

BLIND MULTIUSER DETECTION BASED ON SECOND-ORDER STATISTICS

A Dissertation

Presented to

The Academic Faculty

by

Richard Todd Causey

In Partial Fullfillment

of the Requirements for the Degree of

Doctor of Philosophy in Elecrical Engineering



School of Electrical and Computer Engineering

Georgia Institute of Technology

Atlanta, Georgia

July 30, 1999

Copyright © 1999 by Richard Todd Causey

BLIND MULTIUSER DETECTION BASED ON SECOND-ORDER STATISTICS

Approved:

Professor John R. Barry, Chairman

Professor Mary Ann Ingram

Professor Douglas B. Williams

Date Approved by Chairman: _____

To my wife Gina.

Acknowledgments

I would first of all like to thank my advisor Professor John Barry for his invaluable guidance throughout this journey. I am so very proud to have had the opportunity to work with such an outstanding researcher and teacher.

I am also grateful to those who served on my defense committee: Professors Mary Ann Ingram, Douglas Williams, Steven McLaughlin, and Carl Spruill. I especially appreciate the efforts of Professors Ingram and Williams who agreed to undertake the additional task of reading this thesis. I would like to thank those who served in other capacities as well, either on my proposal or qualifying committees: Professors Stephen B. Wicker, Mark Smith, and John Limb. A special thanks goes to Professor Dale Ray.

I would like to acknowledge the Center for Research in Applied Signal Processing, Hewlett-Packard Labs, and the School of Electrical and Computer Engineering for their financial support of this research.

I would like to thank my friends and colleagues in the Communication Theory Group: Anuj Batra, Alex Yeh, Hyuncheol Park, Renato Lopes, and Abdallah Alahmari.

I would like to thank Dr. Jerry Moore for inspiring me with his example.

I am eternally grateful to my mother and father for the love and support they have given me throughout my life.

Finally, I would like to express my love and appreciation to my wife Gina. She has sacrificed much for me. Her encouragement kept me moving forward. I thank God for her. It is to Gina that I dedicate this thesis, because it would not exist without her smile.

Table of Contents

List of Figures	viii
Summary	xii
1 INTRODUCTION	1
2 BACKGROUND	9
2.1 Mathematical Preliminaries	10
2.1.1 A Review of MIMO System Theory	10
2.1.2 Linear Detectability — Discrete-Time Channels	12
2.1.3 Derivation of the Discrete-Time Channel Model	17
2.1.4 The Generalized Nyquist Criterion	19
2.1.5 Linear Detectability — Continuous-Time Channels	20
2.1.6 Performance Measures for Multiuser Detectors	22
2.2 A Survey of Related Prior Work	24
2.2.1 Classical Blind Equalization	24
2.2.2 Blind Equalization of Cyclostationary Sequences	28
2.2.3 Blind Multiuser Detection	31
3 SUBSPACE METHODS	35
3.1 Signal and Noise Subspaces	36
3.2 An Adaptive Signal-Noise Subspace Separator	42
3.3 Adaptive Singular-Value Decomposition	47
3.4 MMSE Detection	52
3.5 Zero-Forcing Detection	56
3.6 Channel Diagonalization	58

3.7	Chapter Summary	63
	Appendix 3.1: Derivation of the MMSE Linear Multiuser Detector	66
	Appendix 3.2: On the Convergence of (3-25) and (3-20)	68
4	THE WHITEN-ROTATE DETECTOR	71
4.1	Whiten-Rotate Detection	72
4.2	A Project-First Architecture	76
4.3	Blind Adaptive Implementations	79
4.3.1	An Adaptive Whitener	79
4.3.2	An Adaptive Rotator	81
4.4	Experimental Results	83
4.5	Chapter Summary	89
	Appendix 4.1: Proof of Lemma 4-1	91
	Appendix 4.2: Proof of Lemma 4-2	92
	Appendix 4.3: Proof of Lemma 4-3	93
	Appendix 4.4: Proof of Theorem 4-2	95
	Appendix 4.5: A Whiten-Rotate Detector for Noiseless Channels	96
5	LINEAR PREDICTION	98
5.1	Equivalent Channel Models	99
5.1.1	An Autoregressive Channel Model	100
5.1.2	An ARMA Channel Model	103
5.1.3	An Autoprogressive Model	104
5.2	Minimum-Phase Channels	106
5.3	Necessary and Sufficient Conditions	109
5.4	Temporal Linear Prediction	112
5.4.1	One-Step Prediction	113
5.4.2	Multiple-Step Prediction	114

5.4.3	Backward Prediction	115
5.5	Multiuser Detection Using Linear Prediction	117
5.5.1	The Forward LP Detector	117
5.5.2	The Forward-Backward LP Detector	133
5.6	Blind Adaptive Implementations	143
5.7	Experimental Results	147
5.8	Chapter Summary	154
Appendix 5.1:	Proof of Theorem 5-2 and Corollary 5-2.1	156
Appendix 5.2:	Proof of Theorem 5-3 and Corollary 5-3.1	157
Appendix 5.3:	Proof of Theorems 5-8, 5-10, and Corollaries	158
Appendix 5.4:	Derivation of Lemma 5-4	161
Appendix 5.5:	Proof of Lemma 5-5	162
Appendix 5.6:	Proof of Theorem 5-9	164
Appendix 5.7:	Proof of Theorem 5-11	166
Appendix 5.8:	A Backward LP Detector	168
6	STACKED DETECTORS	171
6.1	A Memoryless Channel Model	172
6.2	The Stacked MMSE Detector	174
6.3	The Stacked Zero-Forcing Detector	177
6.4	The Stacked Whiten-Rotate Detector	179
6.5	Signal and Noise Subspaces	190
6.6	Channel Diagonalization and Lossless Precoding	193
6.7	Adaptive Implementations	196
6.7.1	An Adaptive Stacked MMSE or ZF Detector	196
6.7.2	An Adaptive Stacked WR Detector	198
6.7.3	An Adaptive Subspace Separator	199
6.7.4	Selection of the Detector Outputs	200

6.8	Experimental Results	201
6.9	Chapter Summary	205
	Appendix 6.1: An Algorithm for Resolving the Permutation \mathcal{K}	209
7	CONCLUSIONS AND FUTURE WORK	213
7.1	Conclusions	213
7.2	Future Research	216
7.2.1	SVD Convergence	217
7.2.2	Blind Estimation of the LP Index	218
7.2.3	Alternative LP-Based Architectures	221
7.2.4	Fading Channels	222
7.2.5	Correlation Matching	222
	BIBLIOGRAPHY	224
	VITA	239

List of Figures

1-1	A wireless multiuser communication system and discrete-time multiple-input multiple-output (MIMO) channel model.	4
2-1	A successive cancellation detector for Example 2-1.	14
2-2	(a) Original $p \times n$ continuous-time channel and oversampling receiver front end; (b) equivalent $m \times n$ baud-rate discrete-time channel model.	18
2-3	The Bussgang structure for blind equalization.	25
2-4	Donoho's minimum-entropy concept.	27
2-5	Equivalent FIR channel models: (a) an upsampled SISO channel, (b) a SIMO channel, (c) a filterbank.	29
3-1	Equivalent models: (a) a tall channel with a signal-space projector used as the receiver front end, and (b) an equivalent square channel.	42
3-2	An adaptive signal-noise subspace separator.	44
3-3	Convergence of the subspace separator: energy in the rows of the separator channel cascade $\hat{\Theta}_k \mathbf{H}$ versus time k .	46
3-4	Adaptive estimation of singular values: (a) the diagonal elements of $\hat{\mathbf{U}}_k^* \Phi_r \hat{\mathbf{U}}_k$; (b) eigenvalue estimates from (3-29).	51
3-5	A fully blind, adaptive implementation of the minimum-MSE multiuser detector.	54
3-6	The blind adaptive MMSE detector of Experiment 3-3: (a) an MSE learning curve for user 1; (b) the constellations from the last trial, baud 4000 to 5000.	57

3-7	Equivalent models: (a) a memoryless MIMO channel with precoder and front-end rotation filter, and (b) decoupled scalar channels.	60
3-8	A block diagram of adaptive channel-diagonalization algorithm.	61
3-9	Convergence of the diagonalization algorithm: diagonalization metric ζ_k versus time k .	64
4-1	The structure of the whiten-rotate detector as applied to a memoryless multiuser channel.	73
4-2	A comparison of the whiten-rotate detector with the minimum-MSE detector: SNR versus MSE.	77
4-3	An adaptive algorithm for implementing the MMSE rotator of (4-2).	82
4-4	A functional diagram of an adaptive project-first whiten-rotate detector.	84
4-5	Convergence of the project-first adaptive algorithm of Fig. 4-4, showing contributions to MSE from each stage.	86
4-6	The adaptive project-first WR detector applied to a linear antenna array: (a) the MSE learning curve; (b) constellations from the last trial, baud 4000 to 5000.	87
4-7	The adaptive project-first WR detector applied to a synchronous CDMA system: (a) an MSE learning curve; (b) constellations from the last trial, baud 4000 to 5000.	88
5-1	Equivalent models for a tall MIMO channel: (a) moving average (MA) and (b) autoregressive (AR).	101
5-2	An ARMA model for a tall MIMO channel.	103
5-3	An autopgressive model for a tall MIMO channel.	105
5-4	A block diagram of the forward LP detector.	118
5-5	A comparison of the zero-delay FLP and MMSE detectors.	124

5-6	Performance of the zero-delay FLP detector as a function of $\gamma_0^{(1)}$, the fraction of user 1's energy in the zero-th tap.	126
5-7	Comparison of the delay-1 and delay-0 performance as a function of the zero-th tap energy fraction $\gamma_0^{(1)}$.	130
5-8	Comparison of the delay-0 and delay-1 FLP detectors as a function of the energy fraction $\gamma_0^{(1)}$ for 2-tap channels with fixed SNR.	131
5-9	Comparison of several LP-based detectors as a function of the energy fraction $\gamma_0^{(1)}$ for 2-tap channels with fixed SNR.	139
5-10	A comparison of the optimal LP detector with an MMSE detector with equal memory.	141
5-11	A Blind Adaptive Implementation of the LP Detector.	144
5-12	Convergence of LP versus CMA: (a) unshaped 64-QAM; (b) moderately shaped; and (c) heavily shaped.	151
5-13	(a) A two-user asynchronous CDMA system with a chip-rate sampling receiver; (b) An equivalent MIMO FIR channel. (b) equivalent tall (32×2) MIMO model.	153
5-14	Recovered constellations for the asynchronous CDMA system of Experiment 5-8.	153
6-1	Equivalent models: (a) an FIR channel followed by receiver stacking, and (b) a memoryless block-Toeplitz channel preceded by transmitter stacking.	173
6-2	A block diagram of the stacked MMSE detector.	177
6-3	An interpretation of the stacked WR detector showing its relationship to the spatio-temporal prediction-based detectors.	183
6-4	Comparison of the stacked detectors for a single random channel.	186
6-5	Comparison of the stacked detectors: MSE versus delay.	187

6-6	Performance of the stacked detectors as a function of the stacking depth N .	189
6-7	Average MSE penalty of the stacked WR detector relative to the stacked MMSE detector.	191
6-8	(a) A channel-diagonalization technique for tall FIR channels; (b) Equivalent diagonal channel model.	194
6-9	Estimates of the eigenvalues of $\Phi_{\mathbf{R}}$ in Experiment 6-5 produced by the recursion of (6-30) at steady state.	203
6-10	Outputs of the stacked MMSE detector as a function of the signal subspace dimension estimate for Experiment 6-5.	204
6-11	Best outputs of stacked MMSE detector for the asynchronous-CDMA application of Experiment 6-6.	206

Summary

Multiuser detection is the process of recovering information from mutually interfering users of a shared communication channel. Typical applications include wireless networks, bundled cables, and multitrack magnetic recording systems. A multiuser detector exploits the structure of the multiuser interference in order to improve system performance or capacity. Because users sharing the channel usually operate autonomously, it is often desirable for a multiuser detector to function blindly, with no *a priori* knowledge of the channel nor any transmitter cooperation. This thesis addresses the problem of blind multiuser detection.

The research focuses on tall channels, those having more outputs than inputs, because such channels are identifiable up to a memoryless unitary matrix from the second-order statistics of their output. Tall channels arise in a wide variety of applications, including multisensor receivers and code-division multiple access. This research develops adaptive detectors based on linear prediction and subspace decomposition, second-order techniques offering significant advantages over higher-order approaches. These advantages include fast, reliable convergence, low computational complexity, and inherent compatibility with shaped constellations having near-Gaussian distributions.

We first address the special case of memoryless channels. We first propose a novel adaptive signal-noise subspace separator can be used as the front end of any receiver to reduce the complexity of subsequent processing. We extend the technique to perform singular-value decompositions (SVDs) adaptively, and propose blind implementations of the minimum-mean-square-error (MMSE) and zero-forcing (ZF) detectors based on the adap-

tive SVD. We also propose a canonical whiten-rotate (WR) detector offering near-MMSE performance. The adaptive implementation is based on spatial prediction and has very low complexity. For single-user multichannel applications, we propose an adaptive channel-diagonalization algorithm that facilitates transmission approaching capacity.

For channels with memory, we propose a family of detectors based on spatio-temporal prediction. These detectors can be viewed as a generalization of the whiten-rotate detector; they first use one or more temporal predictors to virtually eliminate the channel memory, and then apply spatial algorithms to recover the transmitted signals. These detectors demonstrate like no others the special structure of tall channels, and experimental results confirm the stated benefits of blind detection based on second-order statistics.

We also propose family of so-called stacked detectors for channels with memory. By stacking a sufficient number of receiver observations, these detectors effectively convert a tall FIR channel into a tall memoryless block-Toeplitz channel. The adaptive algorithms for subspace separation, singular-value decomposition, spatial prediction, and channel diagonalization are then readily extended to tall FIR channels. We define stacked versions of the MMSE, ZF, and WR detectors, which effectively implement multiple lower-dimensional detectors for all users at all delays, thereby optimizing the delay of each user. The adaptive implementations we present are robust to their estimate of the signal subspace dimension. Moreover, they need not know the number of users or even the sizes of their signal alphabets. The stacked channel model also leads directly to an information-lossless space-time precoding technique of finite complexity, which can be used to completely eliminate interference in the receiver without noise enhancement.

CHAPTER 1

INTRODUCTION

COMMUNICATION systems in which multiple users share a common channel or transmission medium are known as *multiuser* or multiple-access systems. Wireless cellular networks are perhaps the best example, where many mobile users communicate to a base station through a common transmission medium, the air. Computers connected to a local-area network are another good example. In modern cable television systems, the upstream communication, from the set-top boxes to the headend, is multiple access in nature. In recent years, systems such as these have become a significant and expanding part of the global telecommunications infrastructure. Consequently, multiuser communication theory has become an active area of research with growing importance.

Reliable multiuser communication presents a key new challenge: mitigation of multiuser interference (MUI). In addition to the impairments that plague single-user point-to-point systems, such as intersymbol interference (ISI) and additive noise, multiuser systems must also contend with interference among the users sharing the channel. The conventional approach to the problem of MUI is to coordinate the transmitters so as to avoid interference [1]. For example, frequency-division and time-division multiple access (FDMA and TDMA) are protocols that confine each transmitter to a unique band of frequencies or to a unique slot of time. Code-division multiple access

(CDMA) is a technique that assigns to each user a unique spreading code, chosen from a set of orthogonal codes. All of these techniques seek to eliminate MUI by orthogonalizing the transmitted signals. The hope is that conventional strategies can then be applied in the receiver (or receivers) to detect each user independently.

Multiuser interference, however, can never be perfectly eliminated, only reduced, and consequently, the performance of conventional detection strategies suffers. In CDMA systems such as IS-95 [2], the spreading codes are only approximately orthogonal. Therefore, significant MUI can remain when the received signal power of one transmitter is much greater than that of another; this is known as the near-far effect [3]. Power-control algorithms have had limited success in combatting the problem; nevertheless, the performance of CDMA systems are usually limited by MUI, not noise. Even for perfectly orthogonal codes, lack of transmitter synchronization or the presence of multipath propagation can destroy orthogonality at the receiver. Similar problems exist for multiuser systems, wireless or otherwise, employing frequency or time-division techniques because of imperfections in bandpass filters or system timing. In practice, these systems must inevitably waste a portion of the bandwidth resource in the form of guard bands or guard times in order to keep adjacent channel interference to acceptable levels. The degree to which multiuser interference can be tolerated ultimately impacts system capacity by governing cell sizes, plans for frequency and time-slot allocation, or choice of spreading codes [4].

Even if the imperfections in implementation could be dismissed, interference avoidance is not always desirable because it does not always make the best use of the bandwidth resource. In this sense, attempts to avoid MUI are somewhat reminiscent of

early attempts to avoid ISI in single-user channels by simply reducing the rate of transmission. Clearly, scenarios exist for which interference avoidance is suboptimal in terms of capacity.

Example 1-1. Echo Cancellation. Consider a pair of 2-wire voiceband data modems. Prior to the 1980's, frequency-division duplexing was used to avoid interference between the two transmitters. In contrast, the V.32 [5] standard allows both transmitters to use the same frequency band, and specifies that the resulting interference be mitigated with echo cancellation. The result is a doubling of capacity.

Example 1-2. Space-Division Multiple Access. As another example, consider an indoor wireless network with one antenna in each room of a building with 100 rooms. Suppose that all 100 antennas are tied to a supercomputer in the cellar. The collection of antennas can be viewed as one big super-array. Consider upstream communication from the 100 portable computers to the supercomputer. Suppose that every antenna can hear every portable, at least to some degree. There are a variety of ways to avoid interference (*e.g.*, TDMA, FDMA, each with 100 slots). However, a better idea, in terms of capacity, is to have each portable talk at will and in the entire frequency band, and to use multiuser detection at the supercomputer to resolve the different signals. The multiuser detector in this scenario in effect implements a form of space-division multiple access, but no attempt is made to completely eliminate multiuser interference.

This research addresses the problem of multiuser detection in the presence of multiuser interference, intersymbol interference, and additive noise. The problem is illus-

trated in Fig. 1-1, in which a receiver with one or more antennas observes several digitally modulated signals transmitted at the same nominal baud rate by independent users. We emphasize again that the wireless system of Fig. 1-1 is only one example. Many others exist with analogous impairments, such as bundled cables with crosstalk [6,7] or multi-track magnetic recording systems with inter-track interference [8]. A generic multiuser communication system is described by a multiple-input multiple-output (MIMO) baseband channel model, in which the receiver observes the following $m \times 1$ vector sequence:

$$\mathbf{r}_k = \mathbf{H}_0 \mathbf{x}_k + \mathbf{H}_1 \mathbf{x}_{k-1} + \dots + \mathbf{H}_M \mathbf{x}_{k-M} + \mathbf{n}_k, \quad (1-1)$$

where \mathbf{x}_k is an $n \times 1$ vector of finite-alphabet symbol sequences transmitted by n independent users, where $\mathbf{H}(z) = \mathbf{H}_0 + \mathbf{H}_1 z^{-1} + \dots + \mathbf{H}_M z^{-M}$ is an $m \times n$ channel transfer function with memory M , and where \mathbf{n}_k is noise. The dimension m of the observation depends on the number of sensors p and the number of samples per baud q according to $m = qp$. The multiuser detection problem [9-11] is to determine one or more of the components of \mathbf{x}_k

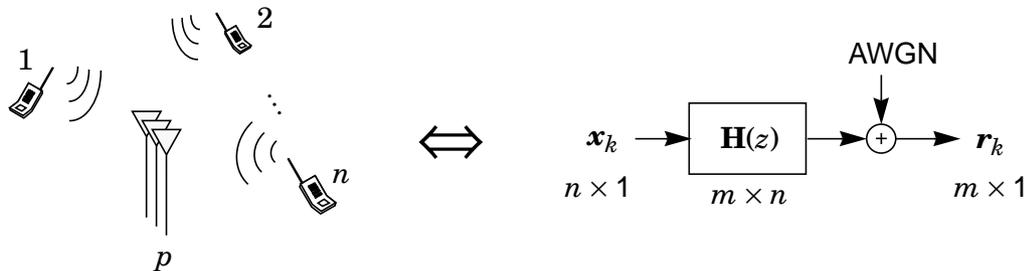


Figure 1-1. A wireless multiuser communication system and discrete-time multiple-input multiple-output (MIMO) channel model.

from the observation \mathbf{r}_k . In other words, the problem is to determine the transmitted sequence of one or more users.

For single-user systems, the transmitter can often assist with the detection task by sending an initial training sequence, which can be used by the receiver to equalize or to identify the channel prior to receiving unknown data. However, for multiuser systems, the use of training sequences is less practical and often impossible. Transmitters generally operate independently and asynchronously. They may be unaware of each other or even unaware of the receiver in question. Consequently, it is desirable for a multiuser detector to function *blindly*, without the cooperation of the transmitters. We assume that the receiver has no prior knowledge of the channel, and that it has only statistical knowledge of \mathbf{x}_k ; for example, it may know the modulation scheme of each user. The blind multiuser detection problem is thus a generalization of the blind equalization problem [12], for which the receiver observation is also described by (1-1) but with $m = 1$ and $n = 1$.

The blind multiuser detection problem we address differs from that addressed by some of the CDMA-specific literature [13,14] in which knowledge of the desired user's spreading code is assumed. In the context of CDMA systems, the model $\mathbf{H}(z)$ of (1-1) captures the effects of the spreading codes, asynchronous transmission, multipath dispersion, transmitter and receiver antenna array patterns, and the phase of the sampler timing. For the special case of synchronous CDMA, chip-rate sampling with perfect timing, a single transmitter antenna, a single receiver antenna, and no multipath dispersion, $\mathbf{H}(z)$ reduces to a memoryless matrix \mathbf{H}_0 whose i -th column is the spreading code of user i . In this case, knowing a spreading code is equivalent to knowing a column of \mathbf{H}_0 . In general, however, the receiver does not know the i -th column of $\mathbf{H}(z)$, even if it

knows the i -th spreading code, because of the distortion caused by the multipath, antenna patterns, and sampling phase. Moreover, we do not wish to limit consideration to CDMA systems. Therefore, our research emphasizes fully blind detectors that do not rely on knowledge of $\mathbf{H}(z)$ in any way.

With neither training nor any channel knowledge, a fully blind detector has no means to distinguish desired users from interfering users; hence, it has no choice but to recover the symbols transmitted by all of the interfering users, and to allow a higher-layer protocol determine who is who. Consequently, the detectors we propose recover the entire vector \mathbf{x}_k .

This research focuses on tall channels, those having more outputs m than inputs n , and on adaptive linear detection algorithms that exploit primarily the *second-order statistics* (SOS) of the observation \mathbf{r}_k . With mild assumptions, tall channels can be identified or equalized up to a memoryless unitary ambiguity by using only the second-order statistics of the channel output [15]. Restricting consideration to second-order techniques is desirable for several reasons. Second-order algorithms are usually more data-efficient [16,17] than higher-order methods, meaning that batch techniques require less data to achieve a given level of accuracy, and that adaptive schemes converge faster. Second-order algorithms do not suffer from problems of ill-convergence that plague many classical methods using non-convex cost functions. Furthermore, second-order algorithms are less sensitive to the channel input distribution and are thus innately compatible with communication systems using shaped constellations having near-Gaussian distributions. The special properties of tall channels lead naturally to

adaptive blind detection algorithms with relatively low complexity and fast, reliable convergence.

This research also inherently addresses certain special cases of the blind single-user detection problem, which can be cast either as a simplification or as a reformulation of the more general multiuser problem. Blind single-user detection (equalization) has importance in multipoint or broadcast channels or in any application where transmitter cooperation is impractical or impossible, *e.g.* satellite, cable television (hybrid fiber-coax [18]), and eavesdropping systems. Oversampling, a technique already widely used because of its numerous other advantages [19-21], preserves the cyclostationarity [22] inherent to most communication signals and produces the tall channel required by second-order detection algorithms. Oversampled single-user systems are modeled by (1-1) with $m > 1$ and $n = 1$; the input sequence \mathbf{x}_k is scalar valued. Moreover, certain multichannel systems can also be modeled by (1-1) by treating the input to each subchannel as a virtual user. For example, discrete multitone (DMT) modulation [23,24], also called orthogonal frequency-division multiplexing (OFDM) [25,26] in certain wireless applications, is a technique that divides the available bandwidth into multiple subchannels and makes all subchannels available to a single user.⁹ By treating each component of \mathbf{x}_k as a virtual user, these systems are easily modeled by (1-1). Single-user wireless systems that transmit with dual polarization [30] or multiple antennas, *e.g.* Bell Labs Layered Space Time (BLAST) [31], can be similarly modeled by defining virtual users. Because we cast these single and multiuser problems

9. See also the American standard for asymmetrical digital subscriber lines (ADSL) [27], which specifies DMT, and the European standards for digital audio and video broadcasting (DAB [28] and DVB [29]), which specify OFDM.

into a common framework, the concepts and algorithms we develop can be applied to any of them.

The remainder of the thesis is organized as follows. In chapter 2, we review linear MIMO system theory and survey prior related work in blind equalization and multiuser detection. Chapters 3 and 4 deal with the special case of memoryless channels. In chapter 3, we present new detectors based on adaptive singular-value decomposition. These include blind implementations of the minimum-mean-square-error (MMSE) and zero-forcing (ZF) detectors as well as a blind channel diagonalization algorithm. We also present a technique for adaptive separation of the signal and noise subspaces that can be used in the front end of any multiuser detector to reduce the complexity of subsequent processing without any information loss. In chapter 4, we present a canonical whiten-rotate structure for multiuser detection with near-MMSE performance. The adaptive implementation is based on spatial linear prediction. Chapters 5 and 6 pertain to channels with memory. In chapter 5, we discuss at length the special properties of tall channels, and we present detectors based on spatio-temporal linear prediction that exploit these properties. In chapter 6, we use a time-to-space mapping procedure to generalize the detectors and concepts of chapters 3 and 4 to channels with memory. We summarize the key contributions of this research and present ideas for future work in chapter 7.

CHAPTER 2

BACKGROUND

MULTIUSER DETECTION strategies can be broadly classified into two groups: linear and nonlinear. Although nonlinear methods can perform better in terms of their probability of symbol error, linear methods are usually lower in complexity and more amenable to blind implementation. Linear detection, however, is not always possible. For a discrete-time channel such as (1-1), we say that the input \mathbf{x}_k is linearly detectable if and only if there exists a stable linear $n \times m$ filter $\mathbf{C}(z)$ that effectively inverts the channel: $\mathbf{C}(z)\mathbf{H}(z) = \mathbf{I}$. The first half of this chapter, section 2.1, reviews key concepts from the theory of linear MIMO systems and multiuser communications including the conditions required for the existence of a linear detector. We express this linear detectability criterion first in section 2.1.2 in terms of the discrete-time model $\mathbf{H}(z)$ of (1-1). Then, using the Nyquist criterion generalized to MIMO systems, we show in section 2.1.5 that the detectability criterion translates into a minimum bandwidth requirement for the underlying continuous-time channel. We review important performance measures including asymptotic multiuser efficiency and near-far resistance in section 2.1.6.

The second half of this chapter, section 2.2, is a survey of related prior work in blind equalization and multiuser detection. In section 2.2.1, we review the classical methods for blind channel identification and equalization, which assume a single-input single-output

(SISO) channel with a stationary input. In section 2.2.2, we survey a newer class of blind equalization algorithms based on second-order cyclostationary statistics. We conclude the chapter with a survey of the literature addressing the general problem of blind multiuser detection, also known as source separation.

2.1 Mathematical Preliminaries

2.1.1 A Review of MIMO System Theory

Most of the mathematical tools of linear system theory, developed in the context of SISO systems [32,33], extend easily to MIMO systems. The *impulse response matrix* \mathbf{H}_k of a linear time-invariant (LTI) discrete-time MIMO system is a matrix of impulse responses whose (i,j) -th element $h_k^{(i,j)}$ is the response at the i -th output to a Kronecker delta δ_k at the j -th input. Equivalently, we can interpret \mathbf{H}_k as a matrix-valued discrete-time sequence. The impulse response \mathbf{H}_k completely describes an LTI MIMO system in the sense that knowledge of \mathbf{H}_k is sufficient to determine the response of the system to any given input:

$$\mathbf{y}_k = \sum_{j=-\infty}^{\infty} \mathbf{H}_{k-j} \mathbf{x}_j. \quad (2-1)$$

A system is said to be *finite impulse response* (FIR) if its impulse response matrix has finite extent; otherwise it is called *infinite impulse response* (IIR). An LTI system is said to be *causal* or *anti-causal* if $\mathbf{H}_k = \mathbf{0}$ for all negative k or all positive k , respectively. We add the qualifier *strictly* (causal or anti-causal) if \mathbf{H}_0 is also zero. The channel model of (1-1) is FIR and both stable and causal. The *z-transform* of \mathbf{H}_k can be interpreted either as a

matrix-valued function of z , or as a matrix of z -transforms as defined in traditional SISO system theory:

$$\mathbf{H}(z) = \sum_{k=-\infty}^{\infty} \mathbf{H}_k z^{-k}. \quad (2-2)$$

We can use the impulse response \mathbf{H}_k or its z -transform $\mathbf{H}(z)$ to establish the stability of a LTI MIMO system. A system is said to be *bounded-input bounded-output (BIBO) stable* if and only if the output \mathbf{y}_k is bounded for all bounded inputs \mathbf{x}_k :

$$|\mathbf{x}_k^{(j)}| < \infty \forall j, k \Rightarrow |\mathbf{y}_k^{(i)}| < \infty \forall i, k. \quad (2-3)$$

It follows that an LTI MIMO system is stable if and only if all elemental impulse responses $h_k^{(i,j)}$ are stable. Applying results from SISO system theory, we conclude that a discrete-time LTI MIMO system is stable if and only if all component sequences $h_k^{(i,j)}$ of its impulse response are absolutely summable:

$$\sum_{k=-\infty}^{\infty} |h_k^{(i,j)}| < \infty \forall i, j. \quad (2-4)$$

In the z -domain, an LTI system is stable if and only if none of the component transforms $H^{(i,j)}(z)$ have poles on the unit circle, or equivalently, if and only if $\mathbf{H}(z)$ converges uniformly on the unit circle; that is, the following summation converges for $|z| = 1$:

$$\sum_{k=-\infty}^{\infty} \|\mathbf{H}_k z^{-k}\|_1 < \infty, \quad (2-5)$$

where ' $\|\cdot\|_1$ ' denotes the matrix 1-norm (the sum of the modulus of all matrix elements). This equivalence follows from the observation that (2-5), evaluated on the unit circle $z = e^{j\theta}$, is identical to (2-4).

2.1.2 Linear Detectability — Discrete-Time Channels

Recall that if a SISO channel $H(z)$ has no zeros on the unit circle, then there exists a stable inverse (having no poles on the unit circle). We say that such a channel is equalizable, and we call its inverse a zero-forcing (ZF) equalizer $C_{ZF}(z)$ because it zeros all ISI: $C_{ZF}(z)H(z) = 1$. For a MIMO channel, an analogous filter, zeroing all ISI and MUI, is called a zero-forcing multiuser *detector*.

Definition 2-1. For the MIMO channel of (1-1), a **zero-forcing detector** is an $n \times m$ stable LTI filter $\mathbf{C}(z)$ satisfying $\mathbf{C}(z)\mathbf{H}(z) = \mathbf{I}$.

In the context of CDMA, a zero-forcing detector is sometimes referred to as a decorrelating detector [10]. The concept of equalizability for SISO channels thus generalizes to linear *detectability* for MIMO channels.

Definition 2-2. For the channel of (1-1), the input \mathbf{x}_k is said to be **linearly detectable** if and only if a zero-forcing detector exists.

The rank of a channel $\mathbf{H}(z)$ on the unit circle determines whether or not linear detection of the channel input is possible.

Theorem 2-1. For the channel of (1-1), the input \mathbf{x}_k is linearly detectable if and only if $\mathbf{H}(z)$ has rank n for all z on the unit circle: $|z| = 1$.

Proof: If $rank[\mathbf{H}(z)] = n$ for all z on the unit circle then $[\mathbf{H}^*(1/z^*)\mathbf{H}(z)]^{-1}$ is stable, and a zero-forcing filter is given by

$$\mathbf{C}_{ZF}(z) = \mathbf{H}^\dagger(z) \equiv [\mathbf{H}^*(1/z^*)\mathbf{H}(z)]^{-1}\mathbf{H}^*(1/z^*). \quad (2-6)$$

(Adding any $n \times m$ filter $\mathbf{V}(z)$ in the left null space of $\mathbf{H}(z)$ to $\mathbf{H}^\dagger(z)$ produces another ZF detector.) If $\text{rank}[\mathbf{H}(z_0)] < n$ for some z_0 on the unit circle, then $\text{rank}[\mathbf{C}(z_0)\mathbf{H}(z_0)] < n$, and hence $\mathbf{C}(z_0)\mathbf{H}(z_0) \neq \mathbf{I}$, for any stable $\mathbf{C}(z)$. \square

The linear detectability criterion of Theorem 2-2 could be expressed in terms of the zeros, rather than the rank, of $\mathbf{H}(z)$ on the unit circle, if zeros were defined as points of rank deficiency. (Similarly poles could be defined in terms of the zeros of the left inverse or pseudoinverse channel.) Although the suggested definitions would strengthen the analogy between equalizability and detectability, we resist formally defining zeros and poles of a MIMO system because there is no universally accepted convention, and ultimately, such definitions are unnecessary for this thesis. Interested readers are referred to Kailath [34] and references therein [35-37].

We now consider three example channels that contrast linear and nonlinear detection methods.

Example 2-1. A Nonlinear Detector. Consider the following 2×2 channel:

$$\mathbf{H}(z) = \begin{bmatrix} 3 & z^{-1} \\ 3z^{-1} & 1 \end{bmatrix}. \quad (2-7)$$

For square channels, the points of rank deficiency are given by the zeros of the determinant. In this case, $\det[\mathbf{H}(z)] = 3 - 3z^{-2}$ is zero at $z = \pm 1$. Therefore, no zero-forcing detector exists. Nevertheless, nonlinear detection methods may still be applicable. The nonlinear detector of Fig. 2-1 is one viable option under certain assumptions. Using the notation of (1-1), the receiver observation is

$$\mathbf{r}_k = \begin{bmatrix} 3 & 0 \\ 0 & 1 \end{bmatrix} \mathbf{x}_k + \begin{bmatrix} 0 & 1 \\ 3 & 0 \end{bmatrix} \mathbf{x}_{k-1} + \mathbf{n}_k. \quad (2-8)$$

If the users are quadrature phase-shift keyed (QPSK) with unity power, and the noise variance σ^2 is sufficiently small, then the first user can be detected from $r_k^{(1)} = 3x_k^{(1)} + x_{k-1}^{(2)} + n_k^{(1)}$ (the first component of \mathbf{r}_k) by treating the contribution from the second user $x_{k-1}^{(2)}$ like additive noise; *i.e.*, the detector scales $r_k^{(1)}$ and then makes a decision:

$$\hat{x}_k^{(1)} = \text{dec}(r_k^{(1)}/3), \quad (2-9)$$

where $\text{dec}(\cdot)$ denotes a decision device. Assuming the decision is correct, the second user can be detected from $r_k^{(2)}$ by cancelling interference from the first:

$$\hat{x}_k^{(2)} = d(r_k^{(2)} - 3\hat{x}_{k-1}^{(1)}). \quad (2-10)$$

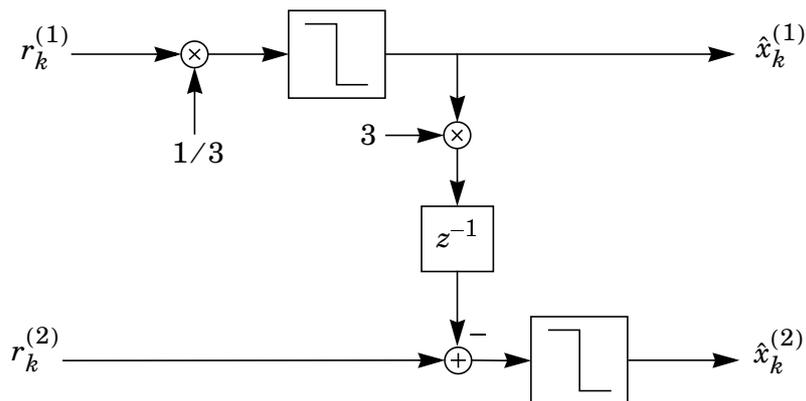


Figure 2-1. A successive cancellation detector for Example 2-1.

The nonlinear detector of Fig. 2-1 uses successive cancellation [38-40]. The technique has several shortcomings, the most immediately obvious of which is that if the decision in (2-9) is incorrect, then it is highly probable that the one in (2-10) is as well. It treats the users asymmetrically or unequally. Blind implementation is difficult because the detector architecture is highly dependent on the channel and its input constellations. There are of course other nonlinear techniques, including multistage detection [41,42] (a symmetric version of successive cancellation), decision-feedback detection [30,43-46], and maximum-likelihood sequence detection [47], each with varying tradeoffs in performance and complexity. If the conditions of Theorem 2-2 are satisfied, however, straightforward linear methods can be used.

Example 2-2. A Linear Detector. Consider another 2×2 channel:

$$\mathbf{H}(z) = \begin{bmatrix} 2 & z^{-1} \\ -1 & 1 \end{bmatrix}. \quad (2-11)$$

On the unit circle, the determinant $\det[\mathbf{H}(z)] = 2 + z^{-1}$ is nonzero and the channel is full rank. Hence, a zero-forcing detector exists and is given by

$$\mathbf{C}_{ZF}(z) = \mathbf{H}^{-1}(z) = \frac{1}{1 + 2z^{-1}} \begin{bmatrix} 1 & -2z^{-1} \\ 1 & 1 \end{bmatrix}. \quad (2-12)$$

A bank of slicers still follows $\mathbf{C}_{ZF}(z)$, but the decision for any particular user in no way impacts that for any other user. For a noiseless channel, the detector perfectly recovers the channel input. Observe that $\mathbf{C}_{ZF}(z)$ is IIR and anti-causal:

$$\frac{1}{1 + 2z^{-1}} = \frac{1}{2} - \frac{1}{4}z + \frac{1}{8}z^2 - \frac{1}{16}z^3 + \dots \quad (2-13)$$

Nevertheless, there exists a delay D (an integer) such that $z^{-D}\mathbf{C}_{ZF}(z)$ is approximately FIR and causal.

As the previous example demonstrates, Theorem 2-2 promises only that if $\mathbf{H}(z)$ is full column rank on the unit circle, then a linear zero-forcing detector $\mathbf{C}_{ZF}(z)$ exists. It does not suggest that the detector is FIR, nor even causal. The fact that an FIR channel has an IIR inverse should not be surprising; it is always the case for SISO channels (assuming an inverse exists). However, as illustrated by the following example, it is not always the case for MIMO channels.

Example 2-3. An FIR Linear Detector. Consider a 3×2 channel:

$$\mathbf{H}(z) = \begin{bmatrix} z^{-1} & -1 - 3z^{-1} \\ 1 - z^{-1} & -1 + z^{-1} \\ z^{-1} & -1 \end{bmatrix}. \quad (2-14)$$

This channel has a zero-forcing detector given by

$$\mathbf{C}_{ZF}(z) = \frac{1}{6} \begin{bmatrix} -3 - 2z^{-1} + 3z^{-2} & 6 + 12z^{-1} + 9z^{-2} & -3 + 5z^{-1} + 6z^{-2} \\ -3 + z^{-2} & 6z^{-1} + 3z^{-2} & -3 + 3z^{-1} - 2z^{-2} \end{bmatrix}. \quad (2-15)$$

It is easy to verify that indeed $\mathbf{C}_{ZF}(z)\mathbf{H}(z) = \mathbf{I}$. Remarkably, both the channel $\mathbf{H}(z)$ and the detector $\mathbf{C}_{ZF}(z)$ are FIR!

We see in chapter 5 that almost all tall FIR channels have causal FIR inverses, which is a direct consequence of the fact that they are full rank *everywhere*, not just on the unit circle. This property greatly simplifies blind implementation of multiuser detectors.

2.1.3 Derivation of the Discrete-Time Channel Model

So far we have considered only discrete-time signals and systems. The underlying physical channel in most communication systems, however, is continuous-time in nature. Nevertheless, the discrete-time model of (1-1) is adequate if the underlying continuous-time channel is stable and bandlimited, with additive white Gaussian noise (AWGN).

Let $\mathbf{H}(t)$ be a $p \times n$ *continuous-time impulse response* matrix. (The (i, j) -th element $H^{(i, j)}(t)$ is the response at the i -th output to a Dirac delta $\delta(t)$ at the j -th input.) The continuous-time observation at the receiver can be expressed in terms of the impulse response matrix as

$$\mathbf{r}(t) = \sum_{k=-\infty}^{\infty} \mathbf{H}(t - kT) \mathbf{x}_k + \mathbf{n}(t), \quad (2-16)$$

where $\mathbf{n}(t)$ is white Gaussian noise.

