Blind Identification of MIMO Systems: Signal Modulation and Channel Estimation

Kai Dietze

Dissertation submitted to the Faculty of the
Virginia Polytechnic Institute and State University
in partial fulfillment of the requirements for the degree of

Doctor of Philosophy
in
Electrical and Computer Engineering

Dr. Warren L. Stutzman, Chair
Dr William A. Davis
Dr. Jeffrey H. Reed
Dr. Annamalai Annamalai
Dr. David L. Russell

August 25, 2005
Blacksburg, Virginia

Keywords: Blind System Identification, Blind Channel Estimation,
Modulation Identification, MIMO

Blind Identification of MIMO Systems: Signal Modulation and Channel Estimation

Kai Dietze

(Abstract)

Present trends in communication links between devices have opted for wireless instead of wired solutions. As a consequence, unlicensed bands have seen a rise in the interference level as more and more devices are introduced into the market place that take advantage of these free bands for their communication needs. Under these conditions, the receiver's ability to recognize and identify the presence of interference becomes increasingly important. In order for the receiver to make an optimal decision on the signal-of-interest, it has to be aware of the type (modulation) of interference as well as how the received signals are affected (channel) by these impediments in order to appropriately mitigate them.

This dissertation addresses the blind (unaided) identification of the signal modulations and the channel in a Multiple Input Multiple Output (MIMO) system. The method presented herein takes advantage of the modulation induced periodicities of the signals in the system and uses higher-order cyclostationary statistics to extract the signal and channel unknowns. This method can be used to identify more signals in the system than antenna elements at the receiver (overloaded case). This dissertation presents a system theoretic analysis of the problem as well as describes the development of an algorithm that can be used in the identification of the channel and the modulation of the signals in the system. Linear and non-linear receivers are examined at the beginning of the manuscript in order to review the *a priori* information that is needed for each receiver configuration to function properly.

Acknowledgements

First of all, I would like to thank my dissertation advisor Dr. Warren Stutzman for serving as the chairman on both my Masters and Ph.D committee and for his invaluable support and guidance during the duration of this research. I would also like to thank my committee members, Dr. William Davis, Dr. Jeffrey Reed, Dr. Annamalai Annamalai, and Dr. David Russell for reviewing my doctoral work.

I would like to thank the students and staff of the Virginia Tech Antenna Group (VTAG) with whom I interacted every day and established strong friendships. I would also like to express my gratitude to the students and staff of MPRG for their collaboration on the numerous projects both of our groups were involved in over the years. I am also indebted to all the friends I made during my stay in Blacksburg for providing encouragement and for making my experience during this period of my life richer and more fulfilling.

Most of all, I am grateful to my parents Carlos and Hilde Dietze and my brother Tassilo who have been a source of inspiration and support throughout my life.

Contents

Appendix

List of Figures

List of Tables

Chapter 1

Introduction

1.1 Motivation

This dissertation addresses the problem of *blind system identification* (BSI). *System* in this case refers to one or multiple coexisting wireless communication links from the point of view of one of the receivers. *Blind* refers to the assumption that no specific information about the environment is known beforehand except for that at its core are independent signals that have been mixed together in a linear fashion to arrive at the received signal. *Identification* is the process of trying to estimate unaided (because of the *blind* assumption) the independent signals or some of their properties as well as to determine the mechanisms through which they contribute to the received signal. The linear process that describes the transfer of energy and information from the original source signals to the receiver will be referred to as the *channel* or *transfer function*. In a communication system, the independent signals that are present in the environment can be divided into three categories: *desired*, *interferers*, and *noise*. The *desired* signal carries information that the receiver is interested in which is in contrast to the *interferers* which are also information bearing signals but directed at other receivers. Both the

desired and the *interfering* signals have *cyclostationary* (periodic statistics) statistics because of their information carrying properties associated with periodicities from their data rates and center frequencies of operation. *Noise*, on the other hand, does not carry any particular information or structure but is instead a result of a combination of random contributions of receiver noise, ambient noise, etc. and therefore can be modeled as a *stationary* signal. Both *interferers* and *noise*, as their names state, are nuisances which hamper the receiver's goal of correctly estimating the signal that has been originally sent by the *desired* source.

The assumption is that the receiver has no knowledge of its environment *a priori*. Environment, in this case, refers to the channel, the interferers with their modulations, the number of signals in the environment, and the noise structure of the environment. This can occur when a receiver is used in bands such as the ISM band, or unlicensed bands. Under such circumstances the signal environment is not known to the receiver because any number of systems could be in use on the same frequency and time slots of the desired signal. These signals might vary in number, modulation, or amount of signal overlap (in time or in frequency), since any system is permitted to use this band at any given time subject to restrictions in power and bandwidth.

Present trends in communication links between devices have opted for wireless instead of wired solutions. As a consequence, unlicensed bands have seen a rise in the interference level as more and more devices are introduced into the market place that take advantage of these free bands for their communications needs. Under these conditions, the receiver's ability to recognize its environment unaided becomes increasingly important. In order for the receiver to make an optimal decision on what the desired signal is depends on the information that is available to it as discussed in Chapter 2. At present times, the receiver knows only the modulation and frequency of the signal it is trying to receive and assumes it is operating in an interference-free environment. This assumption in many cases, and more in the future, is not appropriate, especially if the device is operating in the unlicensed bands where other devices may be in use simultaneously. The receiver needs to know the modulation of all signals in the environment as well as

the instantaneous transfer functions from each of the transmitters to the receiver to achieve maximum performance.

A common solution to this problem is to use a multiple antenna system together with a linear processing technique such as beamforming in order to remove the interference. This method works very well when there is at least the same number of antenna elements on the receiver as signals in the environment [1]. For a small portable device, however, adding more antennas can considerably increase the size, putting a limit on the number of antennas that can be placed on it. In many situations, the number of signals in the environment exceeds the number of antennas on the receiver. These types of scenarios, where the number of signals exceeds the number of sensors, have been referred to as overloaded arrays in the literature [2].

In overloaded environments, a common assumption is that the channel and the modulation of the signals are known to the receiver *a priori*. This is a valid assumption if the receiver is being used in bands were rules apply and everything is structured before hand with a pre-determined modulation (as for example the PCS bands) including training sequences (or an equivalent method) that are in place through a protocol to recover channel information. Maximum likelihood sequence estimation (MLSE) is the method of choice in digital communications systems when the number of unknowns (signals) exceeds the number of equations (antennas). MLSE is able to solve a seemingly underdetermined problem because it takes advantage of the finite alphabet of digital signals. MLSE requires an accurate estimate of the channel from each of the transmitters to each of the receiver antenna elements for proper operation as well as knowledge of the modulation of all of the signals in the system. In more hostile and less structured environments, however, there is no *a priori* information of the number of signals, their bandwidths, their modulation, or the channel. Under these circumstances, the receiver is required to blindly recover all necessary information for the MLSE (or other types of receiver) to provide the best estimate of the desired signal. The identification of these unknowns by the receiver without help is referred to as *blind system identification* in this dissertation.

The goal is to eventually move towards the development of a smart receiver that will be aware of its environment and can be deployed in various types of hostile environments. The receiver can make an optimal decision on what type of processing on the received signals will result in the most reliable estimate of the desired signal based on the particular environment it is in. Such a receiver needs to have the capability of blindly learning all necessary unknowns from the environment and also needs to have a reconfigurable architecture. A smart receiver could be used for the demodulation process in a cognitive radio system.

The concept of cognitive systems encompasses a broader range of features that goes well beyond the specifics of receiver design and adaptation. Cognitive radio is a recent concept [3] in which observations of the operating environment and information from the network is factored into the decision making in order to achieve a certain performance level for the user. The radio becomes self-aware of what the user is trying to do instead of just performing the particular task. The radio is in contact with the main network to share information about the present state of the network and the state of the radio. In order for communications to be transparent, both terminals communicate using a Radio Knowledge Representation Language (RKRL). Under the established language, each of the terminals can ask each other questions on the particular state the radio or network is in or any other information that would improve the service delivered to that particular unit. The idea is to increase the information that is available at the radio and at the network about the operating environment so better decisions can be made and resources can be appropriately allocated. The communications between the devices in the network are transparent to the user.

Cognitive systems focus primarily on the self learning ability at all levels of the radio of which a smart receiver could be an integral part. The details of receiver design are only briefly discussed in this dissertation because of the vast number of possible configurations. Thorough treatments of various types of receivers can be found in literature and will not be the focus of this dissertation. This dissertation will focus on the

identification aspect of the problem, and the following section describes the outline of this dissertation.

1.2 Dissertation Outline

Chapter 2 begins with a review of receiver structures that can be used in estimating the transmitted signal in the presence of interference. Each receiver configuration requires different amounts of *a priori* information such as channel transfer function, the modulation of the signals, and/or the structure of the noise for optimal operation. Chapter 3 introduces some of the statistical definitions and concepts that will be used in this dissertation. The theory behind higher order statistics (HOS) for stationary and cyclostationary signals will be introduced. HOS are commonly measured in terms of cumulants or moments. Cumulants can be related to the moments of the probability density function and vice-versa and can also be used to reconstruct their distribution.

Chapter 4 addresses blind source separation for the non-time dispersive MIMO channel based on independent component analysis (ICA). Blind source separation allows for the identification of as many signals in the system as antenna elements. The identification of the mixing channel can be performed as a two step process based on second-order and then fourth-order statistics. The linear, causal, and time-dispersive multiple-input-multiple-output (MIMO) channel is examined in Chapter 5. The inverse of these types of channels and their decomposition into minimum-phase and all-pass filters are addressed. Inverses of MIMO channels play an important role in the identification of the channel and the signals in the system as will be discussed.

The theoretical background behind the identification of a dispersive channel based on higher-order cyclostationary statistics is presented in Chapter 6. Linearly modulated signals are considered in this chapter to demonstrate the identification of more signals that sensors together with their channels. The identification of the signals and the channel can be performed on a subset of signals at a time using both second and fourth order cyclic statistics. Finally, Chapter 7 presents algorithms that can be used to

determine the various filter coefficients introduced in Chapter 6. At the end of this chapter, examples are explored to demonstrate the effectiveness of blind system identification using higher-order cyclostationary statistics. Chapter 8 concludes the dissertation and list contributions of this work and potential future topics that still need to be addressed.

1.3 Notation and Nomenclature

This section is included to clarify some of the nomenclature used in this dissertation as well as introduce some of the notation that will be used throughout this manuscript. The term *mixing* or its derivatives, in this dissertation, refers to a linear time-invariant operation performed on multiple signals and does *not* refer to a frequency up or down conversion as it is sometimes used in communication systems.

Random signal vectors that change with time are represented by bold lower-case letters (such as $\mathbf{x}_t$, $\mathbf{y}_t$, $\mathbf{x}(z)$, etc.). Matrices that have no randomness associated to them are represented by upper case letters and italicized (such as H_t or $H(z)$ which is used to represent the channel). Plain text upper case letters imply fixed scalars (such as T, N, or M) while lower case letters are used as indexes (such as i, j, k, l). The symbol (*) is used frequently in this dissertation and corresponds to the general conjugate operator (conjugate-transpose) also known as the Hermitian operator. All other operators or variables will be described as they are introduced in this dissertation.

References

[1] R. T. Compton, *Adaptive Antennas: Concepts and Performance*, Englewood Cliffs, N. J.: Prentice-Hall 1988.

[2] J. Hicks, S. Bayram, W. H. Tranter, R. J. Boyle, and J. H. Reed, "Overloaded Array processing with Spatially Reduced Search Joint Detection," *IEEE Journal on Selected Areas in Communications*, vol. 19, no. 8, pp. 1584-1593, August 2001.

[3] J. Mitola III, G. Q. Maguire Jr., "Cognitive Radio: Making Software Radios More Personal," *IEEE Pers. Comm*., pp. 13-18, vol. 6, no. 4, August 1999.

Chapter 2

Receiver Structures for Signal Estimation

This chapter examines some important receiver structures used in estimating the transmitted signal. More significantly, it examines the necessary information that has to be known *a priori* at the receiver for each of these configurations to work properly. The channel transfer function, the modulation of the signals, and the structure of the noise are some of the factors that need to be available at the receiver for optimal operation.

2.1 Signal Estimation

Signal estimation involves extracting a variable that is not directly observable from another variable that is observable. The observable variable could be a function of the estimated variable as well as an arbitrary number of other unknown random variables or constants. Mathematically the observable variable **y** (scalar or vector) can be written as

$$\mathbf{y} = f(\mathbf{s}, \mathbf{n}, H) \tag{2-1}$$

where **s** is the variable of interest, and **n** and H are other unknown variable and constant respectively that are also present in **y**. In this manuscript, the function $f(\bullet)$ of interest is where **y** contains a linear mixing of the signal **s** through the constant matrix H (2-2). The second random variable **n** is an additive noise component to the mixture of **s**. Mathematically, the variable **y** can be expressed as

$$\mathbf{y} = H\mathbf{s} + \mathbf{n} \tag{2-2}$$

where **y**, **s**, and **n** are zero mean random variables. This general form is representative of the linear multiple input multiple output (MIMO) non-time-dispersive wireless communications channel. **s** represents a vector containing the signals from M different sources and **y** represents the received vector at N different antenna elements. H (NxM matrix) is the channel response which includes the combined effects of the channel as well as the antenna effects.

Estimating the variable **s** from the received vector **y** involves a transformation on the observed variable. In general, this function can be linear or non-linear and takes the form

$$\hat{\mathbf{s}} = g(\mathbf{y}) \tag{2-3}$$

where the variable $\hat{\mathbf{s}}$ is an estimate of the random variable **s**. The function $g(\bullet)$ is usually extracted from **y** based on a given criterion that minimizes or maximizes a particular cost function. The mean square error criterion is a commonly used cost function in which the variance of the actual value of **s** and its estimate is minimized over all possible $g(\bullet)$. The optimal estimate of **s** under the mean square error (MMSE) criterion which is given by

$$\arg_{\hat{\mathbf{s}}=g(y)} \min E[(\mathbf{s}-g(\mathbf{y}))(\mathbf{s}-g(\mathbf{y}))] \quad \Rightarrow \quad \hat{\mathbf{s}} = g(\mathbf{y}) \tag{2-4}$$

can be found to be [1]

$$\hat{\mathbf{s}} = E[\mathbf{s} \mid \mathbf{y}] \tag{2-5}$$

where the operator $E[\bullet]$ is the expected value operator. This result make intuitive sense, if the distribution of **s** given **y** was known to the receiver, the most likely result for **s** under those conditions would be the average (or expected value) of that distribution. The MMSE in (2-4) is a cost function that ensures that the error between the true and the estimated signal will have the smallest variance. The expected value operator is defined in the traditional sense as

$$E[\mathbf{y}(\mathbf{x})] = \int \mathbf{y}(\mathbf{x}) f_{\mathbf{x}}(\mathbf{x}) d\mathbf{x}$$

where $f_{\mathbf{x}}(\mathbf{x})$ is the probability density function of the variables **x**. Calculating the expected value in (2-5) requires information of the joint probability distribution of **y** and **s**. This type of information is not usually available at the receiver and other forms of $g(\bullet)$ are usually used in practice.

2.1.1 Linear Estimation

One of the most common forms of estimation is achieved through a linear operation on the observed signal. Linear estimation is used because of its ease of implementation. In the time domain, linear filtering can be implemented through an adaptive time-invariant filter. The matrix form of linear time-invariant estimation is given by

$$\hat{\mathbf{s}}_{linear} = G\mathbf{y} \tag{2-6}$$

The linear estimation matrix G when applied to the variables **y** produces the estimate of **s** above. Under the linear system model considered in this dissertation (2-2), G can be calculated using the minimum-mean-square-error criterion. The optimal value of G in (2-6) under this criterion can be found to be [1]

$$G = R_{sy} R_y^{-1} = R_s H^* \left(H R_s H^* + R_n \right)^{-1} \tag{2-7}$$

where R_y is assumed to be invertible

$$\begin{aligned} R_{sy} &= E\left[\mathbf{s}\mathbf{y}^*\right] \\ R_s &= E\left[\mathbf{s}\mathbf{s}^*\right] \\ R_n &= E\left[\mathbf{n}\mathbf{n}^*\right] \\ R_y &= E\left[\mathbf{y}\mathbf{y}^*\right] \end{aligned} \tag{2-8}$$

The cross correlations and autocorrelations (R_{sy}, R_s, R_n, and R_y) are the second order statistics of the signals in (2-2). The operator * in this dissertation corresponds to the hermitian-conjugate when the operand is a matrix or a vector or just a conjugation when the operand is a scalar. R_y is usually found to be invertible even if R_s is not of full rank due to the full rank nature of R_n. For low signal-to-noise ratios, however, the inverse of R_y can be very close to singular and its inverse can be replaced by a pseudo-inverse to guarantee stability. As can be seen from the above results, linear estimators require information of up to the second order statistics of the random variables which is considerably less than what is needed in the general case given by (2-5) that requires the full description of the distribution. The general case requires knowledge of the joint distribution of **s** given **y** which is considerably harder to estimate or know *a priori* than just the correlation matrices.

In order to approach the optimal value of G in (2-7), the receiver requires an accurate estimate of H and R_s. Usually R_s is information that is known a priori and the matrix H can be estimated from the received signal either blindly or by using a training sequence. Training sequences are set up as part of communications protocols where a previously agreed sequence is transmitted to the receiver and the receiver compares what should be received to what is received and estimates the channel (H) from this information.

2.1.2 Non-linear Estimation

Non-linear estimation performs equal to or better than linear estimation since linear estimation can be thought of as special case of non-linear estimation. Non-linear estimation also assumes more information is known to the receiver a priori about the distributions of **s** and **y** - more than just the second order statistics as in the linear case.

If the joint distribution of **s** and **y** was known at the receiver, one can achieve better performance as illustrated by the following example. Considering the system model given in (2-2) with all real-valued signals

$$\mathbf{y} = H\mathbf{s} + \mathbf{n} \tag{2-9}$$

and assuming that the elements of **n** are independent and identically distributed and are unit variance Gaussian random variables. Additionally, the vector **s** has a binary distribution of either all 1's or -1's. The mixing matrix *H* is assumed to be known to the receiver as well. Using this information, one can construct the joint distribution of **y** and **s**. The development of this example can be seen in the appendix (A.1). The conditional distribution of **s** given **y** can be also determined and under the minimum mean square error criterion the best estimate of **s** can be found using (2-5) to be

$$\hat{\mathbf{s}}_{nonlinear} = E[\mathbf{s} \mid \mathbf{y}] = \tanh\left[\frac{1}{2}\left(h^T\mathbf{y} + \mathbf{y}^T h\right)\right] \tag{2-10}$$

tanh(•) is the hyperbolic tangent and *h* is a column vector with each column element being the sum of that particular row in *H* (*T* corresponds to the transpose operator). The operator tanh performs a nonlinear operation on the observed vector **y**. The $\hat{\mathbf{s}}$ achieved for the non-linear estimator in (2-10) always outperforms the linear estimator given in (2-7) under the same conditions for this example and all other cases.

The improved results on the estimate of **s** in (2-10) over linear case are a direct result of the knowledge available to the receiver. The *a priori* information of the distribution of **s** and **n** allows for a better performing receiver structure. As in the linear case, knowledge of the channel matrix *H* is made at the receiver but in practical situations *H* is also estimated from the received signal.

2.2 Estimation of Finite Alphabet Signals

In digital communication systems, the value of **s** is constrained to a set of predetermined values. In binary communication systems, the elements of **s** are confined to the set {-1,1} while M-ary communication systems have an M symbol alphabet. It is also possible that one of the source elements of the vector **s** is constrained to a particular set of values while another is transmitting using a different set or modulation. For MLSE to work properly, the alphabet of all the elements in **s** has to be known *a priori* to the receiver.

2.2.1 Maximum Likelihood and Zero Forcing Equalizer Receivers

In the development of the maximum likelihood (ML) receiver, the system model of interest given in (2-2) and (2-9) will be exclusively considered. The maximum likelihood receiver determines the transmitted sequence **s** that has the highest probability of being sent given the received vector **y**. Mathematically, this operation can be written as [2]

$$\hat{\mathbf{s}} = \arg_{\mathbf{s}} \max\left[P_{\mathbf{s}|H,\mathbf{y}}\left(\mathbf{s} \mid H, \mathbf{y}\right)\right] \tag{2-11}$$

where $P_{\mathbf{s}|H,\mathbf{y}}\left(\mathbf{s} \mid H, \mathbf{y}\right)$ is the probability of **s** given *H* and **y**. Assuming that all sequences of **s** are equally likely and independent of each other, (2-11) is equivalent to finding the vector **s** that maximizes

$$\hat{\mathbf{s}}_{ML} = \arg_{\mathbf{s}} \max\left[f_{\mathbf{y}|H,\mathbf{s}}\left(\mathbf{y} \mid H, \mathbf{s}\right)\right] \tag{2-12}$$

which can be directly computed from (2-11) using Baye's rule (see 4-17 for details). $f_{\mathbf{y}|H,\mathbf{s}}(\mathbf{y} \mid H, \mathbf{s})$ is the probability density function of **y** given the channel *H* and **s**. Since **y**, **s**, and **n** are related through (2-2) the conditional distribution of **y** can be computed directly. Assuming the elements of the random vector **n** are all identical, independent (or white so that $R_n=\sigma_n^2 I$), and jointly Gaussian distributed with zero mean, $f_{\mathbf{y}|H,\mathbf{s}}(\mathbf{y} \mid H, \mathbf{s})$ becomes

$$f_{\mathbf{y}|H,\mathbf{s}}(\mathbf{y} \mid H, \mathbf{s}) = \frac{1}{(2\pi)^{N/2}|R_n|^{1/2}} e^{-\frac{1}{2}(\mathbf{y}-H\mathbf{s})^* R_n^{-1}(\mathbf{y}-H\mathbf{s})} = \frac{1}{(2\pi)^{N/2}\sigma_n{}^N} \prod_{i=1}^{N} e^{-\frac{(\mathbf{y_i}-[H\mathbf{s}]_i)^2}{2\sigma_n{}^2}} \qquad (2\text{-}13)$$

where N is the dimension of the vector **y**. The reason for using a Gaussian distribution for the vector **n** is a direct result of the central limit theorem where the addition of receiver noise, ambient noise, and other unknown sources that are uncorrelated produces a Gaussian distribution.

In some cases the vector **n** is not white (the elements of **n** are dependent on each other) and has elements with unequal variances as a result of filtering or different amplification of each component. In order to use (2-13) and some of the subsequent results, the vector **n** has to be whitened first. A new observed vector $\tilde{\mathbf{y}}$ can be created through a transformation on **y** which results in white and equal power noise components. A property of joint Gaussian variables is that any linear transformation that uncorrelates its components also makes them independent of each other. Therefore, diagonalizing the matrix R_n would ensure the independence of the noise components. Diagonalization of the hermitian matrix R_n is not unique but it is commonly diagonalized using the Cholesky decomposition into [1]

$$R_n = LL^* \qquad (2\text{-}14)$$

where *L* is a lower triangular matrix. A lower triangular matrix has the property of producing a causal implementation of the whitening process which is advantageous in

real time implementations. The resultant new observed vector $\tilde{\mathbf{y}}$ with uncorrelated and equal variant noise components is given by

$$\tilde{\mathbf{y}} = \tilde{H}\mathbf{s} + \tilde{\mathbf{n}} \tag{2-15}$$

where

$$\begin{aligned} \tilde{H} &= L^{-1}H \\ \tilde{\mathbf{y}} &= L^{-1}\mathbf{y} \\ \tilde{\mathbf{n}} &= L^{-1}\mathbf{n} \end{aligned} \tag{2-16}$$

As can be seen from the above results, in general, it is not necessary to have equal power uncorrelated noise components. A linear transformation of **y** can always be performed to ensure independence and unity variance of the elements of **n**. Some of the subsequent results are developed under the assumption of equal power, white noise elements but this assumption is made to simplify the development of the derivations. The results presented subsequently are therefore also applicable for correlated and unbalanced elements in **n** after the necessary whitening transformations are performed (2-14), (2-15), and (2-16).

Returning to the maximization of (2-13) to achieve the maximum likelihood receiver as required by (2-12) is equivalent to maximizing the logarithm of $f_{\mathbf{y}|H,\mathbf{s}}(\mathbf{y} \mid H, \mathbf{s})$. The logarithm function is a monotonically increasing function and therefore the estimate $\hat{\mathbf{s}}_{ML}$ given by

$$\hat{\mathbf{s}}_{ML} = \arg_{\mathbf{s}} \max \log\left[f_{\mathbf{y}|H,\mathbf{s}}(\mathbf{y} \mid H, \mathbf{s}) \right] \tag{2-17}$$

should result in the same answer as the estimate given in (2-12). Substituting the distribution function $f_{\mathbf{y}|H,\mathbf{s}}(\mathbf{y} \mid H, \mathbf{s})$ in (2-13) into (2-17) results in the maximum likelihood estimate of **s**

$$\hat{\mathbf{s}}_{ML} = \arg_{\mathbf{s}} \max\left[-\left\|\mathbf{y} - H\mathbf{s}\right\|^2\right] = \arg_{\mathbf{s}} \min\left\|\mathbf{y} - H\mathbf{s}\right\|^2 \tag{2-18}$$

for uncorrelated noise components or

$$\hat{\mathbf{s}}_{ML} = \arg_{\mathbf{s}} \min\left\|\mathbf{y} - H\mathbf{s}\right\|^2_{R_n^{-1}} = \arg_{\mathbf{s}} \min\left[\left(\mathbf{y} - H\mathbf{s}\right)^* R_n^{-1}\left(\mathbf{y} - H\mathbf{s}\right)\right] \tag{2-19}$$

for an arbitrary correlation matrix R_n.

The resulting ML estimate $\hat{\mathbf{s}}_{ML}$ is the sequence of **s** which minimizes the distance between the received signal **y** and a reconstructed signal which results when the sequence **s** is applied to the channel. Since the elements of the source signals **s** are selected from a finite alphabet, the search over the possible values of **s** is finite and, therefore, feasible. If on the other hand, the vector **s** contains a single distributed source it would require a search over an infinite number of possibilities making this method impractical.

The performance of maximum likelihood sequence estimation, MLSE, can be calculated as a function of the distance between the elements **s**. The following section addresses the probability of symbol error for a maximum likelihood sequence estimator.

2.2.2 Performance of the MLSE and ZFE Based Receivers

In this section, the performance of the maximum likelihood sequence estimator and the zero forcing equalizer will be examined. The probability of symbol error is different for both these receiver structures and it will be shown that the MLSE will always outperform the ZFE.

Considering the system model of a linear channel where the input sequence **s** is a member of the space $\mathbb{S}$ ($\mathbf{s} \in \mathbb{S}$) and similarly, **y** belongs to the space $\mathbb{Y}$ ($\mathbf{y} \in \mathbb{Y}$) as in (2-2)

$$\mathbf{y} = H\mathbf{s} + \mathbf{n} \tag{2-20}$$

H is assumed to have full column rank to simplify the development of the theory. The vector **y** can be thought of as a mapping of **s** onto the space $\mathbb{Y}$ through the matrix *H* and the resultant mapping is offset by the vector **n**. Figure 2.1 illustrates the mapping that occurs when **s** is mapped onto the space $\mathbb{Y}$. The figure illustrates the case when the space $\mathbb{Y}$ is of larger dimension than the space $\mathbb{S}$ (the overdetermined case N>M). If no noise was present, vector **y** would lie on the range of *H* (R{*H*}) and inverting the mapping *H*, or $H^{\#}$, would result in the correct vector **s**. In the following discussion, in line with the assumptions of a finite alphabet, the vector **s** also belong to the constrained subspace $\mathbb{C}$ of $\mathbb{S}$ which are described by the discrete dark dots $\mathbf{s_C} \in (\mathbf{s_1}, \mathbf{s_2}, \mathbf{s_3})$ in Figure 2.1. Mathematically, $\mathbf{s_C} \in \mathbb{C} \subset \mathbb{S}$.

In the presence of noise, the vector **y** does not necessarily lie on the R{*H*} (see Fig. 2.1). Under these conditions, the value of **s** can only, in general, be approximated. An estimate of **s** (which may or may not be correct) can only be found if **y** lies in the range of *H* (assuming knowledge of *H*). The *least squares* (LS) solution projects the vector **y** onto the R{*H*} so that a solution of **s** can be calculated. The LS projection of **y** on the R{*H*}, $\mathbf{y_H}$, corresponds to the closest point on R{*H*} to the vector **y**. The projected value of **y** on R{*H*}, $\mathbf{y_H}$, in the least square sense is given by

$$\mathbf{y_H} = H\left(H^* H\right)^{-1} H^* \mathbf{y} \tag{2-21}$$

The above $\mathbf{y_H}$ corresponds to the vector $\mathbf{s_H}$ in the $\mathbb{S}$ space

$$\mathbf{s_H} = H^{\#}\mathbf{y} = \left(H^* H\right)^{-1} H^* \mathbf{y} \tag{2-22}$$

where $H^{\#}$ is the pseudo-inverse of the matrix *H*. $H^{\#}$ is a inverse of the mapping *H* of **s** onto the R{*H*}. This can be seen by considering the event when **n**=0 such that $\mathbf{y}=H\mathbf{s_C}$. If the inverse $H^{\#}$ is applied to the noise free case, the resulting estimate for **s** becomes

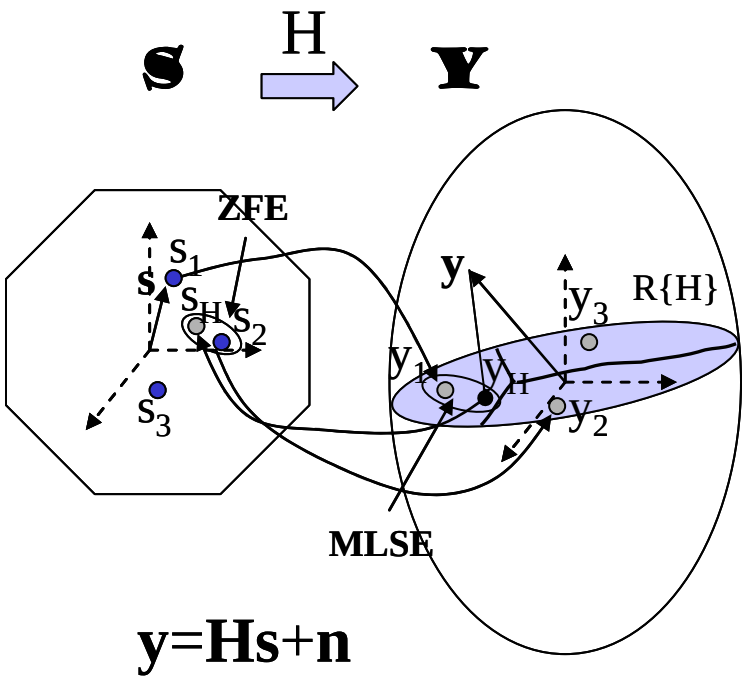

Figure 2.1 Mapping of the signal space $\mathbf{S}$ onto the observed vector space $\mathbf{Y}$

$$\mathbf{s_H} = H^{\#}\mathbf{y} = H^{\#}H\mathbf{s_C} = \left(H^{*}H\right)^{-1}H^{*}H\mathbf{s_C} = \mathbf{s_C} \tag{2-23}$$

$H^{\#}$ is therefore an inverse of the channel H or the *zero forcing equalizer* (ZFE) [3] as is commonly referred to in communication systems. In the presence of noise, the estimate of $\mathbf{s_H}$ becomes

$$\mathbf{s_H} = H^{\#}\mathbf{y} = H^{\#}\left(H\mathbf{s_C} + \mathbf{n}\right) = \mathbf{s_C} + H^{\#}\mathbf{n} = \mathbf{s_C} + \left(H^{*}H\right)^{-1}H^{*}\mathbf{n} \tag{2-24}$$

This is a combination of the true value of the transmitted sequence, $\mathbf{s_C}$, and a filtered version of the noise vector $\mathbf{n}$ through $H^{\#}$. Because of the presence of noise in the system, $\mathbf{s_H}$ may fall somewhere in the space $\mathbf{S}$ but not necessarily inside the constrained

subspace $\mathbb{C}$ which is composed of the values $\mathbf{s_C}$ ($\mathbf{s_1}$, $\mathbf{s_2}$, and $\mathbf{s_3}$ in Fig. 2.1). A ZFE receiver, for example, would pick the value $\mathbf{s}$ in the constrained space $\mathbb{C}$, $\mathbf{s_C}$, closest to the value of the estimate $\mathbf{s_H}$ computed through equalization. In Figure 2.1, for example, the ZFE receiver would pick the vector $\hat{\mathbf{s}}_\mathbf{C} = \mathbf{s_2}$ which is the closest value of $\mathbf{s}$ that lies in $\mathbb{C}$.

$$\hat{\mathbf{s}}_{\mathbf{C}\{ZFE\}} = \arg_{\mathbf{s_C}} \min \left\| \mathbf{s_H} - \mathbf{s_C} \right\|^2 \tag{2-25}$$

The ZFE therefore performs its decision in the space $\mathbb{S}$. The biggest drawback of zero forcing equalizer receivers is that they do not take into account the potential effects of the noise vector $\mathbf{n}$. The filter $H^{\#}$ in (2-24) could potentially correlate an uncorrelated noise vector which could adversely affect the decision of the estimate of the vector $\mathbf{s}$ in the $\mathbb{C}$ space, $\hat{\mathbf{s}}_\mathbf{C}$. The ZFE performs well in environments where the signal-to-noise ratio is high and the biggest deterioration of the received signal is caused by the cross-talk between the signals in $\mathbf{s}$. The ZFE, $H^{\#}$, can also be used for unconstrained values of $\mathbf{s}$ where $\mathbf{s}$ can take on any value in the space $\mathbb{S}$. When the signal $\mathbf{s}$ is unconstrained, the optimal ZFE estimate is $\mathbf{s_H}$ (2-24).

The result for $\mathbf{s_H}$ in (2-19) is a proper estimate for the ZFE which indirectly assumes uncorrelated, equal power noise components. If, on the other hand, the elements of $\mathbf{n}$ are correlated and unbalanced, a better estimate can be achieved by taking into account these factors. By using the matrix R_n, the vector $\mathbf{y}$ can be unbiased to reduce potential errors in the estimate due to an incorrectly modeled noise vector. The resulting $\mathbf{s_H}$ and $\mathbf{y_H}$ which account for noise correlation and varying strength between the elements of $\mathbf{n}$ are given by

$$\begin{aligned} \mathbf{s}_{\mathbf{H,n}} &= \left(H^* R_n^{-1} H\right)^{-1} H^* R_n^{-1} \mathbf{y} \\ \mathbf{y}_{\mathbf{H,n}} &= H\left(H^* R_n^{-1} H\right)^{-1} H^* R_n^{-1} \mathbf{y} \end{aligned} \tag{2-26}$$

The derivations for the above result are presented in the appendix (A.2) and are based on uncorrelating and balancing the noise of the vector **y** using (2-14), (2-15), and (2-16). As stated earlier, the ZFE given in (2-22) and its variant (2-26) perform a decision on the transmitted signal using norms in the space $\mathbb{S}$ (2-25). Maximum likelihood sequence estimation, on the other hand, uses metrics in the space $\mathbb{Y}$ to arrive at an estimate of the transmitted sequence.

Maximum likelihood sequence estimation provides a better alternative to the ZFE receiver in the presence of *inter-symbol-interference* (ISI) and low signal-to-noise ratios for a constrained set of **s**. MLSE takes into account the noise correlation matrix R_n in its decision. MLSE performs it decision in the space $\mathbb{Y}$ and finds the vector $\mathbf{y_C}$ that is closest to the received vector **y** in the metric of R_n^{-1} (2-16). $\mathbf{y_C}$ correspond to the noise free mappings of $\mathbf{s_C}$ onto R{H} through H. It can be shown that $\mathbf{y_{H,n}}$ and $\mathbf{s_{H,n}}$ are both *sufficient statistics* of the received vector **y** as shown in [3]. Any vector that is a transformation on the received signal **y** that contains as much information about the transmitted signal **s** as **y** is a *sufficient statistic*. In other words, a sufficient statistic produces the same estimate $\hat{\mathbf{s}}$ as the original received vector **y** does. The sufficient statistic $\mathbf{s_{H,n}}$ or $\mathbf{y_{H,n}}$ can therefore be used in estimating the input signal **s** instead of the original received signal **y**. Therefore, to evaluate the performance of MLSE it is enough to work with the modified zero-forcing-equalizer estimate, $\mathbf{s_{H,n}}$, or equivalently its $\mathbb{Y}$ space counterpart $\mathbf{y_{H,n}}$.

$$\hat{\mathbf{s}}_{\mathbf{C}\{MLSE\}} = \arg_{\mathbf{s_C}} \min \left\| \mathbf{y_{H,n}} - H\mathbf{s_C} \right\|_{R_n^{-1}}^2 = \arg_{\mathbf{s_C}} \min \left[\left(\mathbf{y_{H,n}} - H\mathbf{s_C} \right)^* R_n^{-1} \left(\mathbf{y_{H,n}} - H\mathbf{s_C} \right) \right] \quad (2\text{-}27)$$

The above norm is performed in the space $\mathbb{Y}$. In order to determine performance of the MLSE it is convenient to re-express the norms in $\mathbb{Y}$ as equivalent norms in $\mathbb{S}$. This procedure will allow the performance measure to be expressed as a function of the separations of the elements of $\mathbf{s_C}$ as will be seen shortly.

The cost function that is minimized in (2-27) can be expressed as an inner product in the space $\mathbb{Y}$ as follows

$$\begin{aligned} \left\|\mathbf{y}_{\mathbf{H,n}} - H\mathbf{s}_{\mathbf{C}}\right\|^2_{R_n^{-1}} &= \left\langle \left(\mathbf{y}_{\mathbf{H,n}} - H\mathbf{s}_{\mathbf{C}}\right), R_n^{-1}\left(\mathbf{y}_{\mathbf{H,n}} - H\mathbf{s}_{\mathbf{C}}\right)\right\rangle_{\mathbb{Y}} \\ &= \left\langle H\left(\mathbf{s}_{\mathbf{H,n}} - \mathbf{s}_{\mathbf{C}}\right), R_n^{-1} H\left(\mathbf{s}_{\mathbf{H,n}} - \mathbf{s}_{\mathbf{C}}\right)\right\rangle_{\mathbb{Y}} \end{aligned} \tag{2-28}$$

The adjoint of the linear map H from space $\mathbb{S}$ to space $\mathbb{Y}$ is a linear map H^* from $\mathbb{Y}$ to $\mathbb{S}$. Hence the inner product in (2-28) in $\mathbb{Y}$ can be written as equivalent inner product in $\mathbb{S}$ using the adjoint matrix H^* as follows

$$\begin{aligned} \left\langle H\left(\mathbf{s}_{\mathbf{H,n}} - \mathbf{s}_{\mathbf{C}}\right), R_n^{-1} H\left(\mathbf{s}_{\mathbf{H,n}} - \mathbf{s}_{\mathbf{C}}\right)\right\rangle_{\mathbb{Y}} &= \left\langle \left(\mathbf{s}_{\mathbf{H,n}} - \mathbf{s}_{\mathbf{C}}\right), H^* R_n^{-1} H\left(\mathbf{s}_{\mathbf{H,n}} - \mathbf{s}_{\mathbf{C}}\right)\right\rangle_{\mathbb{S}} \\ &= \left\|\mathbf{s}_{\mathbf{H,n}} - \mathbf{s}_{\mathbf{C}}\right\|^2_{H^* R_n^{-1} H} \end{aligned} \tag{2-29}$$

The maximum likelihood sequence estimation can now be written as a function of the distances from the sufficient statistic $\mathbf{s}_{\mathbf{H,n}}$ to the constraint values $\mathbf{s}_{\mathbf{C}}$

$$\hat{\mathbf{s}}_{\mathbf{C}\{MLSE\}} = \arg_{\mathbf{s}_{\mathbf{C}}} \min \left\|\mathbf{s}_{\mathbf{H,n}} - \mathbf{s}_{\mathbf{C}}\right\|^2_{H^* R_n^{-1} H} \tag{2-30}$$

where the norm is now performed in $\mathbb{S}$ in the metric of $H^* R_n^{-1} H$. The advantage of (2-30) is that the probability of symbol error can now be computed in terms of the more familiar distances in the space $\mathbb{S}$. The probability of symbol error for MLSE is discussed in the next subsection.

2.2.3 Probability of Error of MLSE

For the development of the average error, we assume $\mathbf{s}_1$ and $\mathbf{s}_2$ are two close neighbors in the constrained space $\mathbb{C}$ in $\mathbb{S}$. The MLSE compares the sufficient statistic $\mathbf{s}_{\mathbf{H,n}}$ to all the

possible vectors of **s** in $\mathbb{C}$ in the metric $H^{*}R_n^{-1}H$ as described by (2-30) and the one with the smallest norm is the ML estimate of the transmitted sequence. If $\mathbf{s_1}$ was sent, an error would occur if $\mathbf{s_2}$ is closer to $\mathbf{s_{H,n}}$ than $\mathbf{s_1}$ in (2-30). This event would occur when

$$\left\|\mathbf{s_{H,n}}-\mathbf{s_2}\right\|^2_{H^{*}R_n^{-1}H} < \left\|\mathbf{s_{H,n}}-\mathbf{s_1}\right\|^2_{H^{*}R_n^{-1}H} \tag{2-31}$$

Equivalently, the above error event can be written as a function of the separation from $\mathbf{s_1}$ to $\mathbf{s_2}$. Simple geometry shows that an error would occur when

$$\psi = \left\langle \left(\mathbf{s_{H,n}}-\mathbf{s_1}\right), H^{*}R_n^{-1}H\left(\mathbf{s_2}-\mathbf{s_1}\right)\right\rangle_{\mathbf{s}} > 0.5\left\|\mathbf{s_2}-\mathbf{s_1}\right\|^2_{H^{*}R_n^{-1}H} \tag{2-32}$$

The calculation of the probability of symbol error for the above event requires the knowledge of the distribution of $\mathbf{s_{H,n}}$-$\mathbf{s_1}$. The value of $\mathbf{s_{H,n}}$ will deviate from the true value of $\mathbf{s_1}$ by the random factor $\bar{\mathbf{n}}$

$$\mathbf{s_{H,n}} = \left(H^{*}R_n^{-1}H\right)^{-1}H^{*}R_n^{-1}\mathbf{y} = \mathbf{s_1} + \left(H^{*}R_n^{-1}H\right)^{-1}H^{*}R_n^{-1}\mathbf{n} = \mathbf{s_1} + \bar{\mathbf{n}} \tag{2-33}$$

The $\mathbf{s_{H,n}}$ is a summation of the transmitted symbol $\mathbf{s_1}$ and a random vector that is a linear transformation on the vector **n**. If the value of $\bar{\mathbf{n}}$ is large enough to place $\mathbf{s_{H,n}}$ closer to $\mathbf{s_2}$ than $\mathbf{s_1}$ the MLSE would cause an error. Since the initial assumption was that **n** is a Gaussian vector, a linear combination of **n** is also Gaussian distributed and therefore so is $\bar{\mathbf{n}}$. The cross-correlation matrix of $\bar{\mathbf{n}}$, $R_{\bar{n}}$, fully describes zero mean jointly distributed Gaussian variables and is given by

$$\begin{aligned} R_{\bar{n}} &= E\left[\bar{\mathbf{n}}\bar{\mathbf{n}}^{*}\right] = E\left[\left(H^{*}R_n^{-1}H\right)^{-1}H^{*}R_n^{-1}\mathbf{n}\mathbf{n}R_n^{-1,*}H\left(H^{*}R_n^{-1}H\right)^{-1,*}\right] \\ R_{\bar{n}} &= \left(H^{*}R_n^{-1}H\right)^{-1}H^{*}R_n^{-1}E\left[\mathbf{n}\mathbf{n}^{*}\right]R_n^{-1,*}H\left(H^{*}R_n^{-1}H\right)^{-1,*} \\ R_{\bar{n}} &= \left(H^{*}R_n^{-1}H\right)^{-1} \end{aligned} \tag{2-34}$$

But what is required to develop the symbol error probability is the distribution of ψ in (2-32) which is a linear function of $\bar{\mathbf{n}}$. The distribution of ψ, which is the norm on the left side of equation (2-32), can be re-written as a function of $\bar{\mathbf{n}}$ as follows

$$\psi = \left\langle \left(\mathbf{s_{H,n}} - \mathbf{s_1}\right), H^* R_n^{-1} H\left(\mathbf{s_2} - \mathbf{s_1}\right)\right\rangle_{\mathbf{g}} = \left\langle \bar{\mathbf{n}}, H^* R_n^{-1} H\left(\mathbf{s_2} - \mathbf{s_1}\right)\right\rangle_{\mathbf{g}} \tag{2-35}$$

Again, ψ being a linear combination of zero mean Gaussian variables is also zero mean Gaussian distributed. Therefore the probability density function of ψ can be fully described from the knowledge of its cross-correlation function. The variance of ψ above can be found to be

$$\begin{aligned} E\left[\psi^H \psi\right] &= E\left[\left\langle \bar{\mathbf{n}}, H^* R_n^{-1} H\left(\mathbf{s_2} - \mathbf{s_1}\right)\right\rangle_{\mathbf{g}}^* \left\langle \bar{\mathbf{n}}, H^* R_n^{-1} H\left(\mathbf{s_2} - \mathbf{s_1}\right)\right\rangle_{\mathbf{g}}\right] \\ &= \left(\mathbf{s_2} - \mathbf{s_1}\right)^* H^* R_n^{-1} H E\left[\bar{\mathbf{n}}\bar{\mathbf{n}}^*\right] H^* R_n^{-1} H\left(\mathbf{s_2} - \mathbf{s_1}\right) \\ \text{var}[\psi] &= \left\langle \left(\mathbf{s_2} - \mathbf{s_1}\right), H^* R_n^{-1} H\left(\mathbf{s_2} - \mathbf{s_1}\right)\right\rangle_{\mathbf{g}} = \left\|\mathbf{s_2} - \mathbf{s_1}\right\|^2_{H^* R_n^{-1} H} \end{aligned} \tag{2-36}$$

Using (2-36), one can replace the above variance into the standard equation for the Gaussian probability distribution [4] and develop a distribution of ψ as follows:

$$f_\psi(\psi) = \frac{1}{\sqrt{2\pi \left\|\mathbf{s_2} - \mathbf{s_1}\right\|^2_{H^H R_n^{-1} H}}} e^{-\frac{\psi^2}{2\left\|\mathbf{s_2} - \mathbf{s_1}\right\|^2_{H^* R_n^{-1} H}}} \tag{2-37}$$

As stated in (2-32) an error occurs when the value of ψ exceeds $0.5\left\|\mathbf{s_2} - \mathbf{s_1}\right\|^2_{H^* R_n^{-1} H}$. This probability of error can be found by integrating the distribution in (2-37) over the error region. The probability of symbol error, $\Pr_{E\{MLSE\}}$, is given by

$$\Pr_{E\{MLSE\}} = \int_{0.5\left\|\mathbf{s_2} - \mathbf{s_1}\right\|^2_{H^* R_n^{-1} H}}^{\infty} f_\psi(\psi) d\psi = Q\left(\frac{0.5\left\|\mathbf{s_2} - \mathbf{s_1}\right\|^2_{H^* R_n^{-1} H}}{\sqrt{\left\|\mathbf{s_2} - \mathbf{s_1}\right\|^2_{H^* R_n^{-1} H}}}\right) = Q\left(0.5\left\|\mathbf{s_2} - \mathbf{s_1}\right\|_{H^* R_n^{-1} H}\right) \tag{2-38}$$

where $Q(\bullet)$ is the Q-function [5]. The above result is the probability of symbol error of MLSE when the constrained space $\mathbb{C}$ is composed of two signals $\mathbf{s_1}$ and $\mathbf{s_2}$. The symbol error was found to depend on the separation of the constrained signals $\mathbf{s_1}$ and $\mathbf{s_2}$ in the metric of $H^* R_n^{-1} H$. To extend the result to more than two signals in $\mathbb{C}$, the probability of symbol error in (2-38) has to be averaged over all possible signals $\mathbf{s_C}$ that are being transmitted.

2.2.4 Special Case: Underdetermined MLSE

Section 2.2.2 addressed the over or exactly determined problem, in this section the underdetermined case will be considered which occurs when N<M. One of the advantages of using maximum likelihood sequence estimation is that this form of estimation can perform well even when the number of signals **s** to be estimated, exceed the number of sensors **y** under the finite alphabet assumption. This case can be better explained through an example. In this example, **y** and **n** have only one component while **s** has two components. The matrix H is therefore 1x2 and the system equation is given by

$$\begin{aligned} \mathbf{y} &= H\mathbf{s} + \mathbf{n} \\ (\mathbf{y})_1 &= \begin{bmatrix} (H)_{11} & (H)_{12} \end{bmatrix} \begin{bmatrix} (\mathbf{s})_1 \\ (\mathbf{s})_2 \end{bmatrix} + (\mathbf{n})_1 \end{aligned} \tag{2-39}$$

Assuming that the vector **s** takes on the binary values {-1,1} for each of its elements, the received value $(\mathbf{y})_1$ is tabulated as a function of $(\mathbf{s})_1$ and $(\mathbf{s})_2$ with no noise present. All possible values of y are enumerated in Table 2-1.

The third column represents the noise-less results of **y** given $H = \begin{bmatrix} 1 & 0.5 \end{bmatrix}$. As can be seen from the third column, all four joint events of $\{(\mathbf{s})_1,(\mathbf{s})_2\}$ can be distinguished from each other even though the system equations are underdetermined (less equations than unknowns to be estimated). The finite alphabet nature of **s** allow for the identification of more unknowns than equations from the knowledge of the constrained values of **s** as well

as the channel transfer function H. This makes MLSE especially attractive in digital communications systems where finite signal constellations are used.

Table 2.1. Example of the Underdetermined Case.

Transmitted sequences $\{(\mathbf{s})_1,(\mathbf{s})_2\}$	**Received signal** $(\mathbf{y})_1$	**Received value** $(\mathbf{y})_1$, for $H = \begin{bmatrix}1 & 0.5\end{bmatrix}$
{-1,-1}	$-(H)_{11}-(H)_{12}$	-1.5
{-1,1}	$-(H)_{11}+(H)_{12}$	-0.5
{1,-1}	$(H)_{11}-(H)_{12}$	0.5
{1,1}	$(H)_{11}+(H)_{12}$	1.5

In the underdetermined case, the maximum likelihood sequence estimation is also dependent on the distances in the space $\mathbb{Y}$ as in (2-27) or with equivalent norm in space $\mathbb{S}$ as in (2-30). In the above example, the four values that the **s** vector can take can be clearly distinguished by selecting the combination of $\{(\mathbf{s})_1,(\mathbf{s})_2\}$ that produces a signal $\mathbf{y_C}$ closest to the received value **y** (norm is performed in $\mathbb{Y}$). The main difference of the underdetermined case as opposed to the equally or over-determined case is that the sufficient statistic given by (2-23) is no longer valid. This can be seen directly by noting that $\left(H^* R_n^{-1} H\right)^{-1}$ in (2-23) is not invertible under underdetermined conditions, H is no longer of full column rank.

A sufficient statistic for the underdetermined MLSE is the minimum norm (MN) solution of $\mathbb{Y}$. The MN sufficient statistics for the underdetermined case are given by

$$\mathbf{s_{H,n}} = H^*\left(HH^*\right)^{-1}\mathbf{y} \qquad (2\text{-}40)$$

$$\mathbf{y_{H,n}} = \mathbf{y}$$

Note that for this case, $\mathbf{s}_{\mathbf{H,n}}$ does not depend on the noise cross-correlation matrix R_n like the overdetermined case given in (2-26). The MLSE estimate of the transmitted sequence **s** is still given by (2-27) and (2-30) for the underdetermined case but with $\mathbf{s}_{\mathbf{H,n}}$ being the statistic given in (2-40).

2.3 Summary

In this chapter the following receiver structures were analyzed: MLSE, the ZFE, and its variants. Depending on the complexity of the receiver, more information has to be available at the receiver which creates a tradeoff of better signal estimation vs. complexity. The ideal receiver based on minimizing the expected value of the transmitted signal and its estimate (2-4) requires information of the joint probability density function of received vector, **y**, and the signal of interest, **s** (2-5). This information is often not available at the receiver in its full description. In general, this receiver is non-linear as shown by the example given by (2-10). A sub-optimal estimator of **s** that is based on a linear estimate of the received variable **y** is given in (2-7) if not as much a priori information is available at the receiver. This receiver requires information of only up to second order moment of **s** and **y** instead of the full distribution. For the linear system model in (2-2) under consideration, the linear receiver requires information of the channel H as well as the a priori cross-correlation matrix of **n** and **s** as can be seen in (2-7).

The zero-forcing-equalizer inverts the effects of the channel H (2-22). It does not consider the possible noise cross-correlations, R_n, in its decision or what happens to the noise when the channel is inverted. For a finite alphabet **s**, the ZFE receiver performs a minimum distance comparison to all the signals in the space $\mathbb{S}$ and chooses the one with the minimum norm. The zero forcing equalizer requires the knowledge of H.

The MLSE receiver takes advantage of the finite alphabet of **s** which is typical of digital communication systems. The advantage of using maximum likelihood sequence

estimation is that it performs well even in underdetermined systems (more signals to be estimated than sensors). The estimate of the transmitted signal is based on a distance norm between **y** and what should have been received given all possible combinations of **s** applied to the channel *H*. The synthesized signal that lies closest to the received signal is the transmitted estimated signal (2-27) and (2-30). The probability of symbol error for MLSE is given by (2-38). The MLSE receiver requires knowledge of *H*, R_n, and the modulation (or alphabet) of each of the elements of **s**.

The performance of the ZFE and the MLSE receivers can only be compared for the over or exactly determined case since the ZFE is based on the LS inversion of the channel which is not possible when the channel is underdetermined. Under the assumption of uncorrelated and equal variant noise signals, the MLSE and ZFE are based on the minimization of the norms $\|\mathbf{s_H} - \mathbf{s_C}\|^2_{H^*H}$ and $\|\mathbf{s_H} - \mathbf{s_C}\|^2$ respectively. The difference in both decision statistics lies in the metric over which the norm is performed. The norm metric H^*H, for MLSE, is equivalent to the norm in space $\mathbf{Y}$ where the noise is uncorrelated. The performance of the ZFE, on the other hand, performs its norm in the space $\mathbf{S}$ where the once uncorrelated noise is now correlated due to inverse operation $H^{\#}$ (2-24). The MLSE will therefore always outperform the ZFE unless the channel matrix *H* is the identity matrix which is a mapping of the space $\mathbf{S}$ back into itself. Additionally, the performance of both receivers will be comparable for high signal-to-noise ratios (small **n**) but for low signal-to-noise ratios (high **n**) the MLSE will show a clear advantage over the ZFE.

The purpose of this chapter is to analyze the necessary information that is required at the receiver for the various receiver structures. The more the receiver knows about its operating environment, the better its achievable performance. The estimates of the channel *H* are usually estimated from training sequences predetermined by the communications protocol and the receiver is usually aware of what types of signals are present in the environment *a priori*. The problem arises if no information were available at the receiver *a priori* and it was left up to the receiver to determine all required

information on its own. The remainder of this dissertation will address this problem and will focus on blindly estimating some of the parameters that need to be available at the receiver a priori such as H, R_n, the number of signals in **s**, and the alphabet of **s**.

References

[1] T. Kailath, A. H. Syed, and B. Hassibi, *Linear Estimation*. Upper Saddle River, New Jersey: Prentice Hall, 2000.

[2] J. G. Proakis, *Digital Communications*. New York, NY: McGraw-Hill, 1995.

[3] L. C. Barbosa, "Maximum Likelihood Sequence Estimators: A Geometric View," *IEEE Trans. on Inf. Theory*, vol. 35, no.2, pp. 419-427, March 1989.

[4] H. Stark and J. W. Woods, *Probability, Random Processes, and Estimation Theory for Engineers*. New Jersey: Prentice Hall,1994.

[5] L. W. Couch, *Digital and Analog Communication Systems*. Upper Saddle River, New Jersey: Prentice Hall, 1997.

Chapter 3

Higher Order Statistics of Joint Random Variables

First and second order statistics are common ways to characterize a random signal. A random variable is usually portrayed, in a simplistic fashion, by its mean (first order statistic) and variance (second order statistic). For temporally varying signals, the time auto-correlation function (second order statistic) is used to quantify the rate of change of the signal versus time. The time auto-correlation of a signal can be quantified as a function of the lag time between the signal and its shifted version for wide-sense stationary (WSS) signals. An equivalent way of examining the rate of change of a signal is in a frequency domain through the power spectral density (PSD) (also a second order statistic) which is the Fourier transform of the auto-correlation function for WSS signals. A signal that has a time auto-correlation function that is an impulse, for example, is referred to as spectrally "white" because it contains all frequencies with equal strength which translates into a constant PSD as a function of frequency.

The relationship between two or more random variables can be quantified through second order statistics using a correlation coefficient. Correlation is a weaker condition on dependence. This is because two signals that are uncorrelated are not necessarily independent but independence does mean both signals are uncorrelated. As the auto-correlation is a measure of the relationship of the same signal at different instances in time, the time cross-correlation function compares the time variations of two different signals. The time cross-correlation function, which is also a second order statistic, is also quantified as a function of a time lag between two signals when both are jointly WSS. The time cross-correlation of two WSS signals **x** and **y** is defined as

$$R_{xy}(\tau) = E\left[\mathbf{x}(t)\mathbf{y}(t+\tau)\right]$$

There is a lot of information that can be extracted from a signal that is not available or resolvable from the first or second order statistics. Higher order statistics (HOS) are commonly measured in terms of cumulants or moments. Cumulants can be related to the moments of the probability density function and vice-versa. Most signals of interest encountered in communication system have higher order statistics (cumulants higher than the second order) that are non-zero. The Gaussian random variable, is an exception however, it is fully described by the first and second order statistic and therefore has zero higher order cumulants as will be explained in this chapter.

The following sections will provide some background on higher order statistics that are required in understanding the subsequent chapters.

3.1 Characteristic Function

The characteristic function is an equivalent representation (a fully invertible transform) of the probability density function. The characteristic function of a random vector **x** is given by

$$\Phi_{\mathbf{X}}(\mathbf{w}) = E\left[e^{j\mathbf{w}^T\mathbf{x}}\right] = \int f_{\mathbf{X}}(\mathbf{x}) e^{j\mathbf{w}^T\mathbf{x}} d\mathbf{x} \tag{3-1a}$$

$$f_{\mathbf{X}}(\mathbf{x}) = \frac{1}{(2\pi)^N} \int \Phi_{\mathbf{X}}(\mathbf{w}) e^{-j\mathbf{w}^T\mathbf{x}} d\mathbf{w} \tag{3-1b}$$

Except for the missing negative sign in (3-1a), the above equation can be recognized as the Fourier transform of the probability density function of **x**, $f_{\mathbf{X}}(\mathbf{x})$. (3-1b) corresponds to the inverse relationship between the characteristic function and the probability density function of **x**. As will be seen in the next section, from the knowledge of $\Phi_{\mathbf{X}}(\mathbf{w})$, one can easily determine the moments of the probability density function.

A second function, which leads to the cumulants of the probability density functions instead of the moments, is given by

$$\Psi_{\mathbf{X}}(\mathbf{w}) = \ln \Phi_{\mathbf{X}}(\mathbf{w}) \tag{3-2}$$

It should be noted that both (3-1) and (3-2) contain the same information and one can be determined from the other. Additionally, from the knowledge of either (3-1) or (3-2) the probability density function of **x** can be fully determined. In order to prevent confusion, the convention adopted in this dissertation for functions of a variable **x** are written with round brackets as with g(**x**) while functions of the distribution of **x** are written with square brackets as with g[**x**].

3.2 Moments and Cumulants for HOS

3.2.1 Moments

The moment of a joint random variable **x** where each component of $(\mathbf{x})_i$ is raised to the corresponding exponent of the element in the vector **r**, $(\mathbf{r})_i$, is given by

$$M[\mathbf{x},\mathbf{r}] = E\left[(\mathbf{x})_1^{(\mathbf{r})_1}(\mathbf{x})_2^{(\mathbf{r})_2}\ldots(\mathbf{x})_i^{(\mathbf{r})_i}\ldots(\mathbf{x})_N^{(\mathbf{r})_N}\right] = \int (\mathbf{x})_1^{(\mathbf{r})_1}(\mathbf{x})_2^{(\mathbf{r})_2}\ldots(\mathbf{x})_i^{(\mathbf{r})_i}\ldots(\mathbf{x})_N^{(\mathbf{r})_N} f_{\mathbf{X}}(\mathbf{x}) d\mathbf{x} \tag{3-3}$$

It is well known [1] that the moments can be determined from the characteristic function. As with the single variable case, the joint characteristic function of several random variables (3-1) can be used to determine the joint moments in (3-3). In order to better appreciate the relationship between $\Phi_{\mathbf{X}}(\mathbf{w})$ and the moments of a distribution, the terms of the integral in (3-1) need to be re-written in an equivalent form. The complex exponential component in (3-1) can be expanded using the Maclaurin series in the following fashion

$$e^{j\mathbf{w}^T\mathbf{x}} = \sum_{(\mathbf{r})_1=0}^{\infty}\sum_{(\mathbf{r})_2=0}^{\infty}\cdots\sum_{(\mathbf{r})_N=0}^{\infty}\frac{\left(j(\mathbf{w})_1(\mathbf{x})_1\right)^{(\mathbf{r})_1}\left(j(\mathbf{w})_2(\mathbf{x})_2\right)^{(\mathbf{r})_2}\ldots\left(j(\mathbf{w})_N(\mathbf{x})_N\right)^{(\mathbf{r})_N}}{(\mathbf{r})_1!(\mathbf{r})_2!\ldots(\mathbf{r})_N!} \tag{3-4}$$

Replacing the series in (3-4) back into (3-1) results in an equivalent representation of the joint characteristic function

$$\Phi_{\mathbf{X}}(\mathbf{w}) = \int f_{\mathbf{X}}(\mathbf{x})\sum_{(\mathbf{r})_1=0}^{\infty}\sum_{(\mathbf{r})_2=0}^{\infty}\cdots\sum_{(\mathbf{r})_N=0}^{\infty}\frac{\left(j(\mathbf{w})_1(\mathbf{x})_1\right)^{(\mathbf{r})_1}\left(j(\mathbf{w})_2(\mathbf{x})_2\right)^{(\mathbf{r})_2}\ldots\left(j(\mathbf{w})_N(\mathbf{x})_N\right)^{(\mathbf{r})_N}}{(\mathbf{r})_1!(\mathbf{r})_2!\ldots(\mathbf{r})_N!}d\mathbf{x} \tag{3-5}$$

$$\Phi_{\mathbf{X}}(\mathbf{w}) = \sum_{(\mathbf{r})_1=0}^{\infty}\sum_{(\mathbf{r})_2=0}^{\infty}\cdots\sum_{(\mathbf{r})_N=0}^{\infty}\frac{\left(j(\mathbf{w})_1\right)^{(\mathbf{r})_1}\left(j(\mathbf{w})_2\right)^{(\mathbf{r})_2}\ldots\left(j(\mathbf{w})_N\right)^{(\mathbf{r})_N}}{(\mathbf{r})_1!(\mathbf{r})_2!\ldots(\mathbf{r})_N!}\int(\mathbf{x})_1^{(\mathbf{r})_1}(\mathbf{x})_2^{(\mathbf{r})_2}\ldots(\mathbf{x})_N^{(\mathbf{r})_N}f_{\mathbf{X}}(\mathbf{x})d\mathbf{x}$$

which is composed of summations over infinite number of terms containing the components of **x** raised to arbitrary powers **r**. The integral over the variables **x** above can be recognized as the joint moment $M[\mathbf{x},\mathbf{r}]$ as defined in (3-3). The characteristic function $\Phi_{\mathbf{X}}(\mathbf{w})$ can therefore be expressed as a function of the sum of the individual joint moments of all orders in **x** as

$$\Phi_{\mathbf{X}}(\mathbf{w}) = \sum_{(\mathbf{r})_1=0}^{\infty}\sum_{(\mathbf{r})_2=0}^{\infty}\cdots\sum_{(\mathbf{r})_N=0}^{\infty}\frac{\left(j(\mathbf{w})_1\right)^{(\mathbf{r})_1}\left(j(\mathbf{w})_2\right)^{(\mathbf{r})_2}\ldots\left(j(\mathbf{w})_N\right)^{(\mathbf{r})_N}}{(\mathbf{r})_1!(\mathbf{r})_2!\ldots(\mathbf{r})_N!}M[\mathbf{x},\mathbf{r}] \tag{3-6}$$

Just as the characteristic function can be expressed as a function of the individual moments (3-6), the joint moments $M[\mathbf{x},\mathbf{r}]$ can be expressed as a function of $\Phi_{\mathbf{X}}(\mathbf{w})$ as follows

$$M[\mathbf{x},\mathbf{r}] = (-j)^{\sum_{i=1}^{N}(\mathbf{r})_i} \Phi^{\mathbf{r}}(0)$$

$$\Phi^{\mathbf{r}}(0) = \frac{\partial^{\sum_{i=1}^{N}(\mathbf{r})_i} \Phi_{\mathbf{X}}(\mathbf{w})}{\partial^{(\mathbf{r})_1}(\mathbf{w})_1 \partial^{(\mathbf{r})_2}(\mathbf{w})_2 \ldots \partial^{(\mathbf{r})_N}(\mathbf{w})_N}\Bigg|_{\mathbf{w}=0} \tag{3-7}$$

Any moment $M[\mathbf{x},\mathbf{r}]$ can be recovered by partial differentiation of the characteristic function with respect to the vector $\mathbf{w}$ applied at $\mathbf{w}=\mathbf{0}$. The characteristic function $\Phi_{\mathbf{X}}(\mathbf{w})$ is therefore also referred to as the moment generating function since from this function alone, all joint moments can be determined.

The relationship between the characteristic function $\Phi_{\mathbf{X}}(\mathbf{w})$ and the joint moment $M[\mathbf{x},\mathbf{r}]$ is given in (3-6) and (3-7). The connection between these two functions establishes a link between the probability density function $f_{\mathbf{X}}(x)$ and its joint moments. From the knowledge of all joint moments, the joint probability density function of the variables $\mathbf{x}$ can be determined with certainty as can be seen from (3-6) and (3-1b). The following subsection will establish a link between cumulants and moments. Cumulants have properties that are very desirable for signal processing applications.

3.2.2 Cumulants

Cumulants can be viewed as the statistical contribution of a particular order moment with the contribution from the lower order moments removed. This is equivalent to the difference between the variance and the second order moment of a signal. The variance of a single random variable is independent of the mean (first order moment) of the same variable while the second order moment does depend on the mean. The variance of a single random variable x is given by

$$\sigma_{\mathrm{x}}{}^2 = E\left[\left(\mathrm{x} - E[\mathrm{x}]\right)\left(\mathrm{x} - E[\mathrm{x}]\right)\right] = E\left[\mathrm{x}^2\right] - E\left[\mathrm{x}\right]^2 = M\left[\mathrm{x}^2\right] - M[\mathrm{x}]^2$$

This variance expression is independent of the mean (first order statistic) because the contribution of the mean ($M[\mathrm{x}]$) to the second order moment ($M[\mathrm{x}^2]$) is subtracted out to form $\sigma_{\mathbf{X}}{}^2$. The variance is, therefore, the second order cumulant of a single random variable where the notation $C[\mathrm{x}^2]=\sigma_{\mathbf{X}}{}^2$ is used here for cumulants. A more general definition of the cumulant follows.

As discussed in the previous section, all joint moments of a random vector can be computed from the moment generating function $\Phi_{\mathbf{X}}(\mathbf{w})$. Similarly, cumulants can be generated through the use of the cumulant generating function $\Psi_{\mathbf{X}}(\mathbf{w})$ defined in (3-2). As with moments (3-7), cumulants can be computed in the following manner from $\Psi_{\mathbf{X}}(\mathbf{w})$

$$\begin{aligned} C\left[\mathbf{x},\mathbf{r}\right] &= (-j)^{\sum_{i=1}^{N}(\mathbf{r})_i}\,\Psi^{\mathbf{r}}(0) \\ \Psi^{\mathbf{r}}(0) &= \left.\frac{\partial^{\sum_{i=1}^{N}(\mathbf{r})_i}\Psi_{\mathbf{X}}(\mathbf{w})}{\partial^{(\mathbf{r})_1}(\mathbf{w})_1\,\partial^{(\mathbf{r})_2}(\mathbf{w})_2 \ldots \partial^{(\mathbf{r})_N}(\mathbf{w})_N}\right|_{\mathbf{w}=0} = \left.\frac{\partial^{\sum_{i=1}^{N}(\mathbf{r})_i}\ln\Phi_{\mathbf{X}}(\mathbf{w})}{\partial^{(\mathbf{r})_1}(\mathbf{w})_1\,\partial^{(\mathbf{r})_2}(\mathbf{w})_2 \ldots \partial^{(\mathbf{r})_N}(\mathbf{w})_N}\right|_{\mathbf{w}=0} \end{aligned} \tag{3-8}$$

where, as with moments, the cumulants are defined as follows

$$C\left[\mathbf{x},\mathbf{r}\right] = C\left[(\mathbf{x})_1^{(\mathbf{r})_1}(\mathbf{x})_2^{(\mathbf{r})_2}\ldots(\mathbf{x})_i^{(\mathbf{r})_i}\ldots(\mathbf{x})_N^{(\mathbf{r})_N}\right] \tag{3-9}$$

Even though the representation of $C[\mathbf{x},\mathbf{r}]$ is useful in relating the cumulants (or moments) of arbitrary order to the cumulant (or moment) generating function, a more suitable representation for a cumulant or moment of a particular order P is more adequate for this dissertation. The order P of a joint cumulant or moment can be related to the vector **r** as

$$\mathrm{P} = \sum_{i=1}^{N} (\mathbf{r})_i \tag{3-10}$$

which is the sum of all the components in **r**. The representation of a cumulant for a particular order P will be written as $C_{\mathrm{P}}[\mathbf{x}]$ where N is the dimesion of **x**. The $C_{\mathrm{P}}[\mathbf{x}]$ is composed of a subset of P elements of **x** which may contain repeating elements. For example, the cumulant of a real vector **x** of length N=2 and order P=2, $C_2[\mathbf{x}]$, can refer to any of the following four cumulants

$$C[(\mathbf{x})_1(\mathbf{x})_2] = C[(\mathbf{x})_2(\mathbf{x})_1], C[(\mathbf{x})_1^2] = C[(\mathbf{x})_1(\mathbf{x})_1], C[(\mathbf{x})_2^2] = C[(\mathbf{x})_2(\mathbf{x})_2] \tag{3-11}$$

Even though the notation in (3-8) and (3-9) is more specific than the chosen $C_{\mathrm{P}}[\mathbf{x}]$ here, the latter is more suitable because of its flexibility and generality. In subsequent chapters, all cumulants of a particular order P are analyzed together and the reason for the adoption of such a notation will become evident. $C_{\mathrm{P}}[\mathbf{x}]$ can be related back to the lower order joint moments as shown in [2]. The cumulant $C_{\mathrm{P}}[\mathbf{x}]$ is given in [2] and can be re-written in the following manner

$$C_{\mathrm{P}}[\mathbf{x}] = C[x_1 x_2 \ldots x_{\mathrm{P}}] = \sum_{p=1}^{\mathrm{P}} \left\{ (-1)^{p-1}(p-1)! \sum_{\substack{\mathcal{L} \in \mathcal{P}(1:\mathrm{P}) \\ \text{all}|\mathcal{L}|=p}} \prod_{j=1}^{p} M[b_j] \right\} \tag{3-12}$$

$$M[b_j] = E[b_j]$$

where $x_1, x_2, \ldots x_{\mathrm{P}}$ can be any of the elements of the vector **x**, $(\mathbf{x})_{\mathrm{i}}$ (including repeated elements). The above cumulant is expressed as a function of moments of order P or lower. The set $\{x_1 \ldots x_{\mathrm{P}}\}$ can be partitioned ($\mathcal{P}(1{:}\mathrm{P})$) in many different ways in a collection of distinctive subsets $\mathcal{L}$ containing elements from 1 to P. The individual components of $\mathcal{L}$, b_j (j=1…$|\mathcal{L}|$), have the following properties

$$b_j \bigcap b_k = \phi$$
$$\{x_1, x_2, ...x_P\} = \bigcup_{j=1}^{|\mathcal{L}|} b_j \tag{3-13}$$

The union of all subsets b_j for a given partition $\mathcal{L}$ forms the complete index set $\{x_1 \ldots x_P\}$. The $|\bullet|$ in (3-12) and (3-13) is the sum of elements of that particular set $\mathcal{L}$. The relationship between cumulants and moments, given in (3-12), is a bit confusing and can be better understood through an example.

If we wish to find the third order cumulant $C_3[\mathbf{x}]=C[x_1x_2x_3]$ from equal or lower order moments, the set $\{x_1,x_2,x_3\}$ can be distinctively partitioned into

$|\mathcal{L}|=p=1$: $b_1=\{x_1x_2x_3\}$

$|\mathcal{L}|=p=2$: $b_1=\{x_1x_2\}$, $b_2=\{x_3\}$

$b_1=\{x_1x_3\}$, $b_2=\{x_2\}$

$b_1=\{x_2x_3\}$, $b_2=\{x_1\}$

$|\mathcal{L}|=p=P=3$: $b_1=\{x_1\}$, $b_2=\{x_2\}$, $b_3=\{x_3\}$

as required in (3-12). Applying the above sets of $\mathcal{L}$ to the equation (3-12) results in the third order cumulant as a function of third, second, and first order moments as follows

$$C[x_1x_2x_3] = M[x_1x_2x_3] - M[x_1x_2]M[x_3] - M[x_1x_3]M[x_2] - M[x_2x_3]M[x_1] + 2M[x_1]M[x_2]M[x_3]$$

The above procedure can be extended for higher-order cumulants but the number of possible partitions increases exponentially with order P and so do the necessary lower order moments that are required. Up to now, the discussion was limited to moments and cumulants of real random variables. For certain conditions, the signal vector **x** is complex, however, as is often the case in signal processing applications. The next subsection will generalize the concept of cumulants to complex variables.

3.2.3 Cumulants and Moments of Complex Signals

There are many ways to express cumulants or moments of complex random variables. A fourth order cumulant and moment of a complex signal, $C_4[\mathbf{x}]$, can be expressed as

$$\begin{array}{ll} C\left[x_1 x_2 x_3 x_4\right] & M\left[x_1 x_2 x_3 x_4\right] \\ C\left[x_1^* x_2 x_3 x_4\right] & M\left[x_1^* x_2 x_3 x_4\right] \\ C\left[x_1^* x_2^* x_3^* x_4\right] & M\left[x_1^* x_2^* x_3^* x_4\right] \\ C\left[x_1 x_2^* x_3^* x_4\right] & M\left[x_1 x_2^* x_3^* x_4\right] \\ \vdots & \end{array}$$

just to name a few. As opposed to real signals, the higher order statistics of complex signals have a non-unique way of being represented. In some cases, a particular feature can be extracted from a collection of signals when the right conjugation is applied to the individual signals. For example, the second order moment (which is also the second order cumulant) of an equal likelihood QPSK signals which can take on the values {1+j,1-j,-1+j,-1-j} illustrates this property. The constellation diagram for such a signal is shown in Figure 3.1.

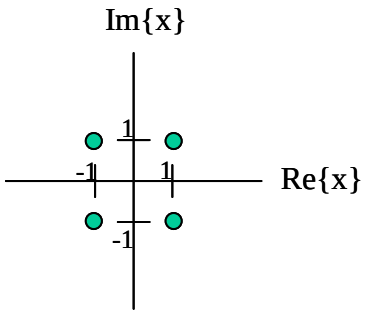

Figure 3.1 Constellation diagram of a QPSK signal.

When each of the signals of the QPSK constellation is equally probable, the second order moment (or cumulant in this case) for two different combinations of conjugation can be found to be

$$M[\mathrm{x}\mathrm{x}^*] = C[\mathrm{x}\mathrm{x}^*] = E[\mathrm{Im}\{\mathrm{x}\}^2] + E[\mathrm{Re}\{\mathrm{x}\}^2]$$
$$M[\mathrm{x}^2] = C[\mathrm{x}^2] = E[\mathrm{Re}\{\mathrm{x}\}^2] - E[\mathrm{Im}\{\mathrm{x}\}^2] + 2jE[\mathrm{Im}\{\mathrm{x}\}\mathrm{Re}\{\mathrm{x}\}] = 0$$

In the first case above, the second order moment has a value which is the summation of the individual marginal order moments. In the latter case, the result is zero since it compares the difference of the second marginal order moments which in this case are equal (the last term in the second equation is zero). The conjugations of the values of x depend on what feature of the probability distribution is to be extracted. Keeping this in mind, in this dissertation, a complex cumulant of order four will be defined as

$$C_4[\mathbf{x}] = C[x_1 x_2^* x_3^* x_4] \tag{3-14}$$

unless otherwise stated. The cumulant tensor, labeled as $\mathcal{C}_4[\mathbf{x}]$, contains all possible fourth order cumulants of the form (3-14) and has symmetric properties that can be exploited. Cumulant tensors will be defined and addressed in more detail in the next section. (3-12) is also valid for complex signals as long as the conjugation state of a particular element $\{x_1,x_2,\ldots,x_P\}$ is maintained throughout the equation. Moments are simpler to compute directly from measured data than cumulants but cumulants have special properties that make them attractive in signal processing applications as will be seen in the following section.

3.2.4 Cumulant and Moment Properties

This section will examine some of the interesting properties of cumulants and why they are useful in the blind source separation problem. Some of these special properties are

also given in [3]. They can all be derived from the definitions of the moments and cumulants presented in this chapter.

3.2.4.1 Scaling

When the individual variables $x_1, x_2, \ldots, x_P$ are pre-multiplied by the constants $a_1, a_2, \ldots, a_P$ respectively, it has the following effect on the moments and the cumulants

$$M\left[a_1 x_1, a_2 x_2, ..., a_P x_P\right] = \prod_{i=1}^{P} a_i M\left[x_1, x_2, ..., x_P\right] \tag{3-15a}$$

$$C\left[a_1 x_1, a_2 x_2, ..., a_P x_P\right] = \prod_{i=1}^{P} a_i C\left[x_1, x_2, ..., x_P\right] \tag{3-15b}$$

The cumulant of the scaled variables of x_i is equivalent to pre-multiplying the cumulant and moment of the set $\{x_1, x_2, \ldots, x_P\}$ by all multiplicative constants a_i.

3.2.4.2 Multi-linearity

Multi-linearity of higher order statistics is a very important property of cumulants and moments. Since a linear channel can be represented by a matrix relating the input to the output of the system so do their respective cumulants and moments but in a multi-linear fashion. This property is central to the development of the theory presented in this dissertation. It relates the cumulants and moments of the input and the output after undergoing a linear mixing operation through the matrix H (NxM).

A linear system relating the input $\mathbf{x}$ to the output $\bar{\mathbf{x}}$ can be written as

$$\bar{\mathbf{x}} = H\mathbf{x} \Rightarrow \left(\bar{\mathbf{x}}\right)_i = \sum_{j=1}^{M} \left(\mathbf{x}\right)_j \left(H\right)_{ij} \tag{3-16}$$

and their respective P order moments and cumulants can be related to each other as follows

$$\left(\mathcal{M}_{\mathrm{P}}\left[\bar{\mathbf{x}}\right]\right)_{ijkl\ldots} = \sum_{p,q,r,s,\ldots=1}^{\mathrm{M}} \left(\mathcal{M}_{\mathrm{P}}\left[\mathbf{x}\right]\right)_{pqrs\ldots} \left(H\right)_{ip} \left(H\right)_{jq} \left(H\right)_{kr} \left(H\right)_{ls} \ldots. \tag{3-17a}$$

$$\left(\boldsymbol{C}_{\mathrm{P}}\left[\bar{\mathbf{x}}\right]\right)_{ijkl\ldots} = \sum_{p,q,r,s,\ldots=1}^{\mathrm{M}} \left(\boldsymbol{C}_{\mathrm{P}}\left[\mathbf{x}\right]\right)_{pqrs\ldots} \left(H\right)_{ip} \left(H\right)_{jq} \left(H\right)_{kr} \left(H\right)_{ls} \ldots. \tag{3-17b}$$

The above equation relates the input and output cumulants and moments of a particular order P and is referred to as mult-linear because it is a higher order extension to the linear relationship given in (3-16). The cumulants and moments $\boldsymbol{C}_{\mathrm{P}}\left[\bar{\mathbf{x}}\right]$, $\mathcal{M}_{\mathrm{P}}\left[\bar{\mathbf{x}}\right]$, $\boldsymbol{C}_{\mathrm{P}}\left[\mathbf{x}\right]$, and $\mathcal{M}_{\mathrm{P}}\left[\mathbf{x}\right]$ are not constants as in (3-3), (3-7), (3-8), and (3-12) but instead are tensors. $\boldsymbol{C}_{\mathrm{P}}\left[\bar{\mathbf{x}}\right]$ and $\mathcal{M}_{\mathrm{P}}\left[\bar{\mathbf{x}}\right]$ are NxNx…N (P dimensions of N) tensors while $\boldsymbol{C}_{\mathrm{P}}\left[\mathbf{x}\right]$ and $\mathcal{M}_{\mathrm{P}}\left[\mathbf{x}\right]$ are MxMx…M (P dimensions of M) tensors. The *i,j,k,l* element of the moment and cumulant tensor can be related to the moments and cumulants in (3-7), (3-8), and (3-12) as follows for the fourth order (P=4) case

$$\left(\mathcal{M}_{4}\left[\mathbf{x}\right]\right)_{ijkl} = M\left[\left(\mathbf{x}\right)_{i}\left(\mathbf{x}\right)_{j}\left(\mathbf{x}\right)_{k}\left(\mathbf{x}\right)_{l}\right] \tag{3-18a}$$

$$\left(\boldsymbol{C}_{4}\left[\mathbf{x}\right]\right)_{ijkl} = C\left[\left(\mathbf{x}\right)_{i}\left(\mathbf{x}\right)_{j}\left(\mathbf{x}\right)_{k}\left(\mathbf{x}\right)_{l}\right] \tag{3-18b}$$

and the above can be generalized to higher or lower orders in a similar fashion. For complex signals, the appropriate conjugation needs to be applied to the elements of **x** in the cumulant function.

3.2.4.3 Symmetry

Moments and cumulants have symmetric properties when an arbitrary permutation is applied to their indices. When the permutation operator Δ is applied to the indexes of the set $\{x_1, x_2, \ldots, x_{\mathrm{P}}\}$, there is no effect on the moment or cumulant when compared to the

un-permuted case for a real valued **x**. This property of cumulants and moments can be mathematically written as

$$M\left[x_1 x_2 \ldots x_{\mathrm{P}}\right] = M\left[x_{\Delta(1)} x_{\Delta(2)} \ldots x_{\Delta(\mathrm{P})}\right] \tag{3-19a}$$

$$C\left[x_1 x_2 \ldots x_{\mathrm{P}}\right] = C\left[x_{\Delta(1)} x_{\Delta(2)} \ldots x_{\Delta(\mathrm{P})}\right] \tag{3-19b}$$

The above moments and cumulants are components of the tensors $\boldsymbol{C}_{\mathrm{P}}\left[\bar{\mathbf{x}}\right]$, $\boldsymbol{\mathcal{M}}_{\mathrm{P}}\left[\bar{\mathbf{x}}\right]$, $\boldsymbol{C}_{\mathrm{P}}\left[\mathbf{x}\right]$, and $\boldsymbol{\mathcal{M}}_{\mathrm{P}}\left[\mathbf{x}\right]$ as described in (3-18). Therefore, these tensors possess super symmetric properties similar to the ones observed in Hermitian matrices. In the complex case, the cumulant and moment tensors have complex conjugate symmetry depending on the complex representation that is being used.

3.2.4.4 Independent Variables

When all the variables associated by $\{x_1, x_2, \ldots, x_{\mathrm{P}}\}$ are independent (the elements of **x** are independent), the joint cumulant of these random signals is zero:

$$C\left[x_1 x_2 \ldots x_{\mathrm{P}}\right] = 0 \tag{3-20}$$

The above result follows from the independence property of the joint probability density function which can be written as the multiplication of its marginal densities. The joint moment, because of independence, separates into a multiplication of the lower order moments. As discussed earlier, cumulants are independent of the lower order moments which are subtracted out resulting in cumulant value of zero.

The property (3-20) applies only to cumulants, however, and does not hold in general for moments of all order. This is another important advantage of cumulants over moments, if the cumulants are zero for all orders, the variables are independent or vice versa.

3.2.4.5 Sum of Independent Variables

Another property that pertains exclusively to cumulants applies to the addition of independent random variables. If the variable associated with the set $\{x_1, x_2, \ldots, x_P\}$ (elements of **x**) are independent of the variables associated with the set $\{y_1, y_2, \ldots, y_P\}$ (elements of **y**), the cumulant of their sum results in the sum of their cumulants.

$$C\left[(y_1 + x_1)(y_2 + x_2)...(y_P + x_P)\right] = C\left[y_1 y_2 ... y_P\right] + C\left[x_1 x_2 ... x_P\right] \tag{3-21}$$

This is another important property of cumulants that does not hold for moments. This property can be derived by noting that the probability density function of the sum of two independent random vectors (**z**=**x**+**y**) is the convolution of their probability density functions or the multiplication of their characteristic functions.

$$\Phi_{\mathbf{Z}}(\mathbf{w}) = \Phi_{\mathbf{Y}}(\mathbf{w})\Phi_{\mathbf{X}}(\mathbf{w}) \tag{3-22}$$

The cumulant generating function $\Psi_{\mathbf{Z}}(\mathbf{w})$, from where the cumulants can be computed, is given in (3-2). For the addition of independent variables, the cumulant generating function separates into the addition of the individual generating functions $\Psi_{\mathbf{X}}(\mathbf{w})$ and $\Psi_{\mathbf{Y}}(\mathbf{w})$.

$$\Psi_{\mathbf{Z}}(\mathbf{w}) = \ln \Phi_{\mathbf{Z}}(\mathbf{w}) = \ln \Phi_{\mathbf{X}}(\mathbf{w}) + \ln \Phi_{\mathbf{Y}}(\mathbf{w}) = \Psi_{\mathbf{X}}(\mathbf{w}) + \Psi_{\mathbf{Y}}(\mathbf{w}) \tag{3-23}$$

The above result leads directly to the property of the cumulant given in (3-21), this property does not apply to joint moments.

3.2.4.6 Cumulants of Gaussian Signals

First of all, since a Gaussian random variable is completely described by its first and second order moments, all higher order moments are a combination of the first and

second order moment. The cumulant of a single random variable for an order larger than 2 (P≥3) is therefore equal to zero.

$$\textit{Gaussian}: \quad C[\mathbf{x}^{P}]=0 \qquad P\geq 3 \tag{3-24}$$

When the above result is considered together with the property described by (3-21) it can be concluded that higher order (P≥3) cumulants are blind to additive Gaussian noise. Cumulant HOS (P≥3) based signal processing makes the desired signal seem noiseless even in the presence of relatively high additive Gaussian noise.

3.2.4.7 Cumulants of Even Distributions

Many probability density functions of signals encountered in digital communication systems have even distributions (they are symmetric about the origin). Some of these modulations are BPSK, QPSK, Q-AM, and M-PSK. For these types of signals, the ones with even probability density function, the resulting odd moments and cumulants are zero.

3.2.5 Cumulants from Measurements (k statistics)

In the previous section some of the most important properties of moments and cumulants were reviewed. Cumulants can be computed from the probability density function directly or from samples drawn from its distribution. When cumulants are computed from a set of measured data they are referred to as k-statistics. The multivariate version of the k-statistic is known as generalized k-statistics. k-statistics ($\hat{C}[x_1 x_2 ... x_P]$) are unbiased estimators of the cumulants with the following properties

$$\begin{aligned} E\left[\hat{C}[x_1 x_2 ... x_P]\right] &= C[x_1 x_2 ... x_P] \\ \hat{C}[x_1 x_2 ... x_P]_{T\to\infty} &= C[x_1 x_2 ... x_P] \end{aligned} \tag{3-25}$$

On the average, the estimate of k-statistic will match the true value of the cumulant. Additionally, when the total number of samples (T) approaches infinity, the k-statistic will equal the true cumulant with certainty. Multivariate k-statistics have been explored in [4] and [5]; in the former an analysis is made on the distribution of the particular k-statistic as a function of time samples T. Knowing the distribution of the k-statistics allows for an estimate on the number of independent samples T that are required to guarantee a certain accuracy of the cumulant.

As an example, a cumulant of order two and three (P=2, P=3) respectively of a vector **x** that contains two random variables is given by

P=2:

$$C[(\mathbf{x})_1(\mathbf{x})_2] = M[(\mathbf{x})_1(\mathbf{x})_2] - M[(\mathbf{x})_1]M[(\mathbf{x})_2]$$

P=3:

$$C[(\mathbf{x})_1{}^2(\mathbf{x})_2] = M[(\mathbf{x})_1{}^2(\mathbf{x})_2] - 2M[(\mathbf{x})_1(\mathbf{x})_2]M[(\mathbf{x})_1] - M[(\mathbf{x})_1{}^2]M[(\mathbf{x})_2] + 2M[(\mathbf{x})_1]^2 M[(\mathbf{x})_2]$$

which can be developed from (3-12). The unbiased estimate of the cumulant (generalized k-statistic) based on T samples of the random vector **x** for the same cases as above can be found in [4] as

P=2:

$$k_{11} = \hat{C}[(\mathbf{x})_1(\mathbf{x})_2] = \frac{1}{\mathrm{T}-1}\left(s_{11} - \frac{1}{\mathrm{T}}s_{10}s_{01}\right)$$

P=3:

$$k_{21} = \hat{C}[(\mathbf{x})_1{}^2(\mathbf{x})_2]_{\mathrm{T}} = \frac{\mathrm{T}}{(\mathrm{T}-1)(\mathrm{T}-2)}\left(s_{21} - \frac{2}{\mathrm{T}}s_{10}s_{11} - \frac{1}{\mathrm{T}}s_{20}s_{01} + \frac{2}{\mathrm{T}}s_{10}^2 s_{01}\right)$$

where

$$s_{ab} = \sum_{t=1}^{\mathrm{T}} (\mathbf{x}(t))_1^a (\mathbf{x}(t))_2^b$$

The resemblance between the cumulant and its estimate is evident and both equal each other when the number of samples T approaches infinity. For large T, the s_{ab}/T estimate is representative of the joint moment $M[(\mathbf{x})_1^a(\mathbf{x})_2^b]$. The complexity of the multivariable k-statistics grows increasingly complex with order and for an arbitrary order the reader is referred to [4] and [5] for the details.

3.2.6 Cumulants for Time-varying HOS

Signals that change with time, in general, also have statistics that are dependent on time. The above sections were mostly concerned with spatial dependence across the elements of a vector **x** with no regard to possible time variations of the signal. The definition of moments and cumulants can be generalized for signals that have time dependencies. In this dissertation, the time-dependent cumulant and moment function will be defined as follows

$$\begin{aligned} C_{\mathrm{P}}[\mathbf{x};t,\boldsymbol{\tau}] &= C[x_1(t-\tau_1)x_2(t-\tau_2)...x_{\mathrm{P}}(t-\tau_{\mathrm{P}})] \\ M_{\mathrm{P}}[\mathbf{x};t,\boldsymbol{\tau}] &= M[x_1(t-\tau_1)x_2(t-\tau_2)...x_{\mathrm{P}}(t-\tau_{\mathrm{P}})] \end{aligned} \tag{3-26}$$

where P is the order of the statistic and $x_1,x_2,..,x_{\mathrm{P}}$ can be any of the elements of the vector $\mathbf{x}(t)$ (with arbitrary conjugations) as in Section 3.2.2. Note that the above equations are in general a function of time t and of the delays τ_p. The signal vector **x** is said to be *stationary* if its cumulants and moments of all order are just a function of the delays $\boldsymbol{\tau}$ and not of t or equivalently

$$\begin{aligned} C_{\mathrm{P}}[\mathbf{x};\boldsymbol{\tau}] &= C_{\mathrm{P}}[\mathbf{x};t,\boldsymbol{\tau}] \\ M_{\mathrm{P}}[\mathbf{x};\boldsymbol{\tau}] &= M_{\mathrm{P}}[\mathbf{x};t,\boldsymbol{\tau}] \end{aligned} \tag{3-27}$$

A similar extension, which incorporates the temporal domain, can be applied to moment and cumulant tensors. The form of the time varying second and fourth order moment and cumulant tensor can be defined as

$$
\begin{aligned}
\left(\mathcal{M}_2\left[\mathbf{x};t,\boldsymbol{\tau}\right]\right)_{ij} &= M\left[\left(\mathbf{x}(t-\tau_1)\right)_i\left(\mathbf{x}(t-\tau_2)\right)_j^*\right] = E\left[\left(\mathbf{x}(t-\tau_1)\right)_i\left(\mathbf{x}(t-\tau_2)\right)_j^*\right] \\
\left(\mathcal{C}_2\left[\mathbf{x};t,\boldsymbol{\tau}\right]\right)_{ij} &= C\left[\left(\mathbf{x}(t-\tau_1)\right)_i\left(\mathbf{x}(t-\tau_2)\right)_j^*\right] = \left(\mathcal{M}_2\left[\mathbf{x};t,\boldsymbol{\tau}\right]\right)_{ij} \\
\left(\mathcal{M}_4\left[\mathbf{x};t,\boldsymbol{\tau}\right]\right)_{ijkl} &= M\left[\left(\mathbf{x}(t-\tau_1)\right)_i\left(\mathbf{x}(t-\tau_2)\right)_j^*\left(\mathbf{x}(t-\tau_3)\right)_k^*\left(\mathbf{x}(t-\tau_4)\right)_l\right] \\
\left(\mathcal{C}_4\left[\mathbf{x};t,\boldsymbol{\tau}\right]\right)_{ijkl} &= C\left[\left(\mathbf{x}(t-\tau_1)\right)_i\left(\mathbf{x}(t-\tau_2)\right)_j^*\left(\mathbf{x}(t-\tau_3)\right)_k^*\left(\mathbf{x}(t-\tau_4)\right)_l\right]
\end{aligned}
\tag{3-28}
$$

which again is a generalization of the spatial definition in (3-18). Only the second and fourth order tensor are shown above because this dissertation focuses on these two statistics, any higher order cumulant tensor can be formed following the same template. Note that the conjugates have been applied to the second and third element of the fourth order cumulant which is consistent with how they will be used in this dissertation unless otherwise specified. Also since the mean of the signal $\mathbf{x}(t)$ is assumed to be zero, the second order cumulant tensor is equal to the second order moment tensor. Each individual element of the cumulant tensors is a function of time t and delay $\boldsymbol{\tau}$ and can be expressed in terms of equal and lower order time-varying moments as follows

$$
\begin{aligned}
\left(\mathcal{C}_2\left[\mathbf{x};t,\boldsymbol{\tau}\right]\right)_{ij} &= M\left[\left(\mathbf{x}(t-\tau_1)\right)_i\left(\mathbf{x}(t-\tau_2)\right)_j^*\right] \\
\left(\mathcal{C}_4\left[\mathbf{x};t,\boldsymbol{\tau}\right]\right)_{ijkl} &= M\left[\left(\mathbf{x}(t-\tau_1)\right)_i\left(\mathbf{x}(t-\tau_2)\right)_j^*\left(\mathbf{x}(t-\tau_3)\right)_k^*\left(\mathbf{x}(t-\tau_4)\right)_l\right] \\
&\quad - M\left[\left(\mathbf{x}(t-\tau_1)\right)_i\left(\mathbf{x}(t-\tau_2)\right)_j^*\right]M\left[\left(\mathbf{x}(t-\tau_3)\right)_k^*\left(\mathbf{x}(t-\tau_4)\right)_l\right] \\
&\quad - M\left[\left(\mathbf{x}(t-\tau_1)\right)_i\left(\mathbf{x}(t-\tau_3)\right)_k^*\right]M\left[\left(\mathbf{x}(t-\tau_2)\right)_j^*\left(\mathbf{x}(t-\tau_4)\right)_l\right] \\
&\quad - M\left[\left(\mathbf{x}(t-\tau_1)\right)_i\left(\mathbf{x}(t-\tau_4)\right)_l\right]M\left[\left(\mathbf{x}(t-\tau_2)\right)_j^*\left(\mathbf{x}(t-\tau_3)\right)_k^*\right]
\end{aligned}
\tag{3-29}
$$

or in short hand form

$$
\begin{aligned}
\left(\mathcal{C}_2\left[\mathbf{x};t,\boldsymbol{\tau}\right]\right)_{ij} &= M_{ij^*}\left[\mathbf{x};t,\boldsymbol{\tau}\right] \\
\left(\mathcal{C}_4\left[\mathbf{x};t,\boldsymbol{\tau}\right]\right)_{ijkl} &= M_{ij^*k^*l}\left[\mathbf{x};t,\boldsymbol{\tau}\right] - M_{ij^*}\left[\mathbf{x};t,\boldsymbol{\tau}\right]M_{k^*l}\left[\mathbf{x};t,\boldsymbol{\tau}\right] \\
&\quad - M_{ik^*}\left[\mathbf{x};t,\boldsymbol{\tau}\right]M_{j^*l}\left[\mathbf{x};t,\boldsymbol{\tau}\right] - M_{il}\left[\mathbf{x};t,\boldsymbol{\tau}\right]M_{j^*k^*}\left[\mathbf{x};t,\boldsymbol{\tau}\right]
\end{aligned}
\tag{3-30}
$$

The above equations describe the general form for the time-varying cumulant statistics. In this dissertation, a special class of signals that have time-variant statistics will be examined which are referred to as *cyclostationary* signals. *Cyclostationary* signals have statistics that change with time but have the special property that its moments and cumulants are periodic in *t*. The next subsection will be devoted to examining these types of signals in more detail.

3.2.4.7 Cumulants and moments of Cyclostatinary Signals

Cyclostationarity is introduced into a random signal usually through a form of periodic time-variant filtering. In communication systems, periodic time-variant filtering is performed by the modulator which introduces periodicities such a data rate and carrier frequency into the signal before transmission. These periodicities also exhibit themselves in the statistics of the signal making the cumulants and moments of the signal periodic in *t* (see Chapter 6). Cyclostationarity is based on time averages (instead of ensemble averages) and numerous works exist on this subject among the more relevant ones to this dissertation are [8] and [9]. In this chapter, a brief summary will be given to introduce the concepts and describe the variables that will be used in subsequent chapters, for more detailed treatment on cyclostationary statistics see [8]. This dissertation is mostly focused in applying these statistics to solve the problem of blind identification. Their cyclic properties allows the second cumulant tensor of joint random cyclostationary signals **x**(t) to be decomposed into a Fourier transform as follows

$$\begin{aligned} \boldsymbol{C}_2[\mathbf{x};t,\boldsymbol{\tau}] &= \sum_{\alpha} \boldsymbol{C}_2^{\alpha}[\mathbf{x};\boldsymbol{\tau}] e^{j2\pi\alpha t} \\ \left(\boldsymbol{C}_2^{\alpha}[\mathbf{x};\boldsymbol{\tau}]\right)_{ij} &= M_{ij^*}^{\alpha}[\mathbf{x};\boldsymbol{\tau}] \\ \left(\boldsymbol{C}_2[\mathbf{x};t,\boldsymbol{\tau}]\right)_{ij} &= \sum_{\alpha} M_{ij^*}^{\alpha}[\mathbf{x};\boldsymbol{\tau}] e^{j2\pi\alpha t} \end{aligned} \tag{3-31}$$

$$M_{ij^*}^{\alpha}[\mathbf{x};\boldsymbol{\tau}] = \lim_{T\to\infty} \frac{1}{T} \int_{-T/2}^{T/2} \left(\mathbf{x}(t-\tau_1)\right)_i \left(\mathbf{x}(t-\tau_2)\right)_j^* e^{-j2\pi\alpha t} dt$$

where $M_{ij^*}^{\alpha}[\mathbf{x};\boldsymbol{\tau}]$ are the Fourier coefficients for the cycle frequency α and can be obtained by a complex-exponential weighted time-average of the product of the corresponding two elements in $\mathbf{x}(t)$. Note that $M_{ij^*}^{\alpha}[\mathbf{x};\boldsymbol{\tau}]$ is non-zero only for a denumerated set of cyclic frequencies α and is only a function of the delays $\boldsymbol{\tau}$. $\boldsymbol{C}_2^{\alpha}[\mathbf{x};\boldsymbol{\tau}]$ corresponds to the *second order cyclic cumulant tensor* (SCCT) and is of size MxM. Equivalently, the fourth order cumulant tensor can be decomposed into a Fourier transform as follows

$$\boldsymbol{C}_4[\mathbf{x};t,\boldsymbol{\tau}] = \sum_{\alpha} \boldsymbol{C}_4^{\alpha}[\mathbf{x};\boldsymbol{\tau}] e^{j2\pi\alpha t} \tag{3-32}$$

where the elements of $\boldsymbol{C}_4^{\alpha}[\mathbf{x};\boldsymbol{\tau}]$ can be found from

$$\begin{aligned}
\left(\boldsymbol{C}_4^{\alpha}[\mathbf{x};\boldsymbol{\tau}]\right)_{ijkl} &= \lim_{T\to\infty} \frac{1}{T} \int_{-T/2}^{T/2} \left(\boldsymbol{C}_4[\mathbf{x};t,\boldsymbol{\tau}]\right)_{ijkl} e^{-j2\pi\alpha t} dt \\
\left(\boldsymbol{C}_4^{\alpha}[\mathbf{x};\boldsymbol{\tau}]\right)_{ijkl} &= M_{ij^*k^*l}^{\alpha}[\mathbf{x};\boldsymbol{\tau}] - \sum_{\beta} M_{ij^*}^{\beta}[\mathbf{x};\boldsymbol{\tau}] M_{k^*l}^{\alpha-\beta}[\mathbf{x};\boldsymbol{\tau}] \\
&\quad - \sum_{\beta} M_{ik^*}^{\beta}[\mathbf{x};\boldsymbol{\tau}] M_{j^*l}^{\alpha-\beta}[\mathbf{x};\boldsymbol{\tau}] - \sum_{\beta} M_{il}^{\beta}[\mathbf{x};\boldsymbol{\tau}] M_{j^*k^*}^{\alpha-\beta}[\mathbf{x};\boldsymbol{\tau}]
\end{aligned} \tag{3-33}$$

The tensor $\boldsymbol{C}_4^{\alpha}[\mathbf{x};\boldsymbol{\tau}]$ is obtained by integrating equation (3-32) as shown in (3-33). $\boldsymbol{C}_4^{\alpha}[\mathbf{x};\boldsymbol{\tau}]$ is the *fourth order cyclic cumulant tensor* (FCCT) which is a function of delay $\boldsymbol{\tau}$ and frequency α and is of size MxMxMxM. Basically, $\boldsymbol{C}_4^{\alpha}[\mathbf{x};\boldsymbol{\tau}]$ corresponds to the Fourier coefficients of the time-varying cumulant tensor $\boldsymbol{C}_4[\mathbf{x};t,\boldsymbol{\tau}]$ at a specific cycle α (3-33). The FCCT above is in turn made up of the cyclic moment functions (CMF) of order four or lower which can be computed by performing a Fourier transform of their respective time-varying moments or equivalently on the time signal $\mathbf{x}(t)$ as follows

$$M_{ij^*k^*l}^{\alpha}[\mathbf{x};\boldsymbol{\tau}] = \lim_{T\to\infty} \frac{1}{T} \int_{-T/2}^{T/2} \left(\mathbf{x}(t-\tau_1)\right)_i \left(\mathbf{x}(t-\tau_2)\right)_j^* \left(\mathbf{x}(t-\tau_3)\right)_k^* \left(\mathbf{x}(t-\tau_4)\right)_l e^{-j2\pi\alpha t} dt \tag{3-34}$$

$$M_{ij^*}^{\alpha}[\mathbf{x};\boldsymbol{\tau}] = \lim_{T\to\infty}\frac{1}{T}\int_{-T/2}^{T/2}\left(\mathbf{x}(t-\tau_1)\right)_i\left(\mathbf{x}(t-\tau_2)\right)_j^* e^{-j2\pi\alpha t}dt$$

$$M_{k^*l}^{\alpha}[\mathbf{x};\boldsymbol{\tau}] = \lim_{T\to\infty}\frac{1}{T}\int_{-T/2}^{T/2}\left(\mathbf{x}(t-\tau_3)\right)_k^*\left(\mathbf{x}(t-\tau_4)\right)_l e^{-j2\pi\alpha t}dt \qquad (3\text{-}35)$$

$$M_{ik^*}^{\alpha}[\mathbf{x};\boldsymbol{\tau}] = \lim_{T\to\infty}\frac{1}{T}\int_{-T/2}^{T/2}\left(\mathbf{x}(t-\tau_1)\right)_i\left(\mathbf{x}(t-\tau_3)\right)_j^* e^{-j2\pi\alpha t}dt$$

$$M_{j^*l}^{\alpha}[\mathbf{x};\boldsymbol{\tau}] = \lim_{T\to\infty}\frac{1}{T}\int_{-T/2}^{T/2}\left(\mathbf{x}(t-\tau_2)\right)_j^*\left(\mathbf{x}(t-\tau_4)\right)_l e^{-j2\pi\alpha t}dt$$

$$M_{il}^{\alpha}[\mathbf{x};\boldsymbol{\tau}] = \lim_{T\to\infty}\frac{1}{T}\int_{-T/2}^{T/2}\left(\mathbf{x}(t-\tau_1)\right)_i\left(\mathbf{x}(t-\tau_4)\right)_l e^{-j2\pi\alpha t}dt$$

$$M_{j^*k^*}^{\alpha}[\mathbf{x};\boldsymbol{\tau}] = \lim_{T\to\infty}\frac{1}{T}\int_{-T/2}^{T/2}\left(\mathbf{x}(t-\tau_2)\right)_j^*\left(\mathbf{x}(t-\tau_4)\right)_k^* e^{-j2\pi\alpha t}dt$$

The properties of *cyclic cumulant tensors* (CCT) are similar to the properties of general cumulant tensors given in section 3.2.4. The scaling, symmetry, independence, and the multi-linearity property can be verified to also apply to CCTs of all order [8].

Some special case results that are of interest in this dissertation will be considered next. It will be important, in subsequent chapters, to know how certain operations on the signals of interest affect the cyclic cumulant and moment tensors. A one dimensional cyclostationary signal x(t) that undergoes time-invariant filtering, for example, as in

$$y(t) = \int_{-\infty}^{\infty} h(\lambda)x(t-\lambda)d\lambda \qquad (3\text{-}36)$$

has the following effect on the cyclic cumulants [8] of y(t)

$$\boldsymbol{C}_2^{\alpha}[y;\boldsymbol{\tau}] = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} h^*(\lambda_2)h(\lambda_1)\boldsymbol{C}_2^{\alpha}[x;\boldsymbol{\tau}-\boldsymbol{\lambda}]d\boldsymbol{\lambda} \qquad (3\text{-}37)$$

$$C_4^{\alpha}[y;\boldsymbol{\tau}] = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} h(\lambda_4)h^*(\lambda_3)h^*(\lambda_2)h(\lambda_1)C_4^{\alpha}[x;\boldsymbol{\tau}-\boldsymbol{\lambda}]d\boldsymbol{\lambda}$$

This is similar to the multi-linear property presented in Section 3.2.4 but extended to time dispersive filters instead of just mixing matrices.

Before ending the discussion on cyclostationarity, it is appropriate to mention that cyclostationarity is based on non-stochastic time averaging instead of ensemble averaging which is a stochastic approach. There are some fundamental differences in these two approaches and for more details the interested reader is directed to the reference [9].

3.3 Measures of Non-Gaussianity

3.3.1 Negentropy

Cumulants as higher order statistics can be viewed as a measure of departure from Gaussian behavior as will be seen shortly. Negentropy quantifies the distance between the probability density function of an arbitrary vector signal ($\mathbf{x}$) and a comparable Gaussian random variable ($\boldsymbol{\phi}$) with the same mean and covariance matrix. The distributions of both signals are compared not directly but through a metric based on differential entropy. Entropy is a mapping of the probability density function onto a real value and can be interpreted as a measure of randomness of a particular distribution. The differential entropy of a signal with probability density function $f_{\mathbf{X}}(\mathbf{x})$ is defined as

$$H_{f_{\mathbf{X}}(\mathbf{x})} = -\int f_{\mathbf{X}}(\mathbf{x})\log\left(f_{\mathbf{X}}(\mathbf{x})\right)d\mathbf{x} \tag{3-38}$$

It is well known that the entropy, (3-38), of jointly Guassian random variables $\boldsymbol{\phi}$ for a given covariance matrix is the largest of all distributions [6]. So the definition of negentropy given by

$$J_{f_{\mathbf{X}}(\mathbf{x})} = H_{f_{\varphi}(\varphi)} - H_{f_{\mathbf{X}}(\mathbf{x})}$$
$$J_{f_{\mathbf{X}}(\mathbf{x})} \geq 0 \tag{3-39}$$

is always greater or equal to zero. In (3-39), $f_{\mathbf{X}}(\mathbf{x})$ and $f_{\boldsymbol{\phi}}(\boldsymbol{\phi})$ are the probability density function of the vector **x** and $\boldsymbol{\phi}$, respectively, under the assumption that both random variables have the same mean and covariance matrix. $J_{f_{\mathbf{X}}(\mathbf{x})}$ is zero only when the distribution of the random variable **x** is jointly Gaussian and therefore equivalent to $\boldsymbol{\phi}$. Negentropy is therefore a measurement that quantifies the deviation from Gaussianity. The less random (less Gaussian) a signal is, the lower its entropy, $H_{f_{\mathbf{X}}(\mathbf{x})}$, and the larger its negentropy, $J_{f_{\mathbf{X}}(\mathbf{x})}$.

