Innovations and Singular Value Decomposition for Blind Sequence Detection in Wireless Channels

by

Sujit Sen

A thesis submitted in conformity with the requirements
for the degree of Master of Applied Science
Graduate Department of Electrical and Computer Engineering
University of Toronto

©Copyright by Sujit Sen 1999

National Library
of Canada

Acquisitions and
Bibliographic Services

395 Wellington Street
Ottawa ON K1A 0N4
Canada

Bibliothèque nationale
du Canada

Acquisitions et
services bibliographiques

395, rue Wellington
Ottawa ON K1A 0N4
Canada

Your file Votre référence

Our file Notre référence

The author has granted a non-exclusive licence allowing the National Library of Canada to reproduce, loan, distribute or sell copies of this thesis in microform, paper or electronic formats.

The author retains ownership of the copyright in this thesis. Neither the thesis nor substantial extracts from it may be printed or otherwise reproduced without the author's permission.

L'auteur a accordé une licence non exclusive permettant à la Bibliothèque nationale du Canada de reproduire, prêter, distribuer ou vendre des copies de cette thèse sous la forme de microfiche/film, de reproduction sur papier ou sur format électronique.

L'auteur conserve la propriété du droit d'auteur qui protège cette thèse. Ni la thèse ni des extraits substantiels de celle-ci ne doivent être imprimés ou autrement reproduits sans son autorisation.

0-612-45997-7

Canada

Innovations and Singular Value Decomposition for Blind Sequence Detection in Wireless Channels

Sujit Sen

A thesis submitted in conformity with the requirements for the Degree of Master of Applied Science, Graduate Department of Electrical and Computer Engineering, in the University of Toronto, 1999

Abstract

A blind sequence detection (BSD) algorithm based on the innovations approach is proposed and its performance in a Rayleigh fading environment is evaluated. A comparison between the innovations and Tong's Singular Value Decomposition (SVD) [25] based blind sequence detection algorithm is also presented. Further insight is gained on how blind sequence detectors behave by examining several parameters such as differential encoding, signal subspace dimension d , correlation delay k and diversity. When diversity is not used, the innovations approach to blind sequence detection clearly outperforms the SVD based blind sequence detector. Overall though, the SVD based blind sequence detection algorithm has a better performance than the innovations approach. However, if one looks at the 'combined performance' and takes into account that the innovations approach is simple, causal, recursive and numerically efficient when compared to SVD, then the performances of the two blind sequence detectors are comparable.

Acknowledgements

I am deeply grateful to my supervisor, Prof. S. Pasupathy, for his guidance and support throughout the course of my graduate studies. His numerous comments and careful reading of the manuscript are greatly appreciated.

I gratefully acknowledge the financial support I received in the form of the University of Toronto Open Masters Fellowship and the research assistantship provided by Prof. S. Pasupathy.

Finally, I thank my parents for their patience, support and encouragement.

Contents

List of Figures	iv
1 Introduction	1
1.1 A Brief Survey	2
1.2 Thesis Organization	5
2 Innovations Processes, Karhunen-Loève Transform and Singular Value Decomposition	7
2.1 Innovations Processes	7
2.2 Karhunen-Loève Transform	11
2.3 Singular Value Decomposition	12
2.4 Discussions	15
2.5 Conclusion	17
3 An Innovations Based Blind Sequence Detector	19
3.1 System Model and Assumptions	19
3.2 Innovations and Channel Orthogonalization	22
3.3 Application of the Viterbi Algorithm	23
3.4 Optimal Estimators for Source Correlation	26
3.5 The Innovations - Blind Sequence Detection Matrix Equation: Solving for “A”	31

3.6	An Innovations Based Blind Sequence Detector: Outline and Implementation	37
3.7	Conclusion	39
4	Results	41
4.1	Simulation Setup	42
4.2	Differential Encoding	46
4.3	Signal Subspace Dimension d	48
4.4	Delay Parameter k	53
4.5	Diversity Reception	54
4.6	Innovations Based Blind Sequence Detector Filters: An Analysis . . .	56
4.7	Error Correcting Codes	58
4.8	Discussion	62
4.9	Conclusion	65
5	Conclusions	67
	Bibliography	69

List of Figures

1.1	A Digital Communication System	2
2.1	Innovations filter (L) and whitening filter (Γ) for $x(t)$	8
3.1	A Mobile Wireless Environment.	22
3.2	The state transition diagram.	25
3.3	The optimal path determined by the Viterbi Algorithm.	25
3.4	Innovations Based Blind Sequence Detector Algorithm	39
4.1	Real parts of the channel impulse responses for the three receivers. . .	45
4.2	Imaginary parts of the channel impulse responses for the three receivers.	45
4.3	Differential Encoding.	47
4.4	Differential Decoding.	47
4.5	Differential encoding for the Innovations Based Blind Sequence Detector with $d=3$, bit delay =2 and $k=1$	48
4.6	Differential encoding for the SVD Based Blind Sequence Detector with $d=2$, bit delay =2 and $k=1$	49
4.7	Varying bit delay for the Innovations Based Blind Sequence Detector with $d=3$ and $k=1$	49
4.8	Varying bit delay for the SVD Based Blind Sequence Detector with $d=2$ and $k=1$	50
4.9	Varying d for the Innovations Based Blind Sequence Detector with bit delay=1 and $k=1$	51

4.10 Varying d for the SVD Based Blind Sequence Detector with bit delay=1 and $k=1$	51
4.11 Varying k for the Innovations Based Blind Sequence Detector with $d=3$ and bit delay=1.	54
4.12 Varying k for the SVD Based Blind Sequence Detector with $d=2$ and bit delay=2.	55
4.13 Diversity Reception for the Innovations Based Blind Sequence Detector with $d=3$, $k=1$ and bit delay=2.	56
4.14 Diversity Reception for the SVD Based Blind Sequence Detector with $d=2$, $k=1$ and bit delay=2.	57
4.15 Proposed blind sequence detector filters with $d=3$, $k=1$ and bit delay =1.	58
4.16 Proposed blind sequence detector filters with $d=3$, $k=1$ and bit delay =1.	59
4.17 (7,4) Hamming Code for Innovations algorithm with $d=3, k=1$ and bit delay=1.	61
4.18 (7,4) Hamming Code for SVD algorithm with $d=2, k=1$ and bit delay=2.	61
4.19 Innovations based blind sequence detection algorithm with $d=3, k=1$ and bit delay=1.	64
4.20 Comparing innovations and SVD blind sequence detection algorithms with a diversity of one.	65

Chapter 1

Introduction

Blind equalization has become very important in the recovery of signals corrupted by channel disturbances. Essentially, blind equalization is where one estimates the channel without knowing what the transmission channel characteristics are. Thus, the goal is to separate $f(n)$ and $x(n)$ by observing $y(n)$ where

$$y(n) = f(n) * x(n) = \sum_k f(k)x(n-k) \quad n, k \in Z. \quad (1.1)$$

In blind equalization, both $f(n)$ (the channel function) and $x(n)$ are unknown or partially known. Receivers are implemented where they learn the channel based upon the signals that are received. This technique involves adaptive detectors where their job is to do "Channel Estimation" and "Data Recovery." Blind sequence detection is a new area that has emerged which is different from many traditional blind equalization methods since it does not require any channel identification at all. As such, an advantage of using blind sequence detection is that the training period of the transmission over an ISI (Intersymbol Interference) channel can be reduced or eliminated, which results in an increase in transmission efficiency. This is particularly attractive in rapidly varying mobile channels whose training signals must be transmitted often for conventional equalizers [25]. For example, the current GSM standard used in Europe has a 28 bits training sequence for every 116 information bits [7].

A significant amount of research and development work has been done in designing receivers to combat these disturbances in digital communication systems. In order to have high-speed reliable communication on a wireless channel, channel identification and equalization are required to combat ISI. Normally, channel identification and equalization are done either by sending training sequences or by designing the equalizer based upon prior knowledge of the channel. Adaptive detectors are used where they do this channel estimation. Unfortunately, in radio communication channels little is known about the channel *a priori*. Also, the standard adaptive detectors used in radio environments waste some of their transmission time on a training sequence. Recently though, blind sequence detection and innovations based estimation techniques have received a lot of attention in recovering a transmitted data sequence corrupted by an unknown channel. Furthermore, blind sequence detection has been found to be very effective in TDMA based wireless systems where short burst errors, that arise from the unknown channel, are very common [11]. This method is also very applicable to many narrow band modulation based mobile systems (i.e. TDMA or GSM) where Rayleigh fading is a characteristic of the channel. Rayleigh fading is a common occurrence for signal degradation. Rayleigh fading is mathematically modeled by multipath channels with randomly time-invariant impulse responses.

Figure 1.1: A Digital Communication System

1.1 A Brief Survey

There are many blind equalization techniques such as Bussgang methods, polyspectra methods, maximum likelihood methods, minimum entropy methods and cyclic spectrum methods. Sato [20] first introduced the concept of self-recovering blind equaliza-

tion. His method optimized certain criteria based on higher order statistics (HOS). Numerous other techniques (e.g. [5] [16] [23]) have also been proposed utilizing HOS. Methods that use higher order statistics of $y(n)$ can identify correctly non-minimum phase channels. Hatzinakos and Nikias [9] have also developed blind channel identification algorithms based on HOS. Unfortunately, simulations have shown that the convergence time (usually thousands of symbols) of these HOS algorithms is too large for mobile channels. Mobile channels require identification and equalization within about 100 symbols [25]. In addition, methods that use HOS are very complex in nature and require long data records. Second-order cyclostationary blind signal processing methods have been proposed ([28] [26] [27]) that have a much lower convergence rate (about 100 symbols). However, these second-order algorithms can not properly identify a channel when the channel has certain special zeros ([27]). Bell and Sejnowski [1] have recently developed an information maximization approach to blind deconvolution. Their technique is based essentially on some basic information theory principles.

Adaptive detectors are widely used in blind sequence detection for channel estimation, sequence detection and data recovery. There are many different philosophies behind adaptive detectors. Ghosh and Weber [4] have proposed a receiver that does joint data and channel estimation in a recursive manner. The data estimation part is accomplished by using maximum likelihood sequence estimation (Viterbi decoding) while the channel estimation is done by the least squares estimation technique. In addition, Ghosh and Weber employ a delayed decision feedback estimation (DDFSE) algorithm [3] to reduce the number of states in the Viterbi trellis in their maximum-likelihood blind equalization method. Seshadri [22] has developed a joint data and channel estimation algorithm using a fast blind trellis search technique. Even though the Viterbi based blind technique developed by Seshadri offers an optimal solution, the price paid is high computational complexity. Raheli *et al.* [17] discuss Per-Survivor Processing (PSP) algorithm which can be used to approximate MLSE algo-

rithms whenever unknown quantities exist that prevent the exact use of the standard Viterbi algorithm. The main principle behind PSP is that the data-aided estimation of unknown parameters can be embedded into the framework of the Viterbi algorithm. Unfortunately, the use of PSP results in a longer computation time. PSP can operate without any training sequences but only in the case where there are enough data for PSP to converge to the channel estimates [2].

Kim and Cox [11] have developed a blind sequence estimator that does not require any training sequences. Their estimator is based on PSP-MLSE where the initial channel estimates needed for the conventional MLSE are obtained blindly. This occurs at the instant when a burst arrives. This blind initial channel estimate aids PSP to converge quickly. An application for this estimator can be used in short burst TDMA formats. Tong [25] has developed a blind sequence estimation technique that is very suitable for use in mobile channels. Tong focuses on blind sequence estimation instead of channel identification. His technique is primarily based upon Singular Value Decomposition (SVD). The main idea is to estimate the source (deterministic) correlation from observation (without knowing the channel). The Viterbi algorithm is then used to reconstruct the input symbols. Tong has shown that estimates of the signal correlation function require only channel orthogonalization. This process is much simpler than estimating the channel. The blind sequence estimation method can be applied to both single and multiple receiver structures. Unfortunately, Tong did not do a thorough performance analysis of his blind sequence detection scheme. As a result, he did not study the various important blind sequence detection parameters such as signal subspace dimension d , correlation delay k , diversity etc.

The innovations process has found numerous applications in the area of communication and signal processing. The innovations approach is essentially a technique for decorrelating a random process. The innovations approach was first used by Kailath [10] for engineering purposes to detect a random signal in AWGN. Rangaswamy *et al.* [18] have developed a method to address the problems of multichannel signal

detection in additive correlated non-Gaussian noise using the innovations approach. Recently, excellent work has been done by Yu and Pasupathy [30] on which they have applied the innovations approach to Rayleigh fading channels, and have developed a general and practical MLSE (Maximum Likelihood Sequence Estimation) receiver which demodulates a received signal recursively in a non-coherent fashion. Their innovations-based MLSE algorithm is implemented by a bank of time-invariant finite impulse response (FIR) filters followed by squaring operations and a Viterbi processor. The Viterbi algorithm is used because it overcomes the exponential complexity of MLSE. The Karhunen-Loève Transform (KLT) is another technique for decorrelating a random process. Yu and Pasupathy [29] have shown that the innovations process and KLT are related by a linear transformation and hence are equivalent in decorrelating coloured noise. However, they chose to use the innovations approach because the causality and computational simplicity of the innovations approach are excellent properties for real time recursive algorithms. So far, in the current literature, nobody has studied the application of an innovations approach to blind sequence detection (especially in the case of fading channels) or compared it to the SVD based blind sequence detection algorithm proposed by Tong [25].

1.2 Thesis Organization

The objectives of this thesis are threefold. First, we seek a good understanding of the theories behind innovation processes, Karhunen-Loève transform (KLT) and Singular Value Decomposition (SVD), and also we must see how these three methods relate to one another. As mentioned earlier, Tong [25] has already developed a blind sequence detector based on SVD. The second objective would be to try a different alternative to blind sequence detection and thus design and implement a blind sequence detector based on the innovations approach for Rayleigh fading channels. One of the main motivations of using an innovations approach for blind sequence detection is that

it has many inherent advantages over SVD. For example, the innovations process is simple, causal and computationally efficient when compared to SVD. Finally, a comparison will be made between Tong's blind sequence detection method based on SVD and the proposed innovations based blind sequence detection algorithm. From this comparison, it is hoped that a better understanding will result of the relationship between these two methods.

In Chapter 2, we will examine three methods that, for discrete time processes, correspond to three kinds of matrix factorizations: the innovations approach; Karhunen-Loève transform (KLT); and Singular Value Decomposition (SVD). We will demonstrate that a transformation exists between the innovations approach and KLT. Furthermore, it will be shown that under certain conditions, KLT and SVD are equivalent. In Chapter 3, the innovations approach to blind sequence detection will be explored. An Innovations Based Blind Sequence Detector will be presented and derived mathematically. Testing and simulation results of the Innovations Based Blind Sequence Detector and Tong's Blind Sequence Detector will be explained and presented in Chapter 4. Furthermore, a comparison of these two methods will be presented. Finally, Chapter 5 will provide some conclusions and suggestions for further study.

Chapter 2

Innovations Processes, Karhunen-Loève Transform and Singular Value Decomposition

Innovations approach, Karhunen-Loève transform (KLT) and Singular Value Decomposition (SVD) are three different techniques that correspond to factoring a matrix. Innovations approach and KLT are two different methods for decorrelating a random process. The innovations approach is a special **LU** factorization of the autocorrelation matrix, while KLT is an eigenvector factorization. SVD is closely associated with eigenvector factorization. It will be shown that the innovations process and KLT can be related by a linear transformation, and hence both techniques are equivalent in decorrelating a discrete time random process. Furthermore, under special conditions, eigenvector factorization is equivalent to SVD.