The *Fourier transform* of $\mathbf{H}(t)$ is a matrix of transforms given by

$$\mathbf{H}(f) = \int_{-\infty}^{\infty} \mathbf{H}(t) e^{-j2\pi ft} dt. \quad (2-17)$$

We assume that the components $H^{(i, j)}(f)$ of $\mathbf{H}(f)$ are bandlimited to $|f| \leq W$, and that the two-sided noise power spectral density (PSD) $\int_{-\infty}^{\infty} \mathbf{E}[\mathbf{n}(t)\mathbf{n}^*(t - \tau)] e^{-j2\pi f\tau} d\tau = N_0 \mathbf{I}$. Fig. 2-2(a), taken from [48], illustrates one possible receiver front end which converts the continuous-time model to the discrete-time model of Fig. 2-2(b) while preserving all signal information. This front end consists of a bank of ideal anti-aliasing low-pass filters, each with cutoff frequency $\frac{q}{2T}$, followed by a sampler with rate q/T . To preserve informa-

tion, the cut-off frequency $\frac{q}{2T}$ must be greater than the bandwidth W , so $q \geq 2WT$. The smallest value of q that ensures the sampling rate is an integer multiple of the baud rate is

$$q = \lceil 2WT \rceil. \quad (2-18)$$

Hence, there are q samples per baud, and each sample is a $p \times 1$ vector. At each baud time k , a serial-to-parallel (S/P) converter stacks q different $p \times 1$ vectors, as follows, to form a new vector:

$$\mathbf{r}_k = \begin{bmatrix} \mathbf{r}(kT) \\ \mathbf{r}((k-1/q)T) \\ \vdots \\ \mathbf{r}((k-(q-1)/q)T) \end{bmatrix}, \quad (2-19)$$

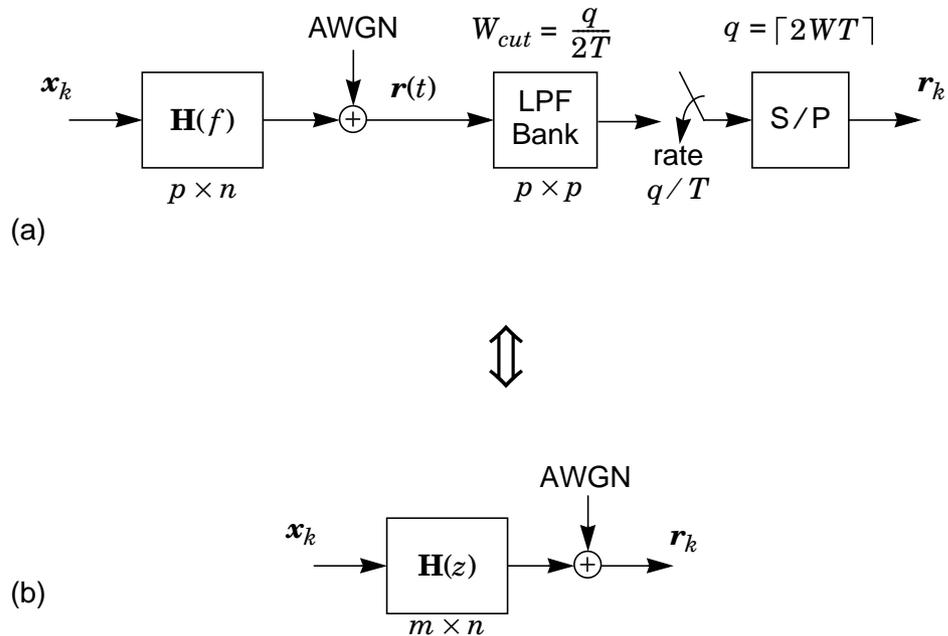


Figure 2-2. [48] (a) Original $p \times n$ continuous-time channel and oversampling receiver front end; (b) equivalent $m \times n$ baud-rate discrete-time channel model.

of dimension $m \times 1$, where

$$m = qp. \quad (2-20)$$

No information-preserving front end can produce a baud-rate output sequence with smaller dimension. This baud-rate output \mathbf{r}_k can be modeled as the output of an $m \times n$ discrete-time channel, shown in Fig. 2-2(b). Following [48], we can express $\mathbf{H}(z)$ as a folded spectrum:

$$\mathbf{H}(z) = \frac{1}{T} \sum_{k=-\infty}^{\infty} \tilde{\mathbf{H}}(f - \frac{k}{T}), \quad (2-21)$$

where $\tilde{\mathbf{H}}(f)$ is formed by stacking phase-shifted versions of $\mathbf{H}(f)$:

$$\tilde{\mathbf{H}}(f) = \begin{bmatrix} \mathbf{H}(f) \\ e^{-j2\pi f T/q} \mathbf{H}(f) \\ \vdots \\ e^{-j2\pi f T(q-1)/q} \mathbf{H}(f) \end{bmatrix}. \quad (2-22)$$

The FIR approximation of (1-1) is reasonable for sufficiently large memory M . The noise of the discrete-time model is also white and Gaussian, with PSD $\sum_{l=-\infty}^{\infty} \mathbf{E}[\mathbf{n}_k \mathbf{n}_{k-l}^*] = \sigma^2 \mathbf{I}$, where $\sigma^2 = N_0 q / T$. This result follows from the ideal *Nyquist* property of the anti-aliasing filterbank.

2.1.4 The Generalized Nyquist Criterion

The familiar Nyquist criterion for pulse or quadrature amplitude modulated (PAM or QAM) signals was generalized to vector-valued signals by Shnidman [49]. Let $\mathbf{x}(t)$ denote a vector-valued QAM signal:

$$\mathbf{x}(t) = \sum_{k=-\infty}^{\infty} \mathbf{P}(t - kT) \mathbf{x}_k, \quad (2-23)$$

where \mathbf{x}_k is an $n \times 1$ vector of symbol sequences, where $\mathbf{P}(t)$ is an $n \times n$ pulse shape (an impulse response matrix), and where T is the baud period. The generalized Nyquist criterion states that samples of $\mathbf{x}(t)$, taken at the baud rate, reproduce \mathbf{x}_k without intersymbol or multiuser interference, *i.e.* $\mathbf{x}(kT) = \mathbf{x}_k$, if and only the Fourier transform of the sampled pulse shape aliases to the identity.

Definition 2-3. [49,50] A pulse shape $\mathbf{P}(t)$ is said to be **Nyquist**, or to satisfy the **generalized Nyquist criterion**, if and only if

$$\mathbf{P}(kT) = \delta_k \mathbf{I}_n \Leftrightarrow \frac{1}{T} \sum_{k=-\infty}^{\infty} \mathbf{P}(f - \frac{k}{T}) = \mathbf{I}_n, \quad (2-24)$$

where $\mathbf{P}(f)$ is the Fourier transform of $\mathbf{P}(t)$.

2.1.5 Linear Detectability — Continuous-Time Channels

Recall that a single scalar-valued QAM signal requires a minimum bandwidth of $W_0 = \frac{1}{2T}$ to avoid ISI [51]. Falconer *et al.* [52] generalized this result to n users. For a $1 \times n$ channel, linear detection of all users is possible only if the signal bandwidth W of each exceeds W_0 by a factor of n . More generally, for a $p \times n$ channel (a receiver with p antennas), linear detection requires that the bandwidth of each signal exceeds W_0 by a factor of n/p .

For a continuous-time MIMO system, such as that illustrated in Fig. 2-2, we define linear detectability as follows.

Definition 2-4. For the $p \times n$ continuous-time channel $\mathbf{H}(f)$ of (2-16), the input \mathbf{x}_k is said to be **linearly detectable** if and only if there exists an $n \times p$ stable linear filter $\mathbf{G}(f)$ such that $\mathbf{P}(f) = \mathbf{G}(f)\mathbf{H}(f)$ satisfies the generalized Nyquist criterion of (2-24).

In other words, the channel can be effectively inverted by the cascade of a linear filter and a baud rate sampler. The following result, due to Falconer *et al.* [52], states necessary and sufficient conditions for linear detection of the channel input.

Theorem 2-2. [52] Suppose $\mathbf{H}(f)$ of (2-16) is a $p \times n$ matrix whose elements have bandwidth W . The following conditions are necessary for linear detection of \mathbf{x}_k :

$$n \leq n_{max} = p \lfloor W/W_0 \rfloor, \quad (2-25)$$

$$W \geq W_{min} = \frac{n}{p} W_0. \quad (2-26)$$

Moreover, if $\mathbf{H}(f)$ is full rank for $|f| \leq W$, then (2-25) is also sufficient.

Proof: [52] The number of nonzero terms in the summation of (2-24) with $\mathbf{P}(f) = \mathbf{G}(f)\mathbf{H}(f)$ is at least $\lfloor 2WT \rfloor$, so (2-24) is a system of n^2 equations and at least $pn \lfloor 2WT \rfloor$ unknowns, given by $\mathbf{G}(f - k/T)$ for each k . There must be at least as many unknowns as equations, which implies the necessity of (2-25). Simple algebra then yields (2-26). If $\mathbf{H}(f)$ is full rank for $|f| \leq W$, then the equations represented by (2-24) are linearly independent, which implies the sufficiency of (2-25). \square

The linear detectability criterion of Theorem 2-2 translates into a minimum bandwidth requirement for the underlying continuous-time channel. Consider the following special cases. For a receiver with only one antenna or sensor, the number of users that can

be linearly detected is roughly equal to the bandwidth expansion W/W_0 , a well-known result in the spread-spectrum community. Multiplying the number of antennas at the receiver by p divides the required bandwidth expansion by the same factor. In systems with less than 100% excess bandwidth ($W < 2W_0$), the maximum number of users that can be linearly detected is equal to the number of sensors p . This agrees with the well-known array processing principle that an array of p antennas can reject $p - 1$ narrowband interferers. Finally, linear multiuser detection using only a single sensor is impossible for systems with less than 100% excess bandwidth. For example, we cannot eavesdrop on the conversation between the two V.32 modems in Example 1-1 via linear means.

2.1.6 Performance Measures for Multiuser Detectors

There are many criteria that can be used to measure the performance of a multiuser detector. Arguably, the best of these measure the probability of some error event, *e.g.*, the probability of symbol error for the i -th user $Pr\{\hat{x}_k^{(i)} \neq x_k^{(i)}\}$, the probability of symbol error for *any* user $Pr\{\hat{\mathbf{x}}_k \neq \mathbf{x}_k\}$, the probability of bit error for user i , etc. However, as for SISO communication systems, we often look for more mathematically tractable, although less precise, measures of performance. The primary figure of merit used throughout this thesis is mean-square error (MSE). The corresponding benchmark detector is the minimum-mean-square-error (MMSE) detector. (See Definitions 3-1 and 6-1.) The MMSE detector is analogous to the MMSE equalizer for SISO channels; it is a compromise between interference (both ISI and MUI) and noise.

Verdú and colleagues [11,53] developed several measures of performance for multiuser detectors in the context of CDMA systems. The measures they developed are intended both to gauge the ability of a detector to reject interferers and also to characterize

its robustness to the near-far problem, a major concern in wireless CDMA systems. The measures are closely related to the probability of symbol error for the user in question.

The presence of interfering users can serve only to increase the probability of symbol error for the desired user. We can attempt to quantify this effect as follows. For the channel $\mathbf{H}(z)$ of (1-1), let P_i denote the power of user i as measured at the receiver. For a given detector $\mathbf{C}(z)$, let $\text{SER}(\sigma) = Pr\{\hat{x}_k^{(i)} \neq x_k^{(i)}\}$ denote the symbol error rate for user i corresponding to the noise variance σ^2 . Following [53], we define the *effective power* ρ_i of user i as the energy required to achieve $\text{SER}(\sigma)$, assuming the same background noise σ^2 , but assuming that none of the interfering users are present.

Definition 2-5. [53] The **multiuser efficiency** for user i is the ratio of the effective power to the actual power: ρ_i / P_i .

The reciprocal of the multiuser efficiency can be interpreted as a power penalty due to the presence of the interferers. The multiuser efficiency in negligible noise is also of interest.

Definition 2-6. [53] The **asymptotic multiuser efficiency** (AME) is the limit of the multiuser efficiency as the noise power goes to zero:

$$\eta_i = \lim_{\sigma^+ \rightarrow 0} \rho_i / P_i. \quad (2-27)$$

The worst-case AME, taken over all possible interference power profiles, characterizes a detector's robustness to the near-far problem.

Definition 2-7. [53] The **near-far resistance** for user i of a detector $\mathbf{C}(z)$ relative to a channel $\mathbf{H}(z)$ is the worst-case AME relative to all channels $\mathbf{H}(z)\mathbf{A}$, where $\mathbf{A} = \text{diag}(A_j)$ with $A_j \geq 0 \forall j \neq i$ and $A_i = 1$:

$$\bar{\eta}_i = \inf_{\mathbf{A}} \eta_i. \quad (2-28)$$

Many of the detectors we develop in subsequent chapters are shown to be *optimally near-far resistant* for all users, meaning that for any channel $\mathbf{H}(z)$, no other multiuser detector has a better near-far resistance.

2.2 A Survey of Related Prior Work

2.2.1 Classical Blind Equalization

We use the term *classical* to describe any blind equalization or channel identification algorithm designed for a SISO channel with a stationary input, such as (1-1) with $m = n = 1$. Classical algorithms share the characteristic that they all use nonlinear processing to exploit the higher-order statistics (HOS) of the observation. HOS are needed to extract the phase information of the channel. SOS alone cannot distinguish between the channels $a + bz^{-1}$ and $b + az^{-1}$, for example. As a consequence, blind identification of a (possibly non-minimum-phase) channel with a stationary input requires that the channel input have non-Gaussian statistics [12], a well-known result.

The most popular blind equalization algorithms are the Bussgang¹⁰ algorithms [55,56], which use the structure of Fig. 2-3. This structure is similar to the deci-

10. In equilibrium, the equalizer output y_k is a Bussgang process, satisfying $E[y_k g^*(y_k)] = E[y_k y_k^*]$ [54].

sion-directed equalizer [57] except that a generic nonlinearity $g(\cdot)$ is used in lieu of a conventional decision device. Nonlinear processing thus arises in the computation of the error used to update the equalizer coefficients: $e_k = g(y_k) - y_k$. The least-mean-square (LMS) algorithm [58,59] is popular because of its simplicity. The decision-directed equalizer is in fact an example of a Bussgang algorithm, but it is poorly suited for initialization or recovery because the decisions are generally unreliable prior to convergence [55].

The first blind equalization algorithm designed expressly for initialization or recovery was proposed by Sato [60] for PAM systems. It uses the structure of Fig. 2-3 with a scaled signum function as its nonlinearity: $g(y_k) = \gamma \operatorname{sgn}(y_k)$, where the scale factor $\gamma = \frac{E[x_k^2]}{E[|x_k|]}$ is based on *a priori* statistical knowledge of x_k . Sato later extended the algorithm to QAM systems [61].

Godard [62] observed that for QAM constellations, ISI cannot produce phase distortion without also producing amplitude distortion. Hence, he proposed an algorithm designed to eliminate amplitude distortion by minimizing the following cost function: $J(y_k) = (|y_k|^p - R_p)^2$, where $R_p = \frac{E[|x_k|^{2p}]}{E[|x_k|^p]}$ for some positive integer p . Treichler and Agee [63] independently proposed the constant-modulus algorithm (CMA), a special case

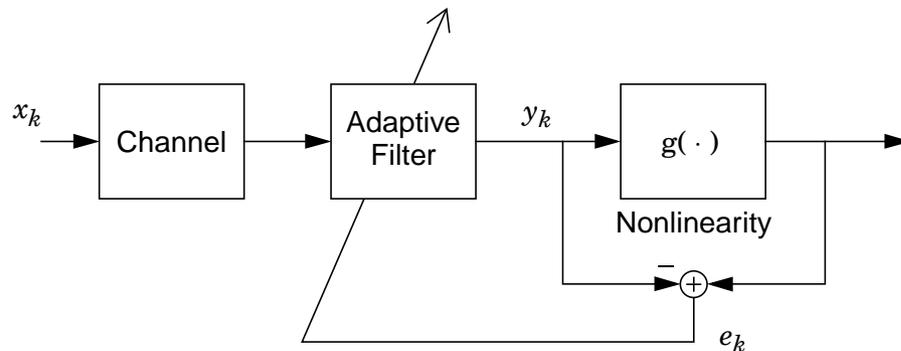


Figure 2-3. The Bussgang structure for blind equalization.

of the Godard algorithm with $p = 2$. CMA was designed originally to restore (at the equalizer output) the constant-modulus property of PSK constellations, but it works for QAM constellations as well, in light of Godard's observation regarding ISI. Godard/CMA can be cast into the Bussgang framework of Fig. 2-3 by proper choice of the nonlinearity $g(\cdot)$.

Shalvi and Weinstein [64] observed that the fourth-order cumulant [65] $c_4(y_k) = E[|y_k|^4] - 2E^2[|y_k|^2] - |E[y_k^2]|^2$ of the equalizer output y_k is bounded by that of the channel input: $|c_4(y_k)| \leq |c_4(x_k)|$. They proposed that $|c_4(y_k)|$ be maximized subject to a power constraint: $E[|y_k|^2] = E[|x_k|^2]$. This algorithm can also be cast into the Bussgang framework by a proper choice of nonlinearity — an interesting exercise in this case because it clearly demonstrates that the use of a nonlinearity $g(\cdot)$ is an implicit use of higher-order statistics.

Bussgang algorithms have several drawbacks. Many are known to suffer from problems of ill-convergence; that is, the coefficients may converge to a suboptimal local minimum because the cost function is often a non-convex function of the equalizer coefficients [66,67]. Moreover, convergence is often slow. Many heuristic methods have been developed to combat these problems. To deal with ill-convergence, Godard [62] suggested a tap-initialization procedure, and Foschini [68] suggested an algorithm to track and center the primary tap. To speed convergence, Benveniste and Goursat [69,70] proposed an algorithm using a weighted sum of the complex-Sato [61] and decision-directed errors. Picchi and Prati [71] suggested a “stop-and-go” algorithm that updates only when the error is reliable. Niki and colleagues proposed CRIMNO [72,73], which uses a nonlinearity with memory to exploit knowledge of symbol sequence correlations.

Alternatives to the Busgang techniques are not as popular in the communications context because of their inherent complexity. Methods involving the explicit estimation of higher-order cumulants [74-76] or their polyspectra [77,78] generally have reliable and fast convergence, but they are numerically complex and therefore usually impractical. The maximum-likelihood channel-estimation algorithms of Seshadri [79] and Ghosh and Weber [80] are also prohibitively complex. Methods based on neural networks have been proposed by Wong and Fine [81] and Chen and Chen [82], but they reportedly suffer from convergence problems.

The fundamental shortcoming of the classical approach to blind equalization is perhaps best expressed by Donoho's minimum-entropy concept [83], illustrated in Fig. 2-4. Donoho introduced the idea of contrast functions to quantify the Gaussianity of a distribution. If the samples of the channel input x_k are stationary, independent, and non-Gaussian, then, from the central limit theorem [84], the channel output r_k is in a sense more Gaussian than x_k . The task of a blind equalizer is thus to drive its output distribution away from Gaussian. This interpretation has immediate implications: blind equalization of channels with Gaussian inputs is impossible, and blind equalization of channels with near-Gaussian

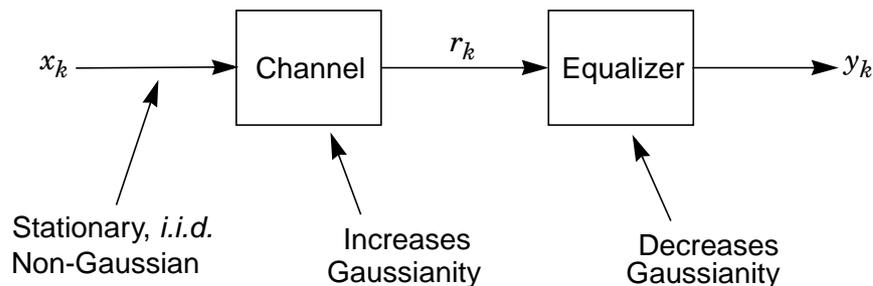


Figure 2-4. Donoho's minimum-entropy concept [83].

inputs is difficult. Robust blind equalization requires that the channel input distribution be far from Gaussian. But, it is well known that Gaussian inputs are necessary to approach Shannon capacity. *Hence, in the classical setting, achieving channel capacity and providing for robust blind equalization are competing interests.* Capacity must be sacrificed to facilitate equalization by the use of either suboptimal transmit shaping or training sequences.

2.2.2 Blind Equalization of Cyclostationary Sequences

The underlying assumption of the classical approaches to blind equalization is that the channel input is stationary. Fortunately, however, most communication signals are *cyclostationary*. Second-order statistics can be sufficient for identifying even non-minimum-phase channels when the input is cyclostationary. Gardner [16] was perhaps the first to understand this; he proposed a channel identification method exploiting the cyclostationary nature of PAM signals. Although the technique does require transmitter cooperation in the form of a pilot tone, no replica of this signal is needed in the receiver.

Tong, Xu, and Kailath (TXK) [17] proposed the first truly blind channel identification algorithm based on second-order statistics. As illustrated in Fig. 2-5, a SISO channel with cyclostationary input can be modeled as a single-input multiple-output (SIMO) channel with a stationary input, as in (1-1) with $m > 1$ and $n = 1$. Exploiting this equivalence, the TXK algorithm defines a stacked observation vector $\mathbf{R}_k^T = [\mathbf{r}_k^T \dots \mathbf{r}_{k-N+1}^T]$ and then estimates its autocorrelation $\Phi_l = E[\mathbf{R}_k \mathbf{R}_{k-l}^*]$ at lags $l = 0$ and $l = 1$. The channel $\mathbf{H}(z)$ can be uniquely determined from these estimates up to an arbitrary complex scalar $e^{j\theta}$ provided that N is sufficiently large and that the correlation estimates are full rank. Although numerically complex, requiring two singular-value decompositions (SVDs), the

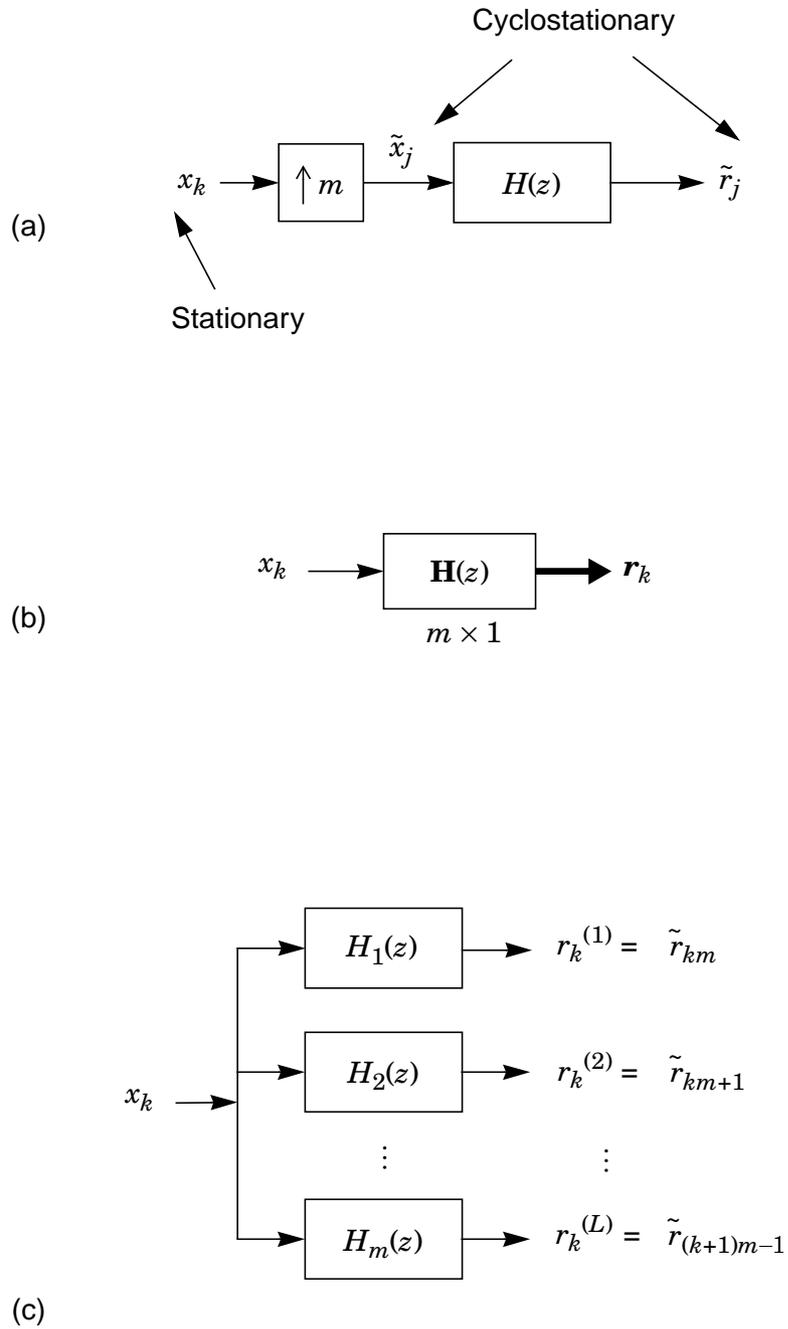


Figure 2-5. Equivalent FIR channel models: (a) an upsampled SISO channel, (b) a SIMO channel, (c) a filterbank.

TXK algorithm was the first to demonstrate the feasibility of second-order blind channel identification in a general context. In subsequent work, Tong *et al.* [85,86] established the necessary and sufficient conditions for the second-order identifiability of SIMO channels.

Using the framework of TXK, Moulines *et al.* [87,88] proposed a channel identification algorithm similar to the well-known MUSIC algorithm [89] that exploits the orthogonality of the signal and noise subspaces of Φ_0 . The algorithm requires only a single SVD, and hence, is less complex than TXK. However, like TXK, it requires an estimate of the channel order, and is reportedly sensitive to inaccuracies [90-92] in this estimate. Liu and Xu [93] independently developed a similar subspace-based algorithm.

Slock and colleagues [94-96] interpreted an oversampled SISO channel as a filterbank, illustrated in Fig. 2-5(c). He showed that the condition required for second-order identifiability can be expressed in terms of the zeros of the filterbank subchannels: the channel is identifiable if and only if the subchannels have no common zeros. This criterion is the same as that required for the existence of a perfect-reconstruction FIR filterbank [97]. The implication is that the output of an FIR SIMO channel simultaneously has not only a moving-average (MA) nature, but also a finitely parameterized autoregressive (AR) nature. Slock was the first to suggest the use of linear prediction to equalize a SIMO channel.

Abed-Meraim [90,91] developed the prediction-error blind identification (PEBI) algorithm, also referred to in the literature as the linear predictive algorithm (LPA). It is a batch-oriented algorithm that first computes a prediction-error filter, and then correlates the prediction error to the receiver observation to determine $\mathbf{H}(z)$. The method is reportedly more robust to over-estimation of the channel order than either TXK or the subspace

approach of Moulines. However, it is still rather complex, requiring a pseudoinverse. Furthermore, its performance reportedly suffers when the channel has weak precursors [92].

Ding [92] developed the outer-product-decomposition algorithm (OPDA) to identify the channel. OPDA and LPA are reportedly similarly robust to over-estimation of the channel order, but OPDA performs better than LPA when channel precursors are small. OPDA requires two SVDs.

All of the preceding SOS-based approaches seek to identify the channel, with the understanding that once the channel is identified, equalizers can then be computed or maximum-likelihood sequence detection can be performed using the Viterbi algorithm. If equalization is the desired goal, a second computation is therefore needed. In contrast, Giannakis and Halford [98] have described methods for directly computing equalizers, both MMSE and ZF, from the observed channel outputs. The zero-delay MMSE equalizer is computed from a batch estimate of Φ_0 . The MMSE equalizer for delays other than zero, as well as the zero-forcing equalizer, requires an estimate of the noiseless autocorrelation, so the noise subspace must be estimated first. Computation of the zero-forcing equalizer requires one pseudoinverse; computation of the MMSE equalizer for general delay requires two. Giannakis and Halford have also described promising adaptive implementations of these direct methods. However, convergence of these algorithms appears to be sensitive to the initial equalizer estimate.

2.2.3 Blind Multiuser Detection

There are many diverse bodies of literature that have relevance to the general blind multiuser detection problem. In some contexts, partial knowledge of the channel is assumed. The problem is greatly simplified when the receiver has partial or full knowledge

of the channel. In array processing [99], knowledge of the array geometry is often assumed; *e.g.*, the generalized sidelobe canceler [100], MUSIC [89], and ESPRIT [101]. In direct-sequence CDMA systems, knowledge of the spreading sequence of the desired user is assumed, *e.g.*, the minimum-output-energy (MOE) detector of Honig *et al.* [13] and the subspace-based detector of Wang and Poor [14]. All of these algorithms exploit channel knowledge in some way, and are thus only partially blind.

The fully blind multiuser detection problem is sometimes called blind source separation.¹¹ Mirroring classical solutions to the single-user problem, many early blind algorithms for multiuser detection or channel identification were based on explicit computation of higher-order statistics. Cardoso proposed algorithms [102-104] for identification of memoryless MIMO channels based on higher-order cumulants. Giannakis [105], Swami [106], and Tugnait [107] proposed similar cumulant-based algorithms for channels with memory. Comon [108,109] generalized the idea of contrast functions to MIMO channels. He showed that, under the assumption that the channel inputs are non-Gaussian and statistically independent, the detector outputs are separated if and only if they are also statistically independent. Linear minimization of the statistical dependence between detector output components is known as independent component analysis (ICA), but the idea really is equivalent to Donoho's idea of minimizing Gaussianity. Moreau and Pesquet [110] also proposed a channel identification algorithm based on generalized contrasts. These algorithms, like their counterparts for SISO channels, are impractical for many applications because of their high computational complexity; they require batch

11. Channels, either with or without memory, are sometimes called mixtures, either convolutive or instantaneous, respectively.

estimates of higher-order statistics. Furthermore, many data samples are needed to produce accurate estimates.

Several adaptive algorithms resembling the classical Bussgang algorithms have also been proposed for blind multiuser detection. These include extensions of CMA to the multiuser problem. For example, pointwise CMA [111,112] is an extension that applies the CMA cost function to each component of the equalizer output: $J_{pw}(\mathbf{y}_k) = \sum_{i=1}^n (|y_k^{(i)}|^2 - R_i)^2$, where $R_i = \frac{E[|x_k^{(i)}|^4]}{E[|x_k^{(i)}|^2]}$, and where $x_k^{(i)}$ and $y_k^{(i)}$ denote the i -th component of the channel input and equalizer output, respectively. Vector CMA [112] applies the CMA cost to the entire equalizer output vector: $J_v(\mathbf{y}_k) = (\|\mathbf{y}_k\|^2 - R)^2$, where $R = \frac{E[\|\mathbf{x}_k\|^4]}{E[\|\mathbf{x}_k\|^2]}$. Combination CMA [112,113] uses a weighted sum of the pointwise and vector-CMA costs. Decorrelating CMA [114,115] adds a term to the pointwise-CMA cost to penalize correlations among equalizer output components at nonzero lags (in an effort to eliminate spurious local minima). These CMA extensions have lower complexity than the aforementioned batch methods, but they can suffer convergence problems. Other adaptive algorithms, proposed for memoryless or unitary channels, include EASI [116] (equivariant adaptive source separation via independence), which is based on a contrast function, and MPLL [117], a multidimensional extension of a decision-directed phase-locked loop.

Another class of blind algorithms, based on the seminal work of Gorokhov, Loubaton, and Moulines [15], has recently been developed for strictly tall channels. Gorokhov and colleagues [15] extended the work of Tong [17] to show that SOS can be sufficient to identify a tall MIMO channel up to a memoryless unitary ambiguity. The residual unitary matrix \mathbf{U} can be interpreted as a generalization of the complex scalar $e^{j\theta}$ left unresolved by the original TXK algorithm. In [15], Gorokhov *et al.* also extended the work of Slock

[94] to show that the output of an identifiable tall channel simultaneously has both an MA and an AR nature. The PEBI algorithm [91] of Abed-Meraim was then generalized, and a similar weighted-least-squares (WLS) algorithm [15, 118] was proposed for blind MIMO channel identification. Although HOS, and thus non-Gaussian channel inputs, are still required to identify \mathbf{U} (or $e^{j\theta}$ for SIMO channels), this is generally a much easier task than identifying $\mathbf{H}(z)$. This statistically decoupled approach reduces the need for HOS to a minimum. Any of the HOS-based methods, in particular, those designed for memoryless or unitary channels, can be applied to identify \mathbf{U} . For example, Icart and Gautier [119] proposed combining PEBI [91] with JADE [103] (joint approximate diagonalization of eigenmatrices). Throughout the course of our research, we adopt as our own this philosophy of *maximal exploitation of SOS*; however, our emphasis is the blind detection of the channel input \mathbf{x}_k rather than the identification of the channel, and our solutions are adaptive rather than batch-oriented.

CHAPTER 3

SUBSPACE METHODS

THIS CHAPTER EXPLORES so-called subspace methods for blind multiuser detection. The receiver observation lies in a vector space known as the receive space. Subspace methods are those which decompose the receive space into two or more smaller orthogonal subspaces, and then exploit their orthogonality. Examples include MUSIC [89], perhaps the best-known subspace algorithm, and the CDMA detectors of Wang and Poor [14] and others [120]. These algorithms are only partially blind in that they assume knowledge of the array manifold or of the desired user's spreading code. The subspace algorithms of Moulines [87-88] and Liu [93], in contrast, are fully blind, but like MUSIC, they are batch oriented and have relatively high complexity. In keeping with our philosophy, the subspace methods we present are adaptive, low complexity, and fully blind.

We limit consideration in this chapter to memoryless channels. A full understanding of the subspace ideas in this context can provide valuable intuition for the more complicated case. Moreover, the memoryless case merits study in its own right because a number of real-world applications are modeled as memoryless systems, *e.g.*, synchronous CDMA or array processing in environments without multipath. We extend the concepts and algorithms presented in this chapter to channels with memory in chapter 6.

In the following section, we briefly review the singular-value decomposition (SVD) and its relationship to the signal and noise subspaces. In section 3.2, we describe a low-complexity algorithm for blindly and adaptively separating the signal subspace from the noise subspace. This subspace separator, as we call it, requires neither batch processing nor explicit singular-value or eigendecompositions. It can be used in the front end of any receiver without loss of signal information. Thus, any blind multiuser detector can then be implemented in the reduced-dimensional signal space, with a commensurate reduction in complexity. Reducing the number of receiver parameters often increases the speed of convergence as well. In section 3.3, we generalize the subspace separation technique to perform an SVD of the channel, again adaptively and without batch processing. Numerical results demonstrate fast convergence and good performance. In section 3.4, we present fully blind implementations of the minimum-mean-square-error (MMSE) and zero-forcing (ZF) multiuser detectors based on the adaptive SVD algorithm. In section 3.4, we present a channel diagonalization algorithm, which facilitates transmission approaching channel capacity in single-user multi-channel applications.

3.1 Signal and Noise Subspaces

We begin by reviewing the singular-value decomposition (SVD) of the channel matrix and its relationship to the signal and noise subspaces. Consider a memoryless channel, for which the model of (1-1) reduces to

$$\mathbf{r}_k = \mathbf{H}\mathbf{x}_k + \mathbf{n}_k, \quad (3-1)$$

where \mathbf{H} is an $m \times n$ memoryless channel matrix. In a narrowband m -sensor linear-array application, the columns of \mathbf{H} represent the steering vectors for the n users, and in a syn-

chronous-CDMA application, with m chips per baud, the columns of \mathbf{H} represent the signature sequences of the n users. As in (1-1), \mathbf{x}_k is a vector of symbol sequences sent by n independent, finite-alphabet transmitters, \mathbf{r}_k is the corresponding receiver observation sequence of dimension m , and \mathbf{n}_k is noise. For the memoryless model of (3-1), we assume that \mathbf{H} has rank n , which implies that the channel is either square or tall ($m \geq n$). We further assume that the signal and noise are independent, zero mean, and satisfy¹² $E[\mathbf{x}_k \mathbf{x}_k^*] = \mathbf{I}$ and $E[\mathbf{n}_k \mathbf{n}_k^*] = \sigma^2 \mathbf{I}$, with $\sigma > 0$.

The observation vector \mathbf{r}_k has dimension m ; therefore, the receive space is \mathbf{C}^m , the set of m -dimensional complex vectors. However, because \mathbf{H} has rank n , the signal term $\mathbf{H}\mathbf{x}_k$ is restricted to an n -dimensional subspace of \mathbf{C}^m , referred to as the signal subspace.

Definition 3-1. For any memoryless channel matrix \mathbf{H} (3-1), the **signal subspace** \mathbf{S} is the range or column span of \mathbf{H} , and the **noise subspace** \mathbf{N} is left null space of \mathbf{H} :

$$\mathbf{S} = \text{range}(\mathbf{H}) = \{ \mathbf{s} : \mathbf{s} = \mathbf{H}\mathbf{x}, \mathbf{x} \in \mathbf{C}^n \} \quad (3-2)$$

$$\mathbf{N} = \text{null}(\mathbf{H}^*) = \{ \mathbf{n} : \mathbf{H}^* \mathbf{n} = \mathbf{0}, \mathbf{n} \in \mathbf{C}^m \}. \quad (3-3)$$

These subspaces are orthogonal complements, $\mathbf{N} = \mathbf{S}^\perp$, meaning that their union is \mathbf{C}^m , but their intersection is empty: $\mathbf{S} \cup \mathbf{N} = \mathbf{C}^m$, and $\mathbf{S} \cap \mathbf{N} = \emptyset$. The signal subspace is closely tied to the singular-value decomposition of \mathbf{H} .

Theorem 3-1. [121] For any complex $m \times n$ matrix \mathbf{H} with rank n , there exists a **singular-value decomposition** (SVD) of the form:

$$\mathbf{H} = \mathbf{U}\mathbf{S}\mathbf{V}^* \quad (3-4)$$

12. No assumption is made regarding source correlation at nonzero lags $E[\mathbf{x}_k \mathbf{x}_{k-l}^*]$, $l \neq 0$.

$$= \sum_{i=1}^n s_i \mathbf{u}_i \mathbf{v}_i^*, \quad (3-5)$$

where $\mathbf{U} = [\mathbf{u}_1 \dots \mathbf{u}_m]$ and $\mathbf{V} = [\mathbf{v}_1 \dots \mathbf{v}_n]$ are $(m \times m$ and $n \times n)$ unitary matrices, $\mathbf{S} = \begin{bmatrix} \tilde{\mathbf{S}} \\ \mathbf{0} \end{bmatrix}$ is $m \times n$, and $\tilde{\mathbf{S}} = \text{diag}(s_1 \dots s_n)$ is a unique $n \times n$ diagonal matrix with real and positive ordered diagonal components $s_1 \geq \dots \geq s_n > 0$, referred to as the **singular values** of \mathbf{H} .

Although \mathbf{S} in (3-4) is unique, the unitary matrices \mathbf{U} and \mathbf{V} are not unique. If $\mathbf{H} = \mathbf{USV}^*$ is a valid SVD then so is $(\mathbf{UT})\mathbf{S}(\mathbf{V}\tilde{\mathbf{T}})^*$ for any unitary matrices \mathbf{T} and $\tilde{\mathbf{T}}$ satisfying $\mathbf{T}\mathbf{S}\tilde{\mathbf{T}}^* = \mathbf{S}$. It can be shown that $\mathbf{T}\mathbf{S}\tilde{\mathbf{T}}^* = \mathbf{S}$ if and only if $\tilde{\mathbf{T}}$ commutes with $\tilde{\mathbf{S}}$ and $\mathbf{T} = \begin{bmatrix} \tilde{\mathbf{T}} & \mathbf{0} \\ \mathbf{0} & \mathbf{Q} \end{bmatrix}$, where \mathbf{Q} is any unitary matrix of dimension $m - n$. A diagonal $\tilde{\mathbf{T}}$ is one example, but $\tilde{\mathbf{T}}$ need not be diagonal if the singular values are not distinct. An explicit description of the SVD ambiguity is provided by the following lemma.

Lemma 3-1. SVD Ambiguity. Let \mathbf{H} be an $m \times n$ matrix with rank n and SVD $\mathbf{H} = \mathbf{USV}^*$. Let $d \leq n$ denote the number of distinct singular values so that $s_1 > s_2 > \dots > s_d > 0$, and let μ_i denote the multiplicity of s_i for $i \in \{1, \dots, d\}$. Any other SVD $\mathbf{H} = \tilde{\mathbf{U}}\tilde{\mathbf{S}}\tilde{\mathbf{V}}^*$ is related to \mathbf{USV}^* as follows:

$$\tilde{\mathbf{U}} = \mathbf{U} \begin{bmatrix} \tilde{\mathbf{T}} & \mathbf{0} \\ \mathbf{0} & \mathbf{Q} \end{bmatrix}, \quad (3-6)$$

$$\tilde{\mathbf{V}} = \mathbf{V}\tilde{\mathbf{T}}, \quad (3-7)$$

where \mathbf{Q} is a unitary matrix of dimension $m - n$, and where $\tilde{\mathbf{T}}$ is a block-diagonal and unitary matrix of dimension $n \times n$:

$$\tilde{\mathbf{T}} = \begin{bmatrix} \mathbf{T}_1 & & \mathbf{0} \\ & \mathbf{T}_2 & \\ \mathbf{0} & & \ddots \\ & & & \mathbf{T}_d \end{bmatrix}. \quad (3-8)$$

The blocks \mathbf{T}_i are unitary submatrices of dimension $\mu_i \times \mu_i$.

Proof: Because $\tilde{\mathbf{T}}$ is block-diagonal and the submatrices of $\tilde{\mathbf{S}}$ are proportional to the identity, $\tilde{\mathbf{S}}$ commutes with $\tilde{\mathbf{T}}$:

$$\tilde{\mathbf{S}}\tilde{\mathbf{T}} = \tilde{\mathbf{T}}\tilde{\mathbf{S}} = \begin{bmatrix} s_1 \mathbf{T}_1 & & \mathbf{0} \\ & s_2 \mathbf{T}_2 & \\ \mathbf{0} & & \ddots \\ & & & s_d \mathbf{T}_d \end{bmatrix}. \quad (3-9)$$

It follows that

$$\tilde{\mathbf{U}}\tilde{\mathbf{S}}\tilde{\mathbf{V}}^* = \mathbf{U} \begin{bmatrix} \tilde{\mathbf{T}} & \mathbf{0} \\ \mathbf{0} & \mathbf{Q} \end{bmatrix} \mathbf{S}\tilde{\mathbf{T}}^* \mathbf{V}^* = \mathbf{U}\tilde{\mathbf{T}} \tilde{\mathbf{T}}^* \mathbf{V}^* = \mathbf{U}\mathbf{S}\mathbf{V}^*. \quad (3-10)$$

Conversely, if (3-8) does not hold then $\tilde{\mathbf{S}}$ and $\tilde{\mathbf{T}}$ do not commute, and $\tilde{\mathbf{U}}\tilde{\mathbf{S}}\tilde{\mathbf{V}}^*$ cannot equal $\mathbf{U}\mathbf{S}\mathbf{V}^*$. \square

In terms of an SVD $\mathbf{H} = \mathbf{U}\mathbf{S}\mathbf{V}^*$, the signal subspace is spanned by the first n columns of \mathbf{U} . Observe from (3-4) that

$$\mathbf{H}\mathbf{x} = \sum_{j=1}^n s_j \mathbf{u}_j \mathbf{v}_j^* \mathbf{x} = \sum_{j=1}^n a_j \mathbf{u}_j, \quad (3-11)$$

where the scalar a_j satisfies $a_j = s_j \mathbf{v}_j^* \mathbf{x}$. Thus, the first n columns $[\mathbf{u}_1 \dots \mathbf{u}_n]$ of \mathbf{U} form a basis for the signal subspace. Consequently, the last $m - n$ columns $[\mathbf{u}_{n+1} \dots \mathbf{u}_m]$ form a

basis for the noise subspace. So, although an SVD is not unique, the signal and noise subspaces are uniquely determined by the left factor \mathbf{U} from any SVD.

Lemma 3-2. For any memoryless channel matrix \mathbf{H} (3-1) with an SVD given by $\mathbf{H} = \mathbf{USV}^*$, the signal and noise subspaces can be expressed in terms of the columns of $\mathbf{U} = [\mathbf{u}_1 \dots \mathbf{u}_m]$ as follows:

$$\mathbf{S} = \text{span}\{\mathbf{u}_1 \dots \mathbf{u}_n\} \quad (3-12)$$

$$\mathbf{N} = \text{span}\{\mathbf{u}_{n+1} \dots \mathbf{u}_m\}. \quad (3-13)$$

Because the signal space is only n -dimensional, we can confine the signal portion of the receiver observation to n components. In other words, we can define a new vector \mathbf{y}_k according to

$$\mathbf{y}_k = \begin{bmatrix} \tilde{\mathbf{r}}_k \\ \mathbf{w}_k \end{bmatrix} = \Theta \mathbf{r}_k, \quad (3-14)$$

where Θ is unitary, such that \mathbf{w}_k , the last $m - n$ components of \mathbf{y}_k , contains no signal contributions. All signal contributions are confined to $\tilde{\mathbf{r}}_k$, the first n components of \mathbf{y}_k . We interpret $\tilde{\mathbf{r}}_k$ as a projection of \mathbf{r}_k onto the signal subspace, and \mathbf{w}_k as a projection of \mathbf{r}_k onto the noise subspace. The sequence $\tilde{\mathbf{r}}_k$ provides sufficient statistics for recovering the channel input; \mathbf{w}_k is irrelevant. Thus, we say that Θ *separates the subspaces*. In general, we define a subspace-separation matrix as follows:

Definition 3-2. For the $m \times n$ memoryless channel \mathbf{H} of (3-1), an $m \times m$ unitary matrix Θ is a **subspace-separation matrix** if and only if the last $m - n$ rows of $\Theta\mathbf{H}$ are identically zero.

Lemma 3-2 suggests one obvious solution: $\Theta = \mathbf{U}^*$. In general, a subspace-separation matrix is related to \mathbf{U}^* as follows.

Lemma 3-3. For the memoryless channel \mathbf{H} of (3-1), a unitary subspace-separation matrix must be of the form:

$$\Theta = \begin{bmatrix} \mathbf{U}_S & \mathbf{0} \\ \mathbf{0} & \mathbf{U}_N \end{bmatrix} \mathbf{U}^*, \quad (3-15)$$

where \mathbf{U}_S and \mathbf{U}_N are arbitrary unitary matrices of dimension n and $m - n$ respectively, and \mathbf{U} is a left factor of any valid channel SVD $\mathbf{H} = \mathbf{USV}^*$.

Proof: The last $m - n$ rows of $\Theta\mathbf{H} = \Theta\mathbf{USV}^*$ must be zero. The SVD theorem implies that the last $m - n$ rows of \mathbf{SV}^* are already zero, so we need only ensure that $\Theta\mathbf{U}$ passes none of the energy from the first n inputs to the last $m - n$ outputs. Hence, $\Theta\mathbf{U}$ must be block diagonal and unitary, implying (3-15). \square

Observe that if Θ satisfies (3-15), then (3-14) reduces to

$$\begin{aligned} \begin{bmatrix} \tilde{\mathbf{r}}_k \\ \mathbf{w}_k \end{bmatrix} &= \Theta\mathbf{H}\mathbf{x}_k + \Theta\mathbf{n}_k \\ &= \begin{bmatrix} \tilde{\mathbf{H}} \\ \mathbf{0} \end{bmatrix} \mathbf{x}_k + \Theta\mathbf{n}_k, \end{aligned} \quad (3-16)$$

where we have introduced an $n \times n$ matrix $\tilde{\mathbf{H}} = \mathbf{U}_S\tilde{\mathbf{S}}\mathbf{V}^*$ with rank n . These last $m - n$ components can be used to estimate the noise variance, if desired, or simply discarded, thereby producing a square channel, as shown in Fig. 3-1 (b), with a reduced receiver observation:

$$\tilde{\mathbf{r}}_k = \tilde{\mathbf{H}}\mathbf{x}_k + \tilde{\mathbf{n}}_k, \quad (3-17)$$

where $E[\tilde{\mathbf{n}}_k\tilde{\mathbf{n}}_k^*] = \sigma^2\mathbf{I}$. Because $\tilde{\mathbf{r}}_k$ is sufficient for estimating \mathbf{x}_k , any multiuser detector, blind or otherwise, can be applied to this new channel $\tilde{\mathbf{H}}$ without compromising performance.

3.2 An Adaptive Signal-Noise Subspace Separator

We now present an algorithm for adaptively implementing the subspace separator Θ of (3-15). Our approach is based on the following observation: on the one hand, the noise energy in the last $m - n$ components of $\mathbf{y}_k = \Theta\mathbf{r}_k$ is the same for any unitary matrix Θ , but

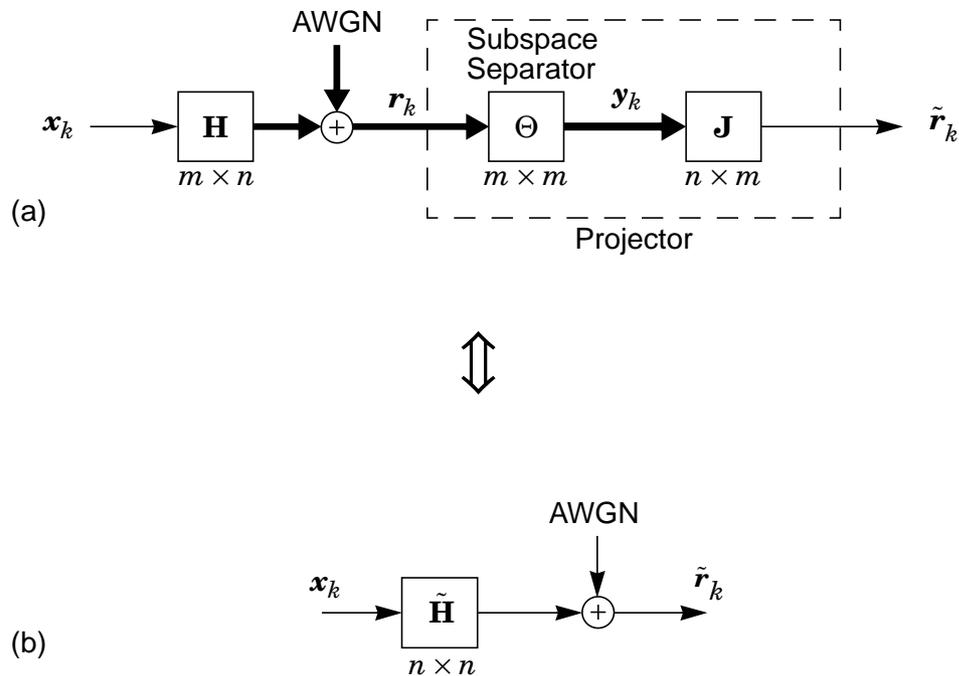


Figure 3-1. Equivalent models: (a) a tall channel with a signal-space projector used as the receiver front end, and (b) an equivalent square channel.

on the other hand, the signal energy in the last $m - n$ components of \mathbf{y}_k is minimized only when Θ satisfies (3-15). Thus we can reformulate the subspace separation problem into that of finding a unitary matrix Θ that minimizes the energy in the last $m - n$ components of its output. We seek a unitary matrix that minimizes the following cost function:

$$E[\|\mathbf{G}\mathbf{y}_k - \mathbf{y}_k\|^2] \quad (3-18)$$

where

$$\mathbf{G} = \begin{bmatrix} \mathbf{I}_n & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{bmatrix}. \quad (3-19)$$

In other words, we seek to rotate \mathbf{r}_k by a matrix Θ such that $\mathbf{y}_k = \Theta\mathbf{r}_k$ is close to $\mathbf{G}\mathbf{y}_k$.

In a manner similar to that used by multidimensional PLL of [117], we can adapt an estimate of Θ iteratively by accumulating unitary matrices that partially rotate $\mathbf{G}\mathbf{y}_k$ to \mathbf{y}_k :

$$\hat{\Theta}_{k+1} = \mathcal{R}^{\lambda}(\mathbf{G}\mathbf{y}_k \rightarrow \mathbf{y}_k)^* \hat{\Theta}_k. \quad (3-20)$$

Following [117], we define a rotation from \mathbf{x} to \mathbf{y} as a unitary matrix \mathcal{R} satisfying $\mathcal{R} \frac{\mathbf{x}}{\|\mathbf{x}\|} = \frac{\mathbf{y}}{\|\mathbf{y}\|}$ and $\mathcal{R}\mathbf{z} = \mathbf{z}$ for all \mathbf{z} orthogonal to $\text{span}\{\mathbf{x}, \mathbf{y}\}$ (the two-dimensional subspace spanned by \mathbf{x} and \mathbf{y}). A closed-form expression for \mathcal{R} is given by

$$\mathcal{R}(\mathbf{x} \rightarrow \mathbf{y}) = \mathbf{I} + \begin{bmatrix} \mathbf{u}, \mathbf{v} \end{bmatrix} \begin{bmatrix} p-1 & \frac{-p}{|p|} \sqrt{1-|p|^2} \\ \sqrt{1-|p|^2} & |p|-1 \end{bmatrix} \begin{bmatrix} \mathbf{u}^* \\ \mathbf{v}^* \end{bmatrix}, \quad (3-21)$$

where p is the normalized inner product, $p = \frac{\mathbf{x}^*\mathbf{y}}{\|\mathbf{x}\|\|\mathbf{y}\|}$, and where $\{\mathbf{u}, \mathbf{v}\}$ is a basis for $\text{span}\{\mathbf{x}, \mathbf{y}\}$: $\mathbf{u} = \frac{\mathbf{x}}{\|\mathbf{x}\|}$ and¹³ $\mathbf{v} = \frac{\mathbf{z}/\|\mathbf{z}\| - p\mathbf{u}}{\sqrt{1-|p|^2}}$. A partial rotation \mathcal{R}^{λ} is then defined as

13. For the case when \mathbf{x} and \mathbf{y} are colinear ($|p| = 1$), we take $\mathbf{v} = \mathbf{0}$ in (3-21).

$$\mathcal{R}^\lambda(\mathbf{x} \rightarrow \mathbf{y}) = \mathcal{R}(\mathbf{x} \rightarrow \lambda\mathbf{y} + (1-\lambda)\mathbf{x}). \quad (3-22)$$

In other words, a partial rotation \mathcal{R}^λ from \mathbf{x} to \mathbf{y} is a rotation \mathcal{R} from \mathbf{x} to an intermediate vector $\lambda\mathbf{y} + (1-\lambda)\mathbf{x}$ lying between \mathbf{x} and \mathbf{y} . The simple recursion of (3-20) with \mathbf{G} defined by (3-19) defines our proposed adaptive subspace separator. A block diagram of the adaptive separator is shown in Fig. 3-2. The block labeled \mathbf{G} takes the place of the decision device¹⁴ in the conventional MPLL. The vector $\mathbf{G}\mathbf{y}_k$ can be interpreted as a projection onto an estimate of the signal space \mathbf{S} . Only valid subspace separators Θ , satisfying (3-15), minimize the projection error.

The recursion of (3-20) requires the multiplication of two $m \times m$ matrices at each iteration. The complexity can be reduced even further by manipulating (3-20) into the following form:

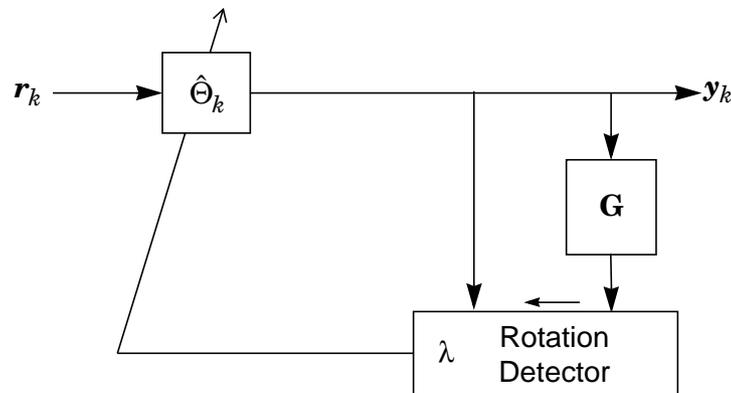


Figure 3-2. An adaptive signal-noise subspace separator.

14. Observe that the structure of the adaptive subspace separator resembles that of the classical Bussgang algorithms, except that \mathbf{G} of (3-19) is linear, rather than nonlinear.

$$\hat{\Theta}_{k+1} = \hat{\Theta}_k + \left[\mathcal{R}^{\lambda}(\mathbf{G}\mathbf{y}_k \rightarrow \mathbf{y}_k)^* - \mathbf{I} \right] \hat{\Theta}_k. \quad (3-23)$$

Observe that the rank-two matrix within the square brackets is the sum of four outer products; therefore, this realization of the algorithm requires the multiplication of only a rank-two matrix with an $m \times m$ matrix, and is thus less complex than (3-20) when $m > 2$. We remark that unlike many of the subspace estimation or tracking algorithms [122-125], the recursion of (3-20) or (3-23) provides a precisely orthonormal estimate of *both* the signal and noise subspaces at each iteration. We demonstrate convergence of the adaptive subspace separator of (3-23) by the following example.

Experiment 3-1. Consider a memoryless system (3-1) with $m = 10$ sensors and $n = 2$ users. According to (3-16), in order to show that (3-23) converges to a subspace separator, we need only demonstrate that the last $m - n$ rows of the separator-channel cascade $\hat{\Theta}_k \mathbf{H}$ converge to zero. Fig. 3-3 shows the energy in each row of $\hat{\Theta}_k \mathbf{H}$ as a function of time k , averaged over 100 randomly selected channels. For each trial, the elements $h_{j,i}$ of the channel matrix are selected independently from a zero-mean, unit-variance, complex Gaussian distribution, and then scaled such that the $\text{SNR}_i = \sum_{j=1}^m |h_{j,i}|^2 / \sigma^2$ of each user i is fixed at 27 dB. (The loop gain is $\lambda = 0.1(2^{-k/300})$ in (3-23).) We see that the energy in each of the last eight rows of $\hat{\Theta}_k \mathbf{H}$ converges quickly to levels of -40 dB or less.

Experiment 3-1 certainly does not prove convergence. Nevertheless, we make the following conjecture.

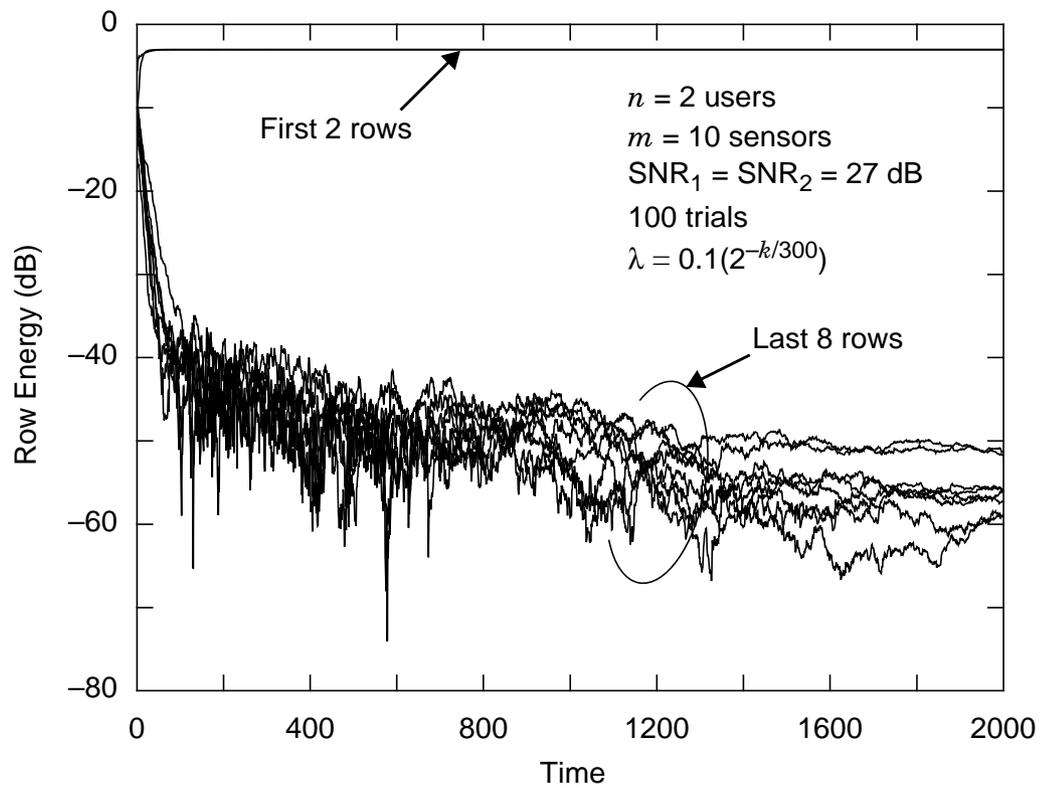


Figure 3-3. Convergence of the subspace separator: energy in the rows of the separator-channel cascade $\hat{\Theta}_k \mathbf{H}$ versus time k .

Conjecture 3-1. The recursion of (3-20) and (3-23) with \mathbf{G} defined by (3-19) converges to a valid subspace-separation matrix as defined by (3-15) for sufficiently small loop gain λ .

Heuristic support of this conjecture is offered in Appendix 3-2, and more experimental results regarding the subspace separator can be found in chapter 4, where it is used in the front end of a blind multiuser detector.

3.3 Adaptive Singular-Value Decomposition

Consider again the m -sensor n -user system of (3-1). According to Conjecture 3-1, the algorithm of (3-20) with \mathbf{G} defined by (3-19) converges to a subspace separator $\Theta = \begin{bmatrix} \mathbf{U}_S & \mathbf{0} \\ \mathbf{0} & \mathbf{U}_N \end{bmatrix} \mathbf{U}^*$, as defined by Lemma 3-3. The block-diagonal ambiguity can be problematic for some applications, where an estimate of \mathbf{U}^* itself is preferred or required. We now propose a simple modification to the subspace-separation algorithm that produces an estimate of \mathbf{U}^* , without any ambiguity other than that described by Lemma 3-1. In other words, we force \mathbf{U}_S to be the identity. (The other ambiguity \mathbf{U}_N can be absorbed as part of \mathbf{U}^* , in light of Lemma 3-1.) Once \mathbf{U} is estimated, it is easy to estimate \mathbf{S} and \mathbf{V} , leading to an adaptive SVD algorithm.

Recall from (3-19) that the first n diagonal elements of \mathbf{G} , corresponding to the signal subspace, are all one, and the last $m - n$ diagonal elements, corresponding to the noise subspace, are all zero. The structure of \mathbf{G} in effect forces the recursion of (3-20) to decompose the receive space into a “large energy” signal subspace and a “small energy” noise subspace. But we need not have only two subspaces. For example, if the singular values are distinct, then the n -dimensional signal subspace $\mathbf{S} = \text{span}\{\mathbf{u}_1 \dots \mathbf{u}_n\}$ can be

further decomposed into n one-dimensional subspaces $\mathbf{S}_i = \text{span}\{\mathbf{u}_i\} \forall i \in \{1, \dots, n\}$. We can distinguish among these smaller subspaces on the basis of the signal energy s_i^2 within each. In general, the signal subspace \mathbf{S} can be decomposed into $d \leq n$ subspaces \mathbf{S}_i of dimension $\mu_i \forall i \in \{1, \dots, d\}$, where d is the number of distinct singular values $s_1 > s_2 > \dots > s_d$, and where μ_i is the multiplicity of s_i . In other words, whereas the subspace separator Θ of (3-15) decomposes the receive space into two subspaces, one signal and one noise, the factor \mathbf{U}^* decomposes the receive space into $d+1$ subspaces, d signal and one noise; we call this a *complete subspace decomposition*. By modifying the diagonal elements of \mathbf{G} , we can perform a complete subspace decomposition and estimate \mathbf{U}^* or \mathbf{U} directly. Specifically, we define \mathbf{G} such that the first n elements are strictly decreasing and strictly greater than the remaining $m - n$ elements:

$$\begin{aligned} \mathbf{G} &= \text{diag}(g_1 \dots g_m) \\ g_1 &> g_2 > \dots > g_n > g_j \geq 0 \quad \forall j \in \{n+1 \dots m\}. \end{aligned} \quad (3-24)$$

The estimate of \mathbf{U} is then updated according to

$$\begin{aligned} \hat{\mathbf{U}}_{k+1} &= \hat{\mathbf{U}}_k \mathcal{R}^\lambda(\mathbf{G}\mathbf{y}_k \rightarrow \mathbf{y}_k) \\ &= \hat{\mathbf{U}}_k + \hat{\mathbf{U}}_k \left[\mathcal{R}^\lambda(\mathbf{G}\mathbf{y}_k \rightarrow \mathbf{y}_k) - \mathbf{I} \right], \end{aligned} \quad (3-25)$$

where $\mathbf{y}_k = \hat{\mathbf{U}}_k^* \mathbf{r}_k$. The block diagram of this adaptive rotator is identical to Fig. 3-2 except that $\hat{\mathbf{U}}_k^*$ replaces $\hat{\Theta}_k$, and \mathbf{G} is defined by (3-24) rather than (3-19).

A demonstration of the convergence of (3-25) is given in Experiment 3-2 later in this section, and again, although this experiment is not a proof, we state the following conjecture.

Conjecture 3-2. The recursion of (3-25) with \mathbf{G} defined by (3-24) converges to a left factor \mathbf{U} in a valid channel SVD $\mathbf{H} = \mathbf{USV}^*$ for sufficiently small loop gain λ .

Conjecture 3-2 does not suggest that (3-25) converges to a unique solution, but rather only to a \mathbf{U} corresponding to one of many possible singular-value decompositions $\mathbf{H} = \mathbf{USV}^*$. See Appendix 3-2 for more discussion regarding convergence.

Invoking Conjecture 3-2, the rotator output converges to $\mathbf{y}_k = \mathbf{U}^* \mathbf{r}_k$, which can be written as

$$\mathbf{y}_k = \begin{bmatrix} \tilde{\mathbf{r}}_k \\ \mathbf{w}_k \end{bmatrix} = \begin{bmatrix} \tilde{\mathbf{H}} \\ \mathbf{0} \end{bmatrix} \mathbf{x}_k + \mathbf{U}^* \mathbf{n}_k, \quad (3-26)$$

where $\tilde{\mathbf{H}} = \tilde{\mathbf{S}} \mathbf{V}^*$. The autocorrelation of \mathbf{y}_k is then given by

$$\Phi_{\mathbf{y}} = \mathbf{S} \mathbf{S}^* + \sigma^2 \mathbf{I}. \quad (3-27)$$

Therefore, the energy in the i -th component of \mathbf{y}_k is

$$\varepsilon_k^{(i)} = E[|y_k^{(i)}|^2] = \begin{cases} s_i^2 + \sigma^2 & i \in \{1, \dots, n\} \\ \sigma^2 & i \in \{n+1, \dots, m\} \end{cases}. \quad (3-28)$$

These energies are in fact the eigenvalues of $\Phi_{\mathbf{r}} = E[\mathbf{r}_k \mathbf{r}_k^*] = \mathbf{H} \mathbf{H}^* + \sigma^2 \mathbf{I}$. The following algorithm can be used for calculating the singular values of the channel \mathbf{H} . At each time k , form an estimate $\hat{\varepsilon}_k^{(i)}$ of the energy in the i -th component $y_k^{(i)}$ of \mathbf{y}_k , using a simple first-order recursion:

$$\hat{\varepsilon}_k^{(i)} = \alpha \hat{\varepsilon}_{k-1}^{(i)} + (1 - \alpha) |y_k^{(i)}|^2, \quad (3-29)$$

where $0 < \alpha < 1$ is a smoothing factor. Then, estimate the singular values s_i of \mathbf{H} and the noise energy σ^2 as follows:

$$\hat{\sigma}_k^2 = \frac{1}{m-n} \sum_{i=n+1}^m \hat{\epsilon}_k^{(i)}, \quad (3-30)$$

$$\hat{s}_i(k) = (\hat{\epsilon}_k^{(i)} - \hat{\sigma}_k^2)^{1/2}. \quad (3-31)$$

In summary, the adaptive SVD algorithm is described by (3-25) and (3-29) through (3-31).

We now demonstrate the algorithm with a computer experiment.

Experiment 3-2. Consider a 4-user QPSK system with 10 sensors, modeled by a randomly generated channel \mathbf{H} of dimension 10×4 , with noise variance $\sigma^2 = 0.01$. Let \mathbf{G} be diagonal with linearly decreasing elements: $\mathbf{G} = \text{diag}(1, 8/9, \dots, 1/9, 0)$. We estimate \mathbf{U} according to (3-25) with $\lambda = 2 \cdot 2^{-k/500}$. We demonstrate convergence of $\hat{\mathbf{U}}_k$ (3-25) in Fig. 3-4(a), where we plot the diagonal elements of $\hat{\mathbf{U}}_k^* \Phi_r \hat{\mathbf{U}}_k$ versus time k . We can interpret these diagonal elements as the true energy in the components of $\mathbf{y}_k = \hat{\mathbf{U}}_k^* \mathbf{r}_k$. We expect that as $\hat{\mathbf{U}}_k$ approaches \mathbf{U} , the matrix $\hat{\mathbf{U}}_k^* \Phi_r \hat{\mathbf{U}}_k$ should approach $\mathbf{S}\mathbf{S}^* + \sigma^2\mathbf{I}$. Fig. 3-4(a) verifies that this is indeed the case. We demonstrate convergence of $\hat{\epsilon}_k^{(i)}$ (3-29) in Fig. 3-4(b), where we plot the estimates $\hat{\epsilon}_k^{(i)} = \mathbf{E}[|y_k^{(i)}|^2]$ versus time with $\alpha = 1 - 0.05/(1 + k/90)$. We see that the simple algorithm is able to accurately estimate the singular values in only a few hundred iterations.

The blind adaptive SVD described in this section is used as a building block in the blind implementations of the MMSE and the ZF detectors of the following sections.

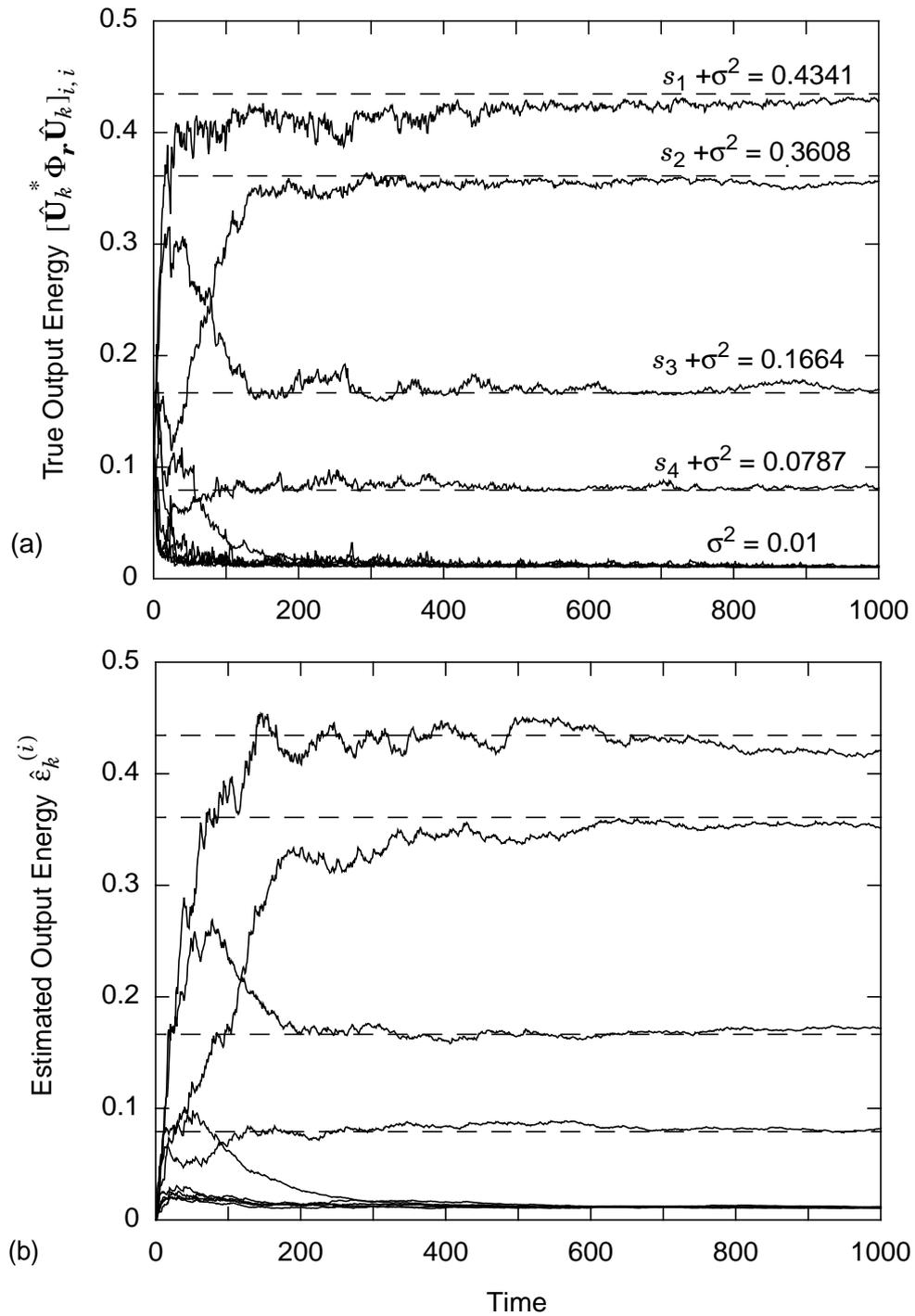


Figure 3-4. Adaptive estimation of singular values: (a) the diagonal elements of $\hat{\mathbf{U}}_k^* \Phi_r \hat{\mathbf{U}}_k$; (b) eigenvalue estimates from (3-29).

3.4 MMSE Detection

Consider again the memoryless multiuser communication system $\mathbf{r}_k = \mathbf{H}\mathbf{x}_k + \mathbf{n}_k$ of (3-1). A linear multiuser detector processes the receiver observation with an $n \times m$ matrix \mathbf{C} , producing $\mathbf{z}_k = \mathbf{C}\mathbf{r}_k$. The resulting MSE for user i is $\text{MSE}_i = E[|z_k^{(i)} - x_k^{(i)}|^2]$. The “best” linear detector is generally considered to be the minimum-mean-square-error (MMSE) detector; that is, the detector that minimizes the MSE for each user.

Definition 3-3. The **minimum-MSE detector** \mathbf{C}_{MMSE} for the channel of (3-1) is the $n \times m$ matrix \mathbf{C} that minimizes the MSE sum $E[\|\mathbf{C}\mathbf{r}_k - \mathbf{x}_k\|^2]$.

Expressions for the MMSE detector have been derived in [126] and [127]. The following lemma expresses the MMSE detector in terms of the memoryless model of (3-1). It also suggests an implementation procedure.

Lemma 3-4. For channel \mathbf{H} of (3-1) with $\sigma^2 > 0$, the MMSE detector can be expressed in three equivalent ways:

$$\mathbf{C}_{MMSE} = \mathbf{H}^*(\mathbf{H}\mathbf{H}^* + \sigma^2\mathbf{I})^{-1} \quad (3-32)$$

$$= (\mathbf{H}^*\mathbf{H} + \sigma^2\mathbf{I})^{-1}\mathbf{H}^* \quad (3-33)$$

$$= \mathbf{V}\mathbf{D}\mathbf{U}^*, \quad (3-34)$$

where \mathbf{V} and \mathbf{U} are unitary factors of a channel SVD $\mathbf{H} = \mathbf{U}\mathbf{S}\mathbf{V}^*$, and where $\mathbf{D} = \mathbf{S}^*(\mathbf{S}\mathbf{S}^* + \sigma^2\mathbf{I})^{-2}$ is a positive diagonal gain matrix of dimension $n \times m$. The MMSE detector is unique.

Proof: See Appendix 3-1.

Note that \mathbf{D} can be decomposed into the form $\mathbf{D} = [\tilde{\mathbf{D}} \ \mathbf{0}]$, where $\tilde{\mathbf{D}} = \tilde{\mathbf{S}}(\tilde{\mathbf{S}}^2 + \sigma^2\mathbf{I})^{-2}$ is diagonal.

The last equivalence (3-34) suggests a realization of the MMSE detector consisting of the cascade of three filters: first, the unitary \mathbf{U}^* , second, the diagonal \mathbf{D} , and last, the unitary \mathbf{V} . In section 3.3 we have already presented a blind method that implements the first filter \mathbf{U}^* . Furthermore, we have also presented a blind method for estimating the singular values \mathbf{S} and noise power σ^2 ; see (3-29) through (3-31). Thus, the second filter $\mathbf{D} = \mathbf{S}^*(\mathbf{S}\mathbf{S}^* + \sigma^2\mathbf{I})^{-2}$ can also be implemented blindly.

A blind implementation of the entire MMSE detector is now apparent; its block diagram is shown in Fig. 3-5. The detector is based on a *rotate-scale-rotate* architecture. The first step is to rotate the observation with an $m \times m$ unitary filter $\hat{\mathbf{U}}_k$ adapted according to (3-25). The second step is to scale the output of $\hat{\mathbf{U}}_k$ by the diagonal $n \times m$ matrix $\hat{\mathbf{D}}_k = \hat{\mathbf{S}}_k^*(\hat{\mathbf{S}}_k\hat{\mathbf{S}}_k^* + \hat{\sigma}_k^2\mathbf{I})^{-2}$, where $\hat{\mathbf{S}}_k$ and $\hat{\sigma}_k^2$ are estimates of \mathbf{S} and σ^2 adapted according to (3-29) through (3-31). The final step is to rotate the output of $\hat{\mathbf{D}}_k$ by an estimate of the $n \times n$ unitary matrix \mathbf{V} . Although the SVD in (3-34) is not unique, once the front-end rotator $\hat{\mathbf{U}}_k$ converges to a *particular* \mathbf{U} , the corresponding factor \mathbf{V} is *uniquely* specified. (To see this substitute \mathbf{C}_{MMSE}^* for \mathbf{H} in Lemma 3-1.) Once $\hat{\mathbf{D}}$ has also converged properly, its output is $\mathbf{w}_k = \mathbf{D}\mathbf{U}^*\mathbf{r}_k$. Hence, of all unitary matrices $\hat{\mathbf{V}}$, only $\hat{\mathbf{V}} = \mathbf{V}$ minimizes the MSE sum $E[\|\hat{\mathbf{V}}\mathbf{w}_k - \mathbf{x}_k\|^2]$. But the problem of finding the best unitary matrix to minimize this MSE sum is precisely the problem for which the MPLL was designed. Then we can estimate \mathbf{V} according to the MPLL recursion [117]:

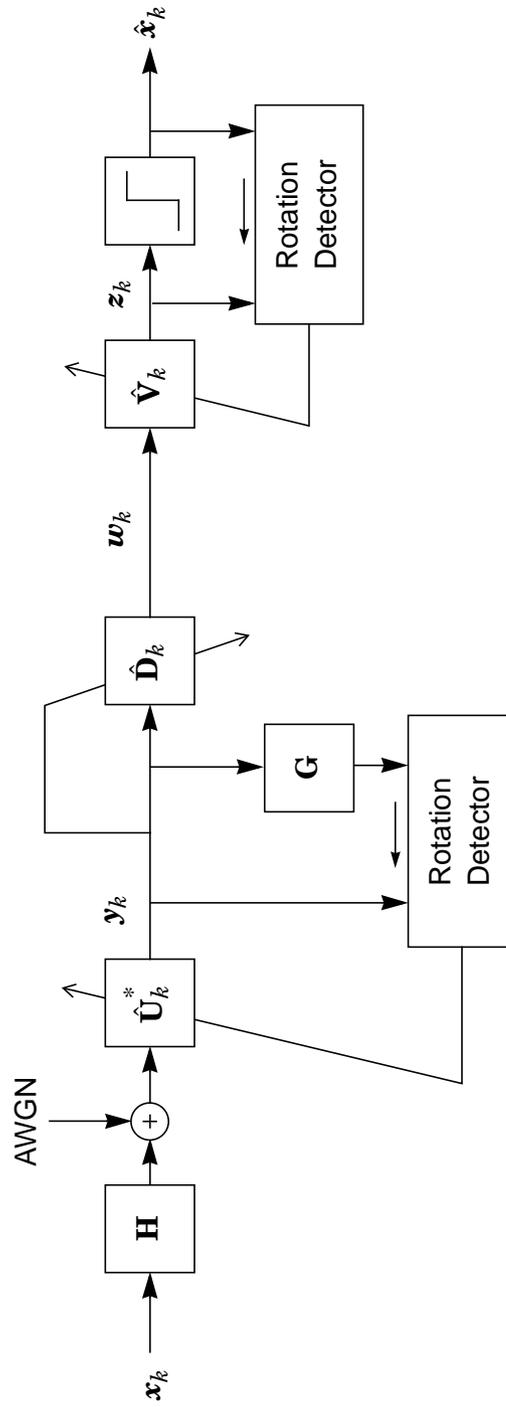


Figure 3-5. A fully blind, adaptive implementation of the minimum-MSE multiuser detector.

$$\begin{aligned}
\hat{\mathbf{V}}_{k+1} &= \mathcal{R}^{\lambda}(\hat{\mathbf{x}}_k \rightarrow \mathbf{z}_k)^* \hat{\mathbf{V}}_k \\
&= \hat{\mathbf{V}}_k + \left[\mathcal{R}^{\lambda}(\hat{\mathbf{x}}_k \rightarrow \mathbf{z}_k)^* - \mathbf{I} \right] \hat{\mathbf{V}}_k,
\end{aligned} \tag{3-35}$$

where $\mathbf{z}_k = \hat{\mathbf{V}}_k \mathbf{w}_k$, and where $\hat{\mathbf{x}}_k = \mathbf{dec}(\mathbf{z}_k)$ is a decision vector. Each component $\hat{x}_k^{(i)}$ of $\hat{\mathbf{x}}_k$ is quantized independently such that $\hat{x}_k^{(i)} = \text{dec}_i(y_k^{(i)})$ is the point in the constellation of user i closest to $y_k^{(i)}$.

Because of the ambiguities inherent in any fully blind detection algorithm, (3-35) may not converge to \mathbf{V} exactly. Identically distributed users are statistically indistinguishable, so they are arbitrarily labeled at the output of any fully blind detector. Moreover, the constellation of each user has rotational symmetries that cannot be blindly resolved. Rotating any square QAM constellation by an integer multiple of 90° , for example, does not change its statistics. In practice, these ambiguities are of little consequence, because they can be resolved by other means.¹⁵ Therefore, it is generally satisfactory if $\hat{\mathbf{x}}_k = \mathbf{K}\mathbf{x}_k$, where $\mathbf{K} = \mathbf{K}_P \mathbf{K}_R$ is the $n \times n$ product of a permutation matrix \mathbf{K}_P and a diagonal unitary matrix $\mathbf{K}_R = \text{diag}(e^{j\theta_i})$, and where the angles θ_i are determined by the rotational symmetries of constellation i . If all users transmit 16-QAM, for example, then \mathbf{K} is a *complex* permutation matrix, *i.e.* a matrix with exactly one nonzero element from $\{\pm 1, \pm j\}$ per row and per column. The following simulation experiment supports the conjecture that the MPLL converges to

$$\hat{\mathbf{V}} = \mathbf{K}\mathbf{V}. \tag{3-36}$$

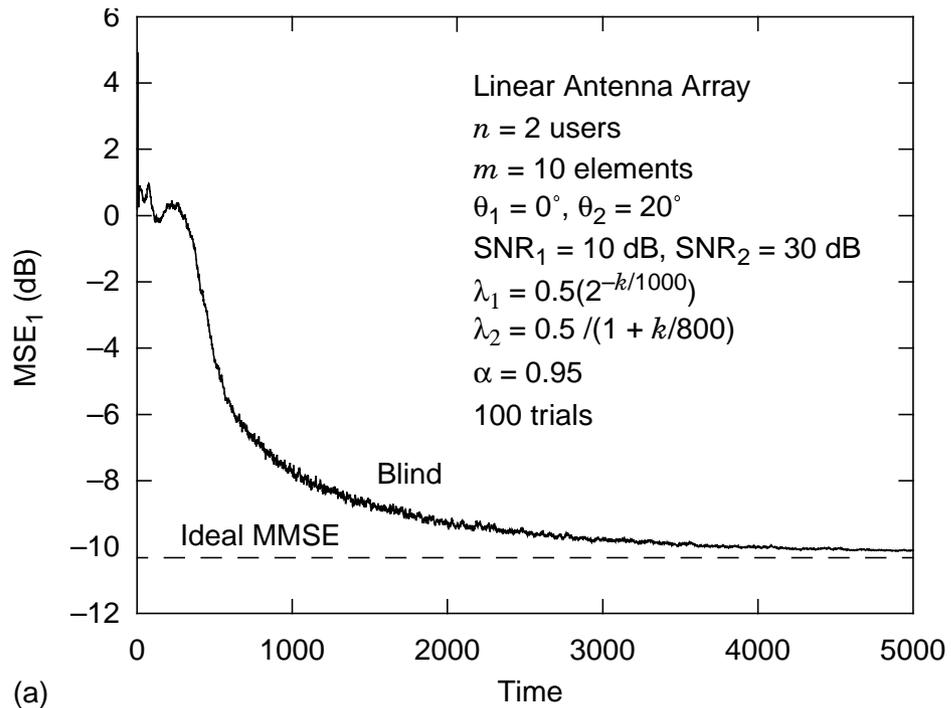
It also demonstrates convergence of the entire blind MMSE detector.

15. For example, differential encoding renders absolute phase irrelevant.

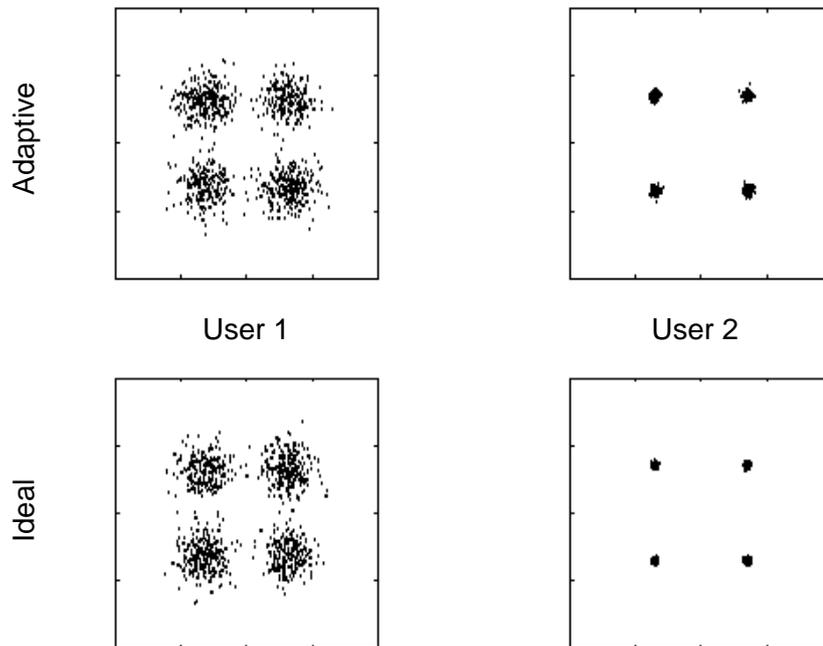
Experiment 3-3. Consider a 10-element linear antenna array with half-wavelength spacing, with two independent QPSK transmitters incident at angles $\theta_1 = 0^\circ$ and $\theta_2 = 20^\circ$, as measured from the broadside. For this arrangement, the channel model is given by (3-1) with $\mathbf{H}_{10 \times 2} = \frac{1}{\sqrt{10}} \mathbf{V}\mathbf{B}$, where $V_{i,l} = \exp\{j\frac{\pi}{\lambda}(i-1)\sin(\theta_l)\}$ [99], and $\mathbf{B} = \text{diag}(B_1 \ B_2)$, and where B_i^2 is the received power of the i -th user. We set $\text{SNR}_1 = 10$ dB and $\text{SNR}_2 = 30$ dB, so that the signal-to-interference ratio for user 1 is $\text{SIR}_1 = -20$ dB. Fig. 3-6(a) shows a plot of the MSE for user 1 versus time for the proposed blind MMSE detector of Fig. 3-5, averaged over 100 input and noise realizations. (The effect of the ambiguous permutation is removed for each trial.) The ideal minimum-MSE benchmark is shown for comparison. Fig. 3-6(b) shows the constellations from time 4000 to time 5000 of the last trial. We see that the proposed algorithm approaches the performance of the minimum-MSE detector, even in the presence of severe near-far interference, without the need for a training sequence and without knowledge of either the channel \mathbf{H} or use of transmitter training sequences. (The receiver parameters are $\lambda_1 = 0.5(2^{-k/1000})$ in (3-25), $\lambda_2 = 0.5/(1 + k/800)$ in (3-35), and $\alpha = 0.95$ in (3-29).)

3.5 Zero-Forcing Detection

The MMSE linear multiuser detector does not completely eliminate multiuser interference, but rather finds the best compromise between interference and noise; it is analogous to the MMSE linear equalizer in the single-user context, which finds the best compromise between intersymbol interference and noise. Complete elimination of multiuser (or intersymbol) interference is generally not a good idea because it can lead to



(a)



(b)

Figure 3-6. The blind adaptive MMSE detector of Experiment 3-3: (a) an MSE learning curve for user 1; (b) the constellations from the last trial, baud 4000 to 5000.

excessive noise enhancement. It can be argued, however, that in some contexts, complete elimination of multiuser interference is desirable, regardless of noise enhancement. Such a detector is called a zero-forcing detector.

Definition 3-4. For the channel \mathbf{H} of (3-1), a **zero-forcing detector** \mathbf{C}_{ZF} is an $n \times m$ matrix satisfying $\mathbf{C}\mathbf{H} = \mathbf{I}$.

The form of a zero-forcing detector is given by the following lemma.

Lemma 3-5. For the channel \mathbf{H} of (3-1), the ZF detector can be expressed in two equivalent ways:

$$\mathbf{C}_{ZF} = \mathbf{H}^\dagger + \mathbf{N}, \quad (3-37)$$

$$= \mathbf{V}\mathbf{S}^\dagger\mathbf{U}^* + \mathbf{N}, \quad (3-38)$$

where \mathbf{V} and \mathbf{U} are unitary factors of an SVD $\mathbf{H} = \mathbf{U}\mathbf{S}\mathbf{V}^*$, $\mathbf{S}^\dagger = [\tilde{\mathbf{S}}^{-1} \ \mathbf{0}]$, and $\mathbf{N}^* \in \text{null}(\mathbf{H}^*)$. If $m > n$, then the ZF detector is not unique.

If we take \mathbf{N} to be $\mathbf{0}$, we can blindly implement a (minimum-norm) ZF detector using the rotate-scale-rotate architecture of the previous section. We need only to replace the estimate of \mathbf{D} with that of \mathbf{S}^\dagger . We can estimate \mathbf{S} using (3-29) through (3-31) just as before, and then invert the singular values to form \mathbf{S}^\dagger .

3.6 Channel Diagonalization

As we discussed in chapter 1, there are some single-user applications that can be cast into a multiuser framework. We now change our focus somewhat and address one such

application. Consider a single-user multichannel communication system consisting of one transmitter with n antennas, and one receiver with m antennas. An example of such a system is the BLAST [31] system of Bell Labs. If the signal bandwidth is much smaller than the channel bandwidth, the transfer function from transmitter to receiver for such a system is memoryless. The observation is given by $\mathbf{r}_k = \mathbf{H}\mathbf{x}_k + \mathbf{n}_k$ of (3-1), but the components of \mathbf{x}_k now correspond to symbols transmitted by the n antennas, rather than n users. Strictly speaking, this is not a multiuser problem, although it can be cast into a multiuser framework by treating the input to each transmitter antenna as a virtual user. However, there are important differences: unlike a true multiuser problem, the virtual users are co-located and thus do not compete for bandwidth. Moreover, they are not necessarily independent and can in fact easily cooperate.

With the total power, averaged over all transmit antennas, constrained according to $\sum_{i=1}^n E[|x_k^{(i)}|^2] \leq P$, we address the question of how to achieve Shannon capacity. Brandenburg and Wyner [128] showed that capacity can be approached by using a channel diagonalization procedure. If the channel matrix \mathbf{H} were known, its SVD $\mathbf{H} = \mathbf{U}\mathbf{S}\mathbf{V}^*$ could be used to design both a transmitter *precoder* \mathbf{V} and a receiver *front-end* filter \mathbf{U}^* such that the overall system is diagonal:

$$\mathbf{S} = \mathbf{U}^* \mathbf{H} \mathbf{V}. \quad (3-39)$$

This system is illustrated in Fig. 3-7. Because the unitary filters are invertible, they are information preserving. The capacity of the resulting diagonal channel \mathbf{S} is thus identical to that of the original channel \mathbf{H} . However, because the subchannels are now decoupled, the transmitter can intelligently (according to the well-known water-pouring procedure [50]) distribute power and information among them.

An adaptive implementation of the channel diagonalization algorithm is shown in Fig. 3-8. Let $\hat{\mathbf{U}}^*$ and $\hat{\mathbf{V}}$ (both unitary) denote an adaptive receiver front-end filter and transmitter precoder, respectively, and let \mathbf{w}_k denote the precoder input, so that the vector of symbols transmitted at time k is $\mathbf{x}_k = \hat{\mathbf{V}}_k \mathbf{w}_k$. Without loss of generality, we take the transmitter power constraint to be $P = n$. We also assume that prior to the convergence of these filters, the power is equally distributed among all transmitter antennas, and that the symbols are uncorrelated across antennas so that $E[\mathbf{w}_k \mathbf{w}_k^*] = \mathbf{I}$. As in section 3.3, we can use the recursion of (3-25) to update $\hat{\mathbf{U}}^*$. Because the precoder $\hat{\mathbf{V}}$ is always unitary, the

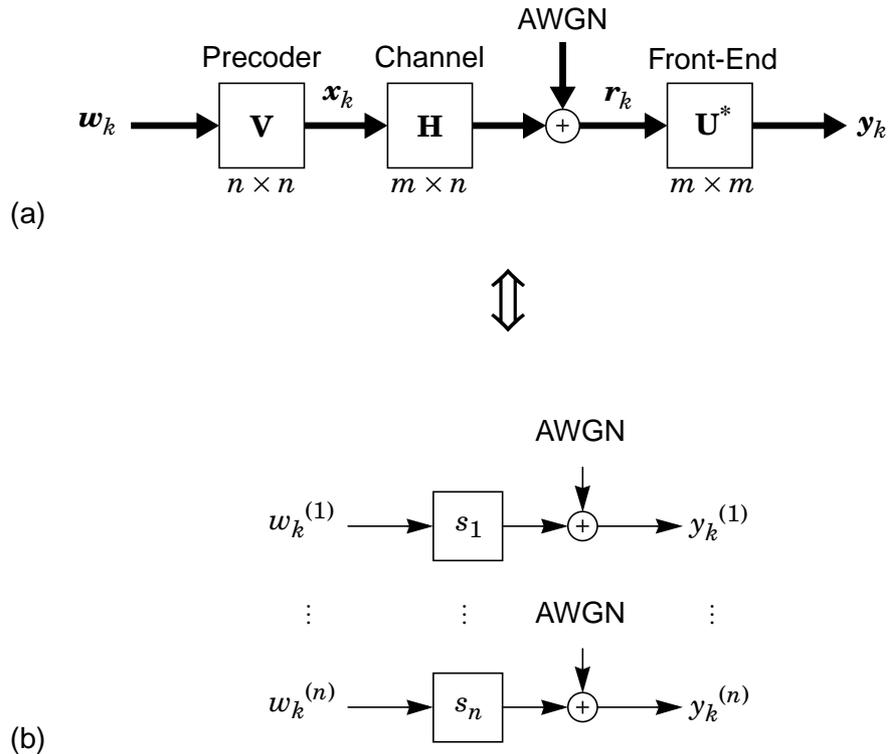


Figure 3-7. Equivalent models: (a) a memoryless MIMO channel with precoder and front-end rotation filter, and (b) decoupled scalar channels.

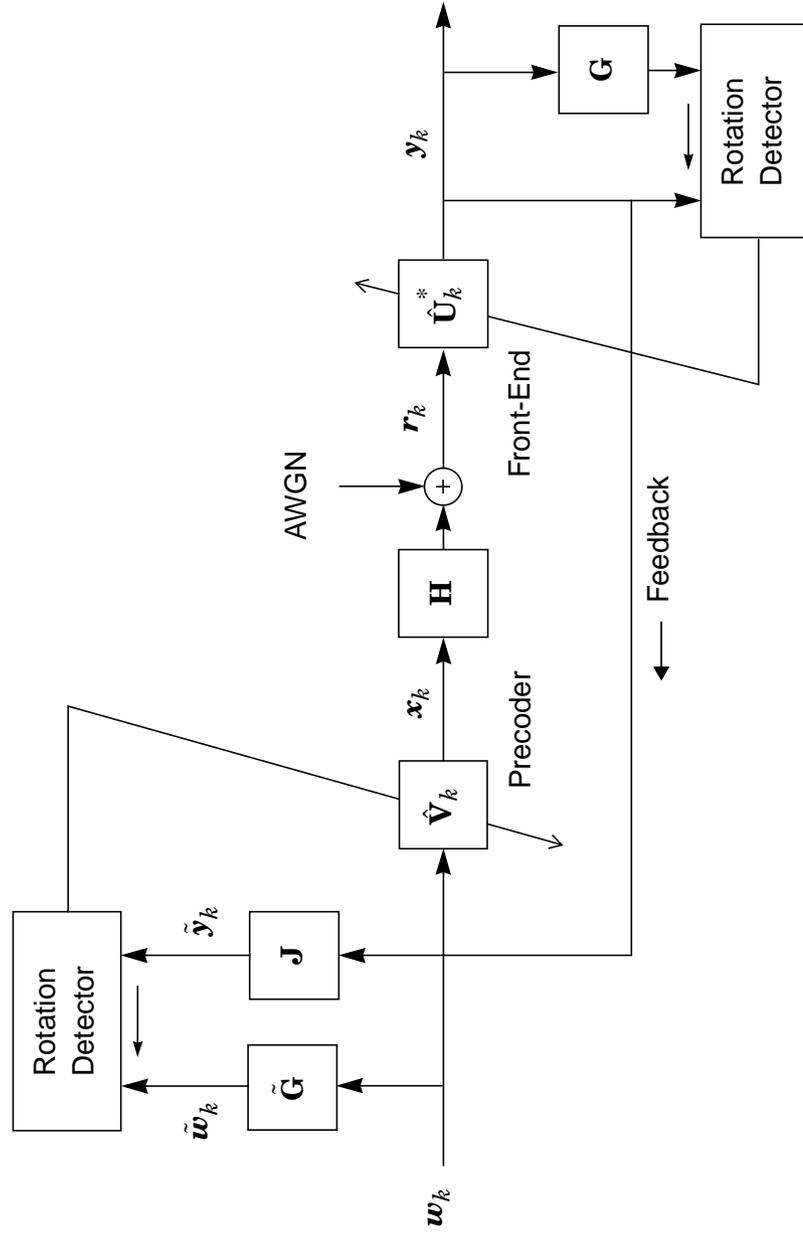


Figure 3-8. A block diagram of adaptive channel-diagonalization algorithm.

second-order statistics of \mathbf{w}_k and \mathbf{x}_k are identical. Therefore, the front-end filter $\hat{\mathbf{U}}^*$ converges to a valid \mathbf{U}^* , without regard to the precoder initialization or transient behavior.

According to Conjecture 3-2, the front-end filter $\hat{\mathbf{U}}^*$ converges to a particular \mathbf{U}^* , and the output of the front-end filter is given by

$$\mathbf{y}_k = \mathbf{S}\mathbf{V}^*\hat{\mathbf{V}}_k\mathbf{w}_k + \tilde{\mathbf{n}}_k, \quad (3-40)$$

where \mathbf{V} is unique, and where $\tilde{\mathbf{n}}_k = \hat{\mathbf{U}}_k^*\mathbf{n}_k$ and \mathbf{n}_k have identical second-order statistics. The output \mathbf{y}_k can be fed back to the transmitter where it can be used to adapt the precoder. Clearly, if $\hat{\mathbf{V}} = \mathbf{V}$, then $\mathbf{y}_k = \mathbf{S}\mathbf{w}_k + \tilde{\mathbf{n}}_k$. In particular, $\tilde{\mathbf{y}}_k = \mathbf{J}\mathbf{y}_k$ is a noisy estimate of $\tilde{\mathbf{S}}\mathbf{w}_k = \mathbf{J}\mathbf{S}\mathbf{w}_k$, where $\mathbf{J} = [\mathbf{I} \ \mathbf{0}]$. So, the transmitter seeks a rotation that maps $\tilde{\mathbf{y}}_k$ to a vector:

$$\tilde{\mathbf{w}}_k = \tilde{\mathbf{S}}\mathbf{w}_k. \quad (3-41)$$

Because the transmitter has access to \mathbf{w}_k , the singular values can be estimated, even for square channels, by averaging as follows:

$$\hat{s}_i(k+1) = \alpha\hat{s}_i(k) + (1 - \alpha) \text{Re}(y_k^{(i)}/w_k^{(i)}), \quad (3-42)$$

where $i = 1 \dots n$, and where $0 < \alpha < 1$ is a smoothing factor. In practice, however, estimation of $\tilde{\mathbf{S}}$ is not necessary for diagonalization; just as in the receiver, any $n \times n$ positive diagonal matrix with strictly decreasing elements, *e.g.* $\tilde{\mathbf{G}} = \mathbf{J}\mathbf{G}\mathbf{J}^T$, can be used in place of $\tilde{\mathbf{S}}$. The estimate of \mathbf{V} is updated by accumulating a matrix that partially rotates $\tilde{\mathbf{y}}_k$ to $\tilde{\mathbf{w}}_k$:

$$\begin{aligned} \hat{\mathbf{V}}_{k+1} &= \hat{\mathbf{V}}_k \mathcal{R}^\lambda(\tilde{\mathbf{y}}_k \rightarrow \tilde{\mathbf{w}}_k) \\ &= \hat{\mathbf{V}}_k + \hat{\mathbf{V}}_k \left[\mathcal{R}^\lambda(\tilde{\mathbf{y}}_k \rightarrow \tilde{\mathbf{w}}_k) - \mathbf{I} \right]. \end{aligned} \quad (3-43)$$

Observe that $\tilde{\mathbf{y}}_k$ and $\tilde{\mathbf{w}}_k$ in (3-43) play the roles of $\mathbf{G}\mathbf{y}_k$ and \mathbf{y}_k , respectively, in (3-25); that is, we accumulate partial rotations from $\tilde{\mathbf{y}}_k$ to $\tilde{\mathbf{w}}_k$ in (3-43), but from $\mathbf{G}\mathbf{y}_k$ to \mathbf{y}_k in (3-25). This role reversal is a consequence of the fact that $\hat{\mathbf{V}}_k$ prefilters the channel, whereas $\hat{\mathbf{U}}_k^*$ postfilters it.

To measure the effectiveness of this algorithm, we introduce a *diagonalization metric* defined as follows:

$$\zeta_k = \|\mathbf{J}\hat{\mathbf{U}}_k^* \mathbf{H}\hat{\mathbf{V}}_k - \tilde{\mathbf{S}}\|_F^2, \quad (3-44)$$

where ‘ $\|\cdot\|_F^2$ ’ indicates a Frobenius norm. The metric is in effect the squared distance between $\mathbf{J}\hat{\mathbf{U}}_k^* \mathbf{H}\hat{\mathbf{V}}_k$ and $\tilde{\mathbf{S}}$. For perfect diagonalization, the metric is zero. The following experiment demonstrates convergence of the algorithm.

Experiment 3-4. We consider randomly (Gaussian) generated channels of dimension 10×4 . All 4 transmitters use QPSK. We conduct 50 trials, adaptively diagonalizing each channel according to the proposed algorithm (with $\lambda_1 = 0.5(1 + (k/700)^2)$ in (3-25) and $\lambda_2 = 0.5(1 + (k/800)^2)$ in (3-43)). Fig. 3-9 shows the ensemble average of the diagonalization metric ζ_k versus time k . We see that the metric quickly approaches -30 dB.

3.7 Chapter Summary

We have proposed to adapt a unitary matrix \mathbf{Q} for operation on an observation \mathbf{r}_k according to the following recursion:

$$\mathbf{Q}_{k+1} = \mathcal{R}^\lambda(\mathbf{G}\mathbf{Q}_k\mathbf{r}_k \rightarrow \mathbf{Q}_k\mathbf{r}_k)^* \mathbf{Q}_k, \quad (3-45)$$

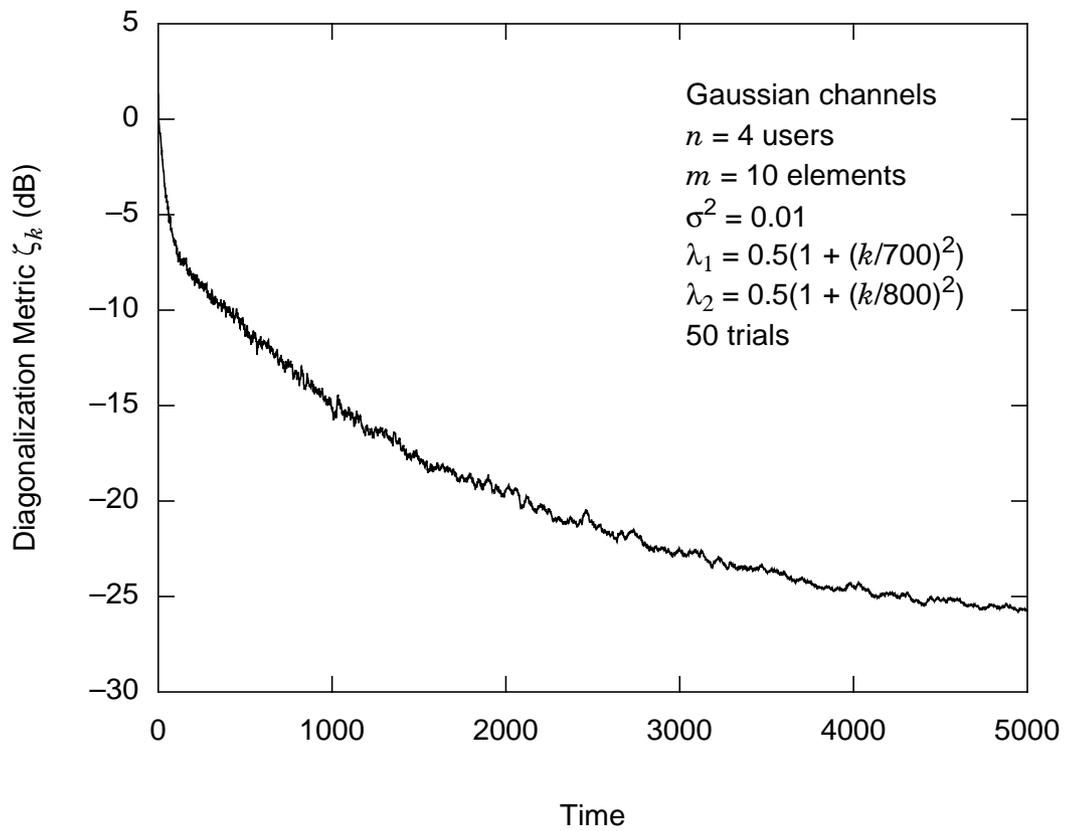


Figure 3-9. Convergence of the diagonalization algorithm: diagonalization metric ζ_k versus time k .

where $\mathcal{R}^\lambda(\mathbf{x} \rightarrow \mathbf{y})$ is a unitary matrix that rotates a fraction λ of the way from $\frac{\mathbf{x}}{\|\mathbf{x}\|}$ to $\frac{\mathbf{y}}{\|\mathbf{y}\|}$, as defined by (3-21) and (3-22). When \mathbf{G} is defined by (3-19), then (3-45) implements the adaptive signal-noise subspace separator of section 3.2, and $\mathbf{Q}_k \equiv \hat{\mathbf{E}}_k$ converges to a matrix Θ of the form given by (3-15). After convergence, the first n components of $\mathbf{y}_k = \mathbf{Q}_k \mathbf{r}_k$ represent the projection of \mathbf{r}_k onto the signal space, and the last $m - n$ components of \mathbf{y}_k represent the projection of \mathbf{r}_k onto the noise subspace. This means that only the first n components of \mathbf{y}_k contain signal energy, and the remaining components can be discarded without loss of information. This subspace separator is used in the next chapter as a means for simultaneously reducing complexity and speeding convergence of the subsequent detector stages.

In contrast, when \mathbf{G} is defined by (3-24), then (3-45) implements the total subspace decomposition of section 3.3; the rotator $\mathbf{Q}_k \equiv \hat{\mathbf{U}}_k^*$ converges to a valid \mathbf{U}^* , where $\mathbf{H} = \mathbf{U}\mathbf{S}\mathbf{V}^*$ is a channel SVD. The singular values \mathbf{S} can then be easily estimated from the energy of the components of \mathbf{y}_k . Using this algorithm, we have proposed fully blind implementations of the minimum-MSE and zero-forcing linear multiuser detectors, offering good performance, low complexity, and fast convergence. In a related application, we have also proposed a means to adaptively diagonalize a memoryless channel.

**APPENDIX 3-1: DERIVATION OF
THE MMSE LINEAR MULTIUSER DETECTOR**

Applying an $n \times m$ memoryless detector \mathbf{C} to the output of the channel defined by (3-1) produces the MSE sum $J(\mathbf{C}) = E[\|\mathbf{C}\mathbf{r}_k - \mathbf{x}_k\|^2]$, which can be expressed as

$$J(\mathbf{C}) = \text{tr}[(\mathbf{C}\mathbf{H} - \mathbf{I})(\mathbf{C}\mathbf{H} - \mathbf{I})^* + \sigma^2\mathbf{C}\mathbf{C}^*] \quad (3-46)$$

$$= \text{tr}[(\mathbf{C}\Phi\mathbf{C}^* - \mathbf{C}\mathbf{H} - \mathbf{H}^*\mathbf{C}^* + \mathbf{I}), \quad (3-47)$$

where $\Phi = \mathbf{H}\mathbf{H}^* + \sigma^2\mathbf{I}$. Since J is quadratic in \mathbf{C} , we can complete the square:

$$J(\mathbf{C}) = \text{tr}[(\mathbf{C} - \mathbf{H}^*\Phi^{-1})\Phi(\mathbf{C} - \mathbf{H}^*\Phi^{-1})^* + \mathbf{I} - \mathbf{H}^*\Phi^{-1}\mathbf{H}]. \quad (3-48)$$

For $\sigma^2 > 0$, Φ is positive definite, so Φ^{-1} exists. The \mathbf{C} that minimizes J is given by (3-32):

$$\mathbf{C}_{MMSE} = \mathbf{H}^*(\mathbf{H}\mathbf{H}^* + \sigma^2\mathbf{I})^{-1}, \quad (3-49)$$

and the corresponding minimum-MSE sum is $J_{min} = \text{tr}[\mathbf{I}_m - \mathbf{H}^*\Phi^{-1}\mathbf{H}]$. Applying the matrix inversion lemma,¹⁶ we obtain

$$\mathbf{C}_{MMSE} = \mathbf{H}^*[\sigma^{-2}\mathbf{I}_m - \sigma^{-2}\mathbf{H}(\sigma^{-2}\mathbf{H}^*\mathbf{H} + \mathbf{I}_n)^{-1}\mathbf{H}^*\sigma^{-2}] \quad (3-50)$$

$$= [\sigma^{-2}\mathbf{H}^* - \sigma^{-2}\mathbf{H}^*\mathbf{H}(\sigma^{-2}\mathbf{H}^*\mathbf{H} + \mathbf{I}_n)^{-1}\mathbf{H}^*\sigma^{-2}]. \quad (3-51)$$

Right-factoring $(\sigma^{-2}\mathbf{H}^*\mathbf{H} + \mathbf{I}_n)^{-1}\mathbf{H}^*\sigma^{-2}$ yields

$$\mathbf{C}_{MMSE} = [(\sigma^{-2}\mathbf{H}^*\mathbf{H} + \mathbf{I}_n) - \sigma^{-2}\mathbf{H}^*\mathbf{H}](\sigma^{-2}\mathbf{H}^*\mathbf{H} + \mathbf{I}_n)^{-1}\mathbf{H}^*\sigma^{-2}, \quad (3-52)$$

which simplifies to (3-33):

$$\mathbf{C}_{MMSE} = (\mathbf{H}^*\mathbf{H} + \sigma^2\mathbf{I}_n)^{-1}\mathbf{H}^*. \quad (3-53)$$

16. $(\mathbf{A}^{-1} + \mathbf{BCD})^{-1} = \mathbf{A} - \mathbf{AB}(\mathbf{DAB} + \mathbf{C}^{-1})^{-1}\mathbf{DA}$.

An alternative expression for the minimum-MSE sum is then given by

$$J_{min} = tr[\mathbf{I} - \mathbf{H}^* \Phi^{-1} \mathbf{H}] \quad (3-54)$$

$$= tr[\mathbf{I} - (\mathbf{H}^* \mathbf{H} + \sigma^2 \mathbf{I}_n)^{-1} \mathbf{H}^* \mathbf{H}] \quad (3-55)$$

$$= tr[(\mathbf{H}^* \mathbf{H} + \sigma^2 \mathbf{I}_n)^{-1} (\mathbf{H}^* \mathbf{H} + \sigma^2 \mathbf{I}_n)^{-1} \mathbf{H}^* \mathbf{H}] \quad (3-56)$$

$$= \sigma^2 tr(\mathbf{H}^* \mathbf{H} + \sigma^2 \mathbf{I}_n)^{-1}. \quad (3-57)$$

Observe that (3-33) or (3-53) also holds for $\sigma^2 = 0$; it reduces to the minimum-norm zero-forcing detector:

$$\mathbf{C}_{ZF} = (\mathbf{H}^* \mathbf{H})^{-1} \mathbf{H}^* \equiv \mathbf{H}^\dagger. \quad \square \quad (3-58)$$

APPENDIX 3-2: ON THE CONVERGENCE OF (3-25) AND (3-20)

In this appendix we present partial heuristic evidence that the recursion $\hat{\mathbf{U}}_{k+1} = \hat{\mathbf{U}}_k \mathcal{R}^\lambda(\mathbf{G}\mathbf{y}_k \rightarrow \mathbf{y}_k)$ of (3-25), where \mathbf{G} is defined by (3-24), converges to a valid left factor \mathbf{U} of a channel SVD $\mathbf{H} = \mathbf{U}\mathbf{S}\mathbf{V}^*$. Viewing the recursion $\hat{\Theta}_{k+1} = \mathcal{R}^\lambda(\mathbf{G}\mathbf{y}_k \rightarrow \mathbf{y}_k)^* \hat{\Theta}_k$ of (3-20), where \mathbf{G} is defined by (3-19), as a special case of (3-25), we also argue that it converges to a valid signal-noise subspace-separation matrix Θ of the form given by (3-15).

The partial rotation \mathcal{R}^λ defined by (3-20) and (3-21) is a function of the normalized inner product $p = \frac{\mathbf{y}^* \mathbf{z}}{\|\mathbf{y}\| \|\mathbf{z}\|}$, where $\mathbf{z} = \lambda \mathbf{G}\mathbf{y} + (1 - \lambda)\mathbf{y}$ (where we have suppressed the subscripts k). We consider first the case when $\lambda = 1$, so that $\mathbf{z} = \mathbf{G}\mathbf{y}$, and $p = \frac{\mathbf{y}^* \mathbf{G}\mathbf{y}}{\|\mathbf{y}\| \|\mathbf{G}\mathbf{y}\|}$. Observe that because \mathbf{G} is real, p is real; p can be interpreted simply as $p = \cos(\theta)$, the cosine of the angle θ between $\mathbf{G}\mathbf{y}$ and \mathbf{y} . Certainly if p is identically 1, then $\mathcal{R} = \mathbf{I}$, and the recursion of (3-25) or (3-20) stops. We argue that these recursions seek to minimize θ , or equivalently to maximize p . We further postulate that if the expected inner product (unnormalized) of \mathbf{y} and $\mathbf{G}\mathbf{y}$ defined by $\beta = E(\mathbf{y}^* \mathbf{G}\mathbf{y})$ is maximized, then these recursions stop on average. This argument in fact applies for any loop gain $0 < \lambda < 1$, because maximizing $E(\mathbf{y}^* \mathbf{z})$, where $\mathbf{z} = \lambda \mathbf{G}\mathbf{y} + (1 - \lambda)\mathbf{y}$, is equivalent to maximizing $E(\mathbf{y}^* \mathbf{G}\mathbf{y})$. This follows from the observation that $E(\mathbf{y}^* \mathbf{z}) = \lambda E(\mathbf{y}^* \mathbf{G}\mathbf{y}) + (1 - \lambda)E[\|\mathbf{y}\|^2]$, and $E[\|\mathbf{y}\|^2]$ is independent of $\hat{\mathbf{U}}$ or $\hat{\Theta}$.

We now show that for the recursion of (3-25), the expected inner product β is maximized only by valid left factors \mathbf{U} in a channel SVD, and that for the recursion of (3-20), β is maximized by valid subspace separators satisfying (3-15). We denote the front-end rota-

tion generically as \mathbf{Q} , whether intended as an estimate of Θ or \mathbf{U}^* . The output \mathbf{y} of the rotator can then be concisely expressed as

$$\mathbf{y} = \mathbf{QUSV}^* \mathbf{x} + \mathbf{Qn}. \quad (3-59)$$

Observe that the unitary matrices \mathbf{V}^* and \mathbf{Q} have no effect on the second-order statistics of \mathbf{x} and \mathbf{n} . Hence, we can further simplify (3-59) as

$$\mathbf{y} = \mathbf{TS}\tilde{\mathbf{x}} + \tilde{\mathbf{n}}, \quad (3-60)$$

where $\mathbf{T} = \mathbf{QU}$, $\tilde{\mathbf{x}} = \mathbf{V}^* \mathbf{x}$, and $\tilde{\mathbf{n}} = \mathbf{Qn}$. The expected inner product of $\mathbf{G}\mathbf{y}$ and \mathbf{y} can then be expressed as follows:

$$E(\mathbf{y}^* \mathbf{G}\mathbf{y}) = E \operatorname{tr}[(\mathbf{GTS}\tilde{\mathbf{x}} + \mathbf{G}\tilde{\mathbf{n}})(\mathbf{TS}\tilde{\mathbf{x}} + \tilde{\mathbf{n}})^*] \quad (3-61)$$

$$= \operatorname{tr}(\mathbf{GTDT}^* + \sigma^2 \mathbf{G}), \quad (3-62)$$

where $\mathbf{D} = \mathbf{SS}^*$, and where we have used the independence of the signal and noise. Because the noise term is irrelevant, it suffices to maximize $\operatorname{tr}(\mathbf{GTDT}^*)$. We can show that this term is upper bounded: $\operatorname{tr}(\mathbf{GTDT}^*) \leq \operatorname{tr}(\mathbf{GD})$. Equivalently, we can show that the following cost function is non-negative:

$$J(\mathbf{T}) = \operatorname{tr}(\mathbf{GD}) - \operatorname{tr}(\mathbf{GTDT}^*) \geq 0. \quad (3-63)$$

Let g_i and d_j denote the i -th and j -th diagonal element of \mathbf{G} and \mathbf{D} , respectively, and let $t_{i,j}$ denote the (i,j) -th element of \mathbf{T} so that

$$J = \sum_i g_i d_i - \sum_i \sum_j g_i d_j |t_{i,j}|^2. \quad (3-64)$$

With some algebraic manipulation, we can express J as

$$J = \sum_i g_i d_i (1 - |t_{i,i}|^2) - \sum_i \sum_{j \neq i} g_i d_j |t_{i,j}|^2. \quad (3-65)$$

Substituting $1 - |t_{i,i}|^2 = \sum_{j \neq i} |t_{i,j}|^2$ produces

$$J = \sum_i g_i d_i \left(\sum_{j \neq i} |t_{i,j}|^2 \right) - \sum_i \sum_{j \neq i} g_i d_j |t_{i,j}|^2 \quad (3-66)$$

$$= \sum_i \sum_{j \neq i} g_i d_i |t_{i,j}|^2 - \sum_i \sum_{j \neq i} g_i d_j |t_{i,j}|^2 \quad (3-67)$$

$$= \sum_i \sum_{j \neq i} g_i (d_i - d_j) |t_{i,j}|^2. \quad (3-68)$$

The inner summation can be expressed as two summations, one for $j > i$, the other for $j < i$, as follows:

$$J = \sum_i \sum_{j > i} g_i (d_i - d_j) |t_{i,j}|^2 + \sum_i \sum_{j < i} g_i (d_i - d_j) |t_{i,j}|^2. \quad (3-69)$$

Exploiting the fact that $|t_{i,j}|^2 = |t_{j,i}|^2$ and reversing the roles of i and j in the second double summation produces

$$J = \sum_i \sum_{j > i} g_i (d_i - d_j) |t_{i,j}|^2 - \sum_i \sum_{j > i} g_j (d_i - d_j) |t_{i,j}|^2 \quad (3-70)$$

$$J = \sum_i \sum_{j > i} (g_i - g_j) (d_i - d_j) |t_{i,j}|^2. \quad (3-71)$$

Observe that $|t_{i,j}|^2 \geq 0$, $d_i - d_j \geq 0$, and $g_i - g_j \geq 0 \forall j > i$ for \mathbf{G} defined according to either (3-19) or (3-24). It follows that $J(\mathbf{T}) \geq 0$ with equality if and only if $|t_{i,j}|^2 = 0 \forall (i,j)$ corresponding to distinct elements $d_i \neq d_j$ and distinct elements $g_i \neq g_j$. If $d_i = d_j$ or $g_i = g_j$, then the corresponding term $|t_{i,j}|^2$ is unconstrained; it cannot contribute to the cost. Therefore, if \mathbf{G} satisfies (3-24), then \mathbf{T} has the form of an SVD ambiguity, given by Lemma 3-1, and $\hat{\mathbf{U}}$ is a valid \mathbf{U} . Similarly, if \mathbf{G} satisfies (3-19), then \mathbf{T} has the form of a subspace-separator ambiguity, given by Lemma 3-3, and $\hat{\Theta}$ is a valid Θ . \square

CHAPTER 4

THE WHITEN-ROTATE DETECTOR

A USEFUL STRATEGY for blind multiuser detection is to decompose the problem into two smaller tasks: *whitening* and *rotation*. Whitening, in the context of (3-1), is simply transforming the receiver observation vector into one whose covariance is the identity. Thus, the whitening step exploits only second-order statistics, and it can easily be implemented blindly. The rotation step can be implemented by a unitary matrix chosen to restore some higher-order statistical property of the channel input. It is well known that the whiten-rotate structure can perfectly invert a noiseless channel [129]. Batch techniques based on the idea were proposed in [108,130,131], and an adaptive algorithm for noiseless channels was presented in [116].

Following the whiten-rotate strategy, and again restricting attention to memoryless channels (3-1), we define a *canonical whiten-rotate detector*, which minimizes the MSE of all users among detectors in its class. In section 4.1, we describe the basic structure of the detector and analyze its performance and properties. We show that the whiten-rotate detector is information lossless and optimally near-far resistant with performance closely approximating that of the MMSE detector. In section 4.2, we define an alternative structure for the detector based on subspace projection. In section 4.3, we detail adaptive blind

implementations of these structures using spatial linear prediction and the subspace-separation algorithm of chapter 3. The implementations we describe offer a good compromise between complexity and convergence speed. We include simulation results for a linear-antenna-array system and a synchronous CDMA system. We generalize the whiten-rotate detector presented here to channels with memory in chapter 6. (Much of chapter 4 appears also in [132,133].)

4.1 Whiten-Rotate Detection

In the context of the system of (3-1), an $n \times m$ matrix \mathbf{C} is said to be a *whitener* if the autocorrelation of $\mathbf{z}_k = \mathbf{C}\mathbf{r}_k$ is the identity matrix, $\mathbf{C}\Phi_r\mathbf{C}^* = \mathbf{I}$, where $\Phi_r = \mathbf{H}\mathbf{H}^* + \sigma^2\mathbf{I}$. We define the whiten-rotate (WR) detector as the whitener with minimal MSE.

Definition 4-1. The **canonical whiten-rotate detector** \mathbf{C}_{WR} for the channel of (3-1) is the $n \times m$ whitener that minimizes the MSE sum $E[\|\mathbf{C}\mathbf{r}_k - \mathbf{x}_k\|^2]$.

Any short whitener \mathbf{C} of dimension $n \times m$ can be expressed as the first n rows of a larger whitener \mathbf{B} of dimension $m \times m$; in particular, we can write $\mathbf{C} = \mathbf{J}\mathbf{B}$, where $\mathbf{J} = [\mathbf{I}_n \ \mathbf{0}]$ is an $n \times m$ truncation matrix, and where $\mathbf{B}\Phi_r\mathbf{B}^* = \mathbf{I}_m$. Recall that, for any given $m \times m$ whitening matrix \mathbf{W} , every other whitening matrix \mathbf{B} can be expressed in the form $\mathbf{B} = \mathbf{Q}\mathbf{W}$ for some $m \times m$ unitary matrix \mathbf{Q} [121]. Thus, given any particular $m \times m$ whitening matrix \mathbf{W} , we can express every $n \times m$ whitener as $\mathbf{C} = \mathbf{J}\mathbf{Q}\mathbf{W}$ for some unitary matrix \mathbf{Q} . This suggests a three-stage implementation of the whiten-rotate detector, as depicted in Fig. 4-1.

Observe from (3-32) that the MMSE detector can be expressed as $\mathbf{C}_{MMSE} = \mathbf{H}^* \mathbf{W}^* \mathbf{W} = (\mathbf{W}\mathbf{H})^* \mathbf{W}$, where we use the identity $\mathbf{W}^* \mathbf{W} = \Phi_r^{-1}$ for any whitener \mathbf{W} . Thus, the MMSE detector could be implemented by following a whitener \mathbf{W} by the $n \times m$ filter $(\mathbf{W}\mathbf{H})^*$. However, by its definition the whiten-rotate detector must follow a whitener by a matrix of the form $\mathbf{J}\mathbf{Q}$. It can be shown that, rather than $(\mathbf{W}\mathbf{H})^*$, the best such filter (minimizing total MSE) is the unique so-called polar factor of $(\mathbf{W}\mathbf{H})^*$ [134], which is simply $(\mathbf{W}\mathbf{H})^*$ with its singular values replaced by unity:

$$\mathbf{J}\mathbf{Q} = \left(\bar{\mathbf{U}}\mathbf{J}^T\bar{\mathbf{V}}^* \right)^*, \quad (4-1)$$

where $\bar{\mathbf{U}}$ and $\bar{\mathbf{V}}$ are $(m \times m$ and $n \times n)$ unitary factors in an SVD of $\mathbf{W}\mathbf{H} = \bar{\mathbf{U}}\bar{\mathbf{S}}\bar{\mathbf{V}}^*$. Note that \mathbf{Q} satisfies (4-1) if and only if it is of the form:

$$\mathbf{Q} = \begin{bmatrix} \bar{\mathbf{V}} & \mathbf{0} \\ \mathbf{0} & \mathbf{V}_N \end{bmatrix} \bar{\mathbf{U}}^*, \quad (4-2)$$

where \mathbf{V}_N is an arbitrary unitary matrix of dimension $m - n$. The rotator \mathbf{Q} of (4-2) performs two tasks. First, it plays the role of the subspace separator of chapter 3 by removing all signal energy from the last $m - n$ components of its output. Second, it also provides the

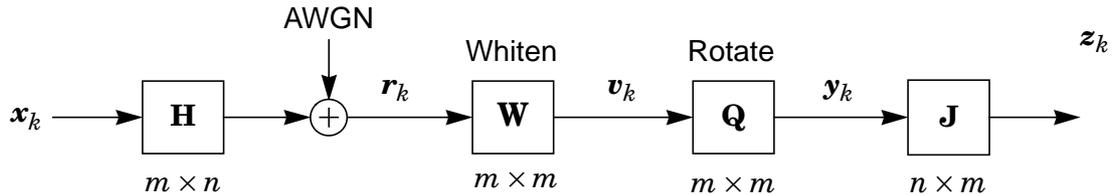


Figure 4-1. The structure of the whiten-rotate detector as applied to a memoryless multiuser channel.

best unitary separation of users within the signal space. In other words, \mathbf{Q} separates signal from noise, and signal from signal.

The following lemma summarizes the form of the whiten-rotate detector.

Lemma 4-1. The whiten-rotate detector of Definition 4-1 is unique, and it can be expressed in three equivalent ways:

$$\mathbf{C}_{WR} = \mathbf{J}\mathbf{Q}\mathbf{W} \quad (4-3)$$

$$= \mathbf{V}\mathbf{J}(\mathbf{S}\mathbf{S}^* + \sigma^2\mathbf{I})^{-1/2}\mathbf{U}^* \quad (4-4)$$

$$= \mathbf{V}(\tilde{\mathbf{S}}^2 + \sigma^2\mathbf{I})^{-1/2}\mathbf{J}\mathbf{U}^*. \quad (4-5)$$

In (4-3), \mathbf{W} is any $m \times m$ whitener (satisfying $\mathbf{W}\Phi_r\mathbf{W}^* = \mathbf{I}$), \mathbf{Q} satisfies (4-2), and $\mathbf{J} = [\mathbf{I} \ \mathbf{0}]$. In (4-4) and (4-5), $\mathbf{H} = \mathbf{U}\mathbf{S}\mathbf{V}^*$ is an SVD, and $\tilde{\mathbf{S}} = \mathbf{J}\mathbf{S}$.

Proof: See Appendix 4-1.

Using this lemma, we derive several properties of the whiten-rotate detector.

Property 4-1. The whiten-rotate detector is information lossless.

This follows from (4-5) by observing that $\mathbf{J}\mathbf{U}^*$ discards no signal energy, and that both \mathbf{V} and $(\tilde{\mathbf{S}}^2 + \sigma^2\mathbf{I})^{-1/2}$ are invertible.

Property 4-2. The whiten-rotate detector approaches the zero-forcing (or decorrelating) detector in the limit as the noise energy goes to zero:

$$\lim_{\sigma \rightarrow 0^+} \mathbf{C}_{WR} = \mathbf{V}\tilde{\mathbf{S}}^{-1}\mathbf{J}\mathbf{U}^* = \mathbf{V}\mathbf{S}^\dagger\mathbf{U}^* = \mathbf{H}^\dagger. \quad (4-6)$$

Property 4-3. The whiten-rotate detector is optimally near-far resistant [53].

Optimal near-far resistance is inherited from the zero-forcing detector.

Lemma 4-2. The MSE for the i -th user of the whiten-rotate detector and MMSE detector, respectively, can be expressed as

$$\text{MSE}_i^{WR} = 2\mathbf{v}_i^* [\mathbf{I} - (\tilde{\mathbf{S}}^2 + \sigma^2\mathbf{I})^{-1/2}\tilde{\mathbf{S}}] \mathbf{v}_i, \quad (4-7)$$

$$\text{MSE}_i^{MMSE} = \sigma^2 \mathbf{v}_i^* (\tilde{\mathbf{S}}^2 + \sigma^2\mathbf{I})^{-1} \mathbf{v}_i, \quad (4-8)$$

where \mathbf{v}_i is the i -th column of \mathbf{V}^* .

Proof: See Appendix 4-2.

Using this lemma, we arrive at the following property of the whiten-rotate detector.

Property 4-4. The MSE of the whiten-rotate detector approaches that of the MMSE detector in the limit as the noise energy goes to zero:

$$\lim_{\sigma \rightarrow 0^+} \frac{\text{MSE}_i^{WR}}{\text{MSE}_i^{MMSE}} = 1. \quad (4-9)$$

Proof: The proof follows from (4-7) and (4-8) and a straightforward application of l'Hôpital's rule.

In the following experiment, we use (4-7) and (4-8) to compare the theoretical performance of the whiten-rotate detector to that of the MMSE detector.

Experiment 4-1. Given a receiver with $m = 10$ sensors, we consider two cases: $n = 2$ users and 10 users. In Fig. 4-2, we plot MSE_1 versus $\text{SNR}_1 = \sum_{j=1}^m |h_{j,1}|^2 / \sigma^2$, averaged over 1000 channels of dimension $10 \times n$. The coefficients of each channel are selected independently from a zero-mean, unit-variance, complex Gaussian distribution, and then the channel columns are scaled so that all odd-numbered users have energy 10 dB below that of even-numbered users. The curves for the two-user case show that even for a low SNR_1 of -10 dB, user 1 incurs only a modest 2-dB MSE penalty for using the WR detector instead of the MMSE detector. Moreover, for $\text{SNR}_1 > 10$ dB, the performance difference between the WR and MMSE detectors is negligible. For the ten-user case, the performance of both detectors degrades; however, the performance difference between the two detectors widens only slightly.

The previous experiment suggests that the MSE performance of the whiten-rotate detector is near to that of the MMSE detector, especially for very tall channels ($m \gg n$) and high signal-to-noise ratios.

4.2 A Project-First Architecture

We can use the subspace separator of chapter 3 to define an alternative structure for the whiten-rotate detector. It is often desirable to separate subspaces before whitening, *i.e.* at the receiver front end, by immediately projecting the m -dimensional receiver observation onto the n -dimensional signal space using a unitary matrix Θ . The advantages of this *project-first* architecture are two: first, it allows all subsequent signal processing to operate in n dimensions rather than m , which reduces the receiver complexity; and second, it

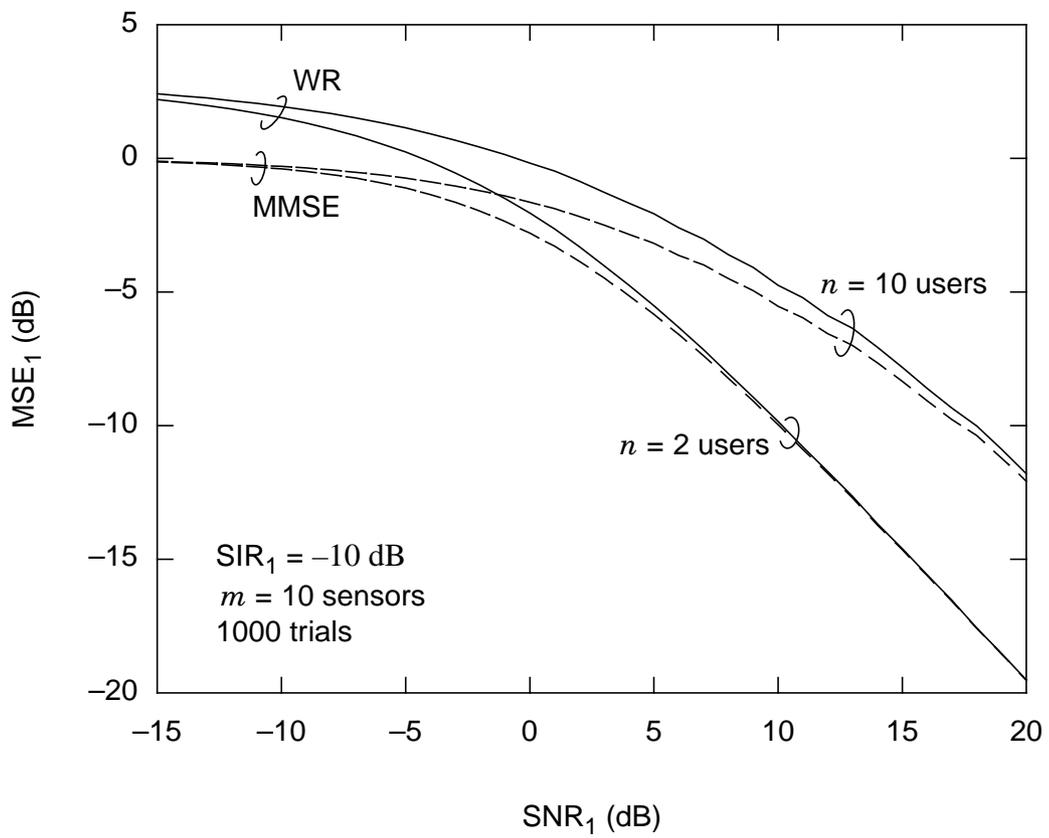


Figure 4-2. A comparison of the whiten-rotate detector with the minimum-MSE detector: SNR versus MSE.

reduces the number of receiver parameters, which often leads to faster receiver convergence.

Recall that for the channel of (3-1), a unitary subspace-separating matrix must be of the form $\Theta = \begin{bmatrix} \mathbf{U}_S & \mathbf{0} \\ \mathbf{0} & \mathbf{U}_N \end{bmatrix} \mathbf{U}^*$, where \mathbf{U}_S and \mathbf{U}_N are arbitrary unitary matrices of dimension n and $m - n$ respectively, and where \mathbf{U} is the left factor of a channel SVD $\mathbf{H} = \mathbf{U}\mathbf{S}\mathbf{V}^*$. Following a subspace separator by a truncation matrix $\mathbf{J} = [\mathbf{I} \ \mathbf{0}]$ produces the square channel of Fig. 3-1(b) with reduced observation:

$$\tilde{\mathbf{r}}_k = \tilde{\mathbf{H}} \mathbf{x}_k + \tilde{\mathbf{n}}_k, \quad (4-10)$$

where $\tilde{\mathbf{r}}_k = \mathbf{J}\Theta \mathbf{r}_k$, $\tilde{\mathbf{H}} = \mathbf{U}_S \tilde{\mathbf{S}} \mathbf{V}^*$, and $E[\tilde{\mathbf{n}}_k \tilde{\mathbf{n}}_k^*] = \sigma^2 \mathbf{I}$. The n components of the new observation vector $\tilde{\mathbf{r}}_k$ form a set of sufficient statistics for estimating \mathbf{x}_k . Thus, the whiten-rotate detector of section 4.1 can be applied to this new channel without compromising performance.

Theorem 4-1. The cascade consisting of a signal-subspace projector $\mathbf{J}\Theta$ followed by a whiten-rotate detector $\tilde{\mathbf{C}}_{WR}$ designed for the reduced channel $\tilde{\mathbf{H}} = \mathbf{J}\mathbf{Q}\mathbf{H}$ precisely implements the WR detector \mathbf{C}_{WR} designed for the original channel \mathbf{H} .

Proof: The proof applies to the MMSE (3-34) and ZF (3-38) detectors as well as to the WR detector (4-5), since all can be expressed in terms of a channel SVD $\mathbf{H} = \mathbf{U}\mathbf{S}\mathbf{V}^*$ as $\mathbf{C} = \mathbf{V}\mathbf{D}\mathbf{J}\mathbf{U}^*$, where $\mathbf{D} = (\tilde{\mathbf{S}}^2 + \sigma^2 \mathbf{I})^{-1/2}$ for the WR detector, $\mathbf{D} = \tilde{\mathbf{S}}^* (\tilde{\mathbf{S}}^2 + \sigma^2 \mathbf{I})^{-1}$ for the MMSE detector, and $\mathbf{D} = \mathbf{S}^\dagger$ for the ZF detector. We need to show that the cascade of $\mathbf{J}\Theta$ and an $n \times n$ detector $\tilde{\mathbf{C}}$ designed for the reduced channel $\tilde{\mathbf{H}}$ is equivalent to the same type of detector designed for the orig-

inal channel \mathbf{H} . In other words, we need to show that $\tilde{\mathbf{C}}\mathbf{J}\Theta = \mathbf{C}$. But based on the SVD $\tilde{\mathbf{H}} = \mathbf{U}_S\tilde{\mathbf{S}}\mathbf{V}^*$ of the reduced channel, we have $\tilde{\mathbf{C}} = \mathbf{V}\mathbf{D}\mathbf{U}_S^*$, so that $\tilde{\mathbf{C}}\mathbf{J}\Theta = \mathbf{V}\mathbf{D}\mathbf{U}_S^*\mathbf{J}\begin{bmatrix} \mathbf{U}_S & \mathbf{0} \\ \mathbf{0} & \mathbf{U}_N \end{bmatrix}\mathbf{U}^* = \mathbf{V}\mathbf{D}\mathbf{J}\mathbf{U}^* = \mathbf{C}$. \square

4.3 Blind Adaptive Implementations

In this section we describe adaptive algorithms for blind implementation of the whiten-rotator detector. The adaptive whitener is based on spatial linear prediction, and the adaptive rotator is based on a simple modification of the MPLL algorithm.

4.3.1 An Adaptive Whitener

A simple way to whiten is to use adaptive linear prediction. Suppose we wish to predict the i -th component $r_k^{(i)}$ of \mathbf{r}_k using a linear combination of the “preceding” components $r_k^{(1)} \dots r_k^{(i-1)}$, yielding an estimate:

$$\hat{\mathbf{r}}_k = \mathbf{P}\mathbf{r}_k, \quad (4-11)$$

where \mathbf{P} is a strictly lower-triangular matrix of prediction coefficients. The prediction error is $\mathbf{e}_k = (\mathbf{I} - \mathbf{P})\mathbf{r}_k$. The best predictor in the least-mean-square sense, *i.e.* minimizing $E[\|\mathbf{e}_k\|^2]$, is closely linked to the Cholesky factorization of the covariance matrix $\Phi_{\mathbf{r}}$ of \mathbf{r}_k .

Lemma 4-3. Generalized Cholesky Factorization. A Hermitian matrix Φ of dimension $m \times m$ and rank $n \leq m$ can be factored in either of two ways:

$$\Phi = \mathbf{G}\mathbf{G}^*, \quad (4-12)$$

$$= \mathbf{M}\mathbf{D}^2\mathbf{M}^*, \quad (4-13)$$

where $\mathbf{G} = \mathbf{M}\mathbf{D}$ is a unique $m \times m$ lower-triangular matrix with real, non-negative diagonal elements, where $\mathbf{D} = \text{diag}(\mathbf{G})$, and where \mathbf{M} is lower triangular with ones on the main diagonal (monic). The matrix \mathbf{M} is unique if and only if the first $m - 1$ rows of Φ are linearly independent.

Proof: See Appendix 4-3.

Theorem 4-2. Linear Prediction. Let \mathbf{r} be a random $m \times 1$ vector with covariance matrix $\Phi = E[\mathbf{r}\mathbf{r}^*]$, and let $\mathbf{e} = (\mathbf{I} - \mathbf{P})\mathbf{r}$ denote the error of a linear predictor, where \mathbf{P} is strictly lower triangular. The \mathbf{P} that minimizes $E[\|\mathbf{e}\|^2]$ is

$$\mathbf{P} = \mathbf{I} - \mathbf{M}^{-1}, \quad (4-14)$$

where \mathbf{M} is any valid monic factor in the generalized Cholesky factorization (4-13) of Φ . The predictor is unique if and only if \mathbf{M} is unique, or equivalently, if and only if the first $m - 1$ rows of Φ are linearly independent.

Proof: See Appendix 4-4.

As long as the noise variance is nonzero, the covariance $\Phi_{\mathbf{r}} = \mathbf{H}\mathbf{H}^* + \sigma^2\mathbf{I}$ is full rank and the Cholesky factorization (4-13) and corresponding predictor (4-14) are both unique. (See Appendix 4-5 for a discussion of the noiseless case.) In practice, the estimate of \mathbf{P} , denoted $\hat{\mathbf{P}}$, can be adapted according to the least-mean-square algorithm:

$$\hat{\mathbf{P}}_{k+1} = (\hat{\mathbf{P}}_k + \mu_p \mathbf{e}_k \mathbf{r}_k^*) \otimes \mathbf{L}, \quad (4-15)$$

where $\mathbf{e}_k = \mathbf{r}_k - \hat{\mathbf{P}}_k \mathbf{r}_k$ is the prediction error, where μ_p is a step size, where ‘ \otimes ’ denotes a component-wise (Schur or Hadamard) product, and where \mathbf{L} is a mask, with ones below

the main diagonal and zeros elsewhere, that constrains $\hat{\mathbf{P}}$ to be strictly lower triangular. We remark that, because (4-15) is derived from a quadratic cost function, convergence to (4-14) is guaranteed for a sufficiently small step size μ_p .

After $\hat{\mathbf{P}}$ converges to (4-14), the covariance of the resulting error $\mathbf{e}_k = (\mathbf{I} - \mathbf{P})\mathbf{r}_k$ is diagonal: $\Phi_{\mathbf{e}} = \mathbf{D}^2$. Therefore, a diagonal gain matrix $\mathbf{A} = \mathbf{D}^{-1}$ converts the prediction error \mathbf{e}_k into the white signal $\mathbf{v}_k = \mathbf{A}\mathbf{e}_k$ with covariance matrix $\Phi_{\mathbf{v}} = \mathbf{I}$. This gain matrix can be implemented adaptively by a bank of independent scalar automatic gain-control (AGC) loops, designed to force the energy at each output to unity. We propose a simple first-order loop for adapting each component of an estimate $\hat{\mathbf{A}} = \text{diag}(\hat{A}^{(1)}, \dots, \hat{A}^{(m)})$:

$$\hat{A}_{k+1}^{(i)} = | \hat{A}_k^{(i)} - \mu_a (|v_k^{(i)}|^2 - 1) |, \quad (4-16)$$

where $v_k^{(i)}$ is the i -th component of $\mathbf{v}_k = \hat{\mathbf{A}}_k \mathbf{e}_k$. In summary, the proposed adaptive whitener is $\hat{\mathbf{W}}_k = \hat{\mathbf{A}}_k (\mathbf{I} - \hat{\mathbf{P}}_k)$, where $\hat{\mathbf{P}}$ and $\hat{\mathbf{A}}$ are adapted according to (4-15) and (4-16), respectively.

4.3.2 An Adaptive Rotator

Recall the structure of the whiten-rotate detector of Fig. 4-1: $\mathbf{C} = \mathbf{J}\mathbf{Q}\mathbf{W}$. Let $\mathbf{v}_k = \hat{\mathbf{W}}_k \mathbf{r}_k$ denote the whitener output with $\hat{\mathbf{W}}$ adapted according to the previous section. It remains to specify an adaptive algorithm to estimate \mathbf{Q} of (4-2). Let the rotator $\hat{\mathbf{Q}}$ be an estimate of \mathbf{Q} , and let $\mathbf{y}_k = \hat{\mathbf{Q}}_k \mathbf{v}_k$ denote the corresponding rotator output. We can modify the MPLL to adaptively implement \mathbf{Q} . We need only modify the decision device. Let $\hat{\mathbf{x}}_k$ denote an *augmented* decision vector:

$$\hat{\mathbf{x}}_k = \begin{bmatrix} \text{dec}(\cdot) & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{bmatrix} \mathbf{y}_k. \quad (4-17)$$

For $i \leq n$, $\hat{x}_k^{(i)} = \text{dec}_i(y_k^{(i)})$ is the point in the constellation of user i closest to $y_k^{(i)}$, but for $i > n$, $\hat{x}_k^{(i)}$ is set to zero. In effect, the decision vector $\hat{\mathbf{x}}_k$ is defined as is there were m users with the last $m - n$ of these transmitting all zeros. The rotator $\hat{\mathbf{Q}}_k$ can then be adapted according to the recursion:

$$\begin{aligned}\hat{\mathbf{Q}}_{k+1} &= \mathcal{R}^{\lambda}(\hat{\mathbf{x}}_k \rightarrow \mathbf{y}_k)^* \hat{\mathbf{Q}}_k \\ &= \hat{\mathbf{Q}}_k + \left[\mathcal{R}^{\lambda}(\hat{\mathbf{x}}_k \rightarrow \mathbf{y}_k)^* - \mathbf{I} \right] \hat{\mathbf{Q}}_k.\end{aligned}\quad (4-18)$$

The modified MPLL is illustrated in Fig. 4-3.

As we discussed in section 3.4 of the previous chapter, there are unavoidable ambiguities that cannot be resolved by any fully blind algorithm. Empirical evidence supports the conjecture that the rotator converges to

$$\hat{\mathbf{Q}} = \begin{bmatrix} \mathbf{K} & \mathbf{0} \\ \mathbf{0} & \mathbf{I}_{m-n} \end{bmatrix} \mathbf{Q}, \quad (4-19)$$

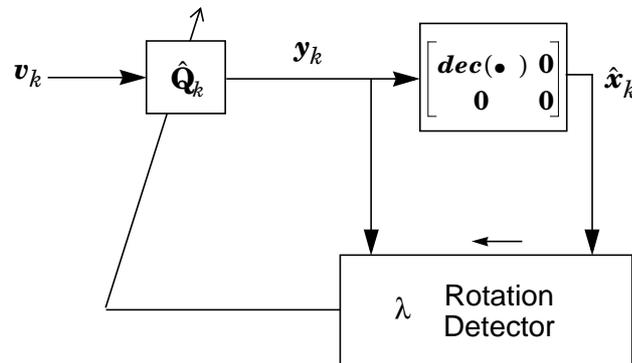


Figure 4-3. An adaptive algorithm for implementing the MMSE rotator of (4-2).

where \mathbf{Q} is given by (4-2), and where, as in (3-36), $\mathbf{K} = \mathbf{K}_p \mathbf{K}_R$ is the $n \times n$ product of a permutation and rotation ambiguity.

4.4 Experimental Results

We conclude with several computer experiments. We use the project-first architecture as depicted in Fig. 4-4. The subspace-separation algorithm, already detailed in section 3.2 of the previous chapter, is used to produce $\tilde{\mathbf{r}}_k$. We then apply the adaptive whitener and rotator of sections 4.3.1 and 4.3.2 to the reduced observation $\tilde{\mathbf{r}}_k$. The first experiment demonstrates convergence of the entire project-first WR detector; in particular, it shows the contributions to MSE from each of the adaptive stages. The second experiment applies the WR detector to an array-processing problem similar to that in Experiment 3-3. The last experiment considers a synchronous CDMA system.

Experiment 4-2. Random Gaussian Channels. We now demonstrate convergence of the entire project-first algorithm of Fig. 4-4. We consider two users, each transmitting 16-QAM with 20-dB SNR. Fig. 4-5 shows $\text{MSE}_1 = E[|z_k^{(1)} - x_k^{(1)}|^2]$ as a function of time, averaged over 1000 realizations of input, noise, and a 10×2 complex Gaussian channel. There are five curves in all. The bottom curve, labeled MMSE, is MSE_1 for the ideal MMSE equalizer. The initial subspace separator is adaptive for the curve above that, but all remaining functions are idealized. Similarly, the other curves are labeled to indicate which components of the algorithm are adaptive. Everything is adaptive for the top curve, with the effect of the ambiguous complex permutation matrix \mathbf{K} removed for each trial. These curves illustrate the MSE contributed by each stage of the project-first algorithm. The subspace separator

converges very quickly and has little impact on MSE. The linear predictor and AGC bank also converge quickly, and the receiver eventually closely approximates the MMSE solution. (The parameters used in the receiver updates are as follows: $\lambda_1 = 0.025$ in (3-23), $\mu_P = 0.03(2^{-k/250})$ in (4-15), $\mu_A = 0.02(2^{-k/300})$ in (4-16), and $\lambda_2 = 0.8(1 + k/200)$ in (4-18).)

Experiment 4-3. A Linear Antenna Array. Consider a 20-sensor linear antenna array with half-wavelength spacing, and suppose two signals are incident at angles $\theta_1 = 0^\circ$ and $\theta_2 = 20^\circ$ (measured from broadside). The channels are generated as in Experiment 3-3 with $\text{SNR}_1 = 15$ dB, and $\text{SNR}_2 = 35$ dB. Fig. 4-6(a) shows a plot of MSE_1 versus time, averaged over 100 input and noise realizations, with the effect of the complex permutation removed. Fig. 4-6(b) shows constellations from time 4000 to time 5000 from the last trial. Once again, we see quick convergence to near-MMSE performance. (The parameters used in the receiver updates are as follows: $\lambda_1 = 0.8/(1 + k/200)$ in (3-23), $\mu_P = 0.1/(1 + k/250)$ in (4-15), $\mu_A = 0.2(2^{-k/1000})$ in (4-16), and $\lambda_2 = 0.8(1 + k/150)$ in (4-18).)

Experiment 4-4. Synchronous CDMA. Consider now a synchronous direct-sequence-CDMA application with three interfering users, each transmitting 16-QAM. Let $c_i \in \{\pm 1\}^{32}$ denote the binary signature sequence with length 32 of the i -th user. If the transmitter pulse-shape filters are Nyquist, and the receiver uses a chip-rate-sampled matched filter followed by a serial-to-parallel converter, the resulting discrete-time channel is given by (3-1) with $\mathbf{H}_{32 \times 3} = \frac{1}{\sqrt{32}} [c_1 \ c_2 \ c_3] \mathbf{B}$, where $\mathbf{B}_{3 \times 3} = \text{diag}(B_1, B_2, B_3)$ is a matrix of signal amplitudes. The signature sequences have normalized correlations $\rho_{i,j} = \frac{1}{32} c_i^T c_j$ of $\rho_{1,2} = -1/8$, $\rho_{1,3} = -1/4$,

1000 trials
 Gaussian channels
 $n = 2$, 16-QAM users
 $m = 10$ sensors
 $\lambda_1 = 0.025$
 $\mu_P = 0.03(2^{-k/250})$
 $\mu_A = 0.02(2^{-k/300})$
 $\lambda_2 = 0.8(1 + k/200)$

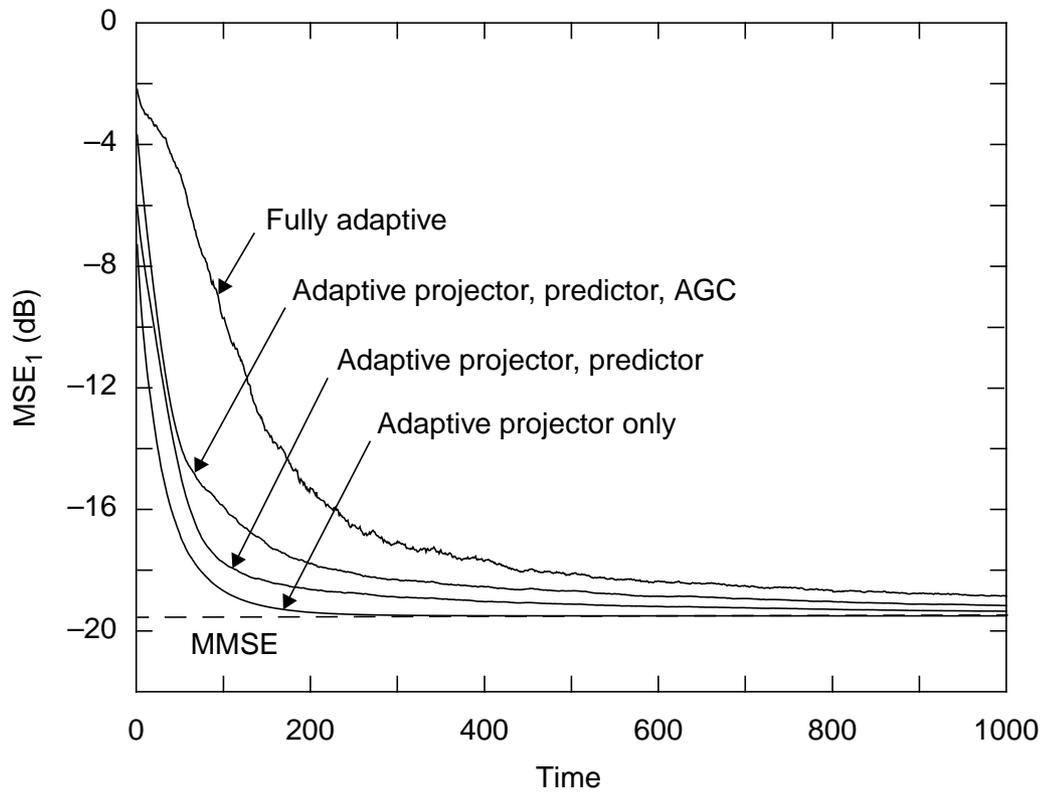


Figure 4-5. Convergence of the project-first adaptive algorithm of Fig. 4-4, showing contributions to MSE from each stage.

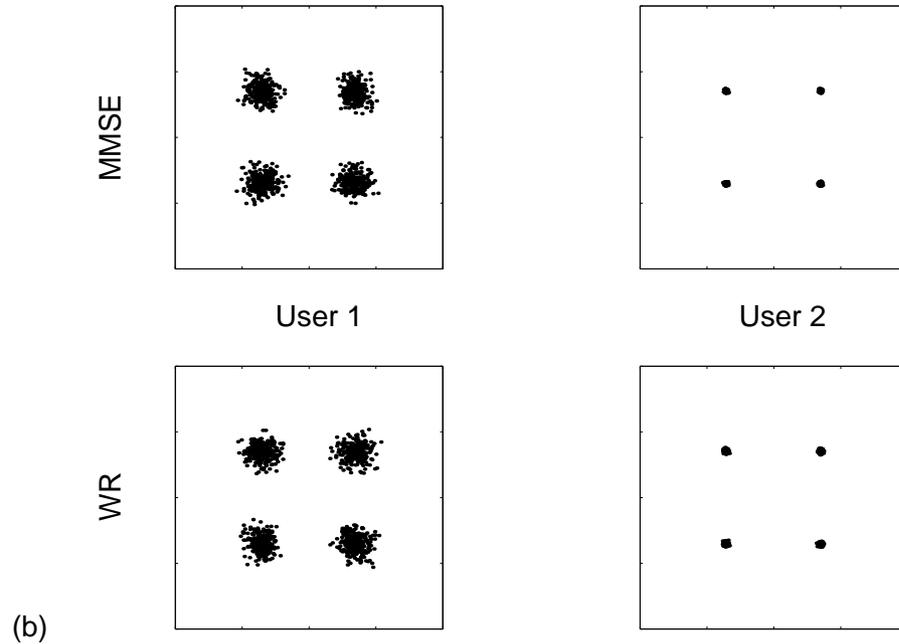
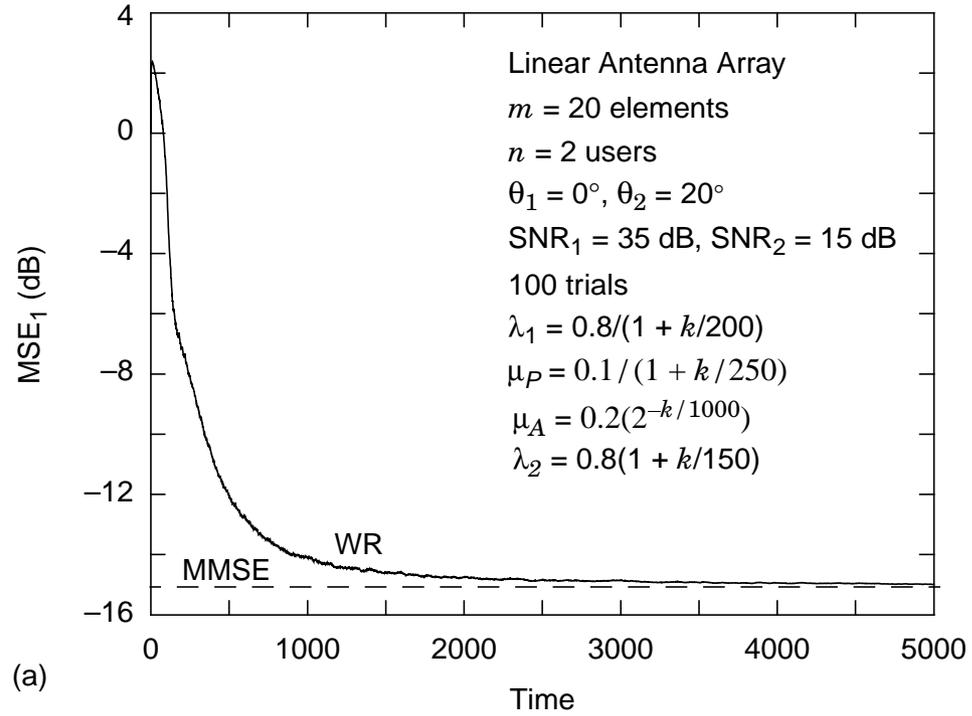


Figure 4-6. The adaptive project-first WR detector applied to a linear antenna array: (a) the MSE learning curve; (b) constellations from the last trial, baud 4000 to 5000.

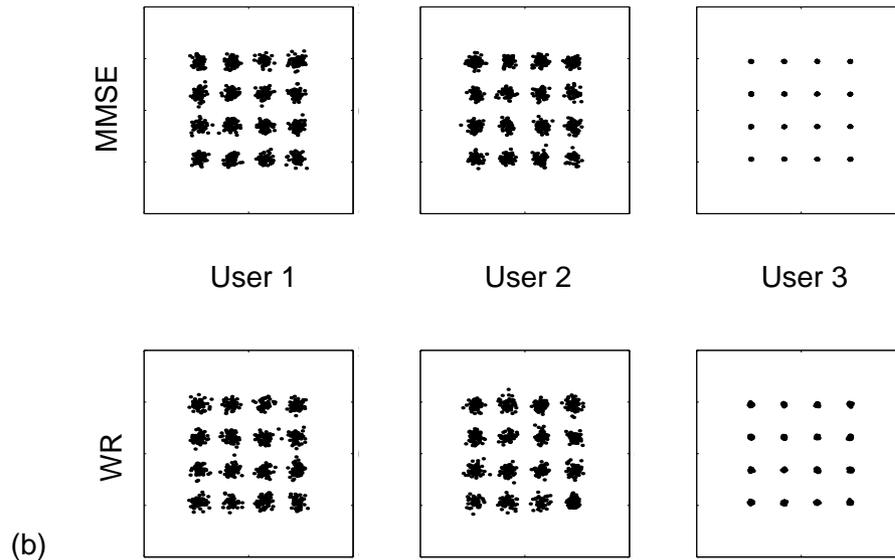
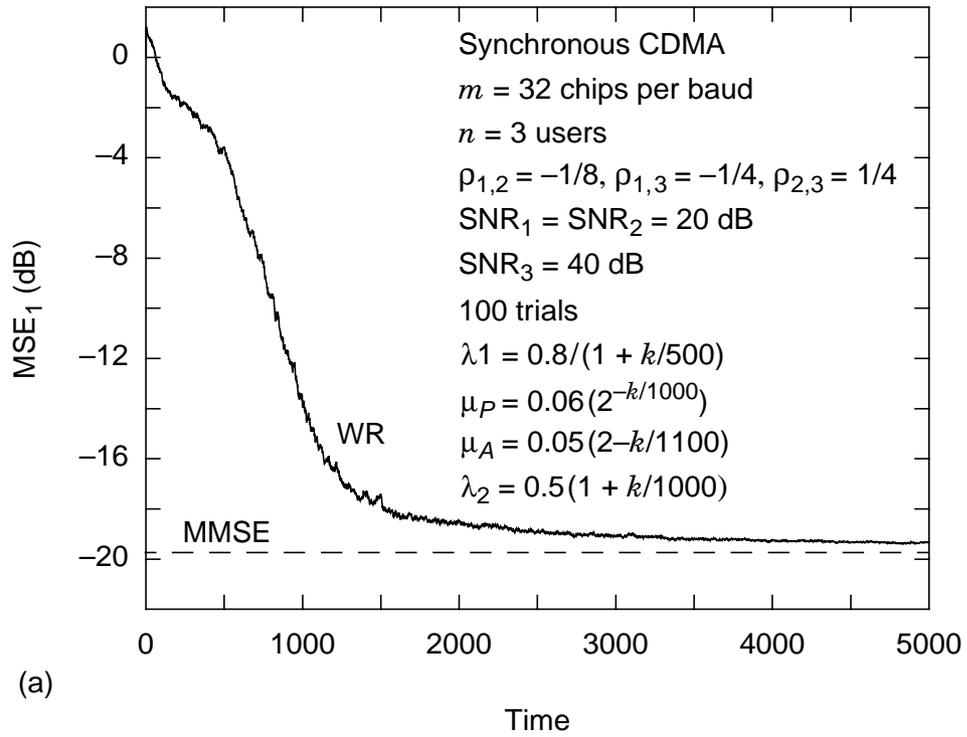


Figure 4-7. The adaptive project-first WR detector applied to a synchronous CDMA system: (a) an MSE learning curve; (b) constellations from the last trial, baud 4000 to 5000.

and $\rho_{2,3} = 1/4$. Fig. 4-7(a) shows an MSE learning curve, averaged over 100 input and noise realizations, with $\text{SNR}_1 = \text{SNR}_2 = 20$ dB and $\text{SNR}_3 = 40$ dB. We see that the algorithm converges quickly, eventually closely approximating the MMSE solution. Fig. 4-7(b) shows constellations of the last 1000 symbols from the last trial. (The receiver parameters are $\lambda_1 = 0.8/(1 + k/500)$ in (3-23), $\mu_P = 0.06(2^{-k/1000})$ in (4-15), $\mu_A = 0.05(2^{-k/1100})$ in (4-16), and $\lambda_2 = 0.5(1 + k/1000)$ in (4-18).)

4.5 Chapter Summary

We have defined a multiuser detector based on a canonical whiten-rotate structure that is information preserving, optimally near-far resistant, and has near-MMSE performance. The WR detector is in fact the whitener with minimal MSE. We have presented two equivalent architectures for the detector, namely the whiten-first and project-first architectures, illustrated in Figs. 4-1 and 4-4, respectively. We have detailed blind implementations for each. For the whiten-first approach, the receiver observation \mathbf{r}_k is whitened via the cascade of a prediction-error filter $\mathbf{I} - \hat{\mathbf{P}}$ and a diagonal gain matrix $\hat{\mathbf{A}}$, where $\hat{\mathbf{P}}$ and $\hat{\mathbf{A}}$ are updated according to (4-15) and (4-16), respectively. The whitener output is then rotated by $\hat{\mathbf{Q}}$, updated according to (4-18). For the project-first approach, the observation \mathbf{r}_k is projected onto the signal space to produce the reduced observation $\tilde{\mathbf{r}}_k = \mathbf{J}\hat{\mathbf{Q}}_k\mathbf{r}_k$, where $\hat{\mathbf{Q}}$ is updated via (3-23). The whitening and rotation algorithms are then applied to $\tilde{\mathbf{r}}_k$. In either architecture, the second-order statistics of the transmitted signals are exploited first. The higher-order statistics are exploited only at the last step (4-18) by finding a unitary matrix that best restores the discrete nature of the channel inputs. The

benefits of this statistically decoupled approach and of linear prediction in particular are dramatically illustrated in the next chapter.

APPENDIX 4-1: PROOF OF LEMMA 4-1

Substituting an SVD $\mathbf{H} = \mathbf{USV}^*$ into $\Phi_r = \mathbf{HH}^* + \sigma^2\mathbf{I}$ yields $\Phi_r = \mathbf{U}\Lambda\mathbf{U}^*$, where $\Lambda = \mathbf{SS}^* + \sigma^2\mathbf{I}$ is diagonal. It follows that $\mathbf{W} = \Lambda^{-1/2}\mathbf{U}^*$ is a whitener, satisfying $\mathbf{W}\Phi_r\mathbf{W}^* = \mathbf{I}$, and that $\mathbf{WH} = \Lambda^{-1/2}\mathbf{SV}^*$. Replacing the singular values of \mathbf{WH} by unity produces its polar factor $\mathbf{J}^T\mathbf{V}^*$. From section 4.1, the WR detector is then $\mathbf{C}_{WR} = \mathbf{JQW}$, where \mathbf{JQ} is the polar factor of $(\mathbf{WH})^*$, or $\mathbf{JQ} = \mathbf{VJ}$; thus, we have $\mathbf{C}_{WR} = \mathbf{VJW} = \mathbf{VJ}\Lambda^{-1/2}\mathbf{U}^*$. This proves (4-4). If we define the diagonal matrix $\tilde{\Lambda} = \mathbf{J}\Lambda\mathbf{J}^T = (\tilde{\mathbf{S}}^2 + \sigma^2\mathbf{I})^{-1/2}$, then (4-5) follows from (4-4) and the identity $\mathbf{J}\Lambda = \tilde{\Lambda}\mathbf{J}$.

We now establish by contradiction the uniqueness of the WR detector. Suppose $\mathbf{C}_1 = \mathcal{P}\{\mathbf{W}_1\mathbf{H}\}^*\mathbf{W}_1$ and $\mathbf{C}_2 = \mathcal{P}\{\mathbf{W}_2\mathbf{H}\}^*\mathbf{W}_2$ denote two distinct WR detectors derived from whiteners \mathbf{W}_1 and \mathbf{W}_2 , respectively, where $\mathcal{P}\{\mathbf{A}\} = \mathbf{U}_A\mathbf{V}_A^*$ denotes the polar factor of $\mathbf{A} = \mathbf{U}_A\mathbf{S}_A\mathbf{V}_A^*$. Since \mathbf{W}_1 and \mathbf{W}_2 are both whiteners, there exists a unitary \mathbf{Q} such that $\mathbf{W}_2 = \mathbf{QW}_1$. It follows that $\mathbf{C}_2 = \mathcal{P}\{\mathbf{QW}_1\mathbf{H}\}^*\mathbf{QW}_1 = (\mathbf{Q}\mathcal{P}\{\mathbf{W}_1\mathbf{H}\})^*\mathbf{QW}_1 = \mathcal{P}\{\mathbf{W}_1\mathbf{H}\}^*\mathbf{W}_1 = \mathbf{C}_1$, a contradiction. \square

APPENDIX 4-2: PROOF OF LEMMA 4-2

Using (4-5), we can express the error of the whiten-rotate detector as

$$\begin{aligned}
 \mathbf{e}_k &= (\mathbf{C}_{WR}\mathbf{H} - \mathbf{I})\mathbf{x}_k + \mathbf{C}_{WR}\mathbf{n}_k \\
 &= \mathbf{V}[\tilde{\Lambda}\tilde{\mathbf{S}} - \mathbf{I}]\mathbf{V}^*\mathbf{x}_k + \mathbf{V}\tilde{\Lambda}\mathbf{J}\mathbf{U}^*\mathbf{n}_k,
 \end{aligned} \tag{4-20}$$

where $\tilde{\Lambda} = (\tilde{\mathbf{S}}^2 + \sigma^2\mathbf{I})^{-1/2}$. The covariance $\Phi_e = E[\mathbf{e}_k\mathbf{e}_k^*]$ of this error is given by

$$\begin{aligned}
 \Phi_e &= \mathbf{V}[\tilde{\Lambda}\tilde{\mathbf{S}} - \mathbf{I}]^2\mathbf{V}^* + \sigma^2\mathbf{V}\tilde{\Lambda}^2\mathbf{V}^* \\
 &= \mathbf{V}[\tilde{\Lambda}^2\tilde{\mathbf{S}}^2 + \mathbf{I} - 2\tilde{\Lambda}\tilde{\mathbf{S}} + \sigma^2\tilde{\Lambda}^2]\mathbf{V}^* \\
 &= 2\mathbf{V}[\mathbf{I} - \tilde{\Lambda}\tilde{\mathbf{S}}]\mathbf{V}^*.
 \end{aligned} \tag{4-21}$$

The MSE of the i -th user $E[|e_k^{(i)}|^2]$ is then given by (4-7).

Similarly, using (3-33), we can express the error of the MMSE detector as

$$\begin{aligned}
 \mathbf{e}_k &= (\mathbf{C}_{MMSE}\mathbf{H} - \mathbf{I})\mathbf{x}_k + \mathbf{C}_{MMSE}\mathbf{n}_k \\
 &= [(\mathbf{H}^*\mathbf{H} + \sigma^2\mathbf{I})^{-1}\mathbf{H}^*\mathbf{H} - \mathbf{I}]\mathbf{x}_k + (\mathbf{H}^*\mathbf{H} + \sigma^2\mathbf{I})^{-1}\mathbf{H}^*\mathbf{n}_k \\
 &= (\mathbf{H}^*\mathbf{H} + \sigma^2\mathbf{I})^{-1}[\mathbf{H}^*\mathbf{H} - (\mathbf{H}^*\mathbf{H} + \sigma^2\mathbf{I})]\mathbf{x}_k + (\mathbf{H}^*\mathbf{H} + \sigma^2\mathbf{I})^{-1}\mathbf{H}^*\mathbf{n}_k \\
 &= -\sigma^2\mathbf{V}\tilde{\Lambda}^2\mathbf{V}^*\mathbf{x}_k + \mathbf{V}\tilde{\Lambda}^2\mathbf{S}^*\mathbf{U}^*\mathbf{n}_k.
 \end{aligned} \tag{4-22}$$

The covariance of this error is

$$\begin{aligned}
 \Phi_e &= \mathbf{V}[\sigma^4\tilde{\Lambda}^4 + \sigma^2\tilde{\Lambda}^4\tilde{\mathbf{S}}^2]\mathbf{V}^* \\
 &= \sigma^2\mathbf{V}\tilde{\Lambda}^2\mathbf{V}^*,
 \end{aligned} \tag{4-23}$$

and the corresponding MSE of the i -th user is then given by (4-8). \square

APPENDIX 4-3: PROOF OF LEMMA 4-3

Given a Hermitian matrix Φ of dimension $m \times m$ and rank $n \leq m$, there exists a square-root matrix \mathbf{S} such that $\mathbf{S}\mathbf{S}^* = \Phi$. Since Φ has rank n , the rows $\{\mathbf{s}_1, \mathbf{s}_2, \dots, \mathbf{s}_m\}$ of \mathbf{S} span an n -dimensional space \mathbf{S} . Performing the Gram-Schmidt orthonormalization procedure on the ordered rows of \mathbf{S} produces a set of m row vectors $\{\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_m\}$, exactly $m - n$ of which are zero, and n of which form an orthonormal basis of \mathbf{S} . Thus, we can write

$$\mathbf{S} = \mathbf{F}\mathbf{V}, \quad (4-24)$$

where the rows of \mathbf{V} are the vectors \mathbf{v}_i , $1 \leq i \leq m$, and \mathbf{F} is a Gram matrix:

$$\mathbf{F} = \begin{bmatrix} F_{1,1} & & & \mathbf{0} \\ F_{2,1} & F_{2,2} & & \\ \vdots & & \ddots & \\ F_{m,1} & F_{m,2} & \dots & F_{m,m} \end{bmatrix}, \quad (4-25)$$

with $F_{i,j} = \langle \mathbf{s}_i, \mathbf{v}_j \rangle$. There exists a set of $m - n$ unit-norm vectors $\{\tilde{\mathbf{v}}_1, \tilde{\mathbf{v}}_2, \dots, \tilde{\mathbf{v}}_{m-n}\}$ orthogonal to the n non-zero rows of \mathbf{V} . Let $\tilde{\mathbf{V}}$ be a unitary matrix formed by replacing the zero rows of \mathbf{V} with this set of vectors. Because the columns of \mathbf{F} multiplying the zero rows of \mathbf{V} in (4-24) are also zero, it follows that $\mathbf{S} = \mathbf{F}\tilde{\mathbf{V}}$. Because \mathbf{F} may have complex diagonal elements, we define $\mathbf{G} = \mathbf{F} \text{diag}\left(\frac{|F_{i,i}|}{F_{i,i}}\right)$ and $\mathbf{U} = \tilde{\mathbf{V}} \text{diag}\left(\frac{F_{i,i}}{|F_{i,i}|}\right)$. It follows that $\mathbf{S} = \mathbf{G}\mathbf{U}$ and that

$$\Phi = \mathbf{G}\mathbf{G}^*. \quad (4-26)$$

The factorization in (4-26) is unique because for any other square root $\tilde{\mathbf{S}} \neq \mathbf{S}$, there exists a unitary matrix \mathbf{Q} such that $\tilde{\mathbf{S}} = \mathbf{S}\mathbf{Q} = \mathbf{G}\mathbf{U}\mathbf{Q} = \mathbf{G}\tilde{\mathbf{U}}$, where $\tilde{\mathbf{U}}$ is also unitary.

The factor \mathbf{G} can be decomposed as $\mathbf{G} = \mathbf{M}\mathbf{D}$, where \mathbf{M} is lower triangular and monic, and where $\mathbf{D} = \text{diag}(\mathbf{G})$, such that

$$\Phi = \mathbf{M}\mathbf{D}^2\mathbf{M}^*. \quad (4-27)$$

If any of the first $m - 1$ diagonal elements $D_{j,j}$, where $j \in \{1, \dots, m - 1\}$ of \mathbf{D} , are zero, then the elements $M_{i,j}$ for $i > j$ of \mathbf{M} are not unique. Therefore, \mathbf{M} is unique if and only if the first $m - 1$ rows of \mathbf{S} are linearly independent, or equivalently, if and only if the first $m - 1$ rows (or columns) of Φ are linearly independent. \square

APPENDIX 4-4: PROOF OF THEOREM 4-2

Given a strictly lower-triangular predictor \mathbf{P} and prediction error $\mathbf{e}_k = (\mathbf{I} - \mathbf{P})\mathbf{r}_k$, the mean-square error $J = E[\|\mathbf{e}_k\|^2]$ can be expressed as

$$J = \text{tr}[(\mathbf{I} - \mathbf{P})\Phi(\mathbf{I} - \mathbf{P})^*], \quad (4-28)$$

where $\Phi = E[\mathbf{r}_k\mathbf{r}_k^*]$. Applying the factorization in (4-13) yields

$$J = \text{tr}[(\mathbf{I} - \mathbf{P})\mathbf{M}\mathbf{D}^2\mathbf{M}^*(\mathbf{I} - \mathbf{P})^*]. \quad (4-29)$$

Because $(\mathbf{I} - \mathbf{P})\mathbf{M}$ is monic and lower triangular, it can be expressed as $\mathbf{I} + \mathbf{B}$, where \mathbf{B} is strictly lower triangular. Thus, the cost function in terms of \mathbf{B} is

$$\begin{aligned} J &= \text{tr}[(\mathbf{I} + \mathbf{B})\mathbf{D}^2(\mathbf{I} + \mathbf{B})^*] \\ &= \text{tr}[\mathbf{D}^2 + \mathbf{B}\mathbf{D}^2\mathbf{B}^* + \mathbf{B}\mathbf{D}^2 + \mathbf{D}^2\mathbf{B}^*]. \end{aligned} \quad (4-30)$$

Because \mathbf{B} is strictly lower triangular, the traces of $\mathbf{B}\mathbf{D}^2$ and $\mathbf{D}^2\mathbf{B}^*$ are both zero in (4-30). Furthermore, the trace of \mathbf{D}^2 is independent of \mathbf{B} . Thus, it suffices to minimize $\text{tr}(\mathbf{B}\mathbf{D}^2\mathbf{B}^*)$, which is clearly accomplished by any strictly lower-triangular $\tilde{\mathbf{B}}$ in the left null space of \mathbf{D} . The best predictor can thus be expressed as

$$\mathbf{P} = \mathbf{I} - (\mathbf{I} + \tilde{\mathbf{B}})\mathbf{M}^{-1}, \quad (4-31)$$

with $\tilde{\mathbf{B}}$ so defined. Observe that $(\mathbf{I} + \tilde{\mathbf{B}})\mathbf{M}^{-1}$ is both monic and lower triangular, and that it diagonalizes Φ :

$$[(\mathbf{I} + \tilde{\mathbf{B}})\mathbf{M}^{-1}]\Phi[(\mathbf{M}^{-1})^*(\mathbf{I} + \tilde{\mathbf{B}})^*] = (\mathbf{I} + \tilde{\mathbf{B}})\mathbf{D}^2(\mathbf{I} + \tilde{\mathbf{B}})^* = \mathbf{D}^2. \quad (4-32)$$

Hence, for any inverse generalized Cholesky factor \mathbf{M}^{-1} , the product $(\mathbf{I} + \tilde{\mathbf{B}})\mathbf{M}^{-1} \equiv \tilde{\mathbf{M}}^{-1}$ is the inverse of some other Cholesky factor. Thus (4-31) reduces to (4-14). \square

APPENDIX 4-5:

A WHITEN-ROTATE DETECTOR FOR NOISELESS CHANNELS

Both the whiten-rotate structure of section 4.1 and its adaptive implementation of section 4.3 are derived under the assumption of nonzero noise. Although a good assumption in practice, the special case of zero noise is also of interest because it provides insight into the behavior of the whiten-rotate detector in the limit of high SNR. Therefore, this appendix contains a brief summary of a whiten-rotate detector for channels without noise. We emphasize that the discussion that follows applies to the whiten-first strategy of section 4.1 only; the project-first strategy of section 4.2 can be applied to noiseless channels without modification.

Without noise, the covariance matrix of the observation vector \mathbf{r}_k in (3-1) is $\Phi_{\mathbf{r}} = \mathbf{H}\mathbf{H}^*$ with rank n . Therefore, if the channel is tall ($m > n$), there does not exist an $m \times m$ whitening matrix \mathbf{W} . Consequently, the whiten-first approach illustrated in Fig. 4-1 and described by (4-3) is not valid. Nevertheless, there does exist a *short* $n \times m$ whitener \mathbf{W} , satisfying $\mathbf{W}\mathbf{H}\mathbf{H}^*\mathbf{W}^* = \mathbf{I}$, and this identity implies that $\mathbf{W}\mathbf{H} = \mathbf{T}$ for some unitary matrix \mathbf{T} . Hence, the whiten-rotate filter $\mathbf{T}^*\mathbf{W}$ achieves zero MSE, which is certainly minimal. In fact, any zero-forcing detector of the form $\mathbf{C}_{ZF} = \mathbf{H}^\dagger + \mathbf{N}$, where $\mathbf{N}^* \in \text{null}(\mathbf{H}^*)$, can be interpreted as a whiten-rotate detector when it is factored according to the QR-factorization theorem [121]: $\mathbf{C}_{ZF} = \mathbf{Q}\mathbf{W}$. Thus, in the absence of noise, the whiten-rotate detector is not unique.

According to (4-13) and (4-14), the least-mean-square linear prediction error $\mathbf{e}_k = (\mathbf{I} - \mathbf{P})\mathbf{r}_k$ has a diagonal covariance matrix $\Phi_{\mathbf{e}} = \mathbf{D}^2$, where $\mathbf{H}\mathbf{H}^* = \mathbf{M}\mathbf{D}^2\mathbf{M}^*$ is a general-

ized Cholesky factorization. Because \mathbf{H} has rank n , exactly $m - n$ of the diagonal elements of \mathbf{D} are zero, which implies that the corresponding components of the error signal \mathbf{e}_k are identically zero. Clearly we lose nothing by discarding these zeros, thereby producing a reduced vector $\tilde{\mathbf{e}}_k$ of dimension n . Let \mathbf{J} denote the $n \times m$ matrix that extracts the nonzero components of \mathbf{e}_k , so that $\tilde{\mathbf{e}}_k = \mathbf{J}\mathbf{e}_k$ has covariance matrix $\Phi_{\tilde{\mathbf{e}}} = \mathbf{J}\mathbf{D}^2\mathbf{J}^* = \tilde{\mathbf{D}}^2$, where $\tilde{\mathbf{D}} = \mathbf{J}\mathbf{D}$ is a full-rank $n \times n$ diagonal matrix containing all of the nonzero components of \mathbf{D} . The diagonal filter $\tilde{\mathbf{D}}^{-1}$ then whitens the reduced error signal. An $n \times m$ whitening matrix can thus be expressed as $\mathbf{W} = \tilde{\mathbf{D}}^{-1}\mathbf{J}(\mathbf{I} - \mathbf{P})$.

The preceding development suggests a method for blind adaptive implementation of \mathbf{W} for the case of low noise. The adaptive predictor of (4-15) is guaranteed to converge to a solution of the form (4-14). After convergence, the $m - n$ components of the prediction error that are nearly zero can be discarded. The n remaining error components can be adaptively scaled according to (4-16). (The MPLL can be applied without modification after the whitener.)

CHAPTER 5

LINEAR PREDICTION

THE CONCEPT of using linear prediction (LP) for blind channel identification and equalization originated with Slock and colleagues [94-96] in the context of fractionally spaced equalization, where a fractionally (T/m) spaced sampler produces a baud-spaced SIMO channel of dimension $m \times 1$. (Recall Fig. 2-5.) Abed-Meraim [91] used these ideas to develop an algorithm for identification of SIMO channels. Use of LP was extended to tall MIMO channels by Gorokhov *et al.* [15][118], Delfosse and Loubaton [135], and Icart and Gautier [119]. All of these algorithms are batch-oriented, where a block of data is collected and used to estimate an autocorrelation matrix, and then a Yule-Walker or similar equation is solved, perhaps by using the Levinson-Durbin [136] algorithm. In contrast, the techniques we detail in this chapter are adaptive with complexity on the order of LMS or one of the various CMA extensions. Moreover, our express goal is always constrained MMSE detection; that is, our detectors are designed to minimize the MSE of all users subject to a set of architectural constraints.

We broaden consideration in this chapter to include channels with memory, as described by (1-1), and we present a family of detectors based on *linear spatio-temporal prediction*. These detectors can be viewed as extensions of the prediction-based whiten-rotate detector of chapter 4; they take the form of a WR detector preceded by one

or more temporal prediction-error filters. Temporal LP is used to eliminate or nearly eliminate the channel memory prior to application of spatial methods. This approach once again conforms to our philosophy of maximal exploitation of second-order statistics; higher-order statistics are exploited only at the last step to estimate a unitary matrix. The virtues of this approach, data efficiency and insensitivity to the channel input distributions, are well documented in this chapter.

In section 5.1, we discuss equivalent representations for the FIR channel of (1-1), including moving-average, autoregressive, and other models. In section 5.2, we extend the notion of minimum phase to MIMO channels. We argue that almost all tall MIMO channels have this minimum-phase property. In section 5.3, we discuss the necessary and sufficient conditions for the models of section 5.1 to exist. We attempt to develop insight into the physical meaning of these conditions. In section 5.4, we discuss temporal prediction in a noiseless environment. We define three predictor architectures, each closely related to one of the channel models of section 5.1. In section 5.5, we formally define multiuser detectors based on linear prediction. We characterize their behavior and performance in the presence of noise, and discuss many of their properties. In section 5.6, we detail blind adaptive implementations of the detectors and demonstrate applications, including an adaptive fractionally spaced equalizer (FSE) for single-user systems using highly shaped signal constellations and a blind multiuser detector for asynchronous CDMA systems.

5.1 Equivalent Channel Models

We again consider channels with memory as described by (1-1), but examine first the noiseless case, for which the receiver observation is given by

$$\mathbf{r}_k = \sum_{i=0}^M \mathbf{H}_i \mathbf{x}_{k-i}. \quad (5-1)$$

We assume that the channel is strictly tall $m > n$, and the input sequence is white: $E[\mathbf{x}_k \mathbf{x}_{k-l}^*] = \mathbf{I} \delta_l$. From linear system theory, \mathbf{r}_k is known as a *moving-average* (MA) process, because it is a weighted average of samples from a white process. We make one additional assumption:

$$\text{rank}[\mathbf{H}_M] = \text{rank}[\mathbf{H}(z)] = n \text{ for all } z \text{ including } \infty. \quad (5-2)$$

Including $z = \infty$ implies that \mathbf{H}_0 also has full column rank. Tall FIR channels satisfying (5-2) have many remarkable properties. For example, the output of such a channel is not only moving average, as in (5-1), but is also finitely autoregressive (AR), which, among other things, implies the existence of an FIR inverse. (Recall Example 2-3.) This is in stark contrast to SISO channels, for which an FIR channel can never have an FIR inverse. Throughout this thesis, we use the labels “moving average” and “autoregressive” and similar terms to refer to the channel itself, not merely to its output, because these equivalent models are innate properties of the channel; the stochastic or deterministic nature of the signals has no relevance. The following section describes the first of three equivalent models for (5-1), all of which are intimately related to linear prediction.

5.1.1 An Autoregressive Channel Model

It is possible to model a tall MA channel (5-1) satisfying (5-2) by the feedback structure of Fig. 5-1(b). The channel output \mathbf{r}_k can be written as

$$\mathbf{r}_k = \sum_{i=1}^N \mathbf{A}_i \mathbf{r}_{k-i} + \mathbf{H}_0 \mathbf{x}_k, \quad (5-3)$$

where the feedback filter $\mathbf{A}(z) = \sum_{i=1}^N \mathbf{A}_i z^{-i}$ is square ($m \times m$) and strictly causal. Such a process is called *autoregressive* (AR), because the output can be expressed in terms of its past. The existence of this model implies that an FIR filter can invert an FIR channel. Observe that the transfer function of the linear feedback system is $[\mathbf{I} - \mathbf{A}(z)]^{-1}$; therefore, $\mathbf{H}(z) = [\mathbf{I} - \mathbf{A}(z)]^{-1} \mathbf{H}_0$, or equivalently $[\mathbf{I} - \mathbf{A}(z)] \mathbf{H}(z) = \mathbf{H}_0$. As shown by the following theorem, the memory N of the feedback filter depends on the channel memory M , but it is always finite when $m > n$.

Theorem 5-1. [15] Let the channel $\mathbf{H}(z)$ of (5-1) with $m > n$ satisfy (5-2). If $N \geq \left\lceil \frac{Mn}{m-n} \right\rceil$, then there exists an $m \times m$ FIR filter $\mathbf{A}(z) = \sum_{i=1}^N \mathbf{A}_i z^{-i}$ with memory N such that $[\mathbf{I} - \mathbf{A}(z)] \mathbf{H}(z) = \mathbf{H}_0$.

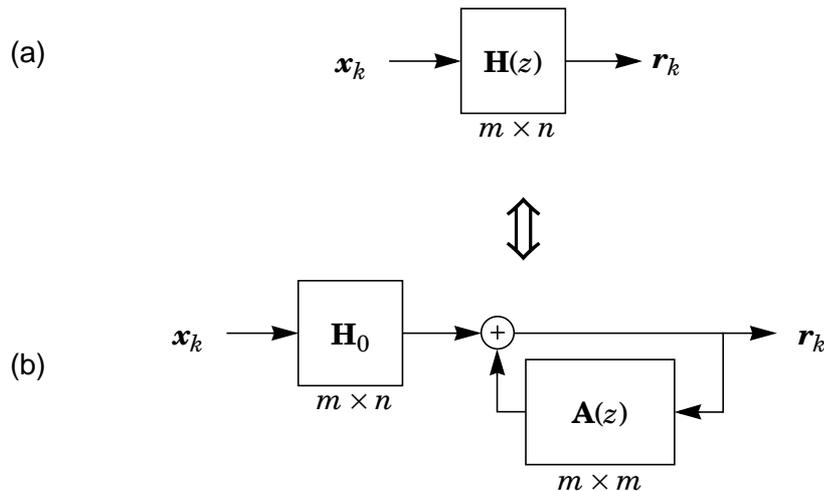


Figure 5-1. Equivalent models for a tall MIMO channel: (a) moving average (MA) and (b) autoregressive (AR).

Proof: [15] We can express $[\mathbf{I} - \mathbf{A}(z)]\mathbf{H}(z) = \mathbf{H}_0$ in block-matrix form as follows:

$$[\mathbf{I} \quad -\mathbf{A}_1 \quad \dots \quad -\mathbf{A}_N] \mathcal{H}_{N+1} = [\mathbf{H}_0 \quad \mathbf{0}_{m \times n(M+N)}], \quad (5-4)$$

where \mathcal{H}_{N+1} is an $m(N+1) \times n(M+N+1)$ block-Toeplitz matrix:

$$\mathcal{H}_{N+1} = \begin{bmatrix} \mathbf{H}_0 & \mathbf{H}_1 & \dots & \mathbf{H}_M & \mathbf{0} & \dots & \mathbf{0} \\ \mathbf{0} & \mathbf{H}_0 & \mathbf{H}_1 & \dots & \mathbf{H}_M & \dots & \mathbf{0} \\ \vdots & & \ddots & \ddots & & \ddots & \\ \mathbf{0} & \dots & \mathbf{0} & \mathbf{H}_0 & \mathbf{H}_1 & \dots & \mathbf{H}_M \end{bmatrix}. \quad (5-5)$$

Subtracting $[\mathbf{I}_m \quad \mathbf{0}_{m \times Nm}] \mathcal{H}_{N+1}$ from both sides of (5-4) yields

$$[\mathbf{0}_{m \times m} \quad -\mathbf{A}_1 \quad \dots \quad -\mathbf{A}_N] \mathcal{H}_{N+1} = -[\mathbf{0}_{m \times m} \quad \mathbf{H}_1 \quad \dots \quad \mathbf{H}_M \quad \mathbf{0}_{m \times Nn}], \quad (5-6)$$

or equivalently,

$$[\mathbf{A}_1 \quad \dots \quad \mathbf{A}_N] \mathcal{H}_N = [\mathbf{H}_1 \quad \dots \quad \mathbf{H}_M \quad \mathbf{0}_{m \times Nn}]. \quad (5-7)$$

This is a system of $mn(M+N)$ scalar equations with m^2N unknowns. From Forney [137], we know that if (5-2) is satisfied then \mathcal{H}_N is full rank for all N , and the equations are linearly independent. Therefore, the system has a solution if $m^2N \geq mn(M+N)$, or equivalently if $N \geq \left\lceil \frac{Mn}{m-n} \right\rceil$. \square

The coefficients $\{\mathbf{A}_i\}$ of $\mathbf{A}(z)$ in (5-3) and Theorem 5-1 are called *AR parameters*; they, together with \mathbf{H}_0 , provide a complete description of the channel. Likewise, the coefficients $\{\mathbf{H}_i\}$ of $\mathbf{H}(z)$ in (5-1) are called *MA parameters*.

Corollary 5-1.1. The coefficients $\{\mathbf{A}_i\}$ of $\mathbf{A}(z) = \sum_{i=1}^N \mathbf{A}_i z^{-i}$ are given by

$$[\mathbf{A}_1 \quad \dots \quad \mathbf{A}_N] = [\mathbf{H}_1 \quad \dots \quad \mathbf{H}_M \quad \mathbf{0}_{m \times Nn}] \mathcal{H}^\dagger + \mathbf{V}, \quad (5-8)$$

where $\mathcal{H} = \mathcal{H}_N$, and for any $\mathbf{V}^* \in \text{null}(\mathcal{H}^*)$. The coefficients are unique if and only if $\frac{Mn}{m-n}$ is an integer and $N = \frac{Mn}{m-n}$.

Proof: The proof follows immediately by verifying that (5-8) satisfies (5-7):

$$\begin{aligned} [\mathbf{A}_1 \dots \mathbf{A}_N] \mathcal{H} &= [\mathbf{H}_1 \dots \mathbf{H}_M \mathbf{0}_{m \times Nn}] \mathcal{H}^\dagger \mathcal{H} + \mathbf{V} \mathcal{H} \\ &= [\mathbf{H}_1 \dots \mathbf{H}_M \mathbf{0}_{m \times Nn}]. \end{aligned} \quad (5-9)$$

The solution is unique if and only if $\text{null}(\mathcal{H}^*)$ is trivial, *i.e.*, if and only if its dimension is zero: $\dim[\text{null}(\mathcal{H}^*)] = mN - n(M+N) = 0$, or equivalently, $N = \frac{Mn}{m-n}$ is an integer. \square

5.1.2 An ARMA Channel Model

It is also possible to model a tall MA channel (5-1) satisfying (5-2) by a hybrid autoregressive-moving-average (ARMA) structure. The model is illustrated in Fig. 5-2, where $\mathbf{H}^L(z) = \sum_{i=0}^L \mathbf{H}_i z^{-i}$ is a truncated version of $\mathbf{H}(z)$, an FIR filter consisting of the first $L \in \{0, 1, \dots, M\}$ taps of $\mathbf{H}(z)$, and $\mathbf{A}^L(z) = \sum_{i=1+L}^{N+L} \mathbf{A}_i z^{-i}$ is a square ($m \times m$) and strictly causal feedback filter. The observation \mathbf{r}_k can thus be expressed as

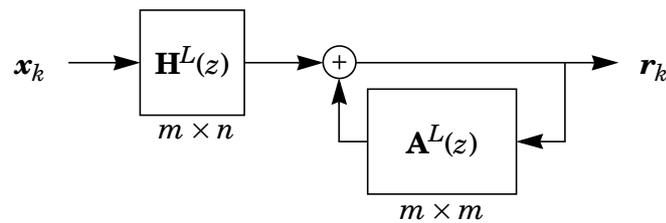


Figure 5-2. An ARMA model for a tall MIMO channel.

$$\mathbf{r}_k = \sum_{i=L+1}^{L+N} \mathbf{A}_i \mathbf{r}_{k-i} + \sum_{i=0}^L \mathbf{H}_i \mathbf{x}_{k-i}. \quad (5-10)$$

The ARMA model reduces to the AR model when $L = 0$. Observe that the linear feedback system has a transfer function given by $[\mathbf{I} - \mathbf{A}^L(z)]^{-1}$, which implies that $[\mathbf{I} - \mathbf{A}^L(z)]\mathbf{H}(z) = \mathbf{H}^L(z)$. Again, the required number N of nonzero taps in the feedback filter is always finite when $m > n$.

Theorem 5-2. Let the channel $\mathbf{H}(z)$ of (5-1) with $m > n$ satisfy (5-2), and let $L \in \{0, 1, \dots, M\}$. If $N \geq \left\lceil \frac{Mn}{m-n} \right\rceil$, then there exists an $m \times m$ FIR filter $\mathbf{A}^L(z) = \sum_{i=L+1}^{L+N} \mathbf{A}_i z^{-i}$ such that $[\mathbf{I} - \mathbf{A}^L(z)]\mathbf{H}(z) = \sum_{i=0}^L \mathbf{H}_i z^{-i}$.

Proof: The proof is similar to that for Theorem 5-1. See Appendix 5-1.

Corollary 5-2.1. The coefficients $\{\mathbf{A}_i\}$ of $\mathbf{A}^L(z) = \sum_{i=L+1}^{L+N} \mathbf{A}_i z^{-i}$ are given by

$$[\mathbf{A}_{L+1} \dots \mathbf{A}_{L+N}] = [\mathbf{H}_{L+1} \dots \mathbf{H}_M \mathbf{0}_{m \times n(L+N)}] \mathcal{H}^\dagger + \mathbf{V}. \quad (5-11)$$

They are unique if and only if $N = \frac{Mn}{m-n}$.

Proof: See Appendix 5-1.

5.1.3 An Autopgressive Model

In addition to the MA, AR, and ARMA models of Fig. 5-1 and Fig. 5-2, a tall MA channel (5-1) satisfying (5-2) can also be modeled as illustrated in Fig. 5-3, where the receiver observation is given by

$$\mathbf{r}_k = \sum_{i=1}^N \mathbf{A}_{-i} \mathbf{r}_{k+i} + \mathbf{H}_M \mathbf{x}_{k-M}. \quad (5-12)$$

We invent the term *autopgressive* (AP) to describe this model because the present output is expressed in terms of *future* outputs; the feedback filter $\mathbf{A}^{\sim 1}(z) = \sum_{i=1}^N \mathbf{A}_{-i} z^i$ is strictly anti-causal. (We denote a negative one step with the superscript ‘ ~ 1 ’ rather than ‘ -1 ’ to avoid confusion with an inverse.) Although the model is somewhat unusual, it is in no way contradictory. The linear feedback section $[\mathbf{I} - \mathbf{A}^{\sim 1}(z)]^{-1}$ is anti-causal, yet the overall channel model $\mathbf{H}(z) = [\mathbf{I} - \mathbf{A}^{\sim 1}(z)]^{-1} \mathbf{H}_M z^{-M}$ is causal. It follows that the FIR filter $[\mathbf{I} - \mathbf{A}^{\sim 1}(z)]$ can be used to isolate the last tap: $[\mathbf{I} - \mathbf{A}^{\sim 1}(z)] \mathbf{H}(z) = \mathbf{H}_M z^{-M}$. Although $[\mathbf{I} - \mathbf{A}^{\sim 1}(z)]$ is anti-causal, and thus cannot be realized, a delayed version $[\mathbf{I} - \mathbf{A}^{\sim 1}(z)] z^{-N}$, which is causal, can be implemented such that $[\mathbf{I} - \mathbf{A}^{\sim 1}(z)] z^{-N} \mathbf{H}(z) = \mathbf{H}_M z^{-M-N}$.

Theorem 5-3. Let the channel $\mathbf{H}(z)$ of (5-1) with $m > n$ satisfy (5-2). If $N \geq \left\lceil \frac{Mn}{m-n} \right\rceil$, then there exists an $m \times m$ FIR filter $\mathbf{A}^{\sim 1}(z) = \sum_{i=1}^N \mathbf{A}_{-i} z^i$ with memory N such that $[\mathbf{I} - \mathbf{A}^{\sim 1}(z)] z^{-N} \mathbf{H}(z) = \mathbf{H}_M z^{-M-N}$.

Proof: See Appendix 5-2.

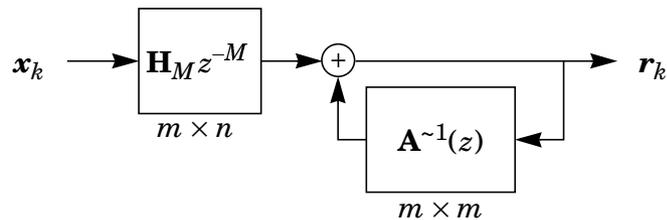


Figure 5-3. An autopgressive model for a tall MIMO channel.

Corollary 5-3.1. The coefficients of $\mathbf{A}^{-1}(z) = \sum_{i=1}^N \mathbf{A}_{-i} z^i$ are given by

$$[\mathbf{A}_{-N} \dots \mathbf{A}_{-1}] = [\mathbf{0}_{m \times Nn} \mathbf{H}_0 \dots \mathbf{H}_{M-1}] \mathcal{H}^\dagger + \mathbf{V}. \quad (5-13)$$

They are unique if and only if $N = \frac{Mn}{m-n}$.

Proof: See Appendix 5-2.

5.2 Minimum-Phase Channels

From traditional linear system theory, we say that a SISO filter or channel $H(z)$ is minimum phase if all its poles and zeros are located inside the unit circle. A minimum-phase channel has many desirable properties. For example, it always has a causal stable inverse. (A channel with all zeros inside the unit circle has an inverse with all poles inside.) A minimum-phase channel also has the property that its energy is maximally concentrated at its zero-th tap. In other words, among all channels with identical spectra $|H(e^{j\theta})|^2$, the minimum-phase channel has the minimum group delay (negative phase derivative) [33]. Delfousse extends the concept of minimum phase to MIMO channels [135] as follows.

Definition 5-1. A MIMO channel $\mathbf{H}(z)$ is called **minimum phase** if and only if there exists a causal stable left-inverse.

This definition reduces to the usual SISO definition when $m = n = 1$. However, a minimum-phase MIMO channel does not have all of the properties of a minimum-phase SISO channel. Its energy is not necessarily concentrated in its zero-th tap. Furthermore, a fractionally spaced channel that is not minimum phase in the SISO sense can be so according

to Definition 5-1, when viewed as a baud-spaced SIMO channel. It might be less confusing to say that a channel is “minimum phase in the MIMO sense” or that it is “causal stable left-invertible.” Nevertheless, we use Definition 5-1, as stated, for its simplicity.

Consider a tall FIR transfer function $\mathbf{H}(z)$ satisfying (5-2) and an associated AR model with FIR feedback filter $\mathbf{A}(z)$ such that $[\mathbf{I} - \mathbf{A}(z)]\mathbf{H}(z) = \mathbf{H}_0$. It follows that $\mathbf{H}(z)$ has a left-inverse $\mathbf{C}(z)$ satisfying $\mathbf{C}(z)\mathbf{H}(z) = \mathbf{I}$; in particular, $\mathbf{C}(z) = \mathbf{H}_0^\dagger[\mathbf{I} - \mathbf{A}(z)]$, where $\mathbf{H}_0^\dagger = (\mathbf{H}_0^* \mathbf{H}_0)^{-1} \mathbf{H}_0^*$. The left-inverse $\mathbf{C}(z)$ is causal and FIR, just like $\mathbf{A}(z)$, and because it is FIR, it is stable. Therefore, Theorem 5-1 implies that every tall FIR channel $\mathbf{H}(z)$ satisfying (5-2) is minimum phase.

Of course, a channel need not be tall to be minimum phase, as illustrated by the following example.

Example 5-1. Consider a 2×2 channel $\mathbf{F}(z)$ with memory $M = 2$:

$$\mathbf{F}(z) = \begin{bmatrix} 2 + z^{-2} & z^{-1} \\ z^{-1} & 1 \end{bmatrix}. \quad (5-14)$$

Observe that $\det[\mathbf{F}(z)] = 2 + z^{-2} - z^{-1}z^{-1} = 2$ is nonzero everywhere. Therefore, an inverse exists, and is given by

$$\mathbf{F}^{-1}(z) = \frac{1}{2} \begin{bmatrix} 1 & -z^{-1} \\ -z^{-1} & 2 + z^{-2} \end{bmatrix}. \quad (5-15)$$

Observe that $\mathbf{F}^{-1}(z)$ is both causal and stable, so $\mathbf{F}(z)$ is minimum phase.

Channels such as those in the previous example belong to an important class of channels known as unimodal channels, all of which are minimum phase.

Definition 5-2. [34] A square FIR channel $\mathbf{F}(z) = \sum_{i=0}^M \mathbf{F}_i z^{-i}$ is called **unimodal** if $\det[\mathbf{F}(z)]$ is nonzero for all z , including ∞ .

Lemma 5-1. All unimodal channels are minimum phase.

Proof: All unimodal channels $\mathbf{F}(z)$ have an inverse because $\det[\mathbf{F}(z)] \neq 0$. The inverse is causal and FIR, and therefore stable, because each element in the inverse is proportional to a cofactor of $\mathbf{F}(z)$, all of which are FIR and stable. \square

In fact, *any* FIR channel, tall or square, that satisfies (5-2) is minimum phase. There is an important distinction, however, between square and tall channels: *almost all tall channels are minimum phase, and almost all square channels are not.* In other words, if we were to select at random¹⁷ a single tall channel $\mathbf{H}(z) = \sum_{i=0}^M \mathbf{H}_i z^{-i}$ from the set of all such channels, it would satisfy (5-2) with probability one. In contrast, a square channel $\mathbf{G}(z) = \sum_{i=0}^M \mathbf{G}_i z^{-i}$ selected in the same manner would satisfy (5-2) with probability zero. In general, $\det[\mathbf{G}(z)]$ would be a polynomial of degree M and therefore would have M zeros. The channel $\mathbf{G}(z)$ would be minimum-phase only if $\det[\mathbf{G}(z)]$ were a constant, a zero-probability event. It is this distinction that makes tall channels remarkable. Of course, it can be argued that “real-world” channels are not generated in this way. A natural question to ask is, “What types of channels satisfy (5-2)?” In the next section, we address this question and attempt to develop some insight into the physical meaning of (5-2).

17. For a given M , select each element of each coefficient \mathbf{H}_i uniformly and independently from the set of complex numbers $|z| < 1$.

5.3 Necessary and Sufficient Conditions

We have seen that (5-2) is sufficient for a tall (or square) channel to be minimum phase, but we have not addressed the physical significance of (5-2). Suppose we have an $n \times n$ channel $\mathbf{G}(z)$ that is not minimum phase. We could follow $\mathbf{G}(z)$ with a tall $m \times n$ minimum-phase filter $\tilde{\mathbf{H}}(z)$ such that the cascade combination is tall: $\mathbf{H}(z) = \tilde{\mathbf{H}}(z)\mathbf{G}(z)$. However, as intuition might suggest, $\mathbf{H}(z)$ does not inherit the minimum-phase property from $\tilde{\mathbf{H}}(z)$. Examining $\mathbf{H}(z)$ more rigorously, we see that since $\mathbf{G}(z)$ is not minimum phase, it is not unimodal, so there exists a $z = z_0$ (possibly ∞) for which $\det[\mathbf{G}(z_0)] = 0$. Consequently, $\mathbf{H}(z_0)$ is not full rank, and $\mathbf{H}(z)$ does not satisfy (5-2). We conclude that in order for a tall channel $\mathbf{H}(z)$ to satisfy (5-2), there can be no square channels, other than minimum-phase channels, hidden within it. Such a channel is said to be irreducible [34].

Definition 5-3. An $m \times n$ channel $\mathbf{H}(z)$ with $m > n$ is said to be **irreducible** if all square right-factors are minimum phase. It is said to be **reducible** if there exists a right-factor that is not minimum phase.

We exclude minimum-phase factors, because any tall channel $\mathbf{H}(z)$ can be expressed as the product of some other tall channel $\tilde{\mathbf{H}}(z)$ and a minimum-phase channel $\mathbf{F}(z)$ according to $\mathbf{H}(z) = \tilde{\mathbf{H}}(z)\mathbf{F}(z)$. This other minimum-phase channel $\tilde{\mathbf{H}}(z)$ is guaranteed to exist because $\mathbf{F}(z)$ is left-invertible: $\tilde{\mathbf{H}}(z) = \mathbf{H}(z)\mathbf{F}^{-1}(z)$. There are infinitely many such factorizations. Irreducibility is thus a generalization of the second-order identifiability condition originally given by Slock [94] for SIMO channels, namely that the subchannels can have no common zeros (no common SISO factors). Observe that an irreducible channel $\mathbf{H}(z)$ is full rank for all nonzero z including ∞ , because the determinant of all square right-factors is

identically nonzero. However, irreducibility does not imply that $\mathbf{H}(0)$ or \mathbf{H}_M is full rank, and thus does not alone satisfy (5-2).

To investigate the rank of \mathbf{H}_M we need to review the final value theorem for MIMO systems [34] as well as the concept of column-reduced channels.

Lemma 5-2. Initial and Final Values. For the channel $\mathbf{H}(z)$ of (5-1), the initial coefficient \mathbf{H}_0 is given by

$$\mathbf{H}_0 = \mathbf{H}(\infty) \equiv \lim_{z \rightarrow \infty} \mathbf{H}(z). \quad (5-16)$$

If the subchannels of all users have memory M , the final coefficient \mathbf{H}_M is given by

$$\mathbf{H}_M = \mathbf{H}(0) = \lim_{z \rightarrow 0} \mathbf{H}(z)z^{-M}. \quad (5-17)$$

If the subchannels do not have the same memory, then

$$\mathbf{H}(0) = \lim_{z \rightarrow 0} \mathbf{H}(z) \text{diag}(z^{M_1} \dots z^{M_n}), \quad (5-18)$$

is a matrix whose i -th column is that of \mathbf{H}_{M_i} , where M_i is the memory in the i -th column of $\mathbf{H}(z)$.

Definition 5-4. [34] The channel $\mathbf{H}(z)$ of (5-1) is said to be **column reduced** if its columns are linearly independent, *i.e.* if $\mathbf{H}(z)\mathbf{D}(z) = \mathbf{0} \Leftrightarrow \mathbf{D}(z) = \mathbf{0}$ where $\mathbf{D}(z) = \sum_{i=0}^{\infty} \mathbf{D}_i z^{-i}$ is $n \times n$.

For the special case of square column-reduced channels, the degree of the determinant is equal to the sum of the memories of the users: $\deg[\det \mathbf{H}(z)] = \sum_{i=1}^n M_i$. The key property of column-reduced channels is that they are full column rank at $z = 0$.

Lemma 5-3. [34] The channel $\mathbf{H}(z)$ of (5-1) is **column reduced** if and only if it is full column rank at $z = 0$.

From Lemmas 5-2 and 5-3, we conclude two things: first, the leading tap \mathbf{H}_0 of an irreducible channel is full rank, and second, the final tap \mathbf{H}_M of a column-reduced channel is full rank provided that all users have memory M . Thus, we arrive at the following result.

Theorem 5-4. The channel $\mathbf{H}(z)$ of (5-1) satisfies (5-2) if and only if it satisfies *all* of the following conditions:

1. $\mathbf{H}(z)$ is irreducible;
2. $\mathbf{H}(z)$ is column reduced;
3. All users have memory M . (5-19)

We stress that while either (5-2) or (5-19) is sufficient for a tall channel $\mathbf{H}(z)$ to be minimum phase, neither is necessary. For example, Gorokhov *et al.* [15] have shown that even if all users do not have identical memory, there still exists an integer $N \leq \sum_{i=1}^n M_i$ such that an AR model exists. In Example 5-1, the users do not have identical memory, and yet the corresponding channel is minimum phase. In fact, the channel in that example is neither column reduced nor tall. Of the three criteria in (5-19), only the first, irreducibility, is necessary for a tall channel to be minimum phase.

If a channel is reducible, then only the irreducible factor can be identified from second-order statistics. Suppose, for example, that $\mathbf{H}(z)$ is tall, but reducible, and thus not minimum phase. We can always factor the channel as $\mathbf{H}(z) = \tilde{\mathbf{H}}(z)\mathbf{U}(z)$, such that $\tilde{\mathbf{H}}(z)$ is minimum phase and $\mathbf{U}(z)$ is allpass: $\mathbf{U}(z)\mathbf{U}^*(1/z^*) = \mathbf{I}$. The channels $\mathbf{H}(z)$ and $\tilde{\mathbf{H}}(z)$ are indistinguishable from their second-order output statistics; *i.e.*, $\mathbf{H}(z)\mathbf{H}^*(1/z^*) =$

$\tilde{\mathbf{H}}(z)\tilde{\mathbf{H}}^*(1/z^*)$. Any second-order algorithm for equalization or identification of $\mathbf{H}(z)$ would behave as if $\tilde{\mathbf{H}}(z)$ were the channel. Therefore it follows that any such algorithm could thus be used to identify or equalize $\tilde{\mathbf{H}}(z)$, the irreducible part of $\mathbf{H}(z)$. This is adequate if the allpass term $\mathbf{U}(z)$ is inconsequential or if it can be resolved by other means. Consider the following example.

Example 5-2. Suppose that a channel $\tilde{\mathbf{H}}(z)$ satisfies (5-2) and is thus minimum phase. This means there exists a causal stable filter $\mathbf{C}(z)$ such that $\mathbf{C}(z)\tilde{\mathbf{H}}(z) = \mathbf{I}$. Now define $\mathbf{H}(z) = \tilde{\mathbf{H}}(z)\mathbf{U}(z)$, where $\mathbf{U}(z)$ is a diagonal delay:

$$\mathbf{U}(z) = \text{diag}(z^{-D_1} \dots z^{-D_n}), \quad (5-20)$$

for integers $D_i > 0$. The diagonal delay is not unimodal because $\det[\mathbf{U}(z)]$ is zero at $z = \infty$. Clearly, the only left-inverse of $\mathbf{U}(z)$ is anti-causal. So $\mathbf{U}(z)$ is not minimum phase. Hence, $\mathbf{H}(z)$ is neither irreducible nor minimum phase. Yet the filter $\mathbf{C}(z)$ is still a viable detector because it leaves only the diagonal delay term, which is harmless: $\mathbf{C}(z)\mathbf{H}(z) = \mathbf{U}(z)$.

5.4 Temporal Linear Prediction

Consider a SISO channel with a white input and with a frequency response $H(e^{j\theta})$. The power spectrum of the output $|H(e^{j\theta})|^2$ conveys no information about the phase of the channel, implying that second-order statistics (SOS) are insufficient for channel identification. However, if the channel were somehow known to be minimum phase, then SOS would be sufficient. In particular, we could use *linear prediction*. We could estimate the present observation r_k from past observations according to $\hat{r}_k = \sum_{i=1}^{\infty} p_i r_{k-i}$, with the

coefficients $\{p_i\}$ of the predictor $P(z) = \sum_{i=1}^{\infty} p_i z^{-i}$ chosen to minimize the variance σ^2 of the prediction error $e_k = r_k - \hat{r}_k$. We could then recover the minimum-phase transfer function (up to a constant factor) via $\hat{H}(z) = \frac{\sigma}{1-P(z)}$. The resulting $\hat{H}(z)$ would differ from $H(z)$ by only an arbitrary complex constant $e^{j\theta}$ with unit magnitude, but this ambiguity would usually not be problematic. (It could be handled at carrier recovery, for example.)

5.4.1 One-Step Prediction

Consider now a tall MIMO channel, which is almost always minimum phase in the sense of Definition 5-1. As we now show, SOS are sufficient for channel identification, up to an arbitrary memoryless unitary matrix \mathbf{U} . Again, we can use linear prediction. We can predict the present observation vector \mathbf{r}_k using a linear combination of the previous N observations: $\hat{\mathbf{r}}_k = \sum_{i=1}^N \mathbf{P}_i \mathbf{r}_{k-i}$, where the coefficients $\{\mathbf{P}_i\}$ are of dimension $m \times m$. We define the one-step predictor of order N as $\mathbf{P}(z) = \sum_{i=1}^N \mathbf{P}_i z^{-i}$. The prediction error is then $\mathbf{e}_k = \mathbf{r}_k - \hat{\mathbf{r}}_k$. If we choose N sufficiently large, then we can use any valid AR representation of (5-3) to express the error as

$$\mathbf{e}_k = \mathbf{H}_0 \mathbf{x}_k + \sum_{i=1}^N (\mathbf{A}_i - \mathbf{P}_i) \mathbf{r}_{k-i}. \quad (5-21)$$

The mean-square prediction error $\xi = E[\|\mathbf{e}_k\|^2] = E\left[\left\|\mathbf{r}_k - \sum_{i=1}^N \mathbf{P}_i \mathbf{r}_{k-i}\right\|^2\right]$ is then

$$\xi = \text{tr} \mathbf{H}_0 \mathbf{H}_0^* + \text{tr} \sum_{i=1}^N \sum_{j=1}^N (\mathbf{A}_i - \mathbf{P}_i) E[\mathbf{r}_{k-i} \mathbf{r}_{k-j}^*] (\mathbf{A}_j - \mathbf{P}_j)^*. \quad (5-22)$$

The double summation is positive semi-definite, so its trace is always non-negative. We can force it to be identically zero by choosing $\mathbf{P}_i = \mathbf{A}_i$ for all $i \in \{1, \dots, N\}$. Therefore, the minimum mean-square prediction error is $\xi_{\min} = \text{tr}(\mathbf{H}_0 \mathbf{H}_0^*)$. In fact, the only solutions

that minimize ξ are $\mathbf{P}(z) = \mathbf{A}(z)$ for any valid feedback filter (satisfying (5-8)) in an AR channel model. The corresponding linear prediction error is $\mathbf{e}_k = \mathbf{H}_0 \mathbf{x}_k$. The channel thus reduces to a memoryless channel, the subject of chapters 3 and 4.

Theorem 5-5. Let the channel $\mathbf{H}(z)$ of (5-1) with $m > n$ satisfy (5-2), and let $\mathbf{P}(z) = \sum_{i=1}^N \mathbf{P}_i z^{-i}$ denote a one-step predictor of order $N \geq \lceil \frac{Mn}{m-n} \rceil$. The coefficients $\{\mathbf{P}_i\}$, minimizing $\xi = E \left[\left\| \mathbf{r}_k - \sum_{i=1}^N \mathbf{P}_i \mathbf{r}_{k-i} \right\|^2 \right]$, are given by (5-8)

$$[\mathbf{P}_1 \dots \mathbf{P}_N] = [\mathbf{H}_1 \dots \mathbf{H}_M \mathbf{0}_{m \times Nn}] \mathcal{H}^\dagger + \mathbf{V}, \quad (5-23)$$

where $\mathbf{V}^* \in \text{null}(\mathcal{H}^*)$. The coefficients are unique if and only if $N = \frac{Mn}{m-n}$.

Let Σ be any $m \times n$ square root of the error autocorrelation matrix $\Phi_e = E[\mathbf{e}_k \mathbf{e}_k^*] = \mathbf{H}_0 \mathbf{H}_0^*$, satisfying $\Sigma \Sigma^* = \mathbf{H}_0 \mathbf{H}_0^*$. This implies that $\Sigma = \mathbf{H}_0 \mathbf{U}$ for some unitary matrix \mathbf{U} . Then the linear prediction channel estimate is $\hat{\mathbf{H}}(z) = [\mathbf{I} - \mathbf{P}(z)]^{-1} \Sigma$. (Notice the similarity to the SISO estimate of $\hat{H}(z) = \frac{\sigma}{1 - P(z)}$.) Since $\mathbf{P}(z) = \mathbf{A}(z)$ and $\Sigma = \mathbf{H}_0 \mathbf{U}$, it follows that $\hat{\mathbf{H}}(z) = [\mathbf{I} - \mathbf{P}(z)]^{-1} \mathbf{H}_0 \mathbf{U}$, and thus $\hat{\mathbf{H}}(z)$ differs from $\mathbf{H}(z) = [\mathbf{I} - \mathbf{P}(z)]^{-1} \mathbf{H}_0$ by only a unitary matrix. So use of second-order output statistics, namely linear prediction, is sufficient for channel identification up to some arbitrary unitary matrix.

5.4.2 Multiple-Step Prediction

Multiple-step prediction is the estimation of the present observation vector \mathbf{r}_k using a linear combination of older observations: $\hat{\mathbf{r}}_k = \sum_{i=L+1}^{L+N} \mathbf{P}_i \mathbf{r}_{k-i}$. We define the $(L+1)$ -step predictor of order N as $\mathbf{P}^L(z) = \sum_{i=L+1}^{L+N} \mathbf{P}_i z^{-i}$. If $L = 0$, this reduces to a one-step predictor. Provided N is sufficiently large, we can use any valid ARMA representation of (5-10), to express the prediction error as

$$\mathbf{e}_k = \sum_{i=0}^L \mathbf{H}_i \mathbf{x}_{k-i} + \sum_{i=L+1}^{L+N} (\mathbf{A}_i - \mathbf{P}_i) \mathbf{r}_{k-i}. \quad (5-24)$$

The mean-square prediction error $\xi = E[\|\mathbf{e}_k\|^2]$ is given by

$$\xi = \text{tr} \sum_{i=0}^L \mathbf{H}_i \mathbf{H}_i^* + \text{tr} \sum_{i=L+1}^{L+N} \sum_{j=L+1}^{L+N} (\mathbf{A}_i - \mathbf{P}_i) E[\mathbf{r}_{k-i} \mathbf{r}_{k-j}^*] (\mathbf{A}_j - \mathbf{P}_j)^*. \quad (5-25)$$

As before, we can force the summand to zero and thus minimize ξ by choosing $\mathbf{P}_i = \mathbf{A}_i$ for all $i \in \{L+1, \dots, L+N\}$, yielding a minimum of $\xi_{\min} = \text{tr} \sum_{i=0}^L \mathbf{H}_i \mathbf{H}_i^*$. Thus, the linear predictor is identical to the feedback filter in the ARMA model of Fig. 5-2, $\mathbf{P}^L(z) = \mathbf{A}^L(z)$, and the corresponding prediction error is $\mathbf{e}_k = \sum_{i=0}^L \mathbf{H}_i \mathbf{x}_{k-i}$.

Theorem 5-6. Let the channel $\mathbf{H}(z)$ of (5-1) with $m > n$ satisfy (5-2), and let $\mathbf{P}^L(z) = \sum_{i=L+1}^{L+N} \mathbf{P}_i z^{-i}$ denote an $(L+1)$ -step predictor of order $N \geq \lceil \frac{Mn}{m-n} \rceil$, where $L \in \{0, 1, \dots, M\}$. The optimal predictor coefficients $\{\mathbf{P}_i\}$ minimizing $\xi = E\left[\left\|\mathbf{r}_k - \sum_{i=L+1}^{L+N} \mathbf{P}_i \mathbf{r}_{k-i}\right\|^2\right]$ are given by (5-11)

$$[\mathbf{P}_{L+1} \cdots \mathbf{P}_{L+N}] = [\mathbf{H}_{L+1} \cdots \mathbf{H}_M \mathbf{0}_{m \times (N+L)n}] \mathcal{H}^\dagger + \mathbf{V}, \quad (5-26)$$

where $\mathbf{V}^* \in \text{null}(\mathcal{H}^*)$. They are unique if and only if $N = \frac{Mn}{m-n}$.

5.4.3 Backward Prediction

We can also estimate the present observation \mathbf{r}_k using future observations: $\hat{\mathbf{r}}_k = \sum_{i=-N}^{-1} \mathbf{P}_i \mathbf{r}_{k-i}$. We call $\mathbf{P}^{-1}(z) = \sum_{i=-N}^{-1} \mathbf{P}_i z^{-i}$ a one-step backward ‘‘predictor’’ of order N for lack of a better name. If N is sufficiently large, we can use the AP representation of (5-12) to express the prediction error as

$$\mathbf{e}_k = \mathbf{H}_M \mathbf{x}_{k-M} + \sum_{i=-N}^{-1} (\mathbf{A}_i - \mathbf{P}_i) \mathbf{r}_{k-i}. \quad (5-27)$$

The mean-square prediction error is

$$\xi = \text{tr} \mathbf{H}_M \mathbf{H}_M^* + \text{tr} \sum_{i=-N}^{-1} \sum_{j=-N}^{-1} (\mathbf{A}_i - \mathbf{P}_i) \mathbf{E}[\mathbf{r}_{k-i} \mathbf{r}_{k-j}^*] (\mathbf{A}_j - \mathbf{P}_j)^*, \quad (5-28)$$

and choosing $\mathbf{P}_i = \mathbf{A}_i$ for all $i \in \{-N, \dots, -1\}$, yields a minimum of $\xi_{\min} = \text{tr} \mathbf{H}_M \mathbf{H}_M^*$. Thus, the optimal backward predictor is identical to the feedback filter in the AP model of Fig. 5-3: $\mathbf{P}^{-1}(z) = \mathbf{A}^{-1}(z)$. The corresponding prediction error is $\mathbf{e}_k = \mathbf{H}_M \mathbf{x}_{k-M}$. Although we cannot implement the predictor $\mathbf{P}^{-1}(z)$ because it is anti-causal, we can implement $z^{-N} \mathbf{P}^{-1}(z)$ with a corresponding delayed prediction error given by $\mathbf{e}_{k-N} = \mathbf{H}_M \mathbf{x}_{k-M-N}$. As was the case with the one-step forward predictor, the effective channel is memoryless; however, the last tap is isolated, rather than the first.

Theorem 5-7. Let the channel $\mathbf{H}(z)$ of (5-1) with $m > n$ satisfy (5-2), and let $\mathbf{P}^{-1}(z) = \sum_{i=-N}^{-1} \mathbf{P}_i z^{-i}$ denote a one-step backward temporal predictor of order $N \geq \left\lceil \frac{Mn}{m-n} \right\rceil$. The optimal coefficients $\{\mathbf{P}_i\}$, minimizing $\xi = E \left[\left\| \mathbf{r}_k - \sum_{i=-1}^{-N} \mathbf{P}_i \mathbf{r}_{k-i} \right\|^2 \right]$, are given by (5-13)

$$[\mathbf{P}_{-N} \dots \mathbf{P}_{-1}] = [\mathbf{0}_{m \times Nn} \mathbf{H}_0 \dots \mathbf{H}_{M-1}] \mathcal{H}^\dagger + \mathbf{V}, \quad (5-29)$$

where $\mathbf{V}^* \in \text{null}(\mathcal{H}^*)$. They are unique if and only if $N = \frac{Mn}{m-n}$.

5.5 Multiuser Detection Using Linear Prediction

In the following sections we present a family of multiuser detectors that combine the temporal LP concepts of section 5.4 and the spatial LP ideas of chapter 4. The first and simplest of these uses a one-step forward temporal predictor. In section 5.5.1, we formally define the forward LP detector and analyze its performance in the presence of noise. In section 5.5.2, we define a generalized forward-backward LP detector, which uses the cascade of a forward and a backward linear predictor, and which in general has better performance, but at the expense of increased complexity.

5.5.1 The Forward LP Detector

So far in this chapter, we have considered only noiseless channels. We have shown that, for the channel $\mathbf{H}(z)$ of (5-1), temporal prediction can be used to eliminate channel memory: $[\mathbf{I} - \mathbf{P}(z)]\mathbf{H}(z) = \mathbf{H}_0$, where $\mathbf{P}(z) = \sum_{i=1}^N \mathbf{P}_i z^{-i}$, as defined by (5-23). Moreover, we have shown in chapter 4 that spatial whitening and rotation can be used to invert the remaining memoryless channel. Recall that, if \mathbf{W} is any $n \times m$ spatial whitener, satisfying in this case $\mathbf{W}\mathbf{H}_0\mathbf{H}_0^* \mathbf{W}^* = \mathbf{I}$, then there exists a unitary matrix \mathbf{U} such that $\mathbf{U}\mathbf{W}\mathbf{H}_0 = \mathbf{I}$. Spatial prediction can be used to implement \mathbf{W} . (See Appendix 4-5.) Therefore, for noiseless channels we can perfectly recover the channel input.

The preceding discussion suggests an architecture for multiuser detection. The *forward linear-predictive (FLP) detector*, illustrated in Fig. 5-4, consists of several stages, the first of which is a one-step forward temporal prediction-error filter $\mathbf{I} - \mathbf{P}(z)$ of order $N \geq \left\lceil \frac{Mn}{m-n} \right\rceil$. The forward prediction error is given by $\mathbf{e}_k = \mathbf{r}_k - \sum_{i=1}^N \mathbf{P}_i \mathbf{r}_{k-i}$, where the coefficients $\{\mathbf{P}_i\}$ are chosen to minimize the mean-square prediction error $\xi = E[\|\mathbf{e}_k\|^2]$.

For nonzero noise, the optimal coefficients are not those of (5-23); they are given below in Theorem 5-8 as a function of the noise variance σ^2 . The next stage of the detector is a memoryless spatial prediction-error filter $\mathbf{I} - \tilde{\mathbf{P}}$. The spatial error is $\tilde{\mathbf{e}}_k = (\mathbf{I} - \tilde{\mathbf{P}})\mathbf{e}_k$, where $\tilde{\mathbf{P}}$ is a strictly lower-triangular $m \times m$ matrix with elements chosen to minimize $\tilde{\xi} = E[\|\tilde{\mathbf{e}}_k\|^2]$. The third stage is a diagonal gain \mathbf{A} chosen such that each component of its output $\mathbf{w}_k = \mathbf{A}\mathbf{e}_k$ has unit energy: $E[\mathbf{w}_k\mathbf{w}_k^*] = \mathbf{I}$. Hence, the cascade $\mathbf{W} = \mathbf{A}(\mathbf{I} - \tilde{\mathbf{P}})$ is an $m \times m$ spatial whitener. The fourth stage of the detector is a unitary rotation matrix \mathbf{Q} chosen to minimize the MSE of all users $E[\|\mathbf{J}\mathbf{Q}\mathbf{w}_k - \tilde{\mathbf{x}}_k\|^2]$, where $\tilde{\mathbf{x}}_k = [x_{k-D_1}^{(1)}, x_{k-D_2}^{(2)}, \dots, x_{k-D_n}^{(n)}]^T$, and where D_i is the delay for user i , also chosen to minimize MSE. We show later in this section that, surprisingly, the optimal delays may not always be zero. The last stage of the detector is the familiar $n \times m$ truncation matrix: $\mathbf{J} = [\mathbf{I} \ \mathbf{0}]$.

A full understanding of the properties and performance of the FLP detector requires additional analysis. We begin by solving for the optimal temporal predictor coefficients in noise.

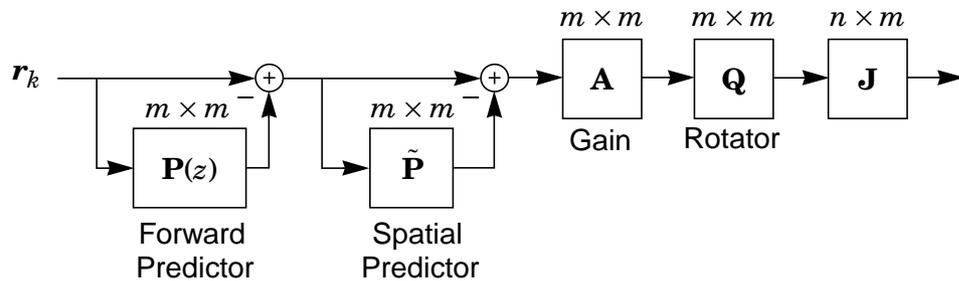


Figure 5-4. A block diagram of the forward LP detector.

Theorem 5-8. Let the channel $\mathbf{H}(z)$ of (1-1) with $m > n$ and $\sigma > 0$ satisfy (5-2), and let $\mathbf{P}(z) = \sum_{i=1}^N \mathbf{P}_i z^{-i}$ denote a one-step predictor of order $N \geq \lceil \frac{Mn}{m-n} \rceil$. The coefficients $\{\mathbf{P}_i\}$, minimizing $\xi = E \left[\left\| \mathbf{r}_k - \sum_{i=-1}^{-N} \mathbf{P}_i \mathbf{r}_{k-i} \right\|^2 \right]$, are uniquely given by

$$\mathbf{P} = [\mathbf{P}_1 \dots \mathbf{P}_N] = \mathbf{K} \mathcal{H}^* (\mathcal{H} \mathcal{H}^* + \sigma^2 \mathbf{I})^{-1} \quad (5-30)$$

$$= \mathbf{K} (\mathcal{H}^* \mathcal{H} + \sigma^2 \mathbf{I})^{-1} \mathcal{H}^*, \quad (5-31)$$

where $\mathbf{K} = [\mathbf{H}_1 \dots \mathbf{H}_M \mathbf{0}_{m \times Nn}]$.

Proof: See Appendix 5-3.

Observe that (5-31) is consistent with (5-23). Because $\lim_{\sigma \rightarrow 0^+} (\mathcal{H}^* \mathcal{H} + \sigma^2 \mathbf{I})^{-1} \mathcal{H}^* = \mathcal{H}^\dagger$, the predictor coefficients for the noisy channel approach the minimum-norm solution for the noiseless channel as the noise variance goes to zero.

We refer to the cascade $\mathbf{F}(z) = [\mathbf{I} - \mathbf{P}(z)]\mathbf{H}(z)$ of the forward prediction-error filter and the channel as the *forward cascade*. We can derive an expression for it using Theorem 5-8.

Corollary 5-8.1. The forward cascade $\mathbf{F}(z) = \sum_{i=0}^{M+N} \mathbf{F}_i z^{-i} = [\mathbf{I} - \mathbf{P}(z)]\mathbf{H}(z)$ has coefficients given by

$$\mathbf{F} = [\mathbf{F}_0 \ \mathbf{F}_1 \ \dots \ \mathbf{F}_{N+M}] = [\mathbf{H}_0 \ | \ [\mathbf{H}_1 \ \dots \ \mathbf{H}_M \ \mathbf{0}_{m \times Nn}] \cdot \Psi], \quad (5-32)$$

where $\Psi = \sigma^2 (\mathcal{H}^* \mathcal{H} + \sigma^2 \mathbf{I})^{-1}$. Furthermore, the prediction error $\mathbf{e}_k = \sum_{i=0}^{M+N} \mathbf{F}_i \mathbf{x}_{k-i}$ has covariance matrix given by

$$\Phi_{\mathbf{e}} = E[\mathbf{e}_k \mathbf{e}_k^*] = \mathbf{H}_0 \mathbf{H}_0^* + \sigma^2 \mathbf{I} + \mathbf{K} \Psi \mathbf{K}^*. \quad (5-33)$$

Proof: See Appendix 5-3.

Again, (5-32) confirms results for the noiseless case; the forward cascade reduces to \mathbf{H}_0 when $\sigma = 0$ because Ψ goes to zero with σ . In general, however, the effect of the prediction error filter is to “scale” the trailing coefficients $\mathbf{K} = [\mathbf{H}_1 \dots \mathbf{H}_M \mathbf{0}_{m \times N_n}]$ by the matrix Ψ , while keeping \mathbf{H}_0 as the leading tap. For $\sigma > 0$, the cascade is not memoryless, but has memory $M + N$. The structure of Ψ has interesting implications for the behavior of the FLP detector; we discuss this in more depth later in this section.

Using (5-33), we can derive expressions for the remaining stages of the FLP detector. From Theorem 4-2, the optimal spatial predictor, minimizing $E[\|(\mathbf{I} - \tilde{\mathbf{P}})\mathbf{e}_k\|^2]$, is $\tilde{\mathbf{P}} = \mathbf{I} - \mathbf{M}^{-1}$, where \mathbf{M} is a term in the Cholesky factorization (4-13) of $\Phi_e = \mathbf{M}\mathbf{D}^2\mathbf{M}^*$. The diagonal gain is $\mathbf{A} = \mathbf{D}^{-1}$. An $m \times m$ spatial whitener is then $\mathbf{W} = \mathbf{A}(\mathbf{I} - \tilde{\mathbf{P}}) = \mathbf{D}^{-1}\mathbf{M}^{-1}$.

Recall that $\mathbf{w}_k = \mathbf{A}(\mathbf{I} - \tilde{\mathbf{P}})\mathbf{e}_k$ is the whitener output, and assume for now that the optimal delay is zero for all users, so that $\tilde{\mathbf{x}}_k = \mathbf{x}_k$. (This assumption is relaxed later in this section.) The best unitary matrix \mathbf{Q}_0 , minimizing $E[\|\mathbf{J}\mathbf{Q}\mathbf{w}_k - \mathbf{x}_k\|^2]$, can then be found in a manner similar to that in chapter 4. Given that the detector front end is $\mathbf{I} - \mathbf{P}(z)$ and that the corresponding temporal prediction error is \mathbf{e}_k , we first find the $n \times m$ matrix \mathbf{C}_0 that minimizes the zero-delay MSE sum $E[\|\mathbf{C}\mathbf{e}_k - \mathbf{x}_k\|^2]$.

Lemma 5-4. Let $\mathbf{e}_k = \sum_{i=0}^{M+N} \mathbf{F}_i \mathbf{x}_k + \mathbf{v}_k$, where \mathbf{x}_k and \mathbf{v}_k are independent with covariances \mathbf{I} and Φ_v , respectively. The unique $n \times m$ matrix \mathbf{C}_D minimizing the D -delay MSE sum $E[\|\mathbf{C}\mathbf{e}_k - \mathbf{x}_{k-D}\|^2]$ is

$$\mathbf{C}_D = \mathbf{F}_D^* \Phi_e^{-1}, \quad (5-34)$$

where $\Phi_e = \sum_{i=0}^{M+N} \mathbf{F}_i \mathbf{F}_i^* + \Phi_v$.

Proof: See Appendix 5-4.

So for a delay of zero, we have $\mathbf{C}_0 = \mathbf{F}_0^* \Phi_e^{-1} = \mathbf{H}_0^* \Phi_e^{-1}$. Furthermore, since $\Phi_e^{-1} = \mathbf{W}^* \mathbf{W}$, it follows that \mathbf{C}_0 can be implemented following the spatial whitener by the short filter $(\mathbf{W}\mathbf{H}_0)^* = (\mathbf{D}^{-1}\mathbf{M}^{-1}\mathbf{H}_0)^*$. Following the discussion of section 4.1, we can express \mathbf{Q}_0 in terms of the polar factor of this short filter:

$$\mathbf{J}\mathbf{Q}_0 = \mathcal{P}^*(\mathbf{D}^{-1}\mathbf{M}^{-1}\mathbf{H}_0), \quad (5-35)$$

where $\mathbf{J} = [\mathbf{I} \ \mathbf{0}]$. We are now in a position to formally define the zero-delay FLP detector.

Definition 5-5. Let the channel $\mathbf{H}(z)$ of (1-1) with $m > n$ and $\sigma > 0$ satisfy (5-2). The $n \times m$ **zero-delay FLP detector** of order $N \geq \left\lceil \frac{Mn}{m-n} \right\rceil$ for $\mathbf{H}(z)$ is uniquely defined as

$$\mathbf{C}_{FLP0}(z) = \mathbf{J}\mathbf{Q}_0\mathbf{D}^{-1}\mathbf{M}^{-1}[\mathbf{I} - \mathbf{P}(z)], \quad (5-36)$$

where $\mathbf{P}(z)$ is given by (5-31), where $\mathbf{M}\mathbf{D}^2\mathbf{M}$ is a Cholesky factorization of Φ_e , given by (5-33), and where \mathbf{Q}_0 satisfies (5-35).

The FLP detector could of course be defined in terms of any square spatial whitener \mathbf{W} , but as we have shown in chapter 4, the spatial whitener based on linear prediction $\mathbf{W} = \mathbf{D}^{-1}\mathbf{M}^{-1}$ has important implementation advantages. Moreover, in the context of (5-36), it has a certain conceptual elegance as well. The cascade of the spatial and temporal prediction-error filters can be interpreted as a single *spatio-temporal prediction* error filter:

$$\mathbf{I} - \tilde{\mathbf{P}}(z) = [\mathbf{I} - \tilde{\mathbf{P}}][\mathbf{I} - \mathbf{P}(z)] = \mathbf{M}^{-1}[\mathbf{I} - \mathbf{P}(z)]. \quad (5-37)$$

This spatio-temporal predictor $\tilde{\mathbf{P}}(z)$ minimizes $\tilde{\xi} = E\left[\left\|\mathbf{r}_k - \sum_{i=0}^N \tilde{\mathbf{P}}_i \mathbf{r}_{k-i}\right\|^2\right]$, where $\tilde{\mathbf{P}}_0$ is strictly lower triangular.

It is of interest to compare the performance of the zero-delay LP detector to that of the zero-delay MMSE detector. It can be shown that the two detectors are closely related.

Lemma 5-5. For the $m \times n$ channel $\mathbf{H}(z)$ of (1-1), satisfying (5-2), with $m > n$ and $\sigma > 0$, let $\mathbf{C}_0(z) = \sum_{i=0}^N \mathbf{C}_i z^{-i}$ denote the unique $n \times m$ ($N+1$)-tap zero-delay MMSE detector, minimizing $E\left[\left\|\mathbf{x}_k - \sum_{i=0}^N \mathbf{C}_i \mathbf{r}_{k-i}\right\|^2\right]$. This detector can be factored as follows:

$$\mathbf{C}_0(z) = \mathbf{C}_0[\mathbf{I} - \mathbf{P}(z)], \quad (5-38)$$

where $\mathbf{P}(z)$ is the optimal one-step temporal predictor of order N defined by (5-31) with prediction error \mathbf{e}_k , and where $\mathbf{C}_0 = \mathbf{H}_0^* \Phi_e^{-1}$ is the memoryless $n \times m$ matrix minimizing $E[\|\mathbf{C}\mathbf{e}_k - \mathbf{x}_k\|^2]$.

Proof: See Appendix 5-5.

In other words, *zero-delay MMSE detection is equivalent to optimal temporal prediction followed by memoryless MMSE detection.* The implication of Lemma 5-5 is that the zero-delay FLP and MMSE detectors should compare much in the same way that the WR and MMSE detectors compare for memoryless channels. This is indeed the case as illustrated by the following computer experiment.

Experiment 5-1. For a system with $n = 2$ users, we consider two receivers: one with $m = 3$ sensors, and the other with $m = 10$. In Fig. 5-5, we plot MSE_1 versus $\text{SNR}_1 = \sum_{j=1}^m \|\mathbf{h}_j^{(1)}\|^2 / \sigma^2$, where $\mathbf{h}_j^{(1)}$ denotes the first column of the j -th channel tap. The curves are the ensemble average of 10000 4-tap channels ($M = 3$) of dimension $m \times 2$. The elements of the channel coefficients are selected independently from a

zero-mean, unit-variance complex Gaussian distribution, and the channel columns are then scaled so that user 1 has energy 10 dB below that of user 2. The curves compare the MSE performance of the zero-delay FLP and MMSE detectors, each with 7 taps ($N = \left\lceil \frac{Mn}{m-n} \right\rceil = 6$). The results illustrate that the performance of the zero-delay FLP detector is near to that of the zero-delay MMSE detector especially for high SNR and for very tall channels ($m \gg n$).

Observe that Fig. 5-5 is similar to Fig. 4-2 from Experiment 4-1. A close comparison of these figures reveals that, for a given SNR, the detectors for channels with memory have significantly higher MSE than those for channels without memory. The reason for this performance difference is that, for channels with memory, energy is distributed over multiple taps, and the temporal predictor essentially discards energy associated with taps \mathbf{H}_1 through \mathbf{H}_M . In Experiment 5-1, there are 4 channel taps, each having, on average, 1/4 of the total channel energy. Hence, the zero-delay FLP detector essentially discards 3/4 of the channel energy. This is true for the zero-delay MMSE detector as well, in light of Lemma 5-5. The performance of both detectors for user i is in fact highly dependent upon the fraction of energy associated with the i -th user that is contained in the zero-th tap:

$$\gamma_0^{(i)} = \frac{\|\mathbf{h}_0^{(i)}\|^2}{\sum_{j=0}^M \|\mathbf{h}_j^{(i)}\|^2}, \quad (5-39)$$

where $\mathbf{h}_j^{(i)}$ denotes the i -th column of the j -th channel tap. The next example demonstrates this dependence more clearly.

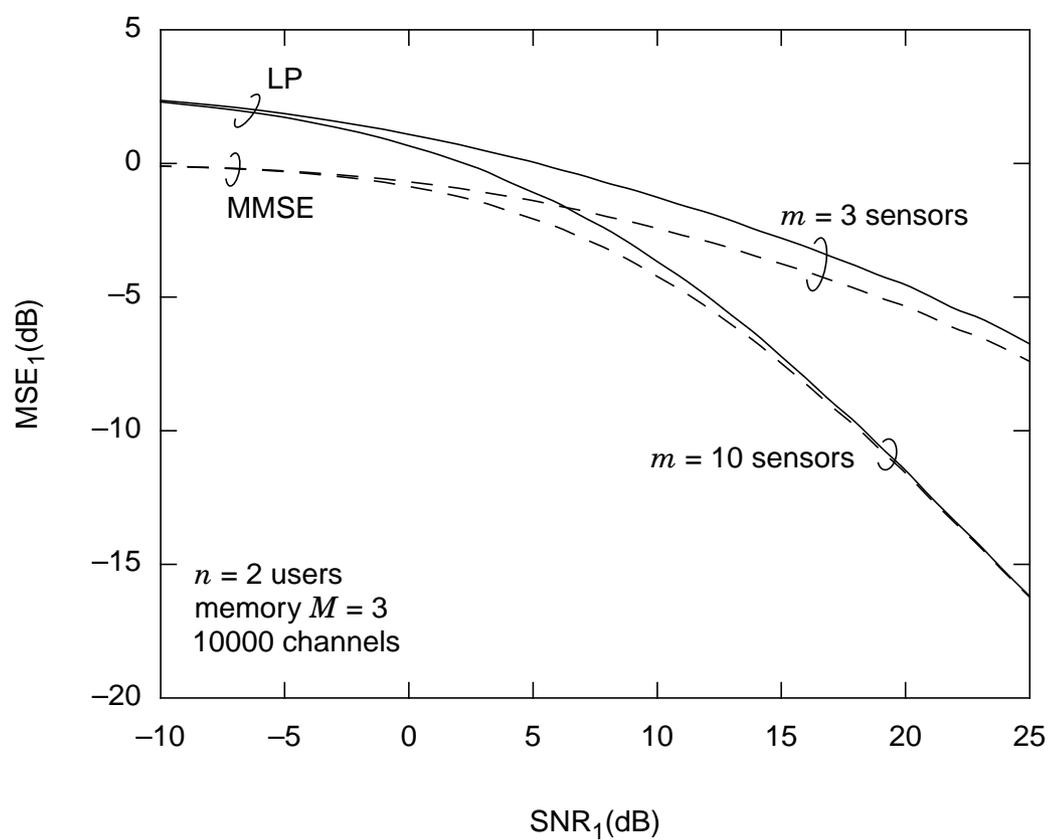


Figure 5-5. A comparison of the zero-delay FLP and MMSE detectors.

Experiment 5-2. In this experiment we analyze the performance of the zero-delay FLP detector as a function of $\gamma_0^{(1)}$, the fraction of user 1's energy contained in the zero-th tap. We consider 1000 random 3-tap ($M = 2$) channels of dimension 5×2 . (The coefficients are selected independently from a zero-mean unit-variance Gaussian distribution, and then $\mathbf{h}_0^{(1)}$ is scaled to control $\gamma_0^{(1)}$, while keeping the total energy associated with user 1 at unity.) The predictor order is $N = \left\lceil \frac{Mn}{m-n} \right\rceil = 2$. Fig. 5-6 plots MSE_1 , averaged over the 1000 trials, versus SNR_1 for different values of $\gamma_0^{(1)}$. Observe that for $\gamma_0^{(1)} \approx 1$, the slope of the curve is approximately -1 , but for $\gamma_0^{(1)} \approx 0$, the slope of the curve is approximately 0. Obviously, the performance is highly dependent upon the energy fraction.

It might seem logical to conclude that if \mathbf{H}_0 is vanishingly small, then the temporal predictor would discard essentially all of the channel energy. However, this is not the case provided that $\sigma^2 > 0$. Recall from Corollary 5-8.1 that the cascade $\mathbf{F}(z)$ of the temporal prediction-error filter $\mathbf{I} - \mathbf{P}(z)$ and the channel $\mathbf{H}(z)$ has coefficients given by $[\mathbf{H}_0 \quad \mathbf{K}\Psi]$, where $\mathbf{K} = [\mathbf{H}_1 \dots \mathbf{H}_M \mathbf{0}_{m \times Nn}]$, and $\Psi = \sigma^2(\mathcal{H}^* \mathcal{H} + \sigma^2 \mathbf{I})^{-1}$. The structure of Ψ has interesting implications for the behavior of the temporal predictor if the leading coefficient \mathbf{H}_0 is sufficiently small. We can express Ψ as follows:

$$\Psi = \sigma^2 \begin{bmatrix} \mathbf{H}_0^* \mathbf{H}_0 + \sigma^2 \mathbf{I} & \mathbf{H}_0^* \tilde{\mathbf{K}} \\ \tilde{\mathbf{K}}^* \mathbf{H}_0 & \tilde{\mathcal{H}}^* \tilde{\mathcal{H}} + \sigma^2 \mathbf{I} \end{bmatrix}^{-1}, \quad (5-40)$$

where $\tilde{\mathbf{K}} = [\mathbf{H}_1 \dots \mathbf{H}_M \mathbf{0}_{m \times n(N-1)}]$, and where $\tilde{\mathcal{H}}$ is an $m(N-1) \times n(M+N-1)$ block Toeplitz matrix given by

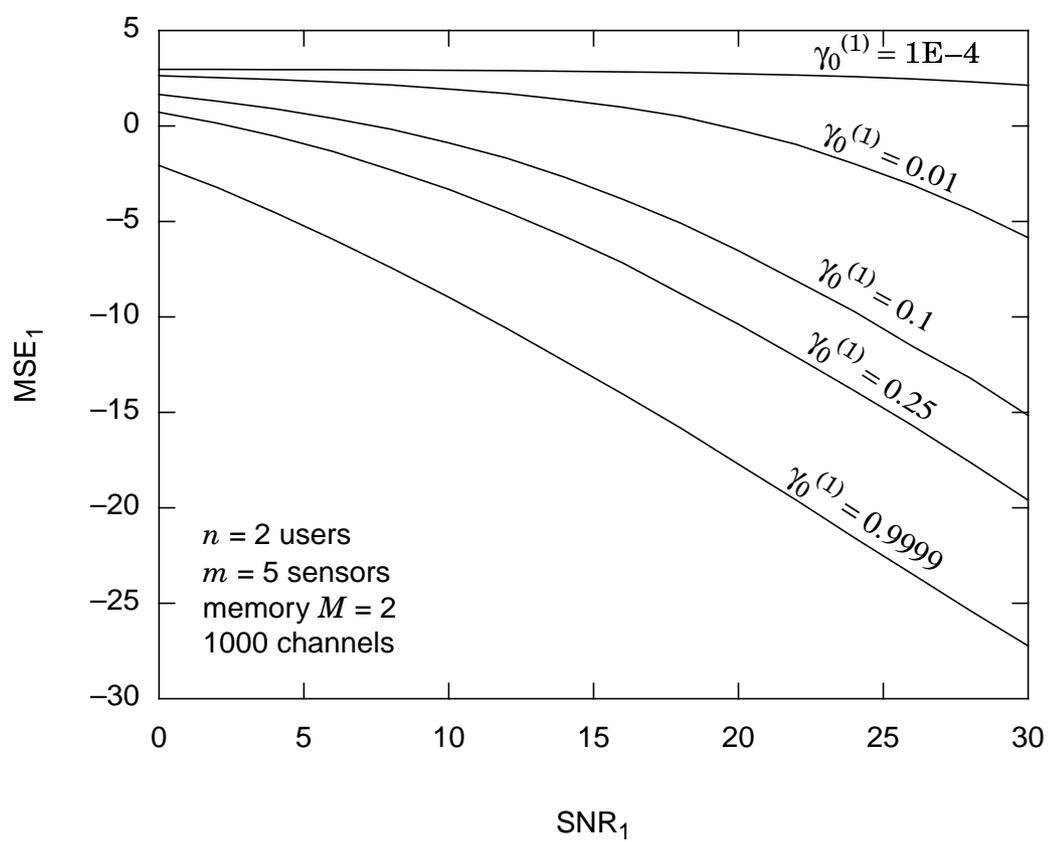


Figure 5-6. Performance of the zero-delay FLP detector as a function of $\gamma_0^{(1)}$, the fraction of user 1's energy in the zero-th tap.

$$\tilde{\mathcal{H}} = \begin{bmatrix} \mathbf{H}_1 & \mathbf{H}_2 & \dots & \mathbf{H}_M & \mathbf{0} & \dots & \mathbf{0} \\ \mathbf{0} & \mathbf{H}_1 & \mathbf{H}_2 & \dots & \mathbf{H}_M & \dots & \mathbf{0} \\ & & \ddots & \ddots & & \ddots & \\ \mathbf{0} & \dots & \mathbf{0} & \mathbf{H}_1 & \mathbf{H}_2 & \dots & \mathbf{H}_M \end{bmatrix}. \quad (5-41)$$

Taking the limit of Ψ as \mathbf{H}_0 goes to zero yields

$$\lim_{\mathbf{H}_0 \rightarrow \mathbf{0}} \Psi = \sigma^2 \begin{bmatrix} \sigma^2 \mathbf{I} & \mathbf{0} \\ \mathbf{0} & \tilde{\mathcal{H}}^* \tilde{\mathcal{H}} + \sigma^2 \mathbf{I} \end{bmatrix}^{-1} = \begin{bmatrix} \mathbf{I} & \mathbf{0} \\ \mathbf{0} & \sigma^2 (\tilde{\mathcal{H}}^* \tilde{\mathcal{H}} + \sigma^2 \mathbf{I})^{-1} \end{bmatrix} \equiv \begin{bmatrix} \mathbf{I}_{n \times n} & \mathbf{0} \\ \mathbf{0} & \tilde{\Psi} \end{bmatrix}. \quad (5-42)$$

Therefore, in the limit, the forward cascade $\mathbf{F}(z)$ becomes

$$\lim_{\mathbf{H}_0 \rightarrow \mathbf{0}} [\mathbf{H}_0 \ [\mathbf{H}_1 \ \dots \ \mathbf{H}_M \ \mathbf{0}_{m \times Nn}] \cdot \Psi] = [\mathbf{0}_{m \times n} \ \mathbf{H}_1 \ [\mathbf{H}_2 \ \dots \ \mathbf{H}_M \ \mathbf{0}_{m \times Nn}] \cdot \tilde{\Psi}], \quad (5-43)$$

and \mathbf{H}_1 , rather than \mathbf{H}_0 , plays the role of leading tap. From (5-40), we observe that if $\mathbf{H}_0 \mathbf{H}_0^*$ is negligible relative to $\sigma^2 \mathbf{I}$, then

$$[\mathbf{H}_0 \ [\mathbf{H}_1 \ \dots \ \mathbf{H}_M \ \mathbf{0}] \cdot \Psi] \approx [\mathbf{0} \ \mathbf{H}_1 \ [\mathbf{H}_2 \ \dots \ \mathbf{H}_M \ \mathbf{0}] \cdot \tilde{\Psi}]. \quad (5-44)$$

In other words, \mathbf{H}_0 is essentially ignored by the predictor if its energy is below the noise floor: $\text{tr}(\mathbf{H}_0 \mathbf{H}_0^*) \ll \sigma^2$; the predictor behaves as if the channel were $\tilde{\mathbf{H}}(z) \approx \sum_{j=1}^M \mathbf{H}_j z^{-j}$. For high SNR, *i.e.* $(\mathcal{H}^* \mathcal{H} + \sigma^2 \mathbf{I})^{-1} \mathcal{H}^* \approx \mathcal{H}^\dagger$, we have that $\mathbf{F}(z) \approx \mathbf{H}_1 z^{-1}$.

This behavior can be described more generally. First, it generalizes to *multiple* leading taps. If the energy in the L leading taps is below the noise floor, $\text{tr} \sum_{j=0}^{L-1} \mathbf{H}_j \mathbf{H}_j^* \ll \sigma^2$, then the predictor essentially ignores all L taps. In this case, the predictor behaves as if the channel were $\tilde{\mathbf{H}}(z) \approx \sum_{j=L}^M \mathbf{H}_j z^{-j}$ such that $\mathbf{F}(z) \approx \mathbf{H}_L z^{-L}$ for high SNR. Second, the behavior in response to each user is independent. If the energy in the i -th column of the L_i leading taps, corresponding to user i , is below the noise floor, $\sum_{j=0}^{L_i-1} \|\mathbf{h}_j^{(i)}\|^2 \ll \sigma^2$, then

the predictor ignores the i -th column of these taps. The following theorem formally states the result in its most general form.

Theorem 5-9. Let the channel $\mathbf{H}(z)$ of (1-1), with $m > n$ and $\sigma > 0$, satisfy (5-2), and let $\mathbf{h}_j^{(i)}$ denote the i -th column of the j -th channel tap \mathbf{H}_j . For each user i , let $L_i \in \{0, \dots, M\}$ denote the length of the precursor Γ_i , defined as $\Gamma_i = [\mathbf{h}_0^{(i)} \ \mathbf{h}_1^{(i)} \ \dots \ \mathbf{h}_{L_i}^{(i)}]$ for $L_i > 0$ or as $\Gamma_i = \mathbf{0}_{m \times 1}$ for $L_i = 0$. Let $\Gamma = [\Gamma_1 \ \dots \ \Gamma_n]$ denote the total precursor. Let $\mathbf{P}(z)$ be the optimal one-step temporal predictor of order $N \geq \lceil \frac{Mn}{m-n} \rceil$ for $\mathbf{H}(z)$, defined by (5-31). The following holds:

$$\lim_{\sigma \rightarrow 0^+} \lim_{\Gamma \rightarrow \mathbf{0}} [\mathbf{I} - \mathbf{P}(z)]\mathbf{H}(z) = \begin{bmatrix} \mathbf{h}_{L_1}^{(1)} & \mathbf{h}_{L_2}^{(2)} & \dots & \mathbf{h}_{L_n}^{(n)} \end{bmatrix} \text{diag} \left[z^{-L_1} \ z^{-L_2} \ \dots \ z^{-L_n} \right]. \quad (5-45)$$

Proof: See Appendix 5-6.

Theorem 5-9 can be roughly paraphrased as follows: *for each user, precursors with energies below the noise floor are ignored.* If the precursor energy of user i , $\gamma_{L_i}^{(i)} = \text{tr}(\Gamma_i \Gamma_i^*)$, is below the noise floor, $\gamma_{L_i}^{(i)} \ll \sigma^2$, then the predictor behaves as if the i -th column of the channel were $\mathbf{h}^{(i)}(z) \approx \sum_{j=L_i}^M \mathbf{h}_j^{(i)} z^{-j}$, and if user i has sufficiently high SNR, then the i -th column $\mathbf{f}^{(i)}(z)$ of $\mathbf{F}(z)$ approaches $\mathbf{h}_{L_i}^{(i)} z^{-L_i}$.

Consider now a channel for which the energy in \mathbf{H}_0 is below the noise floor, $\text{tr}(\mathbf{H}_0 \mathbf{H}_0^*) \ll \sigma^2$, but for which the SNR of all users is high, $(\mathcal{H}^* \mathcal{H} + \sigma^2 \mathbf{I})^{-1} \mathcal{H}^* \approx \mathcal{H}^\dagger$, such that $\mathbf{F}(z) \approx \mathbf{H}_1 z^{-1}$. For this channel the zero-delay LP detector $\mathbf{C}_{FLP0}(z)$, as defined by (5-36), performs poorly. It makes more sense in this case to define a delay-1 LP detector $\mathbf{C}_{FLP1}(z) = \mathbf{J} \mathbf{Q}_1 \mathbf{D}^{-1} \mathbf{M}^{-1} [\mathbf{I} - \mathbf{P}(z)]$, where all terms are defined as in (5-36) except \mathbf{Q}_1 ,

which is now chosen to minimize the delay-1 MSE sum $E[\|\mathbf{Q}\mathbf{w}_k - \mathbf{x}_{k-1}\|^2]$, where we recall that \mathbf{w}_k is the spatially whitened temporal prediction error $\mathbf{w}_k = \mathbf{W}\mathbf{e}_k = \mathbf{D}^{-1}\mathbf{M}^{-1}\mathbf{e}_k$. As before, \mathbf{Q}_1 can be derived from the delay-1 memoryless detector \mathbf{C}_1 given by (5-34). From Lemma 5-4, $\mathbf{C}_1 = \mathbf{F}_1^* \Phi_e^{-1} = \mathbf{F}_1^* \mathbf{W}^* \mathbf{W}$. Hence, \mathbf{Q}_1 satisfies $\mathbf{J}\mathbf{Q}_1 = \mathcal{P}^* (\mathbf{D}^{-1}\mathbf{M}^{-1}\mathbf{F}_1)$. In the following simulation experiments, we show that as the energy in \mathbf{H}_0 becomes vanishingly small, the delay-1 detector begins to outperform the zero-delay detector. Consider the following experiments.

Experiment 5-3. We essentially repeat Experiment 5-2 here, except we implement both the delay-1 and the delay-0 FLP detectors. (We consider 1000 random 3-tap channels of dimension 5×2 , generated as before; again, $N = 2$.) The results are shown in Fig. 5-7, where the solid curves are for the delay-1 detector, and the dashed curves are for the delay-0 detector. The curves clearly show regions in which the delay-1 detector outperforms the delay-0 detector, for instance, when $\gamma_0^{(1)} \leq 0.01$ and $\text{SNR}_1 \leq 24$ dB (see point A) or when $\gamma_0^{(1)} \leq 0.1$ and $\text{SNR}_1 \leq 10$ dB (see point B).

Experiment 5-4. In this experiment we fix the SNR of user 1 at 20 dB, and then vary the zero-th tap energy fraction $\gamma_0^{(1)}$. We consider 1000, 2-tap ($M = 1$) Gaussian channels of dimension 5×2 . The SNR of user 2 is set 10 dB greater than that of user 1, so $\text{SIR}_1 = -10$ dB. The predictor order is $N = \left\lceil \frac{Mn}{m-n} \right\rceil = 1$. Fig. 5-8 plots MSE_1 , averaged over the 1000 trials, versus SNR_1 for both the delay-0 and delay-1 FLP detectors. Observe that the curves intersect when $\gamma_0^{(1)} \approx \text{SNR}_1^{-1}$. To the right of this intersection, the delay-1 detector outperforms the delay-0 detector.

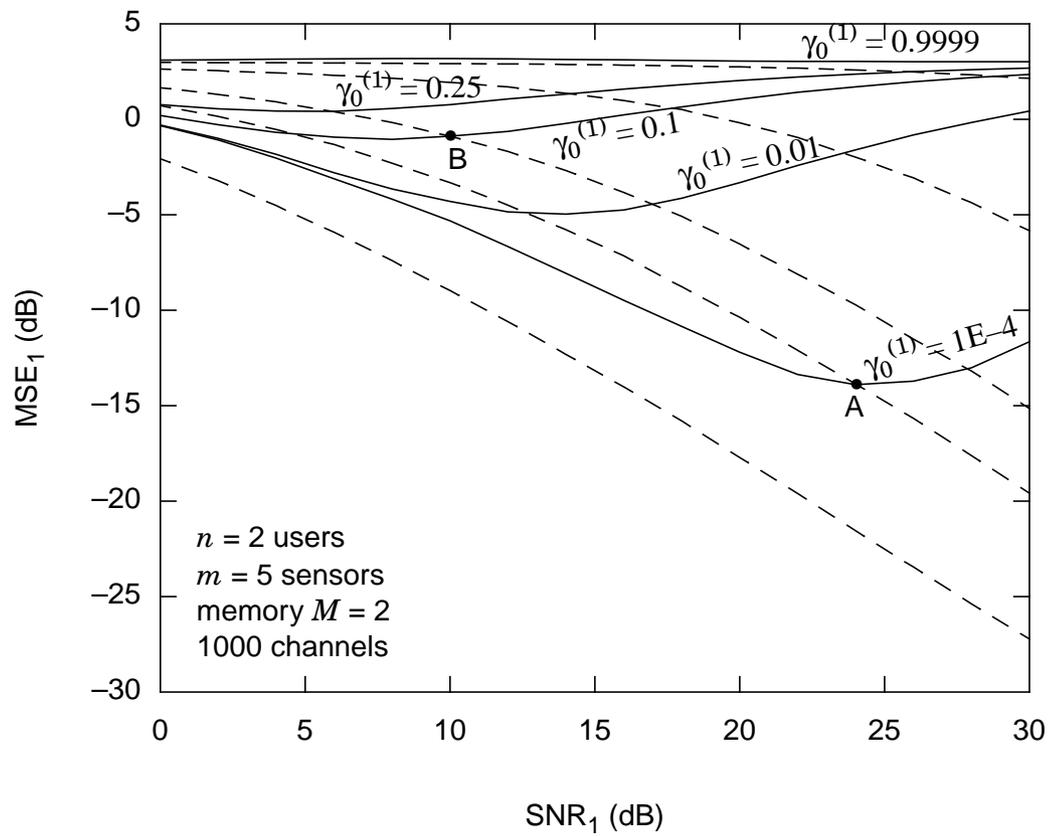


Figure 5-7. Comparison of the delay-1 and delay-0 performance as a function of the zero-th tap energy fraction $\gamma_0^{(1)}$.

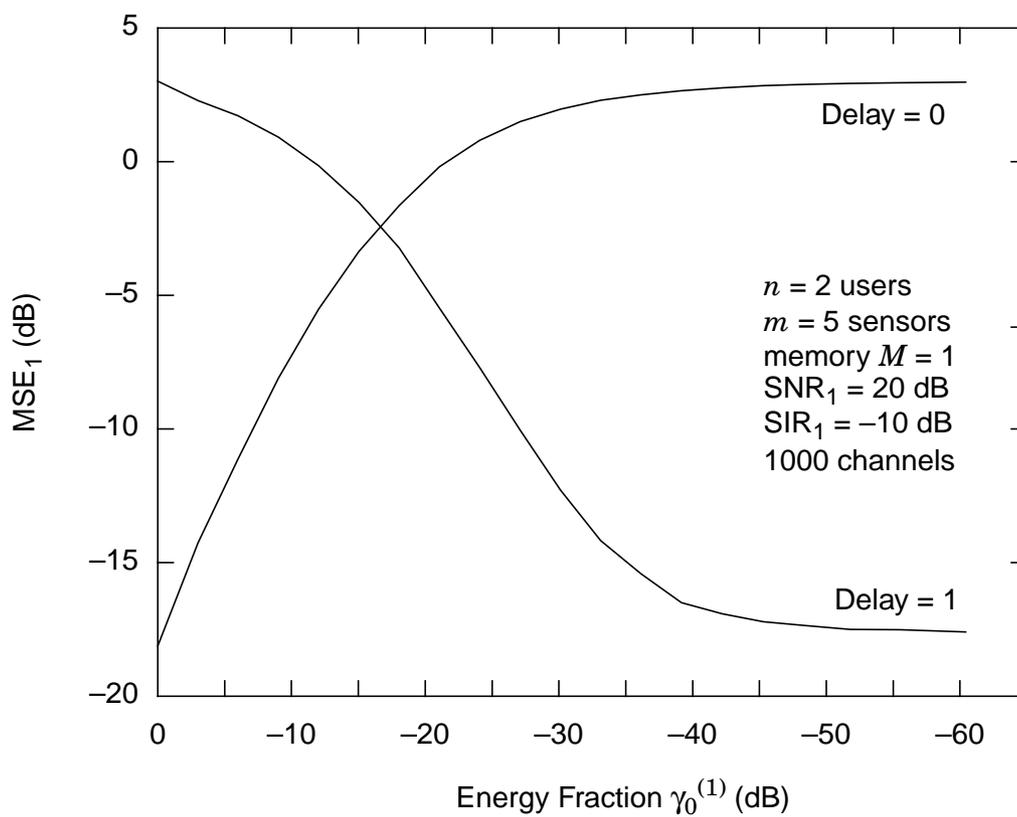


Figure 5-8. Comparison of the delay-0 and delay-1 FLP detectors as a function of the energy fraction $\gamma_0^{(1)}$ for 2-tap channels with fixed SNR.

These examples demonstrate that there exists an optimal delay D_i for each user i , which depends on the energy in the precursor taps corresponding to that user. So the unitary rotation \mathbf{Q} should be defined accordingly. We take $\tilde{\mathbf{x}}_k = [x_{k-D_1}^{(1)}, x_{k-D_2}^{(2)}, \dots, x_{k-D_n}^{(n)}]^T$ and define $\tilde{\mathbf{F}}$ as a $m \times n$ matrix whose i -th column is the i -th column of \mathbf{F}_{D_i} . (Recall that $\mathbf{F}(z) = [\mathbf{I} - \mathbf{P}(z)]\mathbf{H}(z)$.) The unitary matrix $\tilde{\mathbf{Q}}$ that minimizes the MSE sum $E[\|\mathbf{J}\mathbf{Q}\mathbf{w}_k - \tilde{\mathbf{x}}_k\|^2]$ satisfies

$$\mathbf{J}\tilde{\mathbf{Q}} = \mathcal{P}^*(\mathbf{D}^{-1}\mathbf{M}^{-1}\tilde{\mathbf{F}}). \quad (5-46)$$

The definition of the FLP detector with optimal delays is based on the optimal $\tilde{\mathbf{Q}}$.

Definition 5-6. Let the channel $\mathbf{H}(z)$ of (1-1) with $m > n$ and $\sigma > 0$ satisfy (5-2). The (optimal delay) **forward LP detector** of order $N \geq \left\lceil \frac{Mn}{m-n} \right\rceil$ for $\mathbf{H}(z)$ is uniquely defined as

$$\mathbf{C}_{FLP}(z) = \mathbf{J}\tilde{\mathbf{Q}}\mathbf{D}^{-1}\mathbf{M}^{-1}[\mathbf{I} - \mathbf{P}(z)], \quad (5-47)$$

where $\mathbf{P}(z)$ is given by (5-31), where $\mathbf{M}\mathbf{D}^2\mathbf{M}$ is the Cholesky factorization of \mathbf{R}_e , given by (5-33), and where $\tilde{\mathbf{Q}}$ satisfies (5-46).

In the adaptive implementation that follows, $\tilde{\mathbf{Q}}$ is implemented with a decision-directed MPLL. It implicitly finds the optimal delays associated with $\tilde{\mathbf{Q}}$, because these are the ones that produce the smallest slicer error.

Even with optimal delays, the FLP detector may not perform well if the energy in the zero-th tap $\gamma_0^{(1)}$ is small, yet also significant relative to the noise. For instance, in Example 5-4 the average MSE_1 is unacceptable for both the delay-0 and the delay-1 FLP detectors if $\gamma_0^{(1)}$ is near SNR_1^{-1} . One solution to this problem is to modify the linear predictor to

isolate, not the zero-th tap \mathbf{H}_0 , but the L -th tap \mathbf{H}_L , and to choose L such that it corresponds to the tap of greatest energy: $\text{tr}(\mathbf{H}_L \mathbf{H}_L^*) \geq \text{tr}(\mathbf{H}_j \mathbf{H}_j^*) \forall j \neq L$. This can be accomplished by extending the SIMO forward-backward prediction concept mentioned in [96] to MIMO channels. A generalized prediction-based detector based on this idea is presented in the next section.

5.5.2 The Forward-Backward LP Detector

We have shown that a noiseless FIR channel $\mathbf{H}(z) = \sum_{i=0}^M \mathbf{H}_i z^{-i}$ with memory M (5-1) can be equalized to a truncated version $\mathbf{H}^L(z) = \sum_{i=0}^M \mathbf{H}_i z^{-i}$ with memory L by using an $(L+1)$ -step forward linear predictor. We have also shown that an FIR channel $\mathbf{H}^L(z) = \sum_{i=0}^M \mathbf{H}_i z^{-i}$ with memory L can be equalized to its last tap $\mathbf{H}_L z^{-L}$ by using a one-step backward linear predictor. Clearly, the cascade of a $(L+1)$ -step forward predictor and a one-step backward predictor isolates the L -th tap \mathbf{H}_L of the original channel, so that the output of the backward prediction-error filter is $\mathbf{b}_k = \mathbf{H}_L \mathbf{x}_{k-L}$. Spatial whitening and rotation can then be used to invert \mathbf{H}_L .

The previous discussion suggests a generalized architecture for a prediction-based detector. The *forward-backward LP detector*, or simply the *LP detector*, consists first of an $(L+1)$ -step forward temporal prediction-error filter $\mathbf{I} - \mathbf{P}^L(z)$ of order $N \geq \lceil \frac{Mn}{m-n} \rceil$. The forward prediction error is given by $\mathbf{e}_k = \mathbf{r}_k - \sum_{i=L+1}^{L+N} \mathbf{P}_i \mathbf{r}_{k-i}$, where the coefficients $\{\mathbf{P}_i\}$ are chosen to minimize the mean-square prediction error $E[\|\mathbf{e}_k\|^2]$. The next stage is a one-step backward temporal error filter $[\mathbf{I} - \mathbf{P}^L(z)]z^{-\tilde{N}}$ of order $\tilde{N} \geq \lceil \frac{Ln}{m-n} \rceil$. (The delay term $z^{-\tilde{N}}$ ensures causality.) The backward prediction error is given by $\mathbf{b}_k = \mathbf{r}_{k-\tilde{N}} - \sum_{i=-1}^{-N} \mathbf{P}_i \mathbf{r}_{k-i-\tilde{N}}$, where the coefficients are chosen to minimize $E[\|\mathbf{b}_k\|^2]$. The subsequent stages, in order, are a memoryless spatial prediction-error filter $\mathbf{I} - \tilde{\mathbf{P}}$, a diag-

onal gain \mathbf{A} , a unitary rotation matrix \mathbf{Q} , and a truncation matrix $\mathbf{J} = [\mathbf{I} \ \mathbf{0}]$. These are chosen to spatially whiten and optimally rotate \mathbf{b}_k in a manner exactly like that for the forward LP detector.

We now work toward a formal definition of the generalized LP detector. We begin by solving for the optimal $(L+1)$ -step forward predictor coefficients in noise.

Theorem 5-10. Let the channel $\mathbf{H}(z)$ of (1-1) with $m > n$ and $\sigma > 0$ satisfy (5-2), and let $\mathbf{P}(z) = \sum_{i=L+1}^{L+N} \mathbf{P}_i z^{-i}$ denote a $(L+1)$ -step temporal predictor of order $N \geq \lceil \frac{Mn}{m-n} \rceil$, where $L \in \{0, 1, \dots, M\}$. The optimal coefficients $\{\mathbf{P}_i\}$, minimizing $E\left[\left\|\mathbf{r}_k - \sum_{i=L+1}^{L+N} \mathbf{P}_i \mathbf{r}_{k-i}\right\|^2\right]$, are given by

$$[\mathbf{P}_{L+1} \dots \mathbf{P}_{L+N}] = \mathbf{K}^L \mathcal{H}^* (\mathcal{H} \mathcal{H}^* + \sigma^2 \mathbf{I})^{-1} \quad (5-48)$$

$$= \mathbf{K}^L (\mathcal{H}^* \mathcal{H} + \sigma^2 \mathbf{I})^{-1} \mathcal{H}^*, \quad (5-49)$$

where $\mathbf{K}^L = [\mathbf{H}_{L+1} \dots \mathbf{H}_M \ \mathbf{0}_{m \times (L+N)n}]$.

Proof: Appendix 5-3.

Using Theorem 5-10 we can derive an expression for the forward cascade.

Corollary 5-10.1. The forward cascade $\mathbf{F}^L(z) = \sum_{i=0}^{\tilde{M}} \mathbf{F}_i z^{-i} = [\mathbf{I} - \mathbf{P}^L(z)] \mathbf{H}(z)$, with memory $\tilde{M} = M + L + N$, has coefficients $\{\mathbf{F}_i\}$ given by

$$\mathbf{F}^L = [\mathbf{F}_0 \ \mathbf{F}_1 \ \dots \ \mathbf{F}_{\tilde{M}}] = [\mathbf{H}_0 \ \dots \ \mathbf{H}_L \ [\mathbf{H}_{L+1} \ \dots \ \mathbf{H}_M \ \mathbf{0}_{m \times (N+L)n}]] \cdot \Psi, \quad (5-50)$$

where $\Psi = \sigma^2 (\mathcal{H}^* \mathcal{H} + \sigma^2 \mathbf{I})^{-1}$.

Proof: See Appendix 5-3.

Notice that the cascade of (5-50) is similar to that of (5-32); taps 0 though L are the same as those of the original channel, but taps $L+1$ through M are modified by Ψ . Therefore, Corollary 5-10.1 implies that the $(L+1)$ -step predictor and the one-step predictor exhibit the same behavior in response to vanishingly small leading taps.

The output of the forward cascade is the forward prediction error, given by

$$\mathbf{e}_k = \sum_{i=0}^{\tilde{M}} \mathbf{F}_i \mathbf{x}_{k-i} + \mathbf{v}_k, \quad (5-51)$$

where $\mathbf{v}_k = \mathbf{n}_k - \sum_{i=L+1}^{L+N} \mathbf{P}_i \mathbf{n}_{k-i}$. The backward predictor thus sees a channel similar to that of (1-1) except that the noise is colored. The optimal coefficients for the backward predictor, given the observation \mathbf{e}_k , can now be determined. First, we define the following block-Toeplitz matrices formed from the forward cascade and the forward prediction-error filter:

$$\mathcal{F}_{\tilde{N}} = \begin{bmatrix} \mathbf{F}_0 & \mathbf{F}_1 & \dots & \mathbf{F}_{\tilde{M}} & \mathbf{0} & \dots & \mathbf{0} \\ \mathbf{0} & \mathbf{F}_0 & \mathbf{F}_1 & \dots & \mathbf{F}_{\tilde{M}} & \dots & \mathbf{0} \\ \vdots & & & \ddots & & \ddots & \\ \mathbf{0} & \dots & \mathbf{0} & \mathbf{F}_0 & \mathbf{F}_1 & \dots & \mathbf{F}_{\tilde{M}} \end{bmatrix} \quad (5-52)$$

$$\mathcal{E}_{\tilde{N}} = \begin{bmatrix} \mathbf{I}_m & \mathbf{0}_{m \times mL} & -\mathbf{P}_{L+1} & \dots & -\mathbf{P}_{L+N} & \mathbf{0} & \dots & \mathbf{0} \\ \mathbf{0} & \mathbf{I} & \mathbf{0} & -\mathbf{P}_{L+1} & \dots & -\mathbf{P}_{L+N} & \dots & \mathbf{0} \\ \vdots & & & \ddots & & & \ddots & \vdots \\ \mathbf{0} & \dots & \mathbf{0} & \mathbf{I} & \mathbf{0} & \mathbf{P}_{L+1} & \dots & -\mathbf{P}_{L+N} \end{bmatrix}, \quad (5-53)$$

which have dimensions of $m\tilde{N} \times n(\tilde{N} + \tilde{M})$, and $m\tilde{N} \times m(\tilde{N} + N + L)$, respectively. The subscript, indicating the number of block rows, is suppressed when equal to \tilde{N} , so that

$\mathcal{F} = \mathcal{F}_{\tilde{N}}$ and $\mathcal{E} = \mathcal{E}_{\tilde{N}}$. We can now concisely express the optimal backward predictor coefficients as follows:

Theorem 5-11. For the system of (5-51), let $\mathbf{P}^{-1}(z) = \sum_{i=-\tilde{N}}^{-1} \mathbf{P}_i z^{-i}$ denote a one-step backward temporal predictor of order $\tilde{N} \geq \lceil \frac{Ln}{m-n} \rceil$. The coefficients $\{\mathbf{P}_i\}$ minimizing $E\left[\left\|\mathbf{e}_k - \sum_{i=-\tilde{N}}^{-1} \mathbf{P}_i \mathbf{e}_{k-i}\right\|^2\right]$ are uniquely given by

$$\mathbf{P}^{-1} = [\mathbf{P}_{-\tilde{N}} \dots \mathbf{P}_{-1}] = (\mathbf{K}_F \mathcal{F}^* + \sigma^2 \mathbf{K}_E \mathcal{E}^*) (\mathcal{F} \mathcal{F}^* + \sigma^2 \mathcal{E} \mathcal{E}^*)^{-1}, \quad (5-54)$$

where

$$\mathbf{K}_F = [\mathbf{0}_{m \times n \tilde{N}} \mathbf{F}_0 \dots \mathbf{F}_{\tilde{M}-1}] \quad (5-55)$$

$$\mathbf{K}_E = [\mathbf{0}_{m \times m \tilde{N}} \mathbf{I}_m \mathbf{0}_{m \times mL} \mathbf{-P}_{L+1} \dots \mathbf{-P}_{L+N-1}]. \quad (5-56)$$

Proof: See Appendix 5-7.

The *total temporal prediction-error filter* is given by

$$\mathbf{T}(z) = \sum_{i=0}^{N+\tilde{N}} \mathbf{T}_i z^{-i} = [\mathbf{I} - \mathbf{P}^{-1}(z)][\mathbf{I} - \mathbf{P}^L(z)]. \quad (5-57)$$

We refer to the cascade of $\mathbf{T}(z)$ and the channel $\mathbf{H}(z)$, as the *backward cascade*:

$$\mathbf{B}(z) = \sum_{i=0}^{\tilde{M}+\tilde{N}} \mathbf{B}_i z^{-i} = [\mathbf{I} - \mathbf{P}^{-1}(z)][\mathbf{I} - \mathbf{P}^L(z)]\mathbf{H}(z). \quad (5-58)$$

Unfortunately, the coefficients of $\mathbf{B}(z)$ cannot be expressed in terms of the channel coefficients as elegantly as those for $\mathbf{F}(z)$ in (5-32) or (5-50). Nevertheless, a closed-form solution is given by

$$\mathbf{B} = [\mathbf{B}_1 \mathbf{B}_2 \dots \mathbf{B}_{\tilde{M} + \tilde{N}}] = [-\mathbf{P}_{-\tilde{N}} \dots -\mathbf{P}_{-1} \mathbf{I}_m] \mathcal{F}_{\tilde{N}+1}. \quad (5-59)$$

Similarly, the coefficients of $\mathbf{T}(z)$ are given by

$$\mathbf{T} = [\mathbf{T}_1 \mathbf{T}_2 \dots \mathbf{T}_{N+\tilde{N}}] = [-\mathbf{P}_{-\tilde{N}} \dots -\mathbf{P}_{-1} \mathbf{I}_m] \mathcal{E}_{\tilde{N}+1}. \quad (5-60)$$

The covariance of the backward prediction error $\mathbf{b}_k = \mathbf{e}_k - \sum_{i=-\tilde{N}}^{-1} \mathbf{P}_i \mathbf{e}_{k-i}$ can then be expressed in terms of \mathbf{B} and \mathbf{T} as

$$\Phi_{\mathbf{b}} = E[\mathbf{b}_k \mathbf{b}_k^*] = \mathbf{B}\mathbf{B}^* + \sigma^2 \mathbf{T}\mathbf{T}^*. \quad (5-61)$$

All that remains is to specify the spatial whitener and the rotation matrix. Using the Cholesky factorization of $\Phi_{\mathbf{b}} = \mathbf{M}\mathbf{D}^2\mathbf{M}$, we can define a prediction-based whitener as $\mathbf{W} = \mathbf{D}^{-1}\mathbf{M}^{-1}$. We can then derive the optimal rotator as before. We define $\tilde{\mathbf{B}}$ such that its i -th column is that of \mathbf{B}_{D_i} , where D_i is the optimal delay for the i -th user. The optimal rotation $\tilde{\mathbf{Q}}$ then satisfies

$$\mathbf{J}\tilde{\mathbf{Q}} = \mathcal{P}^*(\mathbf{D}^{-1}\mathbf{M}^{-1}\tilde{\mathbf{B}}). \quad (5-62)$$

The generalized LP detector is formally defined as follows.

Definition 5-7. Let the channel $\mathbf{H}(z)$ of (1-1) with $m > n$ and $\sigma > 0$ satisfy (5-2). The (forward-backward) **LP detector** of index $L \in \{0, 1, \dots, M\}$ and order (N, \tilde{N}) , satisfying $N \geq \left\lceil \frac{Mn}{m-n} \right\rceil$ and $\tilde{N} \geq \left\lceil \frac{Ln}{m-n} \right\rceil$, is uniquely defined as

$$\mathbf{C}_{LP}(z) = \mathbf{J}\tilde{\mathbf{Q}}\mathbf{D}^{-1}\mathbf{M}^{-1}z^{-\tilde{N}} [\mathbf{I} - \mathbf{P}^{-1}(z)][\mathbf{I} - \mathbf{P}^L(z)], \quad (5-63)$$

where $\mathbf{P}^L(z)$ is given by (5-49), where $\mathbf{P}^{-1}(z)$ is given by (5-54), where $\mathbf{M}\mathbf{D}^2\mathbf{M}^*$ is a Cholesky factorization of $\Phi_{\mathbf{b}}$, given by (5-61), and where $\tilde{\mathbf{Q}}$ satisfies (5-62).

Observe that for the special case of $L = 0$, there is no backward predictor, and the LP detector reduces to the forward LP detector of Definition 5-6: $\mathbf{C}_{LP}(z) = \mathbf{C}_{FLP}(z)$. Similarly, for the special case of $L = M$, there is no forward predictor, and the LP detector reduces to the *backward LP detector*: $\mathbf{C}_{LP}(z) \equiv \mathbf{C}_{BLP}(z)$. This case is investigated in Appendix 5-8.

We now attempt to quantify the performance of the LP detector with some computer experiments. Recall from Experiment 5-4 that for 2-tap channels, the FLP detector (an LP detector with $L = 0$) produces an unacceptable MSE_i whenever the energy fraction $\gamma_0^{(i)}$ is close to $1/\text{SNR}_i$, regardless of the detector delay. The next experiment considers the same 2-tap channels, but applies an LP detector with a different index L , namely $L = 1$.

Experiment 5-5. We again consider 1000, 2-tap Gaussian channels of dimension 5×2 , as in Experiment 5-4. We vary the zero-th tap energy fraction $\gamma_0^{(1)}$ for SNRs fixed at $\text{SNR}_1 = 20$ dB and $\text{SNR}_2 = 30$ dB, so that $\text{SIR}_1 = -10$ dB. The predictor order is $N = \left\lceil \frac{Mn}{m-n} \right\rceil = 1$. Fig. 5-9, like Fig. 5-8, shows the average MSE_1 versus SNR_1 for the delay-0 and delay-1 FLP detectors, but it adds an additional curve for the LP detector with index $L = 1$ and delay $D = M + N = 2$. Observe that the worst case MSE_1 is less than -13 dB, occurring when $\gamma_0^{(1)} \approx -4$ dB. Regardless of $\gamma_0^{(1)}$, there always exists an LP detector with good performance.

The next experiment more generally quantifies the performