PSEUDO POWER SIGNATURES FOR NONSTATIONARY SIGNAL ANALYSIS AND CLASSIFICATION

A Dissertation

Submitted to the Graduate Faculty of the
Louisiana State University and
Agricultural and Mechanical College
in partial fulfillment of the
requirements for the degree of
Doctor of Philosophy

in

The Department of Electrical and Computer Engineering

by
Vidya Venkatachalam
B.E., Anna University, India, 1992
M.S., Louisiana State University, U.S.A., 1995
December 1998

Acknowledgments

I express my most sincere thanks to my advisor, Dr. Jorge L. Aravena, for his guidance, time, and patience, throughout the period of this work. I am deeply grateful for the support and encouragement I have always received from him, without which this work would not have been possible. I also express my thanks to Dr. Frank Neubrander for initiating me into a better understanding of the fundamentals of mathematical analysis, and to Dr. Charles Harlow, for his valuable contributions and ideas, which helped to increase the applicability of this research. I thank Dr. Guoxiang Gu, and Dr. Kemin Zhou, for providing a rigorous foundation for my research, by the knowledge they have imparted through the various courses I have taken with them; and Dr. Ravi Rau for agreeing to be on my committee. My thanks go to the faculty and staff at the Departments of Electrical Engineering, and Mathematics, who have been friendly and helpful, making my stay here very pleasant; and my family and friends for their support at all times. Finally, I specially thank my husband, Mukund, for enriching my life with his truly colorful and unique personality!

Table of Contents

List of Tables

List of Figures

Notation

W_x	Wigner distribution of a function x
F_x	Short Time Fourier Transform of a function x
SP_x	Spectrogram of a function x, obtained from F_x
ψ	Admissible wavelet function
Ψ	Fourier Transform of ψ
c_ψ^x	Continuous Wavelet Transform of a function x
SC_x	Scalogram of a function x, obtained from c_ψ^x
C_x	Cohen distribution for a function x
PC	Principal component obtained using a singular value analysis
C_ψ^x	Matrix of samples of c_ψ^x
C_ψ	$2\pi \int_{-\infty}^{\infty} \frac{\|\Psi(\omega)\|^2}{\|\omega\|} d\omega$
$\Re$	Real numbers
$\mathcal{Z}$	Integers
$\mathcal{C}$	Complex numbers
$\mathcal{C}^N$	Complex vectors of dimension N
$\mathcal{C}^{L\times N}$	Complex matrices of dimension L by N
$L^2(\Re)$	Hilbert space of square integrable functions, defined as $\{x(t) : \int_{-\infty}^{\infty} \mid x(t) \mid^2 dt < \infty\}$
$l^2(\mathcal{Z})$	Hilbert space of square summable doubly infinite sequences, defined as $\{x(n) : \sum_{n=-\infty}^{\infty} \mid x(n) \mid^2 < \infty\}$
H	Weighted Hilbert space defined as $H = L^2(\Re^2, C_\psi^{-1} \frac{dadb}{a^2})$

$l^2(\mathcal{Z}^2, \frac{1}{2^{4l}})$	Weighted Hilbert space of square summable sequences, defined as $\{x(l,n) : \sum_{l=-\infty}^{\infty} \sum_{n=-\infty}^{\infty} \frac{\lvert x(l,n)\rvert^2}{2^{4l}} < \infty\}$
M	Space of the Continuous Wavelet Transforms
S	Weighted Hilbert space defined as $S = L^2(\Re, C_\psi^{-1} \frac{da}{a^2})$
R	Hilbert space defined as $R = L^2(\Re, db)$
X^*	Conjugate transpose of the matrix X
X^c	Conjugate of the matrix X
X^T	Transpose of the matrix X
$\overline{x}$	Conjugate of the function x
X^{-1}	Inverse of the operator X
X^*	Adjoint of the operator X
$\langle \cdot, \cdot \rangle_X$	Inner product in a Hilbert space X
$\langle \cdot, \cdot \rangle_2$	Inner product in the Hilbert space $L^2(\Re)$ or $l^2(\mathcal{Z})$
$\lVert \cdot \rVert_X$	Norm in a metric space X
$\lVert \cdot \rVert_2$	Norm in the Hilbert space $L^2(\Re)$ or $l^2(\mathcal{Z})$
$X \otimes Y$	Tensor product of the vector spaces X and Y
$X \otimes Y$	Kronecker product of the matrices X and Y
$X \oplus Y$	Direct sum of the vector spaces X and Y
F, G	Low and high pass filtering operators in the filter bank
Γ	Continuous Wavelet Transform Operator
$\mathcal{T}$	Approximation to the sampling operator in H

$\mathcal{S}$	Shensa operator
$\mathcal{K}$	Orthogonal projection operator in H with range M
J	Functional to be minimized
F_2	Frame operator on $L^2(\Re)$
χ_x	Characteristic function of x
$f = \mathcal{O}(g)$	Order of : There exists K such that $f(n) \leq Kg(n)$

Abstract

This research focuses on the analysis and classification of multicomponent non-stationary signals of arbitrary duration. The proposed classification approach has potential applications in areas like moving target detection, object recognition, oil exploration, and speech processing. The wavelet transform is used as the basis for the analysis. The classification technique is based on novel scale energy density functions, called *pseudo power signatures*, which are independent of signal length, and which can be used to characterize the time-scale energy distribution of the signal. These signatures allow for fast classification of signals regardless of their length.

Two approaches to determine pseudo power signatures are presented in this work. The first approach is based on a singular value principal component analysis technique, which, though computationally simple, is not very sensitive to signal characteristics. The second is a more sophisticated approach, and is optimal in a weighted least mean squares sense. The latter technique involves solving an inverse projection problem arising from a nonlinear infinite dimensional minimization, and generates good quality signatures with excellent discriminating capability. An algorithm, with fast convergence, for application to discrete data sets is developed, and a complete analysis of the computational complexity is obtained. Several simulation examples are presented to illustrate the methodology, and its application to practical classification problems. Finally, suggestions for further work in the area are given.

Chapter 1

Introduction

Signal classification is an area of great importance in a wide variety of applications. Representative applications include system identification, moving target detection, oil exploration, and pattern recognition. In these applications, the signals are generally nonstationary in nature; i.e., their statistical properties vary with time. Consequently, nonstationary signal classification is an area of active research in the signal and image processing community. The research in the area usually takes one or both of the following forms :

- The determination of good quality characteristic representations, often called *signatures*, for a signal class.
- The determination of efficient and reliable techniques that use the signatures to perform classification.

In this work, both the above issues are addressed. The main contributions of this work are the introduction of a new concept of signatures for signal classes, called **pseudo power signatures**, which are essentially *independent* of the signal length, the determination and complete analysis of these signatures, and their application to general classification problems. The notion of a signal class in this analysis is based on the energy spread in the time-frequency plane. Since the pseudo power signatures are independent of the signal length, they can be used to characterize arbitrary segments of signals in a class, using the *same characteristic representation*. More generally, the pseudo power signatures allow for the representation of

nonstationary signal classes in a manner that is completely independent of time. At present, no reliable techniques exist which provide for signatures for nonstationary signals that have the time independence property. The importance of this property becomes apparent when one considers a typical application of nonstationary signal classification, which was a motivation for this research, namely, the subsurface classification problem.

1.1 Need for time independent signatures

Subsurface classification is essentially a stratigraphic analysis problem. The objective is to identify, and thus classify, subsurface strata like sand, clay, silty clay, or solid rock. Traditional approaches to performing this classification are mostly invasive, involving drilling operations, and a manual (visual) classification. In addition to being costly, subjective and time consuming, the traditional approaches can introduce negative environmental effects such as creating paths for contaminant migration through the drilling operations, and may not always be feasible.

The modern approach to stratigraphic analysis attempts to develop non-invasive subsurface exploration methods. The approach that motivated this research is an example of a non-invasive technique which uses electromagnetic signals. The principle behind the non-invasive techniques is that electromagnetic signals behave differently when they propagate through different media. Hence, an electromagnetic signal that has propagated through the subsurface should contain information about its structure. This information is recorded in the reflected echoes of the electromagnetic signals. Consequently, one could study the experimental data obtained from the recorded echoes to establish features that are characteristic of the various subsurface strata, and use the presence of these features in a recorded echo of unclassified

strata, to infer the subsurface composition. Further, by either knowing or estimating the signal propagation speed, one could relate the time of occurrence of the features in the echo to the depth (spatial location) of the associated strata.

The problem with the subsurface analysis using the recorded echoes of propagated electromagnetic signals, is that subsurface strata classification presents difficulties which limit the applicability of conventional processing techniques. Layers can be located at any depth, and can be of any thickness, or even be missing. However, it is reasonable to expect that propagation through, say, a sand layer of a certain thickness and depth, will affect the electromagnetic signal in a manner quite different from a similar propagation through a layer of clay of the same thickness and depth. In other words, the propagation of the electromagnetic signal through a layer produces changes in it that are characteristic of the layer itself, and independent of the depth or thickness of the layer. Implicitly, one is postulating that in each layer, there exist *intrinsic characteristic properties which are (ideally) independent of the location and thickness* of the layer. This assumption is supported by experimental evidence. Specially trained expert geologists can infer subsurface composition by analyzing the electromagnetic echoes from the subsurface. The analysis though, is very time consuming and has a strong subjective component. There is therefore a need to formulate tools that can automate the entire classification process.

One approach to classification involves obtaining mathematical models that accurately characterize the associated physical phenomena. However, this approach is very cumbersome and is still in need of basic research to understand all the phenomena involved in the electromagnetic propagation through various media. A signal processing approach to classification appears to be a more attractive alternative. Consequently, many modern automated classification methods use signal processing

analysis techniques to perform the classification. In the classification approach using signal analysis techniques, every layer corresponds to a signal class, and the trace of the recorded echo of the electromagnetic signal propagated through the associated layer constitutes a member of the class. Representative members of the different classes are used to generate **signatures** or characteristic representations for each signal class. The signatures can then be used to classify a composite signal containing components that are members of any of the signal classes, having any arbitrary duration, and located in any random order. For this purpose, one can easily see that the signatures are required to be time independent, both in terms of location and duration of the component signal. The need for signatures which are time independent is not specific to the stratigraphic analysis problem, but is common to several other applications as well. Applications such as moving target detection, sound recognition, and underground mine detection also require characteristic representations that are essentially independent of time. A simple example is the recognition of the letter "A" as spoken by a person on several different occasions in different forms (words). One can clearly see that neither the location of the letter nor the duration remains unchanged in each occurrence. However, a good quality characterization should be capable of identifying the sound regardless of when and for how long it occurs. Thus, the general framework used to develop time independent signatures is very useful in several problems, and contributes to an increased quality of classification algorithms.

1.1.1 Modern classification approaches

Current research in classification essentially focuses on the determination of good quality signatures. For the classification involving nonstationary signal classes, most

modern classification approaches use time-frequency representations to characterize the signal classes. The representations are derived using suitable time-frequency distributions or TFDs ([18]), as it is commonly accepted that TFDs are best suited to characterize and study nonstationary signals ([25]). In the current literature, there exist several classification schemes which use time-frequency representations to perform the classification. One scheme uses signal expansion techniques with specially chosen basis functions ([21],[6],[13],[22]). The main emphasis in this approach is the selection of a basis for a particular signal class, such that the elements of the signal class are well represented using an expansion in terms of the basis. This implies that the defining features of the particular signal class are enhanced when elements of the class are projected onto the basis elements. The basis is then used as being representative of the class. Thus, a signal from a different class cannot be well represented by the basis functions, and, in effect, exhibits a poor 'match' with the basis elements. The degree of 'match' then is the defining criterion for the classification. This approach is very similar to the matched filtering concept commonly used for signal detection. Another scheme uses kernel function design techniques ([8],[9],[10]), where the main principle is the choice of an appropriate kernel to characterize the TFD of a signal class. It was shown by Cohen in [18] that every TFD can be represented as a generalized distribution associated with a specific kernel. The choice of the kernel thus critically determines the properties of the associated TFD. When used as a classification tool, the kernel is designed for a signal class such that the associated TFD for the class is clearly distinguishable from the associated TFD of any other class, using the same kernel. In this case, the kernel is representative of the signal class. Other approaches include principal component analysis techniques ([11],[24]), where the principal components of the

TFD of the signal class are extracted, usually using some form of a singular value, or eigen value analysis, and used to characterize the signal class, and significant feature extraction and mapping techniques ([3],[25]), where a set of well defined measures is used as being representative of the class. Such measures include projections onto predefined convex sets, projections onto eigenspaces, etc. There are also approaches which combine the use of the TFDs with other methods like statistical processing ([12],[23]), and neural networks ([2]) to achieve a better classification. In the first instance, signal classification is achieved based on the statistical properties like second order moments, covariances, etc, as determined from the TFD of the signal. In the neural network approach to classification, the characteristic parameters associated with a signal class, which are determined from the corresponding TFD, are used to train a neural network to identify components of the class. Each of these techniques provides signatures which are essentially time-frequency representations, and are optimal in the context of the specific problem motivating their formulation. However, their applicability to problems where one requires intrinsic representations which are not functions of time, is limited. Hence, even though the individual approaches work very well for specific problems where signal length is a known parameter, there is a very real need to establish a methodology to determine time independent signatures that can be applied to general classification problems.

1.2 Overview of work

This research formulates and solves a general classification problem for signals whose length is unknown. In order to establish the generalized framework, the classification to be performed can be formulated as the following general event detection problem, henceforth referred to as the problem P :

P : *There is a known class of events,* $\{E_k; k = 1, \ldots, n\}$*, which may appear in a given scene for a variable time interval. One has collected data from the scene as a signal* $x(t)$; $t_l \leq t \leq t_h$*, and it is known that only one event is present at any given time. Then there is an unknown partition* $P_x = \{t_l \leq t_1 \leq t_2 \ldots \leq t_r \leq t_{r+1} \ldots \leq t_h\}$*, of* **transition times** *marking the start and end times of an event. The goal is to determine the transition times and the events occurring in each time interval. This process is called* **classification of the signal** $x(t)$.

Chapter 2 provides a brief review of common time-frequency distributions, and develops the notion of pseudo power signatures.

Chapter 3 gives a solution, using a discretized matrix singular value analysis, to the generation of the pseudo power signatures, and demonstrates through examples, the effectiveness and limitations of this approach.

Chapter 4 discusses the theory, problem formulation and the solution procedure of an approach using orthogonal projections in the space of wavelet transforms, to generate the signatures.

Chapter 5 develops a computational algorithm to determine the projection signatures, and illustrates, through examples, the applicability to a general classification problem.

Chapter 6 summarizes the results obtained, and offers suggestions for future work.

Chapter 2

Pseudo Power Signatures

This chapter presents a methodology to determine signatures, for nonstationary signal classes, to meet the requirements of the classification problem P. The problem requires the determination of signatures for nonstationary signal classes which are essentially independent of time. Since the underlying theory behind the generation of signatures for these classes involves the use of time-frequency distributions (TFDs), a brief review of some commonly used TFDs, their relative merits, and their applicability to the classification problem P, is first provided. The problem of determining time independent signatures is then addressed, leading to the formulation of the concept of *pseudo power signatures*.

2.1 Time-frequency Distributions (TFD)

Traditional signal analysis has usually been performed in the frequency domain by the use of *Fourier* techniques. The idea was to separate out the signal into its different frequency components using the Fourier Transform (FT). The signal spectrogram (energy spread of the signal in the frequency domain in terms of the square of the magnitude of its FT) was used to determine the significant frequency components of the signal, which were then used to characterize the signal. However, since the spectrogram did not provide any localization of time events, the time varying information of signals was completely lost when using conventional Fourier techniques. Effectively, a change at any one point in time, was spread out over the entire frequency range. Thus, conventional Fourier analysis was very ineffective for

signals which were nonstationary in nature; i.e., signals whose statistical characteristics varied over time.

Over the past few years, extensive attention has been given to the subject of *time-frequency distributions* for the analysis of nonstationary signals. Cohen, in [1], presents a review of these distributions and their applications. In the most general case, they allow description of the **energy** distribution of the signal in a particular time-frequency region. For the problem P, every event under consideration is regarded as a component of the nonstationary signal $x(t)$. A component of a nonstationary signal is defined as that part of the signal that has a high energy concentration in a localized time-frequency region, which is usually represented by a peak in the time-frequency plane. It follows then that a signal which shows a high energy concentration in a particular region has a significant component in that particular time-frequency region. The different components of a multicomponent signal can then be readily detected by the presence of several well localized peaks in the time-frequency plane. In general, for the analysis of multicomponent nonstationary signals, the choice of any one TFD over another in a given situation is mainly determined by their comparative resolution capabilities ([14]).

2.1.1 The Wigner Distribution

One of the most widely used time-frequency distributions is the **Wigner Distribution** ([15, 16, 17]). This was originally developed for problems in quantum mechanics, and has since been generalized for use in signal processing by Ville, and is often referred to as the Wigner-Ville distribution in signal processing literature. The Wigner Distribution (WD) of a signal $x(t)$ is defined as

$$W_x(t,f) = \int_\tau x(t+\frac{\tau}{2})\overline{x(t-\frac{\tau}{2})}e^{-j2\pi f\tau}d\tau$$

The signal energy over the time-frequency plane is obtained by integrating the WD over the entire plane as

$$E_x = \int_t \int_f W_x(t,f)dtdf$$

Similarly, the signal energy over a given support R is defined as

$$E_x^{(R)} = \int \int_{(t,f)\in R} W_x(t,f)dtdf$$

The energy concentration of the signal in R is then given by

$$\rho_x^{(R)} = \frac{E_x^{(R)}}{E_x}$$

This value is used as being representative of the presence of a component of a signal in a given time-frequency region, and is a criterion to identify the components of a signal in the different regions of the time-frequency plane ([5]). If the energy concentration in a particular time-frequency region exceeds a specified threshold, one can classify a signal component as being present in the region.

The WD is widely used as a basis for classification since it has very desirable properties, such as a high time-frequency concentration which allows a signal component to have a clearly distinguishable peak in the time-frequency plane. However, when dealing with multicomponent signals, as in the problem P, the WD has some very serious limitations. In most cases, the WD is incapable of resolving two components in a signal due to the presence of excessive cross-terms ([7]). One needs very high resolution in order to distinguish two closely spaced components, which brings up another problem in that at high resolutions, the WD produces negative values which are difficult to interpret in energy terms. Thus, the WD is a poor candidate for the decomposition of multicomponent nonstationary signals, and hence is not very useful for solving the problem P.

2.1.2 The Short Time Fourier Transform

Since the FT is such a widely used tool in signal processing, it was attempted to integrate the time-frequency dependency characteristic into the FT by a suitable modification. This led to the *Short Time Fourier Transform* ($STFT$), which effectively is the FT of the signal over a suitable time window $g(t)$. The $STFT$ of a signal $x(t)$ is defined as

$$STFT(t,f) = F_x(t,f) = \int x(\tau)\overline{g(\tau - t)}e^{-2j\pi f\tau}d\tau$$

The $STFT$ can be viewed as the filtering of the signal at all times with a bandpass filter having as impulse response the window function modulated to that frequency ([19]). Thus, the $STFT$ can be viewed as a modulated filter bank.

The energy distribution in the time-frequency plane obtained from the $STFT$ is given by the **spectrogram**(SP), and is defined in terms of the magnitude of the square of the $STFT$. Formally,

$$SP_x(t,f) = \mid F_x(t,f) \mid^2$$

The signal energy over a given support R is defined as

$$E_x^{(R)} = \int\int_{(t,f)\in R} SP_x(t,f)dtdf$$

The spectrogram associated with the $STFT$ can be used to provide very effective representations, even in situations where the WD fails, namely, the resolution of multicomponent signals. It has little cross-term interference, and always assumes only positive values, providing for an unambiguous interpretation. However, the $STFT$ is limited by the fixed time resolution it affords. The time resolution is effectively determined by the shape of the window function used, which is constant at all times. Hence, two time events not separated by an interval larger than the window

length cannot be effectively isolated by the $STFT$. For this reason, though the $STFT$ has much wider applicability than the FT, it is not an attractive candidate for the problem P, where the distance between the signal components is unknown, and the determination of the transition times is very important.

2.1.3 The Continuous Wavelet Transform

To overcome the resolution limitation of the $STFT$, one can imagine letting the resolution vary continuously over the time-frequency plane so that one obtains a variable resolution analysis ([19]). Intuitively, when viewed as a filter bank (similar to the $STFT$ case), this provides continuously increasing time resolution with increasing center frequency of the band pass filters in the analysis section. In other words, the frequency responses of the analysis filters no longer have constant bandwidth as in the case of the $STFT$, but they have instead, bandwidth proportional to their center frequency. This effect is shown in Figure 2.1. Effectively, this implies that, for a band pass filter with center frequency, say f_0, and bandwidth, say BW_0, the relative bandwidth, given by $\frac{BW_0}{f_0} = c$, where c is some constant. Band pass filters, that have this property are referred to in signal processing literature as constant - Q filters.

The constant - Q filtering effect is exactly the effect of the *Continuous Wavelet Transform* (CWT), which can be viewed as another class of TFDs. Wavelet Transform Theory is a recently developed area and is now being used widely for nonstationary signal analysis. The CWT of a function $x(t) \in L^2(\Re)$ is defined as

$$c_\psi^x(a,b) = \langle x, \psi_{a,b} \rangle_2 = \int_{-\infty}^{\infty} x(t)\overline{\psi_{a,b}(t)}dt \tag{2.1}$$

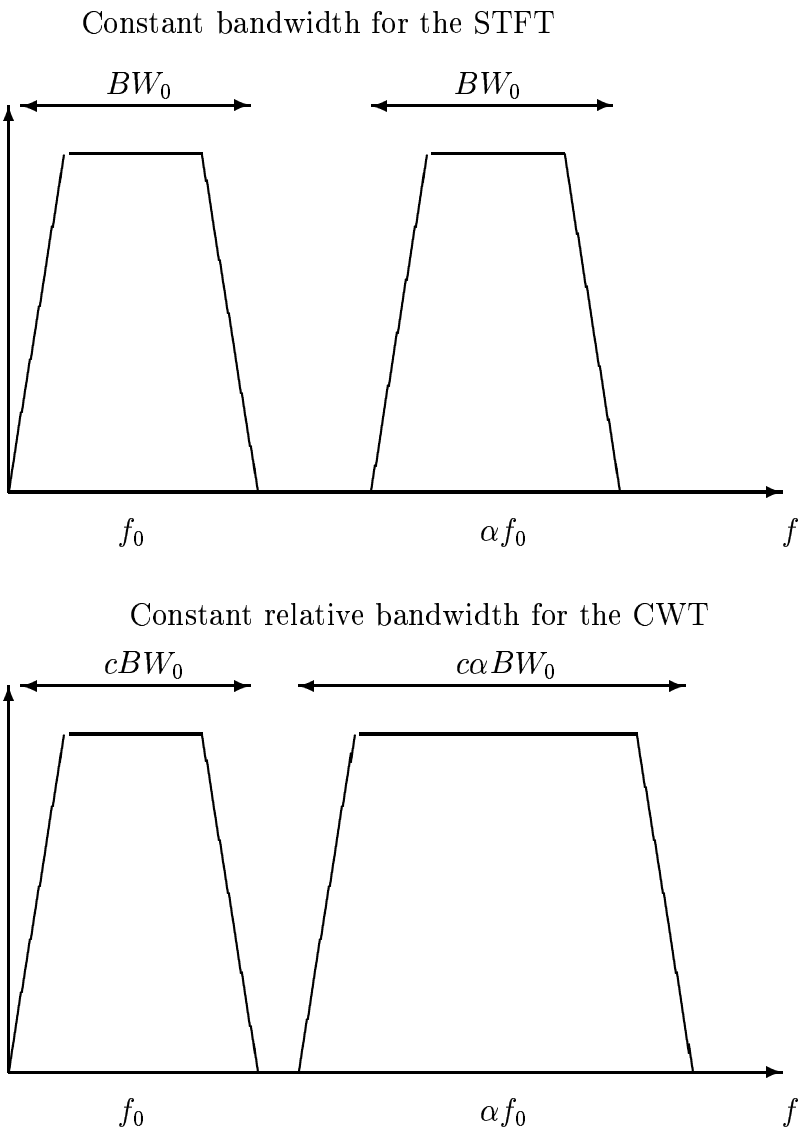

Figure 2.1: Frequency domain coverage by the STFT and the CWT

where $\psi_{a,b}$ is a shifted and dilated version of the basic function $\psi(t) \in L^2(\Re)$, and is defined as

$$\psi_{a,b} =\mid a \mid^{-\frac{1}{2}} \psi(\frac{t-b}{a}) \tag{2.2}$$

The basic function $\psi(t)$ is called the *'mother wavelet'* function. Formally, the CWT can be represented as a mapping $\Gamma : L^2(\Re) \rightarrow H = L^2(\Re^2, C_\psi^{-1}\frac{dadb}{a^2})$ with range M, where M is a closed subspace of H ([26]). The inverse CWT operator $\Gamma^{-1} : M \rightarrow L^2(\Re)$ is defined as follows :

$$\Gamma^{-1}[c_\psi^x](t) = C_\psi^{-1} \int_a \int_b c_\psi^x(a,b)\psi_{a,b}(t)\frac{dbda}{a^2}, \quad c_\psi^x(a,b) \in M \tag{2.3}$$

where $C_\psi = 2\pi \int_{-\infty}^{\infty} \frac{|\Psi(\omega)|^2}{|\omega|} d\omega$ with $\Psi(\omega)$ representing the FT of $\psi(t)$. This implies that the CWT is invertible whenever $C_\psi < \infty$, unlike the FT which has stronger constraints for a valid inversion. This condition for inversion is called the *admissibility condition.*

The energy spread of the CWT in the time-scale plane is defined by the energy distribution called the **scalogram** ([20]). The scalogram of a function $x(t)$ with CWT $c_\psi^x(a,b)$ is defined as

$$SC_\psi^x(a,b) =\mid c_\psi^x(a,b) \mid^2$$

The signal energy over a given support R is defined as

$$E_\psi^{x(R)} = C_\psi^{-1} \int\int_{(a,b)\in R} SC_\psi^x(a,b)\frac{dbda}{a^2}$$

The shifted and dilated wavelets $\psi_{a,b}$ provide a natural localization of a given function $x(t) \in L^2(\Re)$ (finite energy signals) in time and frequency through the use of the time index parameter **b** and the frequency index parameter **a**, more commonly referred to as the scale parameter. In general, the scale **a** is inversely proportional

to the frequency ([20]), which implies that large scales correspond to low frequencies in the signal, and vice versa. Thus, one can visualize the CWT as the extracting of the signal information at higher and higher frequencies (as the scale reduces), with the time localization of this information given by the corresponding **b** i.e. a variable windowing of the signal information with the window shape being determined by $\psi_{a,b}$. Note that $\psi_{a,b}$ can be viewed as the impulse response of a band pass filter with the constant - Q property. Due to the inherent trade - off between time and frequency resolutions, as determined by the Uncertainty principle [1], the CWT has good spectral and poor temporal resolutions at low frequencies which is useful for analyzing low frequency components of long duration, and good temporal and poor spectral resolutions at high frequencies which is useful for analyzing signal components of high frequency and short duration. Since most nonstationary signals encountered in practice are of these forms, the CWT provides a very good representation for these signals. Figure 2.2 represents pictorially the variable windowing (varying time resolution) effect of the CWT, as compared to the fixed windowing (constant time resolution) effect of the $STFT$. Thus, the CWT can clearly isolate two high frequency signals placed close together if a sufficiently high time resolution is used. This excellent *localization capability* of the CWT is extremely useful for solving the problem P, since one can then accurately determine the transition times between events.

[1] **Uncertainty principle (Heisenberg inequality)**([19]) :
If Δt, and Δf, represent the time and frequency resolutions respectively, then their product is bounded below; i.e., the time-bandwidth product $= \Delta t \Delta f \geq \frac{1}{4\pi}$.
This effectively implies that the resolution in both time and frequency cannot be arbitrarily small.