2.1 Innovations Processes

The innovations process has found numerous applications in the area of communications and signal processing. The innovations approach is essentially a technique

for decorrelating a random process. For discrete time processes, the innovations approach is in fact a special **LU** factorization of the autocorrelation matrix. It is also equivalent to the Gram-Schmidt procedure and linear prediction [29]. A brief review of the innovations process will be done.

Let us consider the problem of representing a process $x(t)$ as the response of a causal and linear system, characterized by the causal impulse response $L(t, \tau)$, with a white noise input $y(t)$ which has the unit power spectrum density (PSD). Denote $\Gamma(t, \tau)$ as the inverse of $L(t, \tau)$. In addition, $\Gamma(t, \tau)$ is also a causal and linear system. A process $x(t)$ that can be represented as such is called regular; the white noise $y(t)$ is called the innovations process or innovations of $x(t)$ respectively [15]. From Figure 2.1 it can be seen that a regular process $x(t)$ and its innovations $y(t)$ are causally and linearly equivalent. Thus, from a statistical point of view, the same statistical inferences such as maximum likelihood estimation can be obtained through either processes [30]. The innovations $y(t)$ is preferred for deriving new structures or algorithms since its samples are uncorrelated and easy to manipulate.

Figure 2.1: Innovations filter (L) and whitening filter (Γ) for $x(t)$

We will now focus on a discrete time whitening filter since it is very hard to determine the innovations for a continuous time process. Given a regular discrete time process $\{x_n\}_1^N$ we want to find its innovations process $\{y_n\}_1^N$. For simplicity, we will assume that $\{x_n\}_1^N$ and $\{y_n\}_1^N$ are real random processes since it is trivial to extend the following results to complex-valued random processes. By definition, $\{y_n\}_1^N$ and $\{x_n\}_1^N$, are causally and linearly equivalent and $\{y_n\}_1^N$ is a white noise process. The importance of having a causal system is that it is physically realizable. A causal system is where $h(n) = 0$ for $n < 0$. Since the innovations approach only requires past data samples and not future ones, it is said to be causal and thus

can be implemented in real time applications. We need to find a set of coefficients $\{\gamma_n\}_{1 \leq j \leq i \leq N}$ such that

$$\begin{aligned} y_1 &= \gamma_{11}x_1 \\ &\vdots \\ y_N &= \gamma_{N1}x_1 + \gamma_{N2}x_2 + \dots + \gamma_{NN}x_N \end{aligned} \quad (2.1)$$

and $E(y_i y_j) = \delta_{ij}$. Using matrix notation, (2.1) can be shown as

$$\mathbf{y} = \mathbf{\Gamma} \mathbf{x} \quad (2.2)$$

where $y = (y_1, \dots, y_N)^t$, $x = (x_1, \dots, x_N)^t$ and

$$\mathbf{\Gamma} = \begin{pmatrix} \gamma_{11} & & & \\ \gamma_{21} & \gamma_{22} & & \\ \vdots & & & \\ \gamma_{N1} & \dots & \dots & \gamma_{NN} \end{pmatrix}. \quad (2.3)$$

Thus, $\mathbf{\Gamma}$ is the whitening filter and its inverse $\mathbf{\Gamma}^{-1} = \mathbf{L}$ is the innovations filter given by

$$\mathbf{L} = \mathbf{\Gamma}^{-1} = \begin{pmatrix} l_{11} & & & \\ l_{21} & l_{22} & & \\ \vdots & & & \\ l_{N1} & \dots & \dots & l_{NN} \end{pmatrix}. \quad (2.4)$$

Thus, the innovations process $\{y_n\}_1^N$ of $\{x_n\}_1^N$ is calculated as the response of the causal and linear system $\mathbf{\Gamma}$ with input $\{x_n\}_1^N$. Let the autocorrelation matrix of $\mathbf{x}$ be

$$\mathbf{R}_x = E(\mathbf{x}\mathbf{x}^t) = \begin{pmatrix} R_{11} & R_{12} & \dots & R_{1N} \\ R_{21} & R_{22} & \dots & R_{2N} \\ \vdots & & & \\ R_{N1} & \dots & \dots & R_{NN} \end{pmatrix} \quad (2.5)$$

where $E()$ denotes expectation.

For any matrix $\mathbf{R}$, if all its leading k by k principle submatrices are non singular, then there exists a unit lower triangular matrix $\mathbf{L}_u$ and an upper triangular matrix $\mathbf{U}$ such that $\mathbf{R} = \mathbf{L}_u \mathbf{U}$ [6]. This is called LU factorization of $\mathbf{R}$. If $\mathbf{R}$ is a symmetric matrix, then $\mathbf{R} = \mathbf{L}_u \mathbf{D} \mathbf{L}_u^t$, where $\mathbf{D}$ is a diagonal matrix. This is known as LDL^t factorization of symmetric matrix [8]. If every element of $\mathbf{D}$ is non-negative (or $\mathbf{R}$ is positive semidefinite), $\mathbf{R} = \mathbf{L} \mathbf{L}^t$, where $\mathbf{L}$ is no longer a unit lower triangular matrix. The product $\mathbf{L} \mathbf{L}^t$ is known as Cholesky factorization [15]. The whitening filter $\mathbf{\Gamma}$ is such a lower triangular matrix that transforms the autocorrelation matrix $\mathbf{R}_x$ into the identity as shown below:

$$E(\mathbf{y} \mathbf{y}^t) = E(\mathbf{\Gamma} \mathbf{x} \mathbf{x}^t \mathbf{\Gamma}^t) = \mathbf{\Gamma} \mathbf{R}_x \mathbf{\Gamma}^t = \mathbf{I}. \quad (2.6)$$

Thus, we can express

$$\mathbf{R}_x = \mathbf{L} \mathbf{L}^t \quad (2.7)$$

or

$$\mathbf{R}_x^{-1} = \mathbf{\Gamma}^t \mathbf{\Gamma}. \quad (2.8)$$

This is the Cholesky factorization of $\mathbf{R}_x$. As mentioned earlier, innovations and linear prediction are mathematically equivalent. Innovations can be defined as a series of normalized linear prediction errors [29]. Hence, the Cholesky factorization of the autocorrelation matrix $\mathbf{R}_x$ can also be expressed as

$$\mathbf{R}_x = \mathbf{P}^{-1} \text{diag}(d_1, \dots, d_N) (\mathbf{P}^t)^{-1} = \mathbf{P}^{-1} \mathbf{D} (\mathbf{P}^t)^{-1} \quad (2.9)$$

where $\mathbf{P}$ is a lower triangular matrix composed of all orders of prediction coefficients of $x(t)$ and $\mathbf{D}$ is a diagonal matrix composed of variances of corresponding prediction errors. In the complex domain, we can write this as

$$\mathbf{R}_x = \mathbf{P}^{-1} \text{diag}(d_1, \dots, d_N) (\mathbf{P}^*)^{-1} = \mathbf{P}^{-1} \mathbf{D} (\mathbf{P}^*)^{-1} \quad (2.10)$$

where $*$ denotes complex conjugate transpose. The causal whitening filter can now be written as

$$\mathbf{\Gamma} = \mathbf{D}^{-1/2} \mathbf{P}. \quad (2.11)$$

2.2 Karhunen-Loéve Transform

The Karhunen-Loéve Transform (KLT) is a well known technique for providing a second moment characterization of a random process in terms of uncorrelated random variables [29] [15]. We will provide a brief review of this technique. Suppose we have a real-valued random process $x(t)$, $0 \leq t \leq T$ with zero mean. Define the autocorrelation of this process $x(t)$ to be

$$R_x(t, \tau) = E[x(t)x(\tau)]. \quad (2.12)$$

The KLT is determined by projecting the random process onto an orthogonal basis. This is similar to the Fourier Transform. The basis for the KLT is selected by solving the *eigenequations* or *integral equations* [15],

$$\int_0^T R_x(t, \tau) \varphi_i(\tau) d\tau = \lambda_i \varphi_i(t) \quad 0 < t < T \quad (2.13)$$

where $\varphi_i(t)$ is a set of orthonormal functions in the interval $(0, T)$

$$\int_0^T \varphi_i(t) \varphi_j(t) dt = \delta(i - j) \quad (2.14)$$

and λ_i 's are eigenvalues. From the KLT, the process $x(t)$ can be represented as

$$x(t) = \lim_{N \rightarrow \infty} \sum_{i=1}^N c_i \varphi_i(t). \quad (2.15)$$

The coefficients c_i is a set of uncorrelated random variables given by

$$c_i = \int_0^T x(t) \varphi_i(t) dt. \quad (2.16)$$

For the discrete-time case, the KLT is essentially the eigenvector factorization of the autocorrelation matrix $\mathbf{R}_x$ of a given zero-mean process $\mathbf{x} \in R^N$. Hence the normalized KLT of $\mathbf{x}$ is given by

$$\mathbf{w} = \mathbf{\Lambda}^{-1/2} \mathbf{U}^t \mathbf{x} \quad (2.17)$$

where $\mathbf{U}$ is an orthogonal (unitary for complex process) matrix comprised of N eigenvectors of $\mathbf{R}_x$ (please note that this $\mathbf{U}$ is not the same $\mathbf{U}$ used in section 2.1); $\mathbf{w}$ is a

white noise process; and $\mathbf{\Lambda} = \text{diag}(\lambda_1, \dots, \lambda_N)$ is a diagonal matrix consisting of the eigenvalues of $\mathbf{R}_x$. The autocorrelation matrix $\mathbf{R}_x$ can be factorized as

$$\mathbf{R}_x = \mathbf{U}\mathbf{\Lambda}\mathbf{U}^t = \mathbf{Q}^{-1}\mathbf{Q}^{-t} \quad (2.18)$$

where $\mathbf{Q} = \mathbf{\Lambda}^{-1/2}\mathbf{U}^t$. This type of matrix factorization is known as eigenvector factorization or decomposition.

Hence, in order to calculate the KLT of a given discrete time process $\mathbf{x}$, all we need to do is to find all the eigenvectors and eigenvalues of the autocorrelation matrix $\mathbf{R}_x$. From (2.16), we see that $\mathbf{x}$ and $\mathbf{w}$ are linearly and invertibly dependent and thus are informationally equivalent. This was the same case for the innovations process. However, $\mathbf{Q}$ is not a lower triangular matrix and as such there is not a causal relationship between $\mathbf{x}$ and $\mathbf{w}$. Therefore, by definition, $\mathbf{w}$ is not the innovations of $\mathbf{x}$ [30].

2.3 Singular Value Decomposition

Singular Value Decomposition (SVD) is closely associated with eigenvector factorization of a symmetric matrix $\mathbf{R}_x$: $\mathbf{R}_x = \mathbf{U}\mathbf{\Lambda}\mathbf{U}^t$ where $\mathbf{\Lambda}$ is a diagonal matrix comprising of the eigenvalues of $\mathbf{R}_x$ and $\mathbf{U}$ is an orthogonal matrix consisting of the eigenvectors of $\mathbf{R}_x$. Recall that an orthonormal matrix has the following property: $\mathbf{U}^t\mathbf{U} = \mathbf{I}$. Unfortunately, eigenvector factorization can not be done for many matrices such as a non-symmetric matrix [24]. However, by letting $\mathbf{U}$ on the left and $\mathbf{U}^t$ on the right be any two orthogonal matrices, another factorization can be used which is known as SVD. In addition, the diagonal matrix can be made nonnegative. We denote this matrix as $\mathbf{\Psi}$ and its positive entries as ψ_i which are known as the *singular values* of $\mathbf{R}_x$. The number of singular values, r , is equivalent to the rank of $\mathbf{R}_x$ [24]. A formal definition for the SVD will now be stated:

Singular Value Decomposition 1 Any m by n matrix $\mathbf{R}_x$ can be factored into:

$$\mathbf{R}_x = \mathbf{W}\Psi\mathbf{V}^t = (\text{orthogonal})(\text{diagonal})(\text{orthogonal}). \quad (2.19)$$

$\mathbf{W}$ and $\mathbf{V}$ are orthogonal matrices and the columns of $\mathbf{W}$ (m by m) are the eigenvectors of $\mathbf{R}_x\mathbf{R}_x^t$ and the columns of $\mathbf{V}$ (n by n) are eigenvectors of $\mathbf{R}_x^t\mathbf{R}_x$. The r singular values on the diagonal of Ψ (m by n) are the square roots of the non zero eigenvalues of both $\mathbf{R}_x\mathbf{R}_x^t$ and $\mathbf{R}_x^t\mathbf{R}_x$ [24].

Eigenvalue decompositions are suitable for analyzing a matrix when it is a mapping from a vector space into itself. An example of this is Ordinary Differential Equations. The SVD is suitable for analyzing a mapping from one vector space into another vector space. If $\mathbf{R}_x$ is a square, symmetric and positive definite matrix then SVD and eigenvector factorization are equivalent (i.e. $\mathbf{W}\Psi\mathbf{V}^t = \mathbf{U}\mathbf{\Lambda}\mathbf{U}^t$ [24]). As $\mathbf{R}_x$ departs from symmetry, the difference between the two factorizations increases. The SVD of a real matrix is always real but the eigenvector decomposition of a real non-symmetric matrix might be complex. For indefinite matrices, any negative value in $\mathbf{\Lambda}$ become positive in Ψ and the matrix $\mathbf{W}$ is different from $\mathbf{V}$. In the case of complex matrices, Ψ always remains real but $\mathbf{W}$ and $\mathbf{V}$ become unitary (the complex analogue of orthogonal), $\mathbf{W}\Psi\mathbf{V}^*$.

There are numerous applications for SVD. Image processing is one such area where SVD is being extensively used. Imagine that a satellite takes a picture and wants to send it to earth. Assume the picture is represented by a 1000 by 1000 matrix pixels where we code colours in each pixel from a range of black to white. Thus, when we transmit this information we send out 1 000 000 numbers. It would be much better to just send the essential information in this 1000 by 1000 matrix. Suppose the SVD is known for this matrix. The key idea in order to send out only the essential information is to look at the singular values in Ψ . Some of these singular values are significant while the others are not. If we keep 60 "significant" values and get rid of the remaining 940, we only need to send the corresponding 60 columns of $\mathbf{W}$ and $\mathbf{V}$.

This is commonly referred to as *Signal Subspace Decomposition*. The remaining 940 columns multiplied in $\mathbf{W}\Psi\mathbf{V}^t$ are insignificant because they are multiplied by a small ψ . Since only 60 columns are kept, we send 60 X 2000 numbers instead of 1 000 000. Thus, the amount of information being sent is reduced without affecting the overall picture quality [24].

