Chapter 2

Discrete time series analysis: An overview

We have asked how an analysis of magnetometer records enables us to characterize the cusp region. This characterization is greatly determined by the utility of any analysis technique brought to bear and our ability to interpret the results. In our exploration of the data, we found several discrete time series analysis techniques quite useful; this chapter is presented as an introduction to those techniques. It serves to familiarize the reader with the analytical methods and notational scheme used throughout this thesis. More than just a primer in signal processing, each section offers a physical motivation for the application of a particular technique or formalism to a given suite of data. Since the bulk of this thesis is concerned with the physical interpretation of ground-based magnetometer data, any such motivation stems from processes known to occur or thought to occur in the magnetosphere and ionosphere. While each technique is particularly suited to the analysis of the digital data stream derived from a magnetometer, the scope of its usefulness is not restricted to magnetometer data. The time series analysis techniques described herein are broadly applicable to many types of instrument data, geophysical and otherwise.

Since a comprehensive treatment of discrete time series analysis is beyond the scope of this thesis, some facility with statistical signal processing is assumed on the part of the reader. For a thorough introduction to the theoretical aspects of signal processing, several fine works are available and recommended [e.g., Jenkins and Watts, 1968; Bracewell,
1978; Press et al., 1992; Therrien, 1992]. What follows does, however, provide a sufficient overview for the space physicist interested in expanding his or her ability to gain physical insight through the analysis of instrument data. Our aim is to lay a foundation, one which enables the reader to follow our work. Moreover, considering our approach and examples provided in the ensuing chapters, the foundation should allow the reader to develop and build upon their own intuition, ultimately to pursue their own interests.

2.1 Traditional methods

Traditional approaches in physical signal processing are grounded in Fourier analysis. This is the case because the technique developed by J. Fourier nearly two centuries ago is widely applicable to both the description and analysis of physical systems whose nature is periodic. In addition, Fourier analysis is particularly useful in estimating the behavior of quasi- and non-periodic systems [Jenkins and Watts, 1968]. Both of these properties have direct bearing on the widespread use of Fourier theory in the analysis of geophysical data.

Consider the case of one of our ground-based fluxgate magnetometers. A typical example of the data sampled from such an instrument is shown in Figure 2.1. The data appear to be somewhat sinusoidal, or quasi-sinusoidal, in form; however, the dominant pulsations exist for only a portion of a full cycle. Quite clearly there are other components comprising the signal. Quasi-sinusoidal signals are characteristic of geomagnetic pulsations detected both at the ground and in situ. It is this characteristic which suggests the use of Fourier theory for their analysis.

Barring man-made perturbations, the currents we infer as the source of magnetometer signals are at least 90 km away in the overhead in the ionosphere. The information content of these current-carrying regions is conveyed to our instrument by the magnetic field as a wave phenomenon. Our fluxgate magnetometer is then essentially a current detector. In the magnetosphere-ionosphere system, wave interactions are described, to some approximation, by the equations of MHD theory.

Ultimately, it is the Maxwell equations which govern our understanding of how the magnetic field carries information from a distant current source to a local receiving station. In particular, in source-free regions we find generic wave solutions for the magnetic
Figure 2.1. A typical example of fluxgate magnetometer data. These data were taken from the fluxgate magnetometer at Longyearbyen on 9 January 1991. The data are zero-mean detrended to show deviations from the background geomagnetic field. Shown in the panels are (top) the x and (bottom) y components of the magnetic field. Note the quasi-sinusoidal nature of the data.

Field as $e^{i(kx - \omega t)}$, such that, given the appropriate boundary conditions, the following transformations hold:

$$\nabla \rightarrow ik$$  \hspace{1cm} (2.1)

$$\frac{\partial}{\partial t} \rightarrow -i\omega t$$  \hspace{1cm} (2.2)

where $k$ is the wave vector. Solution of partial differential equations in this manner is identical to Fourier transforming in space and time.

The problem of cusp latitude magnetometer data is amenable to Fourier analysis on two levels. First, the data strike us visually, with the dominant waveforms appearing quasi-sinusoidal. Upon closer inspection, we find a rich variety of behaviors, from intervals of superposed sinusoids of differing frequency to intervals that are apparently completely stochastic. Second, our mathematical intuition concerning the source of the data lends itself to Fourier analysis. The physical, electrodynamical source of our data is fairly well described by a group of partial differential equations, the solution of which can be
had via Fourier transformation.

Both our visual and mathematical perception of the magnetometer data tend to suggest Fourier analysis for an increased understanding. Along with its many uses, Fourier theory does have its limitations; chiefly, its description of non-periodic data is only approximate at best. For that very reason Fourier theory is certainly not the last word in the analysis of geomagnetic signals; however, it is a natural starting point. For us, in presenting the thesis at hand, it is a point of departure.

2.1.1 The Fourier transform

Discrete signals, or time series, can be represented by a vector in bra-ket notation as $|x[t]|$. This denotes a quantity $x$ that is a function of a discrete variable $t$, or time. Each component of the vector is represented by $x[k]$. When the data are sampled at a regular frequency $f_s$, $t$ and $k$ are related by $t = k/f_s$. All Fourier theory applied to discrete signals implies the use of the discrete Fourier transform (DFT). In that vein, the DFT lies at the heart of virtually all the signal processing carried out in this thesis. Since it requires little by way of introduction, we merely define our convention here and offer a brief comment on numerical particulars.

Given a time series $|x[t]|$ of $T$ points, we define the forward and inverse transforms, respectively, as

$$X[q] = \frac{1}{T} \sum_{k=0}^{T-1} x[k] e^{-i2\pi kq/T} \quad (2.3a)$$

$$x[k] = \sum_{q=0}^{T-1} X[q] e^{i2\pi kq/T} \quad (2.3b)$$

Here $f$ and $q$ are related by $f = qf_s/T$. These particular forms (2.3) are a modified and concatenated version of those given by Jenkins and Watts [1968] and Press et al. [1992]. Since the forward (2.3a) and inverse (2.3b) designation can be taken as a matter of convenience, i.e. $q$ and $k$ are interchangeable in (2.3), we select our convention noting that our forward transform is consistent with a causal signal, and that the first Fourier component $X[0]$ is the mean of the times series $|x[t]|$. The latter point is important in considering transient

---

\^1Note that the discrete indices of time $k$ and frequency $q$ are both integer-valued.
phenomenon, in which the mean, or dc component of the signal, is uninteresting and can be made to be zero.

Throughout this thesis, the following conventions hold: the time series data are real and \( T \) is even. While these specifications appear somewhat arbitrary, they are motivated by the fact that our magnetometer data are real, and it is computationally advantageous to consider an even number of samples.\(^2\)

With real time series and \( T \) even, it is of numerical interest to note that \( X[q] = X[T - q]^* \) for \( q \in [1, T/2 - 1] \). Thus only \( T/2 + 1 \) of the \( T \) frequency estimates are independent; these are the so-called positive frequencies. Two of the Fourier components (2.3a) will be real: \( X[0] \) and \( X[T/2] \). The former is simply the mean of \( |x[t]| \); the latter the Nyquist frequency component, or the highest frequency component of the time series which we may resolve. The Nyquist frequency is given by

\[
f_N = \frac{f_s}{2}
\]

where \( f_s \) is the sampling rate.

### 2.1.2 Spectrum estimation

Magnetometer output is sampled in time to provide a time series, here in the form of \( |x[t]| \). This sampling provides us with one particular, and perhaps the most common, description of data, that in the time domain. Application of the forward Fourier transform (2.3a) takes us from the time domain to \( |X[f]| \) in the spectral, or frequency domain.\(^3\) Each domain is suited for certain types of analysis; however, the frequency domain often enables a simpler

---

\(^2\) Some authors, notably Press et al. [1992], would argue that our restriction on \( T \) is too lax. Practically all computation of the DFT is carried out by a fast Fourier transform (FFT) algorithm based on those popularized by J. Cooley and J. Tukey in the mid-1960’s [Press et al., 1992]. The FFT is extremely efficient in that it requires only \( O(T \log_2 T) \) operations to compute the Fourier transform, whereas a brute force calculation of the sum (2.3) requires \( O(T^2) \) operations [Brigham, 1988]. The FFT is known as a radix-2 transform, and as such places a more stringent requirement on the number of samples; namely, \( T = 2^m \) with \( m \in [0, \infty) \). Owing to the great computational efficiency of the FFT, the use of anything other than power-of-two samples is often proscribed; such a requirement is no less arbitrary than our specification that number of samples be even.

While the radix-2 algorithm is certainly efficient, not all practical data lend themselves to division into power-of-two segments. Moreover, sophisticated algorithms exist that are computationally competitive with the FFT. These algorithms are known as mixed-radix transforms, and take advantage of the prime factorization of \( T \) to perform the calculation. For \( T \) prime, of course, the only available option is direct calculation via (2.3). All Fourier transforms in this thesis were computed by the mixed-radix routine of Moler [1992].

\(^3\) From this point forward, we use \( |x[t]| \) and \( |x| \) interchangeably; likewise with \( |X[f]| \) and \( |X| \).
or reduced description of the data [Samson, 1983b]. Geomagnetic data, in particular, lends itself to spectral analysis.

Under the aegis of Fourier analysis, we find the most direct route into the spectral domain, particularly via computation of (2.3a); it is not, however, our only means of entry [Press et al., 1992]. Other, non-Fourier, methods of spectral estimation have been developed, notably the maximum entropy (ME) method of Burg [1975]. The ME method assumes an autoregressive (AR) model of a certain order for the data, thus the spectra produced are model dependent [Therrien, 1992].

In choosing among various methods, we must ask what assumptions underlie our understanding of the physics governing the processes thought to give rise to the signals we receive at the magnetometer station. Too, we must recognize the limitations of each method — no one spectral technique is best for all data or situations. The method and assumptions used to estimate the spectrum have a marked influence on the result [Lacoss, 1971]. In this work we champion the use of Fourier methods to estimate the spectrum. Ostensibly, the data are quasi-periodic in nature, and the equations that govern our understanding of the processes involved may be solved, in principle, by Fourier methods.

A tempting advantage of the ME method is that it can provide arbitrary frequency resolution, as we need only consider an appropriately high order model of the data. Closely spaced spectral peaks that might otherwise evade Fourier methods may be resolvable in this manner. Unfortunately, the magnetometer signals we receive are not thought to be the result of an AR process — in some cases, we are unsure of their exact source mechanism. The ME method is well-suited for the detection and analysis of signals whose source or form is known a priori [Lacoss, 1971]. In the case of geophysical data, justification for the model order is marginal at best [Bracalari and Salusti, 1994], causing us to fall back on classical spectral analysis (J. C. Samson, private communication, 1994).

For a data sequence $|x⟩$ of length $T$, the most simple estimate of the spectrum is given by the magnitude squared Fourier transform as

$$S[f] = \frac{1}{T}|X[f]|^2 \quad (2.5)$$

We refer to this estimate as a raw spectral estimate and it is consistent with the Einstein-Weiner-Khintchine relation [Therrien, 1992]. This estimate is asymptotically unbiased, in
the sense that

$$S[f] \xrightarrow{T \to \infty} \hat{S}(f)$$

(2.6)

where $\hat{S}(f)$ is the true spectrum (note that it is continuous) [Jenkins and Watts, 1968]. For $T$ sufficiently large, we are certain that there is relatively small bias in our estimate.

Unfortunately, the variance in $S[f]$ is rather large

$$\sigma_{S[f]}^2 \approx S^2[f]$$

(2.7)

and does not decrease as $T \to \infty$ [Therrien, 1992]. Finite $T$, which is the case in all practical situations, gives rise to a phenomenon known as leakage [Jenkins and Watts, 1968; Press et al., 1992], in which spectral information from frequency $f$ will bleed into adjacent frequency components from dc to the Nyquist frequency. Such an effect arises because for finite $T$ we are equivalently considering only a windowed portion of an infinite amount of data when we take the Fourier transform [Press et al., 1992].

**Confidence**

We conclude then that the raw estimate (2.5) offers only a very crude approximation of the true spectrum $\hat{S}(f)$. Hence, we turn to various statistical means designed to improve our confidence in the spectral estimate. In this effort we have two principle aims: 1) reduce the bias induced via leakage and 2) reduce the variance in the estimate. Windowing is typically employed to reduce the bias and averaging or smoothing is used to reduce the variance in the spectral estimator [Press et al., 1992; Therrien, 1992]; however, it should be noted that both processes involve some mutual interaction [Chave et al., 1987].

It may seem confusing that windowing is used to reduce bias in the estimator when it is the window, owing to a finite amount of data, that causes leakage in the first place. Simply put, for finite $T$ some sort of window is unavoidable. The windowing operation may be represented as

$$x_w[k] = w[k]x[k]$$

(2.8)

In the simplest case, a rectangular window, $w[k] = 1 \ \forall \ k$. In practical situations, this window introduces the greatest amount of leakage, therefore we choose to construct a window that minimizes this undesirable effect. In this thesis we make use of the following
either of which can be shown to substantially reduce the bias caused by leakage [Press et al., 1992]. In our experience, and that of others [e.g., Jenkins and Watts, 1968; Press et al., 1992], the exact choice of window — there are many — has little effect upon our results.

With regard to variance, it is helpful to consider that spectral estimates are distributed approximately as a chi-square with \( \nu \) degrees of freedom [Jenkins and Watts, 1968]. In the case of the raw estimate (2.5), \( \nu = 2 \). Any step we take to increase the number of degrees of freedom will result in a decrease of the estimate’s variance. As with windows, there are numerous schemes for variance reduction; from among them, we employ the method of Welch [1970].

The Welch method takes the original time series \(|x\rangle\) and breaks it into \( K \), possibly overlapping, subseries \(|x_k\rangle\), each of length \( T/K \). Fourier transforms of the subseries \(|X_k\rangle\) are taken and averaged together, element by element, to form a spectral estimate \(|\bar{x}\rangle\) with elements

\[
\bar{x}[q'] = \sum_{k=0}^{K-1} x_k[q']
\]

Note that there is a cost associated with this procedure. Whereas in the raw estimate we found \( q \in [0, T/2] \) positive frequencies, the Welch method reduces the count of positive frequencies to \( q' \in [0, T/2K] \).

We are more than compensated for the loss of spectral resolution. For non-overlapping subseries, we increase the degrees of freedom to \( \nu = 2K \). Alternately, we reduce the variance in this manner by a factor of \( 1/K \) [Therrien, 1992]. We can further reduce the variance by overlapping the subseries, but only to a point. As we increase \( K \) by overlapping, the subseries become correlated, thus offsetting the reduction in variance [Therrien, 1992; Krause et al., 1994]. For subwindows of length \( T/K \) an overlap of \( T/2K \) points can be shown to be optimal (D. Prichard, private communication, 1994). For this level of overlap, \( \nu \approx 2K \).
Spectral matrices

Univariate signal processing is a luxury seldom afforded the space physicist. In our case, magnetometer records represent a sampling of the geomagnetic field, a vector quantity. As such, we derive at least a 2-dimensional time series from each station. If we consider an array of stations, the need to analyze \( N \)-variate (or \( N \)-dimensional) time series arises. In a series of papers, Samson and Olson showed that the spectral matrix offers a convenient representation of the spectrum of \( N \)-dimensional geophysical time series [Olson and Samson, 1979; Samson and Olson, 1980, 1981; Samson, 1983b].

Before we define the spectral matrix, we introduce the cross spectrum estimate of two time series

\[
S_{ij}[f] = X_i[f]X_j^*[f] \tag{2.12}
\]

Note that \( S_{ii}[f] \) is simply the raw spectrum estimate (2.5); hence, all of our discussion concerning confidence applies [Jenkins and Watts, 1968]. Now, in terms of the cross spectrum estimate, the spectral matrix is defined as

\[
S[f] = \begin{pmatrix}
  S_{00}[f] & \ldots & S_{0,N-1}[f] \\
  \vdots & \ddots & \vdots \\
  S_{N-1,0}[f] & \ldots & S_{N-1,N-1}[f]
\end{pmatrix} \tag{2.13}
\]

This form is equivalent to that of Samson and Olson [1980] and is particularly useful, since the Welch method may be readily employed to calculate it. For \( N \) time series of length \( T \), using length \( T/K \) subspectra in the Welch method, we arrive at \( T/2K + 1 \) spectral matrices of dimension \([N, N]\). One can visualize the result as a stack of matrices, forming a rank three tensor of dimension \([N, N, T/2K + 1]\). Noting that \( S_{ij}[f] = S_{ji}^*[f] \), we see that \( S[f] \) is Hermitian and thus contains only \( N(N+1)/2 \) non-redundant elements.\(^4\)

2.1.3 Frequency-time localization

At any given magnetometer station, we record data from one of the field components for a length of time, resulting in a time series \( x \) of length \( T \). Since \( T \) is finite, we can resolve

\(^4\)It is computationally convenient then to perform some index gymnastics and map the spectral matrices into a reduced vector space as \( S_{ij} \rightarrow s[i + \sum_{\ell=0}^{\ell} \ell(N-i)] \) for each \( f \), where \( j \in [0,N-1] \) and \( i \in [j,N-1] \). This allows us to effectively operate on the rank three tensor in machine memory by handling a simple matrix of dimension \([N(N+1)/2, T/2K + 1]\).
a maximum of $T/2 + 1$ useful spectral estimates. This number will be reduced further via
the Welch method in order to increase our confidence in the spectral estimates. We can in-
crease our resolution by making our observing run appropriately long.\footnote{We can also gain resolution through a process known as zero-padding, in which we append an appropriate number of zeros to $|x\rangle$ \cite{Therrien}. We must not think that by zero-padding we are gaining any information. The process does not add any spectral information to the system, nor does it change the shape of the spectrum; it merely interpolates the original spectral content among the newly opened frequencies. The process is useful however, when computational exigency requires sheer speed, as $T$ can be made to be a power-of-two in this fashion.} Long observation
periods are not an inconvenience, as most magnetic observatories are automated; however,
the processes that give rise to the magnetometer signals are not, in general, long-term.

Fourier spectral estimates are made with the underlying assumption that the data are stationary during the period of observation \cite{Therrien}. Magnetometer data are only approximately stationary. Reliable spectra can be estimated if the first two moments of
the data, the mean and variance, are roughly constant over long time periods, even if
there are local excursions from these baseline values \cite{Chave}. An approximately stationary signal exhibiting this type of behavior — sporadic, short-term episodes of non-
stationarity — is termed wide-sense stationary \cite{Akansu}. Figure 2.2 provides an example of the wide-sense stationary nature of our magnetometer data. Since the data depart locally from stationarity, we are interested in their spectrotemporal characteristics. That is, we wish to know two things: 1) the spectral information carried by the pulsations, and 2) the temporal evolution, or transient response, of that information. Ideally, for a time series $|x\rangle$ we would like to have complete knowledge of $|X\rangle$ for each sample $x[t]$.

An uncertainty principle prevents us from obtaining complete, instantaneous knowl-
dge of the spectrotemporal characteristics of any sort of data, even in principle. One can show that the Fourier transform obeys the following form of the uncertainty principle \cite{Bracewell}

$$\Delta f \Delta t \geq \frac{1}{4\pi}$$

(2.14)

where $f$ and $t$ are continuous. For the discrete time case, the uncertainty becomes a three-
parameter relation \cite{Claerbout}

$$\Delta f \Delta t \zeta^2 \geq 1$$

(2.15)
Figure 2.2. Magnetometer data mean and variance as functions of time. These data were derived from the fluxgate magnetometer at Longyearbyen on 9 January 1991 (a portion of which is shown in Figure 2.1). Each moment is calculated at one-hour intervals and shown for each component of the magnetic field. Since the mean and variance of the data are only approximately constant, we refer to them as wide-sense stationary. The marked excursion from baseline values near 0500 UT is of interest to us, as it presages the passing of magnetic local noon, and hence the cusp region.
where the parameter $\zeta^2_\nu$ is a function of the degrees of freedom $\nu$ in the spectral estimate, and $f$ and $t$ are discrete. The effects of windowing, i.e., leakage, in effect cause us to be completely uncertain about our spectral estimates — spectral information for any frequency $f$ is spread among all $T/2 + 1$ components. At best, we may only have a certain degree of statistical confidence in them. For practical purposes then, we are interested in what sort of frequency resolution $f_r$ we may obtain from a time series. This information is governed by a simple relation

$$\frac{f_r}{f_s} = \frac{1}{T} \quad \text{(2.16)}$$

where $f_s$ is the sampling rate.

Given the relations for uncertainty (2.15) and frequency-time resolution (2.16), the question becomes: How do we balance the gain/loss of time resolution against that of frequency resolution? Our answer lies in one of two signal processing techniques, frequency-time distributions or dynamic spectra. The former include applications of the family of Wigner-Ville distributions [Najmi, 1994] and the generalized Gabor transform [Stockwell et al., 1996]. Much like the ME spectral method, these applications are best suited to problems in which the source and form of the time series is known a priori [Gaunaurd and Strifors, 1996]. It is the latter technique that proves the arbiter of our compromise in frequency-time resolution. A dynamic spectra is also referred to as a short-time Fourier transform [Therrien, 1992].

Consider a process which we wish to analyze, one that is assumed wide-sense stationary over a number of samples $T_s$. We take a time series $|x\rangle$ of length $T > T_s$ and subdivide it into a number of possibly overlapping subseries $|x_i\rangle$, of length $T_i$, where $T > T_i > T_s$. At this point we treat each subseries in isolation, estimating the spectrum via the Welch method, exactly as we did before. In doing so, we are looking for Welch subseries $|x_k\rangle$ of length $T/K \approx T_s$. In practice, $T_s$ is not known, but can perhaps be deduced or estimated after some fashion.

The result of estimating the spectrum over the set $\{|x_i\rangle\}$ is that we now have a conjugate set of spectral estimates $|X_i\rangle$, roughly spanning the original series $|x\rangle$ and its Fourier transform $|X\rangle$ in time and frequency. Graphically, we display this spectrotemporal information as an array, typically called a frequency-time spectrogram. See Figure 2.3 for a typical example of a spectrogram of magnetometer data.
2.1.4 Generalized coherence estimation: Coherence and polarization

When we examine N-dimensional time series, it is often necessary to characterize the signals in two or more of the data channels in terms of their likeness. In the time domain this operation is typically carried out as a cross-correlation [Therrien, 1992]. In the spectral domain, the chief measure of likeness is the coherence between signals. We employ two independent measures of coherence for N-dimensional magnetometer signals. The first is valid only for \( N = 2 \); however, this is often the case when we are considering interstation coherence. The second, a polarization estimator, is appropriate for \( N \geq 2 \) and can be shown to be an estimate of the generalized coherence [Samson, 1983a]. The N-dimensional polarization estimate may further be used to construct data-adaptive pure-state filters that are particularly useful in analyzing wave phenomenon, such as the signals we detect with
our magnetometers [Samson and Olson, 1981].

The first coherence estimator is the traditional magnitude squared coherence function $C_{ij}[f]$, given by Therrien [1992]. It is defined at each frequency $f$ as the ratio of the non-redundant elements of the 2-dimensional spectral matrix

$$C_{ij}[f] = \frac{|S_{ij}[f]|^2}{S_{ii}[f]|S_{jj}[f]}$$

The second is the degree of polarization estimator derived by Samson and Olson [1980]. This estimator is based upon the invariants of the spectral matrix and is given by

$$P[f] = \frac{N[\text{tr} S^2[f]] - [\text{tr} S[f]]^2}{(N - 1)[\text{tr} S[f]]^2}$$

where $N$ is the dimension of the system used to compute the estimate and $\text{tr} S$ is the trace of the spectral matrix. It is important to note that for 2-dimensional data, the polarization estimator (2.18) can be shown to be identical to the classical form given by Fowler et al. [1967].

By definition $C_{ij}, P \in [0, 1]$, and both are biased estimators for finite amounts of data with a small number of degrees of freedom; indeed, for a naive estimate of the spectrum, both are identically 1. When the Welch method of spectrum estimation is used, the magnitude of each coherence estimator descends toward a limiting value, that of the true coherence. In this sense, both $C_{ij}[f]$ and $P[f]$ represent an upper bound to actual coherence present in the signals at each frequency. A central feature of each estimator is that they both tend toward a small value, or noise floor, for white noise processes when $\nu$ is sufficiently large, but finite. This property is a marked advantage in statistical coherence studies where quantization noise in magnetometers may dominate the signal and thus skew the results. Jackson [1996] has shown that such noise may be assumed to be approximately white.

We use both coherence estimators in concert, providing for a more accurate assessment of the true coherence; their advantages and limitations tend to complement one another. Since $P[f]$ is computed from the invariants of the spectral matrix it is independent of sensor orientation; however, a pronounced signal in any one of the $N$ data streams will cause $P[f] \to 1$. In contrast, while $C_{ij}[f]$ does not benefit from the rotational invariance of $P[f]$, it is not biased toward unity for loud, single-channel signals. Thus, the use of both estimators
as a check ensures that we are determining the true coherence present in a set of multi-dimensional signals.

2.2 New directions

The history of geomagnetic signal processing is steeped in Fourier theory. There are sound physical and mathematical reasons underpinning its widespread use in describing magnetometer data. While we are mindful of that success, our search for an improved understanding of the data must not be tradition-bound. There exist in the literature other methods of analysis which have proven themselves, if not for as long, in fields where sampled instrument data share many of the same properties as our magnetometer data.

The generalized coherence measure of Samson and Olson [1980] (2.18) serves as a transition between traditional approaches and new directions. In an age where multi-instrument arrays and collaborative data collection efforts have all but replaced single station measurements, the development of such estimators is important. The polarization estimator $P[f]$ extends classical theory to an $N$-dimensional instrument space that is rapidly characterizing data suites. Still, adoption of the estimator faced resistance after it was introduced to the magnetic pulsation community. Although rigorously developed nearly two decades ago, only recently has its use become widely accepted [e.g., Olson and Fraser, 1994; Waters et al., 1995; Ziesolleck et al., 1996].

In this section we turn from classical Fourier theory to examine the potential application of two transform techniques to the data. The Hilbert and Karhunen-Loève transforms are well-known in electrical engineering and information theory circles, but within the geomagnetic community their use is somewhat novel. Rather than completely abandon the Fourier approach in their favor, we find application of these techniques in support of our classical analysis. And although neither transform was developed from Fourier theory, each may be understood in terms of it.

2.2.1 The Hilbert transform

There are several probable sources of magnetic pulsations likely to produce monochromatic waves, particularly the high-latitude field line resonances [Walker et al., 1992]. Un-
der ideal conditions this assumption may hold, but in a realistic situation, where competing processes are likely to occur, we expect that these pulsations will be detected on the ground as quasi-monochromatic signals at best. The phase properties of such a quasi-monochromatic signal would be of interest to us, since a disruption in phase signature of the signal may indicate a change in the character of the source region [Huang et al., 1992].

The uncertainty relation (2.15) prevents us from examining the spectral characteristics of our data for an arbitrarily short time period. Any information in the frequency domain represents an average over some number of samples in the time domain. For quasi-monochromatic signals however, a technique exists that allows us to examine the phase information of our signal at each point in the time series. This technique is known as the Hilbert transform and allows us to construct a phase-descriptive analytic signal from our original time series [Bracewell, 1978].

Although the Hilbert transform is valid, in general, for any time series [Jackson, 1996], we have found that our confidence in the phase-descriptive properties of the analytic signal is poor when the original time series differs greatly from a monochromatic signal. For this reason, we apply the technique only to processes which either are thought to be, appear to be, or are made to be band-limited.

The Hilbert transform stems from continuous transform theory via the Cauchy integral formula [Morse and Feshbach, 1953]

\[
\psi_H(z) = \frac{1}{\pi} \mathcal{P} \int_{-\infty}^{\infty} \frac{\psi(z_0)}{z_0 - z} \, dz_0
\]

(2.19)

where \(\mathcal{P}\) means take the Cauchy principal value of the integral and \(z\) is complex. In the discrete time case, Hilbert transformation of a time series

\[
|x_H\rangle = \mathcal{H}\{|x\rangle\}
\]

(2.20)

results in a new time series with all of the original frequency components shifted in phase by \(\pm \pi/2\), often called the quadrature of \(|x\rangle\). We use this phase shifting property to define the analytic signal [Bracewell, 1978]

\[
|\tilde{x}\rangle = (1 + i\mathcal{H})|x\rangle
\]

(2.21)

Note that \(|\tilde{x}\rangle\) is in general a complex time series, even for \(|x\rangle\) real.
From the analytic signal, we can determine several useful parameters. The instantaneous amplitude of the original signal is given by the magnitude of the analytic signal $|\tilde{x}[t]|$. More importantly, the instantaneous phase is given by

$$\phi[t] = \tan^{-1}\left(\frac{x_H[t]}{\tilde{x}[t]}\right) = \tan^{-1}\left(\frac{\text{Im} \tilde{x}[t]}{\text{Re} \tilde{x}[t]}\right)$$

(2.22)

The instantaneous frequency may also be estimated via a numerical differencing scheme, $|\delta\phi| \approx \frac{d\phi}{dt}$, but it should only be considered as a local average over several adjacent samples [Krause et al., 1994]. The instantaneous frequency estimate is useful for detecting phase skips in time series.

In practice, we need not calculate $|x_H\rangle$ as an intermediate step in constructing the analytic signal (2.21); rather, Fourier theory provides us with a direct route. Given our usual time series of length $T$, the analytic signal can be calculated as

$$|\tilde{x}\rangle = \mathcal{F}^{-1}\{|X\rangle \otimes |h\rangle\}$$

(2.23)

where $\mathcal{F}^{-1}$ represents the inverse Fourier transform (2.3b) and $\otimes$ is the array (element-by-element) multiplication operator. The vector $|h\rangle$ is akin to the Heaviside step function with elements given by [Krause et al., 1994]

$$h[q] = \begin{cases} 
1 & q = 0, T/2 \\
2 & 0 < q < T/2 \\
0 & T/2 < q \leq T
\end{cases}$$

(2.24)

The weights in $|h\rangle$ are unequal in order that $|x\rangle = \text{Re} |\tilde{x}\rangle$. The analytical signal is then essentially the inverse Fourier transform of only the positive frequencies.

2.2.2 The Karhunen-Loéve transform

Much as with the Hilbert transform, the Karhunen-Loéve transform was initially developed for use with continuous variables [Karhunen, 1947]; however, its useful properties were eventually recognized for the discrete time case and several linear algebraic applications were found (see the review given by Pike et al. [1984]). We make use of the discrete
Karhunen-Loéve transform (DKLT), which is known variously as singular value decomposition, principle orthogonal decomposition, or principle component analysis; it has also been shown to be equivalent to factor analytic methods [Mees et al., 1987].

Use of the DKLT is most common in information theoretic and dynamical systems analyses [Mees et al., 1987]. It has an optimal property in terms of data compression, so it also finds application in image analysis and pattern recognition [Fu, 1968; Akansu and Haddad, 1992]. Determination of the properties of the DKLT is based in matrix theory, so there are strong parallels to the quantum mechanical operator formalism [Gasiorowicz, 1974]. In the geomagnetic community, only Samson [1983a] has examined the concept, from a factor analytic standpoint. His efforts were aimed at optimizing the polarization estimator (2.18) in the presence of noise; eventually, he abandoned the idea in favor of other pursuits (J. C. Samson, private communication, 1995).

The DKLT is similar to the DFT (2.3b) in that we are seeking to expand the original time series \(|x\rangle\) on an orthonormal basis set. The primary difference between the two lies in the fact that the Karhunen-Loéve basis cannot be known \textit{a priori}, it must be determined directly from the data themselves. Additionally, the coefficients of the Karhunen-Loéve expansion form an orthogonal set. The expansion is given by

\[
|x_m\rangle = \sum_{b=0}^{M-1} v_b[m] |a_b\rangle \tag{2.25}
\]

where \(|x_m\rangle\) is one of a set of \(M\) time series, each of length \(T\), \(|a_b\rangle\) is an orthonormal basis vector, and \(v_b[m]\) is an expansion coefficient chosen from among a set of orthogonal vectors. As \(T \to \infty\) the basis functions of the DFT for a general random process are qualitatively similar to those of the DKLT; for periodic random processes, the basis functions of the DFT are identical to those of the DKLT [Therrien, 1992]. However, for finite \(T\), the DKLT provides the optimal description of the data in the mean square sense. For a formal development of the DKLT, see Appendix A.

In terms of transform efficiency, the DKLT is unique in that it is the only transform which perfectly decorrelates the expansion coefficients and optimizes the storage of signal information (energy) among the basis functions [Akansu and Haddad, 1992]. This property has profound implications for pattern matching and feature selection problems in which the exact form of the input is unknown [Fu, 1968]. In the presence of contaminating noise,
or unwanted signal, the DKLT can be used as for effective noise reduction [Broomhead and King, 1986]. Since the coefficients \( v_b[m] \) scale as the information content of the signal, those that fall below some nominal value may be considered insignificant. We caution that this form of noise reduction carries with it the tacit assumption that signals comprise the bulk of the information content in a given data stream.

These properties of the DKLT are significant, considering the central problem we face in analyzing magnetometer data. The spectra we derive are largely the result of an unspecified process. While we may have a basic understanding of the cause of a portion of the spectrum, a large part of the information stems from processes unrelated to those we understand. As a result, we do not have a general mathematical description of the spectra, nor of the noise which appears with it.

In this work we have found two applications of Karhunen-Loève theory in the field of geomagnetic pulsations. The first application takes advantage of its utility in pattern recognition processes, particularly the analysis of a large number of dynamic spectra. In the past, such examinations were conducted merely by eye [e.g., McHarg and Olson, 1992; McHarg, 1993; Engebretson et al., 1995]; we attempt herein to place such analyses on more rigorous ground. The second involves using the noise reduction characteristics of the DKLT, in order to determine a possible characteristic ground-based magnetic signature of the magnetospheric cusp in the time domain. In both cases, we find that the technique provides new and useful physical insights.

### 2.3 Geophysical signal processing

The present work relies heavily on the techniques of discrete time series analysis to gain a physical insight into the processes which characterize the high-latitude coupling of the magnetosphere and ionosphere. The use of such computationally intensive tools can lead us to equate deeper physical understanding with increased processing speed and algorithm efficiency. Nearly three decades ago we were cautioned that such is not the case; what we need is not more computational power, but better interpretive skills.

In [our] experience, the fast computers which are now available are more than adequate for purposes of spectral analysis. Our present computing facilities
are greatly in excess of our ability to make sense of practical data [Jenkins and Watts, 1968, p. 314].

Facility in the art of geophysical signal processing is tempered by two meta-principles. First, it is necessary to have an intimate knowledge of the characteristics of any technique brought to bear on the data. Each technique carries with it a set of advantages and limitations. The best way to develop a sense for these properties is to simply play with the techniques, gaining experience by applying them to both actual and contrived data sets. In this manner, one begins to gain an appreciation for what a particular method can extract from the data. Second, we must have the ability to separate physical results from non-physical ones.

### 2.3.1 Assumptions and results

When we consider our results, we are naturally mindful of the assumptions which we make when we model the physics of the situation. When we rely on signal processing to interpret the data, we must take a further step and become aware of how our signal processing assumptions constrain our results. In the application of a particular technique to data, we must be mindful that results peculiar to that technique can have a radical effect upon our interpretation.

Spectrum estimation provides an excellent example of how assumptions can influence an outcome. Lacoss [1971] showed that the shape of the spectrum varies greatly with the technique used to estimate it. He based his analysis on a known correlation function, so that he could determine the actual spectrum analytically. Employing various methods, he found that the ME-derived spectrum approximated the actual spectrum much more closely than the classical Fourier approach.

He cautions us however, that since the ME spectrum is model dependent, its application is most appropriate in situations where either the spectrum or the model order of the process responsible for the data are known a priori; in our case we have no such knowledge. We have found through numerical experimentation that ME spectra exhibit spurious spectral peaks. The number of ME spectral peaks varies roughly as half the model order. This causes us to find that the assumptions underlying the ME method are untenable, given
what we know of the sources of our magnetometer pulsations.

At some point an assumption must be made concerning noise, or unwanted signal. We use the two terms interchangeably, as any signal not owing to the process we wish to study can be considered noise. In our case, where we have only incomplete knowledge of the input source, this assumption can lead us to discard information about the source as noise, or to include spurious information. Certainly this will have an effect on our results and interpretation. To mitigate this problem, we typically take a statistical approach to the data, so that we can average out the effects of noise and, hopefully, coax the signal to the fore.

Further, since we have incomplete knowledge of the noise in our magnetometer signals — whether it be unwanted signal from uninteresting magnetospheric or ionospheric processes, or instrument noise — we must make assumptions concerning its form. Typically, we assume white noise contamination in the signal, which can be shown to have a spectral matrix proportional to the identity matrix. We are cognizant that this assumption can break down at times, especially in the quantization noise from the instruments at extremely low signal levels, where the bit error level is of order the signal strength [Jackson, 1996]. This source of potentially correlated noise is a problem at high frequencies in fluxgate magnetometers, which have a pronounced rolloff in the Pc 3 frequency range [Primdahl, 1979].

### 2.3.2 Artifacts of processing

Fourier-based estimates taken with a small number of degrees of freedom can lead to unacceptable bias and variance. When we are processing signals with such methods, it is important that we determine whether a result is an artifact of our processing technique or a property of the data. A value of order unity for the polarization estimate (2.18) provides an example. Is the estimate large due to a polarized signal, or is it because we have not made the estimate with enough degrees of freedom? If the latter is true, how do we know when we do have enough degrees of freedom in the estimate?

Such questions cannot be answered directly, but we can gain an intuition toward answering them. Jenkins and Watts [1968] describes a technique known as window closing, in which we are encouraged to explore the parameter space surrounding the construction
of any estimate. This method allows us to gauge whether or not our processing technique will converge on a stable value or not. For the polarization estimate, we typically find that the value descends from order unity to a stable value as the degrees of freedom are increased to $O(10)$ [Olson and Szuberla, 1997].

Finally, we caution against confusing feature-like properties in the results of processing techniques with physical reality. These are perhaps the most difficult artifacts to detect, because they appear to be real. Wolfe et al. [1994] mistook the sidelobes of a particular window function to be peaks in the spectrum associated with a magnetospheric process. Sidelobes from leakage are a property of the window in question, not the physical process. The same window applied to any data would produce the same leakage characteristics in the sidelobes. Problems of this nature can be avoided through the use of surrogate or contrived data, allowing a determination of whether the feature is a function of the data or the technique [e.g., Theiler et al., 1992].
Appendix A

The discrete Karhunen-Loéve transform

Developments of the discrete Karhunen-Loéve transform are typically performed under the assumption of square matrices [e.g., Fu, 1968; Preisendorfer, 1988; Therrien, 1992]. While this simplifies the analysis greatly, real-world data seldom afford us that luxury. What follows is a general development of the DKLT, one that may be applied to typical data sets that do not lend themselves to square analysis. Note that although we have assumed real data throughout this thesis, the following development of the DKLT is quite general and valid for complex data.

We develop a two-pronged eigensystem approach to the DKLT, similar to the singular value decomposition of Akansu and Haddad [1992]. This method is used to avoid the numerical difficulties found by Mees et al. [1987]. For practical computation, such as the analysis of magnetometer time series and spectral data, we find the procedure quite sound.

A.1 Development

Given a set of $N$-dimensional random vectors, $\{|x_m\rangle\}$, with $m \in [0, M - 1]$, we can expand each of the vectors in terms of a finite set of orthonormal basis functions

$$|x_m\rangle = \sum_{b=0}^{M-1} v_b[m] |a_b\rangle$$  \hspace{1cm} (A.1)
In our “bra-ket” notation we define the complex conjugate transpose of a vector (A.2a) and the inner product (A.2b) as

\[ |x_m\rangle = \langle x_m| \tag{A.2a} \]
\[ \langle x_m|x_m \rangle = \sum_{n=0}^{N-1} x_m[n] x_m^*[n] \tag{A.2b} \]

such that the inner product of a “bra”-vector and its corresponding “ket”-vector is positive, semi-definite. The orthonormality requirement

\[ \langle a_i|a_j \rangle = \delta_{ij} \tag{A.3} \]

placed on the basis functions makes the bases form a complete set; they span the \( \mathbb{R}^N \) space of the original set of vectors [Aubry et al., 1991].

Orthonormal bases are characteristic of functional decomposition, as we have seen in the case of the DFT (2.3) and complex exponentials. In the DKLT formalism however, we further require an orthogonal set of vectors formed from the expansion coefficients: \( |v_b\rangle \), where \( |v_b\rangle = (v_b[0] v_b[1] \ldots v_b[M-1])^\dagger \). This orthogonality requirement is given by

\[ \langle v_i|v_j \rangle = c_i \delta_{ij} \tag{A.4} \]

If the random vectors \( \{ |x_m\rangle \} \) are all zero-mean, then (A.4) specifies that the coefficients are uncorrelated [Therrien, 1992]. The restriction (A.4) carries with it several advantages in the information theoretic sense.

The expansion coefficients of the DKLT in (A.1) can be solved for directly, using both (A.3) and (A.4)

\[ \langle a_i|x_m \rangle = \sum_{b=0}^{M-1} v_i[m] \langle a_i|a_b \rangle = v_i[m] \tag{A.5} \]

This relation (A.5) between the coefficients, bases and original random vector is the forward DKLT, and the decomposition in orthogonal basis functions (A.1) is the Karhunen-Loève expansion — also known as the inverse transform.

In the case of the DFT, the orthonormal basis functions were known — complex exponentials or sinusoids. Solving for the expansion coefficients completes the problem of the DFT. In the case of the DKLT however, we have solved for the expansion coefficients in
principle only. Since we seek an orthonormal set \( \{|a_0\rangle\} \) that satisfies (A.4), its form cannot be known \textit{a priori}; rather, the form of the set \( \{|x_m\rangle\} \) uniquely determines the bases into which it can be decomposed.

Before solving for the basis functions of the DKLT, we first make our notation more compact. The original set of random vectors \( \{|x_m\rangle\} \) can be formed into a convenient matrix

\[
X \equiv \left( |x_0\rangle |x_1\rangle \cdots |x_{M-1}\rangle \right)^\dagger \tag{A.6}
\]

This arrangement is commonly known as a data matrix or an embedding matrix [Mees et al., 1987]. Similar matrix arrangements can be made for the basis functions and expansion coefficients

\[
A \equiv \left( |a_0\rangle |a_1\rangle \cdots |a_{M-1}\rangle \right)^\dagger \tag{A.7a}
\]

\[
V \equiv \left( |v_0\rangle |v_1\rangle \cdots |v_{M-1}\rangle \right) \tag{A.7b}
\]

Now, from the orthonormality of the set \( \{|a_0\rangle\} \) (A.3) and the orthogonality of the set \( \{|v_0\rangle\} \) (A.4) we find that

\[
AA^\dagger = I \tag{A.8a}
\]

\[
VV^\dagger = C \tag{A.8b}
\]

where \( I \) is the identity matrix and \( C_{ij} = c_i \delta_{ij} \). Note that \( A \) is unitary only in the special case of square matrices, \( M = N \). We can use the matrices (A.6) and (A.7) together with their relations (A.8) to derive matrix expressions analogous to the inverse (A.1) and forward (A.5) Karhunen-Loéve transforms

\[
X = VA \tag{A.9a}
\]

\[
V =XA^\dagger \tag{A.9b}
\]

Taking the first of these relations (A.9a) and right multiplying by its adjoint gives

\[
XX^\dagger = VAA^\dagger V^\dagger = VV^\dagger \tag{A.10}
\]

Left and right multiplying (A.10) by \( V^\dagger \) and \( V \), respectively, gives

\[
V^\dagger XX^\dagger V = V^\dagger VV^\dagger V = C^2 \tag{A.11}
\]
Here we define an auxiliary Hermitian matrix $\mathbb{P} \equiv XX^\dagger$ to rewrite (A.11) in compact form as an eigenvalue equation

$$V^\dagger \mathbb{P} V = C^2$$  \hspace{1cm} (A.12)

We note that the Hermitian matrix $\mathbb{P}$ is square, which gives it considerable analytic and computational utility. The solution of the eigenvalue problem is straightforward, and once the coefficients $V$ are known, we can solve (A.9a) for the basis functions $A$ as

$$A = C^{-1} V^\dagger X$$  \hspace{1cm} (A.13)

We have then a theoretical machinery developed for solving the DKL T, one which provides us with both the expansion coefficients (A.12) and basis function (A.13). It is also possible to solve an eigenvalue equation directly for the basis functions. We develop the second solution here, and later find a computational motivation for developing two equivalent eigensystem solutions.

Taking the second of the matrix relations (A.9b) and left multiplying by its adjoint gives

$$V^\dagger V = AX^\dagger XA^\dagger$$  \hspace{1cm} (A.14)

Again, we define an auxiliary Hermitian matrix, $\mathbb{R} \equiv X^\dagger X$, which is proportional to the covariance matrix [Broomhead and King, 1986]. We note that $\mathbb{R}$ shares the same square property as $\mathbb{P}$. If the vectors comprising $X$ are all zero-mean, $\mathbb{R}$ is proportional to the correlation matrix. With it we rewrite (A.14) in compact form as another eigenvalue equation

$$ARA^\dagger = C$$  \hspace{1cm} (A.15)

Once again, the basis functions $A$ are solved via the eigenvalue problem (A.15), and we have already a solution for the coefficients $V$ in (A.9b).

The DKL T is unique, in that it is the only orthonormal expansion that results in the orthogonality of the coefficient vectors (A.4) [Therrien, 1992]. This uniqueness stems from the requirement that the basis functions must be determined using the data directly. The uniqueness of the DKL T can be derived from the optimal property that we develop in the next section. Specifically, it is governed by the information stored in the matrices $\mathbb{P}$ and $\mathbb{R}$. From the equivalent eigensystems, we see that the orthonormal bases are simply the
eigenvectors of the matrix $R$ (A.15), and the orthogonal coefficient vectors are the eigenvectors of the matrix $P$ (A.12). Too, the mean squared coefficients (A.4) are encoded in both matrices as the eigenvalues. Note that the eigenvalues of $P$ and $R$ are identical\footnote{Actually, only the non-trivial eigenvalues are identical, since $P$ and $R$ need not be of the same dimension. This property of the auxiliary matrices is central to the practical computation of the DKLT.} — they are the diagonal elements of $C$.

### A.2 Optimal property

In any orthonormal decomposition, we assume that the inverse is true — the decomposed vector can be reconstructed exactly, using all of the expansion terms. In the information theoretic sense, the DKLT expansion (A.1) of a random vector represents the most efficient packing of energy among its expansion coefficients and basis functions [Akansu and Haddad, 1992]. The energy contained in the data is defined as the sum of the eigenvalues of the correlation matrix [Armbruster et al., 1994], which we have seen is proportional to $R$. The uniqueness and optimality of the DKLT stem from the decorrelation of the expansion coefficients in (A.4) [Rao, 1973].\footnote{Strictly speaking, the members of the set $\{|v_b\rangle\}$ are uncorrelated only in the case where the vectors to be decomposed $\{|x_m\rangle\}$ are all zero mean. It is sufficient, but not necessary that this condition be met in order to ensure the information theoretic properties of the DKLT. The statistical orthogonality of the expansion coefficients (A.4) is the necessary and sufficient condition for the uniqueness and optimality of the DKLT.} What this means in the practical sense, is that the DKLT allows us to truncate the expansion and approximate the original vector with the least amount of error. We now show that the truncated Karhunen-Loève representation is the optimal, in terms of the mean-square error introduced.

Given any of the random vectors comprising $\mathcal{X}$, an approximation $|\tilde{x}_m\rangle$ can be made by using only the first $K$ of the $M$ basis functions in $A$. This approximation will introduce some error $|\epsilon_m\rangle$ as

\[
|x_m\rangle = |\tilde{x}_m\rangle + |\epsilon_m\rangle = \sum_{b=0}^{K-1} v_b|m| |a_m\rangle + \sum_{b=K}^{M-1} v_b|m| |a_m\rangle
\]  

(A.16)

Using the expectation operator, $E \equiv \frac{1}{M} \sum_{m=0}^{M-1} \langle \epsilon_m | \epsilon_m \rangle$, the mean-square error in making the approximation is defined as

\[
E \equiv E \{ \langle \epsilon_m | \epsilon_m \rangle \}
\]  

(A.17)
The mean-square error may be expanded

$$E = \mathcal{E}\left\{ \left( \sum_b v_b^* |a_b\rangle \langle c | \langle a_c \right) \right\} \left( \sum_c v_c |a_c\rangle \langle b | \langle a_c \right) \right\}$$

$$= \mathcal{E}\left\{ \sum_{b,c} v_b^* v_c |a_b\rangle \langle a_c \right) \right\}$$

$$= \mathcal{E}\left\{ \sum_b |v_b|^2 \right\}$$

(A.18)

Substituting the forward DKLT (A.5) into (A.18), the mean-square error becomes

$$E = \mathcal{E}\left\{ \sum_b |\langle a_b|x_m\rangle|^2 \right\}$$

$$= \frac{1}{M} \sum_b \langle a_b|x^\dagger x|a_b\rangle$$

$$= \frac{1}{M} \sum_b \langle a_b|\Re|a_b\rangle$$

(A.19)

We then minimize the error $E$ subject to the orthonormality constraint on the basis functions (A.3), $1 - \langle a_b|a_b\rangle = 0$. Using the method of Lagrange’s undetermined multipliers, we form the quantity $\mathcal{L}$

$$\mathcal{L} = \sum_b \langle a_b|\Re|a_b\rangle + \sum_b \varphi_b \left(1 - \langle a_b|a_b\rangle \right)$$

(A.20)

Minimization is accomplished by taking the vector gradient with respect to a basis function and setting the result equal to zero. The notation used here, due to Therrien [1992], is elegant, compact and equivalent to the standard methods of constrained minimization of a complex, scalar quantity [e.g., Morse and Feshbach, 1953; Arfken, 1985]. We take the gradient

$$\nabla_{\langle a_b\rangle} \mathcal{L} = \Re|a_b\rangle - \varphi_b |a_b\rangle = 0$$

(A.21)

and arrive at an eigenvalue equation for $\Re$

$$\Re|a_b\rangle = \varphi_b |a_b\rangle$$

(A.22)

The Lagrange multipliers are thus determined to be the eigenvalues of $\Re$. From (A.15) we know that they are simply the diagonal elements of $\mathcal{C}$, $c_b$. Substituting (A.22) into (A.19),
we rewrite the mean-square error as
\begin{align*}
E &= \frac{1}{M} \sum_b \langle a_b | R | a_b \rangle \\
&= \frac{1}{M} \sum_b \langle a_b | c_b | a_b \rangle \\
&= \frac{1}{M} \sum_b c_b \\
\end{align*}
(A.23)

The minimization for the $R$-based DKLT is now complete; however, since we have claimed the $P$-based approach is equivalent, it must also be shown to be optimal in the mean-square sense.

We can write any of the diagonal elements of the matrix (A.12) as
\begin{equation}
c_b^2 = \langle v_b | P | v_b \rangle \\
(A.24)
\end{equation}
which allows us to recast the mean-square error (A.18) as
\begin{equation}
E = \frac{1}{M} \sum_b \frac{\langle v_b | P | v_b \rangle}{c_b} \\
(A.25)
\end{equation}

Again, we will minimize this quantity, subject to the orthogonality constraint placed on our coefficient vectors (A.4), $1 - \langle v_b | v_b \rangle c_b = 0$. Using the method of Lagrange’s undetermined multipliers, we form the alternate quantity, $L'$,
\begin{equation}
L' = \sum_b \frac{\langle v_b | P | v_b \rangle}{c_b} + \sum_b \vartheta_b \left(1 - \frac{\langle v_b | v_b \rangle}{c_b}\right) \\
(A.26)
\end{equation}

Minimizing $L'$ with respect to a coefficient vector and setting the result equal to zero yields
\begin{equation}
\nabla_\langle v_b | L' = \frac{P | v_b \rangle}{c_b} - \frac{\vartheta_b | v_b \rangle}{c_b} = 0 \\
(A.27)
\end{equation}

which gives us another eigenvalue equation, this time for $P$
\begin{equation}
P | v_b \rangle = \vartheta_b | v_b \rangle \\
(A.28)
\end{equation}
The alternate Lagrange multipliers are now determined to be the eigenvalues of the matrix $P$. Recall however, that in our development of the DKLT we noted that the eigenvalues of $P$ and $R$ were identical; thus $\vartheta_b = c_b$. Using this, we demonstrate the equivalence of
the optimal representation property of both approaches by substituting (A.28) into (A.25), again giving (A.23).

By invoking the singular value decomposition theorem [Rao, 1973], it is easy to show that the eigenvalues of $P$ and $R$ are positive, semi-definite; we also note that this is guaranteed by our constraint on the expansion coefficients (A.4). Knowing this property of the eigenvalues, we can perform the minimization procedure by inspection. From (A.23) the mean-square error in making an approximate expansion of a vector $|\tilde{x}_m\rangle$, with only the first $K$ basis functions of $M$ available, is proportional to the sum of the last $(M - K)$ eigenvalues associated with the ignored basis functions. So the optimal truncated representation of $|x_m\rangle$ is made using the basis functions associated with the $K$ largest eigenvalues of either $P$ or $R$.

### A.3 Computation

Given a set of $N$-dimensional random vectors $\{|x_m\rangle\}$, the problem of interest is to find the matrices $A$ and $V$ such that the terms of the orthonormal decomposition (A.1) are known. In principle, either solution — the eigenvalue problem in $P$ or $R$ — is adequate; however, a consideration of the properties of the data, or embedding matrix $X$ used to construct them provides insight as to how to approach the computation efficiently.

Recall that $X$ is of dimension $[M,N]$, where $M \neq N$, in general. This determines that the auxiliary matrices, $P$ and $R$, will be of dimension $[M,M]$ and $[N,N]$, respectively. The rank of a matrix is defined as the number of its non-zero singular values [Orfanidis, 1988]. Letting $K \equiv \lfloor M,N \rfloor$ and $K' \equiv \lceil M,N \rceil$, the rank of $X$ has the property $r \leq K$. It is a simple matter to show that $P$ and $R$ each inherit the rank of $X$, and by (A.12) and (A.15) they must have the same eigenvalues — these are simply the squared singular values of $X$. The disparate dimension of $P$ and $R$ then guarantees that the additional $(K' - K)$ eigenvalues of the larger of the two are equal to zero. The efficacy of current eigensystem routines, virtually all of which are based on the algorithm of Golub and Reinsch [1970], go as $O(K^3)$ operations for a Hermitian matrix of dimension $K$ [Press et al., 1992]. Since the larger of $P$ and $R$ contains no more information than the smaller, it makes great computational sense to solve the lesser eigenvalue problem: in $P$ for $M < N$, or in $R$ for $M \geq N$. In the event
<table>
<thead>
<tr>
<th>Case</th>
<th>Auxiliary matrix</th>
<th>A</th>
<th>V</th>
</tr>
</thead>
<tbody>
<tr>
<td>$M &lt; N$</td>
<td>$\mathbb{P}: [M, M]$</td>
<td>$[M, N]$</td>
<td>$[M, M]$</td>
</tr>
<tr>
<td>$M \geq N$</td>
<td>$\mathbb{R}: [N, N]$</td>
<td>$[N, N]$</td>
<td>$[M, N]$</td>
</tr>
</tbody>
</table>

Table A.1. Discrete Karhunen-Loève transform matrix dimensions. Shown are the dimensions of the DKLT auxiliary matrices and expansion matrices possible for an embedding matrix $\mathbb{X}$ of dimension $[M, N]$.

$M = N$, we prefer to solve the eigenvalue problem in $\mathbb{R}$, as most eigensystem packages produce normalized eigenvectors on output. This eliminates a renormalization step for $\mathbb{V}$.

The rank of a matrix can alternately be expressed as the number of its linearly independent rows or columns [Therrien, 1992]. For practical data, where additive noise is ubiquitous, the rows and columns of $\mathbb{X}$ are nearly always linearly independent. Thus, to some precision, $r = K$; however, one must be prepared for numerical exceptions. If $r < K$, $\mathbb{C}$ will have at least one diagonal element equal to zero. In this instance, $\mathbb{C}$ is singular and therefore we cannot solve for the basis functions via (A.13). Additional numerical problems can arise if $\mathbb{X}$ is ill-conditioned, but not strictly singular [e.g. Broomhead and King, 1986; Mees et al., 1987]; however, we find these cases easy to handle. If the eigenvalue problems (A.12) and (A.15) prove intractable, one can always turn to singular value methods which attack $\mathbb{X}$ directly [Mees et al., 1987]. We typically find that $\mathbb{X}$ has $r = K$, and so solve the lesser of the two eigenvalue problems, (A.12) or (A.15). Even in the face of numerical anomalies — spurious negative or complex eigenvalues, or exact zeros on the diagonal of $\mathbb{C}$— we find our two-pronged eigensystem approach to be very stable and reliable.

Knowledge of the number of terms available in the expansion (A.1), as a function of the dimensions of $\mathbb{X}$, makes it possible to recognize and avoid most numerical difficulties. According to (A.1), for $M$ vectors comprising $\mathbb{X}$ one should have available $M$ basis functions and coefficient vectors. As can be seen in the summary presented in Table A.1, this is only the case when $M < N$. For $M > N$, we have an overdetermined system, one in which the remaining $(M - N)$ eigenvalues may be thought of as zero. Effectively then, any basis function and coefficient vector associated with a zero eigenvalue may be discarded — therein lies our protocol for handling zeros on the diagonal of $\mathbb{C}$.  


Bibliography


Lacoss, R. T., Data adaptive spectral analysis methods, 36, 661, 1971.

McHarg, M. G., and J. V. Olson, Correlated optical and ULF magnetic observations of the winter cusp – boundary layer system, 19, 817, 1992.


Ziesolleck, C. W. S., Q. Feng, and D. R. McDiarmid, Pc 5 ULF waves observed simultaneously by GOES 7 and the CANOPUS magnetometer array, *101*, 5021, 1996.