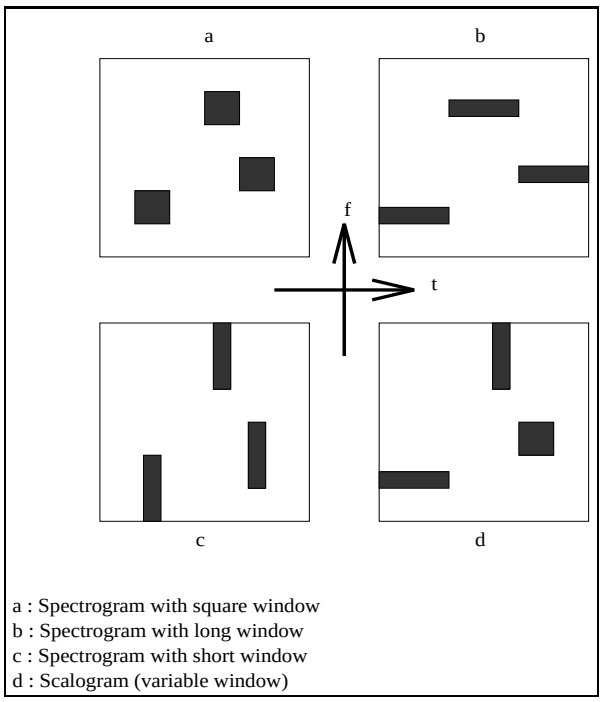

Figure 2.2: Variable windowing of the CWT

A complete generalization to the entire class of time-frequency distributions was given by Cohen ([18]) and is often referred to as the Cohen class of distributions. The Cohen distribution for a function $x(t)$ is given by

$$C_x(t, f; h) = \int \int W_x(\tau, \nu) h(\tau - t, \nu - f) \frac{d\tau d\nu}{2\pi}$$

where $h(t, f)$ is an arbitrary time-frequency characterization function, often referred to as the kernel function. Thus, the Cohen class of distributions is effectively the $2 - D$ correlation of the WD of any function x with the kernel function h. When viewed as a member of this class, the spectrogram associated with the $STFT$ can be represented as

$$SP_x(t, f) = \int \int W_x(\tau, \nu) W_g(\tau - t, \nu - f) \frac{d\tau d\nu}{2\pi}$$

where g is the analyzing window function. Similarly, the scalogram associated with the CWT can be obtained as

$$SC_{\psi}^{x}(a,b) = \int\int W_x(\tau,\nu) W_{\psi}(\frac{\tau - b}{a}, a\nu) \frac{d\tau d\nu}{2\pi}$$

The approach to classification using specially designed kernels which was mentioned in Chapter 1, attempts to design a kernel function $h(t,f)$ for each signal class such that $C_x(t,f;h)$ provides a desired energy characterization for that class. To meet the specific requirements of the classification problem P, the generalized TFD of choice is $C_x = SC_{\psi}^{x}$ which is obtained from the CWT, and corresponds to the kernel $h = W_{\psi}$. It has the properties of excellent time localization which is desirable for detecting the transition times, and efficient computation techniques using filter banks.

2.2 Time independent signatures

This section introduces a methodology for determining characteristic representations for signal classes that are essentially **independent** of the actual duration of each signal component. The methodology uses the concept of spectral energy distribution and develops a representation that allows one to define an "instantaneous energy distribution" which is called the *pseudo power signature*. Owing to the excellent localization capability of the CWT discussed earlier, it is the TFD chosen as the basis for this analysis.

Consider any $x \in L^2(\Re)$ with CWT, $c_{\psi}^{x}(a,b)$, where ψ is an admissible wavelet. The scalogram SC_{ψ}^{x} associated with c_{ψ}^{x} can be interpreted as a time-scale energy density function since one can write

$$\int |x(t)|^2 dt = C_{\psi}^{-1} \int_b \int_a SC_{\psi}^{x}(a,b) \frac{dbda}{a^2}$$

Hence, the function

$$SC_{\psi}^{x}(a,b) = |c_{\psi}^{x}(a,b)|^2$$

can be viewed as the corresponding time-scale power function and the function $SC_{\psi}^{x}(\cdot,b)$ as the "scale power distribution at time b". From the Mean Value Theorem for Integrals [2], one can estimate this function as follows :

Consider any interval $[b_1 \;\; b_2]$ over which $SC_{\psi}^{x}(a,\cdot)$ is continuous. Then, there exists $b_0 \in [b_1 \;\; b_2]$ such that

$$SC_{\psi}^{x}(a,b_0) = \int_{b_1}^{b_2} \frac{|c_{\psi}^{x}(a,b)|^2}{(b_2-b_1)} db$$

If the wavelet has compact support, then as the scale decreases, the value of $SC_{\psi}^{x}(a,b_0)$ is essentially independent of the values of $x(t)$ outside the interval $b_1 \leq t \leq b_2$. Thus, the scale power distribution can be estimated by the scalograms of small segments of the signal, and, for low scales, it would be essentially independent of the length of the record. The lower the scale that one can use, the smaller the segments that are required.

The ideal situation would arise if one could define a wavelet such that for a given class of signals the corresponding wavelet transforms are separable, i.e.

$$c_{\psi}^{x}(a,b) = s_{\psi}^{x}(a) r_{\psi}^{x}(b)$$

---

[2] **Mean Value Theorem for Integrals** : *If a function $f(\cdot)$ is continuous on an interval $[a,b]$, then there is a number $\mu \in [a,b]$ such that*

$$f(\mu)(b-a) = \int_a^b f(t)dt$$

The scale power distribution at time b_0 would then be given by

$$SC_{\psi}^{x}(a, b_0) = |s_{\psi}^{x}(a)|^2 \int_{b_1}^{b_2} \frac{|r_{\psi}^{x}(b)|^2}{(b_2 - b_1)} db; \quad b_0 \in [b_1 \;\; b_2]$$

It is apparent that the normalized distribution would essentially be independent of b_0. Thus, the scale function $s_{\psi}^{x}(a)$, suitably normalized, could be used as the **power signature** to characterize the corresponding signal class in a manner that is independent of duration.

The feasibility of this concept is strengthened by the fact that if one moves away from $L^2(\Re)$ signals, one can find functions whose 'formal' CWT is separable. For example, consider the power signal $x(t) = Ae^{-j\theta t}$. If $\psi(t)$ is an admissible wavelet with Fourier transform, $\Psi(\omega)$, the function

$$c_{\psi}^{x}(a, b) = \int x(t) \frac{1}{\sqrt{a}} \overline{\psi(\frac{t-b}{a})} dt$$

is defined for all values of $a \neq 0, b \in \Re$. Observe then that

$$c_{\psi}^{x}(a, b) = A\sqrt{a} \;\; \overline{\Psi(a\theta)} e^{j\theta b} = s(a) r(b)$$

A natural question that arises based on this result is the following :

Can one determine an admissible wavelet function $\psi \in L^2(\Re)$ that admits a similar result for functions in $L^2(\Re)$?

2.2.1 Approximate power signatures

The answer to the question posed at the end of the last section can be obtained from the Theorem stated below.

Theorem 2.1 *Given any nontrivial function $x \in L^2(\Re)$, and an admissible (nontrivial) wavelet $\psi \in L^2(\Re)$, the space of CWTs of x with respect to ψ does not contain any element of the form given by $c_{\psi}^{x}(a, b) = s(a)r(b)$, where $s \in L^2(\Re, C_{\psi}^{-1} \frac{da}{a^2})$, and $r \in L^2(\Re, db)$.*

PROOF. Let $H = L^2(\Re^2, C_\psi^{-1}\frac{dadb}{a^2})$, and $s \in S = L^2(\Re, C_\psi^{-1}\frac{da}{a^2})$, $r \in R = L^2(\Re, db)$. The space $H = S \otimes R$ ([27]), and hence, it trivially follows that $s \otimes r \in H$. Now, $S \otimes R$ is isomorphic with the space $(S \times R, C_\psi^{-1}\frac{da}{a^2} \otimes db)$, which implies that there exists a unitary operator $\overline{U}$ such that $\overline{U}(s \otimes r) = sr$. Hence, the element $sr \in H$. Let M be the space of the Continuous Wavelet Transforms. Then, M is a closed subspace of H ([26]). Let $c_\psi^x \in M$ denote the CWT of $x \in L^2(\Re)$, where $\psi \in L^2(\Re)$ is some admissible wavelet. Using a proof by contradiction, one can show that this function cannot be of the form $s(\cdot)r(\cdot)$.

By the definition of the CWT,

$$c_\psi^x(a,b) = \int_t x(t)\frac{1}{\sqrt{a}}\overline{\psi(\frac{t-b}{a})}dt$$

Assume that $c_\psi^x(a,b) = s(a)r(b)$. Then,

$$s(a)r(b) = \int_t x(t)\frac{1}{\sqrt{a}}\overline{\psi(\frac{t-b}{a})}dt$$

Keeping a fixed, taking the Fourier Transform on both sides, and using Fubini's Theorem[3] to interchange integrals, one gets

$$\begin{aligned}
s(a)\int_b r(b)e^{-j\omega b}db &= \int_b\int_t x(t)\frac{1}{\sqrt{a}}\overline{\psi(\frac{t-b}{a})}dte^{-j\omega b}db \\
s(a)R(\omega) &= \int_t x(t)\int_b \frac{1}{\sqrt{a}}\overline{\psi(\frac{t-b}{a})}e^{-j\omega b}dbdt, \quad (Fubini) \\
&= \int_t x(t)\sqrt{a}\ \overline{\Psi(a\omega)}e^{-j\omega t}dt \\
&= \sqrt{a}\ \overline{\Psi(a\omega)}\int_t x(t)e^{-j\omega t}dt \\
&= \sqrt{a}\ \overline{\Psi(a\omega)}X(\omega)
\end{aligned}$$

[3] **Fubini's Theorem :** If $\int_x[\int_y \mid f(x,y) \mid dy]dx < \infty$, then, $\int\int f(x,y)dxdy = \int_x[\int_y f(x,y)dy]dx = \int_y[\int_x f(x,y)dx]dy$.

Fix $\omega \in \Omega$, where Ω is the support of $X(\omega)$, and let $p(\omega) = \frac{R(\omega)}{X(\omega)}$. Then,

$$\begin{aligned} p(\omega)\frac{s(a)}{\sqrt{a}} &= \overline{\Psi(a\omega)} \\ \int_a \mid p(\omega)\frac{s(a)}{\sqrt{a}} \mid^2 \frac{da}{a} &= \int_a \frac{\mid \Psi(a\omega) \mid^2}{a} da \\ \mid p(\omega) \mid^2 \int_a \mid s(a) \mid^2 \frac{da}{a^2} &= \frac{C_\psi}{2\pi} \\ \mid p(\omega) \mid^2 \parallel s \parallel_S &= \frac{C_\psi}{2\pi} \end{aligned}$$

The above implies that $\mid p(\omega) \mid$ is a constant function for all $\omega \in \Omega$. This results in the condition $\mid X(\omega) \mid = L \mid R(\omega) \mid$, where $L = \sqrt{2\pi C_\psi^{-1} \parallel s \parallel_S}$. Thus,

$$\frac{\mid s(a) \mid}{L\sqrt{a}} = \mid \Psi(a\omega) \mid, \quad \forall\omega \in \Omega, \quad \forall a \tag{2.4}$$

which implies that $\mid \Psi(a\omega) \mid$ is constant for all $\omega \in \Omega$. Consider $\omega_1, \omega_2 \in \Omega$ with $\omega_1 < \omega_2$. From Equation 2.4,

$$\mid \Psi(a\omega_1) \mid = \mid \Psi(a\omega_2) \mid, \quad \forall a \tag{2.5}$$

Let $\alpha = a\omega_1$. Let $\lambda = \frac{\omega_2}{\omega_1} > 1$. Equation 2.5 can then be rewritten as

$$\mid \Psi(\alpha) \mid = \mid \Psi(\lambda\alpha) \mid, \quad \forall\alpha$$

Define a map $T[\Psi](\alpha) = \Psi(\lambda\alpha)$. Then, $\parallel T[\Psi] \parallel = \frac{1}{\lambda} \parallel \Psi \parallel$. Thus, $\parallel T \parallel = \frac{1}{\lambda} < 1$, i.e. there is a strict contraction here. By the Contraction Mapping Theorem [4], the only fixed point of this transformation is $\Psi(\alpha) = 0$ *a.e.* By the Parseval's Identity,

$$\parallel \Psi \parallel_2^2 = 2\pi \parallel \psi \parallel_2^2$$

which implies that $\psi(t) = 0$ *a.e.*, thus providing the contradiction. □

[4] Let $T : X \to X$ be defined on a complete metric space X, with metric d. Let α satisfy $0 < \alpha < 1$, and $d(T(x), T(y)) \leq \alpha d(x, y)$, for all $x, y \in X$. Then, T has a unique fixed point $\hat{x}$.

The theorem thus establishes the need to obtain a separable *approximation* to a given CWT. The corresponding "instantaneous energy distribution" given by $s_{\psi}^{x}(a)$, suitably normalized, is then referred to as a **pseudo power signature**.

Chapter 3

The Matrix SVD Approach

This chapter presents an approach to generate pseudo power signatures for different signal classes using Principal Component Analysis. First, the general solution approach for signals in $L^2(\Re)$ is addressed. Then, it is shown that the technique can be reduced to the Singular Value Decomposition (SVD) of finite dimensional matrices, when applied to finite discrete time signals. The approach is applied to some artificially generated signals, and its merits and limitations are discussed. The methodology is based on a principal component analysis technique, and is derived from the decomposition of the CWT of a signal as a sum of separable terms. This decomposition is the natural extension of the SVD analysis, and effectively determines the closest separable approximation, in the least mean squares sense, to the CWT given by $c_\psi^x(a,b) \in M \subset H$. This analysis is based on the following result ([29]).

Proposition 3.1 *The CWT can always be expressed as*

$$c_\psi^x(a,b) = \sum_i \sigma_i s_i(a) r_i(b)$$

where $s_i(a) \in S = L^2(\Re, C_\psi^{-1}\frac{da}{a^2})$, and $r_i(b) \in R = L^2(\Re, db)$ for each i. The function sets $\{s_i\}_i$, $\{r_i\}_i$ are complete in S and R respectively.

The principal component of c_ψ^x, denoted by $PC[c_\psi^x]$, is given by $\sigma_1 s_1(a) r_1(b)$. The function s_1 can then be used to define the pseudo power signature for the associated signal x.

In theory, the space $H = S \otimes R$, and the functions s_i, and r_i satisfy the coupled integral equations given by

$$\begin{aligned} s_i(a) &= \frac{1}{\sigma_i} \left\langle c_\psi^x, r_i \right\rangle_R \\ r_i(b) &= \frac{1}{\sigma_i} \left\langle c_\psi^x, s_i \right\rangle_S \end{aligned}$$

The pseudo power signature s_1 can then be determined by solving the above coupled equation system. One can easily deduce that the principal component satisfies the following equation :

$$PC[c_\psi^x](a,b) = \sigma_1 s_1(a) r_1(b) = C_\psi^{-1} \int_\alpha \int_\beta c_\psi^x(a,\beta) c_\psi^x(\alpha,b) \frac{1}{\sigma_1} \overline{s_1(\alpha) r_1(\beta)} \frac{d\beta d\alpha}{\alpha^2}$$

It is worthwhile to note that the concept of obtaining pseudo power signatures using the principal component of the SVD is not limited to the CWT. Any TFD can be represented as the sum of separable components of the form shown in Proposition 3.1 ([29]), and hence, it is feasible to obtain pseudo power signatures using any TFD. For the specific problem P, better performance can be obtained using the CWT, owing to its excellent localization capability. However, for classification problems with different requirements, the use of some other TFD to generate the pseudo power signatures might be more appropriate.

3.1 Computation of the SVD_M signature

The last section presented a technique to determine the pseudo power signatures for signals in $L^2(\Re)$. For computational purposes, however, one usually deals with finite discrete time signals. It is thus necessary to determine the nature of the decomposition given in Proposition 3.1 when applied to finite dimensional discrete signal sets.

In essence, Proposition 3.1 provides a decomposition of the CWT function $c_\psi^x(a,b) \in M$ as a sum of separable terms, with the σ_is representing the weights associated with each term. This decomposition is very similar to the more commonly known SVD [1] applied to finite dimensional matrices. If one can reduce the problem of the determination of the pseudo power signatures for finite discrete signal sets, using principal component analysis, to a standard matrix SVD problem, then one can use any one of the existing standard, and efficient, algorithms for the computation of the SVD. From the SVD, one can extract the principal component, and thus determine the pseudo power signature. Since the signature is obtained from the matrix SVD analysis, it is referred to as the SVD_M signature.

In order to reduce the problem to a standard matrix SVD problem, one needs to make a link between a continuous time function in $L^2(\Re)$ whose samples are given by the finite discrete signal under consideration, the CWT of this continuous function and its discrete equivalent, and the relation between the discrete equivalent to the CWT and the discrete signal itself. Effectively, assume that for some $x(t) \in L^2(\Re)$ with CWT given by c_ψ^x, the discrete signal $x(n) \in l^2$ is obtained by sampling $x(t)$. Then, one would need to find a discrete equivalent to c_ψ^x which can be obtained from $x(n)$, and which completely represents the original signal $x(t)$. Finally, one would need to find a way to represent the discrete equivalent to the CWT as a finite

[1] The SVD applied to finite dimensional matrices is defined as follows. Given a matrix $X \in \mathcal{C}^{L\times N}$, the SVD of X is given by

$$X = U\Sigma V^*$$

where $U \in \mathcal{C}^{L\times L}$ and $V \in \mathcal{C}^{N\times N}$ are unitary matrices, and $\Sigma \in \Re^{L\times N}$ is a positive semidefinite diagonal matrix. The diagonal entries of Σ, $\sigma_1 \geq \sigma_2 \ldots \geq 0$, are referred to as the singular values of X. If u_i, v_i denote the ith columns of the unitary matrices U and V respectively, then the *rank one* matrix given by $\sigma_1 u_1 v_1^*$ determines the principal component of X.

dimensional matrix, and determine an efficient way to compute it. This is exactly the motivation of the next section.

3.1.1 Frames and frame operators

The aim is to obtain a suitable *discretization* of the CWT so that it affords no loss of information, and can be obtained from the finite discrete signal under consideration. The discretized CWT coefficients can then be represented as a finite dimensional matrix. The discretization is achieved by the application of the concept of frames and frame operators in Hilbert spaces ([26]).

Definition 3.1 *A family of functions $\{e_j\}_{j\in\mathcal{Z}}$ in a Hilbert space X is called a frame if there exist $A > 0$, $B < \infty$, such that,*

$$A \parallel x \parallel_X^2 \leq \Sigma_{j\in\mathcal{Z}} \mid \langle x, e_j \rangle_X \mid^2 \leq B \parallel x \parallel_X^2, \quad \forall\, x \in X$$

A and B are called the frame bounds.

Definition 3.2 *If $\{e_j\}_{j\in\mathcal{Z}}$ is a frame in a Hilbert space X, then the frame operator $\mathcal{F}$ is a linear operator from X to $l^2(\mathcal{Z}) = \{c = (c_j)_{j\in\mathcal{Z}};\ \parallel c \parallel^2 = \Sigma_{j\in\mathcal{Z}} \mid c_j \mid^2 < \infty\}$, defined as follows :*

$$(\mathcal{F}x) = c_j = \langle x, e_j \rangle_X, \quad \forall\, x \in X$$

The adjoint frame operator $\mathcal{F}^*$ is then given by

$$\mathcal{F}^* c = \Sigma_{j\in\mathcal{Z}} c_j e_j$$

The application to the discretized CWT arises by noting that, under certain conditions ([26]), one can select the mother wavelet $\psi \in L^2(\Re)$, such that with the

discretization $\{a_l\}_l$, $\{b_n\}_n$, the collection of functions $\{\psi_{l,n}\}_{l,n}$ where (from Equation 2.2)

$$\psi_{l,n} = \mid a_l \mid^{-\frac{1}{2}} \psi(\frac{t - b_n}{a_l})$$

constitutes a tight frame ($A = B$). Note that in this case, the CWT operator acts as a frame operator. Moreover, if one can choose a_l, b_n such that the frame bounds $A = B = 1$, and $\| \psi_{l,n} \|_2 = 1$, then, the collection $\{\psi_{l,n}\}_{l,n}$ constitutes an orthonormal basis (ONB) for $L^2(\Re)$ ([26]). For several commonly used wavelet functions, the choice $a_l = 2^l$, $b_n = n2^l$, results in the generation of an ONB for $L^2(\Re)$. In this case, given any $x \in L^2(\Re)$, the discretized set of CWT coefficients $\{c_{l,n} = c_\psi^x(2^l, n2^l)\}_{l,n}$ defined by

$$c_{l,n} = \Gamma[x](l, n) = \langle x, \psi_{l,n} \rangle_2$$

provides a *complete non-redundant representation* of x in the sense that x can be recovered from this discretized set using the adjoint frame operator. Also, by using finitely many of the discretized coefficients, one can approximate x to any arbitrary precision.

One can represent the finitely many discretized CWT coefficients $c_{l,n} = c_\psi^x(2^l, n2^l)$ as a matrix $C = [c_{l,n}]$. The problem with this representation is that the principal component of the SVD of the matrix C is not really separable in time and scale in the true sense. The time, represented by the variable n, in $c_{l,n}$, is dependent on the associated scale, represented by the variable l. In order to obtain a truly separable approximation, one must have complete independence in the time and scale parameters. This independence can be achieved if one considers the discretization

$$\tilde{c}_{l,n} = c_\psi^x(2^l, n)$$

Such a discretization is redundant, providing an overcomplete representation of x. It has been shown by Shensa in [30], that under certain constraints, this redundant discretization also constitutes a frame.

For most practical applications, c_ψ^x has near compact support in the time-frequency plane. For a signal of finite time support, and a suitably chosen ψ, (where ψ has compact time support), it can be well approximated using finitely many discretized CWT coefficient values $\tilde{c}_{l,n}$. This implies that there exists L, N such that $\tilde{c}_{l,n} \approx 0$, for all $l > L$ and for all $n > N$. One can represent this using a finite dimensional matrix $C_\psi^x = [\tilde{c}_{l,n}]$ of dimension $L \times N$. Applying the SVD to this finite matrix C_ψ^x, one obtains

$$C_\psi^x = U\Sigma V^*$$

and hence,

$$C_\psi^x(l,n) = \sum_i \sigma_i u_i(l)\overline{v_i(n)}$$

The principal component is then obtained by extracting the $rank\ \ one$ matrix $\sigma_1 u_1 v_1^*$, where the vectors $u_1,\ v_1$ are truly separable in time n and scale l. It is shown below that, under certain approximations, the unit vector u_1 is the discrete approximation to the pseudo power signature of x.

The simplest discrete approximation to the pseudo power signature would be the vector obtained from its samples. However, for a general measurable function, there is no guarantee that its samples are bounded, and offer a stable reconstruction. In order to ensure boundedness, and guarantee a stable reconstruction, one can define an approximation to the sampling operator as shown. Let $\mathcal{T}$ denote the operator defined as

$$\mathcal{T}[c](l,n) = \frac{2^{2l}}{4\epsilon^2}\int_{2^l-\epsilon}^{2^l+\epsilon}\int_{n-\epsilon}^{n+\epsilon} c(a,b)\frac{dbda}{a^2},\quad c \in H,\quad \epsilon > 0$$

With this definition, observe that for $\epsilon > 0$,

$$\begin{aligned}
\frac{16\epsilon^4}{(2^{2l})^2} \mid \mathcal{T}[c](l,n) \mid^2 &= \mid \int_{2^l-\epsilon}^{2^l+\epsilon} \int_{n-\epsilon}^{n+\epsilon} c(a,b) \frac{dbda}{a^2} \mid^2 \\
&\leq \int_{2^l-\epsilon}^{2^l+\epsilon} \int_{n-\epsilon}^{n+\epsilon} \mid c(a,b) \mid^2 \frac{dbda}{a^2} \\
\Rightarrow \quad 16\epsilon^4 \mid \frac{\mathcal{T}[c](l,n)}{2^{2l}} \mid^2 &\leq \int_{2^l-\epsilon}^{2^l+\epsilon} \int_{n-\epsilon}^{n+\epsilon} \mid c(a,b) \mid^2 \frac{dbda}{a^2} \\
\Rightarrow \quad 16\epsilon^4 \sum_{l,n} \mid \frac{\mathcal{T}[c](l,n)}{2^{2l}} \mid^2 &\leq \sum_{l,n} \int_{2^l-\epsilon}^{2^l+\epsilon} \int_{n-\epsilon}^{n+\epsilon} \mid c(a,b) \mid^2 \frac{dbda}{a^2} \\
\Rightarrow \quad 4\epsilon^2 \left\| \frac{\mathcal{T}[c](l,n)}{2^{2l}} \right\|_2 &\leq C_\psi \left\| c \right\|_H < \infty
\end{aligned}$$

Since $\epsilon > 0$, this result implies that the weighted sequence $\frac{\mathcal{T}[c](l,n)}{2^{2l}} \in l^2(\mathcal{Z}^2)$. Hence $\mathcal{T} : H \rightarrow l^2(\mathcal{Z}^2, \frac{1}{2^{4l}})$, where the weighted Hilbert space $l^2(\mathcal{Z}^2, \frac{1}{2^{4l}})$ is defined as $l^2(\mathcal{Z}^2, \frac{1}{2^{4l}}) = \{x(l,n) : \sum_{l=-\infty}^{\infty} \sum_{n=-\infty}^{\infty} \frac{|x(l,n)|^2}{2^{4l}} < \infty\}$. Moreover, if $c \in H$ is continuous at $(2^l, n)$, and $\epsilon > 0$ is sufficiently small, $\mathcal{T}[c](l,n) \approx c(2^l, n)$. One can thus view the operator $\mathcal{T}$ as an approximation to the sampling operator, and denote $\mathcal{T}[c](l,n) = c(2^l, n)$ for all $c \in H$, for all l, n. [2] Then, from Proposition 3.1, one has

$$\begin{aligned}
\mathcal{T}[c_\psi^x](l,n) &= \mathcal{T}[\sum_i \sigma_i s_i r_i](l,n) \\
&= \sum_i \sigma_i \mathcal{T}[s_i r_i](l,n) \\
&= \sum_i \sigma_i s_i(2^l) r_i(n)
\end{aligned}$$

The pseudo power signature of x is given by the function s_1. The elements of the matrix C_ψ^x are precisely the elements $\mathcal{T}[c_\psi^x](l,n)$. Thus, the discrete vector $s_{d_1} = [\frac{s_1(2^l)}{2^{2l}}]$ of dimension L, can be directly related to the vector u_1 of dimension L obtained

[2] It is to be understood that the actual value of $\mathcal{T}[c](l,n)$ is obtained using the definition of the operator $\mathcal{T}$. The expression $\mathcal{T}[c](l,n) = c(2^l, n)$ is written for convenience of notation, to emphasize that $\mathcal{T}$ is an approximation to the sampling operator.

from the SVD of the matrix C_ψ^x, as $s_{d_1} = u_1$. Hence, the vector u_1 denotes the discrete approximation to the pseudo power signature of x.

The above analysis implicitly assumes that the functions $s_i \in S$, and $r_i \in R$ are piecewise constant. In general, it does not follow that a rank one matrix is always associated with a separable element in H. However, the analysis makes the assumption that the rank one matrix obtained from the samples of s_i and r_i, maps directly to the separable element $\sigma_i s_i r_i \in H$. This mapping is valid if one imposes the condition that the elements s_i, r_i be piecewise constant. In order to see this more clearly, one can define the following maps :

With the operator $\mathcal{T} : H \rightarrow l^2(\mathcal{Z}^2, \frac{1}{2^{4l}})$ defined as before, the adjoint operator $\mathcal{T}^* : l^2(\mathcal{Z}^2, \frac{1}{2^{4l}}) \rightarrow H$ is given by

$$\mathcal{T}^*[h](a,b) = \sum_{l,n} h(l,n) p_l(a) q_n(b), \quad h \in l^2(\mathcal{Z}^2, \frac{1}{2^{4l}})$$

where the functions $p_l \in S$, and $q_n \in R$ are defined as

$$p_l(a) = \begin{cases} 1, & 2^l \leq a < 2^{l+1} \\ 0, & elsewhere \end{cases}$$

$$q_n(b) = \begin{cases} 1, & n \leq b < n+1 \\ 0, & elsewhere \end{cases}$$

Clearly, if $h(l,n) = s(2^l) r(n)$, then $\mathcal{T}^*[h]$ maps to a separable element in H. Note that $\mathcal{T}^*\mathcal{T}[c_\psi^x]$ is an approximation to the CWT function c_ψ^x which is separately

piecewise constant in both variables. Then,

$$\begin{aligned}
\mathcal{T}^*\mathcal{T}[c_\psi^x](a,b) &= \sum_{l,n} \mathcal{T}[c_\psi^x](l,n)p_l(a)q_n(b) \\
&= \sum_{l,n} c_\psi^x(2^l,n)p_l(a)q_n(b) \\
&= \sum_{l,n}\sum_i \sigma_i s_i(2^l)r_i(n)p_l(a)q_n(b) \\
&= \sum_i \sigma_i \sum_l s_i(2^l)p_l(a) \sum_n r_i(n)q_n(b) \\
&= \sum_i \sigma_i \tilde{s}_i(a)\tilde{r}_i(b)
\end{aligned}$$

If one assumes that the functions $s_i \in S$, and $r_i \in R$ are piecewise constant; i.e., $s_i = \tilde{s}_i$, and $r_i = \tilde{r}_i$, for all i, then $\mathcal{T}^*\mathcal{T} = I$, and the vector $s_{d_1} = [\frac{s_1(2^l)}{2^{2l}}]$ of dimension L, suitably normalized, is given by the vector u_1 obtained from the SVD of the matrix $C_\psi^x = [c_\psi^x(2^l,n)]$. Under these assumptions, it follows that, the discrete approximation to PC takes the form

$$PC[c_\psi^x](a,b) = \sigma_1 s_1(a)r_1(b) \approx \sum_l \sum_n c_\psi^x(2^l,b)c_\psi^x(a,n)\frac{1}{\sigma_1}\overline{s_1(2^l)r_1(n)},$$

and, for a given $x \in L^2(\Re)$, the discrete representation $\sigma_1 u_1 v_1^*$ obtained from the SVD of the discretized CWT matrix C_ψ^x, corresponds to the discrete approximation to the principal component $\sigma_1 s_1 r_1$ of c_ψ^x.