Tong makes use of SVD and *Signal Subspace Decomposition* in his blind sequence estimation algorithm [25]. In his paper, the SVD is done on the autocorrelation matrix $\mathbf{R}_x(0)$ in order to find the Mahalanobis Transform. Tong uses the Mahalanobis Transform as a motivation for finding an optimal estimator in his blind sequence estimation method. The autocorrelation matrix for a vector process $\mathbf{x}(t)$ can be defined as:

$$\mathbf{R}_x(0) = E[\mathbf{x}(t)\mathbf{x}^*(t-k)]. \quad (2.20)$$

This is how the autocorrelation matrix looks like for the model used in Tong's blind sequence estimation [25]:

$$\mathbf{R}_x(0) = \mathbf{H}\mathbf{H}^* + \sigma^2\mathbf{I}. \quad (2.21)$$

Here, $\mathbf{H}$ is the fading channel matrix used in Tong's communication channel model and σ^2 is the variance of the additive white Gaussian noise. A full description of this model will be presented in Chapter 3. The SVD of $\mathbf{R}_x(0)$ can be expressed as

$$\mathbf{R}_x(0) = \mathbf{V}diag(\psi_1^2, \dots, \psi_N^2)\mathbf{V}^* \quad (2.22)$$

$$\mathbf{R}_x(0) = \mathbf{V}_sdiag(\lambda_1^2, \dots, \lambda_d^2)\mathbf{V}_s^* + \sigma^2\mathbf{I} \quad (2.23)$$

where $\mathbf{V}_s$ is the submatrix consisting of the first d columns of $\mathbf{V}$ and $\psi_r^2 = \lambda_r^2 + \sigma^2$. Here, a *Signal Subspace Decomposition* has been done using the SVD where only the essential information of the signal is being used and the rest of the information (i.e. noise) is being discarded. The Mahalanobis Transform [25] can then be found from the SVD of $\mathbf{R}_x(0)$ as follows:

$$\mathbf{T}_m = \Psi_s^{-1/2}\mathbf{V}_s^* \quad (2.24)$$

$$\Psi_s = diag(\lambda_1, \dots, \lambda_d). \quad (2.25)$$

The Mahalanobis Transform is used to orthogonalize the channel matrix $\mathbf{H}$ so that the source correlation is preserved and thus the input symbols can be reconstructed by using the Viterbi Algorithm.

2.4 Discussions

We have discussed three kinds of techniques to decompose a matrix, that is, Cholesky and eigenvector factorization and Singular Value Decomposition. The first two factorizations correspond to two kinds of uncorrelated representations of a random process namely innovations and KLT. The Cholesky decomposition results in a lower triangular matrix $\mathbf{\Gamma}$ with $N(N + 1)/2$ elements while eigenvector factorization has N^2 elements. Furthermore, innovations is a causal system which is useful in recursive algorithms and has about half the number of coefficients when compared to KLT [29]. Triangular factorization is simpler and less expensive than eigenvector factorization and as a result the innovations process is much more computationally efficient than KLT [8].

Yu and Pasupathy [29] have shown that a linear transformation exist between the innovations process and KLT. Unfortunately, they also found out that the KLT can not be calculated by recursive algorithms for innovations process. The following will be a brief summary of their findings. For a regular N -dimensional process $\mathbf{x}$, its innovations $\mathbf{y}$ is calculated as

$$\mathbf{y} = \mathbf{D}^{-1/2}\mathbf{e} \quad (2.26)$$

$$\mathbf{e} = \mathbf{P}\mathbf{x} \quad (2.27)$$

where $\mathbf{P}$ diagonalizes the autocorrelation matrix as shown below:

$$\mathbf{P}\mathbf{R}_{\mathbf{x}}\mathbf{P}^* = \text{diag}(d_1, \dots, d_N) = \mathbf{D}. \quad (2.28)$$

The KLT of $\mathbf{x}$ can be written as

$$\mathbf{v} = \mathbf{U}^*\mathbf{x} \quad (2.29)$$

where $\mathbf{U}$ is a unitary matrix which also diagonalizes $\mathbf{R}_x$:

$$\mathbf{U}\mathbf{R}_x\mathbf{U}^* = \text{diag}(\lambda_1, \dots, \lambda_N) = \mathbf{\Lambda}. \quad (2.30)$$

A linear relationship can be expected between $\mathbf{e}$ and $\mathbf{v}$ since they are both linear transformations of $\mathbf{x}$. Thus we can write

$$\mathbf{e} = \mathbf{\Phi}\mathbf{v} \quad (2.31)$$

where

$$\mathbf{\Phi} = \begin{pmatrix} \phi_{11} & \phi_{12} & \dots & \phi_{1N} \\ \phi_{21} & \phi_{22} & \dots & \phi_{2N} \\ \dots & \dots & \dots & \dots \\ \phi_{N1} & \dots & \dots & \phi_{NN} \end{pmatrix}. \quad (2.32)$$

In order to find $\mathbf{\Phi}$ we observe from (2.26), (2.28) and (2.30) that

$$\mathbf{P} = \mathbf{\Phi}\mathbf{U}^* \quad (2.33)$$

so

$$\mathbf{\Phi} = \mathbf{P}\mathbf{U} \quad (2.34)$$

or

$$\mathbf{\Phi}^* = \mathbf{U}^*\mathbf{P}^*. \quad (2.35)$$

Thus, $\mathbf{\Phi}$ is determined by $\mathbf{U}$ and $\mathbf{P}$. The transformation between $\mathbf{D}$ and $\mathbf{\Gamma}$ is also determined by $\mathbf{\Phi}$ as shown below:

$$\mathbf{D} = \mathbf{\Phi}\mathbf{\Lambda}\mathbf{\Phi}^*. \quad (2.36)$$

Hence, $\mathbf{\Phi}$ describes a relationship between the KLT and innovations. If $\mathbf{\Phi}$ is known, then the KLT can be determined from the innovations process and vice versa. However, we must know both $\mathbf{U}$ and $\mathbf{P}$ in order to calculate $\mathbf{\Phi}$. This implies that the KLT can not be calculated by using the recursive algorithms for innovations process [29]. Fortunately though, $\mathbf{\Phi}$ tells us that a theoretical relationship exist between the KLT and innovations.

From SVD one can obtain the Mahalanobis Transform and from eigenvector factorization one can get the KLT. Recall that if a matrix is square, symmetric and positive definite (which is the case for the autocorrelation function $\mathbf{R}_{\mathbf{x}}$ of a regular process $\mathbf{x}$) then SVD and eigenvector factorization are equivalent [24]. Hence, the KLT and Mahalanobis Transform will be the same where

$$\mathbf{T}_{KLT} = \mathbf{\Lambda}^{-1/2} \mathbf{U}^* = \mathbf{T}_m = \mathbf{\Psi}_s^{-1/2} \mathbf{V}_s^*. \quad (2.37)$$

As mentioned earlier, a linear relationship exist between innovations and KLT. Under special conditions where the matrix is square, symmetric and positive definite, KLT and SVD are the same. We can conclude that a relationship then exists between innovations and SVD. Furthermore, we can state that a transformation must exist between innovations and the Mahalanobis Transform. This is true because the Mahalanobis Transform and KLT are the same under the special conditions mentioned above and thus the arguments used to show that KLT and innovations are related can also be applied to show that the same relationship holds between the Mahalanobis Transform and innovations. Tong uses the Mahalanobis Transform in his blind sequence estimation algorithm. Perhaps innovations can also be used since the two techniques have been shown to have a theoretical relationship.

2.5 Conclusion

We have shown in this Chapter that a theoretical relationship exists between the KLT and innovations. However, we can not use the recursive algorithms for innovations to calculate the KLT. If a matrix is square, symmetric and positive definite it was found that SVD and eigenvector factorization are the same. Furthermore, it was determined that under these same conditions, the KLT and the Mahalanobis Transform are equivalent. Based on this equivalences, it was argued that a relationship exists between the innovations and the Mahalanobis Transform. We conclude that the innovations approach can be used as motivation for deriving a blind sequence estimation

method for fading channels similar to Tong's blind sequence estimation technique. In the following Chapter, we will see that a blind sequence estimation algorithm can indeed be based on the innovations process.

Chapter 3

An Innovations Based Blind Sequence Detector

There appears to be a connection between the work done by Tong [25] and by Yu and Pasupathy [29]. In the previous Chapter, it was shown that a relationship exists between innovations and SVD. In this Chapter, a new approach towards blind sequence detection is proposed based on innovations. This proposed approach does not require any channel identification at all. The fundamental idea behind this methodology is to estimate the source (deterministic) correlation without knowing the channel characteristics, and then to use the Viterbi Algorithm to obtain the input symbols. We will also show that, in order to estimate the signal correlation function, one only has to orthogonalize the channel. The innovations approach is used as a motivation for this orthogonalization.

3.1 System Model and Assumptions

In this section we will outline the system model and parameters used to develop our Innovations Based Blind Sequence Detector. A multipath Rayleigh fading channel is often used to represent the channel characteristic between a transmitter and the i th

mobile receiver in a wireless environment. The multipath Rayleigh fading channel envelope impulse response can be written as

$$c_i(t) = \sum_{k=1}^K \alpha_{ik} \delta(t - \tau_{ik}) \quad (3.1)$$

where K is the number of multipaths, $\alpha_{ik}(t)$ are zero-mean Gaussian processes and τ_{ik} are the associated path delays. The subscript “ i ” denotes the i th receiver in the system. The received baseband signal at the i th receiver can be written as

$$x_i(t) = \sum_n s_n p(t - nT_s) * c_i(t) + n_i(t) \quad (3.2)$$

$$= \sum_n s_n h_i(t - nT_s) + n_i(t) \quad (3.3)$$

where s_n is the symbol sequence, T_s is the symbol interval, $p(t)$ is the impulse response of the pulse shaping filter and $h_i(t)$ is the “composite” channel which includes the channel, pulse shaping and receiver filter and $n_i(t)$ is AWGN. Assume $x(t)$ is bandlimited and sampled at $t = l\Delta$, where $1/\Delta$ is at least twice the bandwidth of $x(t)$. The discrete time model can be expressed as

$$x_i(l) = \sum_n s_n h_i(l - nN) + n_i(l) \quad (3.4)$$

where $N = T_s/\Delta$ is an integer.

In this model, we use oversampling and not symbol rate sampling. Symbol rate sampling is used for a system that utilizes match filters. In order to use a match filter, one must have knowledge of the channel. However, in blind sequence detection the channel is unknown and hence symbol rate sampling can not be used. Oversampling is used when the channel is unknown in order to get sufficient statistics.

Assume $h(l)$ last d symbol intervals or in otherwords $h(l) = 0$ for $l < 0$ and $l > dN$. Thus, the received signal can be written in the following vector representation:

$$\mathbf{x}_i(t) = [x_i(tN), \dots, x_i((t+1)N-1)]^t \quad (3.5)$$

$$\mathbf{n}_i(t) = [n_i(tN), \dots, n_i((t+1)N-1)]^t \quad (3.6)$$

$$\mathbf{h}_i(t) = [h_i(tN), \dots, h_i((t+1)N-1)]^t \quad (3.7)$$

$$\mathbf{H}_i = [\mathbf{h}_i(0), \dots, \mathbf{h}_i(d-1)]^t \quad (3.8)$$

$$\mathbf{s}(t) = [s_t, s_{t-1}, \dots, s_{t-d+1}]^t. \quad (3.9)$$

Therefore, we have

$$\mathbf{x}_i(t) = \mathbf{H}_i \mathbf{s}(t) + \mathbf{n}(t). \quad (3.10)$$

If we have L receivers, we can put all the information from the L receivers into a single form as shown below

$$\mathbf{x}(t) = \begin{pmatrix} \mathbf{x}_1(t) \\ \vdots \\ \mathbf{x}_L(t) \end{pmatrix} \in C^{N \times 1} \quad (3.11)$$

$$\mathbf{n}(t) = \begin{pmatrix} \mathbf{n}_1(t) \\ \vdots \\ \mathbf{n}_L(t) \end{pmatrix} \in C^{N \times 1} \quad (3.12)$$

$$\mathbf{H}(t) = \begin{pmatrix} \mathbf{H}_1(t) \\ \vdots \\ \mathbf{H}_L(t) \end{pmatrix} \in C^{N \times d}. \quad (3.13)$$

In vector form this reduces down to

$$\mathbf{x}(t) = \mathbf{H} \mathbf{s}(t) + \mathbf{n}(t). \quad (3.14)$$

The main idea behind blind sequence detection is to detect the information symbols s_n without knowing explicitly what the channel parameter $\mathbf{H}$ is.

The following assumptions will be made in our model:

(A1) The information symbol sequence s_n is zero mean and $E(s_i s_j^*) = \delta(i - j)$.

(A2) Noise power $n_j()$ is zero mean for all j and $E(n_i(t_1) n_j^*(t_2)) = \sigma^2 \delta(i - j) \delta(t_1 - t_2)$.

(A3) The noise process is uncorrelated with $\{s_n\}$.

(A4) Channel identification matrix $\mathbf{H}$ is an N by d matrix with full column rank.

The first three assumptions are common. The last assumption implies that all channels have finite impulse responses and hence the fading matrix is of finite dimension. For wireless communication channels this is acceptable [25]. Tong has shown that the full rank condition is necessary and sufficient for the channel to be identifiable using only second order statistics [26] [28].

Figure 3.1: A Mobile Wireless Environment.

3.2 Innovations and Channel Orthogonalization

The innovations process plays an important role in our blind sequence detector. The main idea in our approach is to orthogonalize the channels using only the observation data. Let

$$\mathbf{R}_x(k) = E(\mathbf{x}(t)\mathbf{x}^*(t-k)) \quad (3.15)$$

and from (3.15) and assumptions A1 to A4 we get

$$\mathbf{R}_x(0) = \mathbf{H}\mathbf{H}^* + \sigma^2\mathbf{I}. \quad (3.16)$$

By the innovations approach or Cholesky factorization [15] $\mathbf{R}_x(0)$ can be written as

$$\mathbf{R}_x(0) = \mathbf{P}^{-1}\mathbf{D}(\mathbf{P}^*)^{-1} \quad (3.17)$$

where $\mathbf{P}$ is a lower triangle matrix composed of all orders of prediction coefficients of $\mathbf{x}(t)$ and $\mathbf{D}$ is a diagonal matrix where the diagonal entries are the variances of the corresponding prediction errors. If there is no noise in (3.16) ($\sigma^2 = 0$) we then have

$$\mathbf{H}\mathbf{H}^* = \mathbf{P}^{-1}\mathbf{D}(\mathbf{P}^*)^{-1} \quad (3.18)$$

$$= \mathbf{P}^{-1}\mathbf{D}^{1/2}\mathbf{D}^{1/2}(\mathbf{P}^*)^{-1} \quad (3.19)$$

$$(\mathbf{P}^{-1}\mathbf{D}^{1/2})^{-1}\mathbf{H}\mathbf{H}^* = \mathbf{D}^{1/2}(\mathbf{P}^*)^{-1} \quad (3.20)$$

$$(\mathbf{P}^{-1}\mathbf{D}^{1/2})^{-1}\mathbf{H}\mathbf{H}^*(\mathbf{D}^{1/2}(\mathbf{P}^*)^{-1})^{-1} = \mathbf{I} \quad (3.21)$$

$$\mathbf{D}^{-1/2}\mathbf{P}\mathbf{H}\mathbf{H}^*\mathbf{P}^*\mathbf{D}^{-1/2} = \mathbf{I} \quad (3.22)$$

$$(\mathbf{D}^{-1/2}\mathbf{P}\mathbf{H})(\mathbf{D}^{-1/2}\mathbf{P}\mathbf{H})^* = \mathbf{I}. \quad (3.23)$$

Therefore, the transform $\mathbf{\Gamma} = \mathbf{D}^{-1/2}\mathbf{P}$ orthogonalizes the channel $\mathbf{H}$. In other-words, there is an orthogonal matrix $\mathbf{V}$ such that $\mathbf{\Gamma}\mathbf{H} = \mathbf{V}$. The matrix $\mathbf{\Gamma}$ is called the whitening filter and its inverse $\mathbf{L} = \mathbf{\Gamma}^{-1}$ is called the innovations filter of $\mathbf{x}(t)$. When there is no noise, the (deterministic) correlation of the orthogonalized observation

$$\mathbf{y}(t) = \mathbf{\Gamma}\mathbf{x}(t) = \mathbf{V}\mathbf{s}(t) \quad (3.24)$$

preserves the (deterministic) correlation of the source [25]

$$\mathbf{y}^*(t)\mathbf{y}(t-k) = \mathbf{s}^*(t)\mathbf{s}(t-k). \quad (3.25)$$

Thus, the correlation (deterministic) function of $\mathbf{s}(t)$ can be recovered from the received samples without knowing the channel $\mathbf{H}$.