3.3.2 Edgeworth Series

The Edgeworth Series is an expansion of an arbitrary univariable probability density function $f_X(x)$ in the vicinity of a Gaussian density. This type of series is represented in terms of higher order cumulants of the probability density function of the distribution being expanded. Through this series, the cumulants of a probability density function can be related to a best matched Gaussian distribution. From the Edgeworth expansion, the departure from normal for a particular distribution can be quantified in the metric of negentropy.

The Edgeworth Series of a scalar random variable about a standard normal distribution (zero mean, variance equal to one) is given in [7] as

$$\begin{aligned}\frac{f_X(x)}{f_\phi(x)} &= 1 + \frac{1}{3!}c_3 h_3(x) + \frac{1}{4!}c_4 h_4(x) + \frac{10}{6!}c_3{}^2 h_6(x) + \frac{1}{5!}c_3 h_5(x) + \frac{35}{7!}c_3 c_4 h_7(x) \\ &+ \frac{280}{9!}c_3{}^2 h_9(x) + \frac{1}{6!}c_6 h_6(x) + \frac{56}{8!}c_3 c_5 h_8(x) + \frac{35}{8!}c_4{}^2 h_8(x) + \frac{2100}{10!}c_3{}^2 c_4 h_{10}(x) \\ &+ \frac{15400}{12!}c_3{}^4 h_{12}(x) + O(m^{-2})\end{aligned} \tag{3-40}$$

where c_n is the cumulant of order n of the probability density function $f_X(x)$ or in the notation used in this manuscript $c_n=C[x^n]$. The Hermite polynomials of degree n, h_n, are defined in recursion from the equations

$$\begin{aligned} h_0(x) &= 1, \qquad h_1(x) = u, \\ h_{n+1}(x) &= xh_n(x) - \frac{\partial}{\partial x} h_n(x) \end{aligned} \tag{3-41}$$

As stated in [7], the advantage of using the Edgeworth expansion is that it provides an ordering of the components in terms of decreasing significance based on the order m. The order m stems from the central limit theorem which says that a variable x which is a result of the sum of m independent variables with finite cumulants has an nth-order cumulant in the order of $m^{(2-n)/2}$. This theorem is attributed to Cramér [7].

As shown in [7], the negentropy of a probability density function about the standard normal distribution is given by

$$J_{f_X(x)} = \frac{1}{12} k_3{}^2 + \frac{1}{48} k_4{}^2 + \frac{7}{48} k_3{}^4 - \frac{1}{8} k_3{}^2 k_4 + O(m^{-2}) \tag{3-42}$$

To be in line with the definition of the negentropy, the distribution that is being expanded in (3-42) also has zero mean and a variance equal to unity. The above expression is therefore a function of the third order cumulant or higher. The result in (3-30) can be developed from the definition of negentropy as well as the Edgeworth expansion given in (3-40) and (3-41).

3.4 Summary

In this chapter, the theory of higher order statistics was introduced and related to the probability density function of a joint random variable. Cumulants were found to possess several advantages over moments when analyzing joint random signals in signal

processing applications. Cumulants of order larger than two are blind to Gaussian signals (and therefore Gaussian noise) and provide a measure of departure from the Gaussian probability density function seen by (3-29) and (3-30).

Independent random variables have zero cross-cumulants. When signals in **x** are statistically independent the cumulant tensor $\boldsymbol{C}_{\mathrm{P}}[\mathbf{x}]$ is therefore a diagonal tensor. Additionally the cumulant of the sum of two independent random variables is the sum of their individual cumulants. For linear systems, the input and output cumulant tensors of a linear system are related in a multilinear fashion (Section 3.2.4.2).

Cumulants of order one through four are commonly used in blind source separation problems and for zero mean signals ($M[x_1]= M[x_2]= M[x_3]= M[x_4]=0$), are given by

$$\begin{aligned} C[x_1x_2x_3x_4] &= M[x_1x_2x_3x_4] - M[x_1x_2]M[x_3x_4] - \\ &\quad M[x_1x_3]M[x_2x_4] - M[x_1x_4]M[x_2x_3] \\ C[x_1x_2x_3] &= M[x_1x_2x_3] \\ C[x_1x_2] &= M[x_1x_2] \end{aligned} \tag{3-43}$$

The above moment-cumulant relationship can be developed from (3-12). For complex signals, the fourth, third, and second order cumulant for zero mean signals that will be adopted in this dissertation unless otherwise stated (Section 3.2.3) is given by

$$\begin{aligned} C[x_1x_2^*x_3^*x_4] &= M[x_1x_2^*x_3^*x_4] - M[x_1x_2^*]M[x_3^*x_4] - \\ &\quad M[x_1x_3^*]M[x_2^*x_4] - M[x_1x_4]M[x_2^*x_3^*] \\ C[x_1x_2^*x_3] &= M[x_1x_2^*x_3] \\ C[x_1x_2^*] &= M[x_1x_2^*] \end{aligned} \tag{3-44}$$

which can be computed directly from the probability distributions or from measurements. When only samples with a particular distribution are available (and not the distribution), the cumulants can be estimated using k-statistics. k-statistics are unbiased estimates of cumulants as explained in section 2.2.5. The following chapter will examine the blind

source separation problem where the higher order statistics developed here will be used to separate unknown signals into their independent components.

References

[1] H. Stark and J. W. Woods, *Probability, Random Processes, and Estimation Theory for Engineers*. New Jersey: Prentice Hall,1994.

[2] S. R. Jammalamadaka, T. S. Rao, G. Terdik, "Higher Order Cumulants of Random Vectors, Differential Operators, and Applications to Statistical Inference and Time Series." Found at http://www.pstat.ucsb.edu/faculty/jammalam/html/research_publication.htm

[3] L. De Lathauwer, *Signal Processing based on Multilinear Algebra*. Ph.D. Dissertation Katholieke Universiteit Leuven, September 1997.

[4] R. A. Fisher, "Moments and Product Moments of Sampling Distributions," *Proceedings of the London Mathematical Society*, Series 2, 30: pp. 199-238, 1929.

[5] T. P. Speed and H. L. Silock, "Cumulants and Partition Lattices V. Calculating Generalized k-statitics," *J. Austral. Math. Soc.* (Series A) 44, pp. 171-196, 1988.

[6] T. M. Cover and J. A. Thomas, *Elements of Information Theory*. New York: John Wiley & Sons, 1991.

[7] P. Common, "Independent Component Analysis, A New Concept?," *Signal Processing*, Special Issue *Higher Order Statistics*, Vol. 36, No. 3, April 1994, pp. 287-314.

[8] C. M. Spooner, Theory and Application of Higher-Order Cyclostationarity, Ph.D. Dissertation University of California, Davis, June 1992.

[9] W. A. Gardner, *Cyclostationarity in Communications and Signal Processing*. New York: IEEE Press, 1994

Chapter 4

Blind Source Separation for Linear Systems (Non-Dispersive Linear Systems)

In this chapter, the problem of *blind source separation* (BSS) of the non-dispersive linear channel is addressed. Under these circumstances, a receiver has no specific *a priori* information about the signals in the system or the mixing matrix H. The receiver is required extract all the necessary information from the received signal(s) with only the knowledge that the source signals $\mathbf{s}$ are statistically independent of each other. As seen in Chapter 2, ideally, the MLSE receiver requires information of H, R_n, and the alphabet of the transmitted signals to make an optimal decision. The ZFE, on the other hand, bases its decision entirely on the knowledge of the channel matrix H.

The question becomes, can a receiver find out all this information by itself from just monitoring the received vector $\mathbf{y}$? This chapter will examine a method for determining the channel matrix H blindly.

4.1 System Model

The linear system model under consideration is the same as the one addressed in Chapter 2 for the linear wireless channel (2-2). The noiseless case will be examined in this chapter to illustrate the concept of BSS and the system under noisy conditions will be considered in a later chapter. The source signals are assumed to be of zero mean which is typical of wireless communication signals, $E[\mathbf{s}]=0$. The vector $\mathbf{y}$ contains the data received by N sensors (antenna elements) from M signals in $\mathbf{s}$ and H is a full column rank matrix related through

$$\mathbf{y}(t) = H\mathbf{s}(t) \tag{4-1}$$

Additionally, the environments where the elements of $\mathbf{s}$ are equal or less than the number of sensors $\mathbf{y}$ are considered in this chapter or $N \geq M$. Since each of the signals in $\mathbf{s}$ are created from a different source, the elements of $\mathbf{s}$ are all assumed to be statistically independent of each other. An extension to more signals than sensors will be addressed in a subsequent chapter.

The system model in (4-1) is 'underdetermined' in the sense that there are more unknowns than equations. We wish to extract H, which is composed of NxM unknowns, and $\mathbf{s}$, which has M unknowns, for the received vector $\mathbf{y}$ which provides only N equations. At first glance, this problem seems impossible, but if a large number of equations of the form (4-1) are available for independent values of $\mathbf{s}$ a solution can be found statistically. The two methods for BSS addressed in this chapter are *principal component analysis* (PCA) and *independent component analysis* (ICA) both of these methods are statistical in nature. The PCA decorrelates the signals in $\mathbf{y}$ using the second order statistic estimate while ICA requires higher order statistics to transform $\mathbf{y}$ into a set of independent components.

The case of N=M (perfectly determined) will be considered in this Chapter which is the worst case scenario under the conditions $N \geq M$. For the instances when N is larger than

M, an equivalent result in terms of pseudo-inverses (instead of full rank square matrix inverses) or through singular value decompositions can be obtained. The number of equations can also be reduced such that they equal the number of signals **s** in the system.

4.2 Ambiguities of Blind Source Separation

When no information is known at the receiver before hand about the signals in the system (except for the statistical independent assumption), certain ambiguities of the linear model are not resolvable. The first of the indeterminacies is that the variance or power of the signals in **s** cannot be determined with certainty. This can be seen by examining the equation given in (4-1):

$$\mathbf{y}=\begin{bmatrix}(\mathbf{y})_1\\(\mathbf{y})_2\\(\mathbf{y})_3\end{bmatrix}=\begin{bmatrix}(H)_{11}&(H)_{12}&(H)_{13}\\(H)_{21}&(H)_{22}&(H)_{23}\\(H)_{31}&(H)_{32}&(H)_{33}\end{bmatrix}\begin{bmatrix}(\mathbf{s})_1\\(\mathbf{s})_2\\(\mathbf{s})_3\end{bmatrix}=\begin{bmatrix}(H)_{11}&1/a(H)_{12}&(H)_{13}\\(H)_{21}&1/a(H)_{22}&(H)_{23}\\(H)_{31}&1/a(H)_{32}&(H)_{33}\end{bmatrix}\begin{bmatrix}(\mathbf{s})_1\\a(\mathbf{s})_2\\(\mathbf{s})_3\end{bmatrix} \quad (4\text{-}2)$$

From the above equation, one can see that there is an ambiguity in the amplitudes of the components **s**. The scaling of the signal $(\mathbf{s})_2$ above can be compensated for by a corresponding scaling of the second column of the matrix *H*. The true value of *a* cannot be determined from just the observed vector **y**. Therefore the amplitudes of any of the elements in **s** are irresolvable. It is for that reason that the signals in **s** are assumed to be of unit variance or that $E[\mathbf{ss}^*]=I$ in order to set one of the indeterminacies in the following sections.

A second ambiguity is related to the ordering of the independent components **s**. As far as the received signal is concerned, a change in the ordering of the signals does not change the received signal. This is illustrated as follows:

$$\mathbf{y}=\begin{bmatrix}(\mathbf{y})_1\\(\mathbf{y})_2\\(\mathbf{y})_3\end{bmatrix}=\begin{bmatrix}(H)_{11}&(H)_{12}&(H)_{13}\\(H)_{21}&(H)_{22}&(H)_{23}\\(H)_{31}&(H)_{32}&(H)_{33}\end{bmatrix}\begin{bmatrix}(\mathbf{s})_1\\(\mathbf{s})_2\\(\mathbf{s})_3\end{bmatrix}=\begin{bmatrix}(H)_{12}&(H)_{11}&(H)_{13}\\(H)_{22}&(H)_{21}&(H)_{23}\\(H)_{32}&(H)_{31}&(H)_{33}\end{bmatrix}\begin{bmatrix}(\mathbf{s})_2\\(\mathbf{s})_1\\(\mathbf{s})_3\end{bmatrix} \quad (4\text{-}3)$$

A column permutation of the vector H does not affect the received signal as seen by the above equation. From (4-3), the first and second column are permuted and so is the vector **s** and yet there is no perceived difference in the received signal **y**. The determination of the matrix H is therefore irresolvable to an arbitrary permutation of its columns.

4.3 Principal Component Analysis - PCA

PCA can be thought of as a method that operates linearly on the received vector **y**. The linear operator transforms **y** into a different vector **w** that has uncorrelated components. This method can also be thought of as whitening, since this operation is a dual operation to flattening the power spectral density (PSD) in the frequency domain. PCA uses the second order statistic of **y** to create uncorrelated components in **w**.

4.3.1 PCA Theory

The purpose of principal component analysis is to determine a matrix W that when applied to the vector **y** will yield an uncorrelated vector **w** with unit variance components. The reasoning behind this procedure is that if the components of **w** are uncorrelated, they will most likely be closer to **s** since these elements are also uncorrelated. The matrix W is therefore taken as an approximation to the inverse of the matrix H and the vector **w** as an approximation to the signal **s**.

$$\mathbf{w} = W\mathbf{y} = WH\mathbf{s} \tag{4-4}$$

It will be shown subsequently that the matrix W is not a true inverse of the matrix H but rather an approximation. As mentioned earlier, PCA works on the second order statistic of **y**, or R_y. The cross-correlation of the signals in **y** are therefore given by

$$R_{\mathrm{y}} = E\left[\mathbf{y}\mathbf{y}^*\right] = E\left[H\mathbf{s}\mathbf{s}^*H^*\right] = HE\left[\mathbf{s}\mathbf{s}^*\right]H^* = HH^* \tag{4-5}$$

The last step in the above equation is a result of the discussion following equation (4-2) were the $E[\mathbf{ss}^*]$ is arbitrarily set to I to constrain one of the ambiguities. The next step in the PCA algorithm is to factor R_y into H and H^* so that the matrix H can be extracted. As is well known in linear algebra, the factoring of R_y is not unique and therefore extracting the exact value of H cannot be guaranteed. Since the matrix H cannot be determined with certainty, W is not necessarily the inverse of H and there are many matrices W that would produce an uncorrelated vector $\mathbf{w}$ one for each possible factor of R_y. This can be also seen by noting that spatial whiteness of $\mathbf{w}$ imposes pair wise decorrelation, $E[(\mathbf{w})_i(\mathbf{w})_j^*]=0$ $\forall$ $i \neq j$ and $E[(\mathbf{w})_i(\mathbf{w})_i^*]=1$ $\forall$ all i, which can be achieved with N(N+1)/2 constraints and which leaves N(N-1)/2 unknowns not determined.

A common way of factoring the hermitian R_y when H is full rank is through an eigenvalue, eigenvector decomposition. This type of decomposition transforms R_y into a multiplication of a matrix V by a diagonal matrix D^2 and then by V^* as follows

$$R_y = VD^2V^* = (VD)(VD)^* = \tilde{H}\tilde{H}^* \tag{4-6}$$

where the estimate of H is given by

$$\tilde{H} = VD \tag{4-7}$$

As stated earlier, factoring R_y will yield an approximation to the matrix H and most likely not the true value of H. Therefore the matrix W is given by the inverse of the estimate $\tilde{H}$ found in (4-7)

$$W = \tilde{H}^{-1} = D^{-1}V^* \tag{4-8}$$
$$\mathbf{w} = W\mathbf{y}$$

As expected, under the definition of W above, the resultant vector $\mathbf{w}$ has equal power uncorrelated components with variance equal to $E[(\mathbf{w})_i(\mathbf{w})_i^*]=1$ and zero for all cross-terms. This can be verified by calculating the cross-correlation matrix of $\mathbf{w}$ as follows

$$R_{\mathbf{w}} = E\left[\mathbf{w}\mathbf{w}^*\right] = E\left[W\mathbf{y}\mathbf{y}^*W^*\right] = WE\left[\mathbf{y}\mathbf{y}^*\right]W^* = W\left(\tilde{H}\tilde{H}^*\right)W^* = I \tag{4-9}$$

which shows that $E[\mathbf{w}\mathbf{w}^*]$ is as desired the identity matrix. The true value of H is however yet to be determined. For the receivers described in Chapter 2, the exact value of H is required and not just one of the possible solutions to the factorization problem described in (4-5). As stated in the discussion following (4-5), finding a unique solution for H requires finding the remaining N(N-1)/2 undetermined coefficients.

4.3.2 Problems with PCA

The PCA problem, which is based on second order statistics, fails to conclusively find the matrix H. There are still N(N-1)/2 indeterminacies that are still unresolved when using PCA. The structure of the remaining coefficients that set the correct value of H, up to, of course, the inherent ambiguities in H as discussed in Section 4.2, will be addressed in this section.

The structure of the missing components can be found by noting that any matrix H can be written as

$$H = \tilde{H}Q \tag{4-10}$$

where $\tilde{H}$ is an arbitrary matrix and Q is a unitary rotational matrix. The reason for such decomposition will become apparent shortly. The unitary matrix Q has the property

$$QQ^* = I \tag{4-11}$$

where Q^* is the inverse and at the same time the hermitian of the matrix Q. Substituting (4-10) into the cross-correlation of the vector **y** in (4-5) yields the following result:

$$R_{\mathrm{y}} = E\left[\mathbf{y}\mathbf{y}^*\right] = E\left[\tilde{H}Q\mathbf{s}\mathbf{s}^*Q^*\tilde{H}^*\right] = \tilde{H}QE\left[\mathbf{s}\mathbf{s}^*\right]Q^*\tilde{H}^* = \tilde{H}QQ^*\tilde{H}^* = \tilde{H}\tilde{H}^* \quad (4\text{-}12)$$

and from (4-5)

$$R_{\mathrm{y}} = \tilde{H}\tilde{H}^* = HH^* \quad (4\text{-}13)$$

Since the decomposition in (4-10) is valid for any matrix $\tilde{H}$, $\tilde{H}$ can be any of the decompositions of HH^* in (4-13). Therefore, the difference between any non-unique factor $\tilde{H}$ of (4-5 or 13) and the actual value of H is a unitary matrix Q.

Finding the matrix Q will be addressed in the next section. Solving for the matrix Q requires mathematics beyond second order statistics. Second order statistics are blind to the unitary matrix Q in (4-10) as shown in (4-12). As mentioned in Chapter 3, the first and second order statistic of an arbitrary signal is the best Gaussian fit to that particular distribution. Joint Gaussian random variables are invariant to orthonormal rotations because they are fully described by their mean and cross-correlation. This insensitivity to Q is due to the fact that their higher order (greater than three) cumulants are zero. The method of independent component analysis presented in the next section is therefore applicable to non-Gaussian signals.

4.4 Independent Component Analysis – ICA

Independent component analysis is a more sophisticated method of blind channel identification than PCA. It attempts to find the missing unitary matrix Q to resolve the factoring indeterminacies in (4-13). The matrix Q can be found through higher than second order statistics as will be explained in this section. The signals of interest in communication systems are non-Gaussian and usually have even probability distributions. These signals therefore have even cumulants with order larger than three

that are non-zero. The ICA algorithm therefore works with the fourth order cumulant which is the next highest cumulant after the second order since the third order is zero for even distributions (Section 3.2.4.7).

4.4.1 Contrast Functions, Φ[•]

The matrix H and the value of the input signals $\mathbf{s}$ can be found from just the knowledge of the received vector $\mathbf{y}$ with the aid of a contrast function $\mathbf{\Phi}$. The contrast functions are defined as functions that reach a minimum when the estimates of H ($\hat{H}$) and $\mathbf{s}$ ($\hat{\mathbf{s}}$) are the same as the true values of H and $\mathbf{s}$ or any combination of $\hat{H}$ and $\hat{\mathbf{s}}$ that is within the resolvable ambiguity discussed in Section 4.2. For the source separation problem we would like to determine a contrast function which when minimized produces independent components (have sources separated from each other) for a particular random vector. The independent components will be a result of an un-mixing matrix which is the inverse of the channel matrix H. This is equivalent to finding a matrix B (de-mixing matrix) that when applied to $\mathbf{y}$ in (4-1) produces a vector $\mathbf{x}$ that contains independent components as follows:

$$\mathbf{x} = B\mathbf{y} = BH\mathbf{s} \tag{4-14}$$

Independence of the elements in $\mathbf{x}$ occurs, for example, when B is equal to H^{-1}. When B is the inverse of H, the vector $\mathbf{x}$ is equal to the transmitted signal $\mathbf{s}$ and therefore has independent components (from the initial assumption that the elements of $\mathbf{s}$ are statistically independent). The other solutions of B that are possible are the ones where $\mathbf{x}$ is a scaled or permuted version of $\mathbf{s}$ as mentioned in Section 4.2 were the resolvability issues of the linear system were examined. Any vector $\mathbf{x}$ that is a scaled or a permuted version of $\mathbf{s}$ also has independent components.

In order to determine a valid contrast function, $\mathbf{\Phi}$, some of the desirable and necessary properties of such a function have to be considered. As stated in [1], the contrast function when applied to $\mathbf{x}$ in (4-14)

$$\boldsymbol{\Phi}[\mathbf{x}] = \boldsymbol{\Phi}[BH\mathbf{s}] \geq \boldsymbol{\Phi}[\mathbf{s}] \tag{4-15}$$

should have the property of being equal or larger than when the contrast function is applied to the statistically independent vector **s**. Equality is achieved in (4-15), and at the same time the minimum of $\boldsymbol{\Phi}[\mathbf{x}]$, when the matrix *BH* performs a non-mixing operation of **s**. The development of the contrast function $\boldsymbol{\Phi}[\mathbf{x}]$ with the properties described by (4-15) is an important step in blind source separation. The following sections will examine several possible contrast functions that when minimized will result in an estimate of the channel matrix *H* and the transmitted sequence **s**. Each of the contrast functions require different *a priori* information or pre-processing and are based on the distributions or higher order statistics of **x** and/or **s**.

4.4.2 Maximum Likelihood Contrast function $\Phi_{ML}[\bullet]$

This section will examine the maximum likelihood approach to the development of the contrast function $\boldsymbol{\Phi}[\mathbf{x}]$ discussed in the previous section. Under the blind source separation problem, only the vector **y** (4-1) is available at the receiver, and it is this information that is solely used to extract the estimate of the channel $\hat{H}$ and the estimate of the transmitted signal $\hat{\mathbf{s}}$. *H* is an unknown constant matrix and **s** has unknown distribution $f_{\mathbf{S}}(\mathbf{s})$. Since neither the distribution of **s** nor the mixing matrix *H* is known a priori to the receiver, an estimate of the distribution of **y** ($\hat{\mathbf{y}}$) is made by assuming different combinations of *H* and $f_{\mathbf{S}}(\mathbf{s})$ and then finding the resultant synthesized distribution of $\hat{\mathbf{y}}$ that best matches the true measured distribution of **y**. Assuming a distribution of **s**, $f_{\hat{\mathbf{S}}}(\mathbf{s})$, and a mixing matrix $\hat{H}$ (full rank matrix assumed), the resultant distribution of $\hat{\mathbf{y}}$ becomes

$$f_{\hat{\mathbf{Y}}}(\mathbf{y}) = \left|\det \hat{H}^{-1}\right| f_{\hat{\mathbf{S}}}(\hat{H}^{-1}\mathbf{y}) \tag{4-16}$$

which can be developed from (4-1) and the distribution can be found by applying techniques given in [3]. The resultant $f_{\hat{\mathbf{Y}}}(\mathbf{y})$ that most closely matches the measured probability density function of $f_{\mathbf{Y}}(\mathbf{y})$ gives rise to the optimum estimate of H and $f_{\mathbf{S}}(\mathbf{s})$. This common sense approach to derivation to the ML contrast function can also be shown in a more mathematically rigorous fashion. Probabilistically, we would like to set $\hat{H}$ and $f_{\hat{\mathbf{S}}}(\mathbf{s})$ to the values of H, $f_{\mathbf{S}}(\mathbf{s})$ that maximizes the synthesized probability density of $\hat{\mathbf{y}}$ given $\mathbf{y}$ found from the received samples. This corresponds to the maximum a priori probability (MAP)

$$\begin{aligned}\hat{H},\, f_{\hat{\mathbf{S}}}(\mathbf{s}) &= \arg_{H,\, f_{\mathbf{S}}(\mathbf{s})} \max\, f_{\hat{\mathbf{y}}|\mathbf{y}}\left(\hat{\mathbf{y}}\{H,\, f_{\mathbf{S}}(\mathbf{s})\} \mid \mathbf{y}\right) \\ \hat{H},\, f_{\hat{\mathbf{S}}}(\mathbf{s}) &= \arg_{H,\, f_{\mathbf{S}}(\mathbf{s})} \max\left\{\frac{f_{\mathbf{y}|\hat{\mathbf{y}}}\left(\mathbf{y} \mid \hat{\mathbf{y}}\{H,\, f_{\mathbf{S}}(\mathbf{s})\}\right) f_{\hat{\mathbf{y}}}\left(\hat{\mathbf{y}}\{H,\, f_{\mathbf{S}}(\mathbf{s})\}\right)}{f_{\mathbf{y}}(\mathbf{y})}\right\}\end{aligned} \tag{4-17}$$

Under the maximum likelihood assumption, the distribution of $\hat{\mathbf{y}}$ is equally likely or $f_{\hat{\mathbf{y}}}\left(\hat{\mathbf{y}}\{H,\, f_{\mathbf{S}}(\mathbf{s})\}\right) = cst$. Taking in account the ML conditions, (4-17) simplifies to

$$\hat{H},\, f_{\hat{\mathbf{S}}}(\mathbf{s}) = \arg_{H,\, f_{\mathbf{S}}(\mathbf{s})} \max\, f_{\mathbf{y}|\hat{\mathbf{y}}}\left(\mathbf{y} \mid \hat{\mathbf{y}}\{H,\, f_{\mathbf{S}}(\mathbf{s})\}\right) \tag{4-18}$$

The above maximization is equivalent to the log-likelihood function

$$\hat{H},\, f_{\hat{\mathbf{S}}}(\mathbf{s}) = \arg_{H,\, f_{\mathbf{S}}(\mathbf{s})} \max \log\left\{f_{\mathbf{y}|\hat{\mathbf{y}}}\left(\mathbf{y} \mid \hat{\mathbf{y}}\{H,\, f_{\mathbf{S}}(\mathbf{s})\}\right)\right\} \tag{4-19a}$$

$$\hat{H},\, f_{\hat{\mathbf{S}}}(\mathbf{s}) = \arg_{H,\, f_{\mathbf{S}}(\mathbf{s})} \min\left(-\log\left\{f_{\mathbf{y}|\hat{\mathbf{y}}}\left(\mathbf{y} \mid \hat{\mathbf{y}}\{H,\, f_{\mathbf{S}}(\mathbf{s})\}\right)\right\}\right) \tag{4-19b}$$

Since the log function is a monotonically increasing function, the maximization of (4-19) is equivalent to that of (4-18). Additionally, a minimization (instead of maximization) can be performed by introducing a negative sign to the first equation in (4-19) to be consistent with the definition of the contrast function $\mathbf{\Phi}$. The results in (4-19) can be summarized as follows: the ML decision picks the combination of H and $f_{\mathbf{S}}(\mathbf{s})$ that according to (4-16) best matches the distribution of $\mathbf{y}$. The minimum of (4-19b) will

occur when $f_{\mathbf{y}|\hat{\mathbf{y}}}(\mathbf{y}\,|\,\hat{\mathbf{y}}) = 1$ for all values of **y** which corresponds with the event that the assumptions of *H* and distribution of **s** are correct (up to the ambiguity of the model)since they perfectly predict (with probability 1) the event **y**.

In practice, however, the decisions estimates of *H* and **s** arise from a series of independent time samples of $\mathbf{y}=\mathbf{y}_t$ which belong to the set $\{\mathbf{y}_1,\mathbf{y}_2,\mathbf{y}_t, \ldots ,\mathbf{y}_T\}$ such that $f_{\mathbf{y}}(\mathbf{y}_{1:T})=f_{\mathbf{y}}(\mathbf{y}_1)f_{\mathbf{y}}(\mathbf{y}_2)\ldots f_{\mathbf{y}}(\mathbf{y}_t)\ldots f_{\mathbf{y}}(\mathbf{y}_T)$. From a time series perspective we wish to minimize (a variation found in [1])

$$\phi_{ML}[\mathbf{y},\mathbf{s},H] = -\frac{1}{T}\log f_{\mathbf{y}|\hat{\mathbf{y}}}\left(\mathbf{y}_{1:T}\,|\,\hat{\mathbf{y}}\{H,\, f_{\mathbf{S}}(\mathbf{s})\}\right) = -\frac{1}{T}\sum_{t=1}^{T}\log f_{\mathbf{S}}(H^{-1}\mathbf{y}_t) + \log\left|\det H\right| \quad (4\text{-}20)$$

for the entire set of the received signals $\mathbf{y}_t$ on the average based on the assumptions of the distribution of the transmitted sequences **s**, $f_{\mathbf{S}}(\mathbf{s})$, as well as the unknown channel matrix *H*. The right side of the equation in (4-20) is nothing but the log of the assumed probability density function $f_{\hat{\mathbf{Y}}}(\mathbf{y})$ with the assumption of *H* and $f_{\mathbf{S}}(\mathbf{s})$ in (4-16) evaluated at $\mathbf{y}_t$. Equation (4-20) is, therefore, a time series representation of the minimization argument given in (4-19b). It can be verified that when the assumed distribution of **y**, $\hat{\mathbf{y}}$, matches the distribution of the measured **y** the term on the right side of (4-20) is a minimum. If enough time samples of the received signal **y** are available, a time average will eventually approach the true ensemble average. The weak law of large numbers states that for independent identically distributed signals as the sample size *T* approaches infinity [2], the maximum likelihood contrast function $\mathbf{\Phi}[\mathbf{y}]$ becomes

$$\mathbf{\Phi}_{ML}[\mathbf{y};\mathbf{s},H] = \lim_{T\to\infty}\phi_{ML} = -E_{\mathbf{y}}[\log f_{\mathbf{S}}(H^{-1}\mathbf{y}))] + \log\left|\det H\right| \quad (4\text{-}20)$$

The maximum likelihood contrast function $\mathbf{\Phi}_{ML}$ can be written in terms of integral functions over probability density function replacing the expectation operation operations as follows

$$\mathbf{\Phi}_{ML}[\mathbf{y};\mathbf{s},H] = -E_{\mathbf{y}}[\log f_{\mathbf{S}}(H^{-1}\mathbf{y}) - \log|\det H|] = \int f_{\mathbf{Y}}(\mathbf{y}) \log\left(\frac{1}{\left|\det H^{-1}\right| f_{\mathbf{S}}(H^{-1}\mathbf{y})}\right) d\mathbf{y}$$

$$\mathbf{\Phi}_{ML}[\mathbf{y};\mathbf{s},H] = \int f_{\mathbf{Y}}(\mathbf{y}) \log\left(\frac{f_{\mathbf{Y}}(\mathbf{y})}{\left|\det H^{-1}\right| f_{\mathbf{S}}(H^{-1}\mathbf{y})}\right) d\mathbf{y} - \int f_{\mathbf{Y}}(\mathbf{y}) \log f_{\mathbf{Y}}(\mathbf{y}) d\mathbf{y} \tag{4-21}$$

$$\mathbf{\Phi}_{ML}[\mathbf{y};\mathbf{s},H] = D_{f_{\mathbf{Y}}(\mathbf{y})\left\|\det H^{-1}\right| f_{\mathbf{S}}(H^{-1}\mathbf{y})} + H_{f_{\mathbf{Y}}(\mathbf{y})}$$

The first term of (4-21) can be recognized as the *Kullback-Lieber* distance while the second term is the differential entropy of the variable **y**. Noting that the differential entropy of **y**, $H_{f_{\mathbf{Y}}(\mathbf{y})}$, is a constant independent of H or **s** it does not affect the minimization process and therefore the contrast function depends purely on the Kullback-Lieber distance

$$\mathbf{\Phi}_{ML}[\mathbf{y};\mathbf{s},H] = D_{f_{\mathbf{Y}}(\mathbf{y})\left\|\det H^{-1}\right| f_{\mathbf{S}}(H^{-1}\mathbf{y})} = \int f_{\mathbf{Y}}(\mathbf{y}) \log\left(\frac{f_{\mathbf{Y}}(\mathbf{y})}{\left|\det H^{-1}\right| f_{\mathbf{S}}(H^{-1}\mathbf{y})}\right) d\mathbf{y} \tag{4-22}$$

$$\mathbf{\Phi}_{ML}[\mathbf{y};\hat{\mathbf{y}}\{\mathbf{s},H\}] = D_{f_{\mathbf{Y}}(\mathbf{y})\| f_{\hat{\mathbf{Y}}}(\hat{\mathbf{y}})}$$

The equation for the maximum likelihood contrast function shown above results in a intuitive answer. The *Kullback-Lieber* (KL) distance is a measure of how much two probability density function differ from each other and has the property of being greater than or equal to zero [2]. The minimum of the KL distance occurs when the numerator of the argument in the log function is equal to its denominator resulting in $\mathbf{\Phi}_{ML}$ =0 in (4-22). In our case the minimum occurs when the probability density function of the measured received signal **y**, $f_{\mathbf{Y}}(\mathbf{y})$, is equal to the probability density function of the synthesized received signal when the distribution of **s**, $f_{\mathbf{S}}(\mathbf{s})$, and H (4-16) are guessed correctly. Therefore, the true value of H and the true distribution of **s** will result in the ML contrast function in (4-22) reaching its minimum or equivalently

$$H, f_{\mathbf{S}}(\mathbf{s}) = \arg_{H, f_{\mathbf{S}}(\mathbf{s})} \min \mathbf{\Phi}_{ML}[\mathbf{y};\mathbf{s},H] \tag{4-23}$$

A search over all possibilities of H and $f_{\mathbf{S}}(\mathbf{s})$ is not viable because of the infinite combinations that exist for both variables. A contrast function that is less stringent and searches over lesser dimensions would be preferable. In the following subsection the mutual information contrast function will be examined which instead of requiring the synthesized and the measured probabilities to match perfectly, it forces independence on the components of the vector **x** in (4-14). The mutual information contrast function is a suboptimal approach but was shown to produce very good results.

4.4.3 Mutual Information Contrast Function $\Phi_{MI}[\bullet]$

The derivation of the mutual information contrast function starts with the maximal ratio contrast function given in (4-22). Instead of analyzing the contrast function of the variable **y**, however, the $\mathbf{\Phi}_{ML}$ is applied to another related variable **x** defined in (4-14) and repeated again here

$$\mathbf{x} = B\mathbf{y} = BH\mathbf{s} \tag{4-24}$$

We would like **x** to be equal to **s** and in the process find $B=H^{-1}$. The distribution of **x** should therefore match the distribution of **s** if the right B (H^{-1}) is chosen in (4-24). The matrix B is an un-mixing operator and the minimum of the contrast function will be reached when **x** has independent components as described by (4-15). This statement is mathematically equivalent to the maximal ratio contrast function applied to the variable **x** as follows

$$\mathbf{\Phi}_{ML}[\mathbf{x};\mathbf{s}] = D_{f_{\mathbf{X}}(\mathbf{x})\|f_{\mathbf{S}}(\mathbf{s})} = \int f_{\mathbf{X}}(\mathbf{x}) \log\left(\frac{f_{\mathbf{X}}(\mathbf{x})}{f_{\mathbf{S}}(\mathbf{x})}\right) d\mathbf{x} \tag{4-25}$$

which is a comparison between the distributions of **s** and that of **x**. In terms of the measured distribution $f_{\mathbf{Y}}(\mathbf{y})$, the above equation is equivalent to

$$\mathbf{\Phi}_{ML}[\mathbf{x};\mathbf{y},\mathbf{s},B] = D_{\left|\det B^{-1}\right| f_{\mathbf{Y}}(B^{-1}\mathbf{x}) \| f_{\mathbf{S}}(\mathbf{s})} = \int \left|\det B^{-1}\right| f_{\mathbf{Y}}(B^{-1}\mathbf{x}) \log\left(\frac{\left|\det B^{-1}\right| f_{\mathbf{Y}}(B^{-1}\mathbf{x})}{f_{\mathbf{S}}(\mathbf{x})}\right) d\mathbf{s} \quad (4\text{-}26)$$

The maximal likelihood contrast function in (4-25) as explained earlier, requires a search over a virtually infinite combinations of the distribution of $f_{\mathbf{S}}(\mathbf{s})$ and B that would lead to the global minimum of $\mathbf{\Phi}_{ML}[\mathbf{x}]$. A contrast function that does not require such a stringent requirement can be derived based on a minimization of the mutual information of $\mathbf{x}$ which is equivalent to a independence requirement between the elements of the vector $\mathbf{x}$. To get to the mutual information contrast function, the $\mathbf{\Phi}_{ML}[\mathbf{x}]$ in (4-25) can be re-written as follows:

$$\mathbf{\Phi}_{ML}[\mathbf{x};\mathbf{s}] = D_{f_{\mathbf{X}}(\mathbf{x}) \| f_{\mathbf{S}}(\mathbf{s})} = \int f_{\mathbf{X}}(\mathbf{x}) \log\left(\frac{f_{\mathbf{X}}(\mathbf{x}) f_{\tilde{\mathbf{X}}}(\mathbf{x})}{f_{\tilde{\mathbf{X}}}(\mathbf{x}) f_{\mathbf{S}}(\mathbf{x})}\right) d\mathbf{x} \quad (4\text{-}27)$$

The distribution of $\tilde{\mathbf{x}}$, $f_{\tilde{\mathbf{X}}}(\tilde{\mathbf{x}})$, corresponds to a random vector with independent components with each element distributed as its corresponding entry in the vector $\mathbf{x}$. The distribution of $\tilde{\mathbf{x}}$ therefore corresponds to the multiplication of the marginal distribution of the elements of $\mathbf{x}$. Expanding (4-27) further leads to

$$\begin{aligned} \mathbf{\Phi}_{ML}[\mathbf{x};\mathbf{s}] &= D_{f_{\mathbf{X}}(\mathbf{x}) \| f_{\mathbf{S}}(\mathbf{s})} = \int f_{\mathbf{X}}(\mathbf{x}) \log\left(\frac{f_{\mathbf{X}}(\mathbf{x})}{f_{\tilde{\mathbf{X}}}(\mathbf{x})}\right) d\mathbf{x} + \int f_{\mathbf{X}}(\mathbf{x}) \log\left(\frac{f_{\tilde{\mathbf{X}}}(\mathbf{x})}{f_{\mathbf{S}}(\mathbf{x})}\right) d\mathbf{x} \\ \mathbf{\Phi}_{ML}[\mathbf{x};\mathbf{s}] &= \int f_{\mathbf{X}}(\mathbf{x}) \log\left(\frac{f_{\mathbf{X}}(\mathbf{x})}{f_{\tilde{\mathbf{X}}}(\mathbf{x})}\right) d\mathbf{x} + \int f_{\tilde{\mathbf{X}}}(\mathbf{x}) \log\left(\frac{f_{\tilde{\mathbf{X}}}(\mathbf{x})}{f_{\mathbf{S}}(\mathbf{x})}\right) d\mathbf{x} \end{aligned} \quad (4\text{-}28)$$

The second equation in (4-28) is a result from the properties of the independence of $\tilde{\mathbf{x}}$ and of $\mathbf{s}$. In terms of *Kullback-Lieber* distances the maximum likelihood contrast function simplifies to

$$\mathbf{\Phi}_{ML}[\mathbf{x};\mathbf{s}] = D_{f_{\mathbf{X}}(\mathbf{x}) \| f_{\mathbf{S}}(\mathbf{s})} = D_{f_{\mathbf{X}}(\mathbf{x}) \| f_{\tilde{\mathbf{X}}}(\mathbf{s})} + D_{f_{\tilde{\mathbf{X}}}(\mathbf{x}) \| f_{\mathbf{S}}(\mathbf{s})} \quad (4\text{-}29)$$

The first KL term of $\mathbf{\Phi}_{ML}[\mathbf{x}]$, can be recognized as the mutual information of the variables $\mathbf{x}$. The second term is the KL distance between the marginal distributions of $\mathbf{x}$, $\tilde{\mathbf{x}}$, and the source distribution $\mathbf{s}$. Since both $\mathbf{s}$ and $\tilde{\mathbf{x}}$ are independent, the ML contrast function can be expressed in terms of the mutual information contrast function as follows

$$\mathbf{\Phi}_{ML}[\mathbf{x};\mathbf{s}] = \mathbf{\Phi}_{MI}[\mathbf{x}] + \sum_{m=1}^{M} D_{f_{(\tilde{\mathbf{x}})_m}((\mathbf{x})_m)\| f_{(\mathbf{s})_m}((\mathbf{s})_m)} \tag{4-30}$$

$$\mathbf{\Phi}_{MI}[\mathbf{x}] = D_{f_{\mathbf{X}}(\mathbf{x})\| f_{\hat{\mathbf{X}}}(\hat{\mathbf{x}})} = \int f_{\mathbf{X}}(\mathbf{x}) \log\left(\frac{f_{\mathbf{X}}(\mathbf{x})}{f_{\hat{\mathbf{X}}}(\mathbf{x})}\right) d\mathbf{x} \tag{4-31}$$

where $\mathbf{\Phi}_{MI}[\mathbf{x}]$ is the mutual information of the variables $\mathbf{x}$. As the name states, mutual information is a measure of how much the elements of $\mathbf{x}$ are related. Being a KL distance, the MI measure is always greater than or equal to zero. The $\mathbf{\Phi}_{MI}[\mathbf{x}]$ is zero only if there is no mutual information between the components which occurs when the elements of $\mathbf{x}$ are independent of each other. As stated in [1], the maximum likelihood contrast function can be understood as a mismatch function composed of two components as shown in (4-30) which can be interpreted as

$$\begin{pmatrix} Total \\ mismatch \end{pmatrix} = \begin{pmatrix} Deviation\ \ from \\ independence \end{pmatrix} + \begin{pmatrix} Marginal \\ mismatch \end{pmatrix}$$

Both terms that compose the $\mathbf{\Phi}_{ML}[\mathbf{x}]$ above can be minimized independently and when both equal their minimum, the maximum likelihood function will be at its minimum. The advantage of using $\mathbf{\Phi}_{MI}[\mathbf{x}]$ instead of $\mathbf{\Phi}_{ML}[\mathbf{x}]$ is that a search over all possible distributions of the source vector $f_{\mathbf{S}}(\mathbf{s})$ is no longer required. The minimization of $\mathbf{\Phi}_{MI}[\mathbf{x}]$ is a search over the matrix B which produces independent components (mutual information reduction) in the vector $\mathbf{x}$ in (4-24) which is much more tractable than $\mathbf{\Phi}_{ML}[\mathbf{x}]$.

4.4.4 Contrast Functions and Higher Order Statistics

As explained in Chapter 3, the probability density functions of joint random variables can be expressed as a function of their joint cumulants. Equivalently, from the knowledge of all joint cumulants or moments the distribution of a signal is known with certainty. The maximum likelihood contrast function, $\mathbf{\Phi}_{ML}[\mathbf{x}]$, given in (4-27) compares the distribution of the assumed transmitted signal **s** to a signal **x** that is a result of received signal **y** applied to the mixing B (4-14). If the correct channel is estimated (and is invertible), the signal **x** should equal the original source signals **s** as mentioned in the discussion following (4-14). When **x**=**s**, the distributions of both signals should equal and so should their respective joint cumulants. The maximum likelihood contrast function can therefore be solved by comparing cumulants of **x** and **s** assuming the cumulants of **s** are known. An exact match between cumulants will occur when B is equal the inverse of the channel H. In [1], an approximation to the $\mathbf{\Phi}_{ML}[\mathbf{x}]$ contrast function is give in terms of cumulants of **x** and **s** of orders two and four as

$$\mathbf{\Phi}_{ML}[\mathbf{x} \mid \mathbf{s}] \approx \frac{1}{48}\left(12\phi_2[\mathbf{x}] + \phi_4[\mathbf{x}]\right) \tag{4-32}$$

The above equation was derived using the Edgeworth expansion (see Chapter 3) and is valid for **s** and **x** close to Gaussian. **x** and **s** are also assumed to be zero mean random variables with an even distribution so that all odd cumulants are zeros. The functions $\phi_2[\mathbf{x}]$ and $\phi_4[\mathbf{x}]$ are defined as

$$\begin{aligned}
\phi_2[\mathbf{x}] &= \sum_{i,j=1}^{\mathrm{M}} \left(\left(\boldsymbol{C}_2[\mathbf{x}]\right)_{ij} - \left(\boldsymbol{C}_2[\mathbf{s}]\right)_{ij} \right)^2 = \sum_{i,j=1}^{\mathrm{M}} \left(\left(\boldsymbol{C}_2[\mathbf{x}]\right)_{ij} - C\left[(\mathbf{s})_i^{\,2}\right]\delta_{ij} \right)^2 \\
\phi_4[\mathbf{x}] &= \sum_{i,j,k,l=1}^{\mathrm{M}} \left(\left(\boldsymbol{C}_4[\mathbf{x}]\right)_{ijkl} - \left(\boldsymbol{C}_4[\mathbf{s}]\right)_{ijkl} \right)^2 = \sum_{i,j,k,l=1}^{\mathrm{M}} \left(\left(\boldsymbol{C}_4[\mathbf{x}]\right)_{ijkl} - C\left[(\mathbf{s})_i^{\,4}\right]\delta_{ijkl} \right)^2
\end{aligned} \tag{4-33}$$

which is a comparison of all the fourth and second order cumulants of **s** and **x**. The cumulant tensor $\boldsymbol{C}_{\mathrm{P}}[\mathbf{x}]$ and the simple cumulant $C[(\mathbf{s})_i^{\mathrm{P}}]$ are defined in Chapter 3. The minimum of $\boldsymbol{\Phi}_{ML}[\mathbf{x}]$ will occur when all cumulants match and both distributions are equal. As stated in Chapter 3, cumulants follow the multilinear property and since **s** and **x** are related by (4-14)

$$\mathbf{x} = B\mathbf{y} = BH\mathbf{s} \tag{4-34}$$

the fourth order cumulant tensor of **x** is related to the fourth order cumulant tensor of **s** through

$$\left(\boldsymbol{C}_4[\mathbf{x}]\right)_{ijkl} = \sum_{p,q,r,s=1}^{\mathrm{M}} \left(\boldsymbol{C}_4[\mathbf{s}]\right)_{pqrs} \left(BH\right)_{ip} \left(BH\right)_{jq} \left(BH\right)_{kr} \left(BH\right)_{ls} \tag{4-35}$$

as explained in Section 3.2.4.2. From (4-35) above, both fourth order cumulant tensors will match when the mixing matrix $B=H^{-1}$.

4.4.5 Orthogonal Contrast Functions

A search over all values of *B* that would result in the minimization of the contrast function can be very computationally intensive. A 2x2 real matrix *B*, for example has four components (eight components when complex) that need to be searched over to find a minimum of a cumulant based maximum likelihood contrast function like (4-32). A simplification can be made by pre-whitening the received data first as described in Section 4.3. This is equivalent to performing principal component analysis on the vector and leaves the ambiguity of the estimate $\tilde{H}$ to the true value of *H* to a search over an orthonormal matrix *Q*. The matrix *H* can be decomposed into the multiplication

$$H = \tilde{H}Q \tag{4-36}$$

where $\tilde{H}$ was computed during the whitening stage (Section 4.3). A search over an orthonormal matrix is much simpler than a full search over *B*. A 2x2 matrix *Q*, for example is completely characterized by one unknown (two unknowns when complex). This is significantly less complex than a search over *B* but it does require pre-processing of the received data so that the cross-correlation matrix is an identity matrix or equivalently

$$\begin{gathered} \mathbf{w} = W\mathbf{y},\ W = \tilde{H}^{-1} \\ \boldsymbol{C}_2[\mathbf{w}] = R_{\mathbf{w}} = E\left[\mathbf{w}\mathbf{w}^*\right] = E\left[W\mathbf{y}\mathbf{y}^*W^*\right] = I \\ \mathbf{x} = Q^*\mathbf{w} \end{gathered} \tag{4-37}$$

Section 4.3 describes in more detail how the received data **y** can be whitened. Additionally, orthonormal linear transformations do not affect the norm (Forbenius norm) of the cumulant tensors just like the norm of a matrix is invariant to orthonormal rotations. This means that the fourth order contrast function over *Q* can be written as a minimization of

$$\phi_4^Q[\mathbf{x} \mid \mathbf{s}] = \sum_{i,j,k,l=1}^{\mathrm{M}} \left(\left(\boldsymbol{C}_4[\mathbf{x}]\right)_{ijkl} - C\left[(\mathbf{s})_i^{\,4}\right]\delta_{ijkl} \right)^2 = -2\sum_{i=1}^{\mathrm{M}} C\left[(\mathbf{s})_i^{\,4}\right]\left(\boldsymbol{C}_4[\mathbf{x}]\right)_{iiii}^2 \tag{4-38}$$

since $\sum_{i,j,k,l=1}^{\mathrm{M}} \left(\boldsymbol{C}_4[\mathbf{x}]\right)_{ijkl}^2$ is now a constant over *Q* (which was not the case over *B* in (4-33)).

The above orthogonal contrast function is an approximation to the $\boldsymbol{\Phi}_{ML}[\mathbf{x}]$ since it also is a comparison of the higher order statistics of the source to the extracted distributions. In true blind source separation problems, however, the source statistics are not known *a priori*. This requires the development of an alternate contrast function based on mutual information which does not require the knowledge of the source distribution but instead is based on removing the inter-dependencies of the components of the pre-whitened vector **x**.

The vector **x** will have independent components when all cross-cumulants of all orders P of the elements of **x** are zero (Chapter 3). Since the search over a contrast function of a whitened vector is limited to an orthonormal matrix *Q*, minimizing the off-diagonal terms (to reduce independence between the components) of the cumulant tensors $\mathcal{C}_P[\mathbf{x}]$ is equivalent to maximizing the diagonal terms or written in terms of a minimization

$$\phi_{ICA}^{Q}[\mathbf{x}] = -\sum_{i=1}^{N}\left(\mathcal{C}_4[\mathbf{x}]\right)_{iiii}^2 \tag{4-39}$$

This follows directly from the invariance of the Forbenius norm of the tensor $\mathcal{C}_P[\mathbf{x}]$ to a matrix rotation *Q*. The above equation is the basis for independent component analysis (ICA). The second order cumulants of **x** follow the identity matrix (as a result of pre-whitening) and the third order tensor cumulants do not exist for evenly density functions which justifies the use of the fourth order cumulant tensor for the basis of ICA. The contrast function in (4-39) does not require any *a priori* knowledge of the distributions of the source signal **s** except that they are all statistically independent of each other. The fact that ICA requires no specific knowledge of **s** makes it a very powerful blind source separation technique. Also, the ICA contrast function as well as all other orthogonal contrast functions presented in the section work with a whitened version of the received vector **y** which reduces the search to an orthonormal matrix *Q*. The next section will explore a technique for solving for the matrix *Q*. It is based on diagonalizing the fourth order cumulant tensor $\mathcal{C}_4[\mathbf{x}]$. It should be noted that when the tensor $\mathcal{C}_4[\mathbf{x}]$ is diagonal, the ICA contrast function $\phi_{ICA}^{Q}[\mathbf{x}]$ in (4-39) is a minimum.

4.4.6 ICA Through Tensor Diagonalization

The minimization of the ICA contrast function introduced in the previous section is equivalent to diagonalizing the fourth order tensor of **x**. The second order cumulant tensor, $\mathcal{C}_2[\mathbf{x}]$, is a matrix which can be decomposed into a diagonal matrix and an orthogonal matrix containing the eigenvectors. Unfortunately, a generalization of the

eigenvalue decomposition is not available for tensors of order higher than two as far as the author knows, into a matrix and a diagonal tensor. This is because, in general, an arbitrary tensor cannot be diagonalized the way an NxN matrix can [5]. The problem of tensor diagonalization is equivalent to the solution of the fourth order cumulant equation given by

$$\left(\boldsymbol{C}_4[\mathbf{x}]\right)_{ijkl} = \sum_{p,q,r,s=1}^{\mathrm{M}} \left(\boldsymbol{C}_4[\mathbf{w}]\right)_{pqrs} Q_{ip} Q_{jq} Q_{kr} Q_{ls} \tag{4-40}$$

for *Q*. The above equation follows from the multilinear property of cumulants where **w** and **x** are related by (4-37). Ideally, we would want all cross-cumulants of **x** to be zero or in practical applications as small as possible. It should also be noted that because of the pre-whitening operation applied to the received vector **y**, the second order cumulants of **w** and **x** is the identity matrix (*I*) or equivalently

$$\boldsymbol{C}_2[\mathbf{x}] = \boldsymbol{C}_2[\mathbf{w}] = R_w = I \tag{4-41}$$

This means that the average power of the extracted elements of **x** will equal unity.

Comon [4] proposed that (4-40) could be solved by a series of 2x2 Givens rotations performed on each pair of the components of $\boldsymbol{C}_4[\mathbf{w}]$ until the resulting tensor $\boldsymbol{C}_4[\mathbf{x}]$ is diagonal. Comon noticed that the derivative of the contrast function in (4-39) is a function of only two distinct pairs of the elements of the cumulant tensor at a time which motivated this approach. The pairwise rotations are then collected for each step and combined to form the matrix *Q* that is being sought.

A complex pairwise Givens rotation ($J(a,b)$) can be written for the pair *a* and *b* as

$$J^{ab} = c\begin{bmatrix} 1 & \theta \\ -\theta^* & 1 \end{bmatrix} = \begin{bmatrix} \left(J(a,b)\right)_{aa} & \left(J(a,b)\right)_{ab} \\ \left(J(a,b)\right)_{ba} & \left(J(a,b)\right)_{bb} \end{bmatrix},\ c = \frac{1}{\sqrt{1+\theta\theta^*}} \tag{4-42}$$

where the remainder of the components of $J(a,b)$ other than the rows and columns a or b follow the identity matrix. The constant θ is a complex number that is evaluated through this algorithm such that the contrast function $\phi_{ICA}^{Q}[\mathbf{x}]$ is minimized for each pair combination. The equation that relates $\phi_{ICA}^{Q}[\mathbf{x}]$ to the fourth order cumulant tensors of $\mathbf{w}$ and the Givens rotation (4-42) for a given pair a and b follows from (4-40)

$$\phi_{ICA}^{Q}[\mathbf{x};J^{ab}]=\sum_{m=1}^{2}\sum_{i,j,k,l=1}^{2}\left(\boldsymbol{C}_4{}^{ab}[\mathbf{w}]\right)_{ijkl}\left(J^{ab}\right)_{mi}\left(\left(J^{ab}\right)_{mj}\right)^{*}\left(\left(J^{ab}\right)_{mk}\right)^{*}\left(J^{ab}\right)_{ml} \tag{4-43}$$

The tensor $\boldsymbol{C}_4{}^{ab}[\mathbf{x}]$ is of size 2x2x2x2 containing the fourth order cumulants of the pair a and b. In order to simplify the process of solving for the ICA contrast function, two auxiliary variables ξ and b_{ij} are introduced. For a single pair combination, the ICA contrast function in (4-43) can be expanded out and re-written as [4]

$$\phi_{ICA}^{Q}[\mathbf{x};J^{ab}(\xi)]=-\sum_{\substack{i=-2\\0\le i+j\le 4}}^{2}\sum_{j=-2}^{2}b_{ij}\frac{\xi^{i}\xi^{*j}}{\left(\xi\xi^{*}+4\right)^{2}} \tag{4-44}$$

where the matrix J^{ab} shows up in the contrast function in the form of the value ξ which in turn is a function of θ. ξ and θ are related by

$$\xi=\theta-\frac{1}{\theta^{*}} \tag{4-45}$$

The constants b_{ij} of (4-44) are a function of the fourth order cumulants of the whitened vector $\mathbf{w}$ for the corresponding pair a and b as follows

$$\begin{aligned}
&a_0=C[(\mathbf{w})_a,(\mathbf{w})_a,(\mathbf{w})_a,(\mathbf{w})_a], \quad a_1=4C[(\mathbf{w})_a,(\mathbf{w})_a,(\mathbf{w})_a,(\mathbf{w})_b],\\
&a_2=4C[(\mathbf{w})_a,(\mathbf{w})_a,(\mathbf{w})_b,(\mathbf{w})_b], \quad \hat{a}_2=2C[(\mathbf{w})_b,(\mathbf{w})_a,(\mathbf{w})_a,(\mathbf{w})_b],\\
&a_3=4C[(\mathbf{w})_a,(\mathbf{w})_b,(\mathbf{w})_b,(\mathbf{w})_b], \quad a_4=C[(\mathbf{w})_b,(\mathbf{w})_b,(\mathbf{w})_b,(\mathbf{w})_b]
\end{aligned} \tag{4-46}$$

$$
\begin{aligned}
b_{22} &= a_o a_o^* + a_4 a_4^* \\
b_{21} &= \frac{1}{2}\left(a_3^* a_4^* + a_3^* a_4 - a_1 a_0^* - a_0 a_1\right) \\
b_{12} &= b_{21}^* \\
b_{20} &= \frac{1}{4}\left(a_3^{*2} + a_1^2\right) + \frac{1}{2}\left(\hat{a}_2 a_4 + \hat{a}_2 a_0 + \hat{a}_2 a_0^* + \hat{a}_2 a_4^*\right) \\
b_{02} &= b_{20}^* \\
b_{2-1} &= \frac{1}{2}\left(\hat{a}_2 a_3^* - \hat{a}_2 a_1\right) \\
b_{-12} &= b_{2-1}^* \\
b_{2-2} &= \frac{\hat{a}_2^2}{2} \\
b_{-22} &= b_{2-2}^* \\
b_{11} &= 4\left(a_0 a_0^* + a_4 a_4^*\right) + \left(a_2 a_0^* + a_2 a_4^* + a_0 a_2^* + a_4 a_2^*\right) + \frac{1}{2}\left(a_3 a_3^* + a_1 a_1^*\right) \\
b_{10} &= \frac{1}{2}\left(\hat{a}_2 a_3 - a_1 a_2 + a_1 a_4 - \hat{a}_2 a_1^* - a_0 a_3^* - a_0^* a_3^* + a_1 a_4^* - a_1 a_2^* + a_3^* a_2^* + a_2 a_3^*\right) + \\
&\quad \frac{3}{2}\left(a_3^* a_4^* + a_4 a_3^* - a_1 a_0^* - a_0 a_1\right) \\
b_{01} &= b_{10}^* \\
b_{1-1} &= \hat{a}_2\left(a_2 + a_4 + a_0 + a_0^* + a_4^* + a_2^*\right) + a_1 a_3^* + \frac{1}{2}\left(a_3^{*2} + a_1^2\right) \\
b_{-11} &= b_{1-1}^* \\
b_0 &= 2\left(a_0 a_0^* + a_4 a_4^*\right) + a_1 a_1^* + a_3 a_3^* + \hat{a}_2 \hat{a}_2^* + 2 a_2 a_2^* + a_1 a_3 + 2 a_4 a_0^* + 2 a_0 a_4^* \\
&\quad + a_1^* a_3^* + 2\left(a_2\left(a_0^* + a_4^*\right) + a_2^*\left(a_0 + a_4\right)\right)
\end{aligned}
\tag{47}
$$

The goal of this tensor diagonalization algorithm is to minimize (4-44) for all pairs *a* and *b* by applying a series of Givens rotations to all the pairs until the contrast function (4-39) is at its global minimum. As a side note, the expressions for b_{ij} given by Comon in [4] for complex data **y** are not correct and should actually be equal to the expressions in (4-46) and (4-47). The derivation of the constants b_{ij} is straight forward but very tedious and long and will not be presented here. The results of b_{ij} for real data **y** in [4], however, are correct which can be obtained by removing the conjugates operations from both (4-46) and (4-47).

Formally, the procedure of this algorithm can be explained as follows.

(1) Whiten the received vector **y** to create the initial $\mathbf{x}_0$=**w** as described in Section 4.3 to limit the search of the ICA contrast function over an orthonormal matrix Q. Initialize Q^H=I. Initialize $H = \tilde{H}$.

(2) Begin a loop j which that sweeps over all indexed i N(N-1)/2 pairs a and b

(a) Compute the cumulant tensors $\boldsymbol{C}_4^{ab}[\mathbf{x}_i^j]$

(b) Find the ξ_i that minimizes $\phi_{ICA}^{Q}[\mathbf{x}_i^j; J^{ab}(\xi_i^j)]$, (4-44)

(c) Find the angle θ_i^j by solving (4-45), pick the solution where $|\theta_i^j| \in (-1,1]$. Create the matrix $J_i^j(a,b)$ from θ_i^j.

(d) Accumulate the matrix Q^H and H. $Q^H = J_i^j(a,b)Q^H$, $H_i^j = \tilde{H}Q \leftarrow$ estimated channel up to now. Compute $\mathbf{x}_{i+1} = J_i(a,b)\mathbf{x}_i$

(3) Repeat step (2) until the rotational angle for all pairs θ_i of the jth pass of (2) (θ_i^j and ξ_i^j correspond to the jth pass of the ith pair and not to an exponent j) is small according to a preset tolerance level. When for all pairs i, θ_i is within the tolerance limit, stop. Q is the orthonormal matrix sought and H is the estimate of the channel.

The above algorithm will eventually converge to the correct solution for H subject to the ambiguities discussed in Section 4.2 and the vector $\mathbf{x}_i$ will converge to the transmitted signals **s** but normalized to an average power of one. A short example follows in order to illustrate this concept in a more transparent fashion. The example under test is the following

$$\mathbf{y}(t) = H\mathbf{s}(t), \quad H = \begin{bmatrix} 1 & -10 \\ -7 & 3 \end{bmatrix}, \mathbf{s}(t) = \begin{bmatrix} (\mathbf{s}(t))_1 \\ (\mathbf{s}(t))_2 \end{bmatrix}, \mathbf{y}(t) = \begin{bmatrix} (\mathbf{y}(t))_1 \\ (\mathbf{y}(t))_2 \end{bmatrix} \qquad (4\text{-}48)$$

where H is a 2x2 mixing matrix given above. The transmitted signals $\mathbf{s}(t)$ consist of several samples of a random binary sequence composed of $(\mathbf{s}(t))_1,(\mathbf{s}(t))_2 \in \{-1,1\}$ with symbol duration 5 and 8 samples for $(\mathbf{s}(t))_1$ and $(\mathbf{s}(t))_2$, respectively. The random binary sequences were generated using a Matlab random generator. The random generator did not create perfectly zero mean independent signals as can be seen by examining the correlation of both transmitted signals which was equal to $E[\mathbf{s}(t))_1(\mathbf{s}(t))_2]=0.0004$. From this analysis, the matrix H is expected to be correct to less than four significant figures after the decimal point because of the non-ideal independent generation of the random numbers by Matlab. The first step in the ICA algorithm is to whiten the received vector $\mathbf{y}$ and determine the whitening matrix W and its corresponding vector $\mathbf{w}$. The correlation of the vector $\mathbf{y}$ for this example was found to be

$$R_y = E[\mathbf{y}\mathbf{y}^*] = \boldsymbol{C}_2[\mathbf{y}] = \begin{bmatrix} 100.9917 & -36.9697 \\ -36.9697 & 57.9826 \end{bmatrix} \tag{4-49}$$

The whitening matrix W and the initial estimate of H can be found by using the eigen value decomposition as described in Section 4.3. The initial estimate of H based solely on the second order statistics can be computed to be

$$H_0 = \tilde{H} = W^{-1} = \begin{bmatrix} -3.0213 & -9.5846 \\ -5.2526 & 5.5129 \end{bmatrix}, \quad \mathbf{w} = W\mathbf{y} \tag{4-50}$$

And as desired, the correlation matrix of the vector w after whitening is the identity matrix.

$$R_w = E[\mathbf{w}\mathbf{w}^*] = \boldsymbol{C}_2[\mathbf{w}] = \begin{bmatrix} 1.0000 & -0.0000 \\ -0.0000 & 1.0000 \end{bmatrix} \tag{4-51}$$

As discussed in Section 4.3, the channel matrix H can only be estimated up to an orthonormal transformation Q using second order statistics. Step (2) of the algorithm above explains how to determine the remaining matrix Q using the fourth order

cumulants. The fourth order cumulant tensor can be determined for a zero mean vector $\mathbf{x}_i^j$ from

$$\begin{aligned}\left(\boldsymbol{C}_4[\mathbf{x}_i^j]\right)_{pqrs} = {} & E[\left(\mathbf{x}_i^j\right)_p\left(\mathbf{x}_i^j\right)_q\left(\mathbf{x}_i^j\right)_r\left(\mathbf{x}_i^j\right)_s] - E[\left(\mathbf{x}_i^j\right)_p\left(\mathbf{x}_i^j\right)_q]E[\left(\mathbf{x}_i^j\right)_r\left(\mathbf{x}_i^j\right)_s] - \\ & E[\left(\mathbf{x}_i^j\right)_p\left(\mathbf{x}_i^j\right)_r]E[\left(\mathbf{x}_i^j\right)_q\left(\mathbf{x}_i^j\right)_s] - E[\left(\mathbf{x}_i^j\right)_p\left(\mathbf{x}_i^j\right)_s]E[\left(\mathbf{x}_i^j\right)_q\left(\mathbf{x}_i^j\right)_r] \quad p,q,r,s \in \{1,2\}\end{aligned} \tag{4-52}$$

for a vector of length two and is of size 2x2x2x2. In this example, since *H* is 2x2, there is only one pair of distinct indices *i*={1} with *a*=1 and *b*=2. The first pass, *j*=1, requires the minimization of $\phi_{ICA}^{Q}[\mathbf{x}_1^1; J^{12}(\xi_1^1)]$ described by (4-44). The Givens rotation matrix *J*(*1,2*) for the indices *a=1* and *b=2* for a real 2x2 matrix is given by

$$J(1,2) = J^{12} = \frac{1}{\sqrt{1+\theta^2}}\begin{bmatrix} 1 & \theta \\ -\theta & 1 \end{bmatrix} \tag{4-53}$$

When (4-53) is applied to the ICA contrast function, the minimum is reached for a rotational value θ_1^1=-0.9998 in the first *j*th iteration. Updating the matrix *H* with this new information according to step (2) results in

$$H_1^1 = \begin{bmatrix} 4.6399 & -8.9142 \\ -7.6124 & 0.1850 \end{bmatrix} \tag{4-54}$$

In the second pass, *j*=2, a Givens rotational matrix (4-53) with a θ_1^2=0.3999 minimizes the contrast function $\phi_{ICA}^{Q}[\mathbf{x}_1^2; J^{12}(\xi_1^2)]$. Accumulating J_1^2(1,2) to the channel matrix (step (2)) results in an updated estimate of

$$H_1^2 = \begin{bmatrix} 0.9979 & -9.9998 \\ -6.9994 & 2.9985 \end{bmatrix} \tag{4-55}$$

Finally on the third pass, *j*=3, a Givens rotational matrix with θ_1^3=1.3808e-11 minimizes the contrast function $\phi_{ICA}^{Q}[\mathbf{x}_1^3; J^{12}(\xi_1^3)]$. As can be seen from (4-53), the small value θ_1^3

results in a Givens rotation matrix that is very close to the identity matrix. The identity matrix has no effect when aggregated to H and in this pass, there is virtually no change on the updated matrix H which is given by

$$H_1^3 = \begin{bmatrix} 0.9979 & -9.9998 \\ -6.9994 & 2.9985 \end{bmatrix} \Rightarrow H = \begin{bmatrix} 1 & -10 \\ -7 & 3 \end{bmatrix} \qquad (4\text{-}56)$$

The algorithm can now be stopped, since the contribution of the last step is negligible. The difference between H_1^3 and H is around the expected three significant figures after the decimal point which was the limit imposed by the random generator function used by Matlab.

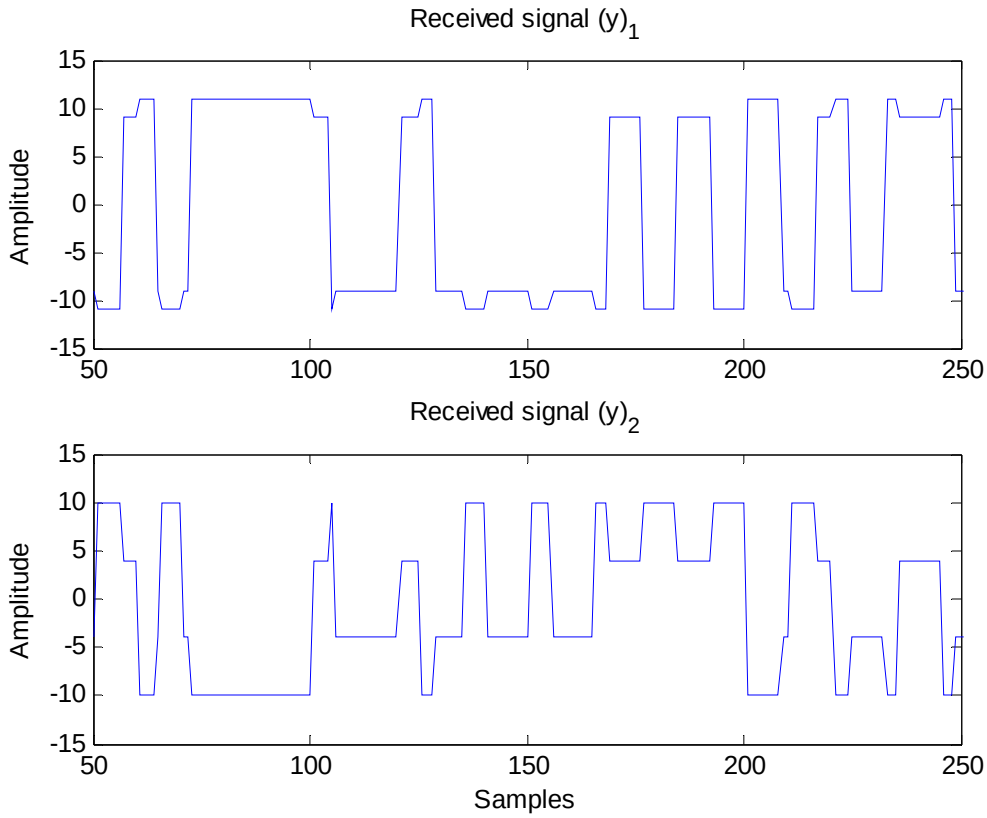

Figure 4.1 Received signal **y** which is related to **s** by **y**=*H***s**.

Figure 4.1 shows a snapshot of the received signal **y**. Even though the original signals are very simple (take only two values), they are not so easy to make out from observed vector **y**. As the transmitted signals **s** increase in number and become more complex in their distribution, distinguishing the signals visually is nearly impossible. Through higher order statistics and through independent component analysis, however, this task proves to be possible.

Figure 4.2, shows the estimated signals **x** that where recovered which match the transmitted signals of **s** perfectly.

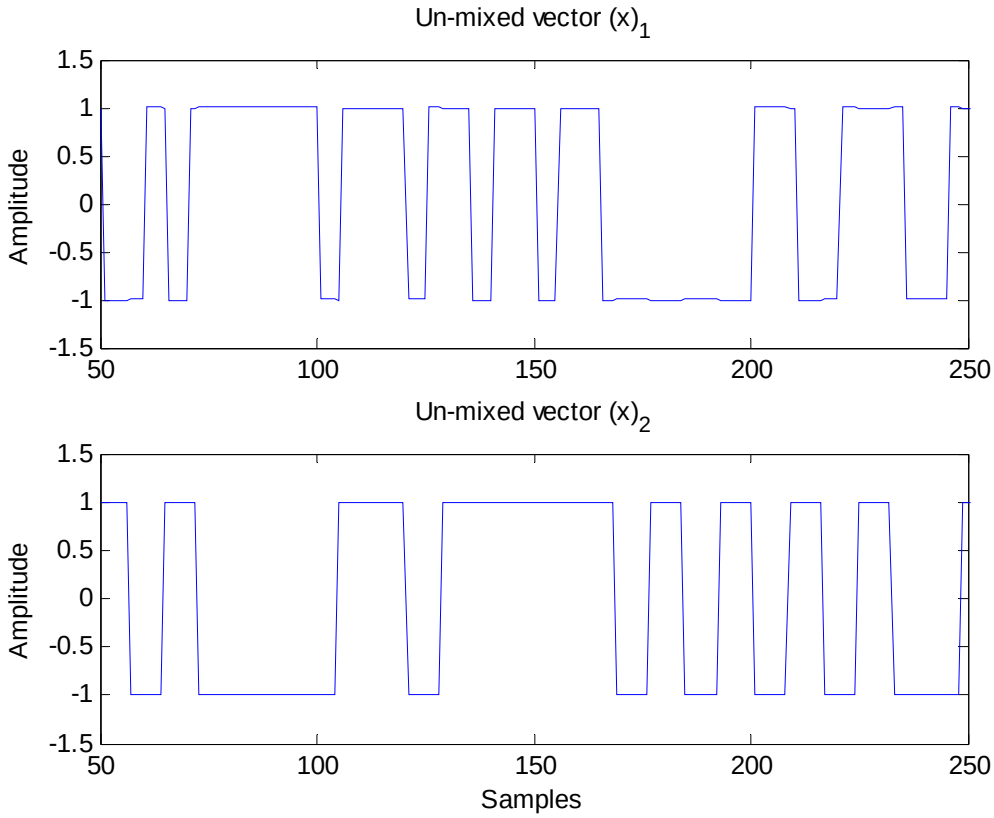

Figure 4.2 Estimates of the transmitted signals **s** described by the vector **x**.

The above example demonstrates that it is possible to extract the channel matrix *H* from just the received signal vector **y** within, of course, the ambiguities of the model. A seemingly underdetermined problem can be readily solved by resorting to higher order

statistics. A series of independent samples of y have to be available at the receiver to make this technique possible. Two equations have been used to find four unknown constants (H) and two unknown signals (($\mathbf{s}$)$_1$,($\mathbf{s}$)$_2$).

4.5 Summary

In this chapter principal component analysis (PCA) and independent component analysis (ICA) were introduced and analyzed. Both are methods for estimating the channel and the transmitted sequences blindly for N≥M. For the linear channel model in (4-1), there are certain irresolvable ambiguities when estimating blindly. First, the amplitude of the signals in $\mathbf{s}$ cannot be resolved with certainty and are therefore arbitrarily assumed to be $E[\mathbf{s}\mathbf{s}^*]=\boldsymbol{C}_2[\mathbf{s}]=I$ (whiteness constraint). Second, the ordering of the components in $\mathbf{s}$ does not change the received vector $\mathbf{y}$ as long as the columns of H are permuted the same way. H is therefore identifiable up to a permutation of its columns and its columns are correct up to a scaling constant that is related to the amplitude constant ambiguity in $\mathbf{s}$.

The difference between PCA and ICA is that the first is based on second order statistics and the approximation of the channel is therefore correct up to an orthonormal matrix Q. ICA on the other hand, goes a bit further and uses the fourth order cumulants to resolve the ambiguity in the matrix Q left unanswered by PCA. ICA can be related to the diagonalization of the fourth order cumulant tensor (subject to the whitening constraint) which is perfectly diagonal when the correct matrix Q is picked. Common [4] came up with a pair wise diagonalization scheme for the fourth order cumulant tensor based on pair wise Givens rotations. As seen from the example, this method works very well in determining the matrix Q as well as diagonalizing the cumulant tensor.

At first glance, the problem of blind signal separations seems impossible because of the large number of unknowns that are being sought relative to the number of equations that are available. The strength in this technique is that it solves the problem by taking advantage of the independent realizations that are available at the receiver through time samples that all are formed from the same equation (4-1) since H never changes. The

solution of the problem can be found statistically with the help of appropriate contrast function.

References

[1] J.-F. Cardoso, "Blind Signal Separation: Statistical Principles," *Proc. of the IEEE*, vol. 86, no. 10, October 1998, pp. 2009-2025.

[2] T. M. Cover and J. A. Thomas, Elements of Information Theory. New York: John Wiley & Sons, 1991.

[3] H. Stark and J. W. Woods, *Probability, Random Processes, and Estimation Theory for Engineers*. New Jersey: Prentice Hall,1994.

[4] P. Common, "Independent Component Analysis, A New Concept?," *Signal Processing*, Special Issue *Higher Order Statistics*, Vol. 36, No. 3, April 1994, pp. 287-314.

[5] L. De Lathauwer, *Signal Processing based on Multilinear Algebra*. PhD. Dissertation Katholieke Universiteit Leuven, September 1997.

Chapter 5

Space-Time MIMO Channels (Dispersive Linear Systems)

The remaining chapters explore the fundamentals of blind signal and channel estimation from the system-theoretic point of view. The limitations of using just second order statistics will be examined and the reasons for resorting to higher order statistics for full channel identification will become clear. This chapter introduces some of the concepts that are needed for this and subsequent chapters. In Chapter 4, the non-dispersive channel was exclusively considered and this chapter treats the time-dispersive channel and its identification in detail. After first presenting some nomenclature that will be used throughout the chapter, analysis of the system identification will be treated. Also, the time-dispersive multiple-input-multiple-output (MIMO) causal linear channel will be explained.

5.1 The Time Dispersive Channel

A linear, causal, and time-dispersive multiple-input-multiple-output (MIMO) channel is examined in this section. These types of channels are often termed convolutive mixtures in literature because they are composed of a series of convolutions that relate the inputs to the outputs of the system. As before, the number of inputs is designated by the letter M and the number of outputs by the letter N. Throughout the chapter, it is assumed that N≥M (i.e., the number of sensors are more or equal to the number of signals in the system) – this condition will be relaxed in subsequent chapters. An additional simplifying assumption is made that the system is noise-less; the noisy system will be addressed later.

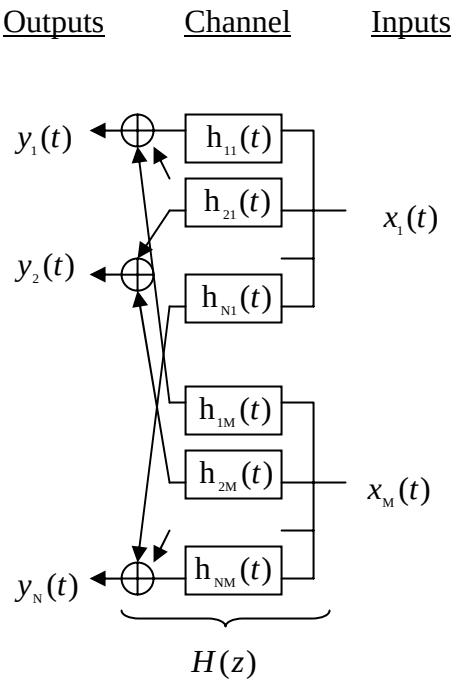

Figure 5.1 Block diagram of MIMO time-dispersive channel.

Figure 5.1 shows a block diagram of a MIMO time-dispersive channel. The contribution of the *j*th source to the *i*th antenna element is modeled by a linear, causal filter with impulse response $h_{ij}(t)$ as seen by the diagram. Mathematically, such a system can be written in matrix convolutional form as

$$\mathbf{y}(t) = \int_{-\infty}^{\infty} H(t-\tau)\mathbf{x}(\tau)d\tau \tag{5-1}$$

where $\mathbf{y}(t)$ is a Nx1 vector, $\mathbf{x}(t)$ is a Mx1 vector, and $H(t)$ is a NxM matrix. The *i*th, *j*th element of $H(t)$, $(H(t))_{ij}$, is the impulse response $h_{ij}(t)$ in Figure 5.1. As is often the case, most processing is currently performed using digital signal processing and at some point the signal $\mathbf{y}(t)$ is sampled by the receiver at sampling rate 1/T before processing. This creates a signal $\mathbf{y}_i$ that exists only for denumerated set of normalized time values *i* which are obtained from the original signal $\mathbf{y}(t)$ through the following process

$$\mathbf{y}_i = \mathbf{y}(t)\big|_{iT} \quad i \in \mathbb{Z} \tag{5-2}$$

where $\mathbb{Z}$ is the set of integers from -∞ to ∞. The sampled equivalent of the system in (5-1) can be written as a summation instead of an integration as

$$\mathbf{y}_i = \sum_{k=-\infty}^{\infty} H_{i-k}\mathbf{x}_k \tag{5-3}$$

where $\mathbf{y}_i \in \mathbb{C}^{Nx1}$, $H_i \in \mathbb{C}^{NxM}$, and $\mathbf{x}_i \in \mathbb{C}^{Mx1}$, where $\mathbb{C}$ represents the field of complex numbers and the superscript NxM correspond to the size of the field. As with Fourier transforms for continuous time signals, some properties of the system in (5-3) can be better appreciated in the z-domain. Also, certain mathematical manipulations of the signal have simpler interpretations in the z-domain than in the time domain. The z-transform adopted in this dissertation for the signals and matrix impulse response is defined as follows:

$$\mathbf{y}(z) = \sum_{i=-\infty}^{\infty} \mathbf{y}_i z^{-i} \tag{5-4}$$

$$\mathbf{x}(z) = \sum_{i=-\infty}^{\infty} \mathbf{x}_i z^{-i}$$

$$H(z) = \sum_{k=-\infty}^{\infty} H_i z^{-i}$$

where the *i*th, *j*th element of $H(z)$, denoted $(H(z))ij$, is the z transform of $h_{ij}(t)$, or $h_{ij}(z)$. The total system response can be written in the z-domain as

$$\mathbf{y}(z) = H(z)\mathbf{x}(z) \tag{5-5}$$

where $\mathbf{y}(z) \in \mathbb{C}[z]^{Nx1}$, $H(z) \in \mathbb{C}[z]^{NxM}$, and $\mathbf{x}(z) \in \mathbb{C}[z]^{Mx1}$ where $\mathbb{C}[z]$ represents the field of rational polynomials in z with complex coefficients of size determined by the superscript MxN. The analysis herein will be limited to rational function in z for $\mathbb{C}[z]$ since these are the functions that arise in the problems addressed in this dissertation. Other functions that are commonly used in this chapter are I_P, which represents an identity matrix of size P, and $\delta(i)$ which represents the Dirac delta.

Blind signal and system estimation depends greatly on the invertibility of the transfer function $H(z)$ as will be seen in this chapter. The following section addresses some differences between single-input-single-output (SISO) and multiple-input-multiple-output (MIMO) system when it comes to $H(z)$ and its inverse. The section also develops the background for understanding much of the analysis presented in this dissertation.

5.2 Invertibility of *H*(z)

In SISO systems, it is well known that a finite impulse response (FIR) filter, $H(z) \in \mathbb{C}[z]$, has an infinite impulse response (IIR) inverse filter, $H(z)^{-1} \in \mathbb{C}[z]$, except for the trivial case [2]. For MIMO systems, on the other hand, non-trivial FIR filters can have finite

inverses. To start the discussion on the inverse of $H(z)$ it is better to first tackle the case of N=M transfer functions and then extend the results to the general NxM case.

5.2.1 Inverses of $H(z)\in\mathbb{C}[z]^{NxN}$

With $H(z)$ being FIR and square, N=M, its left inverse can be computed directly from the elements of $H(z)$ as follows

$$H(z)^{-1} = \frac{adj(H(z))}{\det(H(z))} \qquad H(z)^{-1}H(z) = I_N \ \forall\ z \tag{5-6}$$

which corresponds to the adjoint matrix of $H(z)$ divided by its determinant. From an examination of the properties of the adjoint matrix, one can note that the $adj(H(z))$ will have FIR characteristics since $H(z)$ is assumed to be FIR. This is because the minors of the matrix $H(z)$ are composed of summations and products of the elements in $H(z)$ which have themselves a finite impulse response. Therefore, the only factor that determines if $H(z)^{-1}$ has a finite or infinite impulse response is its determinant for the N=M case.

Since $H(z)$ is FIR, the determinant of $H(z)$ will be a finite polynomial in z or in general $\det(H(z))\in\mathbb{C}[z]$. One of the questions that comes to mind is for what determinant values of $H(z)$ is $H(z)^{-1}$ in (5-6) FIR? By observation, if the determinant of $H(z)$ is a non-zero constant (not a function of z also termed *unimodular* [3]) or if it is of the form cz^i for any non-infinite integer i and non-zero complex constant c (a *monomial* in z), there will be a finite length inverse $H(z)^{-1}$ to the filter $H(z)$. This follows from the previous discussion that the $adj(H(z))$ has elements with FIR properties and if these elements are divided by cz^i, they will also be FIR. These types of systems are usually termed *equalizable* [4] because they are finitely, FIR, invertible. An equilizable system implies that it can be exactly inverted in a finite amount of time by a finite length filter, this is important in the actual implementation of the inverse.

If on the other hand, the determinant of $H(z)$ took on a polynomial form that was not monomial, the elements of $H(z)^{-1}$ would have a pole for each zero of the det($H(z)$). The existence of poles, in general, means that one or more elements of $H(z)$ would have an infinite impulse response which in turn would make $H(z)^{-1}$ IIR. It can be concluded that the inverse filter $H(z)^{-1} \in \mathbb{C}[z]^{NxN}$ of an FIR filter $H(z) \in \mathbb{C}[z]^{NxN}$ is in general IIR except for the special case when the determinant of $H(z)$ is of the form cz^i for a non-zero constant $c \in \mathbb{C}$ and $i \in \mathbb{Z}$. Equivalently, if the determinant of $H(z)$ has no zeros, zeros at $|z|=0$, or zeros at $|z|=\infty$, $H(z)$ is finitely invertible.

5.2.2 Inverses for $H(z) \in \mathbb{C}[z]^{NxM}$

The inverses of the general $H(z) \in \mathbb{C}[z]^{NxM}$ are not as easily computed as in the NxN case. At this point it is appropriate to introduce the definition of rank for matrices composed of polynomial elements. For general size matrices, the *normal rank* of the matrix $H(z) \in \mathbb{C}[z]^{NxM}$ is the rank that $H(z)$ assumes for most of the values of z, which in this dissertation is assumed to be M. For square matrices, N=M, the matrix $H(z)$ will loose rank for a particular set of values of z for which $\det(H(z_k))=0$. Since the determinant is not defined for a non-square matrix, a formal definition for normal rank for the general NxM case can be written as

$$rank(H(z)) = M \;\; \forall \; z \notin \{z_k\}_{k=1}^{K} \;\; \Leftrightarrow \;\; rank(H(z))\Big|_{z=z_k} < M \quad z \in \{z_k\}_{k=1}^{K} \tag{5-7}$$

where the normal rank in this case corresponds to the number of columns, or M, since $N \geq M$ is assumed. The zeros of the matrix $H(z)$ will be defined as the set of z_k's for which the rank $H(z)$ drops below M, which occurs for a finite set of size K.

The discussion of finite length invertible filters in the previous section can be extended to non-square channels as will be explained shortly and it will be shown to depend on the locations of the zeros $\{z_k\}$of the matrix $H(z)$. The left inverse of a non-square matrices $H(z) \in \mathbb{C}[z]^{NxM}$ will be denoted by $H(z)^{\#} \in \mathbb{C}[z]^{MxN}$ and has the following property:

$$H(z)^{\#} H(z) = I_M \quad \forall z \tag{5-8}$$

Note that $H(z)^{\#}$ is in general not unique when N>M. Recall that a matrix $H(z) \in \mathbb{C}[z]^{NxM}$ has a normal rank M and can potentially drop rank for particular values of z which correspond to the zeros of $H(z)$ (5-7). If the matrix $H(z)$ drops rank for a particular set of zeros $\{z_k\}$, the inverse filter $H(z)^{\#}$ must find a way to cancel the effects of these zero otherwise it would not be able to produce a full rank M matrix I_M for all z values in (5-8). Intuitively, the elements $H(z)^{\#}$ must therefore have poles that correspond to the zeros of $H(z)$. Parallels can be drawn to the NxN case mentioned earlier where the det($H(z)$) contains the zeros of the matrix $H(z)$ and appears as the poles (on the denominator) of the inverse matrix $H(z)^{-1}$ as can be seen from (5-6). A finite length left inverse $H(z)^{\#}$ exists if $H(z)$ has no zeros (has rank M for all z), has zeros at $|z|=0$, or at $|z|=\infty$. This is a similar result that was found for the NxN case.

Formally stated, an FIR filter $H(z) \in \mathbb{C}[z]^{NxM}$ is *equalizable* if its left inverse is also FIR, such an inverse exists for transfer functions $H(z)$ that have no zeros, zeros at $|z|=0$, or zeros at $|z|=\infty$. A more rigorous treatment can be made to demonstrate that the zeros of a transfer function $H(z)$ show up as the poles of its left inverse $H(z)^{\#}$ by resorting to greatest common devisors (gcd) of a matrix. To proceed further requires the introduction of some additional terminology. A greatest common right divisor (gcrd), $R(z)$, of a set of K matrices $\{A_k(z)\}$ with same number of columns is defined as [3]

$$A_k(z) = \overline{A}_k(z) R(z) \; \forall k \in \{1,...,K\} \tag{5-9}$$

Note that $R(z)$ is not unique but any two right gcrds must be related by a *unimodular* matrix [3]. Additionally, also from [3], if the concatenated matrix

$$\left[A_1(z)^* \cdots A_k(z)^* \cdots A_K(z)^*\right]^* \quad \text{full column normal rank} \tag{5-10}$$

is full column normal rank, all the gcrds (R(z)) will be nonsingular (in the normal rank sense). Under the assumption that R(z) is a greatest common right divisor of all matrices in the set $\{A_k(z)\}$ which have the property (5-9), requires that the set $\{\overline{A}_k(z)\}$ have no common right factors except for a possible unimodular matrix. Such set $\{\overline{A}_k(z)\}$ of matrices are called *relatively right prime* or *right coprime* [3] and have the property that when concatenated as

$$\left[\overline{A}_1(z)^* \cdots \overline{A}_k(z)^* \cdots \overline{A}_K(z)^*\right]^* \quad \text{full rank } \forall \text{ z} \tag{5-11}$$

produce a matrix with full column rank for *all* z. If we let A_k(z) be the rows of the matrix H(z), so K=N, the greatest common right divisor of all rows of H(z) will be of size $R(z) \in \mathbb{C}[z]^{MxM}$. $H(z) \in \mathbb{C}[z]^{NxM}$ can then be decomposed into

$$H(z) = \overline{H}(z)R(z) \tag{5-12}$$

where $\overline{H}(z) \in \mathbb{C}[z]^{NxM}$ has rows that are relatively right prime which ensures by (5-11) that $\overline{H}(z)$ is full column rank for *all* z. Any matrix $\overline{H}(z) \in \mathbb{C}[z]^{NxM}$ which has full column rank for all z (rank($\overline{H}(z)$)=M $\forall$ z) is termed *irreducible*. Note that since H(z) is of normal rank M by assumption, R(z) also has a normal rank of M and is therefore invertible from the discussion surrounding (5-10). Additionally, since $\overline{H}(z)$ does not lose rank for any value of z, the matrix zeros of H(z) must be contained in R(z). This means that R(z) will lose rank for every z_k of H(z). Any gcrd of H(z) must therefore contain all its zeros, z_k, which means that the det($R(z) \in \mathbb{C}[z]^{MxM}$) will also be zero for $z \in \{z_k\}$.

The left inverse $H(z)^{\#}$ can be re-written in terms of the H(z)'s irreducible component matrix $\overline{H}(z)$ and its greatest common divisor R(z). Using the relationship given in (5-12), the left inverse $H(z)^{\#}$ can be found to be composed of

$$H(z)^{\#} = \left(H(z)\right)^{\#} = \left(\overline{H}(z)R(z)\right)^{\#} = R(z)^{\#}\,\overline{H}(z)^{\#} = R(z)^{-1}\,\overline{H}(z)^{\#} = \frac{adj(R(z))}{\det(R(z))}\,\overline{H}(z)^{\#} \quad (5\text{-}13)$$

the inverse of $R(z)$ and the left inverse of the irreducible component of $H(z)$. Note that the determinant of $R(z)$ divides all the components of $H(z)^{\#}$. Since the determinant of $R(z)$ contains the zeros of $H(z)$, they in turn become the poles of the elements of any left inverse of $H(z)^{\#}$. This relationship between the zeros of a transfer function $H(z)$ and how they show up as poles in the elements of its left inverse $H(z)^{\#}$ is an important observation that will be used again when addressing the topic of minimum-phase filters in Section 5.3.

There have been several papers published on the lengths of the inverse $H(z)^{\#}$ when certain conditions are imposed on $H(z)$. In [6] it was shown that when $H(z)\in\mathbb{C}[z]^{NxM}$ with M<N is *irreducible* and *column reduced* [3], a finite length inverse $H(z)^{\#}$ of highest degree K in z exists for $K \geq \sum_{m=1}^{M} L_m - 1$. L_m correspond to the highest degree of the polynomial in z of the *m*th column of $H(z)$. In [7] the lengths of inverses for non-column reduced matrices are discussed.

5.2.3 Summary of MIMO Invertibility of *H*(z)

Several mathematical results were derived in the foregoing section. This section concludes with a summary of the important results that were derived in this chapter and that will be used again in future parts of this dissertation. In Section 5.2.2, it was determined that the zeros of a MIMO transfer function $H(z)\in\mathbb{C}[z]^{NxM}$, defined in (5-7), appear as poles on the elements of its left inverse $H(z)^{\#}\in\mathbb{C}[z]^{MxN}$. Also proven was that an FIR transfer function $H(z)\in\mathbb{C}[z]^{NxM}$ that has no zeros (is irreducible) has zeros at $|z|=0$, or if it has zeros at $|z|=\infty$ has a finite length left inverse. This means that $H(z)^{\#}$ is also FIR under these conditions.

In practical situations, an FIR $H(z)$ will not necessarily have an inverse that is also FIR (*equalizable* systems). This is a problem in blind system identification since channel identification is done in reverse order by first estimating the inverse $H(z)^{\#}$ and then reconstructing $H(z)$ by inverting $H(z)^{\#}$. The details of this procedure were given in Chapter 4 for non-time dispersive channels and will be addressed in this chapter for the time-dispersive MIMO channels. Practically, this would require an infinite number of samples to accurately identify $H(z)^{\#}$ if $H(z)$ is not equalizable. It is impossible to wait an infinite amount of time in real systems, therefore at some point the number of samples that are being processed need to be truncated. This results in an error in the estimate $\hat{H}(z)^{\#}$ and therefore an error in $\hat{H}(z)$. The error, however, might not be very large since even though $H(z)^{\#}$ is IIR it probably decays fairly quickly and a finite length description $\hat{H}(z)^{\#}$ of the inverse probably matches the original inverse very well. In practical situations, it is probably best to use as many samples as is feasible and to keep in mind that the actual estimated inverse might be only a truncated version of the original inverse. This should, however, not prevent the use of blind signal and system identification but the user should be aware of possibly small errors in the estimate due to truncation in addition to other possible estimation errors.

5.3 Decomposition of a Causal and FIR $H(z)$

Every stable and causal filter or transfer function $H(z) \in \mathbb{C}^{NxM}[z]$ can be decomposed into a *minimum-phase* filter, denoted $H^{mp}(z) \in \mathbb{C}^{NxM}[z]$, and a stable *all-pass* filter, denoted by $H^{ap}(z) \in \mathbb{C}^{MxM}[z]$. That fact that this decomposition can be made will prove useful in understanding blind signal and system identification and will also demonstrate the limitations of system identification using just second order statistics. The decomposition of $H(z)$ for a SISO system into its minimum-phase/all-pass parts is well known but its generalization to MIMO systems has not been addressed in literature to the best of the author's knowledge. Before beginning the discussion of the decomposition of an arbitrary causal and stable filter $H(z)$ into its components, the terms *all-pass* and

minimum-phase and their properties will be briefly reviewed. These properties are important for much of the development for this and subsequent chapters.

5.3.1 Minimum-phase Filters

A filter $H^{mp}(z) \in \mathbb{C}^{NxM}[z]$ is termed *minimum-phase* if it has the special property that itself and its left inverse has an impulse response that is *stable* and *causal* [1]. Defining $H^{mp}(z)^{\#} \in \mathbb{C}^{NxM}[z]$ as the left inverse of $H^{mp}(z)$, both filters are related through an equality which can be mathematically written as

$$H^{mp}(z)^{\#} H^{mp}(z) = I_M \tag{5-14}$$

in the z-domain or

$$\sum_{i=0}^{\infty} H_i^{mp\#} H_{n-i}^{mp} = I_M \delta(n) \tag{5-15}$$

in the time domain were $\delta(n)$ is the Dirac delta function. It should also be noted that the left inverse, $H^{mp}(z)^{\#}$, of $H^{mp}(z)$ is in general not unique unless $H^{mp}(z)$ is a square matrix with full normal rank.

The s*tability* property of minimum-phase filters implies that the impulse response H_i^{mp} (and consequently $H_i^{mp\#}$) is absolutely summable, i.e.

$$\sum_{i=0}^{\infty} \left| H_i^{mp} \right| < \infty, \ \sum_{i=0}^{\infty} \left| H_i^{mp\#} \right| < \infty \tag{5-16}$$

This ensures that when a bounded input is applied to $H^{mp}(z)$ (or $H^{mp}(z)^{\#}$), the output is also guaranteed to be bounded (BIBO stability). The second property of minimum-phase filters is *causality* which is satisfied when the impulse response H_i^{mp} (and

respectively $H_i^{mp\#}$) is zero for values of i<0 (negative time). These properties are easy to see in the time domain but not so intuitive when they are translated into the z-domain. To better understand minimum-phase systems it helps to start with the single-input-single-output (SISO) system, M=N=1, and extend these ideas to multiple-input-multiple-output (MIMO) systems.

5.3.1.1 Minimum-phase in SISO Systems

A single input single output linear time invariant and causal system relating an input $\mathbf{x}_n$ with an output $\mathbf{y}_n$ can be written in time and z-domain as

$$\mathbf{y}_n = \sum_{i=0}^{\infty} H_i \mathbf{x}_{n-i} \quad \Leftrightarrow \quad \mathbf{y}(z) = H(z)\mathbf{x}(z) \tag{5-17}$$

where in this subsection $H(z)$, $\mathbf{y}(z)$, and $\mathbf{x}(z) \in \mathbb{C}[z]$ (since N=M=1). Recall that absolute summability ($|\bullet|$ in the SISO case is the absolute value) of the impulse response coefficients

$$\sum_{i=0}^{\infty} |H_i| < \infty \quad \Rightarrow \quad H(z) = \sum_{i=0}^{\infty} H_i z^{-i} \le \sum_{i=0}^{\infty} |H_i| |z^{-i}| < \infty \ \textit{for} \ |z| \ge 1 \tag{5-18}$$

implies BIBO stability and H_i is nonzero for $i \ge 0$ for causality as shown above. From (5-18), one can see that $H(z)$ is causal and stable if $H(z)$ is analytic for $|z| \ge 1$. The above summation for $H(z)$ will produce a result that is finite as long as $|z| \ge 1$ because absolute summability has been imposed on the coefficients of the impulse response. Analytic means that $H(z)$ is finite for values of $|z| \ge 1$ or that its region of convergence (ROC) is on and outside the unit circle on the z plane. Summarizing, if $H(z)$ is analytic (finite) in the region on and outside of the unit circle on z-plane, it has a causal and stable impulse response H_i.

In the SISO case, $H(z) \in \mathbb{C}[z]$, being analytic for $|z| \geq 1$ translates into a transfer function that has all its poles strictly within the unit circle ($|z|<1$). If $H(z)$ had one or more poles on or outside the unit circle, the summation of $H(z)$ in (5-18) would diverge for these values and either the absolutely summable and/or causality condition would be violated. This implies that a stable and causal impulse response, H_i, of a SISO system has a corresponding z transforms with all its poles inside the unit circle.

In order for $H(z)$ to be minimum-phase, however, its inverse needs also to be causal and stable. This requires that the poles of $H(z)^{-1}$ also be inside the unit circle. It should be noted that the poles of $H(z)$ correspond to the zeros of $H(z)^{-1}$ and vice versa for the SISO system. A minimum-phase filter, $H^{mp}(z) \in \mathbb{C}[z]$, therefore has *all* its poles and zeros strictly inside the unit circle.

The decomposition of an arbitrary $H(z)$ into its minimum-phase and all-pass components in the SISO case is also fairly straight forward [2]. This decomposition is best illustrated through an example. Assuming $H(z)$ is a finite impulse response filter consisting of two zeros one inside and one outside the unit circle as follows

$$H(z) = (1 - \frac{1}{2} z^{-1})(1 - 3z^{-1}) \tag{5-19}$$

which is obviously not minimum-phase according to the definition above since one of its zeros is located at z=3 (outside the unit circle). The procedure consists of reflecting the zeros that are outside the unit circle into the inside of the unit circle. $H(z)$ can then be written as a product of a minimum-phase filter and an all-pass filter as follows:

$$H(z) = \underbrace{(1-\frac{1}{2}z^{-1})(3-z^{-1})}\underbrace{\frac{(1-3z^{-1})}{(3-z^{-1})}}$$

$$H(z) = H^{mp}(z) \quad \bullet \quad H^{ap}(z) \quad note: \left|H^{ap}(e^{j2\pi f})\right| = 1 \qquad (5\text{-}20)$$

$$H^{ap}(z)H^{ap^*}\left(\frac{1}{z^*}\right) = 1$$

$H^{mp}(z)$ contains all the factors (poles and zeros) that are inside the unit circle while the all-pass filter, $H^{ap}(z)$, contains the remaining terms. $H^{ap}(z)$ has the special property that its magnitude response is unity for all frequencies (hence the term all-pass). All-pass filters and its generalization to MIMO systems will be discussed in greater detail in Section 5.3.2.

Extending the concept of minimum-phase from SISO to MIMO transfer functions, $H(z) \in \mathbb{C}^{NxM}[z]$, is not as straight forward and requires special attention. Since MIMO systems are primarily considered in this dissertation, the decomposition of $H(z)$ into its minimum-phase and all-pass components is fundamental in understanding blind system identification. The concepts developed next will be used throughout this and subsequent chapters.

5.3.1.2 Minimum-phase in MIMO systems

A MIMO filter, $H(z) \in \mathbb{C}^{NxM}[z]$, is minimum-phase if it and its left inverse, $H(z)^{\#} \in \mathbb{C}^{MxN}[z]$, are both BIBO stable and causal. As with the SISO case, this implies that the region of convergence of $H(z)$, respectively $H(z)^{\#}$, is on and outside the unit circle in the z-plane since both impulse responses are under the minimum-phase condition absolutely summable

$$\sum_{i=0}^{\infty}\left|H_i^{mp}\right| < \infty \quad \Rightarrow \quad H^{mp}(z) = \sum_{i=0}^{\infty} H_i^{mp} z^{-i} \le \sum_{i=0}^{\infty}\left|H_i^{mp}\right|\left|z^{-i}\right| < \infty \ \ for\ |z| \ge 1 \qquad (5\text{-}21)$$

$$\sum_{i=0}^{\infty}\left|H_i^{mp\,\#}\right|<\infty \quad\Rightarrow\quad H^{mp}(z)^{\#}=\sum_{i=0}^{\infty}H_i^{mp\,\#}z^{-i}\le\sum_{i=0}^{\infty}\left|H_i^{mp\,\#}\right|\left|z^{-i}\right|<\infty \;\; for \;\; |z|\ge 1$$

so $H(z)$, respectively $H(z)^{\#}$, will be finite for $|z|\ge 1$. In the above equation, the norm $|\bullet|$ is the matrix norm which corresponds to the largest singular value of the matrix. Summarizing, $H(z)$ and $H(z)^{\#}$ need to be analytic (finite) for $|z|\ge 1$ for the filter $H(z)$ to be minimum-phase.

This is similar to the result that was derived for SISO. Unlike the SISO case, however, the definitions of poles and zeros as defined for $H(z)\in\mathbb{C}[z]$ do not extend directly to a MIMO filter $H(z)\in\mathbb{C}^{NxM}[z]$. In order for $H(z)$ to be stable and causal, all the elements of $H(z)$ must be stable and causal; so individually each element must have its poles inside the unit circle and only exist for zero or positive time. The question that remains to be answered is what properties of $H(z)\in\mathbb{C}^{NxM}[z]$ need to satisfied for it to be minimum-phase - recall that for the SISO system, $H(z)$ had to have all its poles and zeros inside the unit circle to be minimum-phase.

Recall that in Section 5.2.2 it was shown that the zeros of the transfer function $H(z)$, as defined in (5-7), showed up as the poles in the elements of its left inverse $H(z)^{\#}$. Since this was already demonstrated, this time the same conclusion will be reached but through an example. Recall that $H(z)\in\mathbb{C}^{NxM}[z]$ is FIR and has elements composed of rational polynomials in z and is of normal rank M, its columns can possibly lose rank for a finite set values of z. For example,

$$H(z)=\begin{bmatrix}(1-\frac{1}{2}z^{-1}) & 0\\ (1-3z^{-1}) & 4\end{bmatrix} \tag{5-22}$$

has a *normal* rank of 2 (rank is 2 for most values of z) but when is z=1/2, the rank of $H(z)$ reduces to 1. This means that the above example $H(z)\in\mathbb{C}^{2x2}[z]$ has a single zero at $z_1=1/2$. This could have also been found by computing the zeros of the determinant of

H(z) since the matrix is square (det(H(z))=4(1-1/2z^{-1})). The inverse of H(z) can be readily computed from (5-6) to be

$$H(z)^{-1} = \begin{bmatrix} \dfrac{1}{(1-\frac{1}{2}z^{-1})} & 0 \\ -\dfrac{(1-3z^{-1})}{4(1-\frac{1}{2}z^{-1})} & \dfrac{1}{4} \end{bmatrix} \tag{5-23}$$

Note that some of the elements of H(z)$^{-1}$ have poles that correspond to the zeros of the transfer function of H(z). This should be of no surprise since this possibility was already established in Section 5.2.2 which dealt with MIMO inverses. All components of H(z)$^{-1}$ have poles that are inside the unit circle which ensures that the matrix H(z)$^{-1}$ is analytic (finite) for |z|≥1. The same can be said about H(z) since it is FIR. Therefore, the example filter above is minimum-phase.

In general, an NxM transfer function H(z)∈H^{mp}(z) if it and its left inverse H(z)$^{\#}$ is analytic for |z|≥1. The properties that are required by H(z) to be minimum phase are the following: H(z) has to be analytic (not diverge) for |z|≥1 and all its matrix zeros, defined in (5-7), have to be inside the unit circle. Recall that minimum-phase means that H(z) and its inverse are BIBO stable and causal. Note that because an FIR irreducible (defined in Section 5.2.2) system has no matrix zeros it will always be minimum phase.

The next subsection examines some properties of MIMO all-pass filters. It will be shown that there exists a general generating function for an arbitrary MxM MIMO causal all-pass system. Because of their structure, second order statistics are blind to these types of filters as will be explored in more detail in Chapter 6.

5.3.2 MIMO All-pass Filters

In this section the discussion will be limited to MxM causal MIMO filters because these are the type of filters encountered in blind system identification. Most of the information contained in this section is based on the work done by Vaidyanathan in the book [5]. An all-pass filter is often termed lossless because its magnitude response versus frequency that relates the input to the output of the system is the identity matrix. An MxM all-pass filter $H^{ap}(z)\in\mathbb{C}^{MxM}[z]$ as used in this dissertation has the following property:

$$H^{ap}(z)H^{ap^*}\left(\frac{1}{z^*}\right)=I_M \tag{5-24}$$

this property is termed *paraunitary* [5] and the filters that posses such a property are *paraunitay MIMO filters* which comes from $H^*\left(\frac{1}{z^*}\right)$ which is referred to as the *para-Hermitian conjugate* of $H(z)$ [1]. The above is just a generalization of the lossless condition to the entire complex plane but in terms of the more familiar frequency domain ($z=e^{j\omega}$) it reduces to

$$H^{ap}(e^{j\omega})H^{ap^*}\left(e^{j\omega}\right)=I_M \tag{5-25}$$

Any filter $H(z)\in\mathbb{C}^{MxM}[z]$ which possesses the paraunitary property in (5-24) will be termed *lossless*, *all-pass*, or *paraunitary* interchangeably in this dissertation. The above properties can be seen to be generalizations of what was found for SISO systems in (5-20).

Note also from (5-24) that the para-Hermitian conjugate of an all-pass MIMO MxM filter is also its inverse since their product yields the identity matrix. This result means that the inverse of $H^{ap}(z)$ is given by

$$H^{ap}(z)^{-1} = H^{ap^*}\left(\frac{1}{z^*}\right) \tag{5-26}$$

which is simply computed by replacing z by $1/z^*$ in the matrix $H^{ap}(z)$ and taking the Hermitian transpose. The resulting inverse therefore has the same order as the original filter – this is one of the characteristics of an all-pass filter.

In the time domain, the para-Hermitian conjugation corresponds to flipping about the time axis and Hermitian conjugating the individual impulse response components. For example, the inverse of an all-pass filter $H^{ap}(z)$ of order two given by its time impulse response

$$H^{ap}(z) \quad \Leftrightarrow \quad H_0^{ap} = F_0, H_1^{ap} = F_1, H_2^{ap} = F_2 \tag{5-27}$$

has an inverse of the same length but with its coefficients equal to

$$H^{ap}(z)^{-1} \quad \Leftrightarrow \quad H_{-2}^{ap^{-1}} = F_2^*, H_{-1}^{ap^{-1}} = F_1^*, H_0^{ap^{-1}} = F_2^* \tag{5-28}$$

which correspond to the time flipped and Hermitian conjugated components of the original filter

5.3.2.1 FIR MxM MIMO All-Pass Filters

An FIR causal paraunitary filter can be constructed in the following manner as described in [5]. A finite impulse response filter of order L can be constructed by concatenating a series of degree-one building blocks $V_l(z)$ defined as

$$V_l(z) = I_M - v_l v_l^* + z^{-1} v_l v_l^* \tag{5-29}$$

where v_l is an Mx1 vector with unit norm such that

$$v_l^* v_l = 1 \tag{5-30}$$

Note that the matrix $P_l=v_l v_l^*$ spans the range of v_l and has matrix projection characteristics. Additionally, the matrix that projects onto the space orthogonal to the range of v_l can be found by subtraction $Q_l=I_M-P_l=I_M-v_l v_l^*$. The matrices P_l and Q_l follow the properties of projection matrices onto orthogonal spaces and have the relationships

$$P_l^2=P_l,\ Q_l^2=Q_l,\ P_l=P_l^*,\ Q_l=Q_l^*,\ Q_l P_l=P_l Q_l=0, \quad P_l,Q_l \in \mathbb{C}^{MxM} \tag{5-31}$$

as would be expected. $V_l(z)$ can therefore be thought of as an operator that applies a delay to a resultant projection onto the v_l space but leaves the resultant projection onto the space orthogonal to v_l as is. $V_l(z)$ can be written in terms of P_l and Q_l as follows:

$$V_l(z) = Q_l + z^{-1} P_l \tag{5-32}$$

A general order L causal all-pass FIR filter can be constructed by concatenating a series of degree 1 lossless building blocks starting with a unitary matrix as follows [5]

$$H_L^{ap}(z) = V_L(z) V_{L-1}(z) ... V_1(z) H_0 \tag{5-33}$$

where H_0 is an MxM unitary matrix such that $H_0 H_0^*=I_M$. The determinant of $V_l(z)$ can be computed to be z^{-1} and the determinant of $H_0=\beta$. Since H_0 is unitary, $|\beta|=1$, and the total order L FIR all-pass filter has

$$\det\left(H_L^{ap}(z)\right) = \beta z^{-L} \tag{5-34}$$

The inverse a causal all-pass FIR filter has some important characteristics. The inverse can be computed by performing para-Hermitian conjugation on $H_L^{ap}(z)$ and can be determined to be

$$H_L^{ap}(z)^{-1} = H_L^{ap^*}\left(\frac{1}{z^*}\right) = H_0{}^* V_1(z)^{-1} \ldots V_{L-1}(z)^{-1} V_L(z)^{-1} \tag{5-35}$$

$$V_l(z)^{-1} = V_l^*\left(\frac{1}{z^*}\right) = Q_l{}^* + zP_l{}^* = Q_l + zP_l = V_l(z^{-1})$$

An interesting observation about the inverse $H_L^{ap}(z)^{-1}$ is that it has the same length as the original filter $H_L^{ap}(z)$ which means that an FIR causal all-pass filter is *equalizable*. Also, the filter $H_L^{ap}(z)^{-1}$ is *anticausal* since it is composed of $V_l(z^{-1})$ building blocks which have a negative delay (5-35). Therefore in order to implement the inverse filter would require delaying its response by its length in order to make it realizable which in turn would produce a delay in the recovered signal.

5.3.2.2 IIR MxM MIMO All-Pass Filters

Similarly from [5], a general IIR causal MxM MIMO paraunitary filter can be created using the following building block for $V_l(z)$

$$V_l(z) = I_M - v_l v_l{}^* + \left(\frac{-a_l{}^* + z^{-1}}{1 - a_l z^{-1}}\right) v_l v_l{}^* \quad |a_l| < 1$$
$$V_l(z) = Q_l + \left(\frac{-a_l{}^* + z^{-1}}{1 - a_l z^{-1}}\right) P_l, \quad \det(V_l(z)) = \frac{-a_l{}^* + z^{-1}}{1 - a_l z^{-1}} \tag{5-36}$$

where P_l and Q_l are projection matrices defined as in the previous section. The magnitude of the pole a_l is restricted to be inside the unit circle to ensure stability of the filter. As in the FIR case, an order L all-pass IIR filter can be created by first starting with a unitary matrix H_0 and concatenating to it a series of $V_l(z)$ building blocks until the desired order is reached.

$$H_L^{ap}(z) = V_L(z)V_{L-1}(z)...V_1(z)H_0, \qquad \det\left(H_L^{ap}(z)\right) = \beta\prod_{l=1}^{L}\frac{-a_l^{*} + z^{-1}}{1-a_l z^{-1}} \tag{5-37}$$

Actually, the causal IIR lossless filter also includes as a special case the FIR all-pass filter mentioned in the previous subsection which can be observed by setting a_l to zero in (5-36). The inverse of a causal IIR lossless filter also takes on a note worthy form. Recall that the inverse of a paraunitary filter is its para-Hermitian conjugate or equivalently

$$H_L^{ap}(z)^{-1} = H_L^{ap^*}\left(\frac{1}{z^*}\right) = H_0^{*}V_1(z)^{-1}...V_{L-1}(z)^{-1}V_L(z)^{-1} \tag{5-38}$$

$$V_l(z)^{-1} = V_l^{*}\left(\frac{1}{z^*}\right) = Q^* + \left(\frac{-a_l + z}{1-a_l^{*}z}\right)P^* = Q + \left(\frac{-a_l + z}{1-a_l^{*}z}\right)P$$

Note that since the magnitude of a_l was restricted to be less than 1 in (5-36), the inverse $H_L^{ap}(z)^{-1}$ has poles that are outside the unit circle. At first glance this would seem to give rise to an unstable system, but this is not the case. Poles outside the unit circle would produce unstable filters if the filter was causal, but the above inverse is anti-causal which requires that the poles be strictly outside the unit circle for stability. Stable anti-causal filters have region of convergences on and inside the unit circle (poles must be outside the unit circle) so the inverse system is also BIBO stable.

5.4 Decomposition of MIMO FIR Filters

This section draws from prior results and definitions in this chapter. What follows is a procedure for a general decomposition of a causal and stable $H(z)$ into its minimum-phase and all-pass filter components. As far as the author knows, such decomposition has not been presented in general for FIR MIMO systems in literature. The decomposition that is being sought is of the type

$$H(z)=H^{mp}(z)H^{ap}(z) \tag{5-39}$$

where $H(z)\in\mathbb{C}^{NxM}[z]$, and $H^{mp}(z)\in\mathbb{C}^{NxM}[z]$, and $H^{ap}(z)\in\mathbb{C}^{MxM}[z]$.

An arbitrary FIR stable and causal $H(z)\in\mathbb{C}^{MxM}[z]$ is not necessarily minimum-phase. As was shown in Section 5.2.2 that a MIMO filter is minimum-phase if it has all its zeros inside the unit circle (does not lose rank for $|z|\geq 1$) and is analytic in the region on and outside the unit circle, $|z|\geq 1$ (its ROC is on and outside the unit circle). A causal FIR filter is always analytic in the region outside the unit circle because it is BIBO stable but in general does not have all its zeros within the unit circle as defined in (5-7).

The procedure presented here can be viewed as a generalization of what was done for SISO systems in Section 5.3.1.1. The method is based on reflecting the zeros of $H(z)$ that lie outside the unit circle towards the inside. First, let the set $\{z_k\}$ denote all the zeros with magnitude *greater than* one, $|z_k|>1$, for which $H(z)$ loses rank. These are the zeros that cause $H(z)$ to violate the minimum-phase condition. Therefore for each zero $\{z_k\}$ (possibly repeating zeros) the matrix $H(z=z_k)\in\mathbb{C}^{NxM}$ will lose rank by definition and there will exist a non-zero vector $p_k\in\mathbb{C}^{Mx1}$ in the null space of $H(z=z_k)$ such that

$$H(z=z_k)p_k=0 \quad rank(H(z=z_k))<M, \qquad p_k^{*}p_k=1 \tag{5-40}$$

Without loss of generality let p_k be a unit vector since the interest is in the span of the null space. The product of $H(z)$ and the vector p_k can be written as

$$H(z)p_k=v_k(z)(1-z_k z^{-1}) \tag{5-41}$$

where $v_k(z)\in\mathbb{C}^{Nx1}[z]$ and both sides of the equal sign are zero when evaluated $z=z_k$. The difference between the MIMO and the SISO system is that for the general $H(z)\in\mathbb{C}^{NxM}[z]$, zeros have also an associated direction to them. As mentioned in the previous section $P_k=p_k p_k^{*}$ is a projection matrix that as it name states projects a vector or matrix along the

p_k direction which is the associated space for which $H(z)$ loses rank for the particular zero z_k. It is along the p_k direction that the zero needs to be reflected into the unit circle as was done for SISO systems because along other directions $H(z)$ does not lose rank because of the z_k^{th} zero. First consider z_1 with its associated projection matrix onto the range of p_1, P_1, and the projection matrix orthogonal to the range of p_1, $Q_1=I_M-P_1$. The first zero with magnitude greater than one of $H(z)$ can be projected into the unit circle as follows:

$$\begin{array}{ccc} H(z)\left(Q_1+\left(\dfrac{-1+z_1^{*}z}{z_1-z}\right)P_1\right)=H^{(1)}(z) & & \\ \Uparrow & \Uparrow & \\ \{z_k\}_{k=1}^{K} & \{z_k\}_{k=2}^{K} & |z_k|>1 \end{array} \tag{5-42}$$

Note that only the projection onto the range of p_1 is reflected into the unit circle while in all other directions $H(z)$ is left untouched. The resultant matrix $H^{(1)}(z)$ loses rank for one less zero that is outside the unit circle than $H(z)$. $H^{(1)}(z)$ has zeros that belong to the set $\{z_k\}_{k=2}^{K}$, recall that the set $\{z_k\}$ only contains the zeros that have the property $|z_k|>1$. This procedure can be repeated for all the zeros (possibly repeated) that lie outside the unit circle until all the zeros are reflected into the circle resulting in a minimum-phase system. After K such iterations (for all K zeros of $H(z)$), a resulting minimum-phase system is formed with all its zeros inside the unit circle

$$H(z)\left(Q_1+\left(\frac{-1+z_1^{*}z}{z_1-z}\right)P_1\right)\cdots\left(Q_K+\left(\frac{-1+z_K^{*}z}{z_K-z}\right)P_K\right)=H^{mp}(z) \tag{5-43}$$

The individual blocks that make up the equation above are of the form $V_k(z)^{-1}$ (5-38) which has an inverse given by $V_k(z)$ in (5-36). Moving the individual blocks on the minimum-phase transfer function side results in

$$H(z)=H^{mp}(z)\left(Q_K+\left(\frac{-1+z_K z^{-1}}{z_K^{*}-z^{-1}}\right)P_K\right)\cdots\left(Q_1+\left(\frac{-1+z_1 z^{-1}}{z_1^{*}-z^{-1}}\right)P_1\right) \tag{5-44}$$

The matrices following the minimum-phase filter correspond to the building blocks of an IIR causal lossless system with $H_0=I_M$ as seen from (5-37). This concludes the procedure for demonstrating that a decomposition of an arbitrary FIR causal MIMO system $H(z)$ can be obtained in the form

$$H(z) = H^{mp}(z)H^{ap}(z) \tag{5-45}$$

where $H^{mp}(z)\in\mathbb{C}^{NxM}[z]$ is a minimum-phase system and an all-pass filter $H^{ap}(z)\in\mathbb{C}^{MxM}[z]$.

Special cases that could arise in performing the above decomposition will be treated next. Note that from (5-41) it is the implicit assumption that $H(z)$ loses rank for each zero by only one dimension since p_k is assumed to be a vector. The algorithm above, however, is not restrictive to $H(z)$'s that lose rank by only one dimension for a given zero since the procedure above can be repeated again for the remaining dimensions of the null-space until a matrix that does not drop rank for that particular zero is created. Secondly, repeated zeros are also covered by the above decomposition, even though it is not specifically stated. Each repeated zero will have to be listed separately in the $\{z_k\}$ according to their multiplicity.

Lastly, what remains to be answered is if the decomposition in (5-45) is unique or if other decomposition of the same form exist. To test uniqueness consider two minimum-phase/all-pass factorizations as follows

$$H(z) = H^{mp}(z)H^{ap}(z) = \breve{H}^{mp}(z)\breve{H}^{ap}(z) \tag{5-46}$$

and let $A(z)\in\mathbb{C}^{MxM}[z]$ such that without loss of generality

$$H^{mp}(z) = \breve{H}^{mp}(z)A(z)^{-1} \qquad A(z) = \breve{H}^{ap}(z)H^{ap}(z)^{-1} \tag{5-47}$$

Note that the equation on the right in (5-47) for $A(z)$ is composed of the multiplication of an all-pass filter and an inverse of an all-pass. $A(z)$ is therefore also an all-pass filter and has para-unitary properties and is composed of the building blocks $V_l(z)$ and $V_l(z)^{-1}$ and unitary matrices. Since $A(z)$ is para-unitary it is invertible and its inverse $A(z)^{-1}$ is also para-unitary. The determinants of $A(z)$ and $A(z)^{-1}$ are composed of factors with a pole inside the unit circle and a zero outside the unit circle or vice versa (5-37). Additionally, every zero in the determinant of $A(z)$ is a pole in the determinant of $A(z)^{-1}$ and vice versa. By the previous two argument, $A(z)$, $A(z)^{-1}$, or both will have zeros outside the unit circle and when multiplied with a minimum-phase filter will yield non-minimum-phase matrices so the equalities

$$H^{mp}(z) = \breve{H}^{mp}(z)A(z)^{-1}, \quad A(z)H^{mp}(z) = \breve{H}^{mp}(z) \tag{5-48}$$

are only possible if the para-unitary matrix $A(z)$ is a unitary matrix A. Therefore any two realizations of minimum-phase/all-pass decompositions are related by

$$H^{mp}(z) = \breve{H}^{mp}(z)A^{*} \qquad H^{ap}(z)A = \breve{H}^{ap}(z) \quad where\ A^{*}A = I_M \tag{5-49}$$

which concludes the proof of minimum-phase uniqueness.

5.5 Summary

In this chapter the invertiblity of a causal and stable MIMO FIR matrix $H(z) \in \mathbb{C}^{NxM}[z]$ was addressed. Finite impulse response MIMO systems, unlike SISO systems, can potential have inverses that are FIR as was mentioned in Section 5.2.2 for non-trivial matrices under some circumstances. This special class of filters was termed *equalizable* in [4]. In general, however, the inverse of $H(z)$ can possibly have an infinite impulse response depending on the location of the zeros of $H(z)$. In this chapter, it was also shown that $H(z)$ can be decomposed into a minimum-phase and an all-pass filter (5-45). The definition for these filters and how the decomposition can be made was also discussed in detail.

References

[1] T. Kailath, A. H. Syed, and B. Hassibi, *Linear Estimation*. Upper Saddle River, New Jersey: Prentice Hall, 2000.

[2] John G. Proakis and Dimitris G. Manolakis, *Digital Signal Processing, Principles, Algorithms, and Applications*, Prentice Hall, New Delhi India 2002.

[3] T. Kailath, *Linear Systems*, Prentice Hall, Englewood Cliffs, New Jersey 1980.

[4] Y. Inouye and R.-W. Liu, "A System-Theoretic Foundation for Blind Equalization of and FIR MIMO Channel System," *IEEE Trans. on Circuits and Systems-I*, vol. 49, no. 4, pp. 425-436, April 2002.

[5] P.P. Vaidyanathan, *Multirate Systems and Filter Banks*, Prentice Hall, Englewood Cliffs, New Jersey, 1993.

[6] A. Gorokhov and P. Loubaton, "Subspace-Based Techniques for Blind Separation of Convolutive Mixtures with Temporally Correlated Sources," *IEEE Trans. on Circuits and Systems-I*, vol. 44, no. 9, pp. 813-820, September 1997.

[7] J. K. Tugnait, "On Linear Predictors for MIMO Channels and Related Blind Identification and Equalization," *IEEE Signal Processing Letters*, vol. 5, no. 11, pp. 289-291, November 1998.

Chapter 6

Blind System Identification for Convolutive Mixtures

Blind signal and system estimation relies on identifying the channel impulse response and the signals present in the environment from just the recorded information at the sensors. The wireless channel can be represented as a linear time-dispersive system that can be modeled by a series of transfer functions relating the signals to the sensors as shown in Figure 5.1. The object is to estimate all transfer function unknowns and the original signals using the minimum assumptions possible. The more general and the less restrictive the assumptions needed in solving this problem, the greater the applicability of the analysis.

The material presented in this chapter addresses the time-dispersive linear channel and is an extension of the analysis presented in Chapter 4 where mixing matrices were blindly identified. The identification of the dispersive channel is often referred to as the convolutive mixtures problem because the relationship of the input $\mathbf{x}_i$ to the output $\mathbf{y}_i$ can be written as a convolution of matrices

$$\mathbf{y}_i = \sum_{k=-\infty}^{\infty} H_{i-k}\mathbf{x}_k \tag{6-1}$$

in the time domain where $k \in \mathbb{Z}$, $\mathbf{y}_i \in \mathbb{C}^{Nx1}$, $H_i \in \mathbb{C}^{NxM}$, and $\mathbf{x}_i \in \mathbb{C}^{Mx1}$. Its z-domain representation will also be useful in understanding some of the identification properties associated with different order statistics and in the z-domain the output and input are related through the transfer function as follows

$$\mathbf{y}(z) = H(z)\mathbf{x}(z) \tag{6-2}$$

where $\mathbf{y}(z) \in \mathbb{C}[z]^{Nx1}$, $H(z) \in \mathbb{C}[z]^{NxM}$, and $\mathbf{x}(z) \in \mathbb{C}[z]^{Mx1}$. For now it will be assumed that N≥M (at least the same number of sensors as sources in the environment) but this condition will be relaxed later in the chapter. When N≥M an $H(z)^{\#}$ (the inverse of the channel) can exist that will decouple the signals up to the unresolvable ambiguities which will be covered next.

6.1 Unresolvable Ambiguities

The fact that the inputs are not observed leads to some inherent ambiguities when it comes to the identification of the signals. First of all, the power of the transmitting sources cannot be identified with certainty. At the receiver for a given received amplitude one cannot separate path loss from signal power; for example a weak signal transmitting from a nearby location could produce the same response on the sensors as a strong signal that is far away from the receiver (Section 4.2). To resolve this ambiguity, without loss of generality, the transmitters at their source ($\mathbf{x}_i$) are assumed to be transmitting with unity power in this dissertation and the difference between the transmitted signal and the received power is attributed to path loss. Secondly, the time delay between transmitter and receiver is also indeterminable. The absolute time delay each signal encounters when traveling from its source to the destination is also ambiguous. Lastly, as also discussed in Section 4.2 there is a model ambiguity which

comes from writing the equation in the form given in (6-2). Permuting the columns of $H(z)$ and doing the same with $\mathbf{x}(z)$ results in the same vector $\mathbf{y}(z)$ which causes an unobservable change in the received signal due to the permutation. $H(z)$ and $\mathbf{x}(z)$ are therefore only determinable up to a permutation matrix. $H(z)$ will be referred to as *identified* if an estimated inverse of $H(z)$, $\hat{H}(z)^{\#}$, is found such that

$$\hat{H}(z)^{\#} H(z) = \Lambda(z) D P \quad \text{where } \hat{H}(z)^{\#} \hat{H}(z) = I_{\mathrm{M}} \tag{6-3}$$

where $\Lambda(z) = diag(z^{-l_1}, ..., z^{-l_{\mathrm{M}}})$ and $l_m \in \mathbb{Z}$, D is a diagonal matrix with constant elements, and P is a permutation matrix. Throughout this dissertation, the '^' symbol on top of a variable refers to an estimate of that particular variable. Note that the performance of a properly designed receiver should not be affected by the above unresolvable ambiguities since they involve only scaling and delaying of the signals in the system.

6.2 Blind Identification

At first glance, blind system identification appears as an unsolvable problem. From (6-1), only the vector $\mathbf{y}_i$ is known while the coefficients of H_i and $\mathbf{x}_i$ are all unknowns. There are many more unknowns than equations when thought of it in this way. This is, however, not a stumbling point since one can work with a collection of samples of $\mathbf{y}_i$ to estimate H_i and the properties of $\mathbf{x}_i$ using statistics as will be seen shortly. When working with a set $\{\mathbf{y}_i\}$ this system is identifiable if minimal assumptions are known or made about $\{\mathbf{x}_i\}$.

As mentioned in the introduction, the identification of $\{H_i\}$ and $\{\mathbf{x}_i\}$ from the received sequence $\{\mathbf{y}_i\}$ can only be made by making some assumptions. In order for the analysis to be applicable in a general sense, these assumption need to be as non-specific as possible. Certain properties of the source signals need to be assumed and exploited in order to recover $H(z)$ and $\mathbf{x}(z)$ and in such a way as to remain valid for most (or all) of the

cases that are encountered. The model of the sources should be minimal to have the widest applicability. The type of model that is assumed for the elements of $\mathbf{x}_i$ determines the method, or combination of methods, that can be used to identify the system parameters.

There a different ways of identifying a system blindly based on properties of the signals $\mathbf{x}_i$ in the environment. When the signals $\mathbf{x}_i$ are non-white (colored), the channel is blindly identifiable using up to second-order statistics (SOS) as long as $H(z)$ ($N \geq M$) is irreducible and the power spectra of $\mathbf{x}_i$ is diagonal with distinct polynomial factors [1]. There have been several papers published in this area that specifically consider this problem [1], [2], [3], and [4] for example consider $N \geq M$ exclusively.

When the signals $\mathbf{x}_i$ have non-stationary statistics such as cyclostationarity or other forms of non-stationarity, this property can be used to exploit algorithms based on second-order statistics that identify $H(z)$. Identification of $H(z)$ for cyclostationary signals that have different cycle frequencies has been treated in [5], [6], and [7] using SOS but only under certain conditions of $H(z)$. Other forms of identification based on non-stationarity and non-whiteness of signals are treated in [4] and [8].

Lastly, the non-Gaussianity of the source signals can be exploited. This type of identification resorts to higher-order-statistics (HOS) in order to identify the system parameters. The advantage of using this technique is that signals are not required to be correlated in time or have non-stationary characteristics. The HOS based methods most commonly use cummulants as identification tools. Joint cummulants of several random variables can be related to the higher order joint moments of the signals and vice versa (Chapter 3). The trend in modern communications is to move away from analog modulation schemes and move towards digital formats which have information symbols that are non-Gaussian by nature and usually belong to a finite alphabet set which guarantees that HOS is applicable. It is also possible to use a combination of these three techniques to improve the identifiability of the system. In this chapter, the HOS and the

non-stationarity of communications signals will be exploited to identify the channel and possibly more signals than sensors.

In Chapter 4, no specific information about the behavior of the source signals in the time domain was assumed. This was because the channel under consideration was non-dispersive and higher-order statistics could be used directly on the signals across the sensors (the time dependence was immaterial). The assumptions of Chapter 4 where that the transmitters in the system were independently controlled which in turn produced signals that are independent of each other. This chapter is concerned with the identification of the time-dispersive channel (Figure 5.1) which does require that assumptions be imposed on the time behavior of the source signals in order that the channel and the signals be correctly identified. In this dissertation, the signals $\mathbf{x}_i$ that are sought are the fundamental building blocks of the received signal $\mathbf{y}_i$. This imposes the assumption that the elements of the signal $\mathbf{x}_i$ are independent from transmitter to transmitter (in space) and also from sample to sample (in time). The matrix $H(z)$ is therefore the only operator that causes the space-time independent samples of $\mathbf{x}(z)$ to mix into the signal $\mathbf{y}(z)$ that is received. $\mathbf{x}_i$ is assumed to be space-time independent throughout this dissertation because of the lack of other information available, but if additional *a priori* characteristics of the time dependence of $\mathbf{x}_i$ were know before hand, these properties should be used instead of the independence assumption.

The identification procedure is similar to the analysis presented in Chapter 4. It requires finding an inverse filter that minimizes the mutual information between the estimated source signals not only in space (as was done in Chapter 4) but also in time. Because of the complexity of estimating the probability density functions of the various signals a less computational process can be used based on joint cumulants. Recall from Chapter 3, that when a joint random variable is independent (no mutual information), the joint cumulants of all order are zero.

6.3 Blind Identification of Stationary Systems

The analysis that follows considers only stationary signals. The theory developed in this section will be used as a stepping stone for the section that addresses cyclostationary signals.

6.3.1 Estimation using Second Order Statistics for Stationary Signals

In blind signal and channel estimation, $H(z)$ and $\mathbf{x}(z)$ are not observable and their estimates are recovered solely from the information provided in $\mathbf{y}(z)$. Some minimal assumptions are made about $\mathbf{x}(z)$ that usually involve certain dependence (or independence) conditions of the time sequence $\mathbf{x}_i$ from one sample to the next and from transmitter to transmitter. These assumed properties on the time variations of $\mathbf{x}_i$ are used in order to estimate $H(z)$ and recover $\mathbf{x}_i$ from the received signals $\mathbf{y}_i$ by ways of second and/or higher order statistics of the signal $\mathbf{y}_i$. In this section, only stationary signals will be considered; this assumption will be relaxed later.

In this dissertation, an $\mathbf{x}(z)$ is sought which is space-time independent but otherwise has unrestricted properties because $\mathbf{x}_i$ is assumed to be the fundamental building block of the signal $\mathbf{y}_i$. In the second order sense, the stationary signal $\mathbf{x}_i$ can be characterized by its power spectral density (PSD) or its autocorrelation function. Both the PSD of $\mathbf{x}_i$, $S_{\mathbf{x}}(z)$, and its autocorrelation, $R_{\mathbf{x}}(i)$, are z-transform pairs defined as follows

$$S_{\mathbf{x}}(z) = \sum_{k=-\infty}^{\infty} R_{\mathbf{x}}(k) z^{-k} = I_{\mathrm{M}} \quad \Rightarrow \quad R_{\mathbf{x}}(k) = E\left[\mathbf{x}_i \mathbf{x}_{i-k}^{*}\right] = I_{\mathrm{M}} \delta(k) \tag{6-4}$$

The autocorrelation of $\mathbf{x}_i$ is defined as the expected value ($E[\bullet]$) of $\mathbf{x}_i$ times a Hermitian-conjugate and shifted version of itself. Note that from (6-4) the elements of $\mathbf{x}(z)$ are assumed to have unity power and to be wide-sense stationary (since $R_{\mathbf{x}}(k)$ is only a function of delay k and not of i). Recall that unit power was imposed to resolve the ambiguity in the amplitude of the source signals (Section 6.1). Additionally $\mathbf{x}_i$, by

assumption, has components that are independent (uncorrelated in the second-order sense) in time and in space, which results in a PSD that is an identity matrix and which exists only at k=0. Similarly, the PSD and the autocorrelation of the received signal $\mathbf{y}_i$ can be defined as

$$S_{\mathbf{y}}(z) = \sum_{k=-\infty}^{\infty} R_{\mathbf{y}}(k) z^{-k} \quad \Rightarrow \quad R_{\mathbf{y}}(k) = E\left[\mathbf{y}_i \mathbf{y}_{i-k}^{*}\right] \tag{6-5}$$

where $S_{\mathbf{y}}(z)$ is the z-transform of $R_{\mathbf{y}}(k)$ and $\mathbf{y}_i$ is also wide-sense stationary since $\mathbf{x}_i$ has this property. The second-order statistics of the output $\mathbf{y}_i$ can be related to the second-order statistics of the input $\mathbf{x}_i$ by expressing the signal $\mathbf{y}_i$ above in terms of the convolution (6-1) which describes the system. The autocorrelation of $\mathbf{y}_i$ can be found in terms of the autocorrelation of $\mathbf{x}_i$ as follows

$$\begin{aligned} R_{\mathbf{y}}(k) &= E\left[\sum_{m=-\infty}^{\infty} H_m \mathbf{x}_{i-m} \sum_{n=-\infty}^{\infty} \mathbf{x}_{i-k-n}^{*} H_n^{*}\right] = \sum_{m=-\infty}^{\infty}\sum_{n=-\infty}^{\infty} H_m E\left[\mathbf{x}_{i-m}\mathbf{x}_{i-k-n}^{*}\right] H_n^{*} \\ &= \sum_{m=-\infty}^{\infty}\sum_{n=-\infty}^{\infty} H_m R_{\mathbf{x}}(k-m+n) H_n^{*} \end{aligned} \tag{6-6}$$

and from (6-5), the power spectral density of $\mathbf{y}(z)$ can be found to be

$$\begin{aligned} S_{\mathbf{y}}(z) &= \sum_{k=-\infty}^{\infty}\sum_{m=-\infty}^{\infty}\sum_{n=-\infty}^{\infty} H_m R_{\mathbf{x}}(k-m+n) H_n^{*} z^{-k} = \sum_{p=-\infty}^{\infty}\sum_{m=-\infty}^{\infty}\sum_{n=-\infty}^{\infty} H_m R_{\mathbf{x}}(p) H_n^{*} z^{-(p+m-n)} \\ &= \left(\sum_{m=-\infty}^{\infty} H_m z^{-m}\right)\left(\sum_{p=-\infty}^{\infty} R_{\mathbf{x}}(p) z^{-p}\right)\left(\sum_{n=-\infty}^{\infty} H_n^{*} z^{n}\right) \\ &= H(z) S_{\mathbf{x}}(z) H^{*}\left(\frac{1}{z^{*}}\right) \end{aligned} \tag{6-7}$$

The power spectral density of $\mathbf{y}(z)$, $S_{\mathbf{y}}(z)$, can be found by multiplying the source PSD, $S_{\mathbf{x}}(z)$, on the left by the transfer function $H(z)$ and on the right by the para-Hermitian conjugate of $H(z)$. In general, $S_{\mathbf{x}}(z) \in \mathbb{C}[z]^{MxM}$, $H(z) \in \mathbb{C}[z]^{NxM}$, and $S_{\mathbf{y}}(z) \in \mathbb{C}[z]^{NxN}$. In

short, (6-7) describes what happens to the power spectral density when the signal undergoes a convolution operation.

Recovering $\mathbf{x}_i$ from $\mathbf{y}_i$ involves finding a inverse filter $H(z)^{\#}$ which when applied to $\mathbf{y}_i$ results in the decoupled signal $\mathbf{x}_i$. The problem is that the signal $\mathbf{x}_i$ is not known before hand and therefore one cannot determine if the correct $H(z)^{\#}$ was found by just observing the samples that come out of the inverse filter. Only properties of $\mathbf{x}_i$ are assumed to be known *a priori* (independent components and unit power) and in the second-order sense these properties correspond to the power spectral density of $\mathbf{x}(z)$ which is the identity matrix, I_M in (6-4). Based solely on second-order statistics (SOS), an estimate of $\mathbf{x}_i$, $\mathbf{x}_i^S$ can be found that has the same SOS as $\mathbf{x}_i$. The estimate of $\mathbf{x}_i^S$, therefore, should also have a PSD of

$$S_{\mathbf{x}^S}(z) = \sum_{i=-\infty}^{\infty} R_{\mathbf{x}^S}(k) z^{-k} = I_M \quad \Rightarrow \quad R_{\mathbf{x}^S}(k) = E\left[\mathbf{x}_i^S \mathbf{x}_{i-k}^{S*}\right] = I_M \delta(k) \tag{6-8}$$

which is similar to $S_{\mathbf{x}}(z)$ in (6-4). The logic lies in that the closer the properties of the signal $\mathbf{x}_i^S$ that comes out of the inverse filter are to the original source signal, the closer the time variations of this signal should be to $\mathbf{x}_i$. Figure 6.1 shows a block diagram where an estimate of the inverse filter using only second-order statistics, $H^S(z)^{\#}$, is applied to the signal $\mathbf{y}_i$ and produces the signal $\mathbf{x}_i^S$, which has the same SOS as $\mathbf{x}_i$.

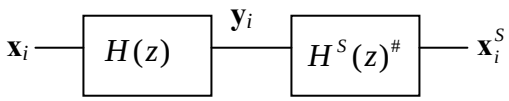

Figure 6.1 Block diagram of the identification of $H(z)$ based only on second-order statistics.

The power spectral density of $\mathbf{x}_i^S$ at the output of the system can be computed with the help of (6-7) to be

$$S_{\mathbf{x}^S}(z) = H^S(z)^{\#} S_{\mathbf{y}}(z) H^{S^*}\left(\frac{1}{z^*}\right)^{\#} = H^S(z)^{\#} H(z) S_{\mathbf{x}}(z) H^*\left(\frac{1}{z^*}\right) H^{S^*}\left(\frac{1}{z^*}\right)^{\#} \qquad (6\text{-}7)$$

Replacing the value of the PSD of $\mathbf{x}_i$ and $\mathbf{x}_i^S$ in (6-4) and (6-8) into the above equation results in the following equality

$$S_{\mathbf{x}^S}(z) = I_M = H^S(z)^{\#} H(z) H^*\left(\frac{1}{z^*}\right) H^{S^*}\left(\frac{1}{z^*}\right)^{\#} \qquad (6\text{-}10)$$

The product of $H^S(z)^{\#}$ and $H(z)$ and their respective para-Hermitian conjugates has to be equal to the identity matrix I_M in order for the properties of $\mathbf{x}_i$ and $\mathbf{x}_i^S$ to be the same in the SOS sense. Recall from Chapter 5 that any $H(z) \in \mathbb{C}[z]^{N \times M}$ can be decomposed into its minimum-phase/all-pass components in the form $H(z)=H^{mp}(z)H^{ap}(z)$. Introducing this equality into (6-10) produces

$$\begin{aligned} S_{\mathbf{x}^S}(z) = I_M &= H^S(z)^{\#} H^{mp}(z) \overbrace{H^{ap}(z) H^{ap^*}\left(\frac{1}{z^*}\right)}^{I_M} H^{mp^*}\left(\frac{1}{z^*}\right) H^{S^*}\left(\frac{1}{z^*}\right)^{\#} \\ I_M &= H^S(z)^{\#} H^{mp}(z) H^{mp^*}\left(\frac{1}{z^*}\right) H^{S^*}\left(\frac{1}{z^*}\right)^{\#} \end{aligned} \qquad (6\text{-}11)$$

where the para-unitary property of all-pass filters (Chapter 5) has been applied. There are therefore an infinite number of solutions for the filter $H^S(z)^{\#}$. From (6-10), it can be seen that if selected correctly $H^S(z)^{\#}$ can be the true inverse $H(z)$ (which is the solution that is desired) and from (6-11), another solution for $H^S(z)^{\#}$ could be the minimum-phase component of $H(z)$. In general, as can be seen from (6-11) $H^S(z)^{\#}$ can be the

inverse of $H^{mp}(z)$ times an arbitrary all-pass filter and still produce a signal $\mathbf{x}_i^S$ that has SOS that match the corresponding statistics of $\mathbf{x}_i$ in (6-4).

From the above discussion, it can be concluded that SOS identify $H(z)$ (or $H(z)^{\#}$) up to an all-pass filter $H^{ap}(z) \in \mathbb{C}[z]^{MxM}$ when the signals $\mathbf{x}_i$ are space-time independent. There is no way to know for sure if the correct inverse has been found using just SOS so it is customary to pick the minimum-phase solution from the set of all possibilities for $H^S(z)$. Recall that minimum-phase filters are causally invertible systems and BIBO stable. The minimum-phase realization can be found through an innovations based approach as explained in the next subsection.

6.3.1.1 Identification of the Minimum-Phase Components using SOS

Identification of an inverse to MIMO filter based solely on the second-order-statistics of $\mathbf{x}_i$ has an infinite number of solutions as discussed previously under the assumptions of (6-4) for $\mathbf{x}_i$. Based on SOS and for lack of other information, a minimum-phase filter is routinely picked as a solution for $H^S(z)^{\#}$ from the set of possible filters that produce a signal $\mathbf{x}_i^S$ with properties as in (6-8). The minimum-phase solution, however, implicitly assumes that the system transfer function, $H(z)$, is minimum-phase which is not true in general. The causality and stability of the filter and its inverse is the main reason for choosing the minimum-phase solution, additionally, the innovations approach is a convenient way at arriving at this solution as will be shown shortly. It should be noted that there is still a potential all-pass component of $H(z)$ that has not been identified yet when following this path. Letting the filter $H^S(z)^{\#}$ be the inverse of a minimum-phase system produces the following relationships

$$H^S(z)^{\#} = H^{mp}(z)^{\#} \quad \Leftrightarrow \quad H^S(z) = H^{mp}(z) \text{ where } H(z) = H^{mp}(z)H^{ap}(z) \qquad (6\text{-}12)$$

$$H^S(z) = \sum_{i=0}^{\infty} H_i^S z^{-i} \;,\; H^S(z)^{\#} = \sum_{i=0}^{\infty} H_i^{S\#} z^{-i} \quad \text{where } H^S(z)^{\#} H^S(z) = I_M$$

where $H^{mp}(z)$ is one of the minimum-phase decompositions of $H(z)$ (see Chapter 5). Note that the index i starts at zero since the minimum-phase solution models a causal system.

The innovations approach is a convenient method to arrive at a minimum-phase realization of the filter $H(z)$. The innovations process can be written as ([9] for example)

$$\mathbf{y}_i = -\sum_{k=1}^{\infty} D_k \mathbf{y}_{i-k} + \mathbf{i}_i \tag{6-13}$$

where the vector $\mathbf{i}_i \in \mathbb{C}^{Nx1}$ is the innovations vector. This method decomposes a signal at any given time into two components; the first contains the contributions of a weighted sum of all past values of the signal (memory) and the second term $\mathbf{i}_i$ which contains the new information that is being added at that instant that has not been seen before. Since $\mathbf{i}_i$ is new information at each time instant, it is uncorrelated with all previous values of $\mathbf{y}_i$ and all past and future values $\mathbf{i}_i$. The correlation properties of the random process $\mathbf{i}_i$ with itself and the process $\mathbf{y}_i$ can me summed up as follows

$$\begin{aligned} R_{\mathbf{i}}(k) &= E\left[\mathbf{i}_i \mathbf{i}_{i-k}^*\right] = C_N \delta(k) \\ R_{\mathbf{yi}}(-k) &= E\left[\mathbf{y}_{i-k} \mathbf{i}_i^*\right] = 0 \quad k > 0 \\ R_{\mathbf{yi}}(0) &= E\left[\mathbf{y}_i \mathbf{i}_i^*\right] = E\left[\mathbf{i}_i \mathbf{i}_i^*\right] \end{aligned} \tag{6-14}$$

which describes mathematically the relationships between both signals in the second-order sense. C_N is an arbitrary NxN matrix with complex coefficients. Also, from Figure 6.1 and (6-12) the signal $\mathbf{x}_i^S$ and $\mathbf{y}_i$ are related through the filter $H^S(z)$ and its inverse $H^S(z)^{\#}$ as follows

$$\begin{aligned} \mathbf{x}_i^S &= \sum_{k=0}^{\infty} H_k^{S\#} \mathbf{y}_{i-k} \Leftrightarrow \mathbf{x}^S(z) = H^S(z)^{\#} \mathbf{y}(z) \\ \mathbf{y}_i &= \sum_{k=0}^{\infty} H_k^S \mathbf{x}_{i-k}^S \Leftrightarrow \mathbf{y}(z) = H^S(z) \mathbf{x}^S(z) \end{aligned} \tag{6-15}$$

which is a second representation of the signal $\mathbf{y}_i$.

Note that the innovations vector $\mathbf{i}_i$ has components that are uncorrelated in time (6-14) which is one of the characteristics of $\mathbf{x}_i^S$. $\mathbf{x}_i^S$, however, has the additional property of also being uncorrelated across its components in space. To create a random process $\mathbf{x}_i^S$ that has the properties given in (6-8) from the innovations vector $\mathbf{i}_i$, requires that the elements of the vector $\mathbf{i}_i$ be also uncorrelated in space. As will be seen shortly, $\mathbf{x}_i^S$ can be obtained from $\mathbf{i}_i$ through a matrix multiplication. The relationship between $\mathbf{x}_i^S$ and $\mathbf{i}_i$ can be obtained by analyzing its cross-correlation with the vector $\mathbf{y}_i$ from (6-13) and (6-15) using the properties (6-14) it can be observed that

$$\begin{aligned} R_{\mathbf{yi}}(0) &= E\left[\mathbf{y}_i \mathbf{i}_i^*\right] = E\left[\mathbf{i}_i \mathbf{i}_i^*\right] \\ &= E\left[\sum_{k=0}^{\infty} H_k^S \mathbf{x}_{i-k}^S \mathbf{i}_i^*\right] = \sum_{k=0}^{\infty} H_k^S E\left[\mathbf{x}_{i-k}^S \mathbf{i}_i^*\right] = H_0^S E\left[\mathbf{x}_i^S \mathbf{i}_i^*\right] \end{aligned} \tag{6-16a}$$

$$\begin{aligned} \mathbf{i}_i &= H_0^S \mathbf{x}_i^S \\ \mathbf{x}_i^S &= H_0^{S\#} \mathbf{i}_i \end{aligned} \tag{6-16b}$$

The vector $\mathbf{x}_i^S$ and $\mathbf{i}_i$ are, therefore, related by the zero filter coefficient H_0^S so the time response of the filter D_i and the filter $H_i^{S\#}$ are also related by this same factor

$$H_k^{S\#} = H_0^{S\#} D_i \quad \text{where } D_0 = I_N \tag{6-17}$$

Once the filter D_i is computed (details in Chapter 7), $H_i^{S\#}$ can be determined using (6-17). H_i^S can also be found directly from $H_i^{S\#}$ and the auto-correlation of the function $\mathbf{y}_i$ as will be found next. An equation relating the impulse response H_i^S to the impulse response of its inverse filter $H_i^{S\#}$ can be found by applying the properties in (6-8) to the cross-correlation of the different signals in the system. The cross-correlation of $\mathbf{x}_i^S$ and $\mathbf{y}_i$, using (6-15) for $\mathbf{y}_i$, results in filter coefficient H_i^{S*} as can be seen from

$$E\left[\mathbf{x}_{i-k}^{S}\mathbf{y}_{i}^{*}\right]=E\left[\mathbf{x}_{i-k}^{S}\sum_{l=0}^{\infty}\mathbf{x}_{i-l}^{S*}H_{l}^{S*}\right]=\sum_{l=0}^{\infty}E\left[\mathbf{x}_{i-k}^{S}\mathbf{x}_{i-l}^{S*}\right]H_{l}^{S*}=\sum_{l=0}^{\infty}I_{M}\delta(l-k)H_{l}^{S*}$$
$$=H_{k}^{S*} \tag{6-18}$$

Using (6-13) and (6-14) a second result for the cross-correlation of $\mathbf{x}_i^S$ and $\mathbf{y}_i$ can be found as a function of the auto-correlation of $\mathbf{y}_i$ and the inverse filter $H_i^{S\#}$.

$$E\left[\mathbf{x}_{i-k}^{S}\mathbf{y}_{i}^{*}\right]=H_{0}^{S\#}E\left[\mathbf{i}_{i-k}\mathbf{y}_{i}^{*}\right]=H_{0}^{S\#}E\left[\left(\mathbf{y}_{i-k}+\sum_{l=1}^{\infty}D_{l}\mathbf{y}_{i-k-l}\right)\mathbf{y}_{i}^{*}\right]$$
$$=H_{0}^{S\#}\left[R_{\mathbf{y}}(-k)+\sum_{l=1}^{\infty}D_{l}R_{\mathbf{y}}(-k-l)\right]=H_{0}^{S\#}\left[R_{\mathbf{y}}^{*}(k)+\sum_{l=1}^{\infty}D_{l}R_{\mathbf{y}}^{*}(k+l)\right] \tag{6-19}$$
$$=H_{0}^{S\#}\left[\sum_{l=0}^{\infty}D_{l}R_{\mathbf{y}}^{*}(k+l)\right]=\sum_{l=0}^{\infty}H_{l}^{S\#}R_{\mathbf{y}}^{*}(k+l)\quad\text{where}\quad D_{0}=I_{\mathrm{N}}$$

Combining both (6-19) and (6-18) results in an equation relating both the filter, H_i^S, and its inverse, $H_i^{S\#}$, through the autocorrelations of $\mathbf{y}_i$ and this equality given by

$$H_{i}^{S*}=H_{0}^{S\#}\left[\sum_{l=0}^{\infty}D_{l}R_{\mathbf{y}}^{*}(i+l)\right]=\sum_{l=0}^{\infty}H_{l}^{S\#}R_{\mathbf{y}}^{*}(i+l)$$
$$\Updownarrow\qquad\text{where } D_{0}=I_{\mathrm{N}} \tag{6-20}$$
$$H_{i}^{S}=\left[\sum_{l=0}^{\infty}R_{\mathbf{y}}(i+l)D_{l}^{*}\right]H_{0}^{S\#*}=\sum_{l=0}^{\infty}R_{\mathbf{y}}(i+l)H_{l}^{S\#*}$$

The above equation relates the filter H_i^S to its inverse through the second-order statistics of the signals recorded at the sensors ($R_{\mathbf{y}}(k)$). The specifics of computing D_i and H_i^S are described in Chapter 7 where algorithms for computing these variable are discussed in more detail.

6.3.1.2 Innovations as a Means for Computing the Minimum-Phase Component

The above discussion describes how to determine the filter H_i^S and how it is related to its inverse $H_i^{S\#}$ using the innovations approach. What remains to be shown, is that the innovations approach arrives at a filter H_i^S that corresponds to a minimum-phase realization of the channel transfer function $H(z)$, $H^{mp}(z)$. To start with the analysis, note that the equation in (6-20) can be written as

$$H_i^S = \sum_{l=-\infty}^{\infty} R_{\mathbf{y}}(i+l) H_l^{S\#*} \quad \text{for} \quad i \geq 0 \tag{6-21}$$

and is only valid for $i \geq 0$ and not for all i which does not permit the direct application of the z-transform to the above equation. The above equation is known as the Wiener-Hopf (WH) equation and can be solved using the WH technique [10]. The approach consists of rewriting the above equation in such a way that it is valid for all i as follows

$$G_i = H_i^S - \sum_{l=-\infty}^{\infty} R_{\mathbf{y}}(i+l) H_l^{S\#*} \quad \text{for} \quad -\infty < i < \infty \tag{6-22}$$

where the signal G_i is strictly anti-causal (zero for $i \geq 0$) so that the equality in (6-21) is conserved. Since this new equation is now valid for all i, the z-transform can be directly applied to each of its terms resulting in

$$\begin{aligned}
&\sum_{i=-\infty}^{\infty} G_i z^{-i} = \sum_{i=-\infty}^{\infty} H_i^S z^{-i} - \sum_{i=-\infty}^{\infty}\sum_{l=-\infty}^{\infty} R_{\mathbf{y}}(i+l) H_l^{S\#*} z^{-i} \\
&G(z) = H^S(z) - \sum_{i=-\infty}^{\infty}\sum_{m=-\infty}^{\infty} R_{\mathbf{y}}(m) H_{m-i}^{S\#*} z^{-i} = H^S(z) - \sum_{m=-\infty}^{\infty}\sum_{p=-\infty}^{\infty} R_{\mathbf{y}}(m) H_p^{S\#*} z^{-(m-p)} \\
&G(z) = H^S(z) - \sum_{m=-\infty}^{\infty} R_{\mathbf{y}}(m) z^{-m} \sum_{p=-\infty}^{\infty} H_p^{S\#*} z^{p} \\
&G(z) = H^S(z) - S_{\mathbf{y}}(z) H^{S^*}\left(\frac{1}{z^*}\right)^{\#}
\end{aligned} \tag{6-23}$$

where $G(z)$ is the z-transform of a strictly anti-causal term, $H^S(z)$ is a causal filter, the para-Hermitian conjugate of $H^S(z)^{\#}$ is anti-causal, and $S_y(z)$ contains terms that are both causal and anti-causal. Expanding $S_y(z)$ into its minimum-phase/all-pass components (6-11) results in

$$\begin{aligned} G(z) &= H^S(z) - H(z)H^*\left(\frac{1}{z^*}\right)H^{S^*}\left(\frac{1}{z^*}\right)^{\#} \\ G(z) &= H^S(z) - H^{mp}(z)H^{ap}(z)H^{ap^*}\left(\frac{1}{z^*}\right)H^{mp^*}\left(\frac{1}{z^*}\right)H^{S^*}\left(\frac{1}{z^*}\right)^{\#} \end{aligned} \tag{6-24}$$

which shows in greater detail the components that make up the equality (6-23). The individual components can be grouped into terms that contain a strictly anti-causal, a causal, and a mixed component as follows

$$\begin{aligned} G(z)\left(H^{mp^*}\left(\frac{1}{z^*}\right)H^{S^*}\left(\frac{1}{z^*}\right)^{\#}\right)^{\#} &= H^S(z)\left(H^{mp^*}\left(\frac{1}{z^*}\right)H^{S^*}\left(\frac{1}{z^*}\right)^{\#}\right)^{\#} - H^{mp}(z) \\ G(z)H^{S^*}\left(\frac{1}{z^*}\right)H^{mp^*}\left(\frac{1}{z^*}\right)^{\#} &= H^S(z)H^{S^*}\left(\frac{1}{z^*}\right)H^{mp^*}\left(\frac{1}{z^*}\right)^{\#} - H^{mp}(z) \end{aligned} \tag{6-25}$$

$$\begin{array}{ccc} \uparrow & \uparrow & \uparrow \\ \text{strictly anti} - \text{causal} & \text{mixed} & \text{causal} \end{array}$$

Note that the para-Hermitian conjugate of a causal filter such as $H^S(z)$ is anti-causal since all its z^{-k} terms are replaced by z^k when performing this operation (the same is true with $H^{mp}(z)^{\#}$). Table 6.1 summarizes the time-domain characteristics of the individual filters that make up the above equation.

Table 6.1 Time Domain Characteristics of the Components in (6-25).

Filter	Type
$H^{S}(z), H^{S}(z)^{\#}$	causal
$H^{S^*}\left(\frac{1}{z^*}\right), H^{S^*}\left(\frac{1}{z^*}\right)^{\#}$	anti-causal
$H^{mp}(z), H^{mp}(z)^{\#}$	causal
$H^{mp^*}\left(\frac{1}{z^*}\right), H^{mp^*}\left(\frac{1}{z^*}\right)^{\#}$	anti-causal
$G(z)$	strictly anti-causal

The minimum-phase component of $H(z)$, $H^{mp}(z)$, corresponds to a causal filter and therefore can only be equal to the causal component of the mixed term in (6-25) since the other term contains strictly anti-causal terms. Mathematically this is written as a expression with a '+' symbol on the bottom [10] as follows,

$$H^{mp}(z) = \left\{ H^{S}(z) H^{S^*}\left(\frac{1}{z^*}\right) H^{mp^*}\left(\frac{1}{z^*}\right)^{\#} \right\}_{+} \tag{6-26}$$

which represents the causal component of the expression in the brackets. The two leftmost components of the above expression correspond to the para-Hermitian conjugate of $H^{mp}(z)^{\#}H^{S}(z)$ which is anti-causal since the product of two causal filters is also causal. So the result corresponds to the causal component of a multiplication of a causal and anti-causal system. Also the minimum-phase component of $H(z)$ shows up inside the brackets in its para-Hermitian conjugate form which makes $H^{mp}(z)$ a function of itself. After some careful analysis, based on the observations just stated and the fact that (6-26) is valid for

any minimum-phase realization, it can be concluded that the only possible solution for $H^{mp}(z)$ is

$$H^{mp}(z) = H^{S}(z) \tag{6-27}$$

This concludes the proof that $H^{S}(z)$ corresponds the minimum-phase component of $H(z)$ when using the innovations approach.

The identification of $H(z)$, however, is not yet complete. Recall that there still exists an all-pass component to $H(z)$ that has not been determined yet to fully identify the channel. The all-pass component, $H^{ap}(z)$, can be identified using higher-order statistics (HOS). Higher-order statistics are not blind to all-pass filters as are second-order statistics as will be explored shortly.

6.3.2 Estimation using Fourth-Order Statistics for Stationary Signals

In the previous section, a minimum-phase component of $H(z)$ was found using second order statistics. In theory, the all-pass component of $H(z)$ can be determined using any statistics higher than two for stationary signals as long as the signals under analysis have non-zero statistics of that order. In digital wireless communications, most signals of interest have even distributions (BPSK, QPSK, QAM, etc.) and therefore have odd cumulants that are zero (Chapter 3). The use of odd higher-order statistics is therefore not recommended to identify the channel for digital communication systems. The next higher even cumulant after the second-order is the fourth-order statistic which will be used to identify $H^{ap}(z) \in \mathbb{C}[z]^{MxM}$ in this dissertation.

In this section, a signal $\mathbf{x}_i^F$ (based on fourth-order statistics) will be found which is a replica of the transmitted signal $\mathbf{x}_i$. The recovered signal $\mathbf{x}_i^F$ should be similar to the original transmitted sequence up to the irresolvable ambiguities of the system or

equivalently (a permutation of its components, an unknown time delay, and an unknown scaling factor)

$$\mathbf{x}^F(z) = \Lambda(z)DP\mathbf{x}(z) \tag{6-28}$$

where $\Lambda(z)$, D, P are defined in Section 6.1. Figure 6.2 shows the last step in the identification of the transfer function $H(z)$ and the recovery of the signal $\mathbf{x}_i$.

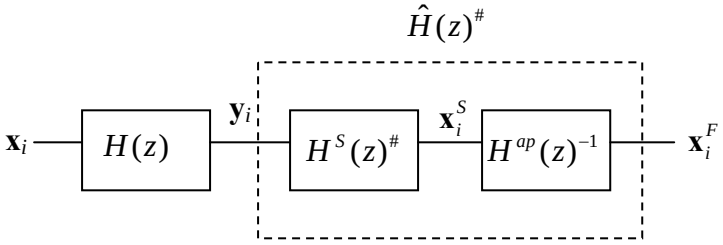

Figure 6.2. Block diagram of the identification process of $H(z)$ based on second-order statistics and then fourth-order statistics.

After second-order identification, all that is left to identify is an all-pass filter to recover an estimate of the transfer function $\hat{H}(z)$ in (6-3). From Figure 6.2 it can be seen that the signal $\mathbf{x}_i^F$ is related to $\mathbf{x}_i^S$ through the all-pass filter $H^{ap}(z)$ as follows

$$\begin{aligned} \mathbf{x}_i^S &= \sum_{k=0}^{L} H_k^{ap}\mathbf{x}_{i-k}^F \quad \Leftrightarrow \quad \mathbf{x}^S(z) = H^{ap}(z)\mathbf{x}^F(z) \\ \mathbf{x}_i^F &= \sum_{k=-L}^{0} H_{-k}^{ap\,*}\mathbf{x}_{i-k}^S \quad \Leftrightarrow \quad \mathbf{x}^F(z) = H^{ap}(z)^{-1}\mathbf{x}^S(z) = H^{ap\,*}\left(\frac{1}{z^*}\right)\mathbf{x}^S(z) \end{aligned} \tag{6-29}$$

In the second equation above, the coefficients of the inverse filter $H^{ap}(z)^{-1}$ are the time flipped and Hermitian conjugated elements of $H^{ap}(z)$ as was discussed in Chapter 5 which is a direct result of para-Hermitian conjugation. The impulse response of the all-pass filter in (6-29) is assumed to be of length L+1 and in the case of IIR filters L can even be infinite. Recall also, as can be seen from the second equation in (6-29), that the inverse

of a causal para-unitary filter is anti-causal an is not realizable unless its appropriately delayed in time. A delay of L will make the filter causal and therefore implementable for real systems, this realizable filter will be designated by

$$H^{(ap)}(z)^{-1} = z^{-L} H^{ap}(z)^{-1} = z^{-L} H^{ap^*}\left(\frac{1}{z^*}\right) \tag{6-30}$$

There are consequences, however, in making $H^{ap}(z)^{-1}$ implementable since the added delay to make this filter realizable will in turn cause a delay in the output $\mathbf{x}_i^F$ by L samples. In the case of IIR filters it would require waiting an infinite amount of time for $\mathbf{x}_i^F$ which is impractical for communication purposes. In practice, infinite impulse response filters at some point should be cut off so that they can be implemented in a finite amount of time and with the smallest truncation error.

Before proceeding with the analysis for determining the filter $H^{ap}(z)$, it will be convenient to rewrite the equality in (6-29) in matrix form as

$$\begin{aligned} \mathbf{x}_i^S &= \begin{bmatrix} H_0^{ap} & \cdots & H_L^{ap} \end{bmatrix} \begin{bmatrix} \mathbf{x}_i^F \\ \vdots \\ \mathbf{x}_{i-L}^F \end{bmatrix} = \tilde{H}^{ap} \tilde{\mathbf{x}}_i^F \\ \mathbf{x}_i^F &= \begin{bmatrix} H_L^{ap^*} & \cdots & H_0^{ap^*} \end{bmatrix} \begin{bmatrix} \mathbf{x}_{i+L}^S \\ \vdots \\ \mathbf{x}_i^S \end{bmatrix} = \tilde{H}^{\overline{ap}} \tilde{\mathbf{x}}_i^S \end{aligned} \tag{6-31}$$

where $\tilde{H}^{ap} \in \mathbb{C}^{\mathrm{MxM(L+1)}}$ ($\tilde{H}^{\overline{ap}} \in \mathbb{C}^{\mathrm{MxM(L+1)}}$) contains the concatenated impulse response of the L+1 size all-pass filter (inverse filter), $H^{ap}(z)$ ($H^{ap}(z)^{-1}$), and the vector $\tilde{\mathbf{x}}_i^F$ ($\tilde{\mathbf{x}}_i^S$) is a stacking of the random variable $\mathbf{x}_i^F$ ($\mathbf{x}_i^S$) form *i* to *i*-L (*i*+L to *i*).

Recall that in order for $\mathbf{x}_i^F$ to be more like the original transmitted sequence than $\mathbf{x}_i^S$ it should be more than space-time uncorrelated, it should have the same properties as the

signal $\mathbf{x}_i$ which is independence of its components in time and in space. From Chapter 3, independence in space means that $\mathbf{x}_i^F$'s fourth-order cumulant tensor should be diagonal. In order for $\mathbf{x}_i^F$ to be also independent in time, the fourth-order cumulant tensor should be zero when any of its components is time shifted with respect to another component. Therefore, if the inverse of the all-pass filter $H^{ap}(z)^{-1}$ is properly found, the two properties mentioned above will have been met.

The synthesized random variable $\tilde{\mathbf{x}}_i^S$ which contains the values of the signal $\mathbf{x}_i^S$ from times i+L to i has components that are related by the filter $H^{ap}(z)$. The fourth-order cumulant of $\tilde{\mathbf{x}}_i^S$ is therefore, in general, not diagonal. A special feature of the concatenated vector $\tilde{\mathbf{x}}_i^S$ is that it combines the space and time dependence of $\mathbf{x}_i^S$ into a single cumulant tensor which eases the analysis. The fourth-order cumulant tensor of $\tilde{\mathbf{x}}_i^S$ has elements indexed by *pqrs* which are defined as

$$\left(\boldsymbol{C}_4\left[\tilde{\mathbf{x}}^S\right]\right)_{pqrs} = C\left[\left(\tilde{\mathbf{x}}_i^S\right)_p \left(\tilde{\mathbf{x}}_i^S\right)_q^* \left(\tilde{\mathbf{x}}_i^S\right)_r^* \left(\tilde{\mathbf{x}}_i^S\right)_s\right] \quad p,q,r,s \in \{1,2,...,M(L+1)\} \qquad (6\text{-}32)$$

where $(\tilde{\mathbf{x}}_i^S)_p$ corresponds to the pth element of the vector $\tilde{\mathbf{x}}_i^S$ and $C[\bullet]$ is the cumulant defined as in Chapter 3.

The signal $\mathbf{x}_i^F$ is recovered by multiplying $\tilde{\mathbf{x}}_i^S$ on the left by the matrix $\tilde{H}^{\overline{ap}}$ which contains the impulse response of $H^{ap}(z)^{-1}$ in concatenated form (6-31). Since cumulants obey the multi-linearity property (Chapter 3), the elements of the fourth-order cumulant tensor of $\mathbf{x}_i^F$ can be found directly from $\tilde{H}^{\overline{ap}}$ and the fourth-order cumulant tensor $\tilde{\mathbf{x}}_i^S$ as follows

$$\left(\boldsymbol{C}_4\left[\mathbf{x}^F\right]\right)_{ijkl} = \sum_{p,q,r,s=1}^{M(L+1)} \left(\boldsymbol{C}_4\left[\tilde{\mathbf{x}}^S\right]\right)_{pqrs} \left(\tilde{H}^{\overline{ap}}\right)_{ip} \left(\tilde{H}^{\overline{ap}}\right)_{jq}^* \left(\tilde{H}^{\overline{ap}}\right)_{kr}^* \left(\tilde{H}^{\overline{ap}}\right)_{ls} \quad i,j,k,l \in \{1,...,M\} \quad (6\text{-}33)$$

The values of $\mathbf{x}_i$ are assumed to have components that are space-time independent which is also a property that has to be satisfied by the signal $\mathbf{x}_i^F$. The fourth order space-time cumulant of $\mathbf{x}_i^F$ (Chapter 3) must therefore be

$$\left(\boldsymbol{C}_4\left[\mathbf{x}^F,\boldsymbol{\tau}\right]\right)_{ijkl} = C\left[\left(\mathbf{x}^F_{i-\tau_1}\right)_i\left(\mathbf{x}^F_{i-\tau_2}\right)^*_j\left(\mathbf{x}^F_{i-\tau_3}\right)^*_k\left(\mathbf{x}^F_{i-\tau_4}\right)_l\right] = \begin{Bmatrix} c_i\delta_{\boldsymbol{\tau}} & \text{for } i=j=k=l \\ 0 & \text{otherwise} \end{Bmatrix} \quad (6\text{-}34)$$

$$\delta_{\boldsymbol{\tau}} = \begin{cases} 1 & \text{for } \tau_1=\tau_2=\tau_3=\tau_4 \\ 0 & \text{otherwise} \end{cases}$$

in order for $\mathbf{x}_i^F$ to be related to $\mathbf{x}_i$ through (6-28) where c_i is a real scalar constant. A generalized form of the filter $H^{ap}(z)^{-1}$ can be constructed using the information presented in the previous chapter. Any para-unitary filter $\tilde{\tilde{H}}^{\overline{ap}}$ as described in (6-31) that leads to a diagonal cumulant that exists for only equal time delays (6-34) will correspond to the all-pass component of the estimated filter $\hat{H}(z)$ in (6-3). Note that there are several such solutions but they should all identify the channel up to the irresolvable ambiguities described in Section 6.1.

The all-pass filter can be solved for in an efficient manner using a contrast function. Contrasts were first introduced in Chapter 4 to solve for unknown unitary matrices and their functionality can be extended so that they also apply for para-unitary filters. It can be shown that properties of all-pass filters can be taken advantage of and can be used to create an appropriate cost function. The details will be discussed in the next chapter.

Until now, the number of signals in the environment were assumed to be less than or equal to the number of sensors (N≥M). This corresponds to having at least the same number of unknowns as equations in a linear system. Under these conditions, the system is invertible (assuming certain rank criteria) and the source signals can be recovered perfectly. When there are more signals than sensors (more unknowns than equations) in the environment, the time domain signals $\mathbf{x}_i$ are in general irrecoverable but properties of each of the source signals are recoverable under some circumstances. If $\mathbf{x}_i$ has certain

characteristics it is possible to identify properties of the signals when there are more signals than sensors even when an $H(z)^{\#}$ that will decouple the time series of the source signals does not exist. Sophisticated receivers such as MLSE are able to estimate the source signals from the knowledge of the modulation (BPSK, QPSK, etc.) property of the elements of $\mathbf{x}_i$ even in the M>N case as discussed in Chapter 2. The following section examines signals that have modulation induced cyclostationary statistics and it will be shown that properties of more signals than sensors can be identified in the system.

6.4 Blind Identification of Cyclostationarity Systems

In order to transmit information from the transmitter to the receiver, the signal is structured in a certain manner based on a pre-determined modulation scheme. In digital communication systems, the transmitted signal consists of a series of symbols which when put together represent the information that is being conveyed. The symbols are taken from a set with finite alphabet that is transferred at a pre-determined constant rate which corresponds to the data rate of the signal. Additionally, the signal is up-converted from baseband to its carrier frequency so that it can be transmitted through an antenna and in order to place it in the correct frequency band of operation. Code division multiple-axis (CDMA) communications systems have an additional pseudo-random code that is modulated on top of the signal which is also periodic with spreading code length. All these artificially induced periodicities induce statistics that are also periodic and are a function of carrier frequency, data rate, code rate, and others periodicities that are imposed on the signal. Information about the structure of the signals is contained at each of its cyclical statistics which offers redundancy and a means of distinguishing signals in the statistical space based on their periodicities. The ability to be able to separate the statistics by frequency allows for the identification of more signals than sensors (M>N). In this dissertation, linearly modulated signals (BPSK, QPSK, QAM, etc.) will be treated exclusively to demonstrate the functionality of the analysis. In this section, the assumptions imposed in the previous section that N≥M will be relaxed, M can now be larger than N.

Modulated signals are not stationary signals but instead have time-varying autocorrelation functions and time-varying higher-order statistics. In fact, they belong to a special class of time-varying systems denoted as periodically-time varying (PTV) systems. The impulse response of PTV systems can be decomposed into time-invariant filters and a denumerated set of associated periodicities which completely characterize the system response. These periodicities and time-invariant filters also show up in the cyclic statistics of the signal as will be seen shortly. PTV systems will not be treated in detail here but are examined in [7]. Linearly modulated systems (modulation and channel) can be written as a time varying impulse response [7]

$$y(t) = \int_{-\infty}^{\infty} h(t,s)x(s)ds \tag{6-35}$$

$$h(t,s) = \sum_{k} q_{\alpha_k}(t-s)e^{j2\pi\alpha_k t}$$

where α_k correspond to the cyclic frequencies of the modulation process and $q_{\alpha_k}(\tau)$ are the time-invariant filters corresponding to each cyclic frequency. A stationary signal x(s) is transformed into a cyclostationary signal through the above operation. Similarly, the autocorrelation function and higher order statistics of $y(t)$ will also be cyclic occurring at the discrete frequencies that are a function of the modulation induced frequencies [7].

As was discussed in Chapter 3, the second and fourth-order cumulant tensors of a cyclostationary vector signal $\mathbf{y}(t)$ can be decomposed into the summation of a denumerated set of cyclic cumulant tensors (3-31),(3-33). This chapter is primarily concerned with the analysis of sampled data instead of continuous time waveforms for which the time-varying cumulants were derived in Chapter 3. Fortunately, the results in Chapter 3 can be extended to discrete signals with only some minor alterations. The main one being replacing integrations in the moment functions by summations. The second-order cumulant tensor of a discrete random variable $\mathbf{x}_t\,(t, \tau_p \in \mathbb{Z})$can be expressed as

$$\boldsymbol{C}_2[\mathbf{x};t,\boldsymbol{\tau}] = \sum_{\alpha} \boldsymbol{C}_2^{\alpha}[\mathbf{x};\boldsymbol{\tau}] e^{j2\pi\alpha t} \tag{6-36a}$$

$$\left(\boldsymbol{C}_2^{\alpha}[\mathbf{x};\boldsymbol{\tau}]\right)_{ij} = M_{ij^*}^{\alpha}[\mathbf{x};\boldsymbol{\tau}] \tag{6-36b}$$

$$M_{ij^*}^{\alpha}[\mathbf{x};\boldsymbol{\tau}] = \lim_{T\to\infty} \frac{1}{2T+1} \sum_{t=-T}^{T} \left(\mathbf{x}_{t-\tau_1}\right)_i \left(\mathbf{x}_{t-\tau_2}\right)_j^* e^{-j2\pi\alpha i} \tag{6-37}$$

were t and τ_p are now discrete variables belonging to the set of integers. Note that the subscript t is now used instead of i for the time index; this distinguishes stationary (index i) from cyclostationary signals (index t). Equivalently, the fourth-order cumulant tensor for a discrete time random vector signal $\mathbf{x}_t$ can be written as

$$\boldsymbol{C}_4[\mathbf{x};t,\boldsymbol{\tau}] = \sum_{\alpha} \boldsymbol{C}_4^{\alpha}[\mathbf{x};\boldsymbol{\tau}] e^{j2\pi\alpha t} \tag{6-38}$$

$$\left(\boldsymbol{C}_4^{\alpha}[\mathbf{x};\boldsymbol{\tau}]\right)_{ijkl} = \lim_{T\to\infty} \frac{1}{T} \sum_{t=-T}^{T} \left(\boldsymbol{C}_4[\mathbf{x};t,\boldsymbol{\tau}]\right)_{ijkl} e^{-j2\pi\alpha t}$$

$$\left(\boldsymbol{C}_4^{\alpha}[\mathbf{x};\boldsymbol{\tau}]\right)_{ijkl} = M_{ij^*k^*l}^{\alpha}[\mathbf{x};\boldsymbol{\tau}] - \sum_{\beta} M_{ij^*}^{\beta}[\mathbf{x};\boldsymbol{\tau}] M_{k^*l}^{\alpha-\beta}[\mathbf{x};\boldsymbol{\tau}]$$
$$- \sum_{\beta} M_{ik^*}^{\beta}[\mathbf{x};\boldsymbol{\tau}] M_{j^*l}^{\alpha-\beta}[\mathbf{x};\boldsymbol{\tau}] - \sum_{\beta} M_{il}^{\beta}[\mathbf{x};\boldsymbol{\tau}] M_{j^*k^*}^{\alpha-\beta}[\mathbf{x};\boldsymbol{\tau}]$$

The moment functions of the above equation can be computed for discrete signals by replacing the integration operation in (3-35) by a summation as was done with (6-37). Additionally, the cyclic frequencies can be constrained to belong to the set $\alpha\in$(-1/2,1/2] because of the normalization and the discreteness of the time axis. Note that α is periodic with period 1 in the cycle domain.

In actuality equation (6-38) should be written in terms of k-statistics to ensure an un-biased estimator (Chapter 3). In order to make the derivations more transparent, however, the above form will be chosen and a large sample size T will be assumed so that the above results will be virtually the same as the result achieved through k-statistics.

When analyzing communication systems, it is common to model the signals at baseband in order to simplify the analysis. Doing this, however, does not take away from the applicability or generality of the analysis. The resulting signal at baseband can always be moved to a carrier frequency by multiplying the result by a complex exponential which has no information content. Linearly modulated signals will be exclusively treated in this chapter to illustrate the effectiveness of cyclostationary statistics based identification. The information carrying signal $\mathbf{x}_t$, has elements which are composed of symbols that belong to a finite set (QPSK, BPSK, QAM, etc.) that are transmitted at a predetermined data rate. The signal $\mathbf{x}_t$ can be thought of being created by a source signal $\mathbf{s}_t$ as shown in Figure 6.3 which has statistics that are stationary up to at least the 4^{th} order. Each of the samples of $\mathbf{s}_t$ contains the information symbols of all the transmitters that are going to be sent over the channel which are independent in time and in space. If the signals were not independent in time, the transmitter would be transmitting redundant information which would reduce the effective data rate of transmission. Inter-dependence of symbols in time is usually removed at the source coding stage of the transmitter. The signal $\mathbf{x}_t$ is created by up-sampling the signal $\mathbf{s}_t$ to the required data rate through T_m which might be different for each of the transmitters. Up-conversion involves filling in with zeros in between two consecutive samples of an element of $\mathbf{s}_t$ until the required data rate is achieved. This operation transforms the stationary signal $\mathbf{s}_t$ into a cylostationary signal $\mathbf{x}_t$ whose elements have statistical periodicities that are equal to its data rate and multiples of its data rate as will be seen shortly.

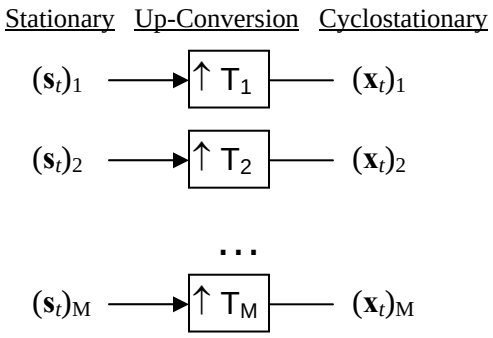

Figure 6.3. Block diagram showing the signals $\mathbf{s}_t$ producing the cyclostationary signal $\mathbf{x}_t$.

For an individual elements of $\mathbf{x}_t$, $(\mathbf{x}_t)_i$, the up-sampling operator can be written as a discrete summation over the symbols of the corresponding element in $\mathbf{s}_t$

$$\left(\mathbf{x}_t\right)_i = \sum_m \left(\mathbf{s}_m\right)_i \delta_{t-mT_i-T_{0,i}} \quad \delta_t = \begin{cases} 1 & \text{when } t = 0 \\ 0 & \text{otherwise} \end{cases} \tag{6-39}$$

where T_i is the reciprocal of the data rate of the transmitter *i* and its symbols are shifted by a delay $T_{0,i}$ with respect to the zero time axis. Because of the simple form of $\mathbf{x}_t$, the statistics of such a time sequence can be computed in a fairly straight forward manner. The time-varying second-order cumulant for just *one* element of $\mathbf{x}_t$ can be readily computed as follows

$$\begin{aligned} C_2\left[\left(\mathbf{x}_t\right)_i; t, \boldsymbol{\tau}\right] &= E\left[\left(\mathbf{x}_{t-\tau_1}\right)_i \left(\mathbf{x}_{t-\tau_1}\right)_i^*\right] = E\left[\left(\mathbf{s}_t\right)_i \left(\mathbf{s}_t\right)_i^*\right] \delta_{\boldsymbol{\tau}} \sum_m \delta_{t-mT_i-T_{0,i}-\tau_1} \\ &= M_{ii^*}[\mathbf{s}] \delta_{\boldsymbol{\tau}} \sum_m \delta_{t-mT_i-T_{0,i}-\tau_1} \\ &, \delta_{\boldsymbol{\tau}} = \begin{cases} 1 & \text{for } \tau_1 = \tau_2 \\ 0 & \text{otherwise} \end{cases} \end{aligned} \tag{6-40}$$

where

$$M_{ii^*}[\mathbf{s}] = \lim_{T\to\infty} \frac{1}{2T+1} \sum_{t=-T}^{t=T} \left(\mathbf{s}_t\right)_i \left(\mathbf{s}_t\right)_i^* \tag{6-41}$$

Note that because $\mathbf{s}_t$ has samples which are independent in time, $C_2\left[\left(\mathbf{x}_t\right)_i; t, \boldsymbol{\tau}\right]$ will be non-zero only when $\tau_1 = \tau_2$. Also, since $\mathbf{s}_t$ is stationary, the second moment of its element *i* is independent of *t* and non-zero only when both delay components of $\boldsymbol{\tau}$ are equal hence the notation $M_{ii^*}[\mathbf{s}]$. The method presented in this dissertation uses the redundancy in the cyclic statistics to identify more signals than sensors in the system so these statistics need

to be computed first from (6-40). The second order cyclic cumulant function of the element *i* of $\mathbf{x}_t$ can be found from (6-40) and (6-36,37) in a fairly straight forward fashion

$$\boldsymbol{C}_2^{\alpha^i}\left[(\mathbf{x}_t)_i;\boldsymbol{\tau}\right]\begin{cases}=\dfrac{1}{T_i}M_{ii^*}[\mathbf{s}]\delta_{\boldsymbol{\tau}}e^{-j2\pi\alpha^i(T_{0,i}+\tau_1)} & \text{for } \alpha^i=\dfrac{k}{T_i},k\in\mathbb{Z} \text{ for } \alpha^i\in\left(-\dfrac{1}{2},\dfrac{1}{2}\right]\\ =\dfrac{1}{T_i}\delta_{\boldsymbol{\tau}}\boldsymbol{C}_2[(\mathbf{s})_i]e^{-j2\pi\alpha^i(T_{0,i}+\tau_1)} & \text{for } \alpha^i=\dfrac{k}{T_i},k\in\mathbb{Z} \text{ for } \alpha^i\in\left(-\dfrac{1}{2},\dfrac{1}{2}\right]\\ =0 \quad \text{For all other } \alpha^i\end{cases} \tag{6-42}$$

Note that the above cyclic cumulant is non-zero only for an integer multiple of its data rate. The redundancy in the cyclic statistic is evident; the second-order cumulant of the element *i* of $\mathbf{s}_t$, $\boldsymbol{C}_2[(\mathbf{s})_i]$, is present at each cyclic frequency α^i of the i^{th} element of $\mathbf{x}_t$, $\boldsymbol{C}_2^{\alpha^i}\left[(\mathbf{x}_t)_i;\boldsymbol{\tau}\right]$. Additionally, the symbol timing information can be recovered for $(\mathbf{x}_t)_i$ when τ_1 and τ_2 are set to zero at any of the cyclic frequencies except $\alpha^i=0$. The phase of the cyclic cumulant is a function of $T_{0,i}$. Summarizing, the second-order statistics of $\mathbf{s}_t$ can be extracted from any of the cyclic frequencies of $\boldsymbol{C}_2^{\alpha^i}\left[(\mathbf{x}_t)_i;\boldsymbol{\tau}\right]$, and symbol timing information ($T_{0,i}$) is also recoverable from the phase for all $\alpha^i\neq 0$.

The fourth-order time-varying cumulant can be computed in a similar fashion using (6-38). This cumulant for the i^{th} element of $\mathbf{x}_t$ can be found to be

$$\begin{aligned}\boldsymbol{C}_4\left[(\mathbf{x}_t)_i;t,\boldsymbol{\tau}\right]=&M_{ii^*i^*i}[\mathbf{s}]\delta_{\boldsymbol{\tau}}\sum_m\delta_{t-mT_i-T_{0,i}-\tau_1}-M_{ii^*}[\mathbf{s}]M_{i^*i}[\mathbf{s}]\delta_{\boldsymbol{\tau}}\sum_m\delta_{t-mT_i-T_{0,i}-\tau_1}\\&-M_{ii^*}[\mathbf{s}]M_{i^*i}[\mathbf{s}]\delta_{\boldsymbol{\tau}}\sum_m\delta_{t-mT_i-T_{0,i}-\tau_1}-M_{ii}[\mathbf{s}]M_{i^*i^*}[\mathbf{s}]\delta_{\boldsymbol{\tau}}\sum_m\delta_{t-mT_i-T_{0,i}-\tau_1}\end{aligned} \tag{6-43}$$

where $\delta_{\boldsymbol{\tau}}$ represents

$$\delta_{\boldsymbol{\tau}}=\begin{cases}1 & \text{for } \tau_1=\tau_2=\tau_3=\tau_4\\ 0 & \text{otherwise}\end{cases} \tag{6-44}$$

and the moments of $\mathbf{s}_t$ are defined as follows

$$M_{ii^*i^*i}[\mathbf{s}] = \lim_{T\to\infty} \frac{1}{2T+1} \sum_{t=-T}^{t=T} (\mathbf{s}_t)_i (\mathbf{s}_t)_i^* (\mathbf{s}_t)_i^* (\mathbf{s}_t)_i \tag{6-45}$$

$$M_{i^{(*)}i^{(*)}}[\mathbf{s}] = \lim_{T\to\infty} \frac{1}{2T+1} \sum_{t=-T}^{t=T} (\mathbf{s}_t)_i^{(*)} (\mathbf{s}_t)_i^{(*)}$$

The symbol (*) denotes the presence or absence of a conjugation such that the conjugation in the subscripts of M matches the corresponding conjugation of $\mathbf{s}_t$ in the summation, see (6-41) for an example. The fourth-order cyclic cumulant can be computed by applying (6-38) to (6-43) and the result is given by

$$C_4^{\alpha}\left[(\mathbf{x}_t)_i;\boldsymbol{\tau}\right] \begin{cases} = \frac{1}{T_i}\delta_{\boldsymbol{\tau}} e^{-j2\pi\alpha^i(T_{0,i}+\tau_1)} \begin{pmatrix} M_{ii^*i^*i}[\mathbf{s}] - M_{ii^*}[\mathbf{s}]M_{i^*i}[\mathbf{s}] \\ -M_{ii^*}[\mathbf{s}]M_{i^*i}[\mathbf{s}] - M_{ii}[\mathbf{s}]M_{i^*i^*}[\mathbf{s}] \end{pmatrix} \\ = \frac{1}{T_i}\delta_{\boldsymbol{\tau}} C_4[(\mathbf{s})_i] e^{-j2\pi\alpha^i(T_{0,i}+\tau_1)} \quad \text{for } \alpha^i = \frac{k}{T_i}, k\in\mathbb{Z} \text{ for } \alpha^i \in \left(-\frac{1}{2},\frac{1}{2}\right] \\ = 0 \qquad \text{For all other } \alpha^i \end{cases} \tag{6-46}$$

Again, the cyclic cumulant is non-zero only for integer multiples of the data rate. As with the second-order cumulant, each of the cyclic frequencies of the fourth order cumulant above contains the fourth-order cumulant of $(\mathbf{s}_t)_i$ and symbol timing can also be recovered from any α^j not equal to zero.

The above results where computed by looking at just one of the elements of $\mathbf{x}_t$ for illustration purposes. In the general sense, the joint second-order cyclic cumulant tensor of all elements of $\mathbf{x}_t$ at a frequency α should be examined (denoted by $C_2^{\alpha}[\mathbf{x}_t;\boldsymbol{\tau}]$ and is of size MxM). $C_2^{\alpha}[\mathbf{x}_t;\boldsymbol{\tau}]$ will pick up only the elements of $\mathbf{x}_t$ whose cumulants have that particular cycle frequency α and ignore the rest (the signals with no statistics at α will be

zero). This type of analysis is a form of selection based on a particular feature of the signal making all signals that do not possess statistics at α seem like they are not present in the system. Similar conclusions can be made about the fourth-order cumulant tensor but this topic will be addressed in more detail later. The advantages of signal selectivity based on cyclic statistics allows for the identification of more signals than sensors because it effectively allows for working with a subset of the signals at a time instead of all of them.

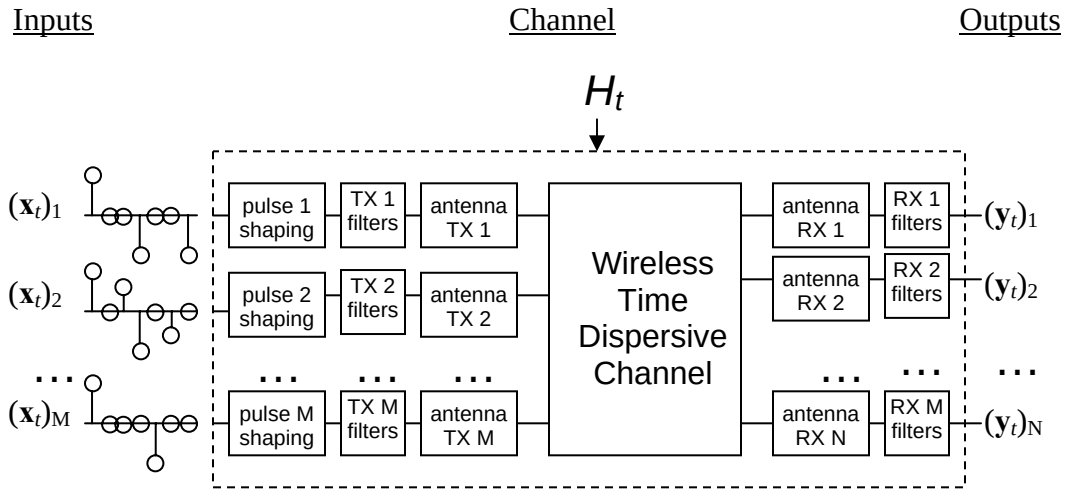

Figure 6.4 Block diagram of baseband model of linearly modulated MIMO system.

To illustrate the effectiveness of using cyclic cumulant tensors for system identification, suppose all M signals in the system have the same center frequency (at baseband) which is a worse-case scenario. If one or more of the signals had a slightly different carrier frequency, there are more distinguishing features between them. These differences make it easier to differentiate one signal from another. As mentioned earlier, linear modulated signals will be exclusively treated in this dissertation to demonstrate the effectiveness of these methods. The block diagram for a linearly modulated communication system, with all signals at baseband, is depicted in Figure 6.4. Each of the elements of $\mathbf{x}_t$ (transmitters) are first pulse shaped, then passed through a series of possible transmit filters and then delivered to the channel through the transmit antenna which also could filter the signal.

The wireless channel has the effect of mixing, scaling, and time dispersing all the signals in the system (as in Figure 5.1). At the receiver, the signal arrives at the receiving antenna and then undergoes some possible filtering after which the sampled signal $\mathbf{y}_t$ is made available for processing. All these filtering operations are assumed to be linear and time-invariant and can be combined into a single effective channel response H_t relating the source signal $\mathbf{x}_t$ to the received signal $\mathbf{y}_t$.

The complete system response can be written as a matrix convolution

$$\mathbf{y}_t = \sum_{k=0}^{K} H_k \mathbf{x}_{t-k} \quad \Leftrightarrow \quad \mathbf{y}(z) = H(z)\mathbf{x}(z) \tag{6-47}$$

or in matrix form

$$\mathbf{y}_t = \begin{bmatrix} H_0 & \cdots & H_K \end{bmatrix} \begin{bmatrix} \mathbf{x}_t \\ \vdots \\ \mathbf{x}_{t-K} \end{bmatrix} = \tilde{H}\tilde{\mathbf{x}}_t \quad \text{where } \tilde{H} = \begin{bmatrix} H_0 & \cdots & H_K \end{bmatrix}, \tilde{\mathbf{x}}_t = \begin{bmatrix} \mathbf{x}_t \\ \vdots \\ \mathbf{x}_{t-K} \end{bmatrix} \tag{6-48}$$

which is similar to (6-1), but the difference in these two systems lies in that fact that $\mathbf{x}_t$ is now a cyclostationary signal instead of just having stationary statistics as previously assumed. This is an important difference in these two analyses as will be seen shortly.

Because of the multi-linearity property of cumulant tensors (Chapter 3), cyclic cumulants contain parts of the channel information at each of the cyclic frequencies α in addition to the statistics of $\mathbf{s}_t$. Recall that stationary signals can be thought of as cyclostationary signals with a single cyclic frequency at α=0. So the methods discussed in the previous section for stationary signals (α=0) can be extended to all other cycle frequencies with some small modifications. Some of the details are omitted because much of the analysis at each cycle frequency follows what was done previously with the stationary case. The problem of identification can now be divided into several smaller problems. Analysis can

be performed on a subset of the signals that have a particular cylic statistic at a time. The second-order time varying cyclic cumulant function of $\mathbf{y}_t$ can be found from (6-48) to be

$$\boldsymbol{C}_2^{\alpha}\left[\mathbf{y}_t;\boldsymbol{\tau}\right]=\begin{bmatrix}H_0 & \cdots & H_K\end{bmatrix}\boldsymbol{C}_2^{\alpha}\left[\tilde{\mathbf{x}}_t;\boldsymbol{\tau}\right]\begin{bmatrix}H_0^*\\ \vdots \\ H_K^*\end{bmatrix} \tag{6-49}$$

Since $\mathbf{x}_t$ is space-time independent, the second-order cyclic cumulant tensor of $\tilde{\mathbf{x}}_t$, $\boldsymbol{C}_2^{\alpha}\left[\tilde{\mathbf{x}}_t;\boldsymbol{\tau}\right]$, is diagonal and takes on the form

$$\boldsymbol{C}_2^{\alpha}\left[\tilde{\mathbf{x}}_t;\boldsymbol{\tau}\right]=\begin{bmatrix}\boldsymbol{C}_2^{\alpha}\left[\mathbf{x}_t;\boldsymbol{\tau}\right] & 0 & 0\\ 0 & \ddots & 0\\ 0 & 0 & \boldsymbol{C}_2^{\alpha}\left[\mathbf{x}_{t-K};\boldsymbol{\tau}\right]\end{bmatrix} \tag{6-50}$$

where $\boldsymbol{C}_2^{\alpha}\left[\mathbf{x}_{t-k};\boldsymbol{\tau}\right]$ are MxM matrices. The $\boldsymbol{C}_2^{\alpha}\left[\mathbf{x}_{t-k};\boldsymbol{\tau}\right]$ matrices are further made up of the statistics of the elements of $\mathbf{x}_t$

$$\boldsymbol{C}_2^{\alpha}\left[\mathbf{x}_{t-k};\boldsymbol{\tau}\right]=\begin{bmatrix}\boldsymbol{C}_2^{\alpha}\left[\left(\mathbf{x}_{t-k}\right)_1;\boldsymbol{\tau}\right] & 0 & 0\\ 0 & \ddots & 0\\ 0 & 0 & \boldsymbol{C}_2^{\alpha}\left[\left(\mathbf{x}_{t-k}\right)_M;\boldsymbol{\tau}\right]\end{bmatrix} \tag{6-51}$$

where $\boldsymbol{C}_2^{\alpha}\left[\left(\mathbf{x}_{t-k}\right)_m;\boldsymbol{\tau}\right]$ are one-dimensional and are non-zero only when the corresponding m^{th} signal in $\mathbf{x}_t$ has a statistic that is non-zero at a cycle frequency α. The cyclic cumulant $\boldsymbol{C}_2^{\alpha}\left[\tilde{\mathbf{x}}_t;\boldsymbol{\tau}\right]$ is therefore mostly a zero matrix except for a few elements in the diagonal that correspond to the transmitters that have statistics at the cycle frequency α. Because of the sparseness in the second order cumulant tensor of $\tilde{\mathbf{x}}_t$, $\boldsymbol{C}_2^{\alpha}\left[\tilde{\mathbf{x}}_t;\boldsymbol{\tau}\right]$, the cyclic statistic $\boldsymbol{C}_2^{\alpha}\left[\mathbf{y}_t;\boldsymbol{\tau}\right]$ has an effect of picking out only the parts of the filter H_k in (6-49) that correspond to that particular subset of transmitters that have non-zero statistics for α.

This property is what will be exploited to identify more signals than sensor with their corresponding channels in this dissertation.

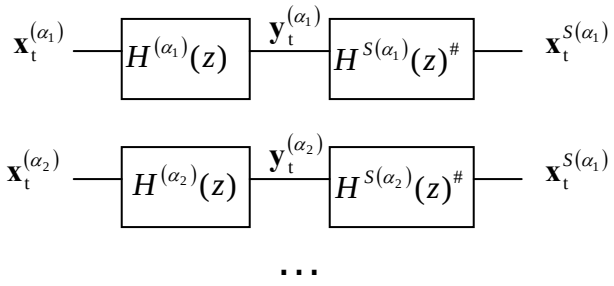

Figure 6.5 Block diagram of the identification of $H^{(\alpha)}$(z) based on cyclic second-order statistics.

Figure 6.5 summarizes the procedure for a method which exploits cyclostationary statistics that can be thought of as an extension of what was done with stationary signals. First of all, a large number of signals and their corresponding channels can be identified using this method, the limit lies in that for a given cyclic statistic α a maximum number of N (number of sensors) signals can be identified with their channels. As long as there exists a cycle frequency at which all linearly modulated signals are available with no more than N overlapping signal statistics, the system is fully identifiable up to the ambiguities of the system. Figure 6.5 describes the procedure where the signals $\mathbf{x}_t$ and $\mathbf{y}_t$ have been separated into smaller subsets (hopefully not larger than N to be identifiable) denoted by $\mathbf{x}_t^{(\alpha)}$ and $\mathbf{y}_t^{(\alpha)}$ based on their statistical cycle frequencies. The procedure then follows a similar path to what was done with stationary signals which involves first identifying the minimum-phase component of the subset transfer function $H^{(\alpha)}$(z) for each group using second-order cyclic statistics. For signals that have a statistic at α (denoted by the superscript α) the equations based solely on second order statistics (denoted by the superscript *S*) can be written as

$$\mathbf{x}_t^{S(\alpha)} = \sum_{k=0}^{\infty} H_k^{S\#(\alpha)} \mathbf{y}_{t-k}^{(\alpha)} \Leftrightarrow \mathbf{x}^{S(\alpha)}(z) = H^{S(\alpha)}(z)^{\#} \mathbf{y}^{(\alpha)}(z)$$
$$\mathbf{y}_t^{(\alpha)} = \sum_{k=0}^{\infty} H_k^{S(\alpha)} \mathbf{x}_{t-k}^{S(\alpha)} \Leftrightarrow \mathbf{y}^{(\alpha)}(z) = H^{S(\alpha)}(z) \mathbf{x}^{S(\alpha)}(z) \quad (6\text{-}52)$$

As with the stationary signals, there are many possible solutions for $H^{S(\alpha)}(z)$ if the only conditions imposed are that the second-order statistics of $\mathbf{x}_t^{(\alpha)}$ match the second-order statistics of $\mathbf{x}_t^{S(\alpha)}$. $H^{S(\alpha)}(z)$ and its inverse $H^{S(\alpha)}(z)^{\#}$, when computed using an innovations type approach corresponds to a minimum-phase component of the $H^{(\alpha)}(z)$ as explained in the previous section. The details of computing the filters based on innovations for cyclostationary signals are left to the next chapter which examines the algorithmic methods to compute the various filter coefficients of $H_k^{S(\alpha)}$.

In order to resolve the ambiguity in power (Section 6.1) the elements of $\mathbf{x}_t$ have been assumed to be transmitting with unity power as was done throughout this dissertation. Note that all second order cyclic components of the element $\left(\mathbf{x}_t^{(\alpha)}\right)_i$ should therefore have a magnitude of unity and so should their respective $\left(\mathbf{x}_t^{S(\alpha)}\right)_i$, as found in (6-40), if their second-order statistics are made to match. For linearly modulated signals at baseband, it is only necessary to have access to the statistics at a single cyclic frequency $\alpha \neq 0$ to determine all remaining statistics of the same order at different cycle frequencies. For the linearly modulated case studied here, this means at integer multiples of the data rate, $1/T_i$, the second-order statistics of the i^{th} element of the signal $\mathbf{x}_t^{S(\alpha)}$ should be of magnitude 1 and phase dependent on α and $T_{0,i}$ and $\boldsymbol{\tau}$ (6-40).

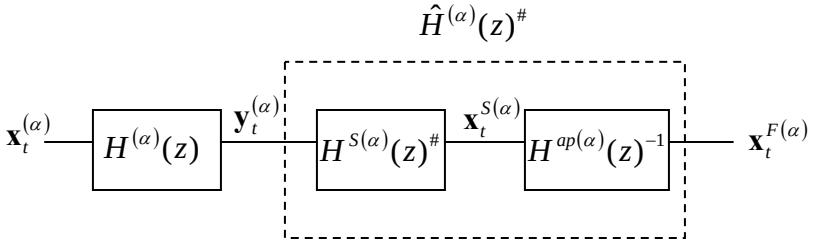

Figure 6.6. Block diagram of the complete identification of $H^{(\alpha)}(z)$ based on cyclic second-order statistics and then cyclic fourth-order statistics.

As with the stationary case, the remaining $H^{ap(\alpha)}(z)$ filter can be computed using the fourth-order cyclic cumulant tensor $\mathcal{C}_4^{\alpha}\left[\mathbf{x}_t^{S(\alpha)};\boldsymbol{\tau}\right]$ on the signal $\mathbf{x}_t^{S(\alpha)}$. Figure 6.6 shows the complete block diagram of the identification procedure where the last remaining block is the identification of the all-pass component of $H^{(\alpha)}(z)$. As was done with second-order statistics, the final stage should also be performed on the same subset of signals for a given cycle frequency α after $H^{S(\alpha)}(z)$ has been identified.

$$\begin{aligned}
\mathbf{x}_t^{S(\alpha)} &= \sum_{k=0}^{L} H_k^{ap(\alpha)} \mathbf{x}_{t-k}^{F(\alpha)} \quad \Leftrightarrow \quad \mathbf{x}^{S(\alpha)}(z) = H^{ap(\alpha)}(z)\mathbf{x}^{F(\alpha)}(z) \\
\mathbf{x}_t^{F(\alpha)} &= \sum_{k=-L}^{0} H_{-k}^{ap(\alpha)^*} \mathbf{x}_{t-k}^{S(\alpha)} \quad \Leftrightarrow \quad \mathbf{x}^{F(\alpha)}(z) = H^{ap(\alpha)}(z)^{-1}\mathbf{x}^{S(\alpha)}(z) = H^{ap(\alpha)^*}\left(\frac{1}{z^*}\right)\mathbf{x}^{S(\alpha)}(z)
\end{aligned} \tag{6-53}$$

The fourth-order cumulant tensor of $\mathbf{x}_t^{S(\alpha)}$ is not in general diagonal because its elements are still related through the filter $H^{ap(\alpha)}(z)$. $\mathcal{C}_4^{\alpha}\left[\mathbf{x}_t^{S(\alpha)};\boldsymbol{\tau}\right]$ which is of size MxMxMxM can be computed using (6-38) and is given by

$$\begin{aligned}
\left(\mathcal{C}_4^{\alpha}\left[\mathbf{x}_t^{S(\alpha)};\boldsymbol{\tau}\right]\right)_{ijkl} &= M_{ij^*k^*l}^{\alpha}\left[\mathbf{x}_t^{S(\alpha)};\boldsymbol{\tau}\right] - \sum_{\beta} M_{ij^*}^{\beta}\left[\mathbf{x}_t^{S(\alpha)};\boldsymbol{\tau}\right] M_{k^*l}^{\alpha-\beta}\left[\mathbf{x}_t^{S(\alpha)};\boldsymbol{\tau}\right] \\
&\quad - \sum_{\beta} M_{ik^*}^{\beta}\left[\mathbf{x}_t^{S(\alpha)};\boldsymbol{\tau}\right] M_{j^*l}^{\alpha-\beta}\left[\mathbf{x}_t^{S(\alpha)};\boldsymbol{\tau}\right] - \sum_{\beta} M_{il}^{\beta}\left[\mathbf{x}_t^{S(\alpha)};\boldsymbol{\tau}\right] M_{j^*k^*}^{\alpha-\beta}\left[\mathbf{x}_t^{S(\alpha)};\boldsymbol{\tau}\right]
\end{aligned} \tag{6-54}$$

where the moments are the elements of $\boldsymbol{C}_2^{\alpha}\left[\mathbf{x}_t^{S(\alpha)};\boldsymbol{\tau}\right]$

$$\left(\boldsymbol{C}_2^{\alpha}\left[\mathbf{x}_t^{S(\alpha)};\boldsymbol{\tau}\right]\right)_{ij} = M_{ij^*}^{\alpha}\left[\mathbf{x}_t^{S(\alpha)};\boldsymbol{\tau}\right] \tag{6-55}$$

Recall that $\boldsymbol{C}_2^{\alpha}\left[\mathbf{x}_t^{S(\alpha)};\boldsymbol{\tau}\right]$ is a unit magnitude diagonal MxM matrix with phases that are a function of $T_{0,i}$, $\boldsymbol{\tau}$ and α (6-42). In (6-54), $\beta,\alpha\in$(-1/2,1/2] and if the cyclic frequency α-β falls out of the range of (-1/2,1/2], it can always be brought back by adding or subtracting integers to the difference until it falls in the correct range. This can be done because the cyclic statistics are periodic in the cycle frequency domain.

The fourth-order cumulant tensor of $\mathbf{x}_t^{F(\alpha)}$, $\boldsymbol{C}_4^{\alpha}\left[\mathbf{x}_t^{F(\alpha)};\boldsymbol{\tau}\right]$, should be diagonal and exist only when $\tau_1=\tau_2=\tau_3=\tau_4$ to be space-time independent. This will ensure that $\mathbf{x}_t^{F(\alpha)}$ be like $\mathbf{x}_t^{(\alpha)}$ up to the irresolvable ambiguities of the system. To remove the dependence of $\mathbf{x}_t^{S(\alpha)}$ it will be convenient to extend the signal into a larger signal space that encompasses the length of $H^{ap(\alpha)}$(z)

$$\begin{aligned}
\mathbf{x}_i^{S(\alpha)} &= \begin{bmatrix} H_0^{ap(\alpha)} & \cdots & H_L^{ap(\alpha)} \end{bmatrix} \begin{bmatrix} \mathbf{x}_i^{F(\alpha)} \\ \vdots \\ \mathbf{x}_{i-L}^{F(\alpha)} \end{bmatrix} = \tilde{H}^{ap(\alpha)} \tilde{\mathbf{x}}_i^{F(\alpha)} \\
\mathbf{x}_i^{F(\alpha)} &= \begin{bmatrix} H_L^{ap(\alpha)^*} & \cdots & H_0^{ap(\alpha)^*} \end{bmatrix} \begin{bmatrix} \mathbf{x}_{i+L}^{S(\alpha)} \\ \vdots \\ \mathbf{x}_i^{S(\alpha)} \end{bmatrix} = \tilde{H}^{\overline{ap}(\alpha)} \tilde{\mathbf{x}}_i^{S(\alpha)}
\end{aligned} \tag{6-56}$$

As with the stationary case, the para-unitary filter $\tilde{H}^{\overline{ap}(\alpha)}$ that diagonalizes the tensor $\boldsymbol{C}_4^{\alpha}\left[\tilde{\mathbf{x}}_t^{S(\alpha)};\boldsymbol{\tau}\right]$ such that $\boldsymbol{C}_4^{\alpha}\left[\mathbf{x}_t^{F(\alpha)};\boldsymbol{\tau}\right]$ is a MxMxMxM diagonal and only exist for $\delta_{\boldsymbol{\tau}}$ can be computed. Chapter 5 described how to create a general form for the all-pass filter $\tilde{H}^{\overline{ap}(\alpha)}$. A contrast function (a statistical version of a cost function) can be used to determine the coefficient of the all-pass filter. The details of this procedure will be explained in the following chapter.

After all signals and their channels have been identified, the estimated channels can be concatenated to arrive at the complete channel response. Recall that the statistics of the signals that that were modulated are contained in each of the cycle frequencies (6-42) and (6-43). The statistical properties of $\mathbf{s}_t$ can be recovered this way and the probability density function of the signals computed. The cumulants of $\mathbf{x}_t^{F(\alpha)}$ contain the cumulants of their respective $\mathbf{s}_t$ so their distribution can be recovered by using to these statistics to reconstruct the probability density function.

6.5 Summary

Higher-order statistics based identification exploit the non-Gaussian distribution of the signals in the system. The identification of $H(z)$ described in this dissertation is based on a two-stage procedure. First, a second-order statistics based identification is used to estimate the minimum-phase component of the channel. The procedure is then followed by a fourth-order cumulant based identification to determine the all-pass component of $H(z)$.

M signals can be recovered with their respective channels as long as $N \geq M$ in a system which contains purely stationary signals. When the signals in the system are cyclostationary, however, more signals than sensors can be identified. If there is at least one cycle frequency where all signals are available with no more than N overlapping signals, the system transfer function and the signal distributions can be completely identified for linearly modulated signals. In order to take advantage of cyclostationarity, the signals in the system need to be oversampled so that their periodic nature can be identified.

In the next chapter algorithms will be developed to determine all the filters coefficients introduced in this chapter. Examples will also be given to show the effectiveness of higher order statistics based identification on simulated data.

References

[1] Y. Hua and J. K. Tugnait, "Blind Identifiability of FIR-MIMO Systems with Colored Input Using Second Order Statistics," *IEEE Signal Processing Letters*, vol.7, no. 12, pp. 348-350, December 2000.

[2] M. Kawamoto and Y. Inouye, "Blind Deconvolution of MIMO-FIR Systems with Colored Inputs Using Second-Order Statistics," *IEICE Trans. Fundamentals*, vol.E86-A, no. 3, pp. 597-604, March 2003.

[3] A. Belouchrani, K. Abed-Meraim, J-F. Cardoso, and E. Moulines, "A Blind Source Separation Technique Using Second-Order Statistics," *IEEE Trans. on Signal Processing*, vol. 45, no. 2, pp. 434-444, February 1997.

[4] H. Buchner, R. Aichner, and W. Kellermann, "A Generalization of a Class of Blind Source Separation Algorithms for Convolutive Mixtures," *Proc. Int. Symp. on Indep. Comp. Analys. (ICA)*, April 2003, Nara, Japan.

[5] L. Tong, G. Xu, and B. Hassibi, "Blind Identification Based on Second-Order Statistics: A Frequency-Domain Approach," *IEEE Trans. on Information Theory*, vol. 41, no. 1, pp. 329-334, January 1995.

[6] L. Tong, G. Xu, and T. Kailath, "Blind Identification and Equalization Based on Second-Order Statistics: A Time Domain Approach," *IEEE Trans. on Information Theory*, vol. 40, pp. 340-349, no. 2, March 1994.

[7] W. A. Gardner, *Cyclostationarity in Communications and Signal Processing*. New York: IEEE Press, 1994

[8] M. Joho, "Convolutive Blind Signal Separation in Acoustics by Joint Approximate Diagonalization of Spatiotemporal Correlation Matrices," *38th*

Asilomar Conf. on Signals, Systems, and Computers, vol. 1, pp. 983-988, November 2004.

[9] J. K. Tugnait and B. Huang, "On a Whitening Approach to Partial Channel Estimation and Blind Equalization of FIR/IIR Multiple-Input Multiple-Output Channels," IEEE Trans. on Signal Processing, vol. 48, no.3, pp. 832-845, March 2000.

[10] T. Kailath, A. H. Syed, and B. Hassibi, *Linear Estimation*. Upper Saddle River, New Jersey: Prentice Hall, 2000.

Chapter 7

Algorithms for Blind System Identification

The previous chapters were devoted to the development of the theory behind blind channel and signal identification for cyclostationary and stationary signals. This chapter addresses the computational aspects associated with determining the transfer function H_t and the distribution and properties of the system signals $\mathbf{x}_t$. Linearly modulated signals at baseband are exclusively treated in this chapter to demonstrate the effectiveness of cyclostationary based identification. In this chapter, there is no restriction on the number of signals M in the system so more signals than antenna elements (which are N in number) are considered.

Algorithms are described that can be used to determine the various filter coefficients introduced in Chapter 6. The methods presented here are not necessarily the only and most efficient ways of computing the unknowns, but they do guarantee consistent and accurate results. At the end of this chapter, examples will be covered to demonstrate the effectiveness of blind system identification using higher-order cyclostationary statistics.

7.1 Subset Model

Recall from Chapter 6 that signals used in digital communication systems have cyclostationary statistics. Such systems can be analyzed and identified by considering a subset of all the signals at a time based on the signal cyclic statistics. This is possible because of the sparseness and the potential differences in the cyclic cumulants of the signals $\mathbf{x}_t$ contained in $\mathbf{y}_t$. In Chapter 6 it was shown that the statistics of linearly modulated source signals $\mathbf{s}_t$ repeat at each of the cycle frequencies. Recall that a cyclostationary signal has cyclic statistics at discrete frequencies α that are a function of data rate, center frequency, or other modulation-induced periodicities. Also, because of the multi-linearity property of cyclic cumulants (Chapter 3), the received signal $\mathbf{y}_t$ at any given cyclic statistics α contains contributions of only the transmitters in $\mathbf{x}_t$ that have the same cycle frequency α modified by their respective channels. By taking advantage of the differences in cyclic statistics of all the system signals (because most likely not all source signals have the same modulation) more signals than sensors can be identified using this method. This was not possible for stationary signals since stationary signals only have a single cyclic frequency; all M signals overlap on each other at α=0. As long as there exists a cyclic frequency α of $\mathbf{y}_t$ where all signals $\mathbf{x}_t$ are available with no more than N overlapping signals, the system is fully identifiable (channel and signal properties) for linearly modulated signals when using this method.

First, recall that for cyclostationary signals, it is possible to work with a reduced number of signals at a time with the same cycle frequencies as discussed in Chapter 6. Consider a subset of the M source signals of $\mathbf{x}_t$ that have non-zero cyclic cumulants at a particular frequency α, denoted $\mathbf{x}_t^{(\alpha)}$, and for identification purposes are less than or equal to N. Considering only these signals, the resulting received signal $\mathbf{y}_t^{(\alpha)}$ based only on the contribution of the signals $\mathbf{x}_t^{(\alpha)}$ would result in

$$\mathbf{y}_t^{(\alpha)} = \sum_{l=0}^{\mathrm{L}} H_t^{(\alpha)} \mathbf{x}_{t-l}^{(\alpha)} \quad \Leftrightarrow \quad \mathbf{y}^{(\alpha)}(z) = H^{(\alpha)}(z)\mathbf{x}^{(\alpha)}(z) \tag{7-1}$$

where $H_t^{(\alpha)}$ is a reduced size version of the original transfer function H_t that only contains the filter coefficient that have to do with the signals $\mathbf{x}_t^{(\alpha)}$. Note that $\mathbf{y}_t^{(\alpha)}$ is not a measurable quantity and its time domain function cannot be extracted from the received signal $\mathbf{y}_t$. The statistics of $\mathbf{y}_t^{(\alpha)}$, however, are separable from $\mathbf{y}_t$ and computable from the time domain function of $\mathbf{y}_t$. The cyclic cumulants of $\mathbf{y}_t^{(\alpha)}$ at α are simply the statistics of $\mathbf{y}_t$ at the cycle frequency α. As will be seen shortly, for system identification purposes, it is enough if the cyclic cumulants of $\mathbf{y}_t^{(\alpha)}$ are computable to determine the filter $H_t^{(\alpha)}$. Note that the cyclic cumulants of signal $\mathbf{y}_t$ at α contain the statistics of just the subset signals $\mathbf{x}_t^{(\alpha)}$ (from which the modulation can be estimated) modified by the channel $H_t^{(\alpha)}$.

In the sections that follow, the theory is developed with the assumptions that N signals need to be identified at each cycle frequency because the actual number of signals $\mathbf{x}_t^{(\alpha)}$, that have statistics at α, is not known before hand. This is not restrictive, however, since the filter coefficients $H_t^{(\alpha)}$ for signals that belong to an element in $\mathbf{x}_t^{(\alpha)}$ that does not exist will be close to zero as will be seen in the example found in Section 7.4.

7.2 Identification of the Minimum-phase Component of $H^{(\alpha)}(z)$

In Chapter 5, it was found that the FIR channel $H^{(\alpha)}(z)$ can be decomposed into a minimum-phase and an all-pass component. In Chapter 6 it was shown that the innovations method is a second-order statistics based method that can be used to determine a minimum-phase component of the channel $H^{(\alpha)}(z)$. This filter based solely on second-order statistics was labeled as $H^{S(\alpha)}(z)$ and whose inverse when applied to $\mathbf{y}_t$ decorrelated (in space and in time) the signals present at a cycle frequency α. The following paragraphs will explain how to computed the filter coefficients of $H^{S(\alpha)}(z)$ in detail from the measured signal $\mathbf{y}_t$.

Firstly, the innovations vector for signals that belong to the subset α can be written as (Chapter 6)

$$\mathbf{y}_t^{(\alpha)} = -\sum_{k=1}^{\mathrm{K}} D_k^{(\alpha)} \mathbf{y}_{t-k}^{(\alpha)} + \mathbf{i}_t^{(\alpha)} \tag{7-2}$$

where $\mathbf{i}_t^{(\alpha)}$ is the innovations vector which by definition is uncorrelated with any past values of $\mathbf{y}_t$ or equivalently

$$E\left[\mathbf{i}_t^{(\alpha)} \mathbf{y}_{t-k}^{(\alpha)*}\right] = 0 \quad \text{for } k \geq 1 \tag{7-3}$$

In the innovations representation in (6-13), the filter $D_k^{(\alpha)}$ represents the memory of system and determines how far and by how much the past values of $\mathbf{y}_t$ make up its present value. The filter $D_k^{(\alpha)}$ is assumed to be of length K+1 and its zero coefficient is given by $D_0^{(\alpha)} = I_{\mathrm{N}}$. In matrix form, the summation in (7-2) can be written as

$$\mathbf{y}_t^{(\alpha)} = -\left[D_1^{(\alpha)} \quad D_2^{(\alpha)} \quad \cdots \quad D_K^{(\alpha)}\right] \begin{bmatrix} \mathbf{y}_{t-1}^{(\alpha)} \\ \mathbf{y}_{t-2}^{(\alpha)} \\ \vdots \\ \mathbf{y}_{t-K}^{(\alpha)} \end{bmatrix} + \mathbf{i}_t^{(\alpha)} \tag{7-4}$$

Multiplying the left and right hand side of the equation by $\left[\mathbf{y}_{t-1}^{*} \quad \mathbf{y}_{t-2}^{*} \quad \cdots \quad \mathbf{y}_{t-K}^{*}\right]$ and extracting only the second-order cyclic cumulant α results in

$$\left[C_2^{\alpha}\left[\mathbf{y}_t;\boldsymbol{\tau}=[0,1]\right]\cdots C_2^{\alpha}\left[\mathbf{y}_t;\boldsymbol{\tau}=[0,K]\right]\right]=-\left[D_1^{(\alpha)}\cdots D_K^{(\alpha)}\right]\begin{bmatrix} C_2^{\alpha}\left[\mathbf{y}_t;\boldsymbol{\tau}=[1,1]\right] & \cdots & C_2^{\alpha}\left[\mathbf{y}_t;\boldsymbol{\tau}=[1,K]\right] \\ \vdots & \ddots & \vdots \\ C_2^{\alpha}\left[\mathbf{y}_t;\boldsymbol{\tau}=[K,1]\right] & \cdots & C_2^{\alpha}\left[\mathbf{y}_t;\boldsymbol{\tau}=[K,K]\right] \end{bmatrix} \tag{7-5}$$

Note that the property (7-3) has been used to arrive at the above equation. Solving (7-5) for the filter coefficients $D_k^{(\alpha)}$ results in

$$\left[D_1^{(\alpha)}\cdots D_K^{(\alpha)}\right]=-\left[C_2^{\alpha}\left[\mathbf{y}_t;\boldsymbol{\tau}=[0,1]\right]\cdots C_2^{\alpha}\left[\mathbf{y}_t;\boldsymbol{\tau}=[0,K]\right]\right]\begin{bmatrix} C_2^{\alpha}\left[\mathbf{y}_t;\boldsymbol{\tau}=[1,1]\right] & \cdots & C_2^{\alpha}\left[\mathbf{y}_t;\boldsymbol{\tau}=[1,K]\right] \\ \vdots & \ddots & \vdots \\ C_2^{\alpha}\left[\mathbf{y}_t;\boldsymbol{\tau}=[K,1]\right] & \cdots & C_2^{\alpha}\left[\mathbf{y}_t;\boldsymbol{\tau}=[K,K]\right] \end{bmatrix}^{\#} \tag{7-6}$$

where the superscript # denotes the inverse as it was used throughout this dissertation. Note that the matrix that is inverted in (7-6) can possibly be rank deficient depending on how many signals of $\mathbf{x}_t$ have cyclic statistics at α. If the above inverse is rank deficient or close to losing rank, proper adjustments should be made to ensure the inverse is computed properly. The coefficients in the matrices given in (7-5) and (7-6) can be computed as explained in Chapter 6 or from

$$C_2^{\alpha}\left[\mathbf{y}_t;\boldsymbol{\tau}\right]=\frac{1}{2T+1}\sum_{t=-T}^{T}\mathbf{y}_{t-\tau_1}\mathbf{y}_{t-\tau_2}^{*}e^{-j2\pi\alpha t} \tag{7-7}$$

$$\left(C_2^{\alpha}\left[\mathbf{y}_t;\boldsymbol{\tau}\right]\right)_{ij}=M_{ij^*}^{\alpha}\left[\mathbf{y}_t;\boldsymbol{\tau}\right]=\frac{1}{2T+1}\sum_{t=-T}^{T}\left(\mathbf{y}_{t-\tau_1}\right)_i\left(\mathbf{y}_{t-\tau_2}\right)_j^{*}e^{-j2\pi\alpha t}$$

Using (7-6) and (7-7) the filter $D_k^{(\alpha)}$ can be computed for the subset signal $\mathbf{y}_t^{(\alpha)}$ from the measured signal $\mathbf{y}_t$. The length of the filter $D_k^{(\alpha)}$ is K+1 where the value of K is not known *a priori*; in practical situations it is best to make K as long as possible or truncate it when the values of the filter $D_k^{(\alpha)}$ fall bellow a pre-determined threshold. The innovations vector can now be computed for the signals that have cyclic statistics at α by

re-arranging (7-2) and solving for $\mathbf{i}_t^{(\alpha)}$ using the computed filter coefficients $D_k^{(\alpha)}$ as follows

$$\mathbf{i}_t = \sum_{k=0}^{K} D_k^{(\alpha)} \mathbf{y}_{t-k} \quad \Leftrightarrow \quad \mathbf{i}_t^{(\alpha)} = \sum_{k=0}^{K} D_k^{(\alpha)} \mathbf{y}_{t-k}^{(\alpha)} \text{ where } D_0^{(\alpha)} = I_{\mathrm{N}}$$

$$\begin{aligned} C_1^{\alpha}\left[\mathbf{i}_t;\boldsymbol{\tau}\right] &= C_1^{\alpha}\left[\mathbf{i}_t^{(\alpha)};\boldsymbol{\tau}\right] \\ C_2^{\alpha}\left[\mathbf{i}_t;\boldsymbol{\tau}\right] &= C_2^{\alpha}\left[\mathbf{i}_t^{(\alpha)};\boldsymbol{\tau}\right] \\ C_3^{\alpha}\left[\mathbf{i}_t;\boldsymbol{\tau}\right] &= C_3^{\alpha}\left[\mathbf{i}_t^{(\alpha)};\boldsymbol{\tau}\right] \\ C_4^{\alpha}\left[\mathbf{i}_t;\boldsymbol{\tau}\right] &= C_4^{\alpha}\left[\mathbf{i}_t^{(\alpha)};\boldsymbol{\tau}\right] \\ &\vdots \end{aligned} \tag{7-8}$$

Note that filter $D_k^{(\alpha)}$ above can be applied directly to the received signal $\mathbf{y}_t$ instead of $\mathbf{y}_t^{(\alpha)}$. This is because $\mathbf{y}_t^{(\alpha)}$ is not a measurable quantity and the application of the filter $D_k^{(\alpha)}$ to either of these two signals produces the same effect of uncorrelating the signals $\mathbf{i}_t^{(\alpha)}$ in the time domain. This is evident by noting that $C_2^{\alpha}\left[\mathbf{i}_t;\boldsymbol{\tau}\right]$ and $C_2^{\alpha}\left[\mathbf{i}_t^{(\alpha)};\boldsymbol{\tau}\right]$ are equal to each other and non-zero only when $\tau_1 = \tau_2$. In this dissertation, the signals $\mathbf{i}_t^{(\alpha)}$ and $\mathbf{i}_t$ are said to be *congruent in the statistical sense at a cycle frequency* α (CSS) if (7-8) holds. The signals $\mathbf{y}_t$ and $\mathbf{y}_t^{(\alpha)}$ as well as $\mathbf{x}_t$ and $\mathbf{x}_t^{(\alpha)}$ are also both *congruent in the statistical sense at a cycle frequency* α. The advantage of using cyclic statistics is evident; it is not necessary to have access to the time domain waveforms of the subset signals of for example $\mathbf{y}_t^{(\alpha)}$, their statistics can be readily extracted from the measurable signals like $\mathbf{y}_t$. The statistics $C_{\mathrm{X}}^{\alpha}\left[\mathbf{i}_t;\boldsymbol{\tau}\right]$ and $C_{\mathrm{X}}^{\alpha}\left[\mathbf{i}_t^{(\alpha)};\boldsymbol{\tau}\right]$ are equivalent and will be used interchangeably in this chapter depending on the point that is being conveyed where X is the order of the cumulant. A similar equivalence in the statistical sense can be extended to all other random signals in the system.

In Chapter 6, it was noted that the signals $\mathbf{i}_t^{(\alpha)}$ are not necessarily uncorrelated in the spatial domain - across the elements of $\mathbf{i}_t^{(\alpha)}$. To decorrelate the signals in both space and

time requires an additional operation. Recall from Chapter 6, that a signal denoted by $\mathbf{x}_t^{S(\alpha)}$ can be found that has similar statistics to $\mathbf{x}_t^{(\alpha)}$ up to the second-order. The space-time uncorrelated signal $\mathbf{x}_t^{S(\alpha)}$ can be computed from the innovations vector $\mathbf{i}_t^{(\alpha)}$, or equivalently $\mathbf{i}_t$ in (7-8), through a matrix multiplication as follows

$$\mathbf{x}_t^{S(\alpha)} = H_0^{S\#(\alpha)}\mathbf{i}_t^{(\alpha)} \quad \Leftrightarrow \quad \mathbf{x}_t^{S} = H_0^{S\#(\alpha)}\mathbf{i}_t \quad \text{CSS at } \alpha \tag{7-9}$$

The desired magnitude of the elements of the second-order cyclic cumulant of $\mathbf{x}_t^{S(\alpha)}$ should be $\left|\boldsymbol{C}_2^{\alpha}\left[\mathbf{x}_t^{S(\alpha)};\boldsymbol{\tau}\right] = I_{\mathrm{N}}\delta_{\boldsymbol{\tau}}\right.$. This problem can be thought of as a unit magnitude diagonalization of the second-order cumulant of $\mathbf{i}_t^{(\alpha)}$ whose eigenvalue-eigenvector decomposition can be written as

$$\boldsymbol{C}_2^{\alpha}\left[\mathbf{i}_t;\tau=[0,0]\right] = VDV^* \tag{7-10}$$

It is fairly straightforward to observe that if $H_0^{S\#(\alpha)}$ in (7-9) is picked to be

$$H_0^{S\#(\alpha)} = \left(V\sqrt{|D|}\right)^{\#} \tag{7-11}$$

the random vector $\mathbf{x}_t^{S(\alpha)}$ will be space-time uncorrelated with unit magnitude diagonal as desired. The complete signal filter response of $H_0^{S\#(\alpha)}$ can be computed by combing (7-11) and (7-6) to produce

$$H_k^{S\#(\alpha)} = H_0^{S\#(\alpha)}D_k^{(\alpha)} \quad \text{where } D_o^{(\alpha)} = I_{\mathrm{N}} \text{ and } k \geq 0 \tag{7-12}$$

Note that the symbol timing information $\mathrm{T}_{0,i}$ for the signals $\mathbf{x}_t^{(\alpha)}$ can be found from the phase of the diagonal elements of $\boldsymbol{C}_2^{\alpha}\left[\mathbf{x}_t^{S};\boldsymbol{\tau}=[0,0]\right]$ from

$$T_{0,i} = \frac{-\,phase\left\{\left(\boldsymbol{C}_2^{\alpha}\left[\mathbf{x}_t^S;\boldsymbol{\tau}=[0,0]\right]\right)_{ii}\right\}}{2\pi\alpha} \tag{7-13}$$

The modulation of the signals, specifically data rate and center frequency, can be estimated by looking at the magnitude of $\boldsymbol{C}_2^{\alpha}\left[\mathbf{x}_t^S;\boldsymbol{\tau}=[0,\tau_1]\right]$ and taking note of all the cycle frequencies were the various elements of $\mathbf{x}_t^S$ have been decorrelated in space and in time. Recall that decorrelating a linearly modulated signal at one of its cycle frequencies, also decorrelates the same signal at all its other cycle frequencies. This effect will be illustrated with an example in Section 7.4.1.

After $H_k^{S\#(\alpha)}$ has been computed, its inverse $H_k^{S(\alpha)}$ - which corresponds to the minimum-phase component of $H_k^{(\alpha)}$ - can be determined by analyzing the cross-correlation of $\mathbf{x}_t^{S(\alpha)}$ and $\mathbf{y}_t^{(\alpha)}$. Two such equation can be found which when combined relate the filter $H_k^{S(\alpha)}$ to its inverse $H_k^{S\#(\alpha)}$. The first equation uses the relationship $\mathbf{y}_t^{(\alpha)} = \sum_{k=0}^{\mathrm{K}} H_t^{S(\alpha)}\mathbf{x}_{t-k}^{S(\alpha)}$ to arrive at the following cross-correlation

$$\begin{aligned} E\left[\mathbf{x}_{t-k}^{S(\alpha)}\mathbf{y}_t^{(\alpha)*}\right] &= E\left[\mathbf{x}_{t-k}^{S(\alpha)}\sum_{l=0}^{\infty}\mathbf{x}_{t-l}^{S(\alpha)*}H_l^{S(\alpha)*}\right] = \sum_{l=0}^{\infty}E\left[\mathbf{x}_{t-k}^{S(\alpha)}\mathbf{x}_{t-l}^{S(\alpha)*}\right]H_l^{S(\alpha)*} \\ &= \sum_{l=0}^{\infty}\boldsymbol{C}_2\left[\mathbf{x}_t^{S(\alpha)};t,\boldsymbol{\tau}=[k,l]\right]H_l^{S(\alpha)*} = \boldsymbol{C}_2\left[\mathbf{x}_t^{S(\alpha)};t,\boldsymbol{\tau}=[k,k]\right]H_k^{S(\alpha)*} \end{aligned} \tag{7-14}$$

A second representation of the cross-correlation $E\left[\mathbf{x}_{t-k}^{S(\alpha)}\mathbf{y}_t^{(\alpha)*}\right]$ can be found using (7-8) and (7-9) which yields the following result

$$\begin{aligned} E\left[\mathbf{x}_{t-k}^{S(\alpha)}\mathbf{y}_t^{(\alpha)*}\right] &= H_0^{S(\alpha)\#}E\left[\mathbf{i}_{t-k}^{(\alpha)}\mathbf{y}_t^{(\alpha)*}\right] = H_0^{S(\alpha)\#}E\left[\left(\sum_{l=0}^{K}D_l^{(\alpha)}\mathbf{y}_{t-k-l}^{(\alpha)}\right)\mathbf{y}_t^{(\alpha)*}\right] \\ &= H_0^{S(\alpha)\#}\sum_{l=0}^{K}D_l^{(\alpha)}\boldsymbol{C}_2\left[\mathbf{y}_t^{(\alpha)};t,\boldsymbol{\tau}=[k+l,0]\right] = \sum_{l=0}^{K}H_l^{S(\alpha)\#}\boldsymbol{C}_2\left[\mathbf{y}_t^{(\alpha)};t,\boldsymbol{\tau}=[k+l,0]\right] \end{aligned} \tag{7-15}$$

Equating (7-14) to (7-15) and considering only the second-order cyclic cumulant at α produces the desired result, an equation relating $H_k^{S(\alpha)}$ to its inverse $H_k^{S(\alpha)\#}$.

$$H_k^{S(\alpha)^*} = \boldsymbol{C}_2^{\alpha}\left[\mathbf{x}_t^S; \boldsymbol{\tau} = [k,k]\right]^{\#} \sum_{l=0}^{K} H_l^{S(\alpha)\#} \boldsymbol{C}_2^{\alpha}\left[\mathbf{y}_t; \boldsymbol{\tau} = [k+l,0]\right] \tag{7-16}$$

The filter $H_k^{S(\alpha)}$ for the subset of signals that have a cycle frequency α in (7-1) can be found solely from the received signal $\mathbf{y}_t$. Recall that the computed $H_k^{S(\alpha)}$ corresponds to the minimum-phase component of the subset transfer function $H_k^{(\alpha)}$. To complete the identification of $H_k^{(\alpha)}$ requires determining also the missing all-pass component, $H_k^{ap(\alpha)}$. The all-pass component of $H_k^{(\alpha)}$ can be computed using higher-order statistics as explained in Chapter 6.

7.3 Identification of the All-pass Component of $H^{(\alpha)}(z)$

The transfer function $H^{(\alpha)}(z)$ is identifiable up to the irresolvable ambiguities of the system (Chapter 6.1) and can be decomposed into

$$\begin{aligned} H^{(\alpha)}(z) &= H^{mp(\alpha)}(z) H^{ap(\alpha)}(z) \\ H^{mp(\alpha)} &= H^{S(\alpha)}(z) \end{aligned} \tag{7-17}$$

The previous section explained how to compute $H_k^{S(\alpha)}$ using the second-order cyclic cumulants. Equivalently, the fourth-order cyclic cumulant can be used to determine the remaining all-pass component of the transfer function. In this section a method for determining $H^{ap(\alpha)}(z)$ based on a minimization of a contrast function will be presented. Contrast functions can be thought of as statistical cost functions which when minimized or maximized yield the desired filter. In Chapter 6, it was found that the inverse of the all-pass filter that completes the identification of $H^{(\alpha)}(z)$ should, when applied to $\tilde{\mathbf{x}}_t^{S(\alpha)}$ as

$$\mathbf{x}_t^{S(\alpha)} = \begin{bmatrix} H_o^{ap(\alpha)} & \cdots & H_{\mathrm{L}}^{ap(\alpha)} \end{bmatrix} \begin{bmatrix} \mathbf{x}_t^{F(\alpha)} \\ \vdots \\ \mathbf{x}_{t-\mathrm{L}}^{F(\alpha)} \end{bmatrix} = \tilde{H}^{ap(\alpha)} \tilde{\mathbf{x}}_t^{F(\alpha)} \quad \Leftrightarrow \quad \mathbf{x}_t^{S} = \tilde{H}^{ap(\alpha)} \tilde{\mathbf{x}}_t^{F} \ \mathrm{CSS\,at}\,\alpha$$

$$\mathbf{x}_t^{F(\alpha)} = \begin{bmatrix} H_{\mathrm{L}}^{ap(\alpha)*} & \cdots & H_0^{ap(\alpha)*} \end{bmatrix} \begin{bmatrix} \mathbf{x}_{t+\mathrm{L}}^{S(\alpha)} \\ \vdots \\ \mathbf{x}_t^{S(\alpha)} \end{bmatrix} = \tilde{H}^{\overline{ap}(\alpha)} \tilde{\mathbf{x}}_t^{S(\alpha)} \quad \Leftrightarrow \quad \mathbf{x}_t^{F} = \tilde{H}^{\overline{ap}(\alpha)} \tilde{\mathbf{x}}_t^{S} \ \mathrm{CSS\,at}\,\alpha \qquad (7\text{-}18)$$

result in a $\boldsymbol{C}_4^{\alpha}\left[\mathbf{x}_t^F;\boldsymbol{\tau}\right]$ that is a diagonal NxNxNxN tensor and is non-zero only when $\tau_1 = \tau_2 = \tau_3 = \tau_4$. To start with the analysis and the development of an algorithm that solves for $H_k^{ap(\alpha)}$ first consider the general form of the all-pass filter. MxM all-pass filters were discussed in detail in Chapter 5 and can be described in general when FIR as

$$H^{ap}(z) = \left(Q_{\mathrm{L}} + z^{-1}P_{\mathrm{L}}\right)..\left(Q_l + z^{-1}P_l\right)..\left(Q_1 + z^{-1}P_1\right)H_0 \qquad (7\text{-}19)$$

where

$$P_l = v_l v_l^*, \quad Q_l = I - P_l \qquad (7\text{-}20)$$

and v_l is a Nx1 unit vector. The filter (7-19) can be transformed into the time domain and concatenated as in (7-18). A general all-pass filter with L=3, for example, has the following structure

$$\tilde{H}_3^{ap} = \begin{bmatrix} Q_3Q_2Q_1H_o & \left(Q_3Q_2P_1 + Q_3P_2Q_1 + P_3Q_2Q_1\right)H_0 & \left(P_3P_2Q_1 + P_3Q_2P_1 + Q_3P_2P_1\right)H_0 & P_3P_2P_1H_0 \end{bmatrix}$$
$$\tilde{H}_3^{ap} = \begin{bmatrix} H_o^{ap} & H_1^{ap} & H_2^{ap} & H_3^{ap} \end{bmatrix} \qquad (7\text{-}21)$$

Note that the first position (H_0^{ap}) of $\tilde{H}^{ap}$ contains the multiplication of just the matrices Q_l from 3 down to 1. The second position (H_1^{ap}) contains the summation of all permutations of length L with just a single matrix P_l and the rest are made up of Q_l. The

third position (H_2^{ap}) contains the summation of all permutations of size L with two of its elements P_l and the rest are filled with the corresponding Q_l. This is a general pattern and is applicable to any all-pass filter of length L+1; the position in $\tilde{H}_{\mathrm{L}}^{ap}$ determines the number of P_l matrices that occur in each term with all possible permutations of this event of size L present in the summation. The last position of $\tilde{H}_{\mathrm{L}}^{ap}$ corresponds to the multiplication of all P_l from L down to 1. The above-mentioned procedure can be used to create and arbitrary length concatenated all-pass filter $\tilde{H}_{\mathrm{L}}^{ap}$.

To develop an algorithm to find $\tilde{H}^{ap(\alpha)}$ in (7-18) requires first examining the general structure of the matrix Q_l, P_l, and H_0 in more detail. Specifically, the 2x2 MIMO all-pass filter will be analyzed in detail here since it will be used in the examples. An arbitrary MxM all-pass filter can be described in a similar manner. Firstly, a general 2x2 unitary matrix H_0 can be constructed as follows

$$H_0 = \frac{1}{\sqrt{1+x_o^2}}\begin{bmatrix} 1 & -x_0 e^{-jy_0} \\ x_0 e^{jy_0} & 1 \end{bmatrix} \tag{7-22}$$

where both x_0 and y_0 are arbitrary real numbers. Equivalently, the 2x1 unit vector v_l takes on the following form

$$v_l = \frac{1}{\sqrt{1+x_l^2}}\begin{bmatrix} 1 \\ x_l e^{jy_l} \end{bmatrix} \tag{7-23}$$

from which the matrices P_l and Q_l can be readily constructed. The resulting P_l and Q_l that make up the filter (7-19) can be found to be

$$P_l = v_l v_l^* = \frac{1}{1+x_l^2}\begin{bmatrix} 1 & x_l e^{-jy_l} \\ x_l e^{jy_l} & x_l^2 \end{bmatrix} \tag{7-24a}$$

$$Q_l = I - P_l = \frac{1}{1+x_l^2}\begin{bmatrix} x_l^2 & -x_l e^{-jy_l} \\ -x_l e^{jy_l} & 1 \end{bmatrix} \tag{7-24b}$$

Using (7-21), (7-22), (7-24), and from the discussion so far as a template, an arbitrary length 2x2 all-pass filter can be readily created. This filter will be a function of the variables x_l and y_l where $l \in \{0,\ldots,\mathrm{L}\}$.

The computation that produces the space-time uncorrelated signal $\mathbf{x}_t^{S(\alpha)}$ was discussed in the previous section. The L+1 shifted versions of the signals $\mathbf{x}_t^S$ can be concatenated in a vector format to create $\tilde{\mathbf{x}}_t^S$ as in (7-18). The length of $\tilde{\mathbf{x}}_t^S$ should ideally span the length of the all-pass filter $\tilde{H}^{\overline{ap}(\alpha)}$ but L is not know *a priori*. The reason for concatenating is to create a vector that extends the time and the space dimension such that the fourth-order cyclic cumulant $\mathcal{C}_4^{\alpha}\left[\tilde{\mathbf{x}}_t^S;\boldsymbol{\tau}\right]$ contains all the components that affect the fourth order cyclic cumulant of $\mathcal{C}_4^{\alpha}\left[\mathbf{x}_t^F;\boldsymbol{\tau}\right]$. A method for determining the length L is by taking advantage of the para-unitary properties of the all-pass filter as will be discussed in the example section. The fourth order cyclic cumulant at α of $\tilde{\mathbf{x}}_t^{S(\alpha)}$ can be computed from the available signals $\mathbf{x}_t^S$ as in (7-19) since

$$\mathcal{C}_4^{\alpha}\left[\tilde{\mathbf{x}}_t^S;\boldsymbol{\tau}\right] = \mathcal{C}_4^{\alpha}\left[\tilde{\mathbf{x}}_t^{S(\alpha)};\boldsymbol{\tau}\right] \tag{7-25}$$

Once L has been determined or picked, the inverse of the all-pass filter $\tilde{H}^{\overline{ap}(\alpha)}$ in (7-18), which is also all-pass can be constructed using the previously described procedure to create a filter that takes the form

$$\tilde{H}^{\overline{ap}(\alpha)} = \begin{bmatrix} Q_{\mathrm{L}}\ldots Q_1 H_o & \cdots & P_{\mathrm{L}}\ldots P_1 H_0 \end{bmatrix} \tag{7-26}$$

which is a function of x_l and y_l since P_l and Q_l are given in (7-24). The correct $\tilde{H}^{\overline{ap}(\alpha)}$ will be found when $\mathcal{C}_4^{\alpha}\left[\mathbf{x}_t^F;\boldsymbol{\tau}\right]$ is a diagonal NxNxNxN tensor and is non-zero only when

$\tau_1 = \tau_2 = \tau_3 = \tau_4$. Using the multi-linearity property of cumulants (Chapter 3), the fourth order cyclic cumulant of $\mathbf{x}_t^F$ can be computed from

$$\left(\boldsymbol{C}_4^{\alpha}\left[\mathbf{x}_t^F;\boldsymbol{\tau}=[0,0,0,0]\right]\right)_{ijkl} = \sum_{p,q,r,s=1}^{\mathrm{N(L+1)}} \left(\boldsymbol{C}_4^{\alpha}\left[\tilde{\mathbf{x}}_t^S;\boldsymbol{\tau}=[0,0,0,0]\right]\right)_{pqrs} \left(\tilde{H}^{\overline{ap}(\alpha)}\right)_{ip} \left(\tilde{H}^{\overline{ap}(\alpha)}\right)_{jq}^{*} \left(\tilde{H}^{\overline{ap}(\alpha)}\right)_{kr}^{*} \left(\tilde{H}^{\overline{ap}(\alpha)}\right)_{ls} \tag{7-27}$$

where $\boldsymbol{C}_4^{\alpha}\left[\mathbf{x}_t^F;\boldsymbol{\tau}=[0,0,0,0]\right]$ is a function of the cyclic cumulant tensor $\boldsymbol{C}_4^{\alpha}\left[\tilde{\mathbf{x}}_t^S;\boldsymbol{\tau}=[0,0,0,0]\right]$ and x_l and y_l. $\boldsymbol{\tau}$=[0,0,0,0] has been chosen for the fourth cyclic cumulant tensor since $\boldsymbol{C}_4^{\alpha}\left[\tilde{\mathbf{x}}_t^S;\boldsymbol{\tau}\right]$ is maximum for $\tau_1 = \tau_2 = \tau_3 = \tau_4$ and the all zero $\boldsymbol{\tau}$ is one such event and that can be conveniently computed.

As mentioned earlier, there is an alternative method for calculating the length of the all-pass filter. It involves first starting with L=0 and creating the corresponding length concatenated vector $\tilde{\mathbf{x}}_t^S$. All possible L=0 all-pass filters can then be applied to the signal $\tilde{\mathbf{x}}_t^S$ to see if any of them produce a fourth order cyclic cumulant $\boldsymbol{C}_4^{\alpha}\left[\mathbf{x}_t^F;\boldsymbol{\tau}\right]$ that satisfies the properties of being an NxNxNxN diagonal tensor that exists only for $\delta_{\boldsymbol{\tau}}$. If the time-space independence property of $\boldsymbol{C}_4^{\alpha}\left[\mathbf{x}_t^F;\boldsymbol{\tau}\right]$ is not satisfied for any of the filters L=0, increase the length of the all-pass filter by one (now L=1) and repeat the procedure. All the L=1 all-pass filter are then applied to the L+1 concatenated vector $\tilde{\mathbf{x}}_t^S$ and the check of time-space independence is performed on the fourth order cyclic cumulant tensor of $\mathbf{x}_t^F$ again. This procedure can be repeated until the correct L is reached which occurs when the cumulant tensor of $\mathbf{x}_t^F$ is in fact diagonal and non-zero only when $\tau_1 = \tau_2 = \tau_3 = \tau_4$. There are infinitely many all-pass filters that have to be cycled through for each L which makes this procedure unmanageable. An algorithm can be developed that solves for the different coefficient for each filter of length L+1 efficiently by taking advantage of properties of the all-pass filter. Recall that the all-pass filter is a para-unitary filter and has the following property

$$\tilde{H}_{\mathrm{L}}^{\overline{ap}(\alpha)}\tilde{H}_{\mathrm{L}}^{\overline{ap}(\alpha)*} = I_{\mathrm{N}} \tag{7-28}$$

This property can be taken advantage of to develop an algorithm that will solve for all the x_l and y_l to arrive at the correct $\tilde{H}^{\overline{ap}(\alpha)}$ through a minimization process. Since $\tilde{H}^{\overline{ap}(\alpha)}$ follows (7-28) and is related to its inverse as in (7-18), it will maximize the cumulant tensor $\boldsymbol{C}_4^{\alpha}\left[\mathbf{x}_t^F;\boldsymbol{\tau}\right]$ along its diagonal when the correct L and x_l and y_l are picked. $\tilde{H}^{\overline{ap}(\alpha)}$ can be thought of gathering the elements of the cyclic cumulant $\boldsymbol{C}_4^{\alpha}\left[\tilde{\mathbf{x}}_t^S;\boldsymbol{\tau}\right]$ and mapping them to the corresponding places in $\boldsymbol{C}_4^{\alpha}\left[\mathbf{x}_t^F;\boldsymbol{\tau}\right]$ while at the same time being constrained to (7-28). The energy in the filter is fixed and is spread throughout the length of the filter, so an algorithm can be implemented that can take advantage of the properties of all-pass filters.

The algorithm starts by first describing an L=0, $\tilde{H}_0^{\overline{ap}(\alpha)}$, all-pass filter which is simply a unitary matrix in terms of x_0 and y_0 and creating $\tilde{\mathbf{x}}_t^S$ which contains just a single time-shifted element of $\mathbf{x}_t^S$. Compute the cyclic cumulant $\boldsymbol{C}_4^{\alpha}\left[\tilde{\mathbf{x}}_t^S;\boldsymbol{\tau}\right]$ using (6-38) and derive an equation $\boldsymbol{C}_4^{\alpha}\left[\mathbf{x}_t^F;\boldsymbol{\tau}\right]$ using (7-27) as a function of x_0 and y_0 and the elements of $\boldsymbol{C}_4^{\alpha}\left[\tilde{\mathbf{x}}_t^S;\boldsymbol{\tau}\right]$. Making use of the properties of all-pass filters mentioned earlier, the closest L=0 all-pass filter to the true $\tilde{H}^{\overline{ap}(\alpha)}$ will maximize the diagonal elements of the cumulant $\boldsymbol{C}_4^{\alpha}\left[\mathbf{x}_t^F;\boldsymbol{\tau}\right]$ or equivalently maximize the contrast function $\phi_{ST}^{ap}\left[\mathbf{x}_t^{F(\alpha)}\right]$ given by

$$\begin{aligned}\phi_{ST}^{ap}\left[\mathbf{x}_t^{F(\alpha)},\mathrm{L}\right] &= \sum_{i=1}^{\mathrm{N}}\left(\boldsymbol{C}_4^{\alpha}\left[\mathbf{x}_t^{F(\alpha)};\boldsymbol{\tau}=[0,0,0,0]\right]\right)_{iiii}\left(\boldsymbol{C}_4^{\alpha}\left[\mathbf{x}_t^{F(\alpha)};\boldsymbol{\tau}=[0,0,0,0]\right]\right)_{iiii}^{*} \\ &= \sum_{i=1}^{\mathrm{N}}\left(\boldsymbol{C}_4^{\alpha}\left[\mathbf{x}_t^{F};\boldsymbol{\tau}=[0,0,0,0]\right]\right)_{iiii}\left(\boldsymbol{C}_4^{\alpha}\left[\mathbf{x}_t^{F};\boldsymbol{\tau}=[0,0,0,0]\right]\right)_{iiii}^{*}\end{aligned} \tag{7-29}$$

The above result can be expanded into its components as function of the all-pass filter of length L+1, $\tilde{H}_{\mathrm{L}}^{\overline{ap}(\alpha)}$

$$\phi_{ST}^{ap}\left[\mathbf{x}_t^{F(\alpha)},\mathrm{L}\right]=\sum_{i=1}^{\mathrm{N}}\left(\boldsymbol{C}_4^{\alpha}\left[\mathbf{x}_t^{F};\boldsymbol{\tau}=[0,0,0,0]\right]\right)_{iiii}\left(\boldsymbol{C}_4^{\alpha}\left[\mathbf{x}_t^{F};\boldsymbol{\tau}=[0,0,0,0]\right]\right)_{iiii}^{*}$$

$$=\sum_{i=1}^{\mathrm{N}}\left\{\begin{array}{l}\sum_{p,q,r,s=1}^{\mathrm{N(L+1)}}\left(\boldsymbol{C}_4^{\alpha}\left[\tilde{\mathbf{x}}_t^{S};\boldsymbol{\tau}=[0,0,0,0]\right]\right)_{pqrs}\left(\tilde{H}_{\mathrm{L}}^{\overline{ap}(\alpha)}\right)_{ip}\left(\tilde{H}_{\mathrm{L}}^{\overline{ap}(\alpha)}\right)_{iq}^{*}\left(\tilde{H}_{\mathrm{L}}^{\overline{ap}(\alpha)}\right)_{ir}^{*}\left(\tilde{H}_{\mathrm{L}}^{\overline{ap}(\alpha)}\right)_{is} \\ *\sum_{p,q,r,s=1}^{\mathrm{N(L+1)}}\left(\boldsymbol{C}_4^{\alpha}\left[\tilde{\mathbf{x}}_t^{S};\boldsymbol{\tau}=[0,0,0,0]\right]\right)_{pqrs}^{*}\left(\tilde{H}_{\mathrm{L}}^{\overline{ap}(\alpha)}\right)_{ip}^{*}\left(\tilde{H}_{\mathrm{L}}^{\overline{ap}(\alpha)}\right)_{iq}\left(\tilde{H}_{\mathrm{L}}^{\overline{ap}(\alpha)}\right)_{ir}\left(\tilde{H}_{\mathrm{L}}^{\overline{ap}(\alpha)}\right)_{is}^{*}\end{array}\right\}$$

(7-30)

For 2x2 MIMO all-pass filters, $\phi_{ST}^{ap}\left[\mathbf{x}_t^{F(\alpha)}\right]$ can be found to be a function of

$$\phi_{ST}^{ap}\left[\mathbf{x}_t^{F(\alpha)},\mathrm{L}\right]=f\left(\left(\boldsymbol{C}_4^{\alpha}\left[\tilde{\mathbf{x}}_t^{S};\boldsymbol{\tau}=[0,0,0,0]\right]\right)_{ijkl},x_{\mathrm{L}},...,x_0,y_{\mathrm{L}},...,y_0\right) \qquad (7\text{-}31)$$

Finding minimums of scalar functions of several variables have been studied extensively and numerical methods are readily available. A steepest decent method or an unconstrained nonlinear optimization procedure, which starts with an initial guess, can be used to arrive at the x_l and y_l that minimize the contrast function in (7-32). The global maximization of $\phi_{ST}^{ap}\left[\mathbf{x}_t^{F(\alpha)},\mathrm{L}\right]$ can be turned into a global minimization problem by taking its reciprocal as follows

$$\overline{\phi}_{ST}^{ap}\left[\mathbf{x}_t^{F(\alpha)},\mathrm{L}\right]=\frac{1}{\phi_{ST}^{ap}\left[\mathbf{x}_t^{F(\alpha)},\mathrm{L}\right]} \qquad (7\text{-}32)$$

Regardless of how the variables x_l and y_l are computed through the minimization process of $\overline{\phi}_{ST}^{ap}\left[\mathbf{x}_t^{F(\alpha)},\mathrm{L}\right]$ above, the resulting $\tilde{H}_{\mathrm{L}}^{\overline{ap}(\alpha)}$ corresponds to the closest L+1 length filter fit to the true $\tilde{H}^{\overline{ap}(\alpha)}$. After the L=0 filter has been optimized (by varying x_0 and y_0) we can move to the L=1 filter and repeat the procedure of creating a general all-pass filter which is now of length L+1=2 and a function of x_1, y_1, x_0 and y_0. The L=1 filter can then be optimized and its minimum value of $\overline{\phi}_{ST}^{ap}\left[\mathbf{x}_t^{F(\alpha)},\mathrm{L}=1\right]$ compared to the minimum $\overline{\phi}_{ST}^{ap}\left[\mathbf{x}_t^{F(\alpha)},\mathrm{L}=0\right]$ of the previous L=0 filter. If the minimum of the L=1 filter is lower

than the minimum value of the L=0 optimization function $\bar{\phi}_{ST}^{ap}\left[\mathbf{x}_t^{F(\alpha)}, \mathrm{L}\right]$ it means that the all-pass filter of length L+1=2 produces a better fit to the true $\widetilde{H}^{\overline{ap}(\alpha)}$. The optimization procedure is then repeated with increasing values of L until the global minimum of the contrast function $\bar{\phi}_{ST}^{ap}\left[\mathbf{x}_t^{F(\alpha)}, \mathrm{L}\right]$ remains constant as the filter length is increased at which the length of the all-pass filter has been determined since no further improvement is evident. Again, the reason this procedure works is because of the para-unitary properties of all-pass filters.

The function $\bar{\phi}_{ST}^{ap}\left[\mathbf{x}_t^{F(\alpha)}, \mathrm{L}\right]$, in general, has several global minimums each of which produce an all-pass filter of length L which together with $H^{S(\alpha)}(z)$ determines the filter $H^{(\alpha)}$(z) up to the irresolvable ambiguities of the system (Chapter 6.1). Any of the all-pass filters that produce global minimums of $\bar{\phi}_{ST}^{ap}\left[\mathbf{x}_t^{F(\alpha)}, \mathrm{L}\right]$ (assuming L+1 is the length of the all-pass filter being identified) are therefore a solution for $\widetilde{H}^{\overline{ap}(\alpha)}$. It should be noted that $\bar{\phi}_{ST}^{ap}\left[\mathbf{x}_t^{F(\alpha)}, \mathrm{L}\right]$ does have local minimums with some of them corresponding to the minimum values of the smaller length filters. To avoid landing in local minimum, several starting points for x_l and y_l should be used to ensure that the algorithm that minimizes $\bar{\phi}_{ST}^{ap}\left[\mathbf{x}_t^{F(\alpha)}, \mathrm{L}\right]$ has found a global and not a local minima.

The statistics of the signals $\mathbf{x}_t^{F(\alpha)}$ are space-time independent at the cycle frequency α and its cyclic cumulants are equivalent to the statistics of the original source signals $\mathbf{x}_t^{(\alpha)}$ at α. The cyclic cumulants of $\mathbf{x}_t^{F(\alpha)}$ therefore correspond to the cyclic cumulants of $\mathbf{x}_t^{(\alpha)}$ from which the sample statistics of $\mathbf{s}_t$ can be computed. Recall from (6-42) and (6-46), that the second and fourth-order cumulant of $\mathbf{s}_t$ for linearly modulated signals can be found from

$$\boldsymbol{C}_2\left[\left(\mathbf{s}_t\right)_i\right] = \boldsymbol{C}_2^{\alpha^i}\left[\left(\mathbf{x}_t\right)_i; \boldsymbol{\tau} = [0,0]\right] * e^{-phase\left[\boldsymbol{C}_2^{\alpha}\left[\left(\mathbf{x}_t\right)_i; \boldsymbol{\tau}=[0,0]\right]\right]} * T_i = T_i \tag{7-33a}$$

$$\boldsymbol{C}_4\left[\left(\mathbf{s}_t\right)_i\right] = \boldsymbol{C}_4^{\alpha^i}\left[\left(\mathbf{x}_t\right)_i; \boldsymbol{\tau} = [0,0]\right] * e^{-phase\left[\boldsymbol{C}_2^{\alpha}\left[\left(\mathbf{x}_t\right)_i; \boldsymbol{\tau}=[0,0]\right]\right]} * T_i \tag{7-33b}$$

and so on for the rest of the cumulants higher than the fourth order. In (7-33a), the unit power assumption of the signals $\mathbf{x}_t$ has been applied to the equation. The distribution of the signals $\mathbf{s}_t$ can be recovered by reconstructing the probability density functions using the estimated cumulants as in (7-33). From the distribution of $\mathbf{s}_t$, the alphabet such as BPSK, QPSK, QAM, etc. of the particular transmitter can be found as will be shown in the example. A common way of reconstructing the distribution of a signal from its cumulants or moments is through the Gram-Charlier series expansion [2]. The Gram-Charlier series can be found in [2] and up to the sixth order is given by

$$f(x) = \sum_{j=0}^{\infty} c_j H_j \alpha(x) \tag{7-34}$$

$$\begin{aligned}
H_0 &= 1 & c_0 &= 1 \\
H_1 &= x & c_1 &= 0 \\
H_2 &= x^2 - 1 & c_2 &= 1/2(\mu_2 - 1) \\
H_3 &= x^3 - 3x & c_3 &= 1/6(\mu_3) \\
H_4 &= x^4 - 6x^2 + 3 & c_4 &= 1/24(\mu_4 - 6\mu_2 + 3) \\
H_5 &= x^5 - 10x^3 + 15x & c_5 &= 1/120(\mu_5 - 10\mu_3) \\
H_6 &= x^6 - 15x^4 + 45x^2 - 15 & c_6 &= 1/720(\mu_6 - 15\mu_4 + 45\mu_2 - 15) \\
&\vdots & &\vdots
\end{aligned}$$

where $\alpha(x)$ is the normal distribution with variance equal to the variance of the distribution that is being estimated. The coefficients μ_X correspond the X order moment of the distribution which can be computed from the cumulants of $\mathbf{x}_t^{F(\alpha)}$ at α in (7-33) as explained in Chapter 3.

The above techniques will be used to determine the unknowns of a system consisting of a two-element array receiver. The following section examines an example where the different parameters like channel and modulation are estimated from just the received signals at the antenna elements.

7.4 Examples of System Identification

Two examples will be addressed in this section. The first case that will be considered is when the numbers of signals M in the system is larger than the number of antenna elements N and these signals are digital signals which have cyclostationary properties. The second example addresses stationary signals where the signals and transfer function can be estimated as long as N≥M.

7.4.1 Cyclostationary Signals

For the example consider a two-element antenna array at the receiver where the vector $\mathbf{y}_t$ contains the time domain fluctuations of the sampled received signal at the antenna elements. In this example, three linearly modulated signals at baseband denoted by vector $\mathbf{x}_t$ are applied to the channel. $\mathbf{x}_t$ contains the symbols being transmitted by each of the transmitters spaced by the reciprocal of their respective data rate (Figure 6.3). The power transmitted by each of the transmitters is normalized to unity. The transfer function H_t contains the aggregate effect of pulse shaping, the transmit/receive filters, the wireless channel which mixes and delays the signals, and finally the effects of the antenna elements as discussed in Chapter 6. The equation relating the received signal to the input can be written as

$$\mathbf{y}_t = H_t \mathbf{x}_t + \mathbf{n}_t \tag{7-35}$$

where $\mathbf{y}_t$ is additionally corrupted by an additive stationary noise signal $\mathbf{n}_t$. In this simulation, $\mathbf{n}_t$ was set to be Gaussian uncorrelated in space and white with a noise power of 0.1 ($1/10^{th}$ of the $\mathbf{x}_t$ powers). In this example, there are M=5 independent signals (2 noise and 3 data carrying sources) and only N=2 antenna elements. Figure 7.1 shows an example snapshot of the signals $\mathbf{y}_t$ which resulted during the simulation.

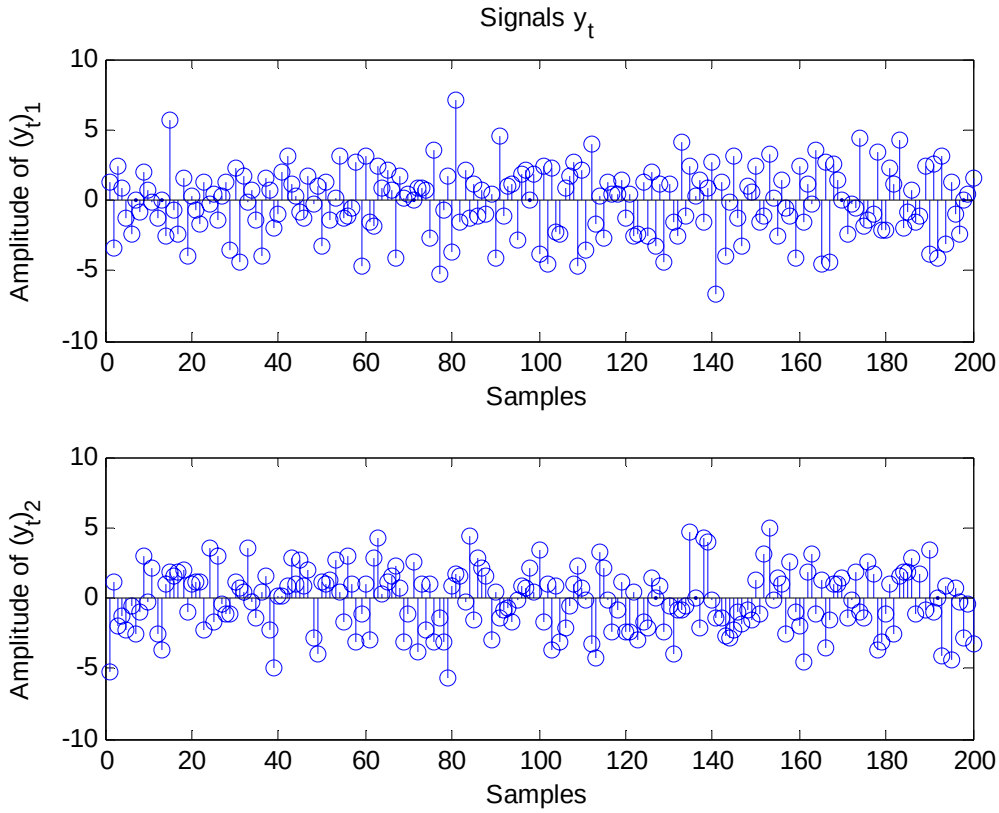

Figure 7.1 Snapshot of signals received at both antenna elements.

From just looking at the time domain functions of $\mathbf{y}_t$ given in Figure 7.1, it is not obvious what type or kind of signals $\mathbf{x}_t$ are lurking in the mixture or what the transfer function H_t is. The second order cyclic cumulant at the first antenna which can be computed from (7-7) and denoted by $\boldsymbol{C}_2^{\alpha}\left[(\mathbf{y}_t)_1;\boldsymbol{\tau}\right]=\left(\boldsymbol{C}_2^{\alpha}\left[\mathbf{y}_t;\boldsymbol{\tau}\right]\right)_{11}$, on the other hand, is more revealing. Figure 7.2 shows the magnitude of the second order cyclic cumulant of the signal received signal at the first antenna element which shows that the signals that make up $(\mathbf{y}_t)_1$ have cycle frequencies at α=-0.5, α=-1/3, α=0, and α=1/3. Recall that the cyclic cumulant is periodic and repeats after α=0.5 so a snapshot $\alpha\in$[-0.5,0.5) completely describes the system.

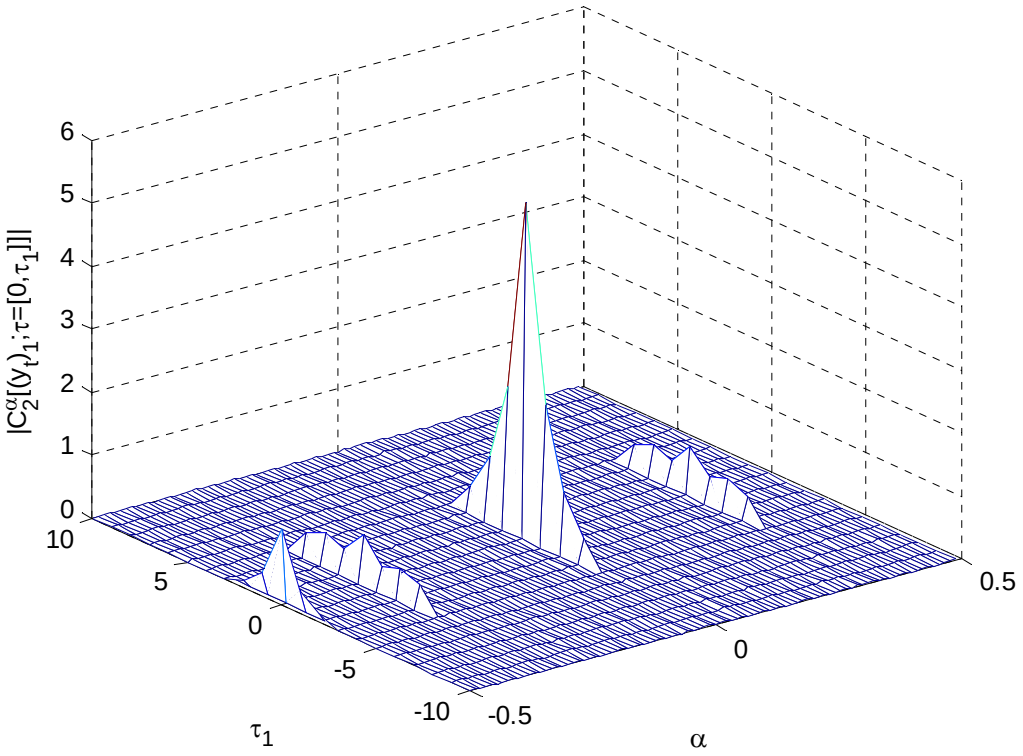

Figure 7.2. Plot of the second-order cyclic cumulant of the signal received at the first antenna element.

To follow the procedure outlined in this chapter, a subset of signals $\mathbf{y}_t^{(\alpha)}$ at the different cycle frequencies $\alpha \in \{-0.5,-1/3,0,1/3\}$ can be worked on one at a time and the transfer function $H_k^{(\alpha)}$ and statistics of $\mathbf{x}_t^{(\alpha)}$ identified as explained previously. To begin with the analysis, start with the subset of signals that have cycle frequencies at α=-0.5 ($\mathbf{x}_t^{(-0.5)}$ and $\mathbf{y}_t^{(-0.5)}$) to determine $H_k^{S(\alpha=-0.5)}$ (or $H_k^{S(-0.5)}$) and then $H_k^{ap(\alpha=-0.5)}$ (or $H_k^{ap(-0.5)}$) which correspond to the filters that make up the subset transfer function $H_k^{(\alpha=-0.5)}$ ($H_k^{(-0.5)}$). Using second-order statistics $D_k^{(-0.5)}$ can be computed as shown in (7-6) which when applied to $\mathbf{y}_t$ uncorrelates in time the subset signals that have cycle frequencies at α=-0.5, $\mathbf{i}_t^{(-0.5)}$. To also decorrelate the signals with cycle frequencies at α in space, the procedure

described in (7-10) and (7-11) estimates the matrix $H_0^{S\#(-0.5)}$ which when applied to $\mathbf{i}_t^{(-0.5)}$ results in the space-time uncorrelated set of signals $\mathbf{x}_t^{S(-0.5)}$. From the filter $D_k^{(-0.5)}$ and $H_0^{S\#(-0.5)}$ the inverse filter $H_k^{S\#(-0.5)}$ can be found and through (7-12) the $H_k^{S(-0.5)}$ based only on second-order statistics can be identified. This concludes the identification of the minimum-phase component of the subset filter $H_k^{S(-0.5)}$ which resulted in

$$\begin{aligned} H_0^{S(-0.5)} &= \begin{bmatrix} 1.23 & -0.54 \\ 0.59 & 1.13 \end{bmatrix}, H_1^{S(-0.5)} = \begin{bmatrix} 0.22 & 0.97 \\ -0.17 & -0.84 \end{bmatrix}, \\ H_2^{S(-0.5)} &= \begin{bmatrix} -0.16 & -0.62 \\ 0.57 & -0.30 \end{bmatrix}, H_3^{S(-0.5)} = \begin{bmatrix} 0.00 & -0.01 \\ 0.00 & 0.00 \end{bmatrix} \end{aligned} \tag{7-36}$$

All components of $H_k^{S(-0.5)}$ after k=4 have elements that are less than 0.01 so the filter can be truncated after k=3. Figure 7.3 shows the magnitude of the second-order cyclic cumulant of the first element of the signal, $\boldsymbol{C}_2^{\alpha}\left[\left(\mathbf{x}_t^S\right)_1;\boldsymbol{\tau}\right] = \left(\boldsymbol{C}_2^{\alpha}\left[\mathbf{x}_t^S;\boldsymbol{\tau}\right]\right)_{11}$, resulting after the filter $H_k^{S\#(-0.5)}$ has been applied to the received signal $\mathbf{y}_t$. The uncorrelated signal at α=-0.5 is evident by noting that the correlation is non-zero only when the delay τ_1=0 is unity as desired. Note that for linearly modulated signals, uncorrelating the signal at one of the cycle frequencies also uncorrelates the same signals at all its other cycle frequencies. Even though it is not evident from Figure 7.3, since all signals in the system have statistics at α=0, the same impulse present at α=-0.5 it also present at α=0. A similar observation can be made of the second-order cumulant of the second element of $\mathbf{x}_t^{S(\alpha)}$. The data rate of the two signals at α=-0.5 is therefore 0.5 or $T_i^{(-0.5)} = 2$ for $i \in \{1,2\}$. The phases of $\boldsymbol{C}_2^{(-0.5)}\left[\mathbf{x}_t^S;\boldsymbol{\tau} = [0,0]\right]$ can be used to find symbol timing information which corresponds to

$$T_{0,1}^{(-0.5)} = \frac{3.1416}{2\pi(-1/2)} = -1 \tag{7-37a}$$

$$T_{0,2}^{(-0.5)} = \frac{3.1416}{2\pi(-1/2)} = -1 \qquad \text{(7-37b)}$$

for both signals.

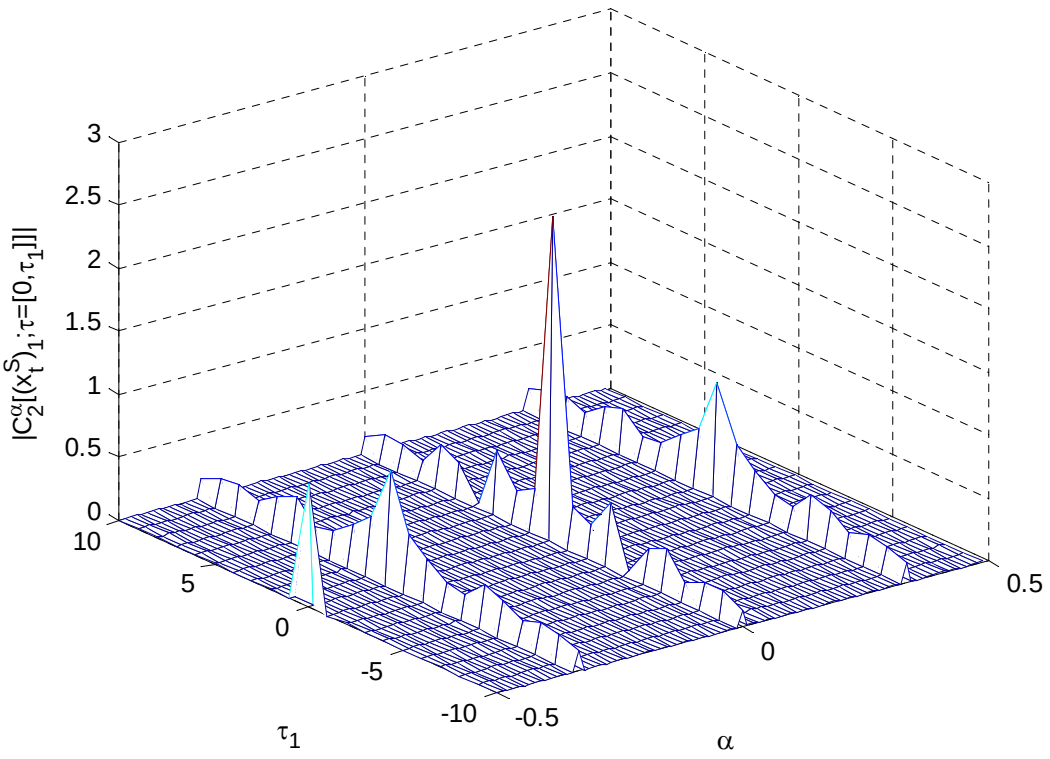

Figure 7.3 Plot of the second order cyclic cumulant of the signal received at $\boldsymbol{C}_2^{\alpha}\left[\left(\mathbf{x}_t^S\right)_1;\boldsymbol{\tau}\right]$ the first antenna element.

The all-pass component of $H_k^{(\alpha)}$ can also be identified using the procedure explained in section 7.3. As described in the procedure, start with the general 2x2 L=0 filter $\tilde{H}_{\mathrm{L}=0}^{\overline{ap}(-0.5)}$ and create the corresponding $\boldsymbol{C}_2^{-0.5}\left[\mathbf{x}_t^F;\boldsymbol{\tau}=[0,0,0,0]\right]$ which will be a function of x_0, y_0, and $\boldsymbol{C}_2^{-0.5}\left[\tilde{\mathbf{x}}_t^S;\boldsymbol{\tau}=[0,0,0,0]\right]$. The contrast function $\bar{\phi}_{ST}^{ap}\left[\mathbf{x}_t^{F(-0.5)},\mathrm{L}=0\right]$ results in a minimum of 0.0361 when x_0=2.20 and y_0=0. The process should then be repeated for a

2x2 L=1 filter $\tilde{H}_{\mathrm{L}=1}^{\overline{ap}(-0.5)}$ and the global minimum of $\bar{\phi}_{ST}^{ap}\left[\mathbf{x}_t^{F(-0.5)}, \mathrm{L}=1\right]$ for this filter is again 0.0361. Since the minimum of $\bar{\phi}_{ST}^{ap}\left[\mathbf{x}_t^{F(-0.5)}, \mathrm{L}=0\right]$ did not change from L=0 to L=1 the all-pass filter is of length L+1=1. The all-pass component $\tilde{H}^{ap(-0.5)}$ can therefore be computed to be (from x_0 and y_0 only)

$$\tilde{H}^{ap(-0.5)} = \begin{bmatrix} 0.41 & 0.91 \\ -0.91 & 0.41 \end{bmatrix} \tag{7-38}$$

for the signals at a cycle frequency α. Combining the filter $H_k^{ap(-0.5)}$ and $H_k^{S(-0.5)}$ produces the estimate of transfer function for the signals that have cycle frequencies at α which results in

$$\begin{aligned} H_0^{(-0.5)} &= \begin{bmatrix} 1.00 & 0.89 \\ -0.79 & 1.01 \end{bmatrix}, H_1^{(-0.5)} = \begin{bmatrix} -0.79 & 0.60 \\ 0.70 & -0.51 \end{bmatrix}, \\ H_2^{(-0.5)} &= \begin{bmatrix} 0.49 & -0.40 \\ 0.51 & 0.39 \end{bmatrix}, H_3^{(-0.5)} = \begin{bmatrix} 0.01 & 0.00 \\ 0.00 & 0.00 \end{bmatrix} \end{aligned} \tag{7-39}$$

All four components of $H_k^{(-0.5)}$ up to k=2 value have significantly large values implying that there are two signals at the cycle frequency α=-0.5 and their respective transfer functions to $\mathbf{y}_t$ can be described by three 2x2 matrices. Also, recall that the statistics of $\boldsymbol{C}_2^{(-0.5)}\left[\mathbf{x}_t^F; \boldsymbol{\tau}\right]$ correspond to the statistics of the original signals $\boldsymbol{C}_2^{(-0.5)}\left[\mathbf{x}_t; \boldsymbol{\tau}\right]$ from which the symbol statistics of $\mathbf{s}_t$ can be estimated. The cumulants of each of the signals can be used to estimate the distribution of the original signals transmitted by the sources. In order to illustrate the effectiveness of this method consider the cumulants of the second element of $\mathbf{x}_t^{F(-0.5)}$ which can be computed using equations in Chapter 6 or from [1]. The cumulants of the distribution of the samples $\mathbf{s}_t$ and the up-converted signal $\mathbf{x}_t$ were shown to be related by $\boldsymbol{C}_X^{(-0.5)}\left[\left(\mathbf{x}_t^F\right)_i; \boldsymbol{\tau}=[0,0]\right] * e^{-phase\left[\boldsymbol{C}_2^{(-0.5)}\left[\left(\mathbf{x}_t^F\right)_i; \boldsymbol{\tau}=[0,0]\right]\right]} * \mathrm{T}_i = \boldsymbol{C}_X^{(-0.5)}\left[\mathbf{s}_t\right]$ where T_i is the reciprocal of the data rate.

Table 7.1 Cumulants up to the order 6 of $\left(\mathbf{x}_t^{F(-0.5)}\right)_2$ compared to the true cumulants.

Actual Cumulants of $\left(\mathbf{x}_t^{(-0.5)}\right)_2$	**Estimated Cumulants of** $\left(\mathbf{x}_t^{F(-0.5)}\right)_2$	**Estimated Cumulants of** $(\mathbf{s}_t)_2$
$C_2^{(-0.5)}\left[(\mathbf{x}_t)_2;\boldsymbol{\tau}=[0,0]\right]=1$	$C_2^{(-0.5)}\left[\left(\mathbf{x}_t^F\right)_2;\boldsymbol{\tau}=[0,0]\right]=1$	$C_2\left[(\mathbf{s}_t)_2\right]=2$
$\left\{C_4^{(-0.5)}\left[(\mathbf{x}_t)_2;\boldsymbol{\tau}=[0,0]\right] * e^{-phase\left[C_2^{(-0.5)}\left[\left(\mathbf{x}_t^F\right)_i;\boldsymbol{\tau}=[0,0]\right]\right]}\right\}=-4$	$\left\{C_4^{(-0.5)}\left[\left(\mathbf{x}_t^F\right)_2;\boldsymbol{\tau}=[0,0]\right] * e^{-phase\left[C_2^{(-0.5)}\left[\left(\mathbf{x}_t^F\right)_i;\boldsymbol{\tau}=[0,0]\right]\right]}\right\}=-4.07$	$C_4\left[(\mathbf{s}_t)_2\right]=-8.14$
$\left\{C_6^{(-0.5)}\left[(\mathbf{x}_t)_2;\boldsymbol{\tau}=[0,0]\right] * e^{-phase\left[C_2^{(-0.5)}\left[\left(\mathbf{x}_t^F\right)_i;\boldsymbol{\tau}=[0,0]\right]\right]}\right\}=64$	$\left\{C_6^{(-0.5)}\left[\left(\mathbf{x}_t^F\right)_2;\boldsymbol{\tau}=[0,0]\right] * e^{-phase\left[C_2^{(-0.5)}\left[\left(\mathbf{x}_t^F\right)_i;\boldsymbol{\tau}=[0,0]\right]\right]}\right\}=68.4$	$C_6\left[(\mathbf{s}_t)_2\right]=136.80$
⋮	⋮	⋮

Table 7.1 lists the cumulants related to the second signal of the $\left(\mathbf{x}_t^{F(-0.5)}\right)_2$**.** First, the true cumulants of the signal $\left(\mathbf{x}_t^{(-0.5)}\right)_2$ are listed followed by the cumulants estimated at the cyclic frequency α=-0.5 in the simulation. The similarity between what was computed and the actual cumulants can be appreciated by comparing both these columns. Finally, the last column corresponds to the estimated cumulants of the distribution of the source samples $(\mathbf{s}_t)_2$ that make up the signal $\left(\mathbf{x}_t^{(-0.5)}\right)_2$ (Chapter 6)**.** In theory when all the moments or cumulants of all order of a signal are available, the probability density function can be reconstructed perfectly. The estimated distribution of $(\mathbf{s}_t)_2$ in Table 1 can be reconstructed using its moments and cumulants by ways of the Gram-Charlier series or the related Edgeworth series [2]. Figure 7.4 shows the reconstruction of the distribution using the Gram-Charlier series of the samples of the second signal $\left(\mathbf{x}_t^{F(-0.5)}\right)_2$ with statistics at α=-0.5 using the cumulants of up to the sixth order. For comparison purposes, the actual distribution of the transmitted signal is also plotted along side the estimated distribution to illustrate that the modulation of the signals can be reconstructed

from the cyclic statistics of $\mathbf{x}_t^F$. In this particular case, the signal $\left(\mathbf{x}_t^{F(-0.5)}\right)_2$ corresponds to a BPSK signal with samples that are $\left(\mathbf{s}_t\right)_2 \in \{\pm\sqrt{2}\}$, and from previous discussions, its data rate was determined to be 0.5 (T_2=2) with a symbol timing $T_{0,2}$=-1 with respect to the zero time axis.

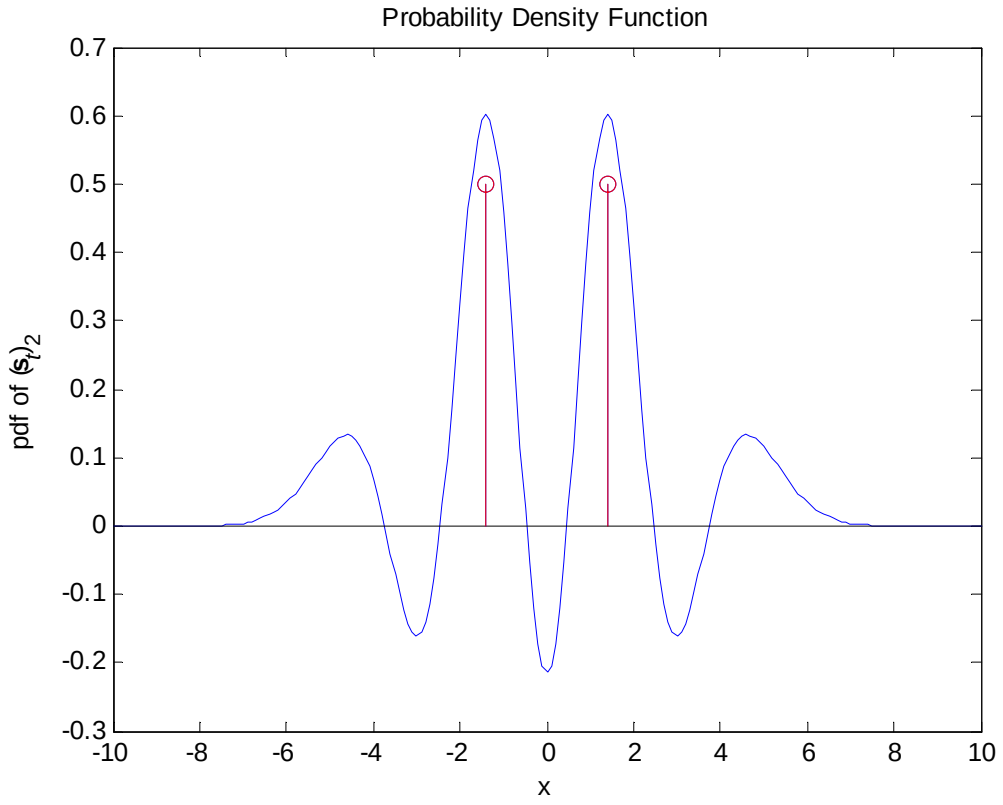

Figure 7.4 Plot of the reconstruction of the probability density function using cumulants up to the order 6.

The identification of the system can be continued by analyzing the signals that have cyclic statistics α=1/3. As before, the first step is to identify $H_k^{S(1/3)}$ based on second-order statistics and compute the corresponding $\mathbf{x}_t^S$ vector by first computing the innovations filter $D_k^{S(1/3)}$ and then the matrix $H_0^{S\#(1/3)}$ which uncorrelates the signals

$\mathbf{i}_t^{(1/3)}$ also in space. An examination of $C_2^{\alpha}\left[\mathbf{x}_t^S;\boldsymbol{\tau}=[0,\tau_1]\right]$, as in Figure 7.3, shows that uncorrelating the signal at α=1/3 also uncorrelates the signals at α=-1/3 and also, but not as obvious, at α=0. These cycle frequencies are therefore related which is characteristic of a 1/3 data rate signal. The transfer function for $H_k^{(1/3)}$ for signals that have cycle frequencies at α=1/3 was found to be

$$\begin{aligned} H_0^{(1/3)} &= \begin{bmatrix} 0.02 & 0.99 \\ -0.06 & 1.00 \end{bmatrix}, H_1^{(1/3)} = \begin{bmatrix} -0.04 & -0.69 \\ 0.04 & 0.80 \end{bmatrix}, \\ H_2^{(1/3)} &= \begin{bmatrix} -0.01 & 0.60 \\ 0.08 & 0.50 \end{bmatrix}, H_3^{(1/3)} = \begin{bmatrix} 0.08 & -0.40 \\ -0.02 & 0.30 \end{bmatrix}, \\ H_4^{(1/3)} &= \begin{bmatrix} 0.02 & 0.01 \\ 0.01 & 0.00 \end{bmatrix} \end{aligned} \tag{7-40}$$

Note that the first column of all components of $H_k^{(1/3)}$ has values that are significantly lower than its second column. Recall that the procedure outlined herein looks for N=2 signals in the system at the cycle frequency α=1/3 but in some cases the number of signals present at a specific cycle frequency will be less than N which results in filter coefficients that are close to zero for non-existing signals. From (7-40), it can be concluded that there is only a single signal with has a cycle frequency at 1/3 in the system which has filter coefficients which corresponds to the second column of $H_k^{(1/3)}$. The reciprocal of the data rate is, as previously discussed, equal to 3 and from the phase of the second-order statistics of $\mathbf{x}_t^S$ the symbol timing can be found to be

$$T_{0,2}^{(1/3)} = \frac{2.0486}{2\pi(1/3)} = 0.98 \Rightarrow 1 \tag{7-41}$$

In conclusion, three cyclostationary signals (more signals than antenna elements) have been identified in this system and the following transfer function was estimated for all the signals

$$\hat{H}_0 = \begin{bmatrix} 1.00 & 0.89 & 0.99 \\ -0.79 & 1.01 & 1.00 \end{bmatrix}, \hat{H}_1 = \begin{bmatrix} -0.79 & 0.60 & -0.69 \\ 0.70 & -0.51 & 0.80 \end{bmatrix},$$
$$\hat{H}_2 = \begin{bmatrix} 0.49 & -0.40 & 0.60 \\ 0.51 & 0.39 & 0.50 \end{bmatrix}, \hat{H}_3 = \begin{bmatrix} 0.01 & 0.00 & -0.40 \\ 0.00 & 0.00 & 0.30 \end{bmatrix} \quad (7\text{-}42)$$

The transfer function above is simply the concatenation of the partial transfer functions $H_k^{(-0.5)}$ in (7-39) and $H_k^{(1/3)}$ in (7-40). The first two signals have data that are spaced 2 (T_1,T_2) samples (data rate is 1/2) apart while the third component has new data every T_3=3 samples (1/3 data rate). The symbol timing for all three signals is given in (7-37) and (7-41) with respect to the zero time axis. As a comparison, the true transfer function used in this simulation was

$$H_0 = \begin{bmatrix} 1.00 & 0.90 & 1.00 \\ -0.80 & 1.00 & 1.00 \end{bmatrix}, H_1 = \begin{bmatrix} -0.80 & 0.60 & -0.70 \\ 0.70 & -0.50 & 0.80 \end{bmatrix},$$
$$H_2 = \begin{bmatrix} 0.50 & -0.40 & 0.60 \\ 0.50 & 0.40 & 0.50 \end{bmatrix}, H_3 = \begin{bmatrix} 0.00 & 0.00 & -0.40 \\ 0.00 & 0.00 & 0.30 \end{bmatrix} \quad (7\text{-}43)$$

which matches the estimated transfer function in (7-42) using cyclic statistics very well. Additionally, as estimated, the first and second signals were both transmitting at a data rate of 1/2 and the third signal at a data rate of 1/3. Lastly, the first signal belonged to a 4-level symbol alphabet $(\mathbf{s}_t)_1 \in \{-2/\sqrt{1.25}, -1/\sqrt{1.25}, 1/\sqrt{1.25}, 2/\sqrt{1.25}\}$, the second and third signals had members that belonged to a two-level symbol set with $(\mathbf{s}_t)_2 \in \{-\sqrt{2}, \sqrt{2}\}$ while $(\mathbf{s}_t)_3 \in \{-\sqrt{3}, \sqrt{3}\}$. These values were picked to produce signals $\mathbf{x}_t$ that had an average power of unity. The estimated distribution of $(\mathbf{s}_t)_2$ was found using the cyclic cumulants determined from the signal $\left(\mathbf{x}_t^{F(-0.5)}\right)_2$ which matched the probability density function of the actual signal that was transmitted. The remaining distributions can be found in a similar manner.

7.4.2 Stationary Signals

When the system contains purely stationary signals, the maximum number of signals that can be identified together with their channels is equal to the number of sensors N. Consider the system

$$\mathbf{y}_t = H_t \mathbf{x}_t \tag{7-44}$$

with a two-element array where $\mathbf{x}_t$ contains two signals that are stationary composed of $(\mathbf{x}_t)_1,(\mathbf{x}_t)_2 \in \{-1,1\}$ both with data rate equal to unity or $T_1,T_2=1$.

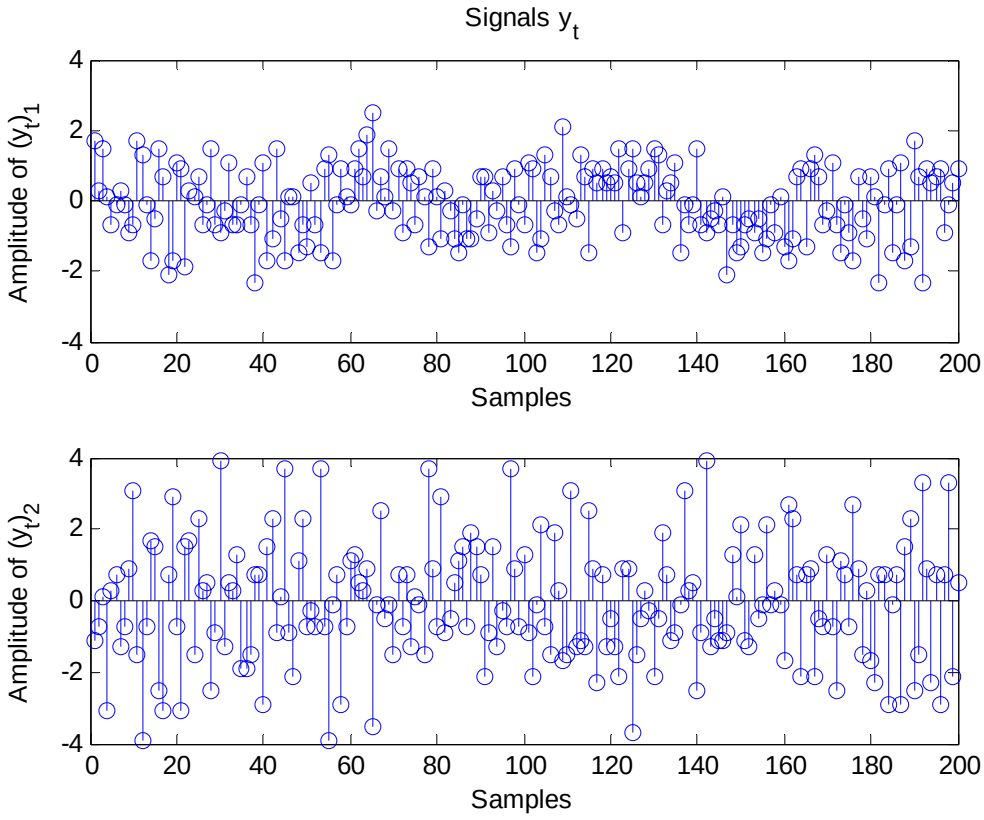

Figure 7.5 Snapshot of the signals received at both antenna elements.

The transfer function H_t was arbitrarily picked to be

$$H_0 = \begin{bmatrix} -0.50 & 1.00 \\ 0.30 & -0.70 \end{bmatrix}, H_1 = \begin{bmatrix} 0.60 & -0.30 \\ 0.70 & -0.10 \end{bmatrix},$$
$$H_2 = \begin{bmatrix} 0.70 & -0.60 \\ 0.10 & -0.40 \end{bmatrix}, H_3 = \begin{bmatrix} 0.20 & 0 \\ 0.20 & -0.30 \end{bmatrix} \tag{7-45}$$

for this example.

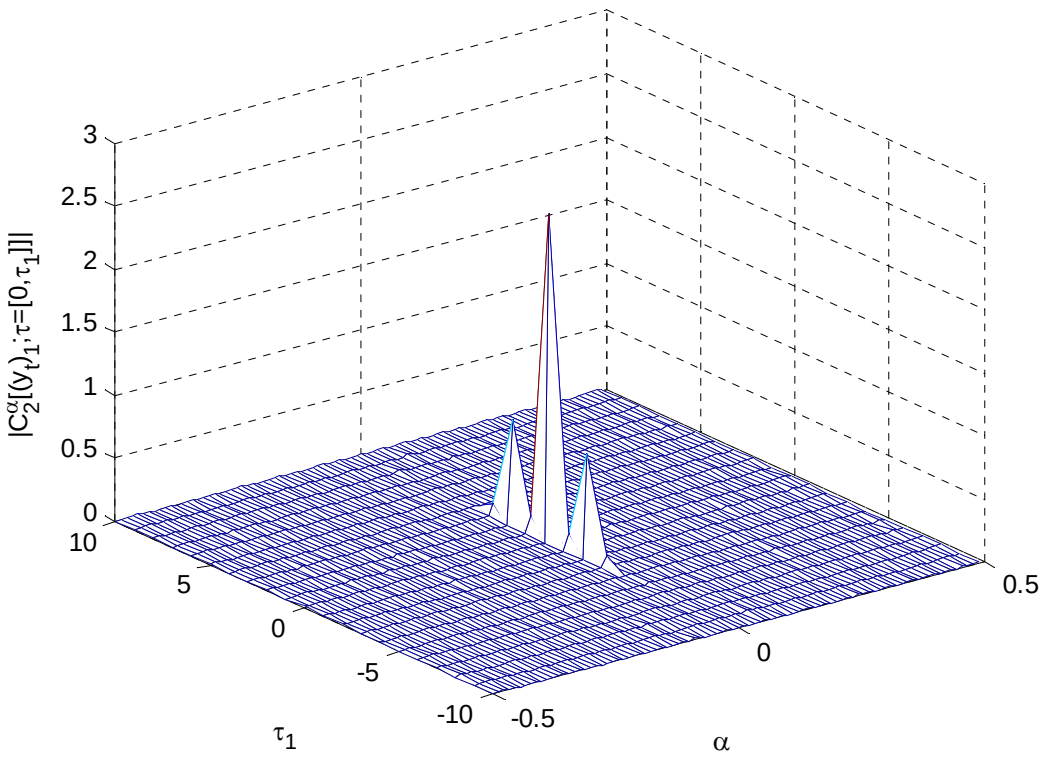

Figure 7.6 Plot of the second order cyclic cumulant of the signal received at the first antenna element.

Figure 7.5 shows a snapshot of the signal received at the antenna elements. As with the previous example the original source signals $\mathbf{x}_t$ and the transfer function H_t are not easily determinable by just looking at the signals $\mathbf{y}_t$. A better idea of the type of signals that are present in the system can be gotten from the second-order cyclic cumulant as was done in the cyclostationary example. The magnitude of the second-order cumulant (Figure 7.6) of the first element of $\mathbf{y}_t$, $\boldsymbol{C}_2^{\alpha}\left[(\mathbf{y}_t)_1;\boldsymbol{\tau}\right]=\left(\boldsymbol{C}_2^{\alpha}\left[\mathbf{y}_t;\boldsymbol{\tau}\right]\right)_{11}$, is non-zero only at a single cyclic cumulant at α=0 which demonstrates that the signals in the system are purely stationary.

The minimum-phase and all-pass components that make up the transfer function H_t can be determined using the procedures outlined in Sections 7.2 and 7.3 at the cycle frequency α=0 from the statistics of the received signal $\mathbf{y}_t$. The filters that make up H_t can be found to be

$$\begin{aligned} H_0^S &= \begin{bmatrix} -0.40 & -1.13 \\ -0.61 & 0.74 \end{bmatrix}, H_1^S = \begin{bmatrix} -0.38 & 0.52 \\ 0.10 & 0.39 \end{bmatrix}, \\ H_2^S &= \begin{bmatrix} -0.19 & 0.84 \\ -0.07 & 0.40 \end{bmatrix}, H_3^S = \begin{bmatrix} 0.01 & 0.09 \\ -0.01 & 0.36 \end{bmatrix} \end{aligned} \tag{7-46}$$

which corresponds to the minimum-phase component and the concatenated all-pass filter in (7-18) was found to be

$$\tilde{\tilde{H}}^{ap} = \begin{bmatrix} 0.01 & -0.03 & 0.90 & 0.43 \\ -0.43 & 0.90 & 0.03 & 0.01 \end{bmatrix} \tag{7-47}$$

Note that the all-pass component in this case is of length 2. Combining both filters in (7-46) and (7-47) arrives at an estimate of H_t which is given by

$$\begin{aligned} \hat{H}_0 &= \begin{bmatrix} 0.48 & -1.01 \\ -0.32 & 0.69 \end{bmatrix}, \hat{H}_1 = \begin{bmatrix} -0.62 & 0.29 \\ -0.70 & 0.10 \end{bmatrix}, \\ \hat{H}_2 &= \begin{bmatrix} -0.69 & 0.60 \\ -0.08 & 0.41 \end{bmatrix}, \hat{H}_3 = \begin{bmatrix} -0.18 & 0.01 \\ -0.20 & 0.30 \end{bmatrix} \end{aligned} \tag{7-48}$$

The estimated transfer function (7-48) is as expected similar to the actual transfer function (7-45) within the ambiguities of the identification (Section 6.1). Since the number of signals in the system is not greater than the number of antenna elements, the time variations of the independent signals can be recovered by inverting the channel. The signals $\mathbf{x}_t^F$ which should be similar to the signals $\mathbf{x}_t$ within (6-28) are plotted in Figure 7.7. These signals correspond to the independent signals present in this system and match the transmitted signals.

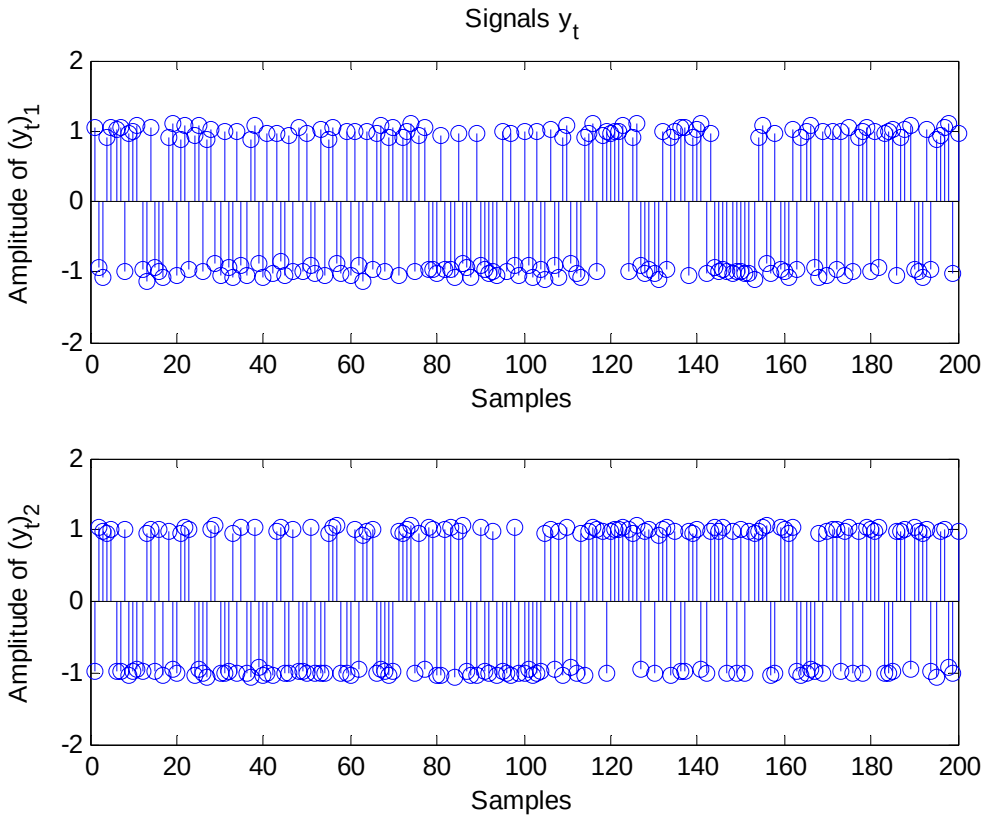

Figure 7.7 Snapshot of the independent signals $\mathbf{x}_t^F$ recovered from the system.

7.4.3 Comments on Computational Time

This section will briefly address the factors affecting the number of samples (T) that are necessary to achieve reliable estimates of the cyclic statistics. The algorithms presented in this dissertation rely on accurate estimates of the higher-order cumulants; obtaining dependable statistics is essential to accurate blind system identification. There does not exist a finite value on the size of sample space that would produce accurate statistics for all cases encountered in wireless communications. An estimate of T can be found, however, once a tolerance of the statistics as well as the environment of operation has been set. The required sample length is environment dependent in the sense that it depends on the number of signals in the system, the type of signals in the system (their periodicities), as well as their relative strengths. Also, the higher the order of the statistics the more time samples are required to achieve an accurate cyclic cumulant.

As T approaches infinity, the cyclic statistics for linearly modulated signals occur only at discrete frequencies which correspond to the cyclic periodicities of the signals present in the system and are zero everywhere else as in Figure 7.3. When a finite sample size is used to compute the cyclic cumulants there is inevitably a certain amount of energy that gets washed over to the adjacent cyclic frequencies around each of the discrete cyclic statistics. The amount of bleed over becomes smaller as T is increased and also decreases when moving away in α from the discrete cyclic frequencies of the signals. This is the reason that two signals can be identified and characterized easier in the single-input single-output case when using a finite T, for example, if their cyclic cumulants are further away from each other in the cyclic domain. Equivalently, the more unlike the signals (data rate and center frequency) are in the system the easier it is to differentiate them given the same sample size that the statistics are computed over.

Note also that the statistics of a signal drop in amplitude for each multiple of the data rate away from the center frequency. Recall that identification can be performed at any of these cyclic statistics and should therefore be performed at or as close to the center frequency of the signal as possible where the cyclic cumulant is the largest and with the

largest SNR. Identification based on a cyclic cumulant that is several multiples of the data rate away from its center frequency requires a larger sample size T to be accurate than using a cyclic cumulant closer to the center frequency. The value of T is therefore also dependent on what cyclic cumulants of the signals are used to fully identify the system. The powers of each of the respective signals, also plays a role in the identification process; weak signals require longer integration times (therefore longer T) to find and characterize than stronger signals.

As discussed above, the sample size that is required for accurate evaluation of the cumulants is a function of the amplitude, number, and structure of the signals in the environment. The analysis which involves estimating the value of T for numerous scenarios under different conditions is beyond the scope this dissertation but can be performed using the information presented herein. This type of analysis is left as a future work for the interested reader.

7.5 Summary of Procedure

This chapter describes how to compute the various filter coefficients introduced in Chapter 6 that are necessary in identifying the complete system. The process starts by first identifying the cycle frequencies for which the cyclic-cumulants of the received signals are non-zero (labeled as $\alpha_1, \ldots, \alpha_i, \ldots$). The signals are then separated based on their cyclic statistics and each non-zero cyclic statistic is analyzed at a time to identify the signals present at each cycle frequency together with their respective channels. The channel can be recovered through a two step procedure as describe by Figure 7.8. It starts first by identifying an inverse to the minimum-phase component of the channel ($H^{mp(\alpha_i)}(z)^{\#}$) using second-order cyclic cumulants followed by the identification of the inverse of the all-pass component ($H^{ap(\alpha_i)}(z)^{-1}$) using the fourth-order cyclic cumulants. Both of these functions are then inverted to obtain an estimate of the channel at each of the cycle frequencies ($H^{(\alpha_i)}(z)$). The filter coefficients of the set $\{H^{(\alpha_i)}(z)\}$ that correspond to the unique signals in the system can then be collected to form the total

channel response $H(z)$ which would identify the channel up to the ambiguities of the system.

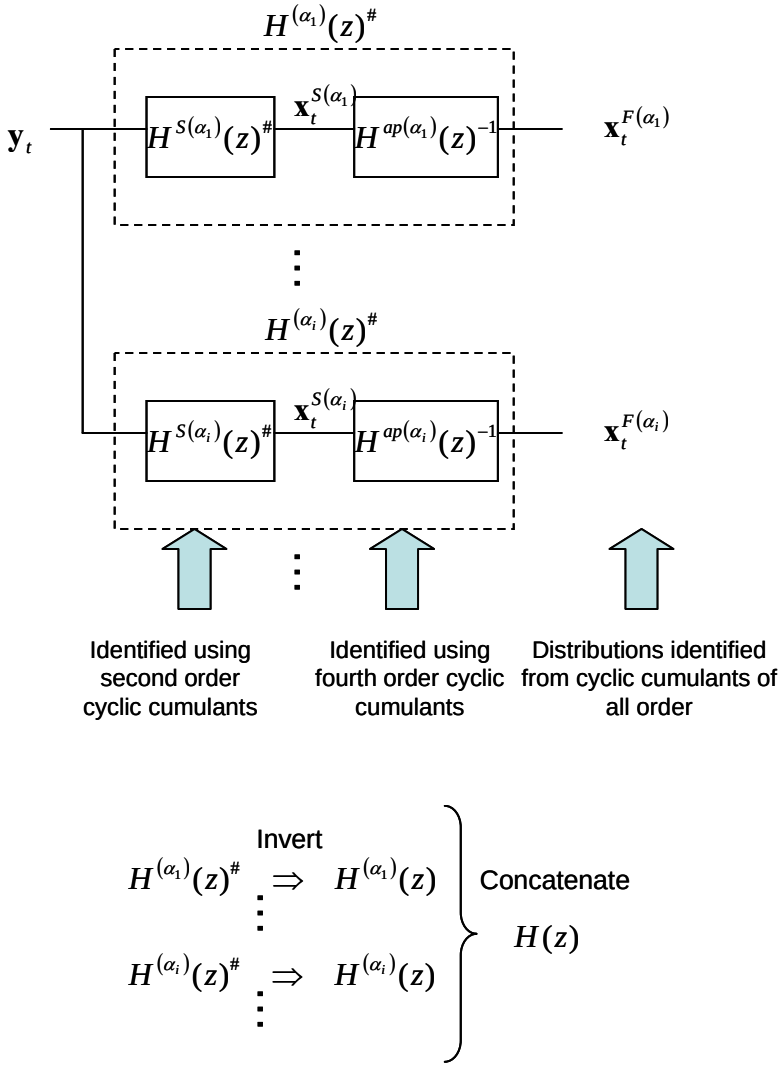

Figure 7.8 Block diagram of the system identification procedure using higher-order cyclic cumulants.

As shown in Chapter 6, the distributions of linearly modulated signals can be extracted at any of its cyclic statistics. A finite number of the higher-order cumulants of each signal in the system can be collected to approximate the signal's distribution using the Gram-Charlier expansion. Also, as long as T is large enough, the identification of

cyclostationary signals and their respective channel is unaffected by the strength of the stationary noise component since there will be a cycle frequency at $\alpha \neq 0$ where the statistics of the signal are present without the presence of the noise.

References

[1] C. M. Spooner, Theory and Application of Higher-Order Cyclostationarity, Ph.D. Dissertation University of California, Davis, June 1992.

[2] A. Stuart and J. K. Ord, *Kendall's Advanced Theory of Statistics Volume 1*, New York, 1994.

Chapter 8

Conclusions

In many cases, receivers are unaware and not designed to operate in the presence of other unknown systems occupying the same frequency spectrum. This, however, commonly occurs in the unlicensed band where multiple systems could potentially be using the same band of operation unknowingly causing interference to each other. As was shown in Chapter 2, knowledge of the interfering signals and how they affect the received signal can be used to improve the decisions on the signal of interest. A receiver that could identify the presence of other systems with no *a priori* knowledge could significantly improve its performance by accounting for interference. Identification of the environment together with the ability of a receiver to reconfigure itself based on learned information plays an important role in turning such a concept into reality.

This dissertation analyzes the blind identification of a multi-input multi-output (MIMO) system. The identification consists of determining the modulation - data rate and alphabet - of the source signals in the environment. Determining an estimate of the time dispersive MIMO channel, which relates all the signals in the system to the signals

sampled at the receiver, is also an important part of the identification process considered in this dissertation.

8.1 Blind Identification for Reconfigurable Receivers

Presently, most receivers are designed with the assumptions of no interference or have a limited ability to account for same system interference through elaborate pre-determined protocols. When other unknown systems are introduced into the environment, however, there is a significant degradation in system performance as a result of the unaccounted interference.

Receivers have also become more complex moving from the traditional single antenna based receivers to ones that employ multiple elements. An interfering environment, in general, should therefore be modeled as a MIMO system. On the input side are the signals in the system and at the output of the system are the signals received on all antenna elements at the receiver. The advantage of having multiple antennas on the receiver are numerous; the received signals can be combined to combat fading, cancel interference through beamforming (smart antennas), or for a more complicated processing a maximum-likelihood (ML) based reception can be used. These three different techniques are listed in order of complexity with the ML based receiver requiring the most *a priori* information.

Smart antennas are linearly based receivers that use beamforming to enhance the signal-of-interest (SOI) while at the same time diminishing the effects of the signals-not-of-interest (SNOI). The limitations of smart antennas lie in its linear based nature where the number of degrees of freedom is equal to the number of antenna elements. Beamforming is equivalent to inverting the effects of the channel (with a proper weighting to also minimize noise) to extract the SOI from the received mixture of all signals. A smart antenna based system can effectively cancel less number of interferers than antenna elements. There are limitations on the number of antennas that can be placed on certain small devices like mobile handsets. The number of elements in turn limits the number of

interferes that can be cancelled through beamforming. There are other non-linear based receivers which use maximum-likelihood estimation that allow for proper identification of the SOI even when the number of interferers outnumbers the antenna elements. Maximum-likelihood receivers use information of not only the channel but also the modulation of all signals in the system to improve the decisions. Using more information, in turn, allows for the identification of more signals than antenna elements on the receiver. Chapter 2 describes these two different types of receivers in more detail.

The usefulness of blind system identification lies in that it permits the receiver to learn from the environment and determine information that was previously unknown to it. The information learned about the signals and the channel in the system can be used to improve the signal reception. When the channel and modulation of the signals are known to the receiver an optimal decision can be made by the receiver to improve its performance. In this dissertation it was shown that all required information can be recovered blindly from the channel.

8.2 Summary and Contributions

In this dissertation a method has been presented that takes advantage of the structure of the digital signals to estimate the impulse response of the channel together with the modulation of the signals in the system. Digital signals used in wireless communication systems have cyclostationary properties with periodic statistics that are a function of data rate, center frequency, and other modulation induced periodicities. More signals than antenna elements can be identified with their channels by taking advantage of the different modulations of the signals in the system.

The theory developed in this dissertation which searches for independent components based on cumulant tensors can be viewed as a generalization of independent component analysis (ICA) to cyclostationary signals. ICA has been developed over the last decade mostly addressing stationary signals. The development of independent component

analysis is discussed in Chapter 4 and can be found in for example [1], [2], and the references herein.

The method developed in this dissertation is based on the assumptions that the signals of the different transmitters are independent of each other. The technique is based on first identifying the inverse channels belonging to all subset of signals that make up the received signals with the same cyclostationary statistics. The inverse channel corresponds to the filter that forces the cyclic cross-cumulants of the signals at a given cycle frequency to be zero. Diagonalization of the cyclic-cumulant tensors (minimization of the cyclic cross-cumulants) is equivalent to minimizing the mutual information of the subset signals which is characteristic of independence which is the property that is being sought. The invertibility of the MIMO channel is therefore an important aspect of blind signal identification and is considered in Chapter 5 where the decomposition of MIMO channels into a minimum-phase and an all-pass component is addressed.

This dissertation focuses on linearly modulated signals to demonstrate the effectiveness of cyclic cumulant based identification of the channel. Chapter 6 describes the identification process as a two step procedure which is performed on all non-zero cyclic statistics one at a time. The second-order cyclic cumulants are used to determine the minimum-phase component of the channel. In Chapter 6 and 7 it was shown that an innovations based decomposition can be used to arrive at the minimum-phase component of the transfer function. Complete identification of the channel of a system with space-time independent signals cannot be performed simply through second-order statistics since these are blind to MIMO all-pass filters. Higher-order statistics, which do not have this downfall, can be used to identify the remaining component of the channel. The fourth-order cumulant tensor was used to determine the MIMO all-pass filter to complete the identification of the channel. The cyclic cumulant at each cycle frequency were found to contain also the cumulants of the original source signals from which the modulation of the source signals can be estimated as described in Chapter 6 and 7.

Finally, Chapter 7 describes algorithms that can be used compute the various channel coefficients given in Chapter 6 that are needed to identify the MIMO channel. Chapter 7 also shows how to recover the signal parameters such as data rate, distribution, and symbol timing. Examples are considered in this chapter to illustrate the effectiveness of the algorithms in the identification of more signals than antenna elements. The methods described in this dissertation are also applicable to stationary signals since these signals can be thought of as cyclostationary signals with only a single statistical cycle frequencies at $\alpha=0$.

8.3 Future Work

There are several possibilities of future work in this area. There is significant work that can be done in the development of efficient algorithms, which might be suboptimal, but can guarantee convergence for less number of samples and/or require fewer computations. Higher-order based signal processing applications require significantly more samples to achieve accurate results than second-order based methods. A related problem could therefore be to perform an analysis on the number of samples that are required to compute higher-order cyclic statistics within a certain tolerance and under different conditions (see Section 7.4.3 for details).

The next step in the analysis is to apply the algorithms developed in this dissertation to actual measured data. Computer based random generators might not simulate transmitters that are truly independent very well since these signals are generated from the same source (usually the internal clock) and therefore difficult to make truly independent. There might therefore be discrepancies in the number of samples that are required for computing higher-order cyclic cumulants in measurements as opposed to simulation. An experiment could be set up in a controlled fashion by turning each system on one at a time and identifying the corresponding channel. The estimated channel can be used as a control when doing the joint estimations of all the transmitters in the system.

The theory in this dissertation was developed for static channels and a contribution could also be made by making the algorithms proposed in this dissertation adaptive to track small changes in the channel. For mobile devices, the ability for the receiver to be adaptive is a must to rake in the full benefits of this method. Potential problems can occur if the number of samples needed to identify the system exceed the time-coherence (the time that the channel can be modeled as static) of the channel. This means that the receiver is unable to collect enough data to accurately estimate the MIMO system because the channel is changing too fast. The sample size T that is necessary to estimate the channel is therefore an important aspect when it comes to non-static channels and the factors affecting this value are discussed in Section 7.4.3.

A related area of research, in which a contribution can be made, is in the development of reconfigurable receivers. Software defined radios are capable of changing their architecture because on their ability to be reprogrammed. Software can be written for the receiver that can both learn from the channel and at the same time incorporate this knowledge into its decision making. The developments of programs that are able to change the receiver architecture are an important part of making the concept of smart receivers a reality.

References

[1] J.-F. Cardoso, "Blind Signal Separation: Statistical Principles," *Proc. of the IEEE*, vol. 86, no. 10, October 1998, pp. 2009-2025.

[2] P. Common, "Independent Component Analysis, A New Concept?," *Signal Processing*, Special Issue *Higher Order Statistics*, Vol. 36, No. 3, April 1994, pp. 287-314.

Appendix

A.1

The optimal non-linear estimate for the system in (2-9) where **n** has identically independent and unit variance Gaussian distributed components and the elements of **s** are either all 1's or -1's is found in this section. Both values of **s** are assumed equally likely.

Recall from (2-5) that $\hat{\mathbf{s}} = E[\mathbf{s} \mid \mathbf{y}]$ is optimal estimate when the distributions are known and can be computed from

$$\hat{\mathbf{s}} = E[\mathbf{s} \mid \mathbf{y}] = \int_{-\infty}^{\infty} \mathbf{s} f_{\mathbf{s}|\mathbf{y}}(\mathbf{s} \mid \mathbf{y}) d\mathbf{s} \tag{A-1}$$

The distribution in the integration can be factored into

$$f_{\mathbf{s}|\mathbf{y}}(\mathbf{s} \mid \mathbf{y}) = \frac{f_{\mathbf{y}|\mathbf{s}}(\mathbf{y} \mid \mathbf{s}) f_{\mathbf{s}}(\mathbf{s})}{f_{\mathbf{y}}(\mathbf{y})} \tag{A-2}$$

where the individual distributions that make up (A-2) can be found to be

$$f_{\mathbf{y}|\mathbf{s}}(\mathbf{y} \mid \mathbf{s}) = \frac{e^{-\frac{(\mathbf{y}-H\mathbf{s})^T(\mathbf{y}-H\mathbf{s})}{2}}}{\left(\sqrt{2\pi}\right)^{\mathbf{N}}} \tag{A-3}$$

$$f_{\mathbf{y}}(\mathbf{y}) = \frac{0.5}{\left(\sqrt{2\pi}\right)^{\mathrm{N}}}\left(e^{-\frac{(\mathbf{y}-H\begin{bmatrix}1\\1\\\vdots\end{bmatrix})^T(\mathbf{y}-H\begin{bmatrix}1\\1\\\vdots\end{bmatrix})}{2}} + e^{-\frac{(\mathbf{y}+H\begin{bmatrix}1\\1\\\vdots\end{bmatrix})^T(\mathbf{y}+H\begin{bmatrix}1\\1\\\vdots\end{bmatrix})}{2}} \right) \tag{A-4}$$

$$f_{\mathbf{s}}(\mathbf{s}) = 0.5\delta(\mathbf{s}-1) + 0.5\delta(\mathbf{s}+1) \tag{A-5}$$

Replacing (A-3), (A-4), and (A-5) into (A-1) and integrating produces

$$\hat{\mathbf{s}} = E[\mathbf{s} \mid \mathbf{y}] = \frac{e^{-\frac{1}{2}(\mathbf{y}-H\begin{bmatrix}1\\1\\\vdots\end{bmatrix})^T(\mathbf{y}-H\begin{bmatrix}1\\1\\\vdots\end{bmatrix})} - e^{-\frac{1}{2}(\mathbf{y}+H\begin{bmatrix}1\\1\\\vdots\end{bmatrix})^T(\mathbf{y}+H\begin{bmatrix}1\\1\\\vdots\end{bmatrix})}}{e^{-\frac{1}{2}(\mathbf{y}-H\begin{bmatrix}1\\1\\\vdots\end{bmatrix})^T(\mathbf{y}-H\begin{bmatrix}1\\1\\\vdots\end{bmatrix})} + e^{-\frac{1}{2}(\mathbf{y}+H\begin{bmatrix}1\\1\\\vdots\end{bmatrix})^T(\mathbf{y}+H\begin{bmatrix}1\\1\\\vdots\end{bmatrix})}} \tag{A-6b}$$

$$\hat{\mathbf{s}} = E[\mathbf{s} \mid \mathbf{y}] = \frac{e^{\frac{1}{2}\left(\left(H\begin{bmatrix}1\\1\\\vdots\end{bmatrix}\right)^T\mathbf{y}+\mathbf{y}^T\left(H\begin{bmatrix}1\\1\\\vdots\end{bmatrix}\right)\right)} - e^{-\frac{1}{2}\left(\left(H\begin{bmatrix}1\\1\\\vdots\end{bmatrix}\right)^T\mathbf{y}+\mathbf{y}^T\left(H\begin{bmatrix}1\\1\\\vdots\end{bmatrix}\right)\right)}}{e^{\frac{1}{2}\left(\left(H\begin{bmatrix}1\\1\\\vdots\end{bmatrix}\right)^T\mathbf{y}+\mathbf{y}^T\left(H\begin{bmatrix}1\\1\\\vdots\end{bmatrix}\right)\right)} + e^{\frac{1}{2}\left(\left(H\begin{bmatrix}1\\1\\\vdots\end{bmatrix}\right)^T\mathbf{y}+\mathbf{y}^T\left(H\begin{bmatrix}1\\1\\\vdots\end{bmatrix}\right)\right)}} \tag{A-6b}$$

In order to simplify the equation above it is convenient to introduce the change of variables

$$h = H\begin{bmatrix}1\\1\\\vdots\end{bmatrix} \tag{A-7}$$

where *h* is a column vector with elements that correspond to the summations of rows of the matrix *H*. The final result can be reached by noting that the numerator in (A-6b) is in

the form of a hyperbolic sine while the denominator has the same operand but can be expressed as a hyperbolic cosine resulting in the final answer

$$\hat{\mathbf{s}} = \tanh\left[\frac{1}{2}\left(h^T\mathbf{y} + \mathbf{y}^T h\right)\right] \tag{A-8}$$

A.2

When the noise signals in **n** are correlated, the decision statistics $\mathbf{y_H}$ and $\mathbf{s_H}$ have to be adjusted accordingly. Recall from (2-14) that the noise auto-correlation can be factored into

$$R_n = LL^* \tag{A-9}$$

so applying the matrix L^{-1} to both sides of the equation uncorrelates the noise in the system

$$\begin{aligned} L^{-1}\mathbf{y} &= L^{-1}H\mathbf{s} + L^{-1}\mathbf{n} \\ \tilde{\mathbf{y}} &= \tilde{H}\mathbf{s} + \tilde{\mathbf{n}} \end{aligned} \tag{A-10}$$

where the auto-correlation of $\tilde{\mathbf{n}}$ is the identity matrix. Now finding the projection of $\tilde{\mathbf{y}}$ onto the range of $\tilde{H}$ as in (2-21) results in

$$\begin{aligned} \tilde{\mathbf{y}}_{\mathbf{H}} &= \tilde{H}\left(\tilde{H}^*\tilde{H}\right)^{-1}\tilde{H}^*\tilde{\mathbf{y}} \\ L^{-1}\mathbf{y}_{\mathbf{H,n}} &= L^{-1}H\left(H^*L^{-1,*}L^{-1}H\right)^{-1}H^*L^{-1,*}L^{-1}\mathbf{y} \end{aligned} \tag{A-11}$$

which is equivalent to performing the following adjusted projection on the signal **y** weighted by auto-correlation of R_n

$$\mathbf{y}_{\mathbf{H,n}} = H\left(H^*R_n{}^{-1}H\right)^{-1}H^*R_n{}^{-1}\mathbf{y} \tag{A-12}$$

Vita

Kai Dietze was born on February 16, 1974 in Frankfurt/Main, Germany. He received his Bachelors of Science in Electrical Engineering in May 1995 from Michigan Technological University, Houghton, Michigan. In May 2001 and December 2005, respectively, he obtained his Masters of Science and Ph.D. degrees in Electrical Engineering from Virginia Tech, Blacksburg, Virginia.

Kai worked as a Research Assistant at the Virginia Tech Antenna Group (VTAG) where he has been involved in developing and testing smart multiple antenna receive and transmit diversity systems for wideband communications. His interests are in signal processing when applied to wireless communications systems. His current research focuses on the areas of adaptive antenna arrays, diversity systems, receiver design, as well as blind system identification.