3.1.2 Implementation technique

The previous section dealt with the problem of approximating the CWT of $x \in L^2(\Re)$ using a discrete set of coefficients, and obtaining the pseudo power signature using this set. For a practical implementation, one needs to determine techniques to efficiently compute the coefficient values. Also, in practice, one usually deals with sampled (discrete) signal sets $x_d \in l^2(\mathcal{Z})$. In the current literature, there exist several very efficient algorithms ([31]) to compute the discretized CWT coefficients. In each

case, the basic structure used for implementation is the Perfect Reconstruction (PR) filter bank. A brief review of the theory of the general PR filter banks is provided in Appendix A. The theory relating the filter bank structure and the discretized CWT coefficient computation is presented below.

Connection of the CWT to filter banks

Several researchers have examined the relationship between wavelets and filter banks in depth ([33], [34]). The underlying theory behind the computation of the discretized CWT coefficients using the filter bank structure is based on a concept called the **Multi Resolution Analysis (MRA)** ([26]).

Definition 3.3 *A multiresolution analysis consists of a sequence of successive approximation closed subspaces $V_l \subset L^2(\Re)$ which satisfy the conditions*

1. $\ldots V_1 \subset V_0 \subset V_{-1} \ldots$
2. $\overline{\bigcup_{l \in \mathcal{Z}} V_l} = L^2(\Re)$
3. $\bigcap_{l \in \mathcal{Z}} V_l = \{0\}$
4. $x \in V_l \Longleftrightarrow x(2^l) \in V_0$
5. $x \in V_0 \Longrightarrow x(\cdot - n) \in V_0, \quad n \in \mathcal{Z}$

   i.e. V_0 is invariant under integer translations.

6. *There exists $\phi \in V_0$ such that the collection*

   $$\{\phi_{0,n};\ \ n \in \mathcal{Z}\}_n$$

   constitutes an ONB for V_0, where $\phi_{l,n}(t) = 2^{-\frac{l}{2}}\phi(2^{-l}t - n)$.

The basic tenet of the MRA is that whenever the above conditions are satisfied, there exists a collection of wavelets $\{\psi_{l,n}\}_{l,n}$; $l, n \in \mathcal{Z}$, which forms an ONB for $L^2(\Re)$, with $\psi_{l,n}(t) = 2^{-\frac{l}{2}}\phi(2^{-l}t - n)$. Moreover, for every $l \in \mathcal{Z}$, with W_l denoting the orthogonal complement of V_l in V_{l-1}, one has

$$V_{l-1} = V_l \oplus W_l; \quad and \quad W_l \perp W_{l'}, \; l \neq l'$$

The result is that one has an orthogonal decomposition of $L^2(\Re)$ given by

$$L^2(\Re) = \bigoplus_{l \in \mathcal{Z}} W_l$$

Also, the basic defining functions associated with the MRA, $\phi(t) = \phi_{0,0}$ and $\psi(t) = \psi_{0,0}$, called the *scaling* and *mother wavelet* functions respectively, satisfy the following two-scale equations.

$$\begin{aligned} \phi(t) &= \Sigma_n f_n \phi_{-1,n}(t) \\ \psi(t) &= \Sigma_n g_n \phi_{-1,n}(t) \end{aligned}$$

The connection to filter banks arises as a consequence of the following result.

Proposition 3.2 *To every MRA, there corresponds a Perfect Reconstruction (PR) filter bank.*

The general form of the PR filter bank associated with the MRA is a tree-structure, as shown in Figure 3.1 for a tree of 3 levels. Here, the discrete input signal x^0 is successively split into finer approximations (narrower frequency bands) at each level as one progresses down the tree. Each level of the tree contains two channels, and hence is maximally decimated. The filters F_a and G_a are low-pass and high-pass filters respectively. The coefficients $\{f_n\}_n$, and $\{g_n\}_n$, associated with the two-scale equations, determine the low-pass and high-pass filter coefficients of the analysis

section of the filter bank. Hence, $F_a(z) = \Sigma_n f_n z^{-n}$, and $G_a(z) = \Sigma_n g_n z^{-n}$. Let F denote the operation of filtering by F_a followed by downsampling by 2, and F^* its adjoint. Similarly, let G denote the operation of filtering by G_a followed by downsampling by 2, and G^* its adjoint. Then, these operators satisfy the conditions

$$\begin{aligned} F^*F + G^*G &= I \\ FF^* &= I \\ GG^* &= I \\ G^*F &= 0 \end{aligned}$$

The filters in the synthesis section can thus be readily derived using the above conditions. In particular, if the wavelet ψ has compact support, the filters F_a and G_a are FIR filters, and the synthesis filters are given by $F_s(z) = F_a(z^{-1})$, $G_s(z) = G_a(z^{-1})$, and the PR condition simplifies to

$$\begin{aligned} F_a(z)F_a(z^{-1}) + G_a(z)G_a(z^{-1}) &= 2 \\ F_a(-z)F_a(z^{-1}) + G_a(-z)G_a(z^{-1}) &= 0 \end{aligned}$$

The determination of the discretized CWT coefficients using this framework is based on the implicit assumption that the signal $x(t) \in V_0$. The input to the filter bank x^0 is then assumed to be given by $x^0(n) = \langle x, \phi_{0,n} \rangle_2$. With this assumption, the output $y^l(n) = c_\psi^x(2^l, n2^l)$. In practice, one usually deals with a sampled signal $x_d \in l^2$. In this case, the above implementation provides exact discretized CWT coefficients for the signal $\tilde{x} \in L^2(\Re)$ given by $\tilde{x}(t) = \Sigma_n x_d(n)\phi(t-n)$.

For the purpose of computing the SVD_M signatures, one needs a discretization which is separable in time and scale. The above approach, which computes the

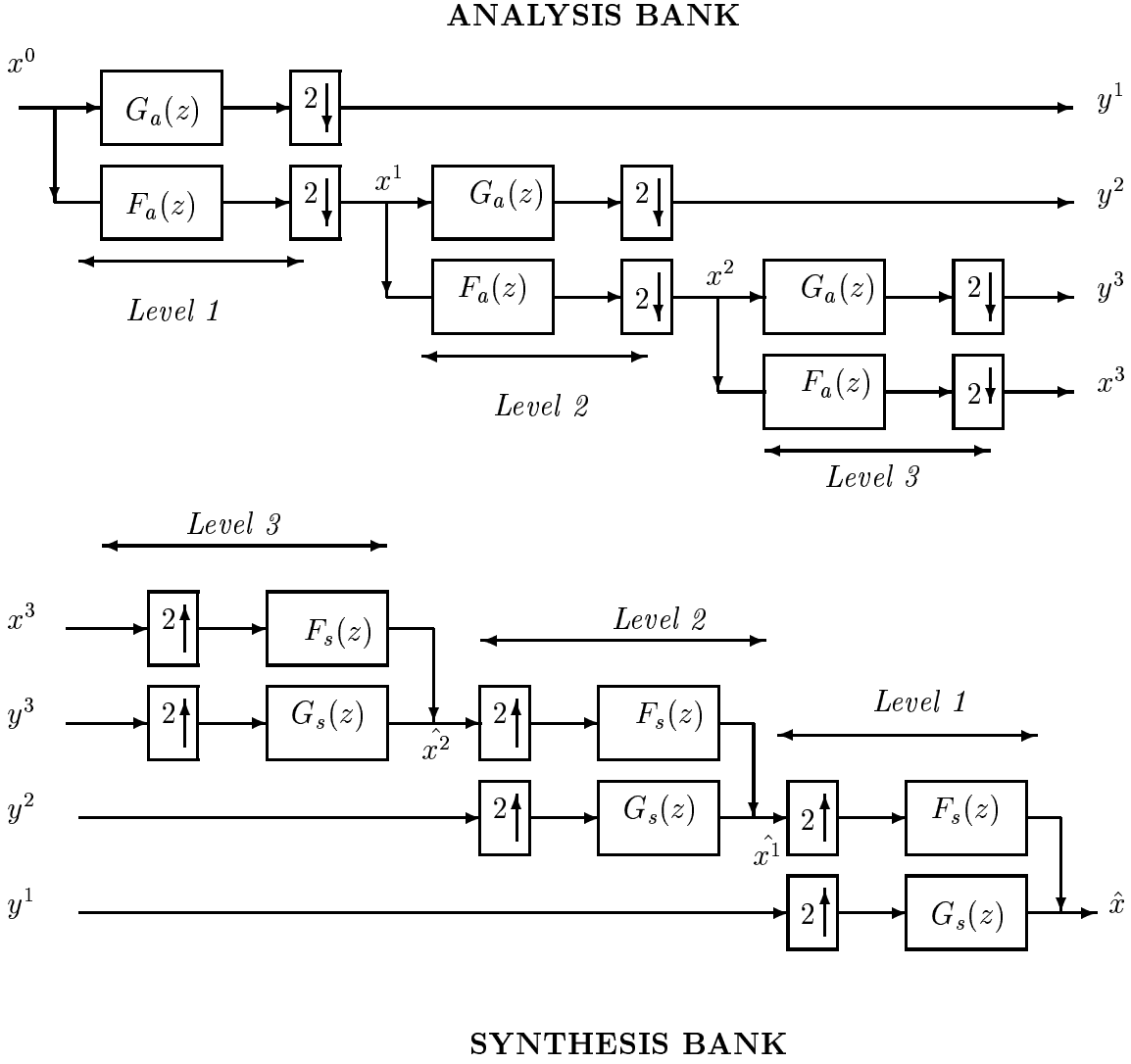

Figure 3.1: Tree-structured filter bank associated with the MRA

CWT coefficients on a dyadic grid in both time and scale, is not separable in the true sense. However, if one computes the CWT coefficients on a grid which is dyadic in scale, but uniform in time, one obtains a separable discretization. The computation of these discretized coefficients can be achieved by using a slight variation of the tree-structured filter bank. It is based on an algorithm referred to as the *Shensa* algorithm ([30]). The Shensa algorithm provides an overcomplete (redundant) representation of x in that it computes the coefficients $c_\psi^x(2^l, n)$; $l, n \in \mathcal{Z}$.

In its basic form, the Shensa algorithm creates a map

$$\mathcal{S} : l^2(\mathcal{Z}) \rightarrow l^2(\mathcal{Z}^2)$$

$\mathcal{S}$ can be implemented using the tree-structured filter bank shown in Figure 3.2. Here, at the lth stage, the analysis filters used are represented by $D^l f$ and $D^l g$ which are obtained by inserting $2^l - 1$ zeros between each pair of filter coefficients in f and g respectively. The decimation at each stage is also done away with in this implementation. In this case, the output $y^l(n) = c_\psi^x(2^l, n)$. Thus, the Shensa algorithm provides a discretization of the CWT which is dyadic in scale, and uniform in time.

3.2 Simulation results

Two sets of computer results are presented in this section. The first result validates the claim that the pseudo power signatures indeed do not depend on the length or the position of data points on the time plane. The second result serves to demonstrate the applicability, and limitations, of the matrix SVD approach to the classification of some artificially generated signals.

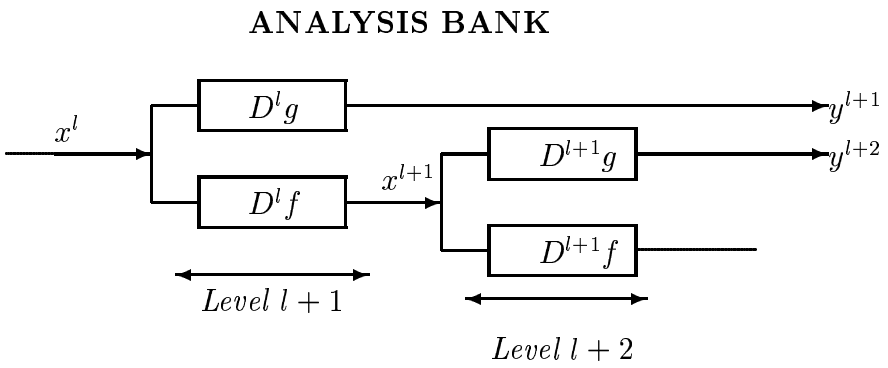

Figure 3.2: Tree-structured filter bank associated with the Shensa Algorithm

For the first case, consider the sample chirp signal shown in Figure 3.3a. This is the Gaussian amplitude modulated chirp signal given by $e^{-.1t^2+j.05t^2+j50t}$. The signature obtained using the matrix SVD analysis is shown in Figure 3.3b (the axis is expressed as a logarithmic function of the scale on a dyadic grid). Figures 3.3c, 3.3e, and 3.3g, show different arbitrarily picked samples of the same chirp signal, varying in length and location on the time plane. Their SVD_M signatures are shown in Figures 3.3d, 3.3f, and 3.3h. These signatures were generated using the $Db4$ wavelet.[3] The Shensa algorithm was used to compute the discretized CWT coefficient matrix. The pseudo power signatures were then readily obtained from the principal component of the SVD of the coefficient matrix. Observe that there is no significant variation in the signature for each sample considered. This test was performed on several different sample signals with similar results. This example is a representative one, used to justify the claim that the concept of using pseudo power signatures to characterize signals independent of time (duration and location) is valid, and applicable to whole classes of nonstationary signals.

[3]This is one of Daubechies' compact support wavelets, and is defined through a two scale equation with 8 coefficients ([26]).

However, there do exist signals which show noticeable variations in their signatures when one considers different sampled data points. A possible explanation for this phenomenon is that these signals are essentially multicomponent, i.e., have several localized disjoint peaks in the time-frequency plane. It is important to note that the entire exercise of representing a signal in a class using one signature pattern is based on the premise that the signal is essentially *monocomponent*. For signals which do not satisfy this premise, and are multicomponent, one would need to extract each component, and apply the above process to it. These signals would then be represented by a *set* of signatures, improving the accuracy of their classification.

For the second example, consider the signals shown in Figure 3.4. These signals are the simple modulated sinc functions $\{x1, x2, x3\}$ given by

$$\begin{aligned} x_1(t) &= e^{j.5\pi t} sinc(\frac{t}{3}) \\ x_2(t) &= e^{j.55\pi t} sinc(\frac{t}{3}) \\ x_3(t) &= e^{j1.55\pi t} sinc(\frac{t}{3}) \end{aligned}$$

Their frequency spectra $\{f1, f2, f3\}$ (the axis is expressed as a fraction of π) and their pseudo power signatures $\{S1, S2, S3\}$ are also shown in the same figure. As before, these signatures were generated using the $Db4$ wavelet. Now consider a signal created by concatenating segments of each signal class: $x1$ over the interval [-125:-50], $x2$ over the interval [-50:50], and $x3$ over the interval [50:115]. The composite signal, its $STFT$, and its discretized CWT are shown in Figure 3.5. Observe that merely examining the signal, its $STFT$, or the CWT is not sufficient to identify either the component signals or the transition points. Furthermore, direct comparison of the $CWTs$ of each signal class with the CWT of the composite signal is also not

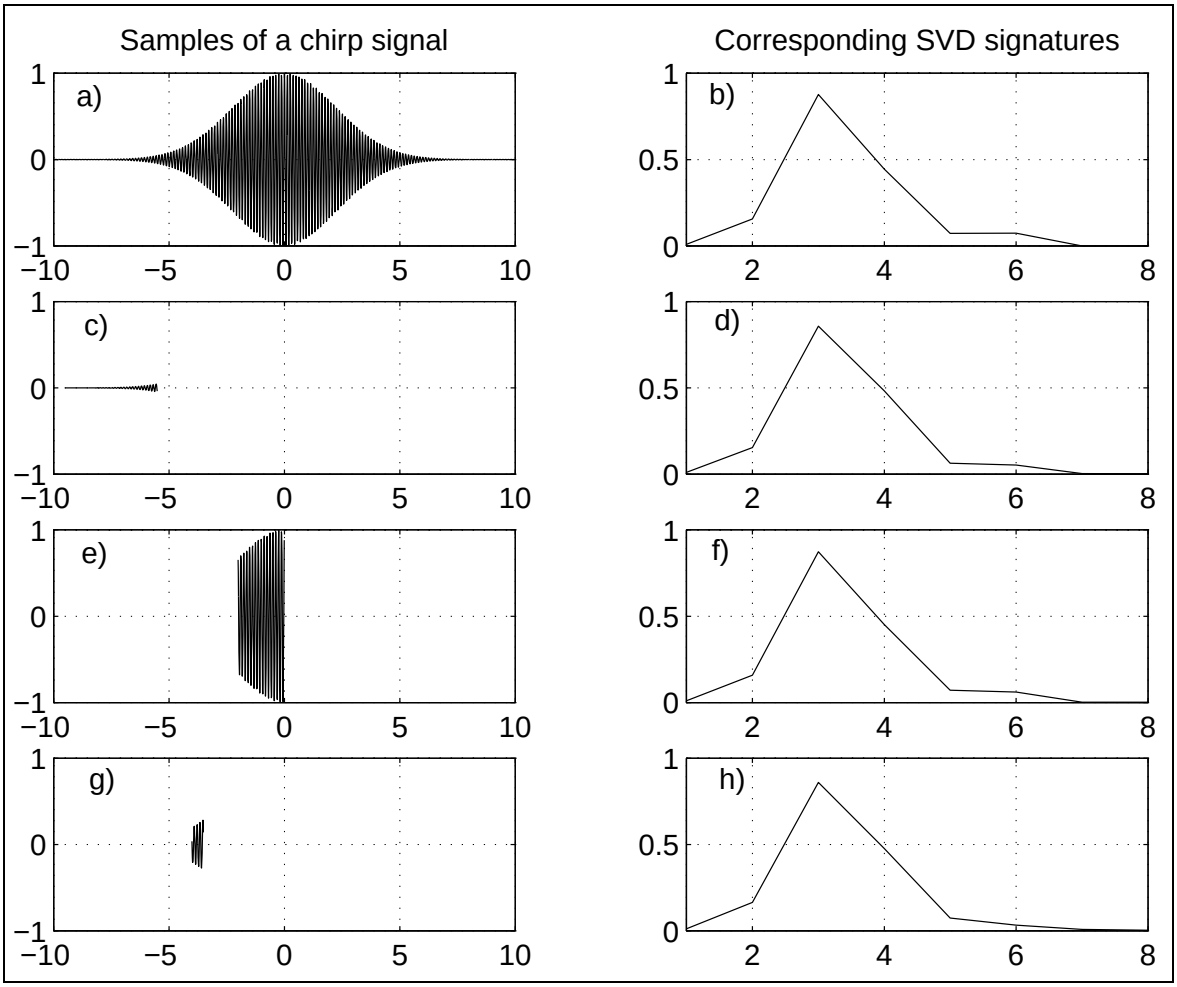

Figure 3.3: Samples of a chirp signal and the corresponding SVD_M signatures

feasible because the CWT support is dependent on the signal duration which is, in general, unknown. For classification purposes, there is a need for a representation which is more intrinsic to each signal class, and is independent of the signal support. These conditions are satisfied by the power signatures shown in Figure 3.4.

The results of the classification using a correlation approach are shown here. Two

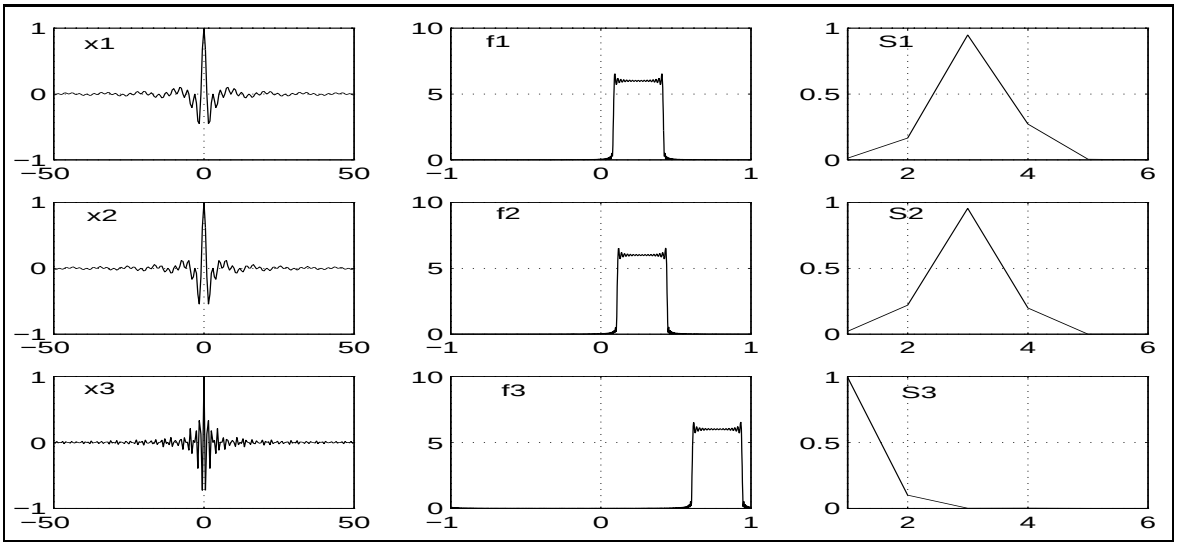

Figure 3.4: The 3 signal classes and their corresponding signatures

assumptions are made in performing this classification.

- All the signal classes are present.
- Only one signal class is present at any given time.

It was established in Chapter 2 that the pseudo power signature Si represents the normalized scale power distribution, and is independent of b. Thus, one can get an accurate picture of the signal composition, with particular reference to the location of the transition points, if one determines the correlation of each Si with the discretized CWT of the composite signal for each b. The results are presented in Figure 3.6. The results show quite clearly that there are 2 transition points in the

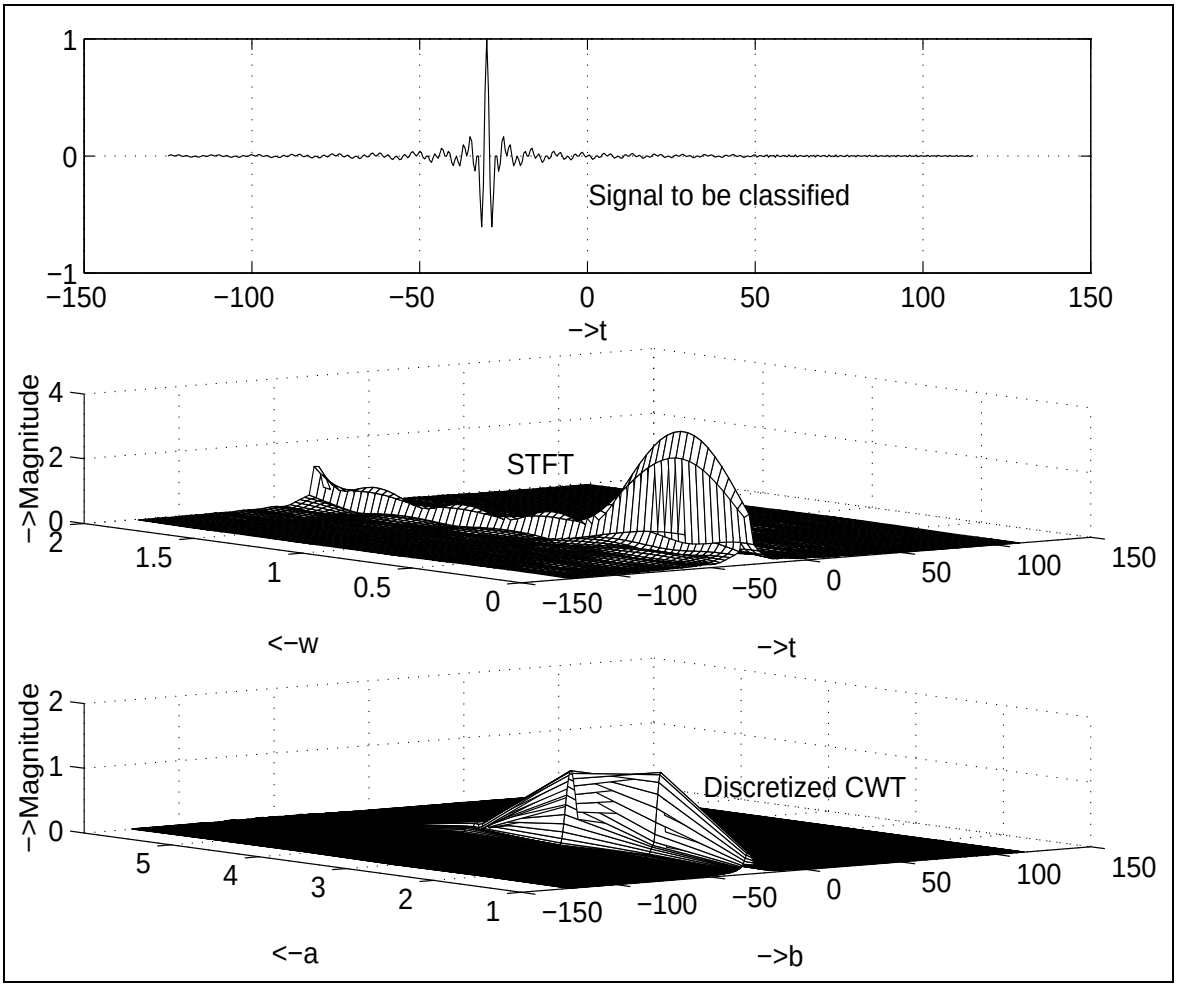

Figure 3.5: The signal, its $STFT$, and its CWT

signal, (the first around -50, and the second around 50), a situation which is not very evident upon examination of the signal. Here, one can make the legitimate assumption that the correlation values must remain fairly constant over a range for the signal to be classified as having support in that range. Hence, one can conclude from the graphs that the support of $x1$ is $[-125 : -50]$, that of $x2$ is $[-50 : 50]$, and that of $x3$ is $[50 : 115]$. Based on the underlying assumptions, the high correlation values of $S1$ in the range $[-50 : 50]$ were disregarded since $S2$ has a higher correlation in that range than $S1$, and is more likely to be present in the range $[-50 : 50]$ than anywhere else.

It is clear from the results presented that the simplistic process of taking the prin-

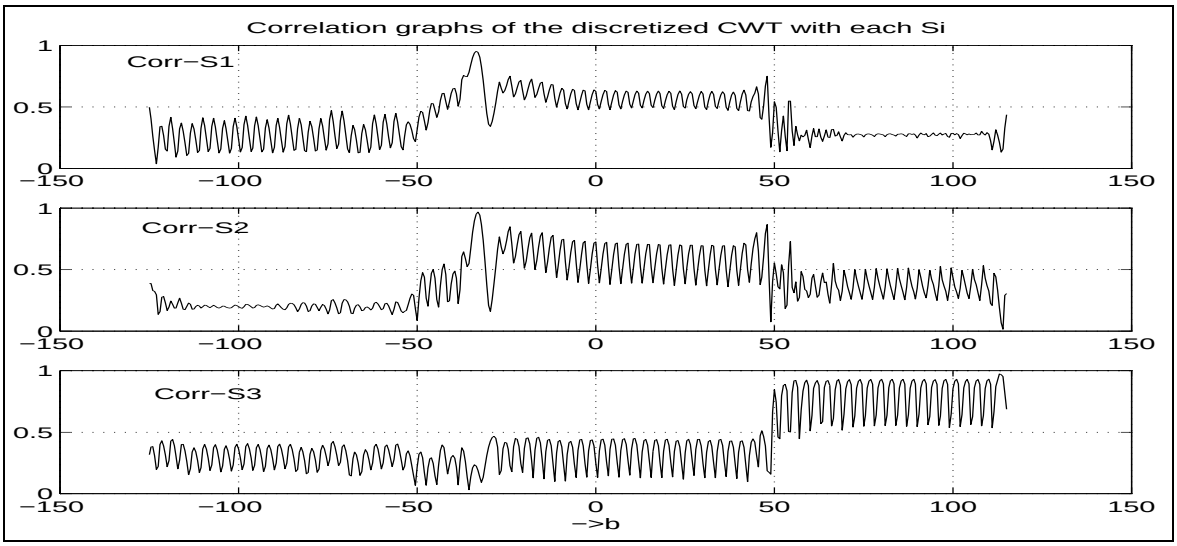

Figure 3.6: Correlation graphs of the discretized CWT

cipal component of the SVD of the discretized c_ψ^x as the pseudo power signature of a signal class can, at times, lead to ambiguous interpretations. The examples presented show that the pseudo power signatures indeed do satisfy the requirement of time independence, and are more discriminating than the Fourier spectra, and more robust than the CWT. However, they lack the ability to capture fine distinctions

between different signal classes, and hence, are not capable of separating signals belonging to two closely spaced classes. This suggests that one needs to determine a more sophisticated technique to find pseudo power signatures with better discriminating capability.

Chapter 4

The Projector Approach

The results shown in the last chapter indicate that the signatures obtained from the principal component of the SVD of the discretized CWT matrix are limited by a lack of fine discriminating capability. In theory, for an element $c_\psi^x \in M$, the principal component of the SVD does, in fact, provide the best separable approximation in the least mean squares sense; i.e., if $s_1 r_1$ denotes the principal component $PC[c_\psi^x]$,[1] of the SVD of c_ψ^x, it satisfies

$$J(s_1, r_1) = \min_{s \in S,\ r \in R} J(s, r), \quad J(s, r) = \left\| c_\psi^x - sr \right\|_H$$

However, the computation of the pseudo power signature using the principal component of the SVD of the discretized coefficient matrix, suffers from a very serious drawback. In addition to the assumptions of piecewise constant forms for the functions s_i, and r_i, it makes the implicit assumption that *the orthogonality of the vectors* $u_i, u_j \in l^2(\mathcal{Z}),\ \ i \neq j$, *implies that the functions* $s_i = \sum_l 2^{2l} u_i(l) p_l(a),\ s_j = \sum_n 2^{2n} u_j(n) p_n(a)$ *are orthogonal in* S. One can readily see that such an assumption is not valid in the space S with its weighted inner product. Specifically, suppose that u_1, u_2 are two orthogonal vectors obtained from the SVD of the discretized CWT matrix. Then $\sum_l u_1(l)\overline{u_2(l)} = 0$. However,

$$\begin{aligned} \langle s_1, s_2 \rangle_S &= C_\psi^{-1} \int_a s_1(a) \overline{s_2(a)} \frac{da}{a^2} \\ &= C_\psi^{-1} \int_a \sum_l 2^{2l} u_1(l) p_l(a) \overline{\sum_n 2^{2n} u_2(n) p_n(a)} \frac{da}{a^2} \end{aligned}$$

[1] The singular value σ_1 has been incorporated into the functions s_1, r_1 for convenience.

$$
\begin{aligned}
&= \sum_l 2^{4l} u_1(l)\overline{u_2(l)} C_\psi^{-1} \int_a p_l(a)\frac{da}{a^2} \\
&= C_\psi^{-1} \sum_l \frac{2^{4l}}{2^{l+1}} u_1(l)\overline{u_2(l)} \\
&\neq 0
\end{aligned}
$$

Note that even if one uses a uniform discretization in scale, the orthogonality assumption is not valid. This occurs because the space S is not the traditional Hilbert space, but is instead, a weighted Hilbert space with the inner product defined as :

$$\langle s_1, s_2 \rangle_S = C_\psi^{-1} \int_a s_1(a)\overline{s_2(a)}\frac{da}{a^2}; \quad \forall \quad s_1, s_2 \in S$$

Consequently, since $H = S \otimes R$, if u_1, u_2, and v_1, v_2, are orthogonal in $l^2(\mathcal{Z})$, it does not follow that the elements $s_1 r_1$, and $s_2 r_2$ in H defined as $s_1 r_1(a,b) = \sum_{l,n} 2^{2l} u_1(l)\overline{v_1(n)} p_l(a) q_n(b)$ and $s_2 r_2(a,b) = \sum_{l,n} 2^{2l} u_2(l)\overline{v_2(n)} p_l(a) q_n(b)$, are orthogonal in H. Hence, given $c_\psi^x \in M$, though $PC[c_\psi^x]$ is such that $\left\langle c_\psi^x, c_\psi^x - PC[c_\psi^x] \right\rangle_H = 0$, this property no longer holds true when one considers the discretization used in Chapter 3, to determine $PC[c_\psi^x]$, resulting in a reduced signature quality. The situation is best described in the Figure 4.1. In the figure, M denotes the closed subspace of CWT functions, and $M^\perp$ its orthogonal subspace in H. The element, $c \in M$, is to be approximated by a separable element of H. In theory, the best separable approximation is provided by the separable element in H that orthogonally projects onto c, and is given by the principal component of c as defined in Proposition 3.1. In the figure, this is represented by the element $sr \in H$. However, the discretized SVD analysis outlined in Chapter 3, yields the element $\tilde{s}\tilde{r} \in H$ as the best separable approximation, where $\tilde{s}\tilde{r} = \sum_{l,n} 2^{2l} u_1(l)\overline{v_1(n)} p_l(a) q_n(b)$. Note however, that in the sense of the weighted inner product defined in H, the element $\tilde{s}\tilde{r}$ does not orthogonally project onto c, but rather onto $\tilde{c} \in M$. If $\| c - \tilde{c} \|_H$ is

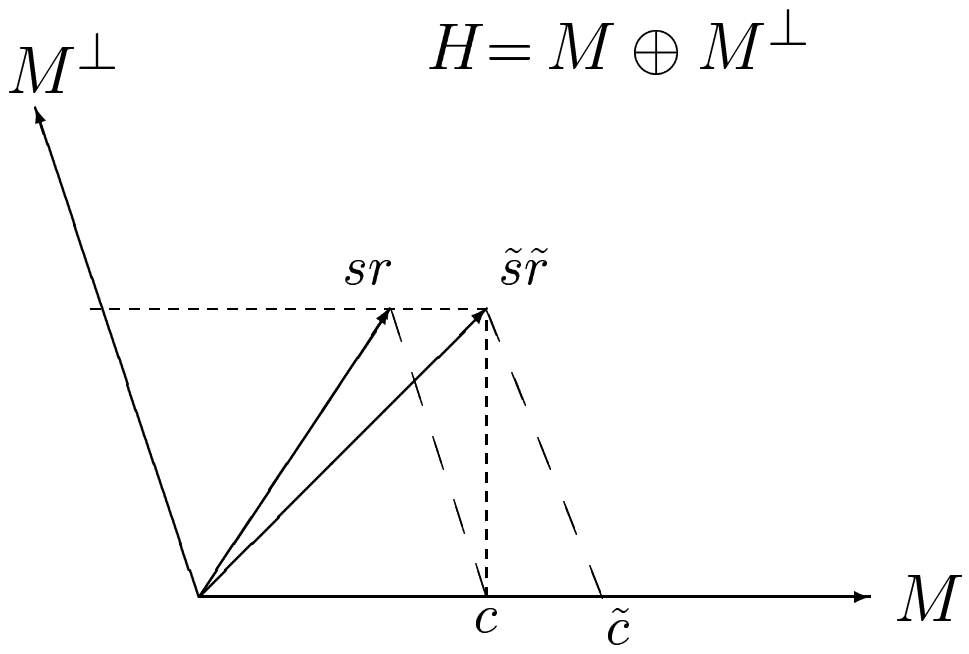

Figure 4.1: Graphical representation of the SVD_M and projection vectors

large, then one can intuitively see that the element $\tilde{s}\tilde{r}$ is a poor approximation to c. What one needs to determine then is the true principal component of c, which is the element $sr \in H$ that orthogonally projects onto c. The normalized function s can then be used to denote the pseudo power signature of the function $x \in L^2(\Re)$ whose CWT is given by c. Since this signature is obtained as a result of a projection, it is referred to as a projection signature. The following sections formulate and solve the problem of directly determining the pseudo power signatures using a suitably defined projection operator.

4.1 Orthogonal projections

The first step in the determination of the projection signature is the definition of a suitable orthogonal projection operator $\mathcal{K} : H \to M$.

Theorem 4.1 *There exists an orthogonal projection operator $\mathcal{K} : H \to M$ defined as follows :*

$$\mathcal{K}[c](a,b) = C_\psi^{-1} \int_\alpha \int_\beta \overline{c_\psi^{\psi_{ab}}(\alpha,\beta)} c(\alpha,\beta) \frac{d\beta d\alpha}{\alpha^2}, \quad \forall\, c \in H$$

PROOF. To prove that $\mathcal{K}$ is an orthogonal projection operator, one needs to show the following.

- $\mathcal{K}$ is well defined, and its range is M.
- $\mathcal{K}$ is a projection, i.e. $\mathcal{K}^2 = \mathcal{K}$.
- $\mathcal{K}$ is self-adjoint, i.e. $\mathcal{K}^* = \mathcal{K}$.

To show that $\mathcal{K}$ is well defined with range M, and is a projection operator.
It is known that M is a reproducing kernel Hilbert space ($r.k.H.s.$) , with the kernel given by $k(a,b;\alpha,\beta) = \overline{c_\psi^{\psi_{ab}}(\alpha,\beta)}$ ([26]). This implies that given any $c \in M$,

$$c(a,b) = C_\psi^{-1} \int_\alpha \int_\beta \overline{c_\psi^{\psi_{ab}}(\alpha,\beta)} c(\alpha,\beta) \frac{d\alpha d\beta}{\alpha^2}$$

Thus, by the definition of $\mathcal{K}$, given any $c \in M$, $\mathcal{K}[c] \in M$. Consider any $c \in H$. Since M is a closed subspace of H, by the projection theorem in Hilbert spaces, $c = c_m + c_{m^\perp}$, where $c_m \in M$, and $c_{m^\perp} \in M^\perp$. Hence,

$$\begin{aligned}
\mathcal{K}[c](a,b) &= \mathcal{K}[c_m](a,b) + \mathcal{K}[c_{m^\perp}](a,b) \\
&= \left\langle c_m, c_\psi^{\psi_{ab}} \right\rangle_H + \left\langle c_{m^\perp}, c_\psi^{\psi_{ab}} \right\rangle_H \\
&= \left\langle c_m, c_\psi^{\psi_{ab}} \right\rangle_H, \quad (Since\ c_\psi^{\psi_{ab}} \in M,\ c_{m^\perp} \in M^\perp,\ \left\langle c_{m^\perp}, c_\psi^{\psi_{ab}} \right\rangle_H = 0) \\
&= c_m(a,b) \quad \in \quad M
\end{aligned}$$

Thus, $\mathcal{K}$ is well defined with range M. Moreover, since M is a $r.k.H.s.$, it directly follows that $\mathcal{K}^2 = \mathcal{K}$. Hence, $\mathcal{K}$ is a projection operator.

To show that $\mathcal{K}$ is self-adjoint, i.e. $\mathcal{K}^* = \mathcal{K}$.

Consider $x, y \in H$.

$$\begin{aligned}
\langle y, \mathcal{K}[x] \rangle_H &= \int_{a,b} y(a,b) \overline{\mathcal{K}[x](a,b)} \frac{dadb}{a^2} \\
&= \int_{a,b} y(a,b) \overline{C_\psi^{-1} \int_{\alpha,\beta} c_\psi^{\psi_{ab}}(\alpha,\beta) x(\alpha,\beta) \frac{d\alpha d\beta}{\alpha^2}} \frac{dadb}{a^2} \\
&= \int_{\alpha,\beta} C_\psi^{-1} \int_{a,b} \overline{c_\psi^{\psi_{\alpha\beta}}(a,b)} y(a,b) \frac{dadb}{a^2} \overline{x(\alpha,\beta)} \frac{d\alpha d\beta}{\alpha^2}; \quad (Fubini) \\
&= \int_{\alpha,\beta} \mathcal{K}[y](\alpha,\beta) \overline{x(\alpha,\beta)} \frac{d\alpha d\beta}{\alpha^2} \\
&= \langle \mathcal{K}[y], x \rangle_H
\end{aligned}$$

Hence, $\mathcal{K}$ is self-adjoint. Since $\mathcal{K}^2 = \mathcal{K}$, and $\mathcal{K}^* = \mathcal{K}$, it follows that $\mathcal{K} : H \to M$ **is an orthogonal projection operator**. □

A result which readily follows from Theorem 4.1 is given by :

Corollary 4.1 *To every $c \in H$, there corresponds one and only one $\hat{x} \in L^2(\Re)$ such that $\Gamma[\hat{x}] = \mathcal{K}[c] \in M$, i.e.*

$$c_\psi^{\hat{x}}(a,b) = \mathcal{K}[c](a,b)$$

For any given $x \in L^2(\Re)$, let $c_\psi^x \in M$ denote its CWT with respect to an admissible $\psi \in L^2(\Re)$. Consider the element $sr \in H$. Let $\hat{c} = \mathcal{K}[sr] \in M$, and $\hat{x} \in L^2(\Re)$ the element associated with sr by Corollary 4.1. It intuitively follows that if one determines $sr \in H$ such that $\| c_\psi^x - \hat{c} \|_M$ is minimized, then one effectively minimizes $\| x - \hat{x} \|_2$. Hence, one can expect that $\hat{c}$, and consequently, sr, will better characterize the intrinsic properties of x. However, it is not known if the orthogonal projection operator $\mathcal{K}$, when restricted to the set of separable elements in H, is one-one. Consequently, there may be more than one separable element $sr \in H$ with the

same projection $\hat{c} \in M$. Thus, in order to ensure the determination of a *unique* projection signature, a regularizing term $\alpha \|sr\|$ is added to the minimization problem. For analysis purposes, $\alpha = 1$. The minimization problem can then be represented as follows :

For a given $c_\psi^x \in M$, find the decomposition $s_\psi^x r_\psi^x \in H$ that minimizes the index

$$J(s_\psi^x, r_\psi^x) = \left\{ \| \ c_\psi^x - \mathcal{K}[s_\psi^x r_\psi^x] \ \|_M^2 + \| \ s_\psi^x r_\psi^x \ \|_H^2 \right\}$$

where $\mathcal{K}$ is the orthogonal projection operator defined earlier.

This is an infinite dimensional nonlinear minimization problem, and requires the solution of the inverse projection problem. The problem formulation and solution procedure for the infinite dimensional case is discussed in Appendix B. However, for a practical application, the problem needs to be reduced to a finite dimensional one, which can then be solved. The problem formulation and the corresponding solution procedure for the finite dimensional case are discussed in the following sections.

4.2 Problem formulation

The first step towards developing a finite dimensional representation for the infinite dimensional minimization problem is to determine a suitable discretization for the elements $c_\psi^x, \mathcal{K}[s_\psi^x r_\psi^x] \in M$, $s_\psi^x \in S$, and $r_\psi^x \in R$. As discussed in Chapter 3, given $c_\psi^x \in M$, one can obtain a discretized equivalent using the concept of frames and frame operators, and a wavelet $\psi \in L^2(\Re)$ of compact support that arises from a MRA. The set of discretized coefficients $\{c_\psi^x(2^l, n)\}_{l,n}$ can then be determined using the Shensa algorithm. As before, one can approximate $\{c_\psi^x(2^l, n)\}_{l,n}$ using finitely many coefficients, and thus obtain a finite dimensional discretized CWT coefficient matrix $C_\psi^x \in \mathcal{C}^{L \times N}$. However, the problem of finding a discrete approximation to

the orthogonal projection operator $\mathcal{K} : H \rightarrow M$, is more involved. The approach followed here approximates $\mathcal{K}$ by using a successive application of the inverse and forward Shensa algorithms, as explained in the next section.

4.2.1 Discrete approximation to the projection operator

In this analysis, it is assumed that the wavelet $\psi \in L^2(\Re)$ arises from a multiresolution. Let $\phi \in V_0 \subset L^2(\Re)$ denote the scaling function associated with the multiresolution (V_0 is one of the spaces of the multiresolution ladder). Note that the collection $\{\phi_{0,n}\}_n$, where $\phi_{0,n} = \phi(t-n)$, constitutes an ONB for V_0. There is an associated implicit assumption that one constrains the computation to the subspace V_0. With these assumptions, one can define a frame operator $F_2 : V_0 \rightarrow l^2(\mathcal{Z})$ as

$$F_2[x](n) = \langle x, \phi_{0,n} \rangle_2 , \quad x \in V_0$$

The adjoint operator $F_2^* : l^2(\mathcal{Z}) \rightarrow V_0$ is then given by

$$F_2^*[z](t) = \sum_n z(n)\phi_{0,n}, \quad z \in l^2(\mathcal{Z})$$

Similarly, using the analysis presented in Chapter 3, one can define a second operator, $\mathcal{T} : H \rightarrow l^2(\mathcal{Z}^2, \frac{1}{2^{4l}})$, which is the approximation to the sampling operator, as

$$\mathcal{T}[c](l,n) = c(2^l, n), \quad c \in H$$

The adjoint operator $\mathcal{T}^* : l^2(\mathcal{Z}^2, \frac{1}{2^{4l}}) \rightarrow H$ is then defined as

$$\mathcal{T}^*[h](a,b) = \sum_l \sum_n h(l,n) q_{l,n}(a,b), \quad h \in l^2(\mathcal{Z}^2, \frac{1}{2^{4l}})$$

where

$$q_{l,n}(a,b) = \begin{cases} 1, & 2^l \le a < 2^{l+1}, \quad n \le b < n+1 \\ 0, & elsewhere \end{cases}$$

Thus, $\mathcal{T}^*\mathcal{T}[c]$ is a piecewise constant approximation to $c \in H$ of the form

$$c(a,b) \approx \mathcal{T}^*\mathcal{T}[c](a,b) = \sum_{l=1}^{L}\sum_{n=1}^{N} c(2^l,n)q_{l,n}(a,b)$$

Let $\mathcal{S} : l^2(\mathcal{Z}) \to l^2(\mathcal{Z}^2)$ denote the forward Shensa operator. Clearly, an element in the range of $\mathcal{S}$ is also in $l^2(Z^2, \frac{1}{2^{4l}})$ (since l assumes only non-negative integer values). Thus, $\mathcal{S} : l^2(\mathcal{Z}) \to l^2(Z^2, \frac{1}{2^{4l}})$. Then, for $x \in V_0$, the map $\mathcal{S}F_2$ defines a matrix C_ψ^x with samples, $c_\psi^x(2^l,n)$, of the wavelet transform of x, i.e. $\mathcal{S}F_2 : V_0 \to l^2(Z^2, \frac{1}{2^{4l}})$ is defined as

$$\mathcal{S}F_2[x](l,n) = \langle x, \psi_{l,n}\rangle_2, \quad x \in V_0$$

The adjoint operator $F_2^*\mathcal{S}^* : l^2(Z^2, \frac{1}{2^{4l}}) \to V_0$ is then obtained as [2]

$$F_2^*\mathcal{S}^*[C](t) = \sum_{l,n} C(l,n)\psi_{l,n}(t)$$

Note that, if C was a matrix of discretized CWT coefficients, $\mathcal{S}^*$ is exactly the inverse Shensa operator. Thus, the adjoint operator $\mathcal{S}^* : l^2(Z^2, \frac{1}{2^{4l}}) \to l^2(\mathcal{Z})$ is

[2] Using the property of adjoints, for $x \in V_0$, $C \in l^2(\mathcal{Z}^2)$,

$$\begin{aligned} \langle \mathcal{S}F_2 x, C\rangle &= \sum_{l,n}\int_t x(t)\overline{\psi_{l,n}(t)C(l,n)} \\ &= \int_t x(t)\overline{\sum_{l,n} C(l,n)\psi_{l,n}(t)} \\ &= \langle x, F_2^*\mathcal{S}^*C\rangle \end{aligned}$$

which implies that $F_2^*\mathcal{S}^*[C](t) = \sum_{l,n} C(l,n)\psi_{l,n}(t)$.

effectively the extended inverse Shensa operator (similar to the infinite dimensional case in Appendix B where $\Gamma^* = \tilde{\Lambda}$). Clearly, $\mathcal{S}^*\mathcal{S} = I$ (analogous to $\Gamma^*\Gamma = I$ as discussed in Appendix B). It is shown below that, in this framework, $\mathcal{S}\mathcal{S}^*$ is the discrete approximation to $\mathcal{K}$ (analogous to the infinite dimensional case where $\Gamma\Gamma^* = \mathcal{K}$).

With the above terminology, the approximation to the operators Γ and Γ^* is given by

$$\begin{aligned} \Gamma &\approx \mathcal{T}^*\mathcal{S}F_2 : V_0 \to H \\ \Gamma^* &\approx F_2^*\mathcal{S}^*\mathcal{T} : H \to V_0 \end{aligned}$$

As shown in Appendix B, the orthogonal projection operator $\mathcal{K} : H \to M$ can be represented as $\mathcal{K} = \Gamma\Gamma^*$. This result can be used to obtain a discrete approximation to $\mathcal{K}$ as given below.

$$\begin{aligned} \mathcal{K} &= \Gamma\Gamma^* \\ &\approx \mathcal{T}^*\mathcal{S}F_2F_2^*\mathcal{S}^*\mathcal{T} \\ &= \mathcal{T}^*\mathcal{S}\mathcal{S}^*\mathcal{T} \end{aligned}$$

It is clear that $\mathcal{T}^*\mathcal{S}\mathcal{S}^*\mathcal{T}$ is not an orthogonal projection, since for any $c \in M$, $\mathcal{T}^*\mathcal{S}\mathcal{S}^*\mathcal{T}c$ only provides a piecewise constant approximation to $c \in M$. However, if one assumes that the CWT is indeed piecewise constant, then $\mathcal{T}^*\mathcal{S}\mathcal{S}^*\mathcal{T}$ can be used as an approximation to the orthogonal projection operator. To see this more clearly, observe that

$$\mathcal{T}^*\mathcal{S}\mathcal{S}^*\mathcal{T}[c](a,b) = \sum_{l',n'} \left\langle \sum_{l,n} c(2^l,n)\psi_{l,n}(t), \psi_{l',n'} \right\rangle_2 q_{l',n'}(a,b), \quad c \in H$$

$$
\begin{aligned}
&= \sum_{l,n} c(2^l, n) \sum_{l',n'} c_{\psi}^{\psi_{l,n}}(2^{l'}, n') q_{l',n'}(a,b) \\
&= \sum_{l,n} c(2^l, n) c_{\psi}^{\psi_{l,n}}(a,b), \quad if\ c_{\psi}^{\psi_{l,n}}\ is\ piecewise\ constant \\
&= c(a,b), \quad if\ c \in M
\end{aligned}
$$

With $s_d(l) = s_{\psi}^x(2^l)$, $s_d \in \mathcal{C}^L$, and $r_d(n) = r_{\psi}^x(n)$, $r_d \in \mathcal{C}^N$, the element $\mathcal{K}[s_{\psi}^x r_{\psi}^x]$ can be approximated by the finite dimensional matrix $\tilde{C} \in \mathcal{C}^{L \times N}$ resulting from the operation $\mathcal{SS}^*[s_d r_d^T]$.

The infinite dimensional minimization can now be formulated as the following finite dimensional problem :

Given a matrix $C_{\psi}^x \in \mathcal{C}^{L \times N}$ of samples on the Shensa grid of the CWT of $x \in L^2(\Re)$, determine the rank one matrix $s_d r_d^T \in \mathcal{C}^{L \times N}$ such that the following functional is minimized

$$
J(s_d, r_d) = \| C_{\psi}^x - \mathcal{SS}^* \left[s_d r_d^T \right] \|_2^2 + \left\| s_d r_d^T \right\|_2^2
$$

4.3 Solution to the minimization problem

This section presents the solution procedure for the finite dimensional minimization problem. The solution technique uses an iterative approach, where, in each iteration, one successively minimizes with respect to the vectors $s_d \in \mathcal{C}^L$, and $r_d \in \mathcal{C}^N$. The basic framework leading to the solution procedure is established below. The entire development is in l^2.

4.3.1 Existence of the minimizer

The first step in developing a solution procedure is to establish the existence of a solution to the finite dimensional minimization problem. The finite dimensional minimization problem has a solution based on the following result.

Theorem 4.2 *There exists* $\bar{s}_d \in B_L = \{s_d \in \mathcal{C}^L; \quad \|s_d\|_2 \leq 1\}$, *and* $\bar{r}_d \in \mathcal{C}^N$ *such that*

$$J(\bar{s}_d, \bar{r}_d) = \inf_{s_d \in B_L, r_d \in \mathcal{C}^N} J(s_d, r_d)$$

PROOF. The proof of the theorem is based on the results established below.

The discrete orthogonal projection approximation operator $\mathcal{S}\mathcal{S}^*$ on the separable finite dimensional Hilbert space $l^2(\mathcal{C}^{L\times N})$, is isometrically equivalent to a square matrix $\tilde{K} : \mathcal{C}^{LN\times LN}$. This isometry, $T : \mathcal{C}^{L\times N} \rightarrow \mathcal{C}^{LN}$, essentially rearranges the elements of an $L \times N$ matrix into a column vector by stacking its rows. Let $T^* : \mathcal{C}^{LN} \rightarrow \mathcal{C}^{L\times N}$ denote the adjoint operator, with the property that $T^*T = I$. Then, $\tilde{K} = T\mathcal{S}\mathcal{S}^*T^*$. Observe that $\tilde{K}$ has the properties that it is Hermitian ($\tilde{K} = \tilde{K}^*$), idempotent ($\tilde{K}^2 = \tilde{K}$), and positive semidefinite ($\tilde{K} \geq 0$). With I_L denoting the identity matrix of size L, and $c = TC_\psi^x$, the functional $J(s_d, r_d)$ can be redefined as

$$\begin{aligned} J(s_d, r_d) &= \| TC_\psi^x - T\mathcal{S}\mathcal{S}^*T^*Ts_d r_d^T \|_2^2 + \left\| Ts_d r_d^T \right\|_2^2 \\ &= \| c - \tilde{K}(r_d \otimes I_L)s_d \|_2^2 + \|(r_d \otimes I_L)s_d\|_2^2 \end{aligned}$$

where $\otimes$ denotes the standard Kronecker product. For a fixed r_d, one can thus define the following subproblem for minimization with respect to s_d :

$$\min_{s_d} J(s_d) = \| c - \tilde{K}(r_d \otimes I_L)s_d \|_2^2 + \|(r_d \otimes I_L)s_d\|_2^2 \tag{4.1}$$

Similarly, denoting $c_T = TC_\psi^{xT}$, and I_N as the identity matrix of size N, one can define the minimization problem with respect to r_d for a fixed s_d as

$$\min_{r_d} J(r_d) = \| c_T - \tilde{K}(s_d \otimes I_N)r_d \|_2^2 + \|(s_d \otimes I_N)r_d\|_2^2 \tag{4.2}$$

Lemma 4.1 *The solutions to the minimization problems defined in Equations 4.1 and 4.2 exist, and are unique.*

PROOF. It can be easily shown that the positive real valued functional $J(\cdot)$ defined in Equations 4.1 and 4.2 is quadratic in the variables s_d and r_d respectively. From Equation 4.1, for a fixed r_d, denote $A_0 = \langle c, c \rangle$, $B = (r_d \otimes I_L)^* c$, and $A = (r_d \otimes I_L)^*(\tilde{K} + I)(r_d \otimes I_L)$. Then, it follows that

$$J(s_d) = A_0 + \langle s_d, As_d \rangle - \langle B, s_d \rangle - \langle s_d, B \rangle$$

which is clearly quadratic in s_d. Moreover, since $A > 0$, the quadratic form is positive definite, and J is convex in s_d [3]. Hence, the minimization problem has a unique solution. A similar result can be shown for r_d. □

From Lemma 4.1, the existence of a unique solution to each of the subproblems defined in Equations 4.1 and 4.2 has been established. Using Calculus of Variations, one can then determine the conditions on the minimizers to each of the subproblems. Since the functional J is separately convex in s_d, and r_d, the first order necessary conditions as determined using Calculus of Variations, become sufficient to determine the minimizers.

For the minimization with respect to s_d, for a fixed r_d, equation 4.1 simplifies to

$$\begin{aligned} J(s_d) &= \left\langle c - \tilde{K}(r_d \otimes I_L)s_d, c - \tilde{K}(r_d \otimes I_L)s_d \right\rangle + \langle (r_d \otimes I_L)s_d, (r_d \otimes I_L)s_d \rangle \\ &= \langle c, c \rangle - \left\langle (r_d \otimes I_L)^* \tilde{K} c, s_d \right\rangle - \left\langle s_d, (r_d \otimes I_L)^* \tilde{K} c \right\rangle \\ &\quad + \left\langle (\tilde{K} + I)(r_d \otimes I_L)s_d, (\tilde{K} + I)(r_d \otimes I_L)s_d \right\rangle \end{aligned}$$

[3] The convexity of J in s_d can be shown from first principles. Consider $J(s_d)$ which is defined on a convex domain. Let $0 < \lambda < 1$, and $s_{d_1}, s_{d_2} \in \mathcal{C}^L$. Then, with $\Delta_s = s_{d_1} - s_{d_2} \neq 0$, and $A = (r_d \otimes I_L)^*(\tilde{K} + I)(r_d \otimes I_L) > 0$, one can readily obtain

$$\begin{aligned} J(\lambda s_{d_1} + (1-\lambda)s_{d_2}) &= \lambda J(s_{d_1}) + (1-\lambda)J(s_{d_2}) - \lambda(1-\lambda)\langle \Delta_s, A\Delta_s \rangle \\ &< \lambda J(s_{d_1}) + (1-\lambda)J(s_{d_2}) \end{aligned}$$

which implies that $J(s_d)$ is strictly convex in s_d. Similarly, one can show that $J(r_d)$ is strictly convex in r_d.

Taking variations with respect to s_d, one obtains

$$\begin{aligned}\delta J_s = &-\left\langle (r_d \otimes I_L)^* \tilde{K} c, \delta s_d \right\rangle - \left\langle \delta s_d, (r_d \otimes I_L)^* \tilde{K} c \right\rangle + \\ &\left\langle (r_d \otimes I_L)^* (\tilde{K} + I)(r_d \otimes I_L) s_d, \delta s_d \right\rangle + \left\langle \delta s_d, (r_d \otimes I_L)^* (\tilde{K} + I)(r_d \otimes I_L) s_d \right\rangle \\ = &\left\langle (r_d \otimes I_L)^* (\tilde{K} + I)(r_d \otimes I_L) s_d - (r_d \otimes I_L)^* \tilde{K} c, \delta s_d \right\rangle + \\ &\left\langle \delta s_d, (r_d \otimes I_L)^* (\tilde{K} + I)(r_d \otimes I_L) s_d - (r_d \otimes I_L)^* \tilde{K} c \right\rangle \end{aligned} \tag{4.3}$$

where δs_d is completely arbitrary. Setting $\delta J_s = 0$, one obtains the necessary condition for minimization with respect to s_d as

$$(r_d \otimes I_L)^* (\tilde{K} + I)(r_d \otimes I_L) s_d - (r_d \otimes I_L)^* c = 0$$

which can be uniquely solved for s_d to yield $\bar{s}_d = Q_{r_d}^{-1} (r_d \otimes I_L)^* c$ where $Q_{r_d} = (r_d \otimes I_L)^* (\tilde{K} + I)(r_d \otimes I_L) > 0$.

Following a similar approach, the necessary condition for the minimization with respect to r_d for a fixed s_d is obtained as

$$(s_d \otimes I_N)^* (\tilde{K} + I)(s_d \otimes I_N) r_d - (s_d \otimes I_N)^* c_T = 0$$

which, in turn, can be uniquely solved for r_d to give $\bar{r}_d = P_{s_d}^{-1} (s_d \otimes I_N)^* c_T$, where $P_{s_d} = (s_d \otimes I_N)^* (\tilde{K} + I)(s_d \otimes I_N) > 0$.

Let $B_L = \{ s_d \in \mathcal{C}^L; \ \ \|s_d\|_2 \leq 1 \}$ denote the closed unit ball in $l^2(\mathcal{Z})$ of dimension L, which is compact. For a fixed s_d of unit norm, $s_d \in B_L$. Let $\bar{r}_d = P_{s_d}^{-1} (s_d \otimes I_N)^* c_T$ be the solution to Equation 4.2, with P_{s_d} defined as before. Then, the functional $J(s_d, \bar{r}_d) = \left\langle c, c - (s_d \otimes I_N) P_{s_d}^{-1} (s_d \otimes I_N)^* c_T \right\rangle$ is effectively a function of $s_d \in B_L$.

Lemma 4.2 *The real valued functional $J(s_d)$ is continuous on the compact set $B_L = \{ s_d \in \mathcal{C}^L; \ \ \|s_d\|_2 \leq 1 \}$.*

PROOF. The real valued functional $J(s_d)$ defined on the compact set B_L is given by

$$J(s_d) = \left\langle c, c - (s_d \otimes I_N) P_{s_d}^{-1} (s_d \otimes I_N)^* c_T \right\rangle, \quad s_d \in B_L$$

Let $\tilde{K} + I = \Lambda^2 > 0$, where Λ is positive definite and self-adjoint. Let $\lambda_{min} > 0$ and $\lambda_{max} > 0$ denote the minimum and maximum eigen values of the positive definite matrix Λ. Then,

$$\lambda_{min}^2 \parallel c \parallel^2 \leq \langle \Lambda c, \Lambda c \rangle \leq \lambda_{max}^2 \parallel c \parallel^2; \ \forall c \in \mathcal{C}^{LN} \tag{4.4}$$

Defining $X(s_d) = \Lambda s_d \otimes I_N$, $P_{s_d} = X^*(s_d) X(s_d)$. Using Equation 4.4, one can readily establish the following results.

$$\lambda_{min} \parallel s_d \parallel \leq \parallel X(s_d) \parallel \leq \lambda_{max} \parallel s_d \parallel \tag{4.5}$$

$$\lambda_{min}^2 \parallel r_d \parallel^2 \leq \langle r_d, P_{s_d} r_d \rangle \leq \lambda_{max}^2 \parallel r_d \parallel^2; \ \forall r_d \in \mathcal{C}^N; s_d \in B_L \tag{4.6}$$

$$\lambda_{max}^{-2} \parallel r_d \parallel^2 \leq \left\langle r_d, P_{s_d}^{-1} r_d \right\rangle \leq \lambda_{min}^{-2} \parallel r_d \parallel^2; \ \forall r_d \in \mathcal{C}^N; s_d \in B_L \tag{4.7}$$

For $s_{d_1}, s_{d_2} \in B_L$, let $\delta_s = s_{d_1} - s_{d_2}$. Then,

$$\|\delta_s\| \leq \|s_{d_1}\| + \|s_{d_2}\| = 2 \tag{4.8}$$

Now,

$$P_{s_{d_1}} - P_{s_{d_2}} = X^*(s_{d_2}) X(\delta_s) + X^*(\delta_s) X(s_{d_2}) + X^*(\delta_s) X(\delta_s)$$

and hence, from Equations 4.5 and 4.8,

$$\left\| P_{s_{d_1}} - P_{s_{d_2}} \right\| \leq 4 \lambda_{max}^2 \parallel \delta_s \parallel \tag{4.9}$$

Also,

$$P_{s_{d_2}}^{-1} - P_{s_{d_1}}^{-1} = P_{s_{d_2}}^{-1} \left(P_{s_{d_1}} - P_{s_{d_2}} \right) P_{s_{d_1}}^{-1}$$

which implies from Equations 4.7 and 4.9

$$\| P_{s_{d_2}}^{-1} - P_{s_{d_1}}^{-1} \| \leq 4\lambda_{min}^{-4}\lambda_{max}^{2} \| \delta_s \|; \ \forall s_{d_1}, s_{d_2} \in B_L$$

From the above results, observe that one can express

$$J(s_d) = \langle c, c \rangle - \frac{1}{4} \left\langle X^*(s_d)\Lambda c, P_{s_d}^{-1} X^*(s_d)\Lambda c_T \right\rangle$$

Denoting $\Delta J = J(s_{d_2}) - J(s_{d_1})$, one can then obtain

$$\begin{aligned} | \Delta J | &= \frac{1}{4}[\left\langle X^*(s_{d_1})\Lambda c, (P_{s_{d_1}}^{-1} - P_{s_{d_2}}^{-1}) X^*(s_{d_1})\Lambda c_T \right\rangle + \\ &\quad \left\langle X^*(s_{d_2})\Lambda c, P_{s_{d_2}}^{-1} X^*(\delta_s)\Lambda c_T \right\rangle + \left\langle X^*(\delta_s)\Lambda c, P_{s_{d_2}}^{-1} X^*(s_{d_1})\Lambda c_T \right\rangle] \\ &\leq 4\lambda_{min}^{-4}\lambda_{max}^{2} \|c\|^2 \| \delta_s \| + \lambda_{min}^{-2} \|c\|^2 \|\delta_s\| + \lambda_{min}^{-2} \|c\|^2 \|\delta_s\| \\ &\leq A \|\delta_s\| \end{aligned}$$

where A is a constant and depends only on $\| c \|$, λ_{min}, and λ_{max}. Then, given any $\epsilon > 0$, there exists $0 < \delta = \frac{\epsilon}{2A}$, such that for all $\| s_{d_1} - s_{d_2} \|_2 < \delta$, $| J(s_{d_2}) - J(s_{d_1}) | < \epsilon$. Thus, J is continuous on B_L. □

The existence of the minimizing solution to the finite dimensional problem is established as follows :

Fix $s_d \in B_L$. Then, by Lemma 4.1, there exists $r_d(s_d) \in \mathcal{C}^N$ such that $J(s_d, r_d(s_d)) = J(s_d) = \inf_{r_d \in \mathcal{C}^N} J(s_d, r_d)$. The real valued functional J is thus defined on B_L which is compact. From Lemma 4.2, J is continuous on the compact set. Hence, it is guaranteed to attain its maximum and minimum on the set, i.e. there exists $\bar{s}_d \in B_L$, such that $J(\bar{s}_d) = \inf_{s_d \in B_L} J(s_d)$, i.e.

$$J(\bar{s}_d, r_d(\bar{s}_d)) = \inf_{s_d \in B_L} J(s_d, r_d(s_d)) = \inf_{s_d \in B_L, r_d \in \mathcal{C}^N} J(s_d, r_d)$$

Denoting $\bar{r}_d = r_d(\bar{s}_d)$, the existence of the minimum is established. □

4.3.2 Iterative procedure

Once the existence of the minimizer has been established, one can develop a procedure for its determination. The procedure followed here is an iterative one, and requires successively solving the necessary conditions given by the equations

$$(r_d \otimes I_L)^*(\tilde{K} + I)(r_d \otimes I_L)s_d - (r_d \otimes I_L)^* c = 0 \tag{4.10}$$

$$(s_d \otimes I_N)^*(\tilde{K} + I)(s_d \otimes I_N)r_d - (s_d \otimes I_N)^* c_T = 0 \tag{4.11}$$

in each iteration. Note that these are only necessary conditions for the minimizing solution, and are not sufficient to guarantee a minimum. The approach is based on the result presented in Lemma 4.1. Effectively, the result ensures that the iterative approach produces a monotonically decreasing cost function sequence. The iterative approach developed for successive minimization with respect to s_d, and r_d is given below.

- Specify a tolerance value *tol* for termination.
- Using randomly picked vectors $s_d^0 \in \mathcal{C}^L$, and $r_d^0 \in \mathcal{C}^N$, compute the cost $J^0 = J(s_d^0, r_d^0)$.
- At the ith stage of the iterative process, with $r_d = r_d^{i-1}$, solve Eq. 4.10 to obtain $\tilde{s}_d^i$. Thus, $\tilde{s}_d^i = Q_{r_d{}^{i-1}}^{-1}(r_d^{i-1} \otimes I_L)^* c$.
- Set $s_d^i = \frac{\tilde{s}_d^i}{\|\tilde{s}_d^i\|_2}$. This constrains s_d^i to the unit ball B_L in the finite dimensional Hilbert space $l^2(\mathcal{Z})$ of dimension L.
- With $s_d = s_d^i$, solve Eq. 4.11 to obtain $r_d^i = P_{s_d^i}^{-1}(s_d^i \otimes I_N)^* c_T$.

- Compute $J^i = J(s_d^i, r_d^i) = \left\langle c, c - (s_d^i \otimes I_N) P_{s_d^i}^{-1} (s_d^i \otimes I_N)^* c_T \right\rangle$. Observe that this is a function of s_d^i only, i.e. $J^i = J(s_d^i)$. Moreover, it is independent of the norm of s_d^i.

- If $J^{i-1} - J^i \leq tol$, terminate.

4.3.3 Convergence to the optimal

From Lemma 4.1, one can immediately see that the iterative procedure outlined produces a monotonically decreasing sequence of costs $\{J^i\}_i$, since one has the result $J(s_d^0, r_d^0) \geq J(\tilde{s}_d^1, r_d^0) \geq \cdots \geq J(s_d^i, r_d^i) \geq \cdots$, from which one can extract the monotonically decreasing sequence

$$\{J^i\}_i; \quad J^i = J(s_d^i) \tag{4.12}$$

defined on the unit ball B_L which is compact. The convergence of the iterative procedure is established based on the following result.

Theorem 4.3 *There exists $\bar{s}_d \in B_L$, and $\bar{J} \geq 0$ such that the sequence $\{J^i\}_i$ defined in Equation 4.12 converges to $\bar{J} = J(\bar{s}_d)$.*

Proof. By Lemma 4.2, the real valued functional $J(s_d)$ is continuous on the compact set B_L. Hence, it is guaranteed to attain its maximum and minimum on B_L. The iterative procedure produces a monotonically decreasing sequence of costs $\{J^i\}_i$, whose limit $\bar{J}$ exists by Lemma 4.2. Hence, $\lim J^i = \bar{J}$. Now, B_L is also sequentially compact. This implies that from the sequence $\{s_d^i\}_i$, one can extract a subsequence $\{s_d^{i_j}\}_{i_j}$ that converges to some $\bar{s}_d \in B_L$, and by the continuity of J,

$$J(\bar{s}_d) = \lim J(s_d^{i_j}) = \lim J^{i_j} = \bar{J}$$

establishing the convergence of the algorithm to the optimal solution. □

An iterative procedure using successive minimization with respect to s_d, and r_d, which converges to the limiting solution whose existence is guaranteed, has thus been developed. It is important to note that, while the procedure attains a minimum for the functional J, it offers no guarantee that the minimum attained is global. This is a general problem in nonlinear minimization techniques. The sufficient condition to ensure that the minimum attained is indeed global is that the functional J be *jointly* convex in the variables s_d and r_d, which is not the case in this problem.

Chapter 5

Simulation Results Using Projection Signatures

The last chapter outlined a solution procedure for the determination of the discrete projection signatures for signal classes. The procedure adopted an iterative approach, involving successive minimization with respect to the variables s_d, and r_d, which were defined in the chapter. In this chapter, the computational algorithm developed based on the procedure is presented, along with some simulation results. These results serve to illustrate the potential capabilities of the projection signatures, and also the limitations of the computational procedure used to determine these signatures.

5.1 Computational algorithm

The computational algorithm used to generate the projection signatures is given below.

1. Select a wavelet $\psi \in L^2(\Re)$ which arises from a MRA, and the number of levels L to be used in the filter bank corresponding to the MRA. Denote the analysis low pass filters as $f \in \mathcal{C}^{N_\psi}$, and the analysis high pass filters as $g \in \mathcal{C}^{N_\psi}$.

2. For the given finite discrete input signal $x \in \mathcal{C}^{N_x}$, determine the discretized CWT coefficient matrix $C_\psi^x \in {}^{L \times N}$ using the forward Shensa algorithm. Note that $N = N_x + (2^L - 1)(N_\psi - 1)$.

3. Based on the scalogram $SC^x_\psi(l,n) = \mid C^x_\psi(l,n) \mid^2$ obtained, modify the value of L such that $C^x_\psi(l,n) \approx 0$, for all $l \geq L$, $n \geq N$. Recompute C^x_ψ using the modified value of L.

4. Pick random vectors $s^0_d \in \mathcal{C}^L$, $r^0_d \in \mathcal{C}^N$, and set a value for tol.

   - At the i th stage, set $r_d = r^{i-1}_d$. Using the conjugate gradient technique ([32]), with gradient given by [1]

     $$\Delta_{s_d} = \Re\{(s_d {r^{i-1}_d}^T + \mathcal{S}\mathcal{S}^* \left[s_d {r^{i-1}_d}^T\right] - C^x_\psi) {r^{i-1}_d}^c\}$$

     solve the minimization problem for s_d. Let $\tilde{s_d}^i$ denote the solution.

   - Set $s^i_d = \frac{\tilde{s_d}^i}{\|\tilde{s_d}^i\|}$.

   - Next, with $s_d = s^i_d$, using the conjugate gradient technique with gradient

     $$\Delta_{r_d} = \Re\{(s^i_d {r_d}^T + \mathcal{S}\mathcal{S}^* \left[s^i_d {r_d}^T\right] - C^x_\psi)^T {s^i_d}^c\}$$

     solve the minimization problem for r_d . Let r^i_d denote the solution.

   - Compute the cost function $J(s^i_d, r^i_d)$. If

     $(J(s^{i-1}_d, r^{i-1}_d) - J(s^i_d, r^i_d)) < tol$, terminate.

5. end.

5.1.1 Computational complexity of the algorithm

The minimization problem under consideration is separately quadratic in s_d, and r_d. Consequently, the use of the conjugate gradient technique for each minimization guarantees convergence for each minimization in $\mathcal{O}(V)$ steps, where V is the size

[1] The gradient can be readily obtained from Equation 4.3 by using the conjugate symmetry property of the inner product in a complex Hilbert space X, i.e. for $x, y \in X$, $\langle x, y\rangle = \overline{\langle y, x\rangle}$. Then, by applying the destacking operator T^*, one can obtain the gradient as shown.

of the vector over which one is minimizing. It is clear that the computational cost associated with the technique depends on the number of iterations I required to reach the optimal solution, which in turn, is a function of tol, and the initial condition. The complexity of each iteration is largely a factor of the complexity of the conjugate gradient technique, and the complexity of the implementation of $\mathcal{S}$ and $\mathcal{S}^*$. The complexity of the conjugate gradient technique is $(6V + 2)$ multiplications and $(6V-3)$ additions per iteration. For a practical implementation, especially when V is large (like, for example, when one is minimizing with respect to r_d), one usually prescribes a termination criteria so that convergence occurs in far fewer than V steps ([32]). The complexity of $\mathcal{S}$ is $2LN_\psi N_x$ multiplications and $2L(N_\psi - 1)N_x$ additions, and that of $\mathcal{S}^*$ is exactly the same ([31]). Thus, the complexity of the Shensa algorithm is a linear function of N_x, L, and N_ψ. If one assumes an average of T_s steps for convergence of the conjugate gradient technique for the minimization with respect to s_d, and T_r steps for the minimization with respect to r_d, the overall cost of the projection algorithm is given by $I[(6L+2+4LN_\psi N_x)T_s+(6N+2+4LN_\psi N_x)T_r]$ multiplications and $I[(6L - 3 + 4L(N_\psi - 1)N_x)T_s + (6N - 3 + 4L(N_\psi - 1)N_x)T_r]$ additions. The overall cost thus depends on the choice of the initial vector r_d^0, since a point closer to the minimum would require fewer iterations to converge to the optimal. However, the choice of s_d^0 does not affect the overall computational cost significantly.

5.2 Simulation results

This section presents the results of applying the iterative solution technique to different nonstationary signals. A summary of the simulation experiments and the corresponding results is given in Table 5.1.

Table 5.1: Simulation results on the projection signatures

EXPERIMENT	ANALYSIS
Signature quality	Highly distinct
Classification	Unambiguous, with sharp transition points
Reliability	Good
Robustness	Not very robust in the presence of much noise
Sensitivity to initial condition	Not affected by perturbations in initial condition, but sensitive to choice of initial condition
Analyzing wavelet	No significant difference in quality
Computational requirements	Very reasonable with fast convergence

5.2.1 Signature quality and application to classification

The first experimental result presented is the classification problem discussed in Chapter 3. The signals used are shown in Figure 5.1, and are the same three signals shown earlier in Figure 3.4. The projection signatures obtained using the $Db4$ wavelet, for $L = 6$, and $r_d^0 = \chi_x$ [2], are also shown in Figure 5.1. Observe how the projection signatures clearly separate the highly correlated signals x_1 and x_2. These signatures were then applied to the classification problem described in Figure 3.5. The composite signal and the correlation graphs of the projection signatures with the discretized CWT of the composite signal are shown in Figure 5.2. From the

[2] $\chi_x \in \mathcal{C}^N$ is the characteristic function of $x \in \mathcal{C}^{N_x}$ defined as

$$\chi_x(n) = \begin{cases} 1, & 1 \le n \le N_x \\ 0, & N_x \le n \le N \end{cases}$$

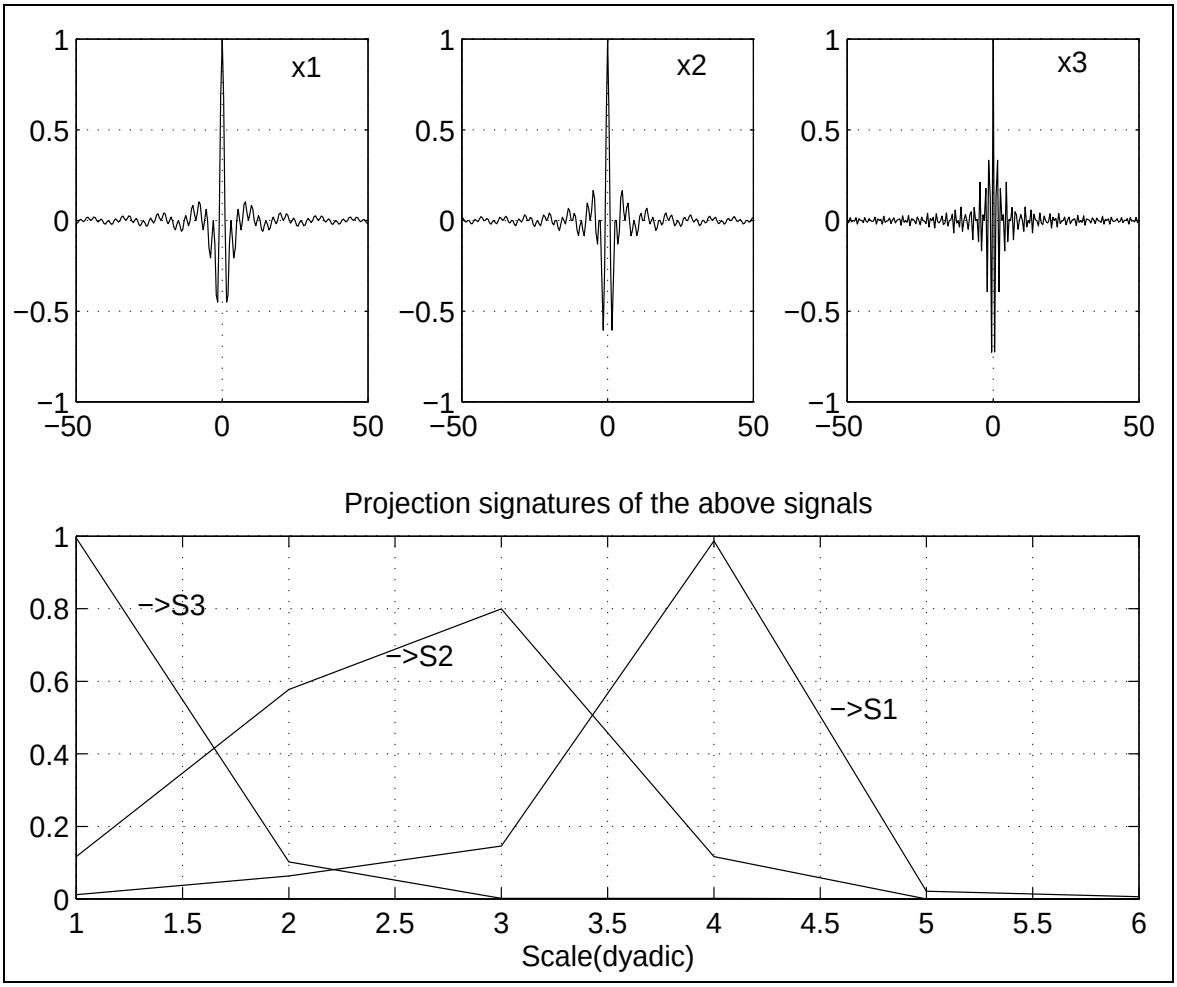

Figure 5.1: The 3 signal classes, and their projection signatures

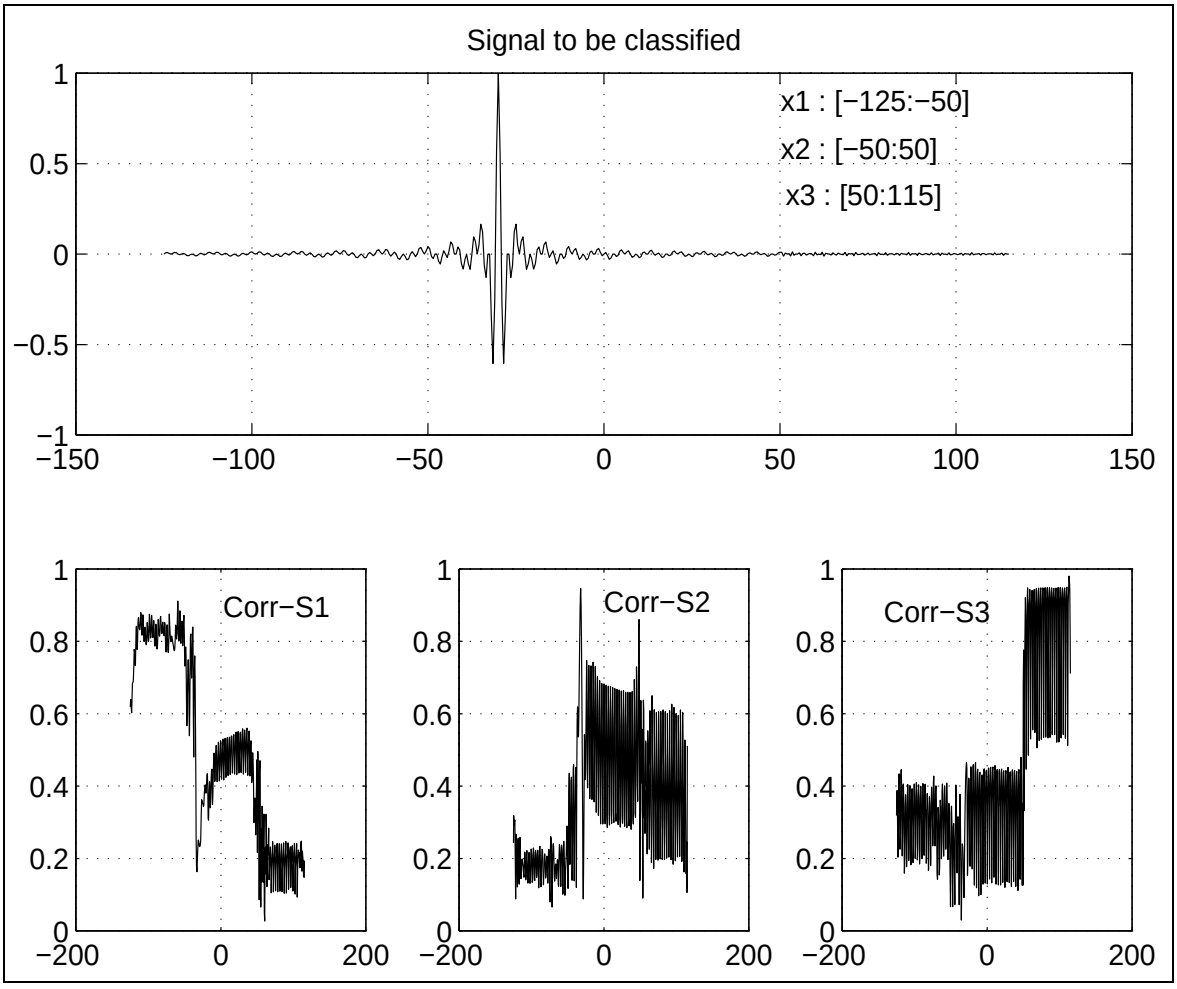

Figure 5.2: The projection signatures applied to the classification problem

correlation graphs, one can conclude with a high degree of confidence that the signal x_1 is present in the segment $[-125 : -50]$, x_2 is present in the segment $[-50 : 50]$, and the signal x_3 is present in the segment $[50 : 115]$. Notice, when compared with Figure 3.6, the high correlation values obtained using the projection signatures, and the clear demarcation of the transition points. The high correlation values are of great importance when one needs to classify a signal where *it is not known a priori if all the events are present.*

5.2.2 Reliability analysis

For use in classification applications, it is extremely important that the projection signatures be reliable measures of the classes they represent. By this, one implies that for samples of different signals in the same signal class, the projection signatures should not show significant variations. Consider a common example of a speech signal, say for example, the letter "A" as spoken by a person on several different occasions. It is reasonable to expect that slight variations will exist between every utterance of the letter by the same person. However, a reliable signature for the signal class (here, the signal class comprises of every sample of the letter "A" spoken by the same person) should be fairly insensitive to these variations. Figure 5.3 presents the results of the reliability experiment using this example. In the figure, Sample #1 represents the projection signatures for 5 different samples of the letter "A" as spoken by a person on a given day, and Sample #2, the projection signatures for samples obtained on a different day. Observe the very high correlations between all the different signatures (the correlation coefficient values ranged between $.9705 - .9950$). The wavelet used was $Db4$ with $L = 8$. As before, in every case, r^0 was taken to be the characteristic function of the sample signal.

The experiment described above is significant for several reasons. The reliability results obtained from the experiment reflect those obtained from several other similar experiments. First, the experiment shows that the projection signatures are true

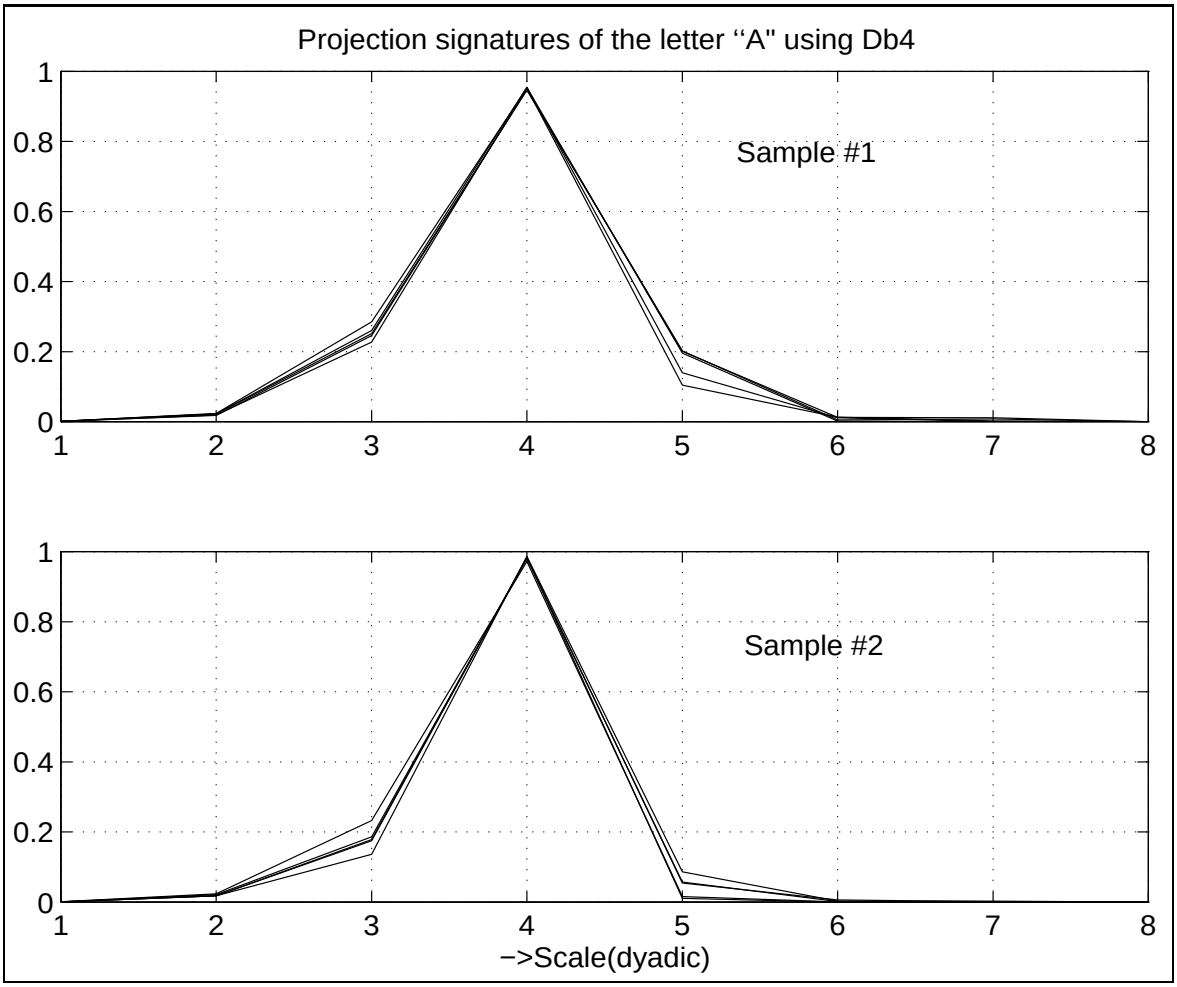

Figure 5.3: Reliability test results for the projection signatures

measures of the signal class they characterize, which is an essential requirement of any representation. Next, it shows that the pseudo power signature technique can be applied to real data signals which may have some background noise, slight anomalies and random variations, even though every signal essentially belongs to the same class. Finally, it reiterates the claim that the signatures are independent

of signal duration, since every signal sample considered in the example was of a duration different from the others.

5.2.3 Robustness analysis

Owing to the assumption of a piecewise constant form for the CWT of the signal in the projection algorithm, one would expect that the algorithm would perform poorly for signals that do not satisfy the premise. However, experiments on a wide variety of signals show surprisingly good results. A possible explanation for this phenomenon is that since the CWT preserves the regularity of the associated signal, a signal that is fairly smooth, and not highly corrupted by noise, is likely to satisfy the premise. When dealing with noisy signals, it might be necessary to prefilter the signal to remove some of the noise before applying the projection technique. A comparative performance of the robustness of the SVD_M and projection signatures in the presence of white Gaussian noise is shown in Figures 5.4 and 5.5. It is important to note that while the matrix SVD approach does not generate very discriminating signatures owing to the invalid orthogonality assumption, this very assumption makes the signatures more robust. The reason for this is that, the incorrect assumption makes for a rather poor representation, which is not very sensitive to the variations occurring in the signal owing to the noise. The signatures computed using the projection algorithm outlined are not as robust, because they better characterize the signal, and hence, are more sensitive to variations in the signal. Also, since the projection algorithm was not guaranteed to attain the global minimum, it is possible that different local minima are attained when the signal is highly corrupted by noise.

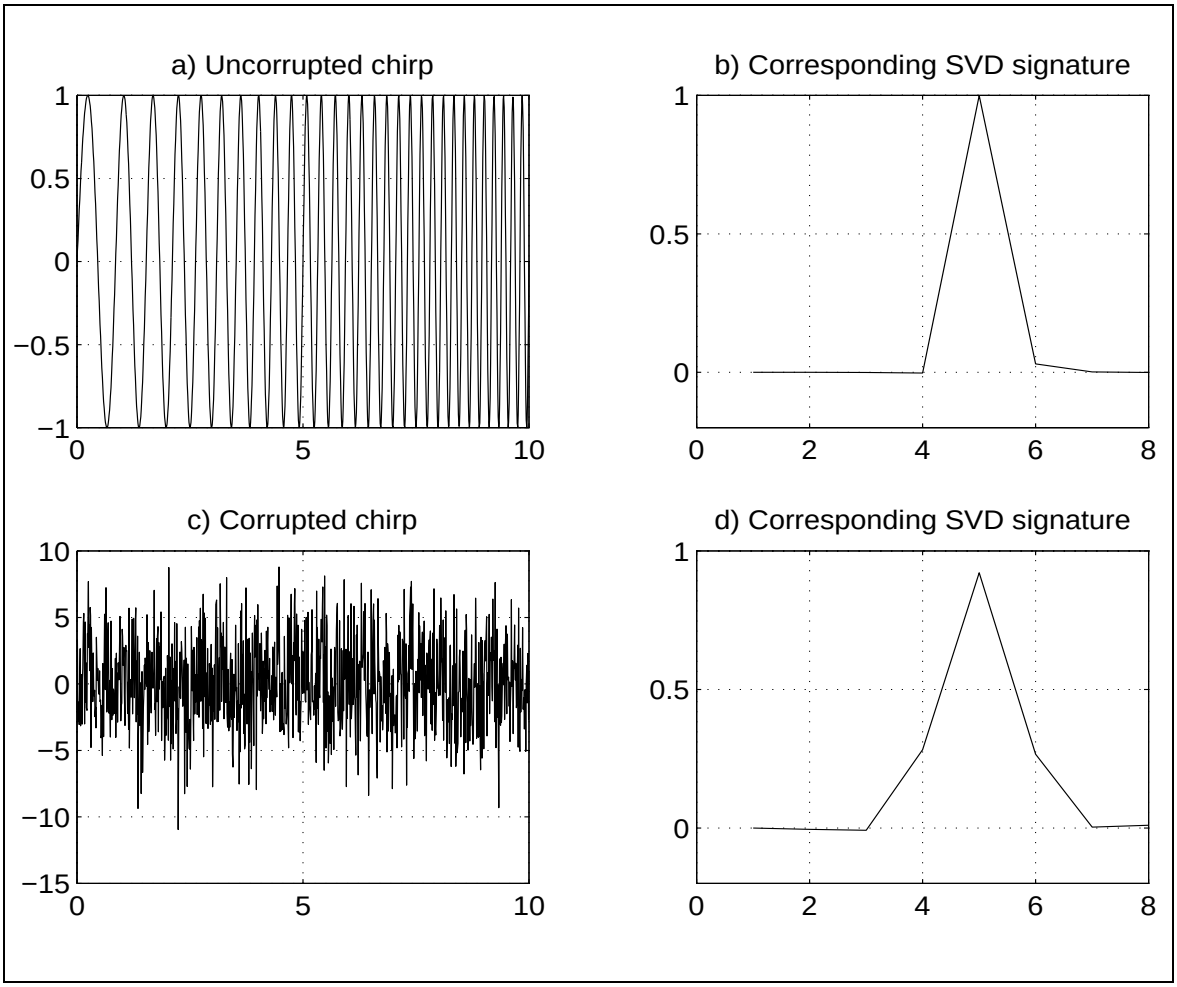

Figure 5.4: Robustness of the SVD_M signatures

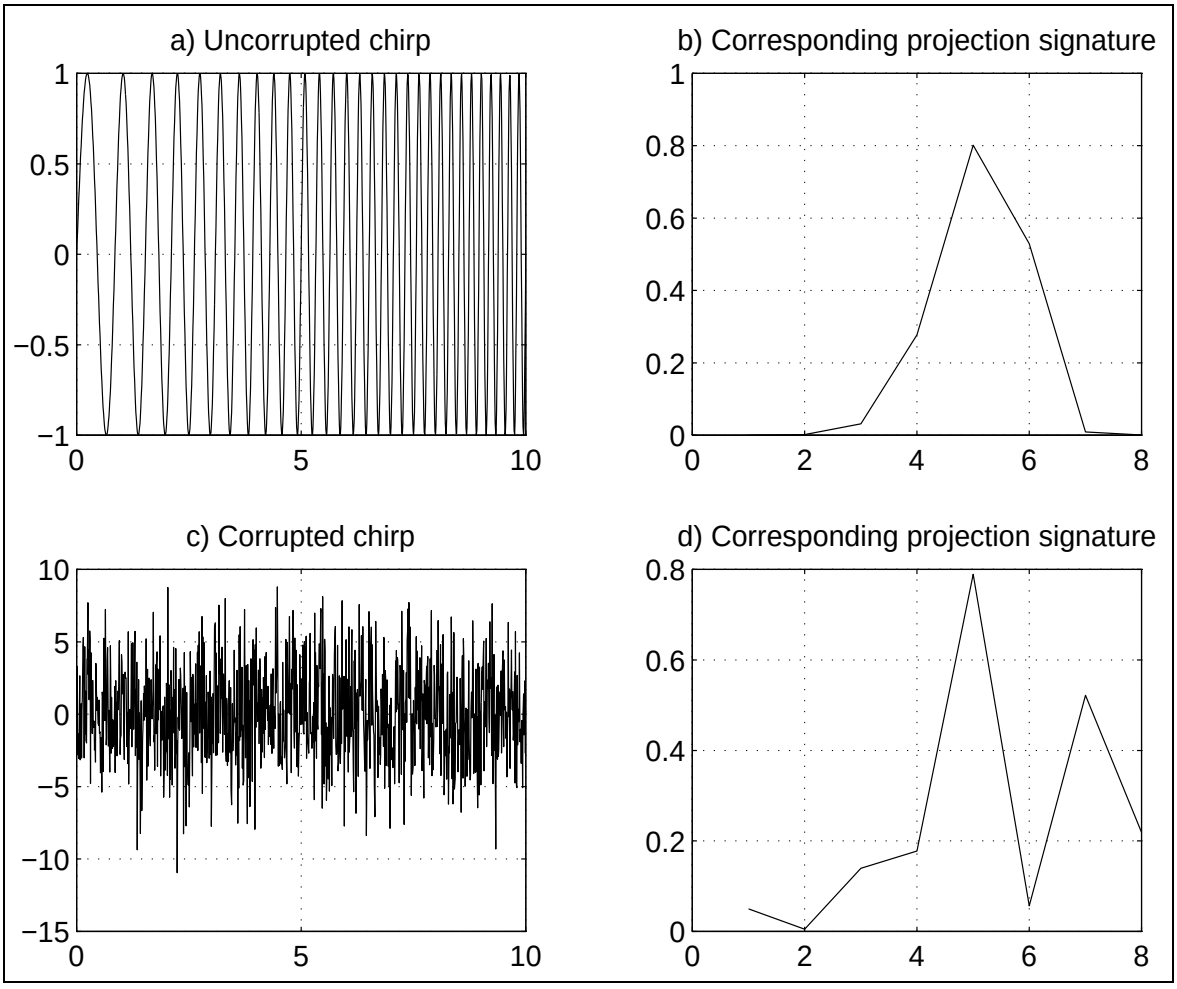

Figure 5.5: Robustness of the projection signatures

Another study of importance in the robustness analysis of the projection signatures is the sensitivity to the initial condition. Since the solution technique offers no guarantee that a global minimum is attained, it is necessary to study the effect of different initial points. From the nature of the solution technique used, the projection signatures do not depend on the initial choice of s_d^0, a fact that was supported by experimental results. The robustness analysis then is effectively based on the sensitivity to the initial vector r_d^0. For the study, one would need to consider two different situations. The first situation is the effect of random perturbations in r_d^0 on the projection signatures. For the signal x shown in Figure 5.6, with $r_d^0 = \chi_x$, the effect of both small ($< 10\%$), and large ($> 75\%$) random perturbations in r_d^0 on the projection signature is shown in the same figure. The results displayed in the figure are representative of several test cases, and indicate that the projection algorithm is quite robust in the face of random perturbations in the initial condition.

The second situation is the case where the basic initial condition r_d^0 itself varies. If one could ensure the achievement of the globally optimal solution using the computational technique, then the solution would be completely independent of the initial condition. In many cases, the projection signatures were completely unaffected by the choice of the initial condition. However, for the three signal classes shown in Figure 5.1, the projection signatures (especially $S1$, and $S2$) show quite some variation when a different initial point is used, as is evident from Figure 5.7. In the figure, the results are shown for r_d^0 chosen randomly, and for r_d^0 obtained from the principal component of the SVD of the discretized CWT matrix for each signal. This result suggests that the solution technique does not converge to the global minimum in all cases. This is a common limitation of nonlinear minimization algorithms, and one usually gets around it by analyzing different initial conditions, and selecting the

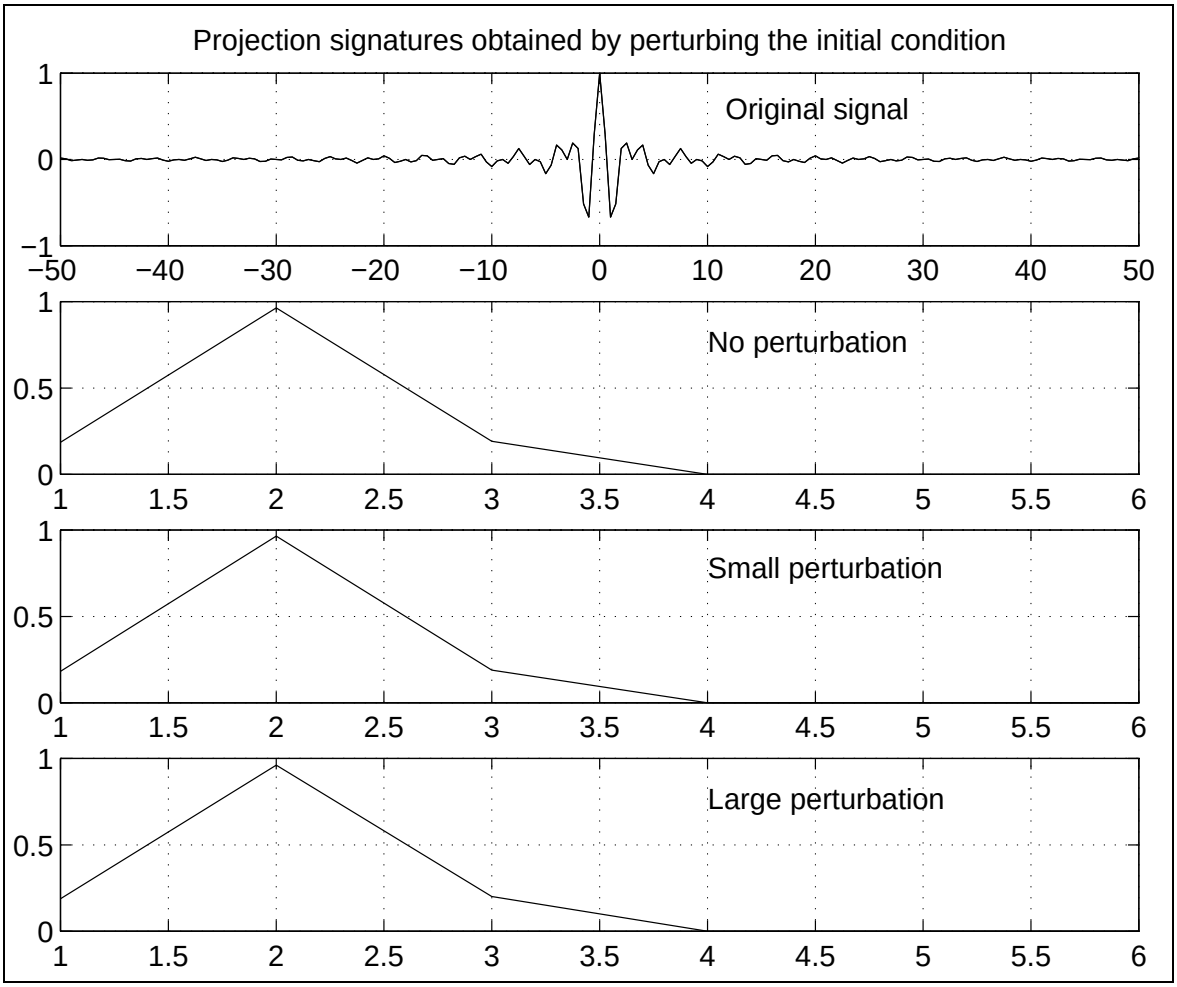

Figure 5.6: Effect of random perturbations in the initial condition r_d^0 on the projection signatures

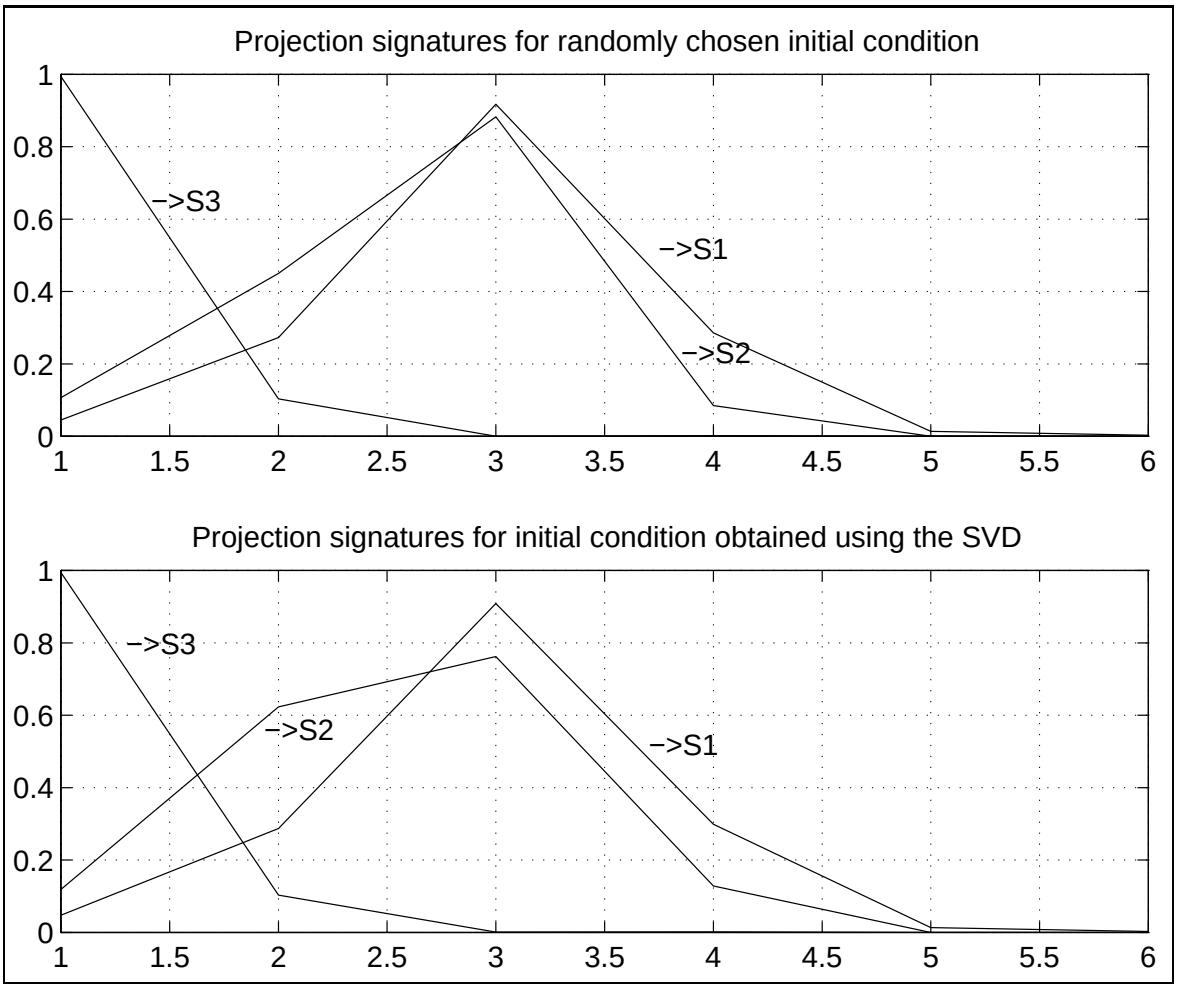

Figure 5.7: Effect of different initial conditions r_d^0 on the projection signatures

one that gives best results. For the classification problem under consideration, one of the main requirements is that the projection signatures obtained should clearly separate even closely spaced signal classes. Intuitively, one can see that for a signal x, if $r_d^0 = \chi_x$, then it has little information pertaining to the intrinsic signal properties, which would 'force' the associated projection signature to capture most of the information about the intrinsic properties of the signal. On this basis, one could conclude that the projection signature would better represent the signal class. This conclusion was borne out by every experimental study made on different signal classes, and a variety of initial conditions. Thus, the initial vector r_d^0 was always set at $r_d^0 = \chi_x$ for the determination of the projection signatures for purposes of classification.

5.2.4 Effect of the analyzing wavelet

The solution procedure used to generate the projection signatures assumes a fixed admissible wavelet $\psi \in L^2(\Re)$ of compact support that arises from a MRA. It is worthwhile to study the effect of using different wavelet functions on the projection signatures of signals. An example of the effect of using different wavelet functions of the Daubechies family ([26]) is illustrated in Figure 5.8. It can be observed that while there are some variations in the projection signatures obtained, they are essentially of the same nature (especially the signatures obtained using $Db6$ and $Db10$).

Now, the Daubechies family of wavelet functions $\{Dbi\}_i$ essentially corresponds to band pass filters with decreasing center frequency as i increases. Hence, one might expect that a high frequency signal will be better represented by $Db2$ than say, by $Db10$. However, the initial results on different classification problems do not indicate any significant differences in the quality of the classification owing to using

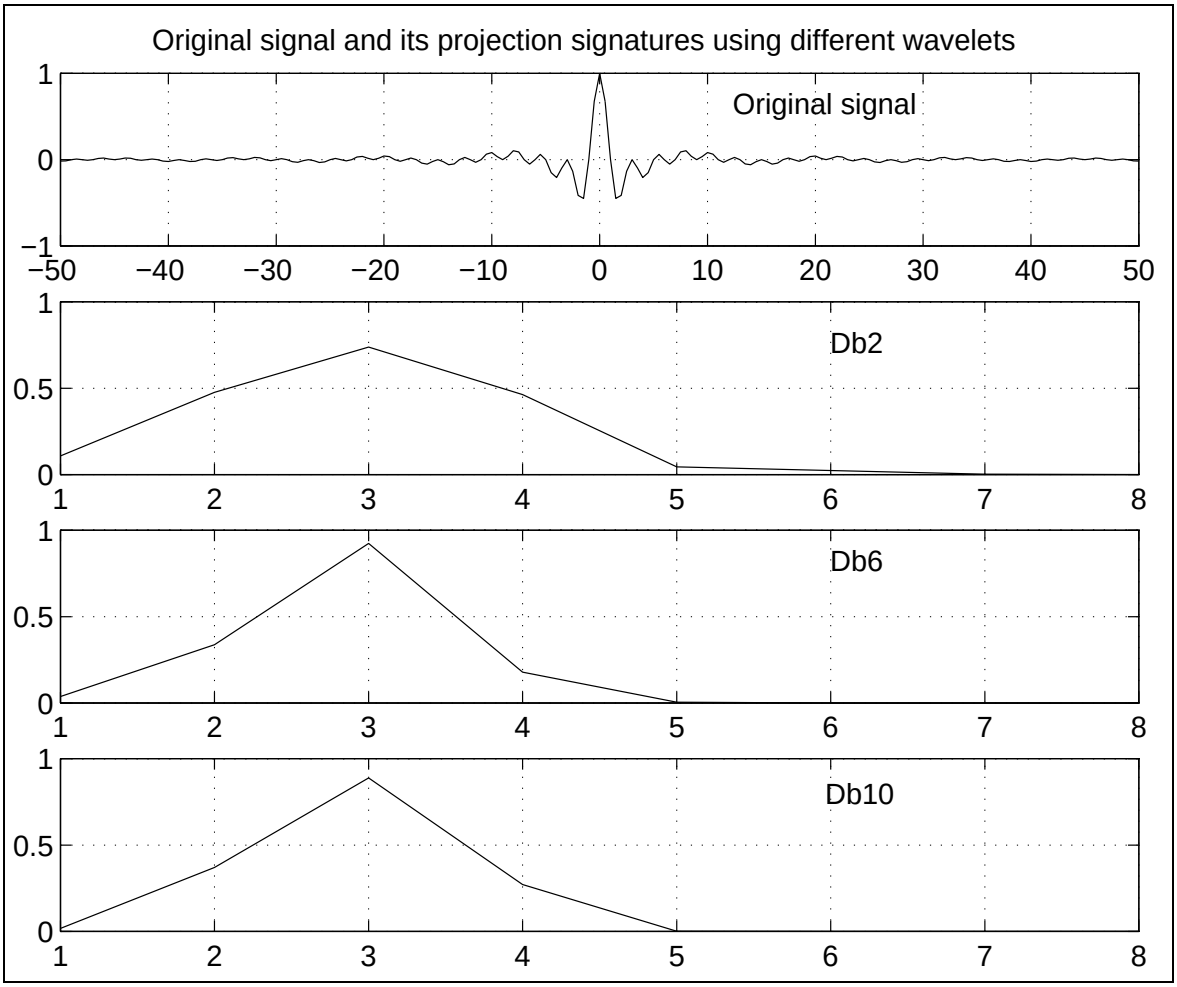

Figure 5.8: Effect of different wavelet functions on the projection signatures

different members of the Daubechies family of wavelets. Nevertheless, the problem of optimal wavelet selection for a better quality classification is an area which needs to be studied in depth. One possible advantage of this study is that if one can indeed select an optimal wavelet, one might be able to obtain sufficiently discriminating signatures using the simple, and more robust, matrix SVD approach.

5.2.5 Convergence issues

In the discussion on the computational complexity of the projection algorithm, an average number of iterations for the convergence of the conjugate gradient algorithm for each minimization was assumed. From several experiments, it was observed that convergence was usually achieved in L steps for the minimization with respect to s_d, but far fewer than N steps (less than 15 steps for $N = 30000$) for the minimization with respect to r_d without the need for a termination criteria. The number of iterations I required for convergence to the minimum showed a wide variation, with values as low as $I = 3$ for a signal of length $N_x = 201$, to values as high as $I = 19$ for a signal of length $N_x = 25000$. In both cases, $N_\psi = 8$, and $L = 8$. For illustration purposes, some sample results on eight different signals of length $N_x = 201$, with $N_\psi = 8$, and $L = 8$, are shown in Figure 5.9. It is seen that, in most of the cases, the maximum reduction in cost is achieved in the first iteration, with only marginal improvements in the subsequent steps. This figure also serves to validate the monotonically decreasing nature of the cost function. From the figure, and several other simulation results, one may then conclude that the algorithm used to generate the projection signatures converges to the minimum in relatively few iterations; i.e., it has a reasonably fast convergence rate.

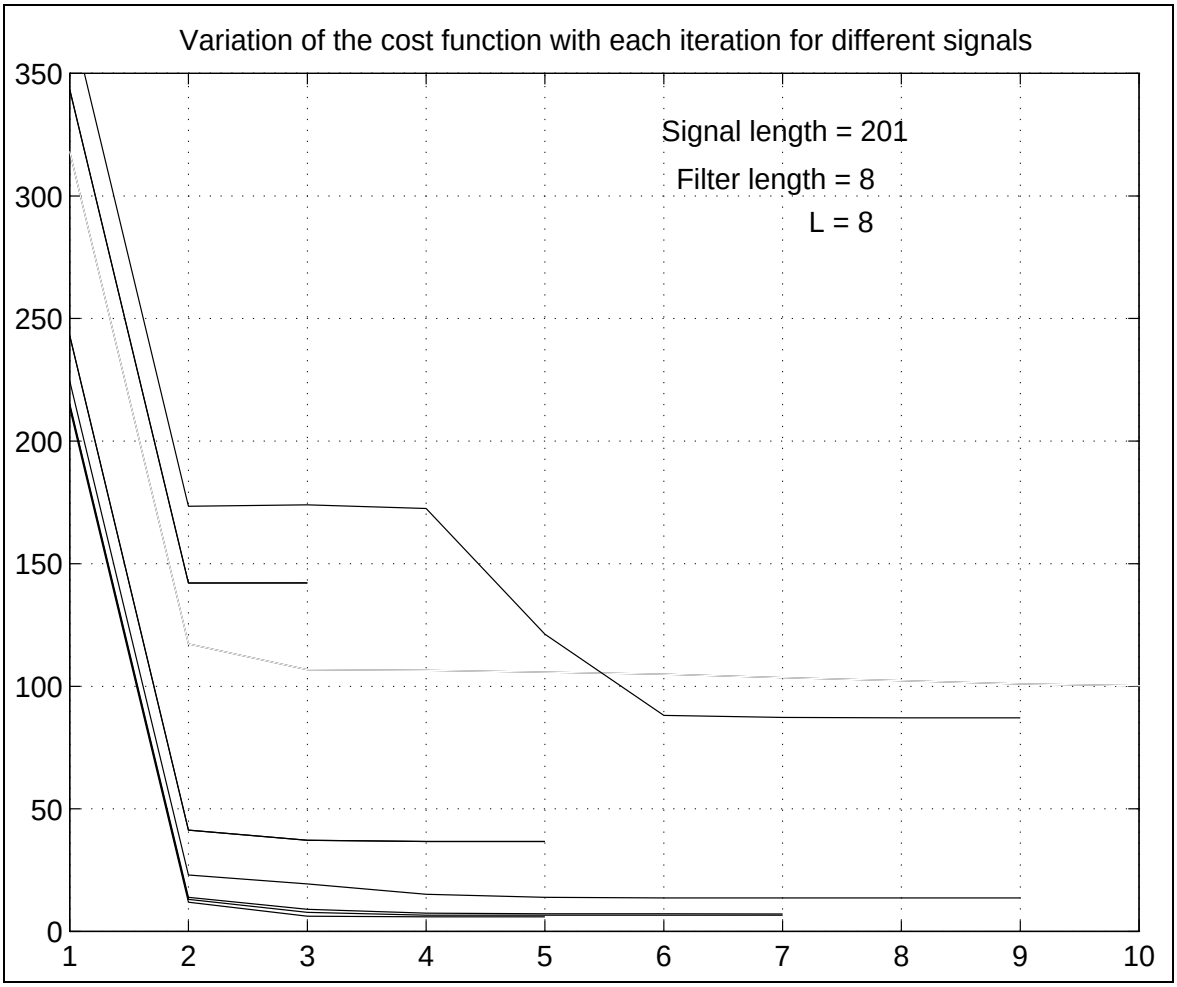

Figure 5.9: Variation of the cost function with each iteration

5.2.6 Quality of the classification

The last issue addressed in this chapter deals with the actual quality of the classification results. For an unambiguous classification, one would ideally like to have correlation graphs which are relatively smooth over different segments, and have sharp transition points. While the second criteria is reasonably met by the correlation graphs shown in Figure 5.2, the first is not. There exist some oscillations in the correlation graphs over the different segments which cannot be easily explained. A curious result that was observed was that these oscillations or fluctuations were greatly reduced when the correlation was performed between the absolute values of the projection signatures, and the discretized CWT, as shown in Figure 5.10. A possible explanation for this could be that the development of the pseudo power signatures was based on the result that if the CWT of a signal was approximated using a separable function of the form $c_{\psi}^{x}(a,b) \approx s(a)r(b)$, then the scalogram SC_{ψ}^{x} (modulus squared of the CWT) could be represented, at some time b_0, by the scale power function $\mid s(a) \mid^2$, suitably normalized. Thus, a correlation in terms of the modulus of the CWT and the pseudo power signature might lead to better quality results.

A second issue regarding the quality of the classification is the use of the simple correlation technique as the basis for the classification. While it is true that the correlation approach gives fairly good results for artificially generated signals, it is reasonable to expect that one might need a more sophisticated technique when dealing with real data signals. At the very least, one might expect to need some additional techniques, along with the straightforward correlation approach. These techniques could take the form of preprocessing the signal to remove noise, and removing cross - correlation effects occurring due to interaction between adjacent

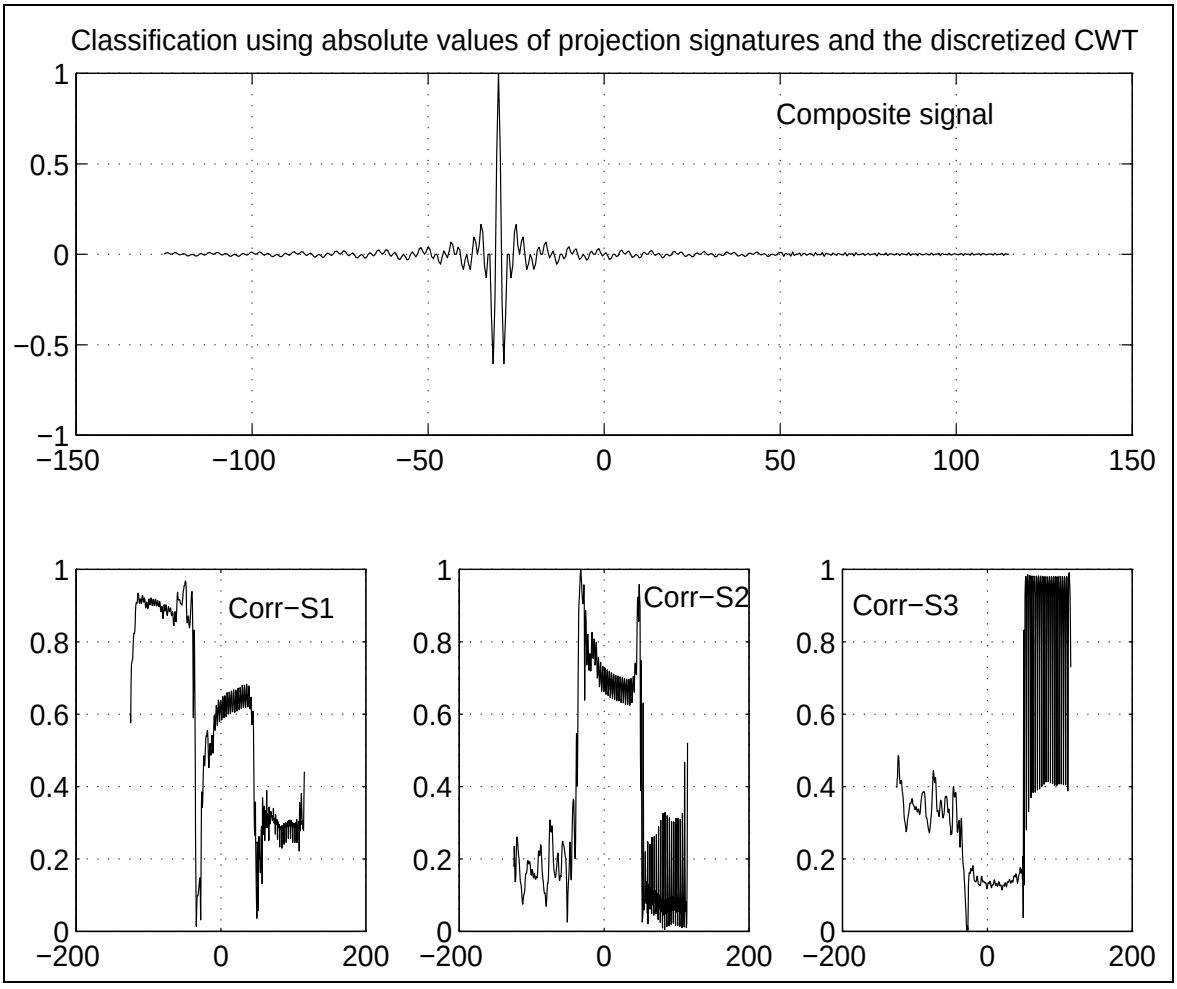

Figure 5.10: The composite signal and its correlation graphs using absolute values

signal classes. The latter technique would be extremely useful when dealing with signals where two or more events could be present at the same time. This is an issue for further study, and will be addressed in greater depth in the next chapter.

Chapter 6

Conclusion

A methodology to perform nonstationary signal classification, using signatures which are essentially independent of signal length, has been discussed in the previous chapters. The methodology has some distinct advantages, though it also suffers from certain limitations. This chapter gives a brief summary and a critical evaluation of the methodology, as well as, suggestions for future work in the area.

6.1 Summary and analysis of the classification methodology

The problem that was addressed in this research was the classification problem P, stated again for convenience.

P : *There is a known class of events,* $\{E_k; k = 1, \ldots, n\}$*, which may appear in a given scene for a variable time interval. One has collected data from the scene as a signal* $x(t)$; $t_l \leq t \leq t_h$*, and it is known that only one event is present at any given time. Then there is an unknown partition* $P_x = \{t_l \leq t_1 \leq t_2 \ldots \leq t_r \leq t_{r+1} \ldots \leq t_h\}$*, of* **transition times** *marking the start and end times of an event. The goal is to determine the transition times and the events occurring in each time interval. This process is called* **classification of the signal** $x(t)$.

The approach to the classification was based on generating signatures or characteristic representations for each event or component signal class. Since the component signals are known to be nonstationary, they are best represented using TFDs. Owing to the excellent localization properties of the CWT, and the efficient techniques available for its computation using filter banks, the CWT was well suited to meet

the particular requirements of the classification problem P, and thus, was the TFD of choice in the determination of the signatures.

Since each component signal could be of any arbitrary duration, one of the primary requirements of the signatures generated was that the signatures should be independent of the signal length. A complete discussion of the need for such time independent signatures was provided in Chapter 1. This requirement led to the formulation of the concept of the *pseudo power signatures* as discussed in Chapter 2. These signatures are essentially approximate scale power distribution functions associated with the time-scale energy density function of the CWT. Since the signatures are approximate 'scale' power distribution functions, they can be used to characterize signal classes in a manner that is independent of time.

Two approaches to compute the pseudo power signatures were presented in this work. The first was a simple approach using a principal component analysis, which essentially reduced to the traditional matrix SVD analysis for discrete signals. The complete solution technique, and experimental results were presented in Chapter 3. It was observed that the signatures generated using the matrix SVD technique were fairly robust, and could successfully be applied to the classification of signals where the components were sufficiently separated in the time-frequency plane. However, the signatures lacked a fine discriminating capability, and were not useful for the classification of signals with closely spaced components. A possible explanation for this could be that the orthogonality assumption made in the discretization technique, is not valid in the space of the CWTs.

A more refined approach to the determination of the pseudo power signatures was proposed in Chapter 4. The approach generated signatures which were obtained using orthogonal projections, and hence, the corresponding signatures were referred

to as projection signatures. The method used to generate the projection signatures essentially involved a nonlinear minimization, and required solving an inverse projection problem. While the approach was computationally more complex than the matrix SVD technique, the signatures obtained were highly discriminating, as could be seen based on some representative experimental results on the signals shown in Chapter 5. Since the signature computation can be done off-line, the computational cost is not a significant drawback when one deals with unknown signals with closely correlated signal components. The more serious problems with the projection approach were that the computational technique made the piecewise constant assumption on the CWT, which may not be valid when working with real data signals, and the technique was not very robust in the presence of noise. These problems and their possible solutions are addressed in the next section, along with some suggestions for future work.

6.2 Areas of future work

One of the first areas of future work is the improvement of the performance of the projection computational algorithm to overcome the problems associated with it. While the problem of making the simplifying assumption of a piecewise constant form for the CWT is of concern, the assumption is necessary to justify the approximation of the orthogonal projection operator $\mathcal{K}$, using the computationally efficient forward and inverse Shensa operators. In order to avoid this, one needs to determine a more accurate technique to compute $\mathcal{K}$ that does not make any assumptions. Moreover, the technique must be reliable, and computationally efficient. A possible solution may be found by exploiting the nature of the operator $\mathcal{K}$. The operator $\mathcal{K}$ is Hilbert-Schmidt, compact, self-adjoint and positive semidefinite. On a separable

Hilbert space, it can be represented as a square matrix ([28]). Moreover, it can be represented as the limit of finite rank compact operators, each of which has a finite matrix representation. It is feasible that a careful analysis of the structure of $\mathcal{K}$ will provide a better finite approximation to $\mathcal{K}$. However, the problem of finding an efficient implementation for this approximation still remains an open issue. This is not an easy problem, and a useful solution may be difficult. A more realistic solution would be to use the same framework as proposed in Chapter 4, but obtain signatures with a finer discretization of the scale. At present, there exist techniques ([31]) for the computation of the CWT on a finer grid in scale. Though this approach will significantly increase the computational cost of the projection algorithm, the piecewise constant assumption on the CWT is more readily met by signals, when one uses a very fine time-scale grid.

The second problem of concern with the projection signatures involved the susceptibility to noise. Though it it true that most real data contains some noise, preprocessing the data set before applying the classification technique to it, usually reduces the problem of noise significantly. It is generally accepted that there is always a trade off between quality and robustness, in the sense that it is unrealistic to expect a high quality or very discriminating signature, (one which very closely represents the associated signal class), to also be very robust. The kind of preprocessing to be done to overcome the lack of robustness problem is an issue that can be further studied. An idea worth exploring would be to apply the robust SVD_M signatures to first obtain a coarse classification, and only use preprocessing and the projection signatures to clarify any ambiguities and refine the classification.

An issue not directly relevant to the signature computation problem, but which is of importance in signature quality, is the selection of an optimal wavelet for the

signature generation, both in the matrix SVD and projection techniques. Initial experimental results indicate that, in some cases, the selection of a wavelet that more closely 'matches' a particular signal class (in terms of time-frequency characteristics), provides a better quality signature. However, in order to make a conclusive analysis, a more thorough study needs to be made. The study might conceivably involve solving another minimization, but this time, with respect to the wavelet function. A possible advantage of solving this problem is that one might be able to obtain fairly discriminating signatures using the simple and efficient matrix SVD technique if one uses a more 'optimal' wavelet.

Another significant issue that needs further research is one that is directly applicable to real world signals. This issue is the one where the basic assumption in P, that only one event is present at any given time, is dropped. In a real world scenario, it is unrealistic to expect that adjacent signal components do not interact with each other, and that a composite signal consisting of several different components, will not exhibit some characteristics which result from these interactions. Essentially, in order to obtain a more realistic and meaningful classification, one needs to modify the basic problem P to allow for the *presence of two or more events at a time.*

The problem then is one of obtaining a more realistic model for the classification. In signal processing terminology, one can relate the modeling problem to the stratigraphic classification problem which motivated this research. As shown in Figure 6.1, one can view each stratigraphic layer as a filter, and the original known trace as the filter response to the input pulse. These responses could then be seen as the impulse responses e_i of each layer. Every event in P then can be understood to represent the filtered response y_i of the filter with impulse response e_i. The goal of the classification would be the identification of the e_i associated with each event y_i.

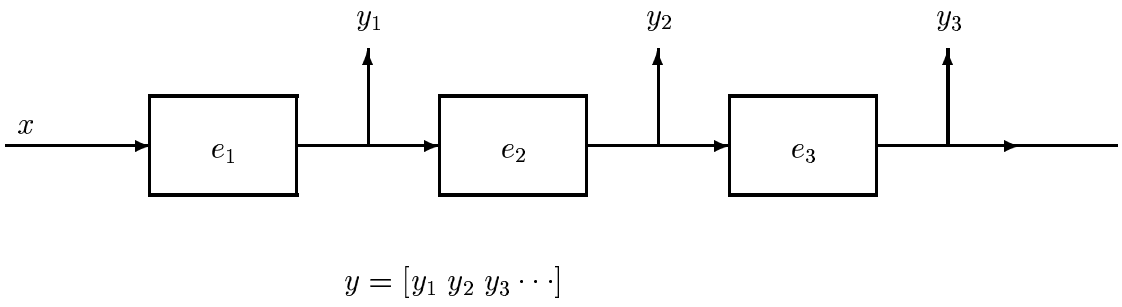

Figure 6.1: Modeling the classification problem

With this understanding, in Figure 6.1, the response y_i of each filter e_i represents each event in P (each stratigraphic layer), with x representing the input pulse. The important difference between the problem modeling in P, and the modeling here, is that the *filters are all connected in series*, which automatically implies that the reflected echo y_i not only depends on x, and e_i, but also on $e_j,\ j < i$. The unknown signal to be classified is given by $y = [y_1\ y_2\ y_3 \cdots]$. It is intuitive that this modeling more closely depicts a practical situation.

Once the problem has been modeled, the issue of the signal classification can be addressed. Suppose one has determined the signatures corresponding to each event e_i. From the nature of the model, it is obvious that the straightforward correlation technique used to perform the classification may, in general, not yield good results. One needs to take into account the effects of interaction of the different events or signal classes which might affect the correlation values. Effectively, one would need to determine the cross correlation effects between different classes, and process them out. The problem becomes quite complex when one deals with a large number of possible events which may occur in any random order. However, this is an important problem, and its solution would greatly increase the applicability of the classification technique to a wide variety of problems.

In conclusion, a methodology to perform nonstationary signal classification for signals of arbitrary unknown duration has been formulated through the use of time independent pseudo power signatures. It is worthwhile to note that the actual classification process can be performed quickly because the signatures are vectors of very small dimension. The method has wide applicability since it permits the separation of highly correlated signals. The potential areas of application are in fields as diverse as oil exploration, hidden mine detection, moving target detection, system identification, and pattern recognition.

Bibliography

[1] Leon Cohen, " Time-frequency Distributions - A Review," *Proc. IEEE*, Vol. 77, No. 7, pp 941-977, July 1989.

[2] Mackay A. E. Okure and Michael A. Peshkin, " Quantitative Evaluation Of Neural Networks For NDE Applications Using The ROC Curve," *Proc. of the 21st Annual Review of Progress in Quantitative Nondestructive Evaluation*, Snowmass, CO, 1994.

[3] Irwin W. Sandberg and Ajit T. Dingankar, " Signal Classification And Relative Compactness," *Proc. of the 1995 IEEE Workshop on Nonlinear Signal and Image Processing*, Greece.

[4] Anthony J. Pohl, " Adaptive And Fixed Wavelet Features For Narrowband Signal Classification," *M.S. Thesis*, Air University, 1995.

[5] F. Hlawatsch and W. Kozek, " Time-frequency Analysis Of Linear Signal Spaces," *Proc. IEEE ICASSP*, 1991.

[6] F. Hlawatsch and W. Kozek, " Time-frequency Projection Filters And Time-frequency Signal Expansions," *IEEE Trans. on Signal Processing*, Vol. 42, No. 12, pp 3321-3334, Dec. 1994.

[7] F. Hlawatsch, " Interference Terms In The Wigner Distribution," *Proc. 1984 Int. Conf. on Digital Signal Processing*, pp 363-367, Florence, Italy, Sep. 1984.

[8] Y. Zhao, Les E. Atlas and R. Marks, " The Use Of Cone-shaped Kernels For Generalized Time-frequency Representations Of Nonstationary Signals," *IEEE Trans. ASSP*, Vol. 38, No. 7, pp 1084-1091, July 1990.

[9] Jack McLaughlin and Les E. Atlas, " Applications Of Operator Theory To Time-frequency Analysis And Classification," *submitted to IEEE Trans. on Signal Processing.*

[10] E. J. Zalubas and M. G. Amin, " Time-frequency Kernel Design By The Two-dimensional Frequency Transformation Method," *IEEE Trans. on Signal Processing*, Vol. 43, No. 9, pp 2198-2202, Sep. 1995.

[11] Daniel Beyerbach and Hamid Nawab, " Principal Components Analysis Of The Short Time Fourier Transform," *Proc. IEEE ICASSP*, 1991.

[12] Mark L. Fowler and Leon H. Sibul, " Signal Detection Using Group Transforms," *Proc. IEEE ICASSP*, 1991.

[13] Nurgun Erdol and Feng Bao, " Wavelet Transform Based Optimal Receivers," *Proc. of the Asilomar Conference*, 1993.

[14] Douglas L. Jones and Thomas W. Parks, " A Resolution Comparison Of Several Time-frequency Representations," *IEEE Trans. on Signal Processing*, Vol. 40, No. 2, pp 413-420, Feb. 1992.

[15] T.A.C.M. Claasen and W.F.G. Mecklenbrauker, " The Wigner Distribution - A Tool For Time-frequency Signal Analysis, Part I : Continuous Time Signals," *Philips Journal Of Research*, Vol. 35, No. 3, pp 217-250, 1980.

[16] T.A.C.M. Claasen and W.F.G. Mecklenbrauker, " The Wigner Distribution - A Tool For Time-frequency Signal Analysis, Part II : Discrete Time Signals," *Philips Journal Of Research*, Vol. 35, No. 3, pp 256-299, 1980.

[17] T.A.C.M. Claasen and W.F.G. Mecklenbrauker, " The Wigner Distribution - A Tool For Time-frequency Signal Analysis, Part III : Relations With Other Time-frequency Signal Transformations," *Philips Journal Of Research*, Vol. 35, No. 6, pp 372-389, 1980.

[18] Leon Cohen, " Generalized Phase-space Distribution Functions," *J. Math. Phys.*, Vol. 7, No. 5, pp 781-786, 1966.

[19] Olivier Rioul and Martin Vetterli, " Wavelets And Signal Processing," *IEEE Signal Processing Magazine*, pp 14-38, Oct. 1991.

[20] Olivier Rioul and Patrick Flandrin, " Time-Scale Energy Distributions: A General Class Extending Wavelet Transforms," *IEEE Trans. on Signal Processing*, Vol. 40, No. 7, pp 1746-1757, July 1992.

[21] Stephane Mallat and Zhifeng Zhang, " Matching Pursuit With Time-frequency Dictionaries," *IEEE Trans. on Signal Processing*, Vol. 41, No. 12, pp 3397-3415, Dec. 1993.

[22] R. Coifman and V. Wickerhauser, " Entropy-based Algorithms For Best Basis Selection," *IEEE Trans. on Information Theory*, Vol. 38, No. 2, Mar. 1992.

[23] X. Xu, I.S. Reed, W. Kraske, and A.D. Stocker, " A Robust Multi-spectral Object Detection By Using Wavelet Transform," *Proc. IEEE ICASSP*, 1992.

[24] V. Gouaillier and L. Gagnon, " Ship Silhouette Recognition Using Principal Components Analysis," *Proc. SPIE Conf. on Applications of Digital Image Processing XX*, San Diego, 1997.

[25] Rachel L. Learned and Alan S. Willsky, " A Wavelet Packet Approach To Transient Signal Classification," *LIDS Technical Report, LIDS-P-2199*, submitted to Applied and Computational Harmonic Analysis.

[26] I. Daubechies, " Ten Lectures On Wavelets," *Regional Conference Series in Applied Mathematics*, SIAM, Philadelphia, 1992.

[27] M. Reed and B. Simon, " Methods Of Modern Mathematical Physics 1 : Functional Analysis," *Academic Press, Inc.*, San Diego, 1980.

[28] B.V. Limaye, " Functional Analysis," *John Wiley & Sons*, 1981.

[29] L. Cohen, " Time-frequency Analysis," *Prentice Hall Signal Processing Series*, 1995.

[30] Mark J. Shensa, " The Discrete Wavelet Transform: Wedding The À Trous And Mallat Algorithms," *IEEE Trans. on Signal Processing*, Vol. 40, pp. 2464-2482, Oct. 1992.

[31] Olivier Rioul and Pierre Duhamel, " Fast Algorithms For Discrete And Continuous Wavelet Transforms," *IEEE Trans. on Information Theory*, Vol. 38, No. 2, pp 569 - 586, Mar. 1992.

[32] G. H. Golub and C. F. Van Loan, " Matrix Computations - II Edition," *The Johns Hopkins University Press*, 1989.

[33] Anand K. Soman and P.P. Vaidyanathan, " On Orthonormal Wavelets And Paraunitary Filter Banks," *IEEE Trans. on Signal Processing*, Vol. 41, No. 3, Mar. 1993.

[34] Martin Vetterli and Cormac Herley, " Wavelets And Filter Banks : Relationships And New Results," *Proc. IEEE ICASSP*, 1990.

[35] Truong Q. Nguyen, " Digital Filter Bank Design : Quadratic-constrained Formulation," *IEEE Trans. on Signal Processing*, Vol. 43, No. 9, pp 2103-2108, Sep. 1995.

[36] Oktay Alkin and Hakan Caglar, " Design Of Efficient M-band Coders With Linear Phase And Perfect Reconstruction Properties," *IEEE Trans. on Signal Processing*, Vol. 43, No. 9, pp 1579-1590, July 1995.

[37] Arthur Wouk, " A Course Of Applied Functional Analysis," *John Wiley & Sons*, 1979.

Appendix A

Filter banks are a classical implementation of the principle of splitting a signal and then reconstructing it. The typical scheme is shown in Figure A.1. There are two sections to any filter bank - the analysis section, and the synthesis section. The input signal $x(n)$ is divided into M frequency bands using the filters $\tilde{F}_k$ in the analysis section. The channel component signals are then decimated by a factor R which results in the aliased channel outputs. The two operations can be represented by an equivalent linear operation F_k. The system is said to be *maximally decimated* when $R = M$. For most applications, the analysis filters are band pass filters which are decimated to their nominal Nyquist rates. Thus these signals form a set of critically sampled representations of the original signal $x(n)$.
The output signal $\hat{x}(n)$ is a reconstructed version of the input signal. The reconstruction is achieved by linear operations G_k which consist of upsampling the channel signals to their original sampling rate, and passing each through the interpolation filters $\tilde{G}_k$ in the synthesis bank, and then summing the results.
The most common family of filter banks is the *Quadrature Mirror Filters(QMF)* where the analysis and synthesis banks are designed using a *perfect reconstruction criterion*. There is abundant literature on the design of QMFs and it is still an active research topic ([35],[36]). The basic idea behind the perfect reconstruction filter bank is the following. For ideal band pass filters in the analysis and synthesis sections, perfect reconstruction is possible. Unfortunately, ideal filters are not realizable, and in practice, the individual channel outputs are always aliased. However, the specialty of the filter bank environment is that information about the aliased signals in one channel is available in other channel signals. Hence, it is possible to obtain perfect reconstruction systems even though the individual channel signals are aliased if the filters F_k and G_k satisfy certain conditions. These conditions are :

- The perfect reconstruction (PR) condition :

$$\sum_{k=1}^{M} F_k G_k = I$$

- The aliasing cancellation condition :

$$F_k G_l = 0, \quad k \neq l$$

The main concentration of effort in the research in this area has been in the design of optimal, high quality filters (both FIR and IIR), which afford perfect reconstruction (upto a time delay). The theory of MRA provides one approach to design PR filter banks since all MRAs are associated with a PR filter bank. It is worthwhile to note that the converse is not true.

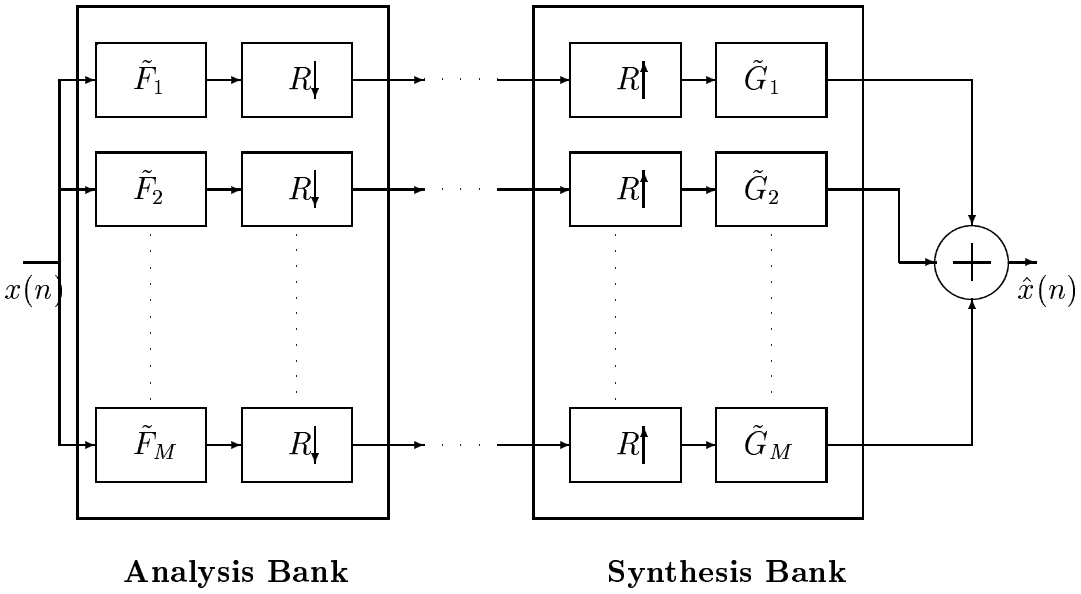

Figure A.1: M band analysis/synthesis filter bank system with decimation factor R

Appendix B

B.1 The infinite dimensional problem formulation

Let $H = (\Re^2, C_\psi^{-1}\frac{dbda}{a^2})$, and $M \subset H$ denote the space of the CWTs. The problem to be solved is as follows :
For a given $c_\psi^x \in M$, find the decomposition $s_\psi^x r_\psi^x \in H$ that minimizes the index

$$J(s_\psi^x, r_\psi^x) = \left\{ \| \ c_\psi^x - \mathcal{K}[s_\psi^x r_\psi^x] \ \|_M^2 + \| \ s_\psi^x r_\psi^x \ \|_H^2 \right\}$$

where $\mathcal{K} : H \to M$ is the orthogonal projection operator defined as :

$$\mathcal{K}[c](a,b) = C_\psi^{-1} \int_\alpha \int_\beta \overline{c_\psi^{\psi_{ab}}(\alpha,\beta)} c(\alpha,\beta) \frac{d\beta d\alpha}{\alpha^2}, \quad \forall \ c \in H$$

To formulate the above problem in a manner that lends itself to a solution, one can define the following maps.

1. The Continuous Wavelet Transform, $\Gamma : L^2(\Re) \to H$, defined by

   $$[\Gamma x](a,b) = \langle x, \psi_{a,b} \rangle \ ; \ x \in L^2(\Re)$$

   where $\psi \in L^2(\Re)$ is an admissible wavelet.

2. The map, $\tilde{\Lambda} : H \to L^2(\Re)$, defined by

   $$[\tilde{\Lambda} c](t) = C_\psi^{-1} \int_a \int_b c(a,b) \psi_{a,b}(t) \frac{dbda}{a^2}; \ c \in H$$

   This is effectively the inverse CWT operator, extended to all of H by defining $M^\perp$ to be its null space.

3. The family of maps, $\Theta_s : L^2(\Re) \to H$, defined for each $s \in S$ by

   $$[\Theta_s r](a,b) = s(a) r(b)$$

4. A dual family of maps, $\Upsilon_r : S \to H$, defined for each $r \in R$ by

   $$[\Upsilon_r s](a,b) = s(a) r(b)$$

Note that the Hilbert spaces H and S are not self - dual. However, by the Riesz isometry, one can always identify a Hilbert space with its dual. In the following analysis, this property is used to justify the definition of the adjoint maps.

Theorem B.1 *If $\psi \in L^2(\Re)$ is an admissible wavelet, then $\tilde{\Lambda}$ is a well defined map. Moreover, $\tilde{\Lambda}$ is the adjoint map from H to $L^2(\Re)$, i.e.*

$$\tilde{\Lambda} = \Gamma^*$$

PROOF. Let $\psi \in L^2(\Re)$ be an admissible wavelet. The CWT operator $\Gamma : L^2(\Re) \to H$, and hence its adjoint operator $\Gamma^* : H \to L^2(\Re)$. Thus, for any $c, d \in H$ where $c = \Gamma x$, using adjoints in Hilbert spaces, one obtains,

$$\langle \Gamma x, d \rangle_H = \langle x, \Gamma^* d \rangle_2$$

Hence,

$$\begin{aligned}
\langle \Gamma x, d \rangle_H &= C_\psi^{-1} \int_a \int_b \Gamma x(a,b) \overline{d(a,b)} \frac{dbda}{a^2} \\
&= C_\psi^{-1} \int_a \int_b \int_t x(t) \overline{\psi_{ab}(t)} dt \overline{d(a,b)} \frac{dbda}{a^2} \\
&= \int_t x(t) C_\psi^{-1} \int_a \int_b \overline{\psi_{ab}(t) d(a,b)} \frac{dbda}{a^2} dt; \quad (Fubini) \\
&= \int_t x(t) \overline{C_\psi^{-1} \int_a \int_b d(a,b) \psi_{ab}(t) \frac{dbda}{a^2}} dt \\
&= \langle x, \Gamma^* d \rangle_2
\end{aligned}$$

where

$$\begin{aligned}
\Gamma^* d &= C_\psi^{-1} \int_a \int_b d(a,b) \psi_{ab}(t) \frac{dbda}{a^2} \\
&= \tilde{\Lambda} d
\end{aligned}$$

Hence, $\tilde{\Lambda} = \Gamma^*$ on H. □

Theorem B.2 *The maps Θ_s, Υ_r are well defined and their adjoints are given by*
$\Theta_s^* : H \to L^2(\Re)$ defined by $\Theta_s^* c = \langle c, s \rangle_S$, $\forall c \in H$
$\Upsilon_r^* : H \to S$ defined by $\Upsilon_r^* c = \langle c, r \rangle_2$, $\forall c \in H$

PROOF. By definition,

$$[\Theta_s r](a,b) = s(a) r(b)$$

Using adjoints, one gets

$$\langle \Theta_s r, c \rangle_H = \langle r, \Theta_s^* c \rangle_2$$

Now,

$$\begin{aligned}
\langle \Theta_s r, c \rangle_H &= C_\psi^{-1} \int_a \int_b s(a) r(b) \overline{c(a,b)} \frac{dbda}{a^2} \\
&= \int_b r(b) C_\psi^{-1} \int_a s(a) \overline{c(a,b)} \frac{da}{a^2} db; \quad (Fubini) \\
&= \int_b r(b) \overline{C_\psi^{-1} \int_a c(a,b) \overline{s(a)} \frac{da}{a^2}} db \\
&= \langle r, \Theta_s^* c \rangle_2
\end{aligned}$$

which implies that

$$\begin{aligned}\Theta_s^* c &= C_\psi^{-1} \int_a c(a,b)\overline{s(a)}\frac{da}{a^2} \\ &= \langle c, s\rangle_S\end{aligned}$$

The derivation of the expression of Υ_r^* follows along exactly similar lines. □

From the above definitions, it readily follows that

- $\tilde{\Lambda}\Gamma = \Gamma^*\Gamma : L^2(\Re) \rightarrow L^2(\Re)$
  is the identity transformation, since $\Gamma^* = \Gamma^{-1}$ on M.
- $\Gamma\tilde{\Lambda} = \Gamma\Gamma^* : H \rightarrow H$
  is the orthogonal projection operator $\mathcal{K} : H \rightarrow M$ defined earlier [1].

B.2 Solution to the minimization problem

The nonlinear infinite dimensional minimization problem is solved using an iterative approach based on a successive minimization in each iteration, with respect to the variables $s \in S$ and $r \in R$. First, the existence of a minimizing solution to the problem is established.

B.2.1 Existence of the minimizer

Theorem B.3 *There exists* $\bar{s} \in B_S = \{s \in S;\ \|s\|_S \leq 1\}$, *and* $\bar{r} \in R$ *such that*

$$J(\bar{s}, \bar{r}) = \inf_{s \in B_S, r \in R} J(s, r)$$

[1] The result can be shown as follows: For $c \in H$,

$$\begin{aligned}\Gamma\Gamma^*[c](a,b) &= \int_t \Gamma^*[c](t)\overline{\psi_{ab}(t)}dt \\ &= \int_t C_\psi^{-1}\int_\alpha\int_\beta c(\alpha,\beta)\psi_{\alpha\beta}(t)\frac{d\beta d\alpha}{\alpha^2}\overline{\psi_{ab}(t)}dt \\ &= C_\psi^{-1}\int_\alpha\int_\beta c(\alpha,\beta)\int_t \psi_{\alpha\beta}(t)\overline{\psi_{ab}(t)}dt\frac{d\beta d\alpha}{\alpha^2}, \quad (Fubini) \\ &= C_\psi^{-1}\int_\alpha\int_\beta c(\alpha,\beta)\overline{\int_t \psi_{ab}(t)\overline{\psi_{\alpha\beta}(t)}dt}\frac{d\beta d\alpha}{\alpha^2} \\ &= C_\psi^{-1}\int_\alpha\int_\beta c(\alpha,\beta)\overline{c_\psi^{\psi_{ab}}(\alpha,\beta)}\frac{d\beta d\alpha}{\alpha^2} \\ &= \mathcal{K}[c](a,b)\end{aligned}$$

PROOF. The proof of the theorem is based on the results established below.
One can break the minimization problem into the following two subproblems. For a fixed $r \in R$, define the subproblem :

$$\min_s J(s) = \left\langle c_\psi^x - \mathcal{K}\Upsilon_r s, c_\psi^x - \mathcal{K}\Upsilon_r s \right\rangle_M + \langle \Upsilon_r s, \Upsilon_r s \rangle_H \tag{B.1}$$

Similarly, for a fixed $s \in S$, define the subproblem :

$$\min_r J(r) = \left\langle c_\psi^x - \mathcal{K}\Theta_s r, c_\psi^x - \mathcal{K}\Theta_s r \right\rangle_M + \langle \Theta_s r, \Theta_s r \rangle_H \tag{B.2}$$

Lemma B.1 *The solutions to the minimization problems defined in Equations B.1 and B.2 exist, and are unique.*

PROOF. It can be easily shown that the positive real valued functional $J(\cdot)$ defined in Equations B.1 and B.2 is quadratic in the variables s and r respectively. From Equation B.1, for a fixed r, denote $A_0 = \left\langle c_\psi^x, c_\psi^x \right\rangle$, $B = \Upsilon_r^* c_\psi^x$, and $A = \Upsilon_r^*(\mathcal{K}+I)\Upsilon_r$. Then, it follows that

$$J(s) = A_0 + \langle s, As \rangle - \langle B, s \rangle - \langle s, B \rangle$$

which is clearly quadratic in s. Moreover, since $A > 0$, the quadratic form is positive definite, and J is convex in s [2]. Hence, the minimization problem has a unique solution. A similar result can be shown for r. □

From Lemma B.1, the existence of a unique solution to each of the subproblems defined in Equations B.1 and B.2 is assured. Using Calculus of Variations, one can then determine the conditions on the minimizers to each of the subproblems. Since the functional J is separately convex in s, and r, the first order necessary conditions as determined using Calculus of Variations, become sufficient to determine the minimizers.
Consider the subproblem given in Equation B.1. For a fixed $r \in R$,

$$\begin{aligned} J_r(s) &= \left\langle c_\psi^x - \mathcal{K}\Upsilon_r s, c_\psi^x - \mathcal{K}\Upsilon_r s \right\rangle_M + \langle \Upsilon_r s, \Upsilon_r s \rangle_H \\ &= \left\langle c_\psi^x, c_\psi^x \right\rangle_M - \left\langle c_\psi^x, \mathcal{K}\Upsilon_r s \right\rangle_M - \left\langle \mathcal{K}\Upsilon_r s, c_\psi^x \right\rangle_M + \langle s, \Upsilon_r^*(\mathcal{K}+I)\Upsilon_r s \rangle_H \end{aligned}$$

[2] The convexity of J in s can be shown from first principles. Consider $J(s)$ which is defined on a convex domain. Let $0 < \lambda < 1$, and $s_1, s_2 \in S$. Then, with $\Delta_s = s_1 - s_2 \neq 0$, and $A = \Upsilon_r^*(\mathcal{K}+I)\Upsilon_r > 0$, one can readily obtain

$$\begin{aligned} J(\lambda s_1 + (1-\lambda)s_2) &= \lambda J(s_1) + (1-\lambda)J(s_2) - \lambda(1-\lambda)\langle \Delta_s, A\Delta_s \rangle \\ &< \lambda J(s_1) + (1-\lambda)J(s_2) \end{aligned}$$

which implies that $J(s)$ is strictly convex in s. Similarly, one can show that $J(r)$ is strictly convex in r.

Taking variations with respect to s, one obtains

$$\delta J_s = \left\langle \Upsilon_r^*(\mathcal{K}+I)\Upsilon_r s - \Upsilon_r^* c_\psi^x, \delta s \right\rangle_S + \left\langle \delta s, \Upsilon_r^*(\mathcal{K}+I)\Upsilon_r s - \Upsilon_r^* c_\psi^x \right\rangle_S$$

where δs is completely arbitrary. Setting $\delta J_s = 0$, one can then obtain the necessary condition for minimization with respect to s as

$$\Upsilon_r^*(\mathcal{K}+I)\Upsilon_r s - \Upsilon_r^* c_\psi^x = 0 \tag{B.3}$$

Let $Q_r = \Upsilon_r^*(\mathcal{K}+I)\Upsilon_r$. Since Q is a positive definite operator, there exists a unique solution to this equation given by $\bar{s} = Q_r^{-1}\Upsilon_r^* c_\psi^x$.
Following a similar approach, for a fixed $s \in S$, taking variations δJ_r with respect to r, one can obtain the necessary condition for minimization with respect to r by setting $\delta J_r = 0$. The condition is given by

$$\Theta_s^*(\mathcal{K}+I)\Theta_s r - \Theta_s^* c_\psi^x = 0 \tag{B.4}$$

If $P_s = \Theta_s^*(\mathcal{K}+I)\Theta_s$, there exists a unique solution to this equation given by $\bar{r} = P_s^{-1}\Theta_s^* c_\psi^x$ since P is a positive definite operator.
Let $B_S = \{s \in S; \ \|s\|_S \leq 1\}$ denote the closed unit ball in S. Since the Hilbert space S is reflexive ($S^{**} = S$), B_S is weakly compact. For a fixed s of unit norm, $s \in B_S$. Let $\bar{r} = P_s^{-1}\Theta_s^* c_\psi^x$ be the solution to Equation B.2, with P_s defined as before. Then, the functional $J(s,\bar{r}) = \left\langle c_\psi^x, c_\psi^x - \Theta_s P_s^{-1}\Theta_s^* c_\psi^x \right\rangle$ is effectively a function of $s \in B_S$.

Lemma B.2 *The real valued functional $J(s)$ is continuous on the weakly compact set $B_S = \{s \in S; \ \|s\|_S \leq 1\}$.*

PROOF. The real valued functional $J(s)$ defined on the weakly compact set B_S is given by

$$J(s) = \left\langle c_\psi^x, c_\psi^x - \Theta_{s^i} P_{s^i}^{-1}\Theta_{s^i}^* c_\psi^x \right\rangle, \quad s \in B_S$$

Let $\mathcal{K} + I = \Lambda^2 > 0$, where Λ is positive definite and self-adjoint. Let $\lambda_{min} > 0$ and $\lambda_{max} > 0$ denote the minimum and maximum eigen values of the positive definite operator Λ. Then,

$$\lambda_{min}^2 \parallel c \parallel_H^2 \leq \langle \Lambda c, \Lambda c \rangle_H \leq \lambda_{max}^2 \parallel c \parallel_H^2; \ \forall c \in H \tag{B.5}$$

Defining $X(s) = \Lambda\Theta_s$, $P_s = X^*(s)X(s)$. Using Equation B.5, one can readily establish the following results.

$$\lambda_{min} \parallel s \parallel_S \leq \parallel X(s) \parallel_S \leq \lambda_{max} \parallel s \parallel_S \tag{B.6}$$

$$\lambda_{min}^2 \parallel r \parallel_R^2 \leq \langle r, P_s r \rangle_R \leq \lambda_{max}^2 \parallel r \parallel_R^2; \ \forall r \in R; s \in B_S \tag{B.7}$$

$$\lambda_{max}^{-2} \parallel r \parallel_R^2 \leq \left\langle r, P_s^{-1} r \right\rangle_R \leq \lambda_{min}^{-2} \parallel r \parallel_R^2; \ \forall r \in R; s \in B_S \tag{B.8}$$

For $s_1, s_2 \in B_S$, let $\delta_s = s_1 - s_2$. Then,

$$\|\delta_s\|_S \leq \|s_1\|_S + \|s_2\|_S = 2 \tag{B.9}$$

Now,

$$P_{s_1} - P_{s_2} = X^*(s_2)X(\delta_s) + X^*(\delta_s)X(s_2) + X^*(\delta_s)X(\delta_s)$$

and hence, from Equations B.6 and B.9,

$$\|P_{s_1} - P_{s_2}\| \leq 4\lambda_{max}^2 \parallel \delta_s \parallel_S \tag{B.10}$$

Also,

$$P_{s_2}^{-1} - P_{s_1}^{-1} = P_{s_2}^{-1}\,(P_{s_1} - P_{s_2})\,P_{s_1}^{-1}$$

which implies from Equations B.8 and B.10

$$\parallel P_{s_2}^{-1} - P_{s_1}^{-1} \parallel \leq 4\lambda_{min}^{-4}\lambda_{max}^2 \parallel \delta_s \parallel_S;\ \forall s_1, s_2 \in B_S$$

From the above results, observe that one can express

$$J(s) = \left\langle c_\psi^x, c_\psi^x \right\rangle_M - \frac{1}{4}\left\langle X^*(s)\Lambda c_\psi^x, P_s^{-1}X^*(s)\Lambda c_\psi^x \right\rangle_R$$

Denoting $\Delta J = J(s_2) - J(s_1)$, one can then obtain

$$\begin{aligned} \mid \Delta J \mid \ &= \ \frac{1}{4}[\left\langle X^*(s_1)\Lambda c_\psi^x, (P_{s_1}^{-1} - P_{s_2}^{-1})X^*(s_1)\Lambda c_\psi^x \right\rangle_R + \\ &\quad \left\langle X^*(s_2)\Lambda c_\psi^x, P_{s_2}^{-1}X^*(\delta_s)\Lambda c_\psi^x \right\rangle_R + \left\langle X^*(\delta_s)\Lambda c_\psi^x, P_{s_2}^{-1}X^*(s_1)\Lambda c_\psi^x \right\rangle_R] \\ &\leq \ 4\lambda_{min}^{-4}\lambda_{max}^2 \left\| c_\psi^x \right\|_M^2 \parallel \delta_s \parallel_S + \lambda_{min}^{-2} \left\| c_\psi^x \right\|_M^2 \|\delta_s\|_S + \lambda_{min}^{-2} \left\| c_\psi^x \right\|_M^2 \|\delta_s\|_S \\ &\leq \ A\,\|\delta_s\| \end{aligned}$$

where A is a constant and depends only on $\parallel c_\psi^x \parallel_M$, λ_{min}, and λ_{max}. Then, given any $\epsilon > 0$, there exists $0 < \delta = \frac{\epsilon}{2A}$, such that for all $\parallel s_1 - s_2 \parallel_S < \delta$, $\mid J(s_2) - J(s_1) \mid < \epsilon$. Thus, J is continuous on B_S. □

The existence of the minimizing solution to the infinite dimensional problem is established as follows.
Fix $s \in B_S$. Then, by Lemma B.1, there exists $r(s) \in R$ such that $J(s, r(s)) = J(s) = \inf_{r \in R} J(s, r)$. The real valued functional J is thus defined on B_S which is weakly compact. From Lemma B.2, J is continuous on the weakly compact set. Hence, it is guaranteed to attain its maximum and minimum on the set ([37]); i.e., there exists $\bar{s} \in B_s$, such that $J(\bar{s}) = \inf_{s \in B_S} J(s)$. Now, $J(\bar{s}) = J(\bar{s}, r(\bar{s}))$. Hence,

$$J(\bar{s}, r(\bar{s})) = \inf_{s \in B_S} J(s, r(s)) = \inf_{s \in B_S, r \in R} J(s, r)$$

With $\bar{r} = r(\bar{s})$, the existence of the minimum is established. □

B.2.2 Iterative procedure

Once the existence of the minimizer has been established, one can develop a procedure for its determination. The procedure followed here is an iterative one, and requires successively solving the necessary conditions given in Equations B.3 and B.4 in each iteration. The approach is based on the result presented in Lemma B.1. Effectively, the result ensures that the iterative approach produces a monotonically decreasing cost function sequence. The iterative approach developed for successive minimization with respect to s, and r is given below.

- Pick random elements $s^0 \in S$, and $r^0 \in R$, and compute $J^0 = J(s^0, r^0)$.
- At the ith stage, with $r = r^{i-1}$, solve Eq. B.3 to obtain $\tilde{s}^i$. Thus, $\tilde{s}^i = Q^{-1}_{r^{i-1}} \Upsilon^*_{r^{i-1}} c^x_\psi$.
- Set $s^i = \frac{\tilde{s}^i}{\|\tilde{s}^i\|_S}$. The element $s^i \in S$ is thus constrained to the unit ball $B_S \subset S$.
- With $s = s^i$, solve Eq. B.4 to obtain r^i. Then, $r^i = P^{-1}_{s^i} \Theta^*_{s^i} c^x_\psi$.
- Compute $J^i = J(s^i, r^i) = \left\langle c^x_\psi, c^x_\psi - \Theta_{s^i} P^{-1}_{s^i} \Theta^*_{s^i} c^x_\psi \right\rangle$. J^i is thus seen to be only a function of $s^i \in B_S$, i.e. $J^i = J(s^i)$. Moreover, it is independent of the norm of $s^i \in B_S$, i.e. $J^i = J(s^i, r^i) = J(\tilde{s}^i, \| \tilde{s}^i \|_S r^i)$, which justifies the unit norm constraint imposed on $s^i \in B_S$.

B.2.3 Convergence to the optimal

From Lemma B.1, one can immediately see that the iterative procedure outlined produces a monotonically decreasing sequence of costs $\{J^i\}_i$, since one has the result $J(s^0, r^0) \geq J(\tilde{s}^1, r^0) \geq \cdots \geq J(s^i, r^i) \geq \cdots$, from which one can extract the monotonically decreasing sequence

$$\{J^i\}_i; \quad J^i = J(s^i) \tag{B.11}$$

defined on the unit ball B_S . The convergence of the iterative procedure is established based on the following result.

Theorem B.4 *There exists $\bar{s} \in B_S$, and $\bar{J} \geq 0$ such that the sequence $\{J^i\}_i$ defined in Equation B.11 converges to $\bar{J} = J(\bar{s})$.*

Proof. By Lemma B.2, the real valued functional $J(s)$ is continuous on the weakly compact set B_S. Hence, it is guaranteed to attain its maximum and minimum on B_S. The iterative procedure produces a monotonically decreasing sequence of costs $\{J^i\}_i$, whose limit $\bar{J}$ exists by Lemma B.2. Hence, $\lim J^i = \bar{J}$. The set B_S is also weakly sequentially compact.[3] This implies that from the sequence $\{s^i\}_i$, one can

[3] **Eberlein's Theorem :** Every bounded sequence in a reflexive Banach space X has a weakly convergent subsequence.

extract a subsequence $\{s^{i_j}\}_{i_j}$ that converges weakly to some $\bar{s} \in B_S$, such that by the continuity of J,

$$J(\bar{s}) = \lim J(s^{i_j}) = \lim J^{i_j} = \bar{J}$$

establishing the convergence to the optimal solution. □

An iterative procedure using successive minimization with respect to s, and r, which converges (weakly) to the limiting solution whose existence is guaranteed, has thus been developed. It is important to note that, while the procedure attains a minimum for the functional J, it offers no guarantee that the minimum attained is global. This is a general problem in nonlinear minimization techniques. The only way in which one can ensure that the minimum attained is indeed global is if the functional J is known to be *jointly* convex in the variables s and r, which is not the case in this problem.

Vita

Vidya Venkatachalam was born in Madras, India, on October 6, 1969. She received her bachelor of engineering degree in Electrical and Electronics Engineering from Anna University, Madras, India, in June 1992. She joined the graduate program in Louisiana State University in August 1992, and received the master of science degree in Electrical Engineering, in May 1995. She is currently a candidate for the degree of Doctor of Philosophy at Louisiana State University, which will be conferred at the December commencement, 1998.

Venkatachalam, Vidya, B.E., Anna University, 1992
M.S.E.E., Louisiana State University, 1995
Doctor of Philosophy, Fall Commencement, 1998
Major: Electrical Engineering; Minor: Mathematics
Pseudo Power Signatures for Nonstationary Signal Analysis and Classification
Dissertation directed by Professor Jorge L. Aravena
Pages in dissertation, 114. Words in abstract, 220.

ABSTRACT

This research focuses on the analysis and classification of multicomponent nonstationary signals of arbitrary duration. The proposed classification approach has potential applications in areas like moving target detection, object recognition, oil exploration, and speech processing. The wavelet transform is used as the basis for the analysis. The classification technique is based on novel scale energy density functions, called *pseudo power signatures*, which are independent of signal length, and which can be used to characterize the time-scale energy distribution of the signal. These signatures allow for fast classification of signals regardless of their length.

Two approaches to determine pseudo power signatures are presented in this work. The first approach is based on a singular value principal component analysis technique, which, though computationally simple, is not very sensitive to signal characteristics. The second is a more sophisticated approach, and is optimal in a weighted least mean squares sense. The latter technique involves solving an inverse projection problem arising from a nonlinear infinite dimensional minimization, and generates good quality signatures with excellent discriminating capability. An algorithm, with fast convergence, for application to discrete data sets is developed, and a complete analysis of the computational complexity is obtained. Several simulation examples are presented to illustrate the methodology, and its application to practical classification problems. Finally, suggestions for further work in the area are given.