3.3 Application of the Viterbi Algorithm

The important inner product preserving property (3.25) is the key point of the blind sequence detection algorithm. The Viterbi Algorithm can be used for a noisy channel to calculate the estimated sequence which results in the minimum correlation

(deterministic) difference between the Γ -transformed received sample $\mathbf{y}(t)$ and the estimated sequences. Denoting

$$r_y^{(k)}(t) = \mathbf{y}^*(t)\mathbf{y}(t-k) \quad (3.26)$$

$$r_s^{(k)}(t) = \mathbf{s}^*(t)\mathbf{s}(t-k) \quad (3.27)$$

$$= \sum_{l=0}^{d-1} s_{t-l}^* s_{t-l-k} \quad (3.28)$$

one obtains

$$r_y^{(k)}(t) = r_s^{(k)}(t) + w^{(k)}(t) \quad (3.29)$$

where $w^{(k)}(t)$ represents the noise

$$w^{(k)}(t) = \mathbf{s}^*(t)\mathbf{H}\Gamma^*\Gamma\mathbf{n}(t-k) + \mathbf{n}^*(t)\Gamma^*\Gamma\mathbf{H}\mathbf{s}(t-k) + \mathbf{n}^*(t)\Gamma^*\Gamma\mathbf{n}(t-k). \quad (3.30)$$

An optimum sequence detection can be defined as [25]:

$$\min_t \sum_l |r_y^{(k)}(t) - r_s^{(k)}(t)|^2 \quad (3.31)$$

which can be solved by applying the Viterbi Algorithm to a K^{d+k-1} state trellis for a K-QAM signalling. The delay parameter k should be selected in such a fashion so that the number of states in the Viterbi Algorithm is minimum. For $k=0$, $\{s_k\}$ cannot be uniquely determined from $r_y^{(0)}(t)$ since each of the symbols can have a random phase without affecting $r_y^{(0)}(t)$ [25].

We will now consider an example of blind sequence detection for a BPSK source with channel dimension $d=2$. With this example we can write

$$r_y^{(1)}(t) = s_t s_{t-1} + s_{t-1} s_{t-2} + w^{(1)}(t). \quad (3.32)$$

The state δ_n will be defined as

$$\delta_n = (s_{n-1}, s_n). \quad (3.33)$$

Figure (3.1) displays the state transition diagram. In Figure (3.2), a typical sequence is shown where $s_n = \{-1, -1, 1, -1, 1, 1\}$. Its correlation sequence $r_y^{(1)}(t)$ is the estimate of $r_s^{(1)}(t)$ and the corresponding optimal path is determined by the Viterbi Algorithm.

Figure 3.2: The state transition diagram.

Figure 3.3: The optimal path determined by the Viterbi Algorithm.

3.4 Optimal Estimators for Source Correlation

The innovations process only allows us to find the source correlation of $\mathbf{s}(t)$ from the correlation of $\mathbf{y}(t)$ in the absence of noise. In many communication systems, noise is a common phenomena and must be dealt with. Unfortunately, the whitening filter Γ is not optimum in the presence of noise.

The main goal of the Innovations Based Blind Sequence Detector is to obtain an optimum estimation of the source correlation

$$r_s(t) = \mathbf{s}^*(t)\mathbf{s}(t-1). \quad (3.34)$$

The Viterbi Algorithm can then be used with the estimated $r_s(t)$ to recover the input symbols. We can use the innovations process as a *motivation* to estimate $r_s(t)$ from the correlation function of the transformed observations. Define

$$\mathbf{y}(t) = \Gamma_o \mathbf{x}(t) \quad (3.35)$$

$$r_y(t) = \mathbf{y}^*(t)\mathbf{y}(t-1). \quad (3.36)$$

The goal is to find an optimum filter Γ_o that minimizes

$$J(\Gamma) = E(|r_y^{(k)}(t) - r_s^{(k)}(t)|^2). \quad (3.37)$$

We will now mathematically derive $J(\Gamma)$. Recall that

$$\mathbf{x}(t) = \mathbf{H}\mathbf{s}(t) + \mathbf{n}(t) \quad (3.38)$$

$$\mathbf{y}(t) = \Gamma \mathbf{x}(t) \quad (3.39)$$

$$r_s = \mathbf{s}^*(t)\mathbf{s}(t-1) \quad (3.40)$$

$$r_y = \mathbf{y}^*(t)\mathbf{y}(t-1) \quad (3.41)$$

$$\mathbf{A} = \Gamma^* \Gamma \quad (3.42)$$

$$\mathbf{Q} = \mathbf{H}\mathbf{H}^* \quad (3.43)$$

$$J = E(|r_y - r_s|^2). \quad (3.44)$$

We have

$$J = E(|r_y - r_s|^2) \quad (3.45)$$

$$= E(|r_y|^2) - E(r_y r_s^*) - E(r_s r_y^*) + E(|r_s|^2). \quad (3.46)$$

We will now evaluate each term in (3.46).

Computation of $E(|r_y|^2)$: From (3.38), (3.39) and (3.42) we have

$$r_y = \mathbf{y}(t)^* \mathbf{y}(t-1) \quad (3.47)$$

$$= \mathbf{x}^*(t) \mathbf{A} \mathbf{x}(t-1) \quad (3.48)$$

$$= r_y^{(i)} + r_y^{(ii)} + r_y^{(iii)} + r_y^{(iv)} \quad (3.49)$$

where

$$r_y^{(i)} = \mathbf{s}^*(t) \mathbf{H}^* \mathbf{A} \mathbf{H} \mathbf{s}(t-1) \quad (3.50)$$

$$r_y^{(ii)} = \mathbf{n}^*(t) \mathbf{A} \mathbf{n}(t-1) \quad (3.51)$$

$$r_y^{(iii)} = \mathbf{s}^*(t) \mathbf{H}^* \mathbf{A} \mathbf{H} \mathbf{n}(t-1) \quad (3.52)$$

$$r_y^{(iv)} = \mathbf{n}^*(t) \mathbf{A} \mathbf{H} \mathbf{s}(t-1). \quad (3.53)$$

In addition to the previous assumptions (A1 to A4), assume $\{s_k\}$ is i.i.d. and the noise $n_i(\cdot)$ is independent of $n_j(\cdot)$ and the source $\{s_k\}$ [25]. Taking these assumptions into consideration we can show that

$$E(|r_y|^2) = E(|r_y^{(i)}|^2) + E(|r_y^{(ii)}|^2) + E(|r_y^{(iii)}|^2) + E(|r_y^{(iv)}|^2). \quad (3.54)$$

We can write $E(|r_y^{(i)}|^2)$ as

$$E(|r_y^{(i)}|^2) = E(\mathbf{s}^*(t) \mathbf{H}^* \mathbf{A} \mathbf{H} \mathbf{s}(t-1) \mathbf{s}^*(t-1) \mathbf{H}^* \mathbf{A} \mathbf{H} \mathbf{s}(t)). \quad (3.55)$$

Identity 1 *If X is a $1 \times N$ matrix, Y is a $N \times N$ matrix and Z is a $N \times 1$ matrix then*

$$XYZ = \text{tr}(YZX). \quad (3.56)$$

Using this identity in (3.56) we can express (3.55) as

$$E(|r_y^{(i)}|^2) = \text{tr}(\mathbf{H}^* \mathbf{A} \mathbf{H} E(\mathbf{s}(t-1) \mathbf{s}^*(t-1) \mathbf{H}^* \mathbf{A} \mathbf{H} \mathbf{s}(t) \mathbf{s}^*(t))). \quad (3.57)$$

Let $\mathbf{\Pi} = \mathbf{H}^* \mathbf{A} \mathbf{H} = [\pi_{ij}]$, where π_{ij} is the (i, j) th component of $\mathbf{\Pi}$. We now have

$$E(|r_y^{(i)}|^2) = \text{tr}(\mathbf{\Pi}^* E(\mathbf{s}(t-1) \mathbf{s}^*(t-1) \mathbf{\Pi} \mathbf{s}(t) \mathbf{s}^*(t))). \quad (3.58)$$

It can be shown that the (i, j) th component of

$$\mathbf{s}(t-1) \mathbf{s}^*(t-1) \mathbf{\Pi} \mathbf{s}(t) \mathbf{s}^*(t) = s_{t-i} s_{t-j+1}^* \sum_{k,l} \pi_{k,l} s_{t-l+1} s_{t-k}^*. \quad (3.59)$$

Thus the (i, j) th component of $E(\mathbf{s}(t-1) \mathbf{s}^*(t-1) \mathbf{\Pi} \mathbf{s}(t) \mathbf{s}^*(t))$ is given by

$$E(s_{t-i} s_{t-j+1}^* \sum_{k,l} \pi_{k,l} s_{t-l+1} s_{t-k}^*) = \pi_{ij}. \quad (3.60)$$

Since s_k is an i.i.d. sequence the above equation holds. Therefore we can write

$$E(|r_y^{(i)}|^2) = \text{tr}(\mathbf{\Pi}^* E(\mathbf{s}(t-1) \mathbf{s}^*(t-1) \mathbf{\Pi} \mathbf{s}(t) \mathbf{s}^*(t))) \quad (3.61)$$

$$= \text{tr}(\mathbf{\Pi}^* \mathbf{\Pi}) \quad (3.62)$$

$$= \text{tr}(\mathbf{H}^* \mathbf{A}^* \mathbf{H} \mathbf{H}^* \mathbf{A} \mathbf{H}) \quad (3.63)$$

$$= \text{tr}(\mathbf{A}^* \mathbf{H} \mathbf{H}^* \mathbf{A} \mathbf{H} \mathbf{H}^*) \quad (3.64)$$

$$= \text{tr}(\mathbf{A}^* \mathbf{Q} \mathbf{A} \mathbf{Q}^*). \quad (3.65)$$

We will now determine $E(|r_y^{(ii)}|^2)$.

$$E(|r_y^{(ii)}|^2) = E(\mathbf{n}^*(t) \mathbf{A} \mathbf{n}(t-1) \mathbf{n}^*(t-1) \mathbf{A}^* \mathbf{n}(t)) \quad (3.66)$$

$$= \text{tr}(\mathbf{A} E(\mathbf{n}(t-1) \mathbf{n}^*(t-1) \mathbf{A}^* \mathbf{n}(t) \mathbf{n}^*(t))). \quad (3.67)$$

Since the noise is uncorrelated and white and using the same techniques as before we have

$$E(|r_y^{(ii)}|^2) = \sigma^4 \text{tr}(\mathbf{A} \mathbf{A}^*). \quad (3.68)$$

The remaining 3rd and 4th terms are evaluated in a similar fashion:

$$E(|r_y^{(iii)}|^2) = E(\mathbf{s}^*(t)\mathbf{H}^*\mathbf{A}\mathbf{n}(t-1)\mathbf{n}^*(t-1)\mathbf{A}^*\mathbf{H}\mathbf{s}(t)) \quad (3.69)$$

$$= \text{tr}(\mathbf{H}^*\mathbf{A}E(\mathbf{n}(t-1)\mathbf{n}^*(t-1))\mathbf{A}^*\mathbf{H}\mathbf{s}(t)\mathbf{s}^*(t)) \quad (3.70)$$

$$= \text{tr}(\mathbf{H}^*\mathbf{A}E(\mathbf{n}(t-1)\mathbf{n}^*(t-1))\mathbf{A}^*\mathbf{H}E(\mathbf{s}(t)\mathbf{s}^*(t))) \quad (3.71)$$

$$= \sigma^2 \text{tr}(\mathbf{H}^*\mathbf{A}\mathbf{A}^*\mathbf{H}) \quad (3.72)$$

$$= \sigma^2 \text{tr}(\mathbf{A}\mathbf{A}^*\mathbf{Q}) \quad (3.73)$$

$$E(|r_y^{(iv)}|^2) = E(\mathbf{n}^*(t)\mathbf{A}^*\mathbf{H}\mathbf{s}(t-1)\mathbf{s}^*(t-1)\mathbf{H}^*\mathbf{A}\mathbf{n}(t)) \quad (3.74)$$

$$= \text{tr}(\mathbf{A}^*\mathbf{H}E(\mathbf{s}(t-1)\mathbf{s}^*(t-1))\mathbf{H}^*\mathbf{A}E(\mathbf{n}(t)\mathbf{n}^*(t))) \quad (3.75)$$

$$= \sigma^2 \text{tr}(\mathbf{A}^*\mathbf{H}\mathbf{H}^*\mathbf{A}) \quad (3.76)$$

$$= \sigma^2 \text{tr}(\mathbf{A}\mathbf{A}^*\mathbf{Q}). \quad (3.77)$$

Therefore we can say that

$$E(|r_y|^2) = \text{tr}(\mathbf{A}\mathbf{Q}\mathbf{A}\mathbf{Q}) + \sigma^4 \text{tr}(\mathbf{A}^2) + 2\sigma^2 \text{tr}(\mathbf{A}^2\mathbf{Q}). \quad (3.78)$$

Computation of $E(r_s r_y^)$ and $E(r_y r_s^*)$:* Using the previously mentioned assumptions we have

$$E(r_s r_y^*) = E(\mathbf{s}^*(t)\mathbf{s}(t-1)\mathbf{s}^*(t-1)\mathbf{H}^*\mathbf{A}\mathbf{H}\mathbf{s}(t)) \quad (3.79)$$

$$+ \mathbf{s}^*(t)\mathbf{s}(t-1)\mathbf{s}^*(t-1)\mathbf{H}^*\mathbf{A}\mathbf{n}(t)\mathbf{s}^*(t)\mathbf{s}(t-1)\mathbf{n}^*(t-1)\mathbf{A}\mathbf{H}\mathbf{s}(t) \quad (3.80)$$

$$+ \mathbf{s}^*(t)\mathbf{s}(t-1)\mathbf{n}^*(t-1)\mathbf{A}\mathbf{n}(t)) \quad (3.81)$$

$$= E(\mathbf{s}^*(t)\mathbf{s}(t-1)\mathbf{s}^*(t-1)\mathbf{H}^*\mathbf{A}\mathbf{H}\mathbf{s}(t)) \quad (3.82)$$

$$= \text{tr}(E(\mathbf{s}(t-1)\mathbf{s}^*(t-1))\mathbf{H}^*\mathbf{A}\mathbf{H}E(\mathbf{s}(t)\mathbf{s}^*(t))) \quad (3.83)$$

$$= \text{tr}(\mathbf{H}^*\mathbf{A}\mathbf{H}) \quad (3.84)$$

$$= \text{tr}(\mathbf{A}\mathbf{Q}) \quad (3.85)$$

and similarly

$$E(r_y r_s^*) = E(\mathbf{s}^*(t)\mathbf{H}^*\mathbf{A}\mathbf{H}\mathbf{s}(t-1)\mathbf{s}^*(t-1)\mathbf{s}(t)) \quad (3.86)$$

$$+ \mathbf{s}^*(t) \mathbf{H}^* \mathbf{A} \mathbf{n}(t-1) \mathbf{s}^*(t-1) \mathbf{s}(t) \quad (3.87)$$

$$\mathbf{n}^*(t) \mathbf{A} \mathbf{H} \mathbf{s}(t-1) \mathbf{s}^*(t-1) \mathbf{s}(t) + \mathbf{n}^*(t) \mathbf{A} \mathbf{n}(t-1) \mathbf{s}^*(t-1) \mathbf{s}(t)) \quad (3.88)$$

$$= E(\mathbf{s}^*(t) \mathbf{H}^* \mathbf{A} \mathbf{H} \mathbf{s}(t-1) \mathbf{s}^*(t-1) \mathbf{s}(t)) \quad (3.89)$$

$$= \text{tr}(E(\mathbf{H}^* \mathbf{A} \mathbf{H} \mathbf{s}(t-1) \mathbf{s}^*(t-1) \mathbf{s}(t) \mathbf{s}^*(t))) \quad (3.90)$$

$$= \text{tr}(\mathbf{H}^* \mathbf{A} \mathbf{H}) \quad (3.91)$$

$$= \text{tr}(\mathbf{A} \mathbf{Q}). \quad (3.92)$$

Computation of $E(|r_s|^2)$: Assuming that the source symbols are independent, we have

$$E(|r_s|^2) = E(r_s r_s^*) \quad (3.93)$$

$$= E(\mathbf{s}^*(t) \mathbf{s}(t-1) \mathbf{s}^*(t-1) \mathbf{s}(t)) \quad (3.94)$$

$$= E\left(\sum_{i,j=1}^d s_{t-i+1}^* s_{t-i} s_{t-j}^* s_{t-j+1}\right) \quad (3.95)$$

$$= \sum_{i,j=1}^d E(s_{t-i+1}^* s_{t-i} s_{t-j}^* s_{t-j+1}) \quad (3.96)$$

$$= \sum_{i=1}^d E(|s_{t-i+1}|^2 |s_{t-i}|^2) \quad (3.97)$$

$$= \sum_{i=1}^d E(|s_{t-i+1}|^2) E(|s_{t-i}|^2) \quad (3.98)$$

$$= \sum_{i=1}^d 1 \quad (3.99)$$

$$= d. \quad (3.100)$$

Therefore we can finally show that

$$J = \text{tr}(\mathbf{A} \mathbf{Q} \mathbf{A} \mathbf{Q}) + \sigma^4 \text{tr}(\mathbf{A}^2) + 2\sigma^2 \text{tr}(\mathbf{A}^2 \mathbf{Q}) - 2\text{tr}(\mathbf{A} \mathbf{Q}) + d \quad (3.101)$$

where $\mathbf{A}$ and $\mathbf{Q}$ are Hermitian matrices.

$$\mathbf{A} = \mathbf{\Gamma}^* \mathbf{\Gamma} \quad (3.102)$$

$$\mathbf{Q} = \mathbf{H} \mathbf{H}^*. \quad (3.103)$$

In order to find Γ_o , we have to find the optimum A_o that minimizes $J(\Gamma)$ in (3.37) by letting

$$\frac{\partial J(\Gamma)}{\partial \mathbf{A}} = 0. \quad (3.104)$$

By doing this we get [25]:

$$\mathbf{QAQ} + \sigma^4 \mathbf{A} + 2\sigma^2 \mathbf{AQ} - \mathbf{Q} = 0. \quad (3.105)$$

From $\mathbf{A}_o$ the optimum filter Γ_o can be found. Solving for $\mathbf{A}$ in this matrix equation is not straightforward and to date no closed form solution has been found. However, in the next section some possible solutions for $\mathbf{A}$ will be presented given certain circumstances and assumptions.

3.5 The Innovations - Blind Sequence Detection Matrix Equation: Solving for “A”

In this section we will attempt to provide some approximate solutions to (3.105) and propose some suboptimal filters for the given approximations. The problem is to solve for $\mathbf{A}$ in the following equation:

$$\mathbf{QAQ} + \sigma^4 \mathbf{A} + 2\sigma^2 \mathbf{AQ} - \mathbf{Q} = 0 \quad (3.106)$$

$$\mathbf{A} = \mathbf{\Gamma}^* \mathbf{\Gamma} \quad (3.107)$$

$$\mathbf{Q} = \mathbf{HH}^*. \quad (3.108)$$

We also have the following identity where

$$\mathbf{R}_x(0) = \mathbf{HH}^* + \sigma^2 \mathbf{I} \quad (3.109)$$

$$= \mathbf{P}^{-1} \mathbf{D} (\mathbf{P}^*)^{-1}. \quad (3.110)$$

Therefore, we can write

$$\mathbf{Q} = \mathbf{P}^{-1} \mathbf{D} (\mathbf{P}^*)^{-1} - \sigma^2 \mathbf{I} \quad (3.111)$$

$$\mathbf{Q} = \mathbf{P}^{-1}(\mathbf{D} - \sigma^2 \mathbf{P} \mathbf{P}^*)(\mathbf{P}^*)^{-1} \quad (3.112)$$

$$\mathbf{C} = \mathbf{D} - \sigma^2 \mathbf{P} \mathbf{P}^* \quad (3.113)$$

$$\mathbf{Q} = \mathbf{P}^* \mathbf{C} (\mathbf{P}^*)^{-1}. \quad (3.114)$$

Here, $\mathbf{C}$ is a Hermitian matrix, $\mathbf{P}$ is a lower triangular matrix (prediction coefficients) and $\mathbf{D}$ is a diagonal matrix (variances of prediction error coefficient).

We will first vary the σ^2 parameter and see what solutions arise. Let us assume that the SNR is *very high* or $\sigma^2 = 0$. We then have

$$\mathbf{Q} \mathbf{A} \mathbf{Q} - \mathbf{Q} = 0. \quad (3.115)$$

$$\mathbf{Q} \mathbf{A} \mathbf{Q} = \mathbf{Q} \quad (3.116)$$

$$\mathbf{Q} \mathbf{A} = \mathbf{I} \quad (3.117)$$

$$\mathbf{A} = \mathbf{Q}^{-1} \quad (3.118)$$

$$= (\mathbf{P}^{-1} \mathbf{D} (\mathbf{P}^*)^{-1})^{-1} \quad (3.119)$$

$$= \mathbf{P}^* \mathbf{D}^{-1} \mathbf{P} \quad (3.120)$$

$$= \mathbf{\Gamma}_o^* \mathbf{\Gamma}_o. \quad (3.121)$$

Solving for $\mathbf{\Gamma}_o$, we have

$$\mathbf{\Gamma}_o = \mathbf{D}^{-1/2} \mathbf{P} \quad (3.122)$$

which is the whitening filter as expected when $\sigma^2 = 0$.

Let us assume that we are in a *high* SNR environment. We can then reduce (3.105) to

$$\mathbf{Q} \mathbf{A} \mathbf{Q} + 2\sigma^2 \mathbf{A} \mathbf{Q} - \mathbf{Q} = 0. \quad (3.123)$$

Solving for $\mathbf{A}$ in (3.123) we have

$$\mathbf{Q} \mathbf{A} + 2\sigma^2 \mathbf{A} \mathbf{Q} = \mathbf{Q} \quad (3.124)$$

$$(\mathbf{Q} + 2\sigma^2 \mathbf{I}) \mathbf{A} = \mathbf{I} \quad (3.125)$$

$$\mathbf{A} = (\mathbf{Q} + 2\sigma^2 \mathbf{I})^{-1} \quad (3.126)$$

$$= (\mathbf{P}^{-1} \mathbf{D} (\mathbf{P}^*)^{-1} + \sigma^2 \mathbf{I})^{-1} \quad (3.127)$$

$$= (\mathbf{P}^{-1}(\mathbf{D} + \sigma^2 \mathbf{P} \mathbf{P}^*)(\mathbf{P}^*)^{-1})^{-1} \quad (3.128)$$

$$= (\mathbf{P}^{-1}(\mathbf{D} + \sigma^2 \mathbf{P} \mathbf{P}^*)^{\frac{1}{2}}(\mathbf{D} + \sigma^2 \mathbf{P} \mathbf{P}^*)^{\frac{1}{2}}(\mathbf{P}^*)^{-1})^{-1} \quad (3.129)$$

$$= \mathbf{P}^*(\mathbf{D} + \sigma^2 \mathbf{P} \mathbf{P}^*)^{-\frac{1}{2}}(\mathbf{D} + \sigma^2 \mathbf{P} \mathbf{P}^*)^{-\frac{1}{2}}\mathbf{P}. \quad (3.130)$$

Solving for Γ_o , we have

$$\Gamma_o = (\mathbf{D} + \sigma^2 \mathbf{P} \mathbf{P}^*)^{-\frac{1}{2}}\mathbf{P}. \quad (3.131)$$

We will now further investigate (3.131) and find a binomial approximation to Γ_o in (3.131).

Binomial Theorem 1 *If k is any real number and $|x| < 1$, then*

$$(1 + x)^k = \sum_{n=0}^{\infty} \binom{k}{n} x^n \quad (3.132)$$

$$\binom{k}{n} = \frac{k(k-1)\dots(k-n+1)}{n!} \quad (3.133)$$

where $n \geq 1$ and $\binom{k}{0} = 1$.

Applying the Binomial Theorem to (3.131) we have:

$$(\mathbf{D}^2 + \sigma^2 \mathbf{P} \mathbf{P}^*)^{-\frac{1}{2}}\mathbf{P} = ((\mathbf{I} + \sigma^2 \mathbf{P} \mathbf{P}^* \mathbf{D}^{-2})\mathbf{D}^2)^{-\frac{1}{2}}\mathbf{P} \quad (3.134)$$

$$= (\mathbf{I} + \sigma^2 \mathbf{P} \mathbf{P}^* \mathbf{D}^{-2})^{-\frac{1}{2}}\mathbf{D}^{-1}\mathbf{P} \quad (3.135)$$

$$= \left(\sum_{n=0}^{\infty} \binom{-\frac{1}{2}}{n} (\sigma^2 \mathbf{P} \mathbf{P}^* \mathbf{D}^{-1})^n \right) \mathbf{D}^{-1/2}\mathbf{P} \quad (3.136)$$

$$\approx (1 - \frac{1}{2}(\sigma^2 \mathbf{P} \mathbf{P}^* \mathbf{D}^{-1}))\mathbf{D}^{-1/2}\mathbf{P} \quad (3.137)$$

$$= (\mathbf{D} - \frac{1}{2}\sigma^2 \mathbf{P} \mathbf{P}^*)\mathbf{D}^{-3/2}\mathbf{P}. \quad (3.138)$$

Notice that as $\sigma^2 \rightarrow 0$, $\Gamma_o \rightarrow \mathbf{D}^{-1}\mathbf{P}$ which is the whitening filter.

Finally, let us investigate the situation where the SNR is very low (the parameter σ^2 is thus high). With this assumption, we can reduce (3.105) to the following

$$\mathbf{Q} \mathbf{A} \mathbf{Q} + \sigma^4 \mathbf{A} = \mathbf{Q}. \quad (3.139)$$

Solving for $\mathbf{A}$ we have

$$\mathbf{A}\mathbf{Q} + \sigma^4\mathbf{A}\mathbf{Q}^{-1} = \mathbf{I} \quad (3.140)$$

$$\mathbf{A}(\mathbf{Q} + \sigma^4\mathbf{Q}^{-1}) = \mathbf{I} \quad (3.141)$$

$$\mathbf{A} = (\mathbf{Q} + \sigma^4\mathbf{Q}^{-1})^{-1} \quad (3.142)$$

$$= (\mathbf{P}^{-1}\mathbf{C}(\mathbf{P}^*)^{-1} + \sigma^4(\mathbf{P}^{-1}\mathbf{C}(\mathbf{P}^*)^{-1})^{-1})^{-1} \quad (3.143)$$

$$= (\mathbf{P}^{-1}\mathbf{C}(\mathbf{P}^*)^{-1} + \sigma^4(\mathbf{P}^*\mathbf{C}^{-1}\mathbf{P})^{-1}). \quad (3.144)$$

Unfortunately, we can not factor $\mathbf{A}$ into $\mathbf{\Gamma}^*\mathbf{\Gamma}$ and thus we can not find $\mathbf{\Gamma}_o$.

We will now solve the matrix equation (3.105) for the one dimensional case. In the one dimensional case, (3.105) can be written as

$$q^2a + \sigma^4a + 2\sigma^2aq = q. \quad (3.145)$$

Solving for a we get

$$a(q^2 + 2\sigma^2q + \sigma^4) = q \quad (3.146)$$

$$a = \frac{q}{q^2 + 2\sigma^2q + \sigma^4} \quad (3.147)$$

$$= \frac{q}{(q + \sigma^2)^2} \quad (3.148)$$

$$= \frac{p^{-1}c(p^*)^{-1}}{(p^{-1}c(p^*)^{-1} + \sigma^2)^2} \quad (3.149)$$

$$= p^{-1}(p^{-1}c(p^*)^{-1} + \sigma^2)^{-2}c(p^*)^{-1} \quad (3.150)$$

$$= p^{-1}(p^{-1}d(p^*)^{-1})^{-2}c(p^*)^{-1} \quad (3.151)$$

$$= p^{-1}c^{1/2}(p^{-1}d(p^*)^{-1})^{-1}(p^{-1}d(p^*)^{-1})^{-1}c^{1/2}(p^*)^{-1}. \quad (3.152)$$

Thus, for $\mathbf{\Gamma}_o$ we have

$$\mathbf{\Gamma}_o = (p^{-1}d(p^*)^{-1})^{-1}c^{1/2}(p^*)^{-1} \quad (3.153)$$

$$= (d - \sigma^2pp^*)^{1/2}d^{-1}p. \quad (3.154)$$

A minimum variance estimate $\hat{\mathbf{s}}(t)$ of $\mathbf{s}(t)$ will be done in order to estimate $r_s(t) = \mathbf{s}^*(t)\mathbf{s}(t-1)$ by using $r_{\hat{s}}(t) = \hat{\mathbf{s}}^*(t)\hat{\mathbf{s}}(t-1)$. Suppose we have a system

$$y = W\beta + \epsilon \quad (3.155)$$

where W is a known matrix and β and ϵ are random vectors. Then the minimum variance linear estimate of β is given by [14]

$$\hat{\beta} = E(\beta y^T)(E(yy^T))^{-1}y. \quad (3.156)$$

Let us now translate this problem to the innovations based BSD. Here, our system is

$$\mathbf{x}(t) = \mathbf{H}\mathbf{s}(t) + \mathbf{n}(t) \quad (3.157)$$

where $\mathbf{H}$ is W and $\mathbf{n}(t)$ is ϵ . Recall that

$$R_x(0) = \mathbf{H}\mathbf{H}^* + \sigma^2\mathbf{I} \quad (3.158)$$

$$= \mathbf{P}^{-1}\mathbf{D}(\mathbf{P}^*)^{-1}. \quad (3.159)$$

Therefore, we can write

$$\mathbf{H}\mathbf{H}^* = \mathbf{P}^{-1}\mathbf{D}(\mathbf{P}^*)^{-1} - \sigma^2\mathbf{I} \quad (3.160)$$

$$= \mathbf{P}^{-1}(\mathbf{D} - \sigma^2\mathbf{P}\mathbf{P}^*)^{1/2}(\mathbf{D} - \sigma^2\mathbf{P}\mathbf{P}^*)^{1/2}(\mathbf{P}^*)^{-1} \quad (3.161)$$

and thus

$$\mathbf{H} = \mathbf{P}^{-1}(\mathbf{D} - \sigma^2\mathbf{P}\mathbf{P}^*)^{1/2}. \quad (3.162)$$

Rewriting (3.157) we get

$$\mathbf{x}(t) = \mathbf{H}\mathbf{s}(t) + \mathbf{n}(t) \quad (3.163)$$

$$= \mathbf{P}^{-1}(\mathbf{D} - \sigma^2\mathbf{P}\mathbf{P}^*)^{1/2}\mathbf{s}(t) + \mathbf{n}(t) \quad (3.164)$$

and thus we have

$$W = \mathbf{P}^{-1}(\mathbf{D} - \sigma^2\mathbf{P}\mathbf{P}^*)^{1/2} \quad (3.165)$$

$$\epsilon = \mathbf{n}(t) \quad (3.166)$$

$$\beta = \mathbf{s}(t). \quad (3.167)$$

Our task now is to estimate $\mathbf{s}(t)$ from $\hat{\mathbf{s}}(t)$. Using (3.156) we have

$$\hat{\mathbf{s}}(t) = E(\mathbf{s}(t)\mathbf{x}^*(t))\mathbf{R}_x^{-1}(0)\mathbf{x}(t) \quad (3.168)$$

$$= E(\mathbf{s}(t)(\mathbf{s}^*(t)(\mathbf{D} - \sigma^2\mathbf{P}\mathbf{P}^*)^{1/2}(\mathbf{P}^*)^{-1} \quad (3.169)$$

$$+\mathbf{n}^*(t))(\mathbf{P}^{-1}\mathbf{D}(\mathbf{P}^*)^{-1})^{-1}\mathbf{x}(t) \quad (3.170)$$

$$= E(\mathbf{s}(t)(\mathbf{s}^*(t)(\mathbf{D} - \sigma^2\mathbf{P}\mathbf{P}^*)^{1/2}(\mathbf{P}^*)^{-1} + \mathbf{n}^*(t))\mathbf{P}^*\mathbf{D}^{-1}\mathbf{P}\mathbf{x}(t) \quad (3.171)$$

$$= (\mathbf{D} - \sigma^2\mathbf{P}\mathbf{P}^*)^{1/2}(\mathbf{P}^*)^{-1}\mathbf{P}^*\mathbf{D}^{-1}\mathbf{P}\mathbf{x}(t) \quad (3.172)$$

$$= (\mathbf{D} - \sigma^2\mathbf{P}\mathbf{P}^*)^{1/2}\mathbf{D}^{-1}\mathbf{P}\mathbf{x}(t). \quad (3.173)$$

Thus, the linear transform Γ_{mv} that gives the minimum variance estimate is

$$\Gamma_{mv} = (\mathbf{D} - \sigma^2\mathbf{P}\mathbf{P}^*)^{1/2}\mathbf{D}^{-1}\mathbf{P} \quad (3.174)$$

which is the same matrix that provides the same minimum variance estimate of the source correlation r_s (refer to (3.154)) for a one dimensional system.

Table 3.1 provides a summary of the various suboptimal filters that can be used in the Innovations Based Blind Sequence Detector algorithm.

Assumption	Γ_o
Very High SNR, $\sigma^2 = 0$	$\mathbf{D}^{-1/2}\mathbf{P}$
High SNR	$(\mathbf{D} + \sigma^2\mathbf{P}\mathbf{P}^*)^{-\frac{1}{2}}\mathbf{P}$
Binomial 1st Order Approx. to High SNR	$(\mathbf{D} - \frac{1}{2}\sigma^2\mathbf{P}\mathbf{P}^*)\mathbf{D}^{-3/2}\mathbf{P}$
Low SNR	No Solution
1-Dimensional System	$(d - \sigma^2 p p^*)^{1/2} d^{-1} p$
Minimum Variance Estimate	$(\mathbf{D} - \sigma^2\mathbf{P}\mathbf{P}^*)^{1/2}\mathbf{D}^{-1}\mathbf{P}$

Table 3.1: List of suggested filters for use in the innovations based blind sequence detection algorithm.

In addition to the above theoretically derived filters, we propose other filters or transforms derived on a purely intuitive approach. Let us examines Tong's optimal

estimator based on SVD [25]:

$$\mathbf{T}_o = (\sigma^2 \mathbf{I} + \mathbf{\Psi}_s)^{-1} \mathbf{\Psi}_s^{1/2} \mathbf{V}_s^*. \quad (3.175)$$

We propose the following filter based on an innovations approach that closely resembles Tong's BSD filter in (3.175):

$$\mathbf{\Gamma} = (\sigma^2 \mathbf{I} + \mathbf{D})^{-1} \mathbf{D}^{1/2} \mathbf{P}. \quad (3.176)$$

We shall refer to the filter in (3.176) as Filter 1. Combining (3.176) and the Minimum Variance Estimate filter in (3.174), we can “colourize” the noise parameter in (3.176) and thus suggest the following filter structure:

$$\mathbf{\Gamma} = (\sigma^2 \mathbf{P} \mathbf{P}^* + \mathbf{D})^{-1} \mathbf{D}^{1/2} \mathbf{P}. \quad (3.177)$$

We shall refer to the filter in (3.177) as Filter 2. Finally, using the Binomial filter in (3.138) as a motivation, we propose the following filter:

$$\mathbf{\Gamma} = (\mathbf{D}^{1/2} - 1/2\sigma^2 \mathbf{P} \mathbf{P}^*) \mathbf{D}^{-1} \mathbf{P}. \quad (3.178)$$

We shall refer to this filter as Filter 3.

3.6 An Innovations Based Blind Sequence Detector: Outline and Implementation

A schematic of the innovations approach to blind sequence detection is shown in Figure 3.4. The following is a brief outline of the innovations based blind sequence detection algorithm.

1. Form the data vectors $\mathbf{x}_i(t)$ by sampling the received signal at each receiver.

$$\mathbf{x}_i(t) = [\mathbf{x}_i(tN), \dots, \mathbf{x}_i((t+1)N-1)]^t. \quad (3.179)$$

2. Form the total data vector $\mathbf{x}(t)$ by

$$\mathbf{x}(t) = [\mathbf{x}_1^t(t), \dots, \mathbf{x}_M^t(t)]^t. \quad (3.180)$$

3. Estimate $R_x(0)$ by

$$\widehat{\mathbf{R}}_x(0) = 1/L \sum_{i=0}^{L-1} \mathbf{x}(t-i) \mathbf{x}^*(t-i). \quad (3.181)$$

4. Compute the Cholesky factorization of $\widehat{\mathbf{R}}_x(0)$

$$\widehat{\mathbf{R}}_x(0) = \mathbf{P}^{-1} \mathbf{D} (\mathbf{P}^*)^{-1}. \quad (3.182)$$

5. Estimate the signal dimension d .

6. Form the optimal transform matrix $\mathbf{\Gamma}$. Obtain $\mathbf{\Gamma}$ according to Table 3.1 (see discussion below on how $\mathbf{\Gamma}$ was obtained).

7. Form the “reduced” optimal $\mathbf{\Gamma}_o$ by selecting the first “d” rows of $\mathbf{\Gamma}$.

8. Apply the linear transform to the data vector

$$\mathbf{y}(t) = \mathbf{\Gamma}_o \mathbf{x}(t). \quad (3.183)$$

9. Compute the correlation

$$r_y(t) = \mathbf{y}^*(t) \mathbf{y}(t-1). \quad (3.184)$$

10. Estimate the source symbol sequence using the Viterbi Algorithm by minimizing

$$\min \sum_t |r_y(t) - r_s(t)|^2 \quad (3.185)$$

$$r_s(t) = \sum_{i=0}^d s_{t-i}^* s_{t-i-1}. \quad (3.186)$$

Estimating the dimension of the signal subspace d is very important. Using a lower dimension approximation is often satisfactory [25]. The smaller d results in a smaller trellis for the Viterbi Algorithm. Unfortunately, Tong did not investigate what the effects are of varying d on the performance of blind sequence detectors. In the current literature, there are many detection schemes which can be used as a starting point to estimate d [25] [12]. In the next Chapter we will investigate the effects of varying d and determine an optimal “ d ” for both algorithms based on our simulation results. For both the innovations and SVD approach to blind sequence detection we have assumed that σ^2 is known. This was done to make a fair comparison between the two estimators. A closed form solution for the innovations based blind sequence detection matrix equation (3.105) could not be found. Thus, an optimal transform Γ_o could not be determined. However, several approximate solutions have been proposed to (3.105) and as a result some filters have been suggested. In the following Chapter, we will test each of these filters and determine which one of them has the best performance for a given environment.

Figure 3.4: Innovations Based Blind Sequence Detector Algorithm

3.7 Conclusion

We have shown in this Chapter that the innovations approach and the Viterbi Algorithm can be used in a blind sequence detection algorithm for a Rayleigh fading channel. The whitening filter, Γ , can be used to orthogonalize the channel, $\mathbf{H}$, in

the absence of noise. As such, the correlation of the orthogonalized observation, r_y , preserves the source correlation, r_s . The Viterbi Algorithm can be used to reconstruct the input symbols, and is particularly effective in the case of noisy channels. Unfortunately, the innovations process is not optimum in the presence of noise. An attempt was made to determine an optimal filter based on the innovations process for a noisy channel. However, a closed form solution could not be found to the “Innovations - Blind Sequence Detection Matrix Equation” (3.105) and hence an optimum filter could not be determined. Several approximate solutions to equation (3.105) were suggested. Based on these solutions, some transform filters have been proposed. In the next Chapter, these filters will be tested in a Rayleigh fading environment and their performance will be assessed.

Chapter 4

Results

The performance of a digital receiver is measured by the probability of error. In digital communications, we are interested in finding the probability of a single symbol or bit error rate. We can estimate this probability as the ratio of the number of bit errors to the total number of transmitted bits. Unfortunately, it is extremely difficult to determine analytically these error probabilities for our Innovations Based Blind Sequence Detector. However, computer simulation is an approach that we can use in order to assess the performance of our digital receiver. Computer simulations are particularly attractive when it is impossible to do an analytical evaluation of the performance of a digital communication system. In this chapter, we will employ this technique to approximate the probability of error or bit error rate for the Innovations Based Blind Sequence Detector and Tong's SVD based blind sequence detection algorithm. We will first describe and set up the model that will be used in the simulations. In the subsequent sections, several parameters, such as differential encoding bit delay, signal subspace dimension d , correlation delay k , and diversity will be varied and tested for both the innovations and SVD based blind sequence detection algorithms. We will also see if either technique benefits from using a single error correcting (7,4) Hamming code.

4.1 Simulation Setup

In this section we present a description of the model used in our simulation runs. This model is similar to the one used by Tong [25]. Three ray multipath channels were used where we assumed to have three receivers. The sampling rate at each receiver was four times the symbol rate ($T=4$). The composite channels ($h_i(t)$) used in our model can be described as

$$h_i(t) = \alpha_{i1}p(t) + \alpha_{i2}p(t - \tau_1) + \alpha_{i3}p(t - \tau_2) \quad i = 1, 2, 3 \quad (4.1)$$

where $p()$ was a raised-cosine pulse with 90% roll off and the time delays were $\tau_1 = 0.7T$ and $\tau_2 = 1.2T$. The gain α_{ij} at the i th receiver with respect to the j th multipath is

$$[\alpha_{ij}] = \begin{pmatrix} 1.2094 + 0.8260j & -0.5123 - 1.3070j & 0.8922 + 0.9049j \\ 0.6251 - 0.1450j & 0.1163 + 0.3929j & -0.0075 - 0.0219j \\ -0.9660 + 1.6894j & 0.2007 - 0.0755j & 0.1018 + 0.1685j \end{pmatrix}. \quad (4.2)$$

The gain (α_{ij}) was generated from a zero mean unit variance Gaussian distribution. The channel matrix $\mathbf{H}$, was obtained by truncating the channel impulse responses to

six symbol intervals. Thus, we can represent $\mathbf{H}$ as

$$f_1(t) = \begin{pmatrix} 0.0009 + 0.0028j \\ -0.0055 - 0.0024j \\ -0.0100 - 0.0078j \\ -0.0010 - 0.0003j \\ -0.0004 - 0.0009j \\ -0.0031 + 0.0000j \\ -0.0040 + 0.0011j \\ -0.0001 - 0.0003j \\ -0.0002 - 0.0003j \\ 0.0041 - 0.0079j \\ 0.0061 - 0.0113j \\ -0.0004 - 0.0002j \end{pmatrix} \quad f_2(t) = \begin{pmatrix} 0.0009 + 0.0028j \\ -0.0055 - 0.0024j \\ -0.0100 - 0.0078j \\ -0.0010 - 0.0003j \\ -0.0004 - 0.0009j \\ -0.0031 + 0.0000j \\ -0.0040 + 0.0011j \\ -0.0001 - 0.0003j \\ -0.0002 - 0.0003j \\ 0.0041 - 0.0079j \\ 0.0061 - 0.0113j \\ -0.0004 - 0.0002j \end{pmatrix} \quad (4.3)$$

$$f_3(t) = \begin{pmatrix} 0.0009 + 0.0028j \\ -0.0055 - 0.0024j \\ -0.0100 - 0.0078j \\ -0.0010 - 0.0003j \\ -0.0004 - 0.0009j \\ -0.0031 + 0.0000j \\ -0.0040 + 0.0011j \\ -0.0001 - 0.0003j \\ -0.0002 - 0.0003j \\ 0.0041 - 0.0079j \\ 0.0061 - 0.0113j \\ -0.0004 - 0.0002j \end{pmatrix} \quad f_4(t) = \begin{pmatrix} 0.0009 + 0.0028j \\ -0.0055 - 0.0024j \\ -0.0100 - 0.0078j \\ -0.0010 - 0.0003j \\ -0.0004 - 0.0009j \\ -0.0031 + 0.0000j \\ -0.0040 + 0.0011j \\ -0.0001 - 0.0003j \\ -0.0002 - 0.0003j \\ 0.0041 - 0.0079j \\ 0.0061 - 0.0113j \\ -0.0004 - 0.0002j \end{pmatrix} \quad (4.4)$$

$$\begin{aligned}
f_5(t) &= \begin{pmatrix} 0.0009 + 0.0028j \\ -0.0055 - 0.0024j \\ -0.0100 - 0.0078j \\ -0.0010 - 0.0003j \\ -0.0004 - 0.0009j \\ -0.0031 + 0.0000j \\ -0.0040 + 0.0011j \\ -0.0001 - 0.0003j \\ -0.0002 - 0.0003j \\ 0.0041 - 0.0079j \\ 0.0061 - 0.0113j \\ -0.0004 - 0.0002j \end{pmatrix} & f_6(t) &= \begin{pmatrix} 0.0009 + 0.0028j \\ -0.0055 - 0.0024j \\ -0.0100 - 0.0078j \\ -0.0010 - 0.0003j \\ -0.0004 - 0.0009j \\ -0.0031 + 0.0000j \\ -0.0040 + 0.0011j \\ -0.0001 - 0.0003j \\ -0.0002 - 0.0003j \\ 0.0041 - 0.0079j \\ 0.0061 - 0.0113j \\ -0.0004 - 0.0002j \end{pmatrix} \quad (4.5) \\
\mathbf{H} &= [f_1(t) \ f_2(t) \ f_3(t) \ f_4(t) \ f_5(t) \ f_6(t)]. \quad (4.6)
\end{aligned}$$

Additive white Gaussian noise was added at the receiver end in our digital communication system model. We used the following definition for the signal to noise ratio (SNR) in our simulations:

$$SNR = 10 \log_{10} \frac{E(||\mathbf{H}\mathbf{s}(t)||^2)}{E(||\mathbf{n}(t)||^2)}. \quad (4.7)$$

A Binary Phase Shift Keying (BPSK) signal constellation was used as the input source to our communication system model.

The covariance matrix $\mathbf{R}_x(0)$ was estimated everytime 100 bits of data was received. The transformation matrix $\mathbf{\Gamma}_o$ was calculated from the Cholesky factorization of the estimated covariance matrix $\hat{\mathbf{R}}_o(0)$. The received data $\mathbf{x}(t)$ was transformed by $\mathbf{\Gamma}_o$ and its correlation was computed. The correlation of the transformed samples were then fed to the Viterbi Algorithm in order to reconstruct the input symbols. The initial state of the trellis was known and this was done by sending a few flag bits at the beginning of transmission. A minimum of 10^5 bits were tested for all SNR values. If 100 bit errors did not occur for a particular SNR value, the total number of bits transmitted was increased until 100 bit errors occurred.

Figure 4.1: Real parts of the channel impulse responses for the three receivers.

Figure 4.2: Imaginary parts of the channel impulse responses for the three receivers.

Many filters for Γ_o have been proposed for the Innovations Based Blind Sequence Detector since a closed form solution to (3.105) could not be determined. In the following sections, only the “best” results from each of these “suggested” filters will be shown. In Chapter 3, seven possible filters have been suggested for Γ_o . So, for example, in the case where we investigate the signal subspace dimension d , we will test all seven filters but only show the best result that arises from these filters for each SNR value. Our intention here is to first determine what the optimal d , k and bit delay values are for the Innovations Based Blind Sequence Detector. Once the optimum values have been determined for these parameters, an analysis will then be done in Section 4.6. comparing each of the seven proposed filters for Γ_o .

4.2 Differential Encoding

In many signal constellations, if the constellation is rotated, there is no way that the receiver can distinguish it from a valid constellation, unless it knows what the actual transmitted data symbols are. Unfortunately, the receiver has no knowledge of the transmitted bits. As a result, the receiver may decide on a receive phase corresponding to a rotated constellation which will in turn result in the information bits being incorrectly decoded. However, a technique known as *differential encoding* can be used to eliminate this problem. In this scheme, the information is encoded by the change in constellation position between symbols rather than absolute position [13]. This type of encoding technique is often used on channels with rapid fading such as mobile radio channels. In these type of channels, the phase can vary rapidly especially during deep amplitude fade [13].

In our simulations, we used differential encoding for a BPSK signal. Mathematically speaking, the information bits $x(n)$ can be differentially encoded as:

$$y(n) = x(n) \star x(n - 1) \quad (4.8)$$

where $y(n)$ and $x(n)$ are discrete values and can be either 1 or -1. The block diagram

in Figure 4.3 illustrates how the information bits were differentially encoded in our simulation program. On the receiver end, the encoded sequence $y(n)$ can be decoded to obtain the original information bits $x(n)$ by using the following equation:

$$x(n) = y(n) \star y(n - 1). \quad (4.9)$$

Figure 4.4 demonstrates how the received bit sequence was decoded.

Figure 4.3: Differential Encoding.

Figure 4.4: Differential Decoding.

From Figures 4.5 and 4.6 it can clearly be seen that differential encoding improves the performance of both the Innovations and SVD Based Blind Sequence Detector. When the bit delay was varied for the innovations based blind sequence detection algorithm, a bit delay of 1 seemed to give the best overall result (see Figure 4.7). However, for the SVD based blind sequence detection algorithm, the optimal performance was achieved for a bit delay of 2 (see Figure 4.8). Differential encoding is an effective technique to use in blind sequence detection. In blind sequence detection, $\{s_k\}$ and $\{s_k e^{j\Theta}\}$ have the same autocorrelations. Hence, an ambiguity exists and this can only be eliminated by knowing more information of the source. By differentially encoding the information bits, this ambiguity can be eliminated, and as shown in the

Figure 4.5: Differential encoding for the Innovations Based Blind Sequence Detector with $d=3$, bit delay =2 and $k=1$.

simulation results, it improves the overall performance of the digital receiver. For the following sections, we used differential encoding in our simulation runs.

4.3 Signal Subspace Dimension d

As shown in our simulations, determining the signal subspace dimension d is very important for blind sequence detection algorithms. Channel impulse responses have small tails so it is often impossible to exactly determine d . However, an exact value of d is not necessary for blind sequence detection algorithms. As our simulation results will show, using a lower approximation for d is adequate enough for blind sequence detection algorithms. Furthermore, a smaller d results in a smaller trellis in the Viterbi algorithm making it more computationally efficient. For both the innovations and SVD algorithms, increasing d did not always result in a lower bit error rate (BER) for a particular SNR value. In Figure 4.9 we can see that for the innovations

Figure 4.6: Differential encoding for the SVD Based Blind Sequence Detector with $d=2$, bit delay =2 and $k=1$.

Figure 4.7: Varying bit delay for the Innovations Based Blind Sequence Detector with $d=3$ and $k=1$.

Figure 4.8: Varying bit delay for the SVD Based Blind Sequence Detector with $d=2$ and $k=1$.

receiver, $d = 1$ gave the best performance for SNR values up to 12. For SNR values greater than 12, $d = 3$ provided the best BER. In the SVD algorithm, $d = 1$ had the lowest BER for SNR values up to 3 while $d = 2$ provided the lowest BER for SNR values greater than 3 dB.

The simulation results have shown that an optimum d exists for each of the blind algorithms studied. One would assume that by using a larger d , a better performance would result but this was not the case as shown in Figures 4.9 and 4.10. Generally speaking, the BER does decrease as d increases until an optimum value is reached. Increasing d beyond this optimum value results in a poorer performance of the receiver. We will provide a possible explanation as to why both the innovations and SVD blind sequence detection algorithms behave in this manner. Let us look at an example where the channel $\mathbf{H}$ has a dimension length of 6 (i.e. $d=6$). In this example, we used rows 1 to 3 and rows 5 to 7 of the channel matrix described in (4.6) as our channel fading matrix $\mathbf{H}$. Hence, we assumed to have 2 receivers and the sampling

Figure 4.9: Varying d for the Innovations Based Blind Sequence Detector with bit delay=1 and $k=1$.

Figure 4.10: Varying d for the SVD Based Blind Sequence Detector with bit delay=1 and $k=1$.

rate at each receiver was 3 times the symbol rate ($T = 3$).

For 100 bits of data the estimated autocorrelation matrix $\widehat{\mathbf{R}}_x(0)$ was determined to be:

$$\widehat{\mathbf{R}}_x(0) = \begin{pmatrix} 1.3 & 1.5 + 0.2j & 1.1 + 0.1j & 0.4 - 0.3j & 0.2 - 0.1j & 0.2 + 0.1j \\ 1.5 - 0.2j & 2.4 & 1.8 + 0.1j & 0.2 - 0.3j & 0.2 + 0.2j & 0.4 + 0.4j \\ 1.1 - 0.1j & 1.8 - 0.1j & 1.6 & 0.4 - 0.1j & 0.5 + 0.2j & 0.6 + 0.2j \\ 0.4 + 0.3j & 0.2 + 0.3j & 0.4 + 0.1j & 0.4 & 0.3 - 0.1j & 0.2 - 0.2j \\ 0.2 + 0.1j & 0.2 - 0.2j & 0.5 - 0.2j & 0.3 + 0.1j & 0.5 & 0.4 - 0.1j \\ 0.2 - 0.1j & 0.4 - 0.4j & 0.6 - 0.2j & 0.2 + 0.2j & 0.4 + 0.1j & 0.5 \end{pmatrix}. \quad (4.10)$$

The singular values of $\widehat{\mathbf{R}}_x(0)$ were

$$\mathbf{S}_x = \begin{pmatrix} 5.0993 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1.3018 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0.2602 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0.0129 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0.0103 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0.0091 \end{pmatrix}. \quad (4.11)$$

From (4.9), most of the signal is contained within the first two columns or dimensions (i.e. $d=2$) while the remaining dimensions contain noise. Hence, for the SVD blind sequence algorithm, the best performance occurs when $d=2$. If d is increased, a better performance does not result since all we are doing is taking into account more noise in our transform calculation and not gaining any new information about the signal. For the innovations approach, the signal and noise are evenly spread throughout the

entire space. The singular values for the innovations approach were

$$\mathbf{S}_y = \begin{pmatrix} 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 \end{pmatrix}. \quad (4.12)$$

From (4.10), the signal and noise are *equalized* for all the 6 dimensions. Thus, as d is increased more information about the signal is known and a better performance results. However, if d is greater than the optimal value, the performance of the receiver gets worse because too much noise has been taken into account.

4.4 Delay Parameter k

The delay parameter k in (3.31) should be chosen in such a manner so that the number of states in the Viterbi algorithm is minimum. From the simulation results, it was found that setting $k=1$ was sufficient to determine the information sequence $\{s_k\}$ (see Figures 4.11 and 4.12). Increasing k did not improve the performance of the receiver for both the Innovation and SVD Based Blind Sequence Detector. In fact, as k was increased, a poorer performance resulted (see Figures 4.11 and 4.12)

In blind sequence detection we are interested in estimating the *correlation* of the source. In fact, the correlation is used as a branch metric for the Viterbi Algorithm. The longer the correlation delay (i.e. higher the k) the less correlated the samples will be. As $k \rightarrow \infty$, the samples will asymptotically become independent. Since we are estimating the source correlation, the more correlated the samples are (i.e. the smaller the delay k is), the better the estimate. So naturally, one would choose to have a small delay in order to obtain better estimates of the source correlation. We conclude that the added complexities of a higher k are not beneficial to blind sequence

Figure 4.11: Varying k for the Innovations Based Blind Sequence Detector with $d=3$ and bit delay=1.

detection algorithms.

4.5 Diversity Reception

Diversity is an important communication receiver technique that provides wireless link improvement at relatively low cost [19]. Diversity requires no training bits because a training sequence is not needed by the transmitter. Currently, there are many diversity techniques which are very practical to implement and also at the same time provide significant link improvement with little added cost. Diversity techniques make use of the random nature of radio propagation by locating independent signal paths. In Rayleigh fading environments, it is very common for a transmitted signal to have multiple-delayed copies or reflections of itself occurring because of the nature of the mobile wireless environment. Diversity is a relatively easy concept to understand and works as follows. If one of the signal paths undergoes a deep fade then it is possible

Figure 4.12: Varying k for the SVD Based Blind Sequence Detector with $d=2$ and bit delay=2.

that another *independent* path may not have had a deep fade and as such contains a strong signal in it. When the receiver has more than one path to choose from, its performance in terms of instantaneous and average SNRs may improve by as much as 20 dB to 30 dB [19].

For the innovations based blind sequence detection algorithm, a small improvement occurred in the BER for diversity reception. However, increasing the number of “diversity paths” beyond a certain number did not result in the receiver performing better. It was determined that a diversity of 2 had the best overall BER for the innovations approach. Unfortunately, increasing the diversity reception beyond 2 resulted in a poorer performance (see Figure 4.13). For very high SNR values ($SNR > 25$) a diversity of 1 had the lowest BER. The innovations approach seems to be sensitive to diversity technique because of the positive definite property of the autocorrelation matrix $\mathbf{R}_x(0)$. This point is further explained in section 4.8 (Discussions). A signifi-

Figure 4.13: Diversity Reception for the Innovations Based Blind Sequence Detector with $d=3$, $k=1$ and bit delay=2.

cant improvement in the BER occurred for the SVD based blind sequence detection algorithm when diversity reception was used (see Figure 4.14). Diversity is much more effective for SVD based blind sequence detection algorithms than innovations based blind sequence detection algorithms.

4.6 Innovations Based Blind Sequence Detector Filters: An Analysis

In Chapter 3, a closed form solution to the innovations-blind sequence detection matrix equation (3.105) could not be analytically determined. However, several approximate solutions were suggested. From these solutions, seven filters were proposed for Γ_o . These seven filters are shown in Table 4.1. From Figures 4.15 and 4.16 the Minimum Variance Estimate filter had the best overall performance or lowest BER

Figure 4.14: Diversity Reception for the SVD Based Blind Sequence Detector with $d=2$, $k=1$ and bit delay=2.

Assumption	Γ_o
Very High SNR, $\sigma^2 = 0$	$\mathbf{D}^{-1/2}\mathbf{P}$
High SNR	$(\mathbf{D} + \sigma^2\mathbf{P}\mathbf{P}^*)^{-1/2}\mathbf{P}$
Binomial 1st Order Approx. to High SNR	$(\mathbf{D} - \frac{1}{2}\sigma^2\mathbf{P}\mathbf{P}^*)\mathbf{D}^{-3/2}\mathbf{P}$
Minimum Variance Estimate	$(\mathbf{D} - \sigma^2\mathbf{P}\mathbf{P}^*)^{1/2}\mathbf{D}^{-1}\mathbf{P}$
Filter 1	$(\sigma^2\mathbf{I} + \mathbf{D})^{-1}\mathbf{D}^{1/2}\mathbf{P}$
Filter 2	$(\sigma^2\mathbf{P}\mathbf{P}^* + \mathbf{D})^{-1}\mathbf{D}^{1/2}\mathbf{P}$
Filter 3	$(\mathbf{D}^{1/2} - 1/2\sigma^2\mathbf{P}\mathbf{P}^*)\mathbf{D}^{-1}\mathbf{P}$

Table 4.1: List of the seven proposed filters.

Figure 4.15: Proposed blind sequence detector filters with $d=3$, $k=1$ and bit delay $=1$.

of the seven proposed filters. The 1st Order Binomial filter had the best performance for SNR values less than 15 dB (see Figure 4.15). For SNR values greater than 15 dB, the Minimum Variance Estimate filter outperformed the other filters.

4.7 Error Correcting Codes

A great deal of work has been done in the area of error correcting codes. In this section, we will examine the effect of using a simple single error correcting (7,4)Hamming code for our Innovations and SVD Based Blind Sequence Detectors. We hope to accomplish two things by using a simple single error correcting code. First of all, we would like to see if coding can overall improve the BER for both blind sequence detection algorithms. Secondly and more importantly, we would like to investigate and get a better feel of the type of errors produced by both methods. Are the errors random or bursty? If the errors are randomly distributed, then using a single error

Figure 4.16: Proposed blind sequence detector filters with $d=3$, $k=1$ and bit delay $=1$.

correcting code will improve the performance of both receivers. However, if we have burst errors, then this type of coding will not result in an overall better BER.

