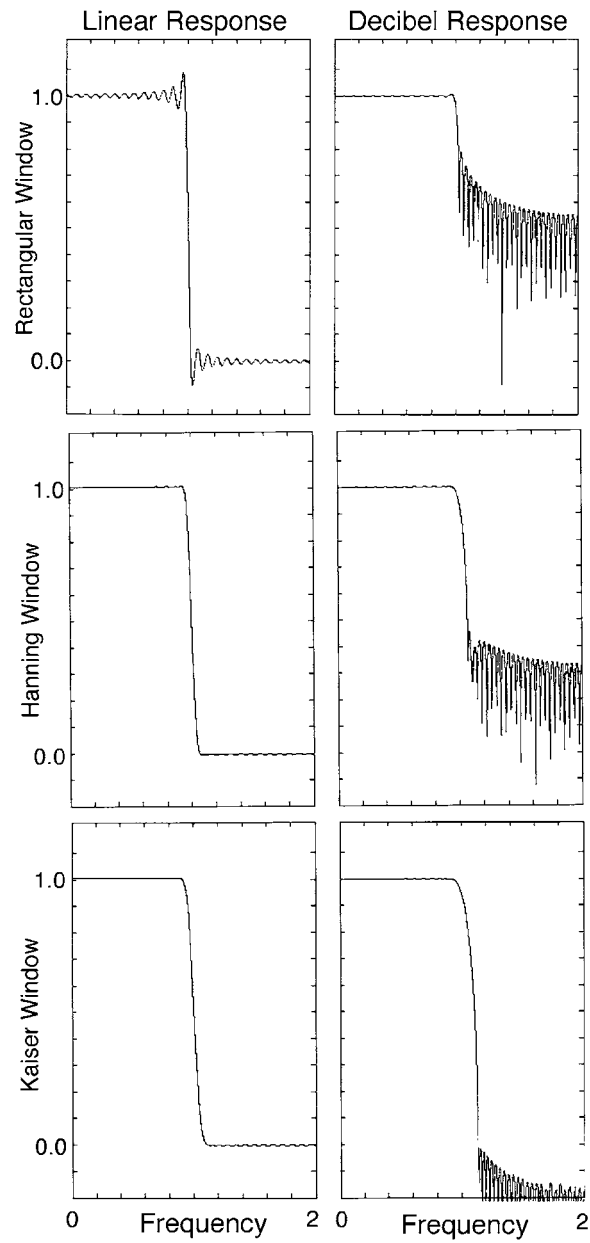


Figure 3.2. Response functions of three low-pass filter windows. (a) Rectangular window, (b) Hanning window, and (c) Kaiser window. The left panels show the linear response and right panels show decibel response. The Nyquist frequency is 2 Hz and the high-cutoff frequency is 1 Hz.

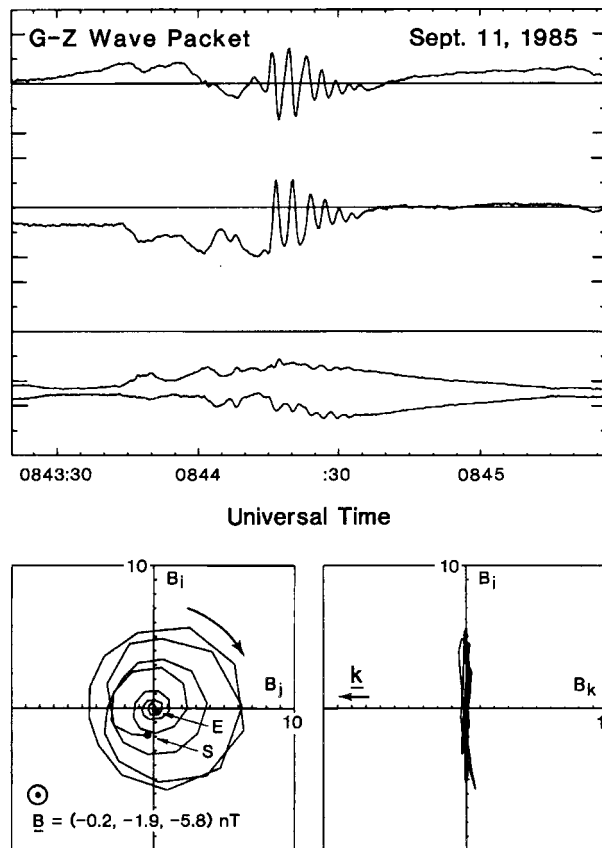


Figure 3.3. Hodograms of the wave packets upstream of the bow shock of comet Giacobini-Zinner (Le et al., 1989). The upper frame shows the magnetic field measurements in minimum variance coordinates. The lower two frames show the hodograms with the average field removed.

3.2.8. Hodogram

The polarization and the decay or growth of a wave can be most clearly presented in form of hodograms (e.g., Russell et al., 1971; Le et al., 1989), see Figure 3.3 for example. Hodograms can be made either before or after filtering the data. In particular, hodograms should be able to illustrate clearly the properties derived from the wave analysis. If not, one should check over his/her analysis carefully.

3.2.9. Dynamic Spectra

A particular useful display of spectra is called the dynamic spectrum a time series of power spectra. It contains the information discussed above in Sections 3.2.3 and 3.2.7. One may also examine the time evolution of the coherence and ellipticity of the waves. To understand the properties of a wave, one still has to perform the principal axis analysis, either explicitly or implicitly.

To make a dynamic spectrum, one needs to determine the number of data points, or the length of the window, for each individual spectrum and the overlap of the time series for successive spectra. Although a broad frequency range is presented in a dynamic spectrum, the window length should be determined according to the wave of most interest and the guidelines discussed in Section 3.2.3. The lower frequency portion is in general under-represented and the higher frequency portion may be less coherent. If a wave is present only over a finite time interval, in the dynamic spectrum, it may ‘propagate’ outside of the finite interval because of the length of the moving window. The presence of a discontinuity or a single spike in the time series can cause much serious problems in a dynamic spectrum analysis. The Fourier spectrum of a discontinuity or spike is broad-banded. It usually appears as a strip across all frequencies. Because an automated procedure to make a dynamic spectrum does not recognize a discontinuity, a strip with enhanced power will extend to a width of the window length on each side of the discontinuity (from the start to the end when the window includes the discontinuity). This could be potentially confused with a broadband wave of a finite region (unfortunately, this is not a hypothetical problem.) Therefore, it is extremely important to examine the original time series data and not to simply rely on a computerized automated dynamic spectral analysis, for example, in a statistical study.

3.3. MODE IDENTIFICATION

With the routine wave analysis discussed in Section 3.2, we have determined the frequency range of the wave of interest and overall properties of the wave, such as the propagation direction, polarization, and frequency. However, since these results are obtained in the spacecraft frame, the frequency in the plasma frame, wavelength, and phase velocity of the wave in general remain unknown. The magnetic field measurements from a single spacecraft alone in principle cannot determine these wave parameters. To further understand a wave, one needs measurements either from separated spacecraft or from plasma instruments.

Two major parameters of interest are the phase velocity and the frequency in the plasma frame (the wave number then can be calculated). In this section we will discuss how to identify a wave mode which gives the range of the phase velocity and physical functions of a wave. In the next section, we will resolve the Doppler shift and determine the frequency in the plasma frame.

3.3.1. *MHD Modes in Homogeneous Plasmas*

Waves at frequencies much lower than the ion gyrofrequency can be treated as MHD waves in low beta plasma ($\beta \leq 0.1$). We note that it is common to refer to these low frequency waves as MHD waves even when the wave dispersion cannot be derived by the MHD assumption. In the recent years, there has been an increasing awareness that the dispersion and other properties of a wave described by the MHD theory are different from that by the kinetic theory and may not be

appropriate for moderate and high beta plasmas, with which we are often dealing in space. Nevertheless, we base our discussion on MHD theory and point out where caution should be taken.

In MHD linear theory for isotropic uniform plasmas, there are four modes, the fast, intermediate, slow and entropy modes (e.g., Kantrowitz and Petschek, 1966; Kivelson and Russell, 1995). There is an ongoing debate on the existence of the MHD slow mode. In linear kinetic theory, the slow mode is strongly Landau damped in high β plasmas when the electron temperature is smaller than the ion temperature, as it is in the magnetosheath, plasma sheet and most of the magnetosphere, and its phase velocity is greater than the intermediate mode velocity (Gary, 1992). However, in data analyses an observation should not eliminate the slow mode from consideration solely on the basis of theory, because conditions may exist that maintain this mode in the face of damping. In the following discussion, we will refer to the intermediate mode as the Alfvén mode to avoid the implication of the ordering in the phase speeds.

The entropy mode is a nonpropagating perturbation with a zero phase velocity. The phase velocities for the other three modes are derived from linear MHD theory, to be,

$$v_{\text{phase}(f,s)}^2 = \frac{1}{2}[(C_S^2 + C_A^2) \pm \sqrt{(C_S^2 + C_A^2)^2 - 4C_S^2 C_A^2 \cos^2 \theta}], \quad (3.11a)$$

$$v_{\text{phase}(a)} = C_A \cos \theta, \quad (3.11b)$$

where C_S , C_A , and θ are the sound speed, Alfvén speed and the angle between the wave vector and the background field. The subscripts f and s stand for the fast and slow modes and correspond to the plus and minus signs on the right-hand-side of Equation (3.11a). The subscript A stands for the Alfvén mode. Their dependence on θ and $\beta = 2C_S^2/C_A^2 \gamma$ is given in Figure 3.4.

Characteristic features of each mode can be obtained from the perturbation relations. The Alfvén mode is incompressible, and thus the perturbations should be transverse. Both fast and slow modes are compressible and hence contain perturbations in both field pressure and plasma pressure. For the fast (slow) mode, the two pressures vary in (180° out of) phase. These characteristics can be easily differentiated in the power spectra (see Section 3.2.3) and coherence analysis (see Section 3.2.4). In inhomogeneous plasmas, the Alfvén mode may contain variations in density and the field strength. However, it requires that the variations in the total pressure, the sum of the thermal and magnetic pressures, be zero. Here we can see how important it is to have an accurate intercalibration between the plasma and field measurements. Otherwise, an inhomogeneous Alfvén wave may be misidentified as a homogeneous slow wave, and vice versa. Under this situation, careful examination of the perturbation and propagation directions is extremely important.

The fast mode propagates more isotropically than the other two propagating modes. Since the other two modes do not propagate perpendicular to the magnetic

FRIEDRICHS DIAGRAMS

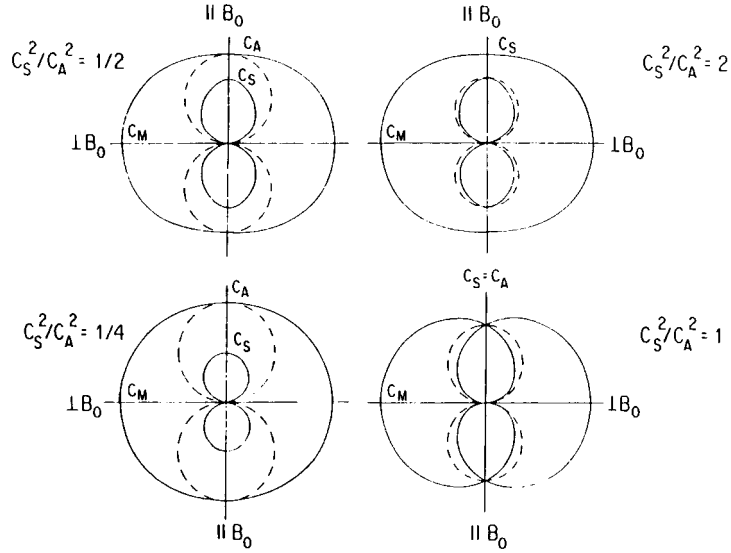


Figure 3.4. MHD phase velocities as functions of β and field direction. The field direction is along y . Ratio C_S^2/C_A^2 equals $\gamma\beta/2$, quantities C_A and C_S are the Alfvén and sound speeds, respectively. The phase velocity for the entropy mode is zero, i.e., at the origin of each frame. C_M is the magnetosonic speed for perpendicular propagation equal to the square root of the sum of the squares of the Alfvén and sound speeds.

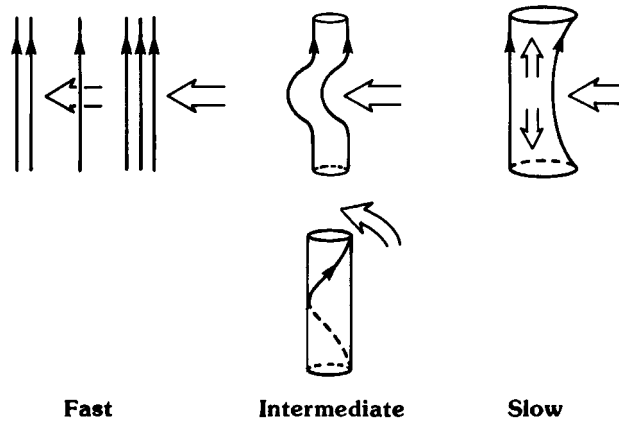


Figure 3.5. Physical functions of the three propagating MHD modes.

field, the fast mode most efficiently transmits the pressure perpendicular to the field. The Alfvén mode bends and twists the magnetic field and the plasma motion. The function of the slow mode is more interesting. As shown in Figure 3.5, if one applies a pressure perturbation perpendicular to a flux tube, the field strength will increase to conserve the flux while the cross-section of the tube decreases. If this is done slowly then, the thermal pressure will decrease because the pressure has a chance to equilibrate. Then the thermal pressure is anticorrelated with the magnetic field pressure, $B^2/2\mu_0$. This is accomplished by the plasma moving away from the compressed region. Thus the role of the slow mode is to convert the perpendicular pressure perturbations to parallel pressure perturbations. With these physical pictures in mind, one can more readily understand why a particular mode exists in a certain region.

3.3.2. *Mirror Mode*

The mirror mode instability can be derived from the MHD slow mode branch with inclusion of a finite temperature anisotropy, T_{\perp}/T_{\parallel} . However, Southwood and Kivelson (1993) showed that the instability is a result of kinetic effects. The unstable condition is

$$\frac{T_{\perp}}{T_{\parallel}} \geq 1 + \frac{1}{\beta_{\perp}}, \quad (3.12)$$

where β_{\perp} is the plasma beta evaluated using the perpendicular temperature. The mirror mode is a purely growing mode with zero frequency in the frame of plasma. It has a maximum growth rate when \mathbf{k} is about 70° from the field direction (Gary, 1992). The perturbations resulting from the instability convect with the plasma flow and are observed as oscillating waves. Krauss-Varban et al. (1994) showed that these perturbations correspond to the entropy mode in MHD. The mirror mode is expected to exist downstream of the bow shock (Crooker and Siscoe, 1977; Lee et al., 1988) and in outer magnetosphere where the plasma condition usually meets the mirror instability criterion. The perturbations associated with the mirror mode are similar to that of the perpendicular propagating slow modes. A major difference between the two is that the phase velocity is zero for the former but non-zero for the latter. However, when the slow mode phase velocity is small and the flow velocity is dominant, the two modes are difficult to distinguish as discussed in Section 3.4. The schemes discussed in the next subsection are designed to differentiate the slow and mirror modes.

3.3.3. *Transport Ratios*

One way to identify a wave mode is to measure the phase velocity and compare it with the expected dispersion relation (Hoppe et al., 1981). However, as will be discussed in Section 3.4, to measure the phase velocity needs at least two spacecraft and 3-D plasma measurements. A different approach is to determine the mode according to the ratios among perturbations of different quantities. These ratios

are referred to as transport ratios (e.g., Gary and Winske, 1992). Because of the difference in the roles of waves, each mode has a particular set of values of the transport ratios which can be calculated from theory.

3.3.3.1. *Transverse ratio.* The transverse ratio is defined as the ratio of the transverse component of magnetic wave power to the compressional component power, or

$$T_R = \delta \mathbf{B}_\perp \cdot \delta \mathbf{B}_\perp / \delta B_\parallel^2 \quad (3.13)$$

where $\delta \mathbf{B}_\perp \cdot \delta \mathbf{B}_\perp = \delta \mathbf{B} \cdot \delta \mathbf{B} - \delta B_\parallel^2$, and δB_\parallel is the amplitude in the magnetic field strength. When $T_R \gg 1$, the wave is transversely polarized and when $T_R \ll 1$, compressionaly polarized.

3.3.3.2. *Compressional ratio.* The compressional ratio is defined as the ratio of the compression in the plasma to the magnetic field perturbation, or

$$C_R = \frac{\delta N^2}{N_0^2} \frac{B_0^2}{\delta \mathbf{B} \cdot \delta \mathbf{B}}. \quad (3.14)$$

It represents the partition of the wave power between the plasma (density) and magnetic field. Since usually the error in the thermal pressure measurements is smaller than that in the density, as discussed in Section 1.3.3, it is recommended to use $\delta P/P_0$ to replace $\delta N/N_0$.

3.3.3.3. *Phase ratio.* It has been defined as

$$P_R = \frac{\delta P_i}{P_{i0}} \frac{P_{B0}}{\delta P_B} \quad (3.15a)$$

and also by

$$P_R = \frac{\delta N}{N_0} \frac{B_0}{\delta B_\parallel}. \quad (3.15b)$$

Since only the sign of the ratio concerns us, the two definitions are essentially the same. In practice, the noise is lower for the first definition. This ratio determines whether the compression in the magnetic field and plasma is in phase or out of phase. It is useful only for compressional waves. For incompressible waves, the two perturbations are dominated by noise.

3.3.3.4. *Alfvén ratio.* The Alfvén ratio is one of the earliest recognized transport ratios, and is defined as

$$A_R = \frac{\delta \mathbf{v} \cdot \delta \mathbf{v}}{C_A^2} \frac{B_0^2}{\delta \mathbf{B} \cdot \delta \mathbf{B}}. \quad (3.16)$$

It is particularly useful for theoretical investigations.

3.3.3.5. *Doppler ratio.* The Doppler ratio is defined as

$$D_R = \frac{\delta \mathbf{v} \cdot \delta \mathbf{v}}{v_0^2} \frac{B_0^2}{\delta \mathbf{B} \cdot \delta \mathbf{B}}. \quad (3.17)$$

It is different from the Alfvén ratio by a factor of M_A^2 where M_A is the Alfvén Mach number. This ratio is important because it contains not only the information on the velocity fluctuations but also on the background flow itself, which is not included in the Alfvén ratio. The background velocity becomes crucial when one is to determine whether the phase velocity is greater or less than, or is significant at all compared with the flow velocity. Under some approximations, the Doppler ratio is related to the ratio of the frequency in the rest frame of the plasma to that in the spacecraft frame.

An important feature of the above transport ratios is that they are independent of frequency. Therefore, they can be generalized in the frequency domain by assuming a spectrum is a simple superposition of waves of different frequencies. They are also independent of the propagation direction, and hence isolated from the errors and uncertainty that may occur when determining the propagation direction.

When comparing the observed transport ratios with those calculated from theory, one could find that none of the modes in theory completely match the observations. In order to characterize a fluctuation with a mode, one has to choose a mode that is ‘most likely’ to represent the wave. Different schemes will evaluate the ‘likelihood’ from different angles. Song et al. (1994) first introduced a hierarchical scheme; Denton et al. (1995) proposed a parallel scheme. Recently, Omid and Winske (1995) systematically investigated the mode identification problem using data from computer simulations.

3.3.4. *Hierarchical Scheme*

The scheme proposed by Song et al. (1994) is a qualitative and deterministic scheme. One follows the chart shown in Figure 3.6 and makes a yes or no decision at each point. The levels of the boxes have been determined according to the accuracies of the measurements.

In the solar wind, the Alfvén velocity is much less than the flow velocity and the lowest branch of this scheme does not apply.

3.3.5. *Parallel Scheme*

Denton et al. (1995) proposed a parallel scheme. In this scheme, all observed transport ratios are treated to be equally accurate. Each observed ratio is then compared with the theoretical values for all modes with different possible propagation angles. A mode is identified as the one with the smallest sum of the differences between theoretical and observational values of a select set of ratios.

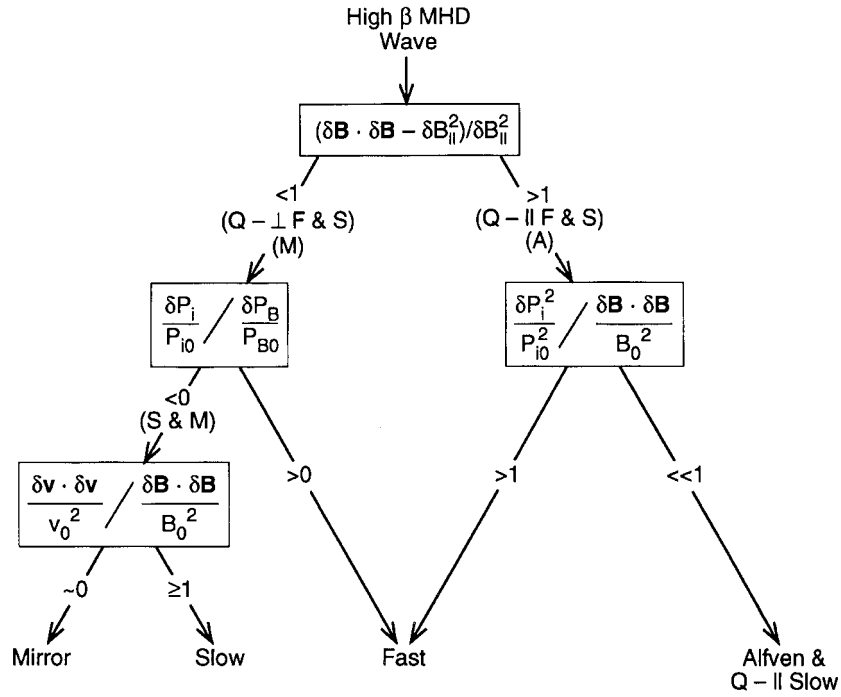


Figure 3.6. A hierarchical scheme of wave mode identification (Song et al., 1994). A given fluctuation can be distinguished among four different modes. At each step, the user makes a yes–no decision. The level of a box is determined according to the accuracy of the measurements. The scheme can be implemented in the frequency domain assuming that the waves are linear superpositions.

3.4. FREQUENCY AND PHASE VELOCITY

If the frequency of a wave in the plasma frame is ω and the plasma flows with a velocity \mathbf{V} relative to an observer, the frequency measured by the observer is

$$\omega' = \omega + \mathbf{k} \cdot \mathbf{V}. \quad (3.18)$$

This relationship can be derived from Equation (1.1) with a Fourier transformation. The second term on the right is the Doppler shift. Dividing both sides of Equation (17) by \mathbf{k} , we have

$$v' = v_{\text{phase}} + V \cos \eta, \quad (3.19)$$

where $v' = \omega'/|\mathbf{k}|$, v_{phase} , and η are the apparent velocity of the wave to the spacecraft, the phase velocity of the wave and the angle between the flow velocity and the propagation direction. The direction of the apparent velocity is along the propagation direction.

With plasma measurements and the routine wave analysis discussed in the last section, $V \cos \eta$, the Doppler velocity can be determined. The apparent velocity can be measured if there are two separated spacecraft,

$$v' = \frac{\mathbf{L} \cdot \mathbf{k}^0}{\Delta t}, \quad (3.20)$$

where \mathbf{L} and Δt are defined the same as Equation (2.20) and will be discussed further in Section 4. This method fails when the separation of the two space spacecraft is along the wave front, or $\mathbf{L} \cdot \mathbf{k}^0 = 0$. If the uncertainty in the determination of propagation direction which affects both η and v' is large, the result has an extremely large uncertainty.

There have been debates about whether the apparent velocity defined by Equation (3.20) corresponds to the phase velocity or group velocity. Because the phase velocity is the propagation velocity of the oscillations and the group velocity is the propagation velocity of the envelope of a group of oscillations, the group velocity corresponds to a time scale significantly longer than that of oscillations. Therefore, if the timing delay is measured from the frequency range of the oscillations, the apparent velocity corresponds to the phase velocity.

With measured frequency from the spectral analysis, the wave number, phase velocity and the frequency in the plasma frame can then in principle be determined. Without both 3-D plasma velocity and the timing difference measurements, to resolve the phase velocity, one has to make assumptions.

In the solar wind, the phase velocity of fluctuations is usually much smaller than the Doppler velocity. The phase velocity for entropy modes is zero in theory. In these two cases, the apparent velocity is similar to the Doppler velocity. The wavelength can be derived with only one spacecraft. It is important to point out however that to assume a fluctuation to be an entropy mode is to make a substantial physical assumption. For example, some oscillations in the magnetosheath which have been assumed to be mirror modes actually have a significant phase velocity (Song et al., 1992).

For quasi-standing waves, the apparent velocity is much smaller than the phase velocity. In this case, the Doppler velocity and the phase velocity are similar in magnitude but opposite in direction.

In summary, wave properties can be determined by combining plasma and field measurements, if a wave is nearly purely one of the four MHD modes. With more than one spacecraft, the apparent velocity of the wave to the spacecraft and hence the wavelength can be determined. The Doppler shift can be determined from 3-D plasma measurements and the minimum variance analysis. Therefore, the properties of the wave can be quantitatively determined with two of the above three determinations. The uncertainties for the three determinations are different. The determination of the apparent velocity has the least uncertainty but requires two spacecraft. The determination of the Doppler shift has the next to least uncertainty. However, if the flow velocity measurements are two dimensional, an additional assumption is needed and hence the uncertainty becomes larger. The phase velocity calculated from MHD dispersion relations has a relatively large uncertainty since the wave may not be purely a single mode and may not be in the linear stage, and may not be quantitatively described by MHD at all.

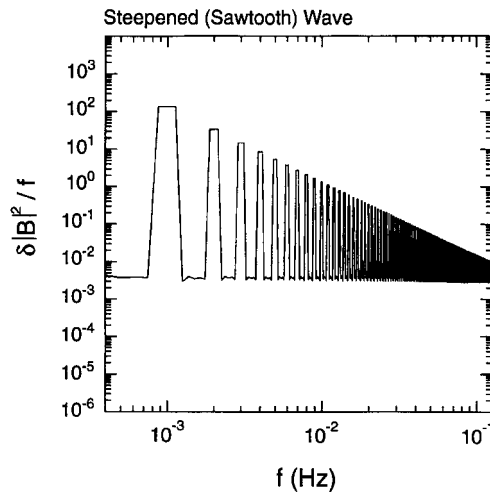


Figure 3.7. The Fourier spectrum of a steepened, or saw-tooth, wave. The frequency of the primary wave is 1×10^{-3} Hz and the amplitude of the perturbation equals the background field strength.

3.5. NONLINEAR EFFECTS

In the previous sections, we have discussed how to analyze waves of small amplitudes. If the amplitude is large, nonlinear effects will become important.

3.5.1. Steepened Wave

If nonlinear effects involve the temperature, a wave will steepen into saw-tooth type profile. Physically, the steepening process is a harmonics generation process: the power at harmonics of the fundamental wave frequency is enhanced. The Fourier spectrum of a saw-tooth wave, Figure 3.7, shows significant power in all harmonics. Therefore peaks at high frequencies in a spectrum may not necessarily indicate a set of waves, instead they could be simply the result of a single steepened wave. A visual inspection of time series data should effectively prove or disprove the possibility.

3.5.2. Large Amplitude Transverse Waves

It is possible that nonlinearity occurs only in the magnetic field in particular when the plasma beta is high. The nonlinearity in the magnetic field alone may not lead to wave steepening. Both the positive and negative variations in the transverse components increase the magnitude of the magnetic field. As the result, there are new issues in analyzing linearly polarized perturbations.

One of the most important issues in treating large amplitude field fluctuations is $\langle |\mathbf{B}| \rangle \neq |\langle \mathbf{B} \rangle|$ where $\langle \rangle$ denotes averaging in time domain. The difference is significant in evaluation of the Alfvén velocity, plasma β and relative wave amplitude of the field. Given a linearly polarized field perturbation $\mathbf{B} = (B_{x0} +$

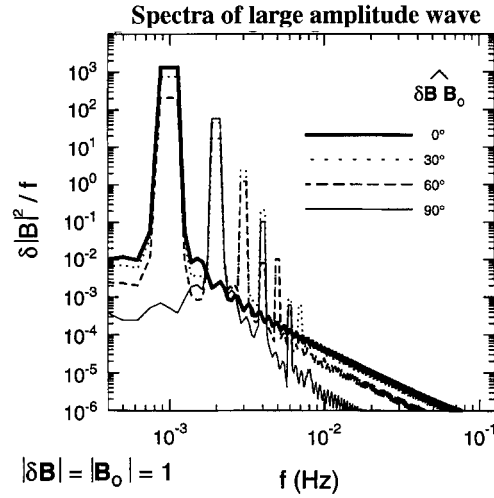


Figure 3.8. Power spectrum of the field strength for large amplitude field fluctuations. The frequency of the primary wave is 1×10^{-3} Hz and the amplitude of the perturbation equals the background field strength. The angles given in the legend are the angle between the background field and the perturbed field. Peaks appear at twice the fundamental and higher harmonics of the primary wave when the wave is not purely compressional.

$B_1 \cos \omega t, B_{y0}, 0$), the average field is $\langle \mathbf{B} \rangle = \mathbf{B}_0 = (B_{x0}, B_{y0}, 0)$ and the magnitude of the average field is

$$|\langle \mathbf{B} \rangle| = \sqrt{B_{x0}^2 + B_{y0}^2}. \quad (3.21)$$

The magnitude of the field $B = |\mathbf{B}| = \sqrt{(B_{x0} + B_1 \cos \omega t)^2 + B_{y0}^2}$. The average field strength is

$$\langle |\mathbf{B}| \rangle = \frac{1}{T} \int_{-T/2}^{T/2} |\mathbf{B}| dt. \quad (3.22)$$

The difference between $\langle |\mathbf{B}| \rangle$ and $|\langle \mathbf{B} \rangle|$ depends on the propagation angle. In general, $\langle |\mathbf{B}| \rangle$ is greater than $|\langle \mathbf{B} \rangle|$. They become the same for perpendicular propagation when $B_{x0} > B_1$. When comparing observations with theory, one has to decide which definition of the field strength corresponds to the value in theory. Most theories of waves assume that a wave is superposed on an average field. In these cases, one should use the definition given by Equation (3.21). Namely, one should use the strength of the average field, instead of the average of the field strength.

A most important phenomenon when analyzing large amplitude field fluctuations is the appearance of the higher harmonics in the field strength in particular the second harmonic because when a component of the field varies between positive

and negative, both positive and negative perturbations contribute positively to the field strength. The amplitudes of the harmonics in the field strength are

$$|B|_m = \frac{2}{T} \int_{-T/2}^{T/2} B \cos m\omega t \, dt \quad m = 1, 2, 3, \dots \quad (3.23)$$

The harmonics and their amplitudes in the field strength can be significantly different from those in the components. There is no general simple analytical expression for the integral in Equation (3.23). The first harmonic in the field strength would be dominant if $B_{x0} > B_1$. The second harmonic would be important if $B_{x0} < B_1$. Therefore, one may expect a significant presence of the second harmonic in the field strength for parallel propagation but not for perpendicular propagation. Figure 3.8 shows the Fourier spectra of the field strength that result from monochromatic large amplitude field fluctuations in one component ($B_0 = |\mathbf{B}_1| = 1$). The spectrum in the field strength depends strongly on the angle between the background field and the perturbed field, namely the angle between \mathbf{B}_0 and \mathbf{B}_1 . The power in the first harmonic decreases and the power in the second harmonic increases with the angle. Figure 3.9 shows an example in which wave power does not show in the field strength in the fundamental frequency but in the second harmonic. The dashed lines indicate spectra of white noise the energy density of which is independent of frequency. The most significant peak at the low frequency end of each component, as marked as F_1 , F_2 and F_3 has no significant corresponding peak in the spectrum of the field strength. This is consistent with nearly parallel propagating large amplitude waves as shown by the thin solid line in Figure 3.8. Peaks in the spectrum of the field strength clearly appear at twice the frequencies of the peaks in the components, as marked at $2F_1$, $2F_2$ and $2F_3$. At least one of them, $2F_2$, has no corresponding peak in the spectra of the components. Therefore, the second harmonic in the field strength is a major indicator of the nonlinear effects and creates problems in using wave analysis scheme based on linear theory, such as in Figure 3.6. If one examined only the spectrum of the field strength, this second harmonic peak could be mistakenly identified as a compressional wave. Therefore, it is important to check the spectra of the components: for a compressional wave, enhanced power should appear in at least one of the components.

3.6. PLASMA WAVE ANALYSES

Electric field measurements are often recorded with a dipole antenna mounted perpendicular to the spacecraft spin axis. During the spacecraft rotation, the electric potential differences along the sensors are measured. The electric field variations with scales longer than or comparable to the length of the antenna can be derived by examining these potential differences with respect to the spin modulation and antenna's orientation. These signals are often Fourier analyzed on-board and the Fourier components are transmitted to the ground. The magnetic field fluctuations

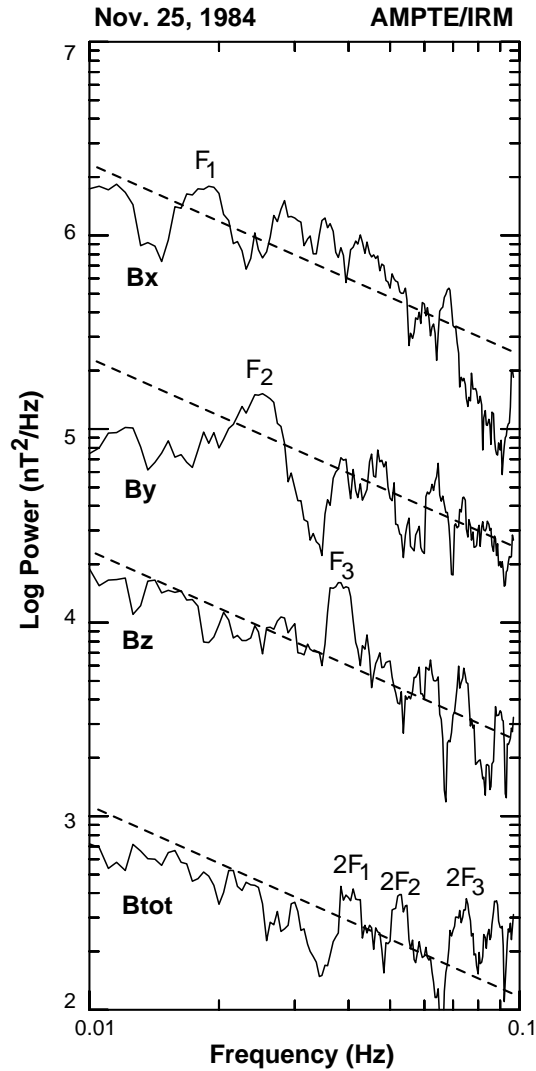


Figure 3.9. Fourier spectra of large amplitude (\sim average field) magnetic field perturbations. The scales for the three components have been shifted upward by one decade consecutively. The dashed lines indicate spectra of a slope of -1 . The most significant peak at the low frequency end of each component, as marked as F_1 , F_2 and F_3 has no corresponding peak in the spectrum of the field strength. But there are peaks at twice these frequencies in the field strength, as marked at $2F_1$, $2F_2$ and $2F_3$.

are usually measured by a search coil magnetometer in the same instrument package. If all six components of the electric and magnetic fields are available, one is able to determine the Poynting flux of each wave. Analyzing plasma wave measurements largely relies on the dynamic spectral analysis as discussed in Sections 3.2.3 and 3.2.9.

Here are some general guidelines to reading the spectra. If the enhancements appear in both the electric and magnetic fields, the waves are electromagnetic. From the Faraday's law, loosely speaking, the ratio between the power in the two components is the square of the phase velocity. For example, if the two spectra have a similar slope in a frequency range, the dispersion is weak in this range. If the strength of the electric component increases faster with frequency than the magnetic component, the phase velocity increases, and vice versa. If the enhancements occur only in the electric component, the waves are electrostatic. In this case, from Faraday's law, the wavevector is parallel to the electric perturbations and the phase velocity cannot be derived by using the ratio between the electric and (zero) magnetic amplitudes. Observationally, the latter may be largely determined by noise. Polarizations of the wave and the propagation direction (of electromagnetic waves) may be derived by examining the spin modulation of the signals (Scarf and Russell, 1988; Strangeway, 1991; Song et al., 1998).

To interpret a plasma wave is generally more difficult than low frequency waves. The first difficulty one encounters is whether the wave is generated where it is observed or not. Near the source region, a wave does not need to satisfy any particular dispersion because of the combination of the wave growth and possible strong spatial damping. One cannot assume a region of greater wave amplitude is the source region because a greater amplitude does not necessarily mean a greater wave energy flux. The phase velocity of a plasma wave is sensitive to the local plasma conditions. Strong waves usually occur in the region of rapid changes in plasma parameters. Many wave modes may be reflected at these boundaries which complicates the possible interpretations. One has to examine carefully the plasma parameters associated with the wave and estimate the unstable conditions for each of the possible modes.

4. Spatial Correlation Analyses

Correlation analysis studies the relationship between measurements from different observation points. The observers can be either separated spacecraft in space or different ground stations. When a disturbance or a wave travels from one observer to another, it is observed with a timing difference in the time domain or a phase difference in the frequency domain between the two observers plus any change of the disturbance during the travel time period. From the timing or phase difference and the separation between the two observers, one is able to determine the wave number of the wave. The change in the signal while propagating reduces the level

of the correlation between the two observed signals and thus the time or phase difference may be a function of frequency and possibly time.

4.1. BACKGROUND

4.1.1. Correlation Analysis in the Time Domain

If two observers measure the same quantity at different locations and the two time series data are $Q(0, t)$ and $Q(\mathbf{L}, t)$, where we have assumed that the location of the first observer as the origin and the separation between the two observers is \mathbf{L} , the cross-correlation coefficient is defined as

$$\alpha_{12}(\tau) = \frac{1}{K} \int_{t_1}^{t_2} Q(0, t - 1/2\tau) Q(\mathbf{L}, t + 1/2\tau) dt, \quad (4.1)$$

where t_1 and t_2 are the start and stop times of the selected interval, and K is the normalization factor and is

$$K = \left[\int_{t_1-1/2\tau}^{t_2-1/2\tau} Q^2(0, t) dt \int_{t_1+1/2\tau}^{t_2+1/2\tau} Q^2(\mathbf{L}, t) dt \right]^{1/2}.$$

The time lag with the highest cross-correlation coefficient can be used as the timing difference, Δt , between the two observers. The integration is realized through the summation over the time window. In general, a narrower width of the window provides a greater maximum of the correlation coefficient because the noise within the window width may be likely weaker. For discontinuity analysis, one should select a narrower window because for an ideal discontinuity, the cross-correlation coefficient decreases as $[1 + (|\Delta t - \tau|)/(t_2 - t_1)]^{-1}$. When the window is much greater than the time shift, the correlation coefficient appears flat. For wave analysis, a wider window, which includes several wave cycles will be statistically more significant.

If a wave is monochromatic, the correlation coefficient will periodically increase to unity and hence there is an ambiguity in determining the time lag. Fortunately, the waves in space are not purely monochromatic and one usually has no difficulty to find a peak in the correlation coefficient with the greatest value. One still has to be careful since in some situations it is possible that the greatest correlation occurs associated with observations of different wave fronts.

What limits the resolution of the timing difference? If a wave is purely monochromatic, the resolution of the timing difference may be infinitely high, and depends on the accuracy of the clocks rather than the resolution of the data. However, because the waves in space are not monochromatic, as just discussed above, and also because a wave may change slightly as it propagates from one observer to another, or be slightly different at some distance along the wavefront where the second observation is made, there is uncertainty in the timing difference. This

uncertainty is anti-correlated with the value of the highest correlation coefficient and equal to or smaller than the resolution of the data. For example, since a discontinuity departs most from an ideal monochromatic wave, the uncertainty in the timing difference is the highest and equals half of the time resolution of the data. To minimize the uncertainty in timing discontinuities, one should use the data with the highest resolution and filter the noise before an analysis. Otherwise, in common practice cross-correlation analysis is replaced by simply timing the delay of the discontinuity between the two observers. When multiple waves are present, each of them may have a different time delay. One needs to band-pass filter the data before the correlation analysis. Otherwise, the time delay will be determined by the wave of greatest power and the correlation coefficient will be smaller.

4.1.2. Correlation Analysis in the Frequency Domain

In the frequency domain, correlation analysis studies the phase difference between observations at two locations. From this phase difference, one may be able to determine the wave number.

If two time series data are $Q(0, t)$ and $Q(\mathbf{L}, t)$ and their Fourier spectra are $Q(0, \omega)$ and $Q(\mathbf{L}, \omega)$, the cross-correlation coefficient is

$$C_{12}(\omega) = \frac{\langle Q(0, \omega) Q^*(\mathbf{L}, \omega) \rangle}{K'(\omega)}, \quad (4.2)$$

$$K' = [\langle Q^2(0, \omega) \rangle \langle Q^2(\mathbf{L}, \omega) \rangle]^{1/2},$$

where the asterisk stands for the conjugate and $\langle \rangle$ indicates an ensemble average. The ensemble average is performed in time domain and equals the Fourier spectrum of the corresponding quantity in Equation (4.1) (e.g. Jenkins and Watts, 1968). Note here that Equation (4.2) is different from that given in most theoretical studies. In theoretical studies, the ensemble average is performed over wavevector. While the wave vectors are obviously known in theory, observationally we believe that with a very limited number of observers we are still far from being able to resolve multiple wave modes at a single frequency.

Since the correlation coefficient in the frequency domain is complex, it can be written as

$$C_{12}(\omega) = \kappa^2(\omega) e^{i\phi(\omega)}, \quad (4.3)$$

where $\kappa(\omega)$ and $\phi(\omega)$ are coherence and phase lag, respectively. The coherence gives a measure of wave versus noise. The noise can be either due to random fluctuations or due to waves with a finite coherence length compared with the separation between the two observers. The coherence length along a wavefront is the spatial scale of the wave. The coherence length normal to the wavefront is roughly the traveling distance of the wave, including both convection and propagation, within its decay time. For example, since the convection speed of the solar wind is much greater than the phase velocity, the waves in the solar wind often have

larger coherence length along the flow than transverse to the flow. The difference between random fluctuations and waves with a coherence length shorter than the separation is that the power spectra are broad for the former and at least one of them is peaked for the latter.

If the noise is due to random fluctuations, the phase difference of a wave between two observers is

$$\phi(\omega) = \mathbf{k} \cdot \mathbf{L} + h2\pi, \quad (4.4)$$

where $h = 0, \pm 1, \dots$. Thus, the wave number can be determined, but with the so-called 2π ambiguity. Here the direction of the wave vector is determined from the methods discussed in Sections 3.1 and 3.2. Since the frequency in Equation (4.4) is in the spacecraft frame, further determination of the wave properties needs to consider the Doppler shift, Equation (3.18). The time lag can be calculated from $\Delta t = \phi(\omega)/\omega$, which is a function of frequency.

In summary, correlation analysis can determine either the propagation speed of a discontinuity or the wavelength of a wave. For a discontinuity, the correlation analysis has to be performed in the time domain. For a wave, the analysis can be performed in either the time domain or the frequency domain. The analysis in the frequency domain contains the information about the phase as functions of frequency and hence can provide a better overall view of the waves at different frequencies. The result of an analysis in the frequency domain is in principle the same as in the time domain with prefiltering. These analyses are restricted to waves of a single source.

4.1.3. Relationship between the Quantities in Satellite Frame and Plasma Frame

In Section 1.1, we discussed the relationship between a quantity in the satellite frame and in the plasma frame for a single spacecraft. When comparing observations from two satellites at different times, the variations other than simple convection from one satellite to the other will affect the data analysis. These variations include spatial (between the two satellites) and temporal (between the two times) variations due to the physical evolution of the phenomenon being studied, and changes in the velocity and spacecraft separation.

The difference between the quantity Q measured by satellite i at location \mathbf{r} at time t and by satellite j separated by a distance \mathbf{L}_{ij} from satellite i at time $t + \tau$ is, see Figure 4.1 for the geometry,

$$\Delta_{\tau}^{ij} Q = Q_j(\mathbf{r} + \mathbf{L}_{ij} - \mathbf{v}\tau, t + \tau) - Q_i(\mathbf{r}, t), \quad (4.5)$$

where $\tau\mathbf{v}$ should be understood as $\int \mathbf{v} dt$. A time domain correlation analysis is to provide time shift $\Delta t = \tau$ with a minimal difference, i.e., $\Delta_{\tau}^{ij} Q$ goes to zero. Taylor expanding Equation (4.5) and retaining the terms up to the second order, the difference operator is

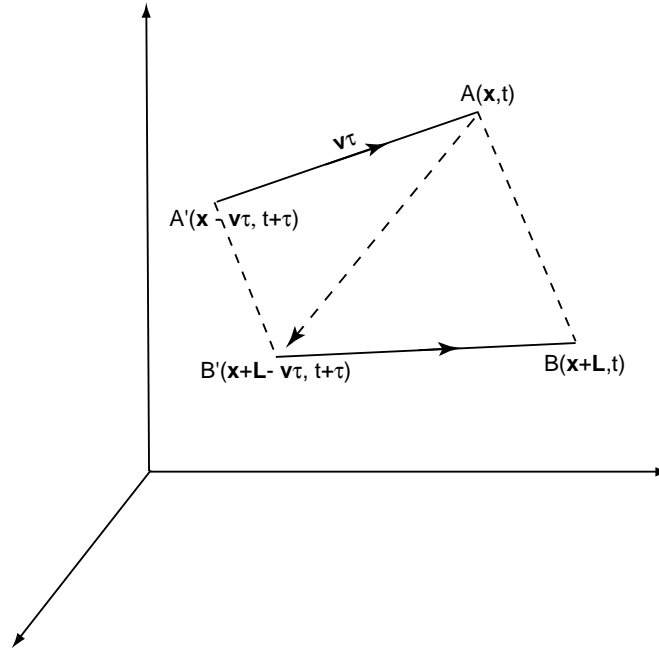


Figure 4.1. Geometry of two separated spacecraft relative to a wave front. Satellite A is located at point A and satellite B at point B. The plasma element that is measured by satellite A at time t and the plasma element at point A' is measured at $t + \tau$. Similarly, the plasma element that is measured by satellite B at $t + \tau$ is from point B' . A correlation analysis compares measurements at different locations and/or different times.

$$\begin{aligned}
 \Delta_{\tau}^{ij} = & \left[\left(\tau \frac{\partial}{\partial t} - \tau \mathbf{v} \cdot \nabla \right) + \mathbf{L}_{ij} \cdot \nabla \right] + \\
 & + \left[\frac{\tau}{2} \left(\tau \frac{\partial}{\partial t} - \tau \mathbf{v} \cdot \nabla + \mathbf{L}_{ij} \cdot \nabla \right) \frac{\partial}{\partial t} \right] - \\
 & - \left[\frac{\tau}{2} \left(\tau \frac{\partial}{\partial t} - \tau \mathbf{v} \cdot \nabla + \mathbf{L}_{ij} \cdot \nabla \right) (\mathbf{v} \cdot \nabla) \right] + \\
 & + \left[\frac{1}{2} \left(\tau \frac{\partial}{\partial t} - \tau \mathbf{v} \cdot \nabla + \mathbf{L}_{ij} \cdot \nabla \right) (\mathbf{L}_{ij} \cdot \nabla) \right].
 \end{aligned} \tag{4.6}$$

We should point out that the validity of the Taylor expansion relies on $\mathbf{L}_{ij} - \mathbf{v}\tau$ to be much smaller than L_c and τ to be much smaller than t_c , where L_c and t_c are the coherence length and coherence time, respectively. Whereas the measured quantity can contain discontinuities. Namely, \mathbf{L}_{ij} is not necessarily much smaller than the scale of the discontinuities (in the normal direction). As long as the correlation analysis provides a meaningful correlation, the Taylor expansion is valid. The terms in the first bracket are due to the linear effects. The main differences observed by the two spacecraft are caused by the temporal variations in the plasma frame and

the spatial non-uniformity of the plasma along the motion direction and along the separation direction of the two spacecraft. The function of the correlation analysis is to provide a τ that minimizes the value in this bracket. Namely, the correlation analysis converts linearly the temporal variations in the spacecraft frame into spatial variations in the plasma frame of reference. It is less capable of providing information about spatially static structures in the spacecraft frame. In general the nonlinearities in the temporal and spatial variations will result in a residue for these terms. This residue reduces the correlation coefficient of the analysis. The terms in the second bracket are due to the nonlinear effects of temporal variations. The terms in the third bracket are due to the nonlinear effects of nonsteady state, nonuniform plasma motion. Similarly, the terms in the fourth bracket are due to the temporal and spatial changes in the separation between the two spacecraft.

In most cases, the variations in the separation vector (the fourth bracket) are negligibly small. The velocity variation (the third bracket) could become significant near an oscillating boundary and for a circularly polarized wave the velocity of which could be twisted by the wave perturbations, in particular when the background velocity is small.

For a single spacecraft, $L = 0$, the first order terms recovers Equation (1.1). Note that the velocity in Equation (4.6) is not necessarily the plasma velocity but could be the velocity of any frame relative to the spacecraft. In a shock frame or a wave frame, $\partial/\partial t = 0$. The first order term is $(\mathbf{L}_{ij} - \mathbf{v}\tau) \cdot \nabla$. In the ideal 1-D case, $\Delta_{\tau}^{ij} = 0$ where τ is defined as the time delay with the maximum correlation coefficient, and it yields equations (2.20) and (3.20) by noting that the speed of the wave frame relative to the spacecraft is the sum of the phase velocity and the plasma frame speed.

4.1.4. *Data Analysis of Observations from a Cluster of Satellites*

A ‘cluster’ of spacecraft refers to four closely separated satellites. The relative locations of the satellites can be represented in general by an irregular tetrahedron. Regulated by orbital dynamics, the shape of the tetrahedron changes with time. If the change in their relative distance is much less than in the relative distance itself during the interval of the event being studied, the change of the tetrahedron may be neglected. The uncertainties in the determination of the relative locations and in the intracalibration of the instruments among the satellites determine the lower limit of the uncertainty of any analysis.

In principle, when the cluster is not coplanar, the propagation direction of a 1-D discontinuity or a 1-D steady state wave can be determined without the assumption that the propagation direction is along the minimum variance direction. With the approximation that the propagation direction is along the minimum variance direction, either the unsteadiness of the wave (in the wave frame), or a 2-D steady state structure or wave can be resolved under certain conditions. A moving flux tube with a constant shape is an example of a 2-D steady state structure. A purely periodic surface wave in the wave frame is an example of the 2-D steady state

wave. For a given perturbation, even with a noncoplanar cluster of satellites, there is no general conclusive interpretation. One has to make approximations about each of the extension in three dimensions (bounded, periodic, or uniform) and its time dependence (pulsating, periodic, or steady). Differences in these approximations lead to different interpretations, which has been the major cause of existing controversies in the science community. Each of these approximations, however, can be tested both using either the information from other instruments or techniques that are not included in this article as standard techniques, and using the constraints provided by statistics and physical consequences.

In the following sections, we discuss the application of correlation analysis to multiple spacecraft observations. Most discussion is based on the authors' experience from previous missions and recent results from other groups. We emphasize that the problems with massive multiple spacecraft data analysis have not yet been fully exposed. Significant modification may be needed when applying the following principles.

4.2. ONE-DIMENSIONAL DISCONTINUITY ANALYSES

As discussed in Section 2.1.1, to completely determine observationally the properties of a 1-D steady state discontinuity, more than one satellite is needed. The central issue is to determine the direction and magnitude of the velocity of the shock frame relative to the spacecraft without imposing the conservation laws which should be verified by the results. The timing difference between spacecraft i and j is

$$\Delta t_{ij} = \mathbf{L}_{ij} \cdot \mathbf{V}_{\text{disc}} / V_{\text{disc}}^2. \quad (4.7)$$

Note that, as discussed at the end of Section 4.1.3, the above expression is accurate to the first order and only for an ideal steady-state 1-D discontinuity. In comparison, Equation (1.2) does not require the steadiness approximation. With four spacecraft, \mathbf{V}_{disc} can be uniquely determined if the four spacecraft are not coplanar. With five noncoplanar spacecraft, Equation (4.7) can become overdetermined and \mathbf{V}_{disc} can be determined using best-fit, for example, or average over the several \mathbf{V}_{disc} obtained from each subset of the four time differences (Russell et al., 1983). In practice, one may obtain several time delays crossing a discontinuity for each pair of spacecraft, for example, one at the leading edge, one in the middle and one at the trailing edge, when using high resolution data. If one assumes that the difference between the two time delays is not due to temporal variation of the thickness, averaging these sets will improve the statistical significance of the results.

Using the resultant \mathbf{V}_{disc} and Equation (2.7), one can derive the quantities in the R–H relations. The accuracy of the time delay is limited by the temporal resolution of the data when the fluctuations within the discontinuity cannot be resolved. A greater time delay compared with the temporal resolution implies a

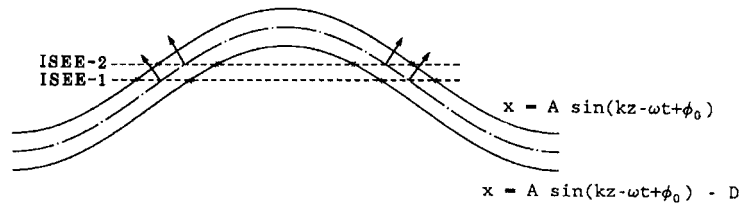


Figure 4.2. As a surface wave passes by two spacecraft, a few timing differences and normal directions can be obtained in order to determine the parameters of the surface wave.

smaller uncertainty in the timing, whereas during this longer interval, there is a greater chance for the discontinuity to change its properties. If the resolution is sufficiently high, the thickness of the discontinuity then can be determined from the duration of the crossing. If the spacecraft are relatively close, a major uncertainty arises from the determination of the relative locations of the spacecraft.

If the type of the discontinuity is known or the techniques discussed in Section 2 are applicable, a combination of the timing method with those in Section 2 will improve the statistical significance of the results. In addition, one may be able to find intervals when the discontinuity lies between the spacecraft and hence some of them provide the upstream conditions while others the downstream ones. Quantities in Equations (2.10) and (2.11) can be substituted with those simultaneously measured by different satellites instead of those measured at different times from the same satellite.

Since each method has its own uncertainty, when averaging results from different methods, how to weigh each of them becomes important. Russell et al. (1983) documented several cases and provided comprehensive discussions on the uncertainties.

Nonsteadiness of a discontinuity is observed as the difference in the duration of the discontinuity crossing registered among different satellites and the difference in the time shifts. It can be treated as the variation around the average. The nonsteadiness includes the temporal variations in the discontinuity's orientation (i.e., its normal direction) and the thickness, and the changes in the shock velocity relative to the satellites.

4.3. TWO DIMENSIONAL STRUCTURES

4.3.1. Surface Wave Analyses

Surface waves in magnetospheric physics refer to the wavy motion of a discontinuity. To describe such a motion, the most important parameters are the wavelength and the displacement, as well as the period. The wavelength is measured along the discontinuity surface in the direction of the surface wave propagation and the displacement is normal to the average surface. Simple boundary oscillations are a 1-D approximation when the wavelength is infinity. With a finite wavelength, the discontinuity is bent. As expected, the boundary normal varies. If the thickness of

the boundary is much smaller than the curvature of the boundary the discontinuity analysis methods discussed in Section 2 are applicable. The normal direction of the boundary can be determined by each of the crossings. The variation in the normal direction can then be measured by the satellites, see Figure 4.2 for example. With multiple satellites, one is able to determine the characteristics of a surface wave (Song et al., 1989). Mottez and Chanteur (1994) discussed a similar technique to analyze a general 3-D thin surface using a cluster of satellites.

A different approach has been recently developed using the 'remote sensing' technique (Walthour et al., 1993). As a boundary oscillates, the plasma and field parameters in the regions adjacent to the boundary vary. For example, the mathematical description of a surface wave is an exponential decay in the amplitude of the perturbation away from the boundary (e.g., Southwood, 1968). From these variations, a satellite can remotely sense the boundary and its motion when it approaches to and departs from the boundary. This new technique requires high quality measurements. It has been tested by using two-satellite data (Walthour et al., 1994).

4.3.2. *Flux Tube*

When the plasma beta is not extremely large, the anisotropy in a plasma introduced by the field often results in structures that have similar sizes in the two directions across the field and their greatest scale size along the field. Usually such a structure has properties distinct from its surrounding plasmas. Russell et al. (1990) provided a collection of comprehensive discussions on flux tubes from various regions in space. The majority of previous observations are based on single spacecraft measurements. The geometry of a structure, for example, a flux transfer event (Russell and Elphic, 1978) or a plasmoid, is often deduced from time series data (which provides the variation in one dimension), the region where it is observed on a statistical basis (which provides the scale of the second dimension), and the assumption that the structure is very long in the third dimension. Sometimes, observations from multiple satellites are available. Examination of these events confirms the general geometry people have developed for these phenomena.

A cluster of spacecraft can provide much more detailed information about these flux tubes. A general approach to analyzing these structures is similar to that of surface waves.

It is worth mentioning that since a flux tube does not satisfy the 1-D approximation, the minimum variance direction across the event may not have any definite meaning. For example, unless there are other indications, one should not take the minimum variance direction as the direction of the symmetric axis because the axial field may increase toward the axis due to a pinch process that arises from an axis-aligned current within the tube.

4.4. DETERMINATION OF ELECTRIC CURRENTS

The electric current is a very important physical quantity in space physics. However it is extremely difficult to measure in space. Because the conductivity in a collisionless plasma is poorly known, Ohm's law cannot be used to evaluate the current. In theory, there are two ways to measure the current. They are

$$\mathbf{J} = \sum_i q_i \rho_i \mathbf{v}_i , \quad (4.8)$$

$$\mathbf{J} = \frac{1}{\mu_0} \nabla \times \mathbf{B} , \quad (4.9)$$

where subscript i denotes each charge species. As discussed in Section 1.3.3, it is difficult to measure accurately the fluxes of the species that carry most of the current due to the intrinsic limitation of particle detectors and spacecraft charging. This is particularly true in collisionless plasmas where particle gyro motions are dominant. In most regions in the magnetosphere, the relative drift between ions and electrons is small. The limitation of instruments further reduces the ability of resolving such a small difference. For example, as shown in Equation (1.10), the effective cutoff velocity of a detector depends on the species. The spacecraft charge will have much greater effects on electron detection than on ion detection. When the spacecraft is strongly negatively charged, the difference in the errors in the velocity measurements will produce a large error in evaluating Equation (4.8) even if the detector has a similar small lower cutoff energies for electrons and ions. Although there are several reports on the determination of the electric current using Equation (4.8), see Frank et al. (1981), for a partially successful test, we remain cautious about quantitative use of the current from direct particle measurements.

4.4.1. *Single Satellite*

Equation (4.9) has been widely used to estimate the current. Almost all previous single satellite usage was based on a 1-D current sheet assumption. In 1-D

$$\begin{aligned} J_m &= \frac{1}{\mu_0} \Delta B_l / D , \\ J_l &= -\frac{1}{\mu_0} \Delta B_m / D , \end{aligned} \quad (4.10)$$

where D is the thickness of the current sheet across which $\Delta \mathbf{B}$ is evaluated. To determine D from a time difference, one needs to know the velocity of the current layer relative to the spacecraft. Therefore all discussions given in Section 2 on determination of the shock frame apply here.

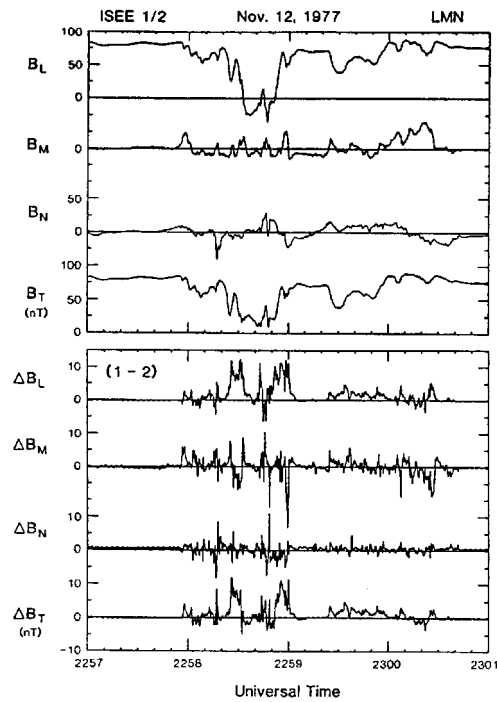


Figure 4.3. Field differences at two satellite locations which can be used to infer the current (Elphic, 1989). The upper four panels show the magnetic field measurements from two magnetopause crossings from ISEE 1. The lower four panels show the differences between the ISEE 1 and 2 field measurements.

4.4.2. Dual Satellites

When measurements from two satellites are available, there are two simple approaches to estimate the current. The first approach is to take the difference between the simultaneous field measurements from the two satellites assuming that the current sheet is one dimensional and lies between the two spacecraft. This corresponds to the $\tau = 0$ case in Equation (4.6). The first-order terms gives the same expression of the current as in Equation (4.10), but the field differences are evaluated from simultaneous measurements at different locations. The current can then be determined as functions of time (Elphic, 1989). The resultant time series is often highly fluctuating and spiky, see Figure 4.3, for example. The causes for such fluctuations could be multiple: there may exist physically meaningful small scale structures within the current layer, the relative motion among sub-current sheets in the normal direction can result in varying time-space conversion, and the 1-D approximation to these structures may break down. When the 1-D approximation breaks down, the time series of the field differences cannot be simply interpreted as currents.

Taking an orthogonal approach, one may compare the timing of the same field measured by the two satellites. This corresponds to setting the left-hand side of

Equation (4.6) zero. Under the assumption that the structure of the current is stationary, but the motion of the current changes with time, retaining the first order terms, the velocity of the current layer can then be calculated by dividing the satellite separation along the normal of the current by the timing difference between the two satellites' measuring of the same magnetic field, as shown by Equation (4.7). When the velocity of the current layer is known, from Equation (4.10) one can determine the current according to the time series of each satellite, see Figure 4.4. In theory, the currents derived from the two satellites should be the same if the current is in steady state. In fact, the resultant currents are different, indicating that the current is not in a steady state. How to reconcile this inconsistency remains to be investigated.

Often the two spacecraft are on the same side of the current sheet (after passing through or before the crossing) at different times. These data on the same side can be used to help test the 1-D time stationary assumption by combining the two approaches above.

4.4.3. Multiple Satellites

The above techniques rely on the 1-D assumption of the current. The measurements from a cluster of satellites will enable us to determine the spatial derivatives and hence the current in Equation (4.9) without the 1-D assumption. The spatial derivative method applied to a quasi-stationary current system is very different from the analysis of a fast moving current structure. In the former case, correlation analysis may not be able to provide a meaningful timing difference. Therefore the approach of setting τ equal to zero is used. In the latter case, τ can be determined and hence more information is available to measure the current and its motion velocity.

4.4.3.1. *Quasi-stationary currents.* A current system can be considered as quasi-stationary if $|\Delta^{ij}\mathbf{B}| \gg |\Delta_\tau\mathbf{B}|L_{ij}/\tau V$, where V is the velocity of the current system relative to the spacecraft (note here that when $V = 0$, τ goes to infinity). In other words, the measured temporal variation in the field is much smaller than the difference among satellites. This assumption is applicable for a cluster moving slowly through a current region that is thicker than the separation of the satellites. The electric current density can be determined by (Dunlop et al., 1988),

$$\mathbf{J}_{ijk} = \frac{1}{\mu_0} \frac{\Delta^{ij}\mathbf{B} \cdot \mathbf{L}_{jk} - \Delta^{jk}\mathbf{B} \cdot \mathbf{L}_{ij}}{(\mathbf{L}_{ij} \times \mathbf{L}_{jk})^2}. \quad (4.11)$$

The direction of \mathbf{J}_{ijk} is outward along the normal direction of the surface formed by spacecraft i , j , and k . Four such \mathbf{J}_{ijk} can be obtained. The current density within the cluster can be calculated by averaging the four current densities. The direction of the current, $\hat{\mathbf{J}}$ is parallel to the vector summation of \mathbf{J}_{ijk} s, and its magnitude is

$$J = \frac{(J_{ijk}|\mathbf{L}_{ij} \times \mathbf{L}_{jk}|)}{\hat{\mathbf{J}} \cdot (\mathbf{L}_{ij} \times \mathbf{L}_{jk})}. \quad (4.12)$$

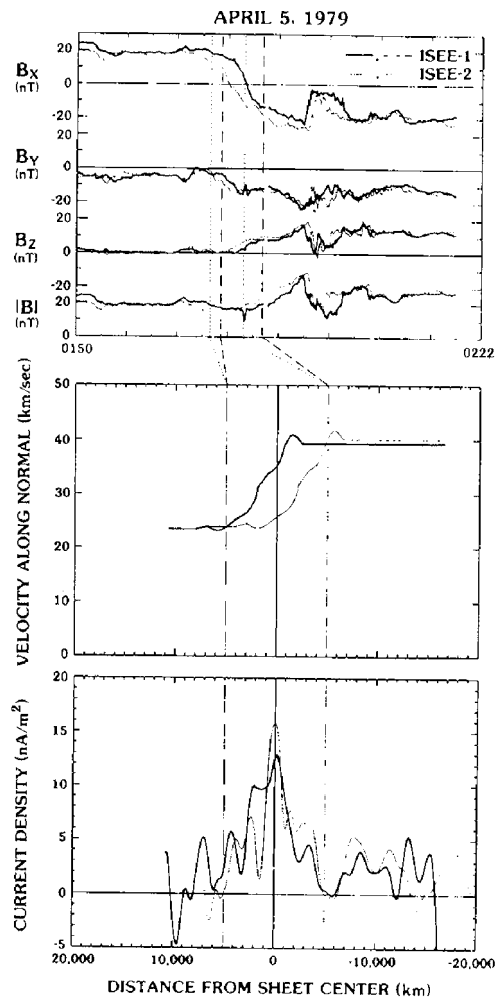


Figure 4.4. Two-satellite determination of the current (McComas et al., 1986). Under the 1-D steady state assumption of the current structure, the velocity of the current motion can be determined, middle panel, and then the current density can be derived, lower panel.

An important test of the result is to check whether the sum of $\mathbf{J}_{ijk} \cdot (\mathbf{L}_{ij} \times \mathbf{L}_{jk})$ is much smaller than each of them.

We should emphasize that the approach under the steady state approximation is not the same as the approach when the spacecraft separations are much smaller than the spatial scale of the current structure. When two spacecraft see basically different things, the correlation analysis does not provide anything meaningful.

4.4.3.2. *Traveling current structures.* Current systems in space often move quickly relative to the satellites. A boundary or surface, that is neither necessarily thinner than the spacecraft separation nor planar, can pass a cluster of spacecraft quickly

without significant changes in itself. Under this circumstance, the correlation analysis can provide very important information. Note here that more than one time shift (here one may use the measured timing difference directly instead of performing a correlation analysis because to perform the correlation analysis at each of the edges is usually very difficult) can be determined for a single passage of the current for each pair of spacecraft, at its leading and trailing edges. Averaging the velocities obtained from Equation (4.7) for all satellite pairs, one yields the velocity of the current structure. Here we recall that there is no assumption made on the shape or normal of the structure when using Equation (4.7) and hence it is applicable to structures of any regular shape. The geometry of the current structure can be examined according to the timing difference of each pair from the average and their relative locations.

With the derived velocity \mathbf{V}_0 , the current density at spacecraft i can be calculated by

$$\mathbf{J}_i = \frac{1}{\mu_0} \nabla \times \mathbf{B}_i = \frac{1}{\mu_0} \mathbf{V}_0^{-1} \times \frac{\delta \mathbf{B}_i}{\delta t_i}, \quad (4.13)$$

where prefix δ denotes the difference measured from the same spacecraft at different times, and $\mathbf{V}_0^{-1} = (1/V_x, 1/V_y, 1/V_z)$. Although this expression is meaningful in theory, in practice the fluctuations in the smallest component of the velocity and the magnetic fluctuations perpendicular to the smallest velocity component may cause a very large uncertainty in the calculation. One should keep caution in using it.

4.5. WAVE ANALYSES

In wave analysis, the time derivatives in Equation (4.6) are, in general, not zero in the spacecraft frame. The approach that sets $\tau = 0$ will eliminate all the terms of frequency and hence cannot be used to analyze waves. With a non-coplanar cluster of four spacecraft, in principle, one is able to determine the frequency, wavelength and phase velocity of a planar wave if its wavelength is of the order of spacecraft separation. For waves with a wavelength that is either much greater or smaller than the spacecraft separation, there is no general method to derive all these three wave parameters.

For perturbations that can be approximated as 1-D, but not necessarily monochromatic, one can follow the general steps.

(1) Select the perturbations of interest and band-pass filter the time series data. Note that if the wave is not monochromatic, the spectrum will have finite width.

(2) Perform wave analysis to determine the direction of the wave propagation for each spacecraft, and take the average of these vectors as the direction of the wave vector.

(3) Perform correlation analysis for each component of the filtered data using each pair of spacecraft.

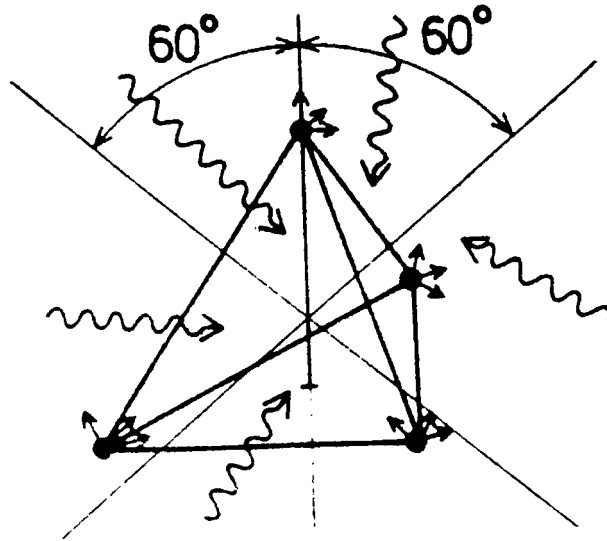


Figure 4.5. Many wave modes pass across a cluster of satellites (Neubauer and Glassmeier, 1990).

(4) Calculate the apparent velocity of the perturbation from the correlation analysis and compare its direction with the direction of the wave derived from the wave analysis.

(5) Analyze the deviation of each quantity from its corresponding average. In general, these deviations can be attributed to temporal variations that may not be directly related to the wave.

The biggest difficulty one may face here is that there is too much information and some of these pieces are in conflict. While one could use average and normalization to enhance the robustness of the results, we suggest one leave enough room for different possible interpretations because we think that at this moment we have very limited knowledge about the problems inherent in an analysis of such complexity.

Neubauer and Glassmeier (1990) proposed the concept of wave-telescope. As illustrated in Figure 4.5, when a wave propagates across a cluster of satellites, each satellite observes a different phase according to the wave propagation direction and the relative location among them. From these phase shifts, in theory, one is able to derive the wave properties. There are established techniques in radio science to analyze such information. While this proposal is intriguing, we note that in space wave analysis the signal-noise ratio is often low. Multiple modes in space plasma often exist, see Section 3.3 for more details, and are difficult to distinguish. Often when testing with a known wave, a scheme may work, but there may be multiple solutions when the input is unknown.

5. Discussion and Summary

Although it seems mathematically trivial, Equation (1.2) is one of the most important results in our review. It describes all physical effects on the measured magnetic field and provides a rigorous theoretical foundation for minimum variance analyses of the magnetic field. It shows that the minimum variance analyses are based on a one-dimensional assumption but do not require a steady state. If the minimum variance analyses also depended on the steady state assumption as many people perceived since it was first developed to analyze discontinuity, it would not have been applicable to wave analysis.

Another fundamental result is the derivation of Equation (4.6) which provides a unified theory for correlation analyses. Every existing correlation analysis technique can be found to be a special case of this equation. This expression specifies clearly the assumptions upon which each technique is based. In developing data analysis methods for Cluster missions, one needs to compare the methods against this expression in order to identify the effects that are neglected in the methods and justify for the simplification. We emphasize that at the practical level there has not yet been sufficient analysis of multiple satellites data to test and refine many of the techniques discussed herein.

On one hand, discontinuity analysis is, among time series data analysis methods, best understood and its computer programs are most user-friendly. It is widely used. However, on the other hand, the errors of the analysis have not been recognized as widely as it should. Occasionally, one may see in publication a discontinuity is presented in boundary normal coordinates with an unequal normal magnetic field across it.

Multiple parameter best fits are often used to determine unmeasured shock parameters. As we now know that this method in practice usually has a large uncertainty and may provide a large range of solutions with a small difference in input parameters. Here we again urge caution in using this method.

In wave analysis, a common trend is that fewer and fewer people actually analyze the time series data itself, and more and more people rely on dynamic spectrograms. Occasionally, one sees that a field rotation is interpreted as broadband emission that last for a period of the window length. Here we strongly urge analysts to examine the original time series data. On the other hand, there is a significant room for further improving the techniques of wave analysis.

In addition to mastering the various analysis methods, an analyst has to possess adequate knowledge of the data set he/she is using. From our experience, direct collaboration with an builder of the instrument is an effective way to ensure that the data is utilized correctly. However this is still no substitute for the proper documentation of the instruments and the data stream.

Acknowledgments

This work was supported by NASA under research grants NAGW–3948, NAGW–3974, NAG5–4066 and NAG5–3880 to UCLA, and by National Science Foundation/Office of Naval Research under research grant NSF–ATM 9713492. The authors are grateful to Bengt Sonnerup for useful discussions.

References

- Anderson, T. W.: 1978, *An Introduction to Multivariate Statistical Analysis*, John Wiley and Sons, New York.
- Anderson, B. J., Fuselier, S. A. and Murr, D.: 1991, 'Electromagnetic Ion Cyclotron Waves Observed in the Plasma Depletion Layer', *Geophys. Res. Letters*, **18**, 1955.
- Arthur, C. W., McPherron, R. L. and Means, J. D.: 1976, 'A Comparative Study of Three Techniques for Using the Spectral Matrix in Wave Analysis', *Radio Sci.* **11**, 833.
- Barnes, A. and Hollweg, J. V.: 1974, 'Large-Amplitude Hydromagnetic Waves', *J. Geophys. Res.* **79**, 2302.
- Born, M. and Wolf, E.: 1965, *Principles of Optics*, Pergamon Press Inc., Oxford.
- Chao, J. K.: 1970, *Interplanetary Collisionless Shock Waves*, Rep. CSR TR-70-3, Mass. Inst. of Technology, Center for Space Res., Cambridge, Mass.
- Chao, J. K.: 1995, 'Intermediate Shocks: Observations', *Adv. Space Res.* **15** (8/9), 521.
- Chao, J. K., Zhang, X. X. and Song, P.: 1995, 'Derivation of Temperature Anisotropy from Shock Jump Relations: Theory and Observations', *Geophys. Res. Letters* **22**, 2409.
- Colburn, D. S. and Sonett, C. P.: 1966, 'Discontinuities in the Solar Wind', *Space Sci. Rev.* **5**, 439.
- Crooker, N. U. and Siscoe, G. L.: 1979, 'A Mechanism for Pressure Anisotropy and Mirror Instability in the Dayside Magnetosheath', *J. Geophys. Res.* **82**, 185.
- Denton, R. E., Gary, S. P., Lin, X., Anderson, B. J., LaBelle, J. W. and Lessard, M.: 1995, 'Low-Frequency Fluctuations in the Magnetosheath Near the Magnetopause', *J. Geophys. Res.* **100**, 5665.
- Dunlop, M. W., Southwood, D. J., Glassmeier, K.-H. and Neubauer, F. M.: 1988, 'Analysis of Multipoint Magnetometer Data', *Adv. Space Res.* **8**, (9), 273.
- Elphic, R. C.: 1989, 'Multipoint Observations of the Magnetopause: Results from ISEE and AMPTE', *Adv. Space Res.* **8**, (9), 223.
- Frank, L. A., McPherron, R. L., DeCoster, R. J., Burek, B. G., Ackerson, K. L. and Russell, C. T.: 1981, 'Field-Aligned Currents in the Earth's Magnetotail', *J. Geophys. Res.* **86**, 687.
- Gary, S. P.: 1992, 'The Mirror and Ion Cyclotron Anisotropy Instabilities', *J. Geophys. Res.* **97**, 8519.
- Gary, S. P. and Winske, D.: 1992, 'Correlation Function Ratios and the Identification of Space Plasma Instabilities', *J. Geophys. Res.* **97**, 3013.
- Harvey, C. C., Etcheto, J., De Javel, Y., Manning, R. and Petit, M.: 1978, 'The ISEE Electron Density Experiment', *IEEE Trans. Geosci. Electron.* **GE-16**, 231.
- Hoppe, M. M., Russell, C. T., Frank, L. A., Eastman, T. E. and Greenstadt, E. W.: 1981, 'Upstream Hydromagnetic Waves and Their Association with Backstreaming Ion Populations: ISEE 1 and 3 Observations', *J. Geophys. Res.* **86**, 4471.
- Jenkins, G. M. and Watts, D. G.: 1968, *Spectral Analysis and Its Applications*, Holden-Day, London, p. 525.
- Kantrowitz, A. and Petschek, H. E.: 1966, in W. B. Kunkel (ed.), 'MHD Characteristics and Shock Waves', *Plasma Physics in Theory and Application*, McGraw-Hill Book Co., New York, p. 494.
- Kawano, H. and Higuchi, T.: 1995, 'The Bootstrap Method in Space Physics: Error Estimation for the Minimum Variance Analysis', *Geophys. Res. Letters* **22**, 307.

- Kepko, E., Khurana, K. K. and Kivelson, M. G.: 1966, 'Accurate Determination of Magnetic Field Gradients from Four Point Vector Measurements: 1. Use of Natural Constraints on Vector Data Obtained from Single Spinning Spacecraft', *IEEE Trans. Geosci. Electron.* **32**, 377.
- Khurana, K. K., Kepko, E., Kivelson, M. G. and Elphic, R. C.: 1996, 'Accurate Determination of Magnetic Field Gradients from Four Point Vector Measurements: 2. Use of Natural Constraints on Vector Data Obtained from Four Spinning Spacecraft', *IEEE Trans. Geosci. Electron.* **32**, 5193.
- Kivelson, M. G. and Russell, C. T.: 1995, *Introduction to Space Physics*, Cambridge University Press, p. 567.
- Kodera, K., Gendrin, R. and De Villedary, C.: 1977, 'Complex Representation of a Polarized Signal and Its Application to the Analysis of ULF Waves', *J. Geophys. Res.* **82**, 1245.
- Krauss-Varban, D. and Omid, N.: 1993, 'Propagation Characteristics of Wave Upstream and Downstream of Quasi-Parallel Shocks', *Geophys. Res. Letters*. **20**, 1007.
- Krauss-Varban, D., Omid, N. and Quest, K. B.: 1994, 'Mode Properties of Low-Frequency Waves: Kinetic Theory versus Hall-MHD', *J. Geophys. Res.* **99**, 5987.
- LaBelle, J. and Treumann, R. A.: 1988, 'Plasma Waves at the Dayside Magnetopause', *Space Sci. Rev.* **47**, 175.
- Le, G., Russell, C. T. and Smith, E. J.: 1989, 'Discrete Wave Packets Upstream from the Earth and Comets', *J. Geophys. Res.* **94**, 3755.
- Lee, L. C., Price, C. P. and C. S. Wu: 1988, 'A Study of Mirror Waves Generated Downstream of a Quasi-Perpendicular Shock', *J. Geophys. Res.* **93**, 274.
- Lepidi, S., Villante, U., Lazarus, A. J., Szabo, A. and Paularena, K.: 1996, 'Observations of Bow Shock Motion During Times of Variable Solar Wind Conditions', *J. Geophys. Res.* **101**, 11107.
- Lepping, R. P. and Behannon, K. W.: 1980, 'Magnetic Field Directional Discontinuities: 1. Minimum Variance Errors', *J. Geophys. Res.* **85**, 4695.
- Lin, Y. and Lee, L. C.: 1994, 'Structure of Reconnection Layers in the Magnetosphere', *Space Sci. Rev.* **65**, 59.
- McComas, D. J., Russell, C. T., Elphic, R. C. and Bame, S. J.: 1986, 'The Near-Earth Cross-Tail Current Sheet: Detailed ISEE 1 and 2 Case Studies', *J. Geophys. Res.* **91**, 4287.
- McPherron, R. L., Russell, C. T. and Coleman, P. J., Jr.: 1972, 'Fluctuating Magnetic Fields in the Magnetosphere: II ULF Waves', *Space Sci. Rev.* **13**, 411.
- Means, J. D.: 1972, 'Use the Three-Dimensional Covariance Matrix in Analyzing the Polarization Properties of Plane Waves', *J. Geophys. Res.* **77**, 5551.
- Mellott, M. M. and Greenstadt, E. W.: 1984, 'The Structure of the Oblique Subcritical Bow Shock: ISEE 1 and 2 Observations', *J. Geophys. Res.* **89**, 2151.
- Mottez, F. and Chanteur, G.: 1994, 'Surface Crossing by a Group of Satellites: A Theoretical Study', *J. Geophys. Res.* **99**, 13499.
- Mozer, F. S., Hones, E. W., Jr. and Birn, J.: 1983, 'Comparison of Spherical Double Probe Electric Field Measurements with Plasma Bulk Flows in Plasmas Having Densities Less than 1 cm^{-3} ', *Geophys. Res. Letters* **10**, 737.
- Mozer, F. S., Cattell, C. A., Temerin, M., Torbert, R. B., Von Glinski, S., Woldorff, M. and Wygant, J.: 1979, 'The dc and ac Electric Field, Plasma Density, Plasma Temperature, and Field-Aligned Current Experiments on the S3-3 Satellite', *J. Geophys. Res.* **84**, 5875.
- Neubauer, F. M. and Glassmeier, K.-H.: 1990, 'Use of an Array of Satellites as a Wave Telescope', *J. Geophys. Res.* **95**, 19115.
- Newbury, J. A., Russell, C. T. and Gedalin, M.: 1997, 'The Determination of Shock Ramp Thickness using the Noncoplanar Magnetic Field Component', *Geophys. Res. Letters* **24**, 1975.
- Omid, N. and Winske, D.: 1995, 'Structure of the Magnetopause Inferred from the Kinetic Riemann Problem', *J. Geophys. Res.* **100**, 11935.
- Petrinec, S. M. and Russell, C. T.: 1993, 'Intercalibration of Solar Wind Instruments During the International Magnetospheric Study', *J. Geophys. Res.* **98**, 18963.

- Rankin, D. and Kurtz, R.: 1970, 'Statistical Study of Micropulsation Polarizations', *J. Geophys. Res.* **75**, 5445.
- Russell, C. T.: 1988, BANAL: A Computer Program for the Display and Analysis of Magnetometer Data, April 1988.
- Russell, C. T.: 1983, 'Interactive Analysis of Magnetic Field Data', *Adv. Space Res.* **2** (7), 173.
- Russell, C. T. and Elphic, R. C.: 1978, 'Initial ISEE Magnetometer Results: Magnetopause Observations', *Space Sci. Rev.* **22**, 681.
- Russell, C. T., Childers, D. D. and Coleman, P. J., Jr.: 1971, 'OGO-5 Observations of Upstream Waves in the Interplanetary Medium: Discrete Wave Packets', *J. Geophys. Res.* **76**, 845.
- Russell, C. T., Mellott, M. M., Smith, E. J. and King, J. H.: 1983, 'Multiple Spacecraft Observations of Interplanetary Shocks: Four Spacecraft Determination of Shock Normals', *J. Geophys. Res.* **88**, 4739.
- Russell, C. T., Riedler, W., Schwingenschuh, K. and Yeroshenko, Y.: 1987, 'Mirror Instability in the Magnetosphere of Comet Halley', *Geophys. Res. Letters* **14**, 644.
- Russell, C. T., Priest, E. R. and Lee, L. C.: 1990, *Physics of Magnetic Flux Ropes*, Geophysical Monograph, Vol. 58, AGU, Washington.
- Scarf, F. L. and Russell, C. T.: 1988, 'Evidence of Lightning and Volcanic Activity on Venus', *Science* **240**, 222.
- Scokopke, N., Paschmann, G., Brinca, A. L., Carlson, C. W. and Luhr, H.: 1990, 'Ion Thermalization in Quasi-Perpendicular Shocks Involving Reflected Ions', *J. Geophys. Res.* **95**, 6337.
- Song, P., Elphic, R. C., Russell, C. T.: 1989, 'Multi-Spacecraft Observations of Magnetopause Surface Waves: ISEE 1 and 2 Determinations of Amplitude, Wavelength and Period', *Adv. Space Res.* **8**, (9), 245.
- Song, P., Russell, C. T. and Thomsen, M. F.: 1992, 'Waves in the Inner Magnetosheath: A Case Study', *Geophys. Res. Letters* **22**, 2191.
- Song, P., Russell, C. T., Strangeway, R. J., Wygant, J. R., Cattell, C. A., Fitzenreiter, R. J. and Anderson, R. R.: 1993, 'Wave Properties Near the Subsolar Magnetopause: Pc 3-4 Wave Energy Coupling for Northward IMF', *J. Geophys. Res.* **98**, 187.
- Song, P., Russell, C. T. and Gary, S. P.: 1994, 'Identification of Low Frequency Fluctuations in the Terrestrial Magnetosheath', *J. Geophys. Res.* **99**, 6011.
- Song, P., Zhang, X. X. and Paschmann, G.: 1997, 'Uncertainties in Plasma Measurements: Effects of Lower Cutoff Energy and Spacecraft Charge', *Planet. Space Sci.* **45**, 255.
- Song, P., Zhu, Z., Russell, C. T., Anderson, R. R., Gurnett, D. A., Ogilvie, K. W. and Strangeway, R. J.: 1998, 'Properties of the ELF Emissions in the Dayside Magnetopause', *J. Geophys. Res.* submitted.
- Sonnerup, B. U. Ö.: 1971, 'Magnetopause Structure During the Magnetic Storm of September 24, 1961', *J. Geophys. Res.* **76**, 6717.
- Sonnerup, B. U. Ö. and Cahill, L. J.: 1967, 'Magnetopause Structure and Attitude from Explorer 12 Observations', *J. Geophys. Res.* **72**, 171.
- Sonnerup, B. U. Ö., Papamastorakis, I., Paschmann, G. and Luhr, H.: 1987, 'Magnetopause Properties from AMPTE/IRM Observations of the Convection Electric Field: Method Development', *J. Geophys. Res.* **92**, 12137.
- Sonnerup, B. U. Ö., Papamastorakis, I., Paschmann, G. and Luhr, H.: 1990, 'The Magnetopause for Large Magnetic Shear: Analysis of Convection Electric Fields from AMPTE/IRM', *J. Geophys. Res.* **95**, 12137.
- Sonnerup, B. U. Ö., Paschmann, G. and Phan, T.-D.: 1995, in P. Song, B. U. Ö. Sonnerup and M. F. Thomsen, 'Fluid Aspects of Reconnection at the Magnetopause: *In situ* Observations', *Physics of the Magnetopause*, Geophysical Monograph, Vol. 90, AGU, Washington.
- Southwood, D. J.: 1968, 'The Hydromagnetic Instability of the Magnetospheric Boundary', *Planet. Space Sci.* **16**, 587.

- Southwood, D. J. and Kivelson, M. G.: 1993, 'Mirror Instability 1: The Physical Mechanism of Linear Instability', *J. Geophys. Res.* **98**, 9181.
- Strangeway, R. J.: 1991, 'Polarization of the Impulsive Signals Observed in the Nightside Ionosphere of Venus', *J. Geophys. Res.* **96**, 22741.
- Takahashi, K., Anderson, B. J. and Strangeway, R. J.: 1990, 'AMPTE CCE Observations of Pc 3-4 Pulsations at L=2-6', *J. Geophys. Res.* **95**, 17179.
- Walthour, D. W., Sonnerup, B. U. Ö., Paschmann, G., Luhr, H., Klumpp, D. and Potemra, T.: 1993, 'Remote Sensing of Two-Dimensional Magnetopause Structures', *J. Geophys. Res.* **98**, 1489.
- Walthour, D. W., Sonnerup, B. U. Ö., Elphic, R. C. and Russell, C. T.: 1994, 'Double Vision: Remote Sensing of a Flux Transfer Event with ISEE 1 and 2', *J. Geophys. Res.* **99**, 8555.
- Young, D. T., Perraut, S., Roux, A., de Villedary, C., Gendrin, R., Korth, A., Kremser, H. and Jones, D.: 1981, 'Wave-Particle Interaction Near Ω_{He^+} Observed on GEOS 1 and 2: 1. Propagation of Ion Cyclotron Waves in He⁺-Rich Plasma', *J. Geophys. Res.* **86**, 6755.