The generator matrix $\mathbf{G}$ used for the (7,4) Hamming code in our simulation was:

$$\mathbf{G} = \begin{pmatrix} 1 & 0 & 0 & 0 & 1 & 1 & 1 \\ 0 & 1 & 0 & 0 & 1 & 1 & 0 \\ 0 & 0 & 1 & 0 & 1 & 0 & 1 \\ 0 & 0 & 0 & 1 & 0 & 1 & 1 \end{pmatrix}. \quad (4.13)$$

The transpose of the parity check matrix $\mathbf{H}$ used for our (7,4) Hamming code was

$$\mathbf{H}^T = \begin{pmatrix} 1 & 1 & 1 \\ 1 & 1 & 0 \\ 1 & 0 & 1 \\ 0 & 1 & 1 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}. \quad (4.14)$$

For SNR values up to 13 dB, there was no benefit in using coding for the innovations based blind sequence detection algorithm. However, for SNR values greater than 13 dB, the performance of the receiver improved greatly (see Figure 4.17). We can conclude that for low SNR environments, coding is not effective. For SVD algorithm, there was roughly a 3 dB to 5 dB improvement in performance when coding was used (see Figure 4.18).

Generally speaking, coding is beneficial to use in a high SNR environment. Normally for low SNR environments, coding is not useful. For the innovations approach, this was the case and as such the coding scheme used was dependent on the SNR values. However, for the SVD case, this was not so. From Figure 4.18, it appears that the coding scheme used for the SVD based blind sequence detection algorithm was independent of the SNR values. In other words, coding was beneficial regardless of whether it was a low or high SNR environment. For both the innovations and SVD based blind sequence detectors, when a single error correcting code was used a general improvement resulted for the BER. This suggest that the errors are randomly distributed and are not of the bursty type.

Figure 4.17: (7,4) Hamming Code for Innovations algorithm with $d=3, k=1$ and bit delay=1.

Figure 4.18: (7,4) Hamming Code for SVD algorithm with $d=2, k=1$ and bit delay=2.

4.8 Discussion

From our comparisons of the innovations and SVD blind sequence detection algorithms, it is evident that the SVD technique has a better performance than the innovations approach. Perhaps a reason for this is that the innovations procedure requires that the autocorrelation matrix $\mathbf{R}_x(0)$ be positive definite. However, for SVD, this requirement is not needed (see Chapter 2). In many cases, the $\mathbf{R}_x(0)$ is only positive definite because noise is present in the system. This issue will be discussed later on. The innovations approach to blind sequence detection appears to be quite sensitive to this positive definiteness property. Let us take a close look as to why this is so.

Suppose we have a one dimensional fading matrix $\mathbf{H}$ given by:

$$\mathbf{H} = \begin{pmatrix} h_{11} \\ h_{21} \\ h_{31} \end{pmatrix}. \quad (4.15)$$

Noise is then added to the system and the resulting autocorrelation matrix $\mathbf{R}_x(0)$ of $\mathbf{x}$ where $\mathbf{x} = \mathbf{H}\mathbf{s} + \mathbf{n}$ is:

$$\mathbf{R}_x(0) = \mathbf{H}\mathbf{H}^* + \sigma^2\mathbf{I}. \quad (4.16)$$

The eigenvalues of $\mathbf{R}_x(0)$ are:

$$(\sigma^2 + h_{11}^2 + h_{21}^2 + h_{31}^2, \sigma^2, \sigma^2). \quad (4.17)$$

Notice that if $\sigma^2 = 0$, the matrix is not positive definite. As a result, the Cholesky factorization of $\mathbf{R}_x(0)$ can not be done (see Chapter 2). However, by adding some noise to the system, the innovations of $\mathbf{x}$ can be determined. Unfortunately, by adding noise to this system, we are *artificially* making $\mathbf{R}_x(0)$ positive definite and as such the innovations approach becomes very sensitive to this. Depending on the value of σ^2 , we can either be moving away or towards the matrix being positive definite. If we are far from positive definiteness, Cholesky factorization can not be done properly. Simply

adding lots of noise to the system and making $\mathbf{R}_x(0)$ more positive definite doesn't help since the increased noise will affect the performance of our receiver. M. Schmidt and G. Fettweis have also stated that Cholesky factorization of $\mathbf{R}_x(0)$ is strongly influenced by σ^2 and the positive definiteness of the $\mathbf{R}_x(0)$ matrix is determined by the channel impulse response [21].

These inherent difficulties limit the performance of the Innovations Based Blind Sequence Detector as can be seen by the simulation results. For example, referring to Figure 4.19, in low SNR environments where $\mathbf{R}_x(0)$ is positive definite the BER is high. As σ^2 is decreased or as the SNR is increased, performance does improve but then an error floor is reached. Eventhough the SNR is increasing, the BER does not improve because the $\mathbf{R}_x(0)$ is less positive definite and as such proper Cholesky factorization can not be done making the receiver less robust. Even if the dimension of the fading matrix $\mathbf{H}$ is increased such that

$$\mathbf{H} = \begin{pmatrix} h_{11} & h_{12} \\ h_{21} & h_{22} \\ h_{31} & h_{32} \end{pmatrix} \quad (4.18)$$

the $\mathbf{R}_x(0)$ for the system will not be positive definite if $\sigma^2 = 0$ since one of the eigenvalues of $\mathbf{R}_x(0)$ is 0. The autocorrelation matrix $\mathbf{R}_x(0)$ determined in our simulation runs was based on a 12 by 6 fading matrix $\mathbf{H}$ (see equation (4.6)). Here $\mathbf{R}_x(0)$ was positive definite only when $\sigma^2 \neq 0$. If $\sigma^2 = 0$ (i.e. noise free environment) Cholesky factorization could not be performed on $\mathbf{R}_x(0)$.

Using diversity techniques can also affect the positive definiteness of a matrix. The channel model matrix $\mathbf{H}$ can change depending on how many receivers are used. For most of our simulations we assumed to have 3 receivers. If we wanted to use two receivers, this would have corresponded to using only the first 8 rows of $\mathbf{H}$ in equation 4.6 while one receiver would have related to the first four rows of $\mathbf{H}$. Thus, depending on how many receivers we use will affect how many rows $\mathbf{H}$ will have and as such this will affect the positive definiteness of $\mathbf{R}_x(0)$. This can be seen from the

Figure 4.19: Innovations based blind sequence detection algorithm with $d=3, k=1$ and bit delay=1.

results obtained in section 4.5 where diversity did not improve the overall BER for the innovations approach (see Figure 4.13) since Cholesky factorization and positive definiteness are sensitive to the channel model used.

Another possible reason as to why the SVD blind algorithm is better than the innovations one is that in SVD, the signal subspace can be *isolated* from the noise subspace. In the innovations approach, the signal and noise are evenly spread out in each dimension. Thus, in SVD, by isolating the signal subspace we can obtain the maximum amount of information about the signal and use it for calculating the optimal transform. Furthermore, we can avoid using most of the noise components in the system since the signal and noise have been separated (see equation (4.9) for an example of signal and noise separation). For the innovations' case, the signal and noise are evenly mixed and as such we can not obtain the maximum information about the signal for the optimal transform calculations.

Figure 4.20: Comparing innovations and SVD blind sequence detection algorithms with a diversity of one.

When diversity reception was not used, it was observed that the innovations approach to blind sequence detection had an overall better performance than the SVD based blind sequence detection algorithm. In fact, for SNR values greater than 10 dB, the innovations approach was much more superior than SVD (please refer to Figure 4.20). For example, for a fixed SNR of 30 dB, the SVD's BER was roughly 100 times greater than the innovations' BER.

4.9 Conclusion

Error performance of the Innovations and SVD Based Blind Sequence Detectors were studied. Differential encoding enhanced the digital receiver performance for both the blind sequence detection techniques. Increasing the signal subspace dimension d did not always result in a lower BER. It was determined that an optimal d existed for each SNR value for both the innovations and SVD blind algorithms. Our studies

also showed that the added complexities of using a higher delay parameter k were not beneficial to blind sequence detection algorithms. Diversity reception was an effective technique for the SVD based blind sequence detection algorithm. However, this was not the case for the innovations approach. Single error correcting codes were found to be beneficial when employed in blind sequence detectors for both the innovation technique and the SVD technique. Overall, SVD based blind sequence detection algorithms have a better performance than innovation based ones except for the case where diversity reception is not used.

Chapter 5

Conclusions

In this thesis, an innovations approach to blind sequence detection was studied and compared to SVD based blind sequence detection. Three different techniques for factorizing a matrix were examined - the innovations approach (Cholesky factorization), the Karhunen-Loeve transform (eigenvector factorization) and Singular Value Decomposition (SVD). It was found that a linear relationship existed between these three methods. Under special conditions - when the matrix is square, symmetric, and positive definite,- eigenvector factorization is equivalent to SVD. Hence, it was shown that a linear transformation existed between Cholesky factorization and SVD. Each of these techniques has its advantages and disadvantages. The benefits of using the innovations approach are that it is a causal, simpler, less expensive and computationally more efficient system than the other methods. SVD can orthogonalize a fading channel. It was also shown that Cholesky factorization can orthogonalize a fading channel, thus making it suitable for blind sequence detection. An Innovations Based Blind Sequence Detector for Rayleigh fading channels was proposed and its performance was studied using computer simulations. Simulations were also carried out to compare the error performance of both the SVD and innovations based blind sequence detection algorithms. It was observed that, in general, SVD performed better than the innovations approach. However, when diversity reception was not used,

the innovations receiver outperformed the SVD based blind sequence detector.

Further work still needs to be done in blind sequence detection. The performance of both the innovations and SVD blind sequence detectors should be evaluated using other channel models. Perhaps there are other channels where one method clearly outforms the other one. One could try using the innovations approach on a channel where Tong's blind sequence detector will not work (refer to the discussion in [25] where such channels are discussed). In the absense of diversity , the innovations approach to blind sequence detection clearly has a better performance than the SVD based blind sequence detector. One could explore this avenue and explain why SVD is sensitive to diversity while the innovations approach is not. Depending on the environment, it can be advantageous to use either SVD or innovations for blind sequence detection. Each technique has its own advantages and disadvantages. Perhaps an algorithm can be developed which will permit combination of these two techniques and which therefore can be applied to blind sequence detection, so that this new combination can exploit the advantages of both SVD and innovations. It is to be hoped that this new combination will result in improved performance for the digital receiver. Simulations were implemented to approximate the bit error rate (BER). In order to get a more accurate assessment of blind sequence detectors, it would be desirable to derive analytically an exact probability of error curve for both the innovations and SVD based blind sequence detection algorithms. If this is not possible, perhaps an upper bound on the bit error probability could be derived.

Bibliography

- [1] A.J. Bell and T.J. Sejnowski. An information maximization approach to blind separation and blind deconvolution. *Neural Computation*, vol. 7:pp. 1129–1159, 1995.
- [2] K. M. Chugg. Acquisition performance of blind sequence detectors using per-survivor-processing. *Proc. VTC97*, 2:539–543, May 1997.
- [3] A. Duel-Hallen and C. Heegard. Delayed decision-feedback sequence estimation. *IEEE Trans. Commun.*, COM -37(5):pp. 428–436, 1989.
- [4] M. Ghosh and C.L. Weber. Maximum-likelihood blind equalization. *Optical Engineering*, vol. 31(6):pp. 1224–1228, June 1992.
- [5] D.N. Godard. Self-recovering equalization and carrier tracking in two-dimensional data communication systems. *IEEE Trans. Commun.*, vol. COM-28:pp. 1867–1875, Nov. 1980.
- [6] G.H. Golub and P.V. Deoren. Numerical linear algebra, digital signal processing and parallel algorithms. *Computer and System Sciences*, 70, 1988.
- [7] D.J. Goodman. Trends in cellular and cordless communications. *IEEE Commun. Mag.*, pages 31–40, June 1991.
- [8] W. Hager. *Applied Numerical Linear Algebra*. Prentice Hall, Englewood Cliffs, New Jersey, 1988.

- [9] D. Hatzinakos and C. Nikias. Estimation of multiple channel response in frequency selective channels. *IEEE J. Select. Areas Commun.*, vol. 7:pp. 12–19, Jan. 1989.
- [10] T. Kailath. A general likelihood-ratio for random signals in gaussian noise. *IEEE Trans. Inform. Theory*, vol. 15:pp. 350–361, May 1969.
- [11] B.J. Kim and D.C. Cox. Blind sequence estimation for short burst communications over time varying wireless channels. *Proc. ICUPC 97*, 2:713–717, 1997.
- [12] K. Konstantinides and K. Yao. Statistical analysis of effective singular values in matrix rank determination. *IEEE Trans. ASSP*, vol. 36:757–763, May 1988.
- [13] E. A. Lee and D. G. Messerschmitt. *Digital Communication*. Kluwer Academic Publishers, Boston, 2nd edition, 1994.
- [14] D. G. Luenberger. *Optimization by Vector Space Methods*. John Wiley, New York, 2nd edition, 1969.
- [15] A. Papoulis. *Probability, Random Variables and Stochastic Processes*. McGraw-Hill, New York, 3rd edition, 1991.
- [16] B. Porat and B. Friedlander. Blind equalization of digital communication channels using higher order moments. *IEEE Trans. Signal Processing*, vol. 39:pp. 522–526, 1991.
- [17] R. Raheli, A. Polydoros, and C. Tzou. Per-survivor processing: a general approach to MLSE in uncertain environments. *IEEE Trans. Commun.*, 43(2/3/4):354–364, February/March/April 1995.
- [18] M. Rangaswamy, J. Michels, and D. Weiner. Multichannel detection for correlated non-gaussian random processes based on innovations. *IEEE Trans. Signal Processing*, vol. 43:1915–1922, August 1995.

- [19] Theodore S. Rappaport. *Wireless Communication: Principles and Practice*. Prentice Hall, New Jersey, 1st edition, 1996.
- [20] Y. Sato. A method of self-recovering equalization for multilevel amplitude-modulation. *IEEE Trans. Commun.*, vol. COM-43:pp. 679–682, June 1975.
- [21] M Schmidt and G. Fettweis. Fir prefiltering for near minimum phase target channels. *The Sixth Canadian Workshop on Information Theory*, pages 67–70, 1999.
- [22] N. Seshadri. Joint data and channel estimation using fast blind trellis search techniques. *in Proc. Globecom*, pages 1659–1663, 1991.
- [23] D. Shalvi and E. Weinstein. New criteria for blind deconvolution of non minimum phase systems (channels). *IEEE Inform. Theory*, vol. 36:pp. 312–320, March 1990.
- [24] G. Strang. *Linear Algebra and its Applications*. Harcourt Brace Jovanich Inc., 3rd edition, 1998.
- [25] L. Tong. Blind sequence estimation. *IEEE Trans. Commun.*, vol. 43:pp. 2986–2994, Dec. 1995.
- [26] L. Tong, G. Xu, and T. Kailath. Fast blind equalization of multipath channels via antenna arrays. *in Proc. ICASSP 1993*.
- [27] L. Tong, G. Xu, and T. Kailath. Blind identification and equalization based on second order statistics: a time domain approach. *IEEE Trans. Inform. Theory*, vol. 40, March 1994.
- [28] L. Tong, G. Xu, and T. Kailath. Blind identification and equalization of multipath channels: a frequency domain approach. *IEEE Trans. Inform. Theory*, Jan. 1995.

- [29] X. Yu. *Innovations Based Maximum Likelihood Sequence Estimation for Rayleigh Channels*. PhD thesis, University of Toronto, Toronto, 1995.
- [30] X. Yu and S. Pasupathy. Innovations based MLSE for rayleigh fading channels. *IEEE Trans. Commun.*, vol. 43(2/3/4):1534–1544, February/March/April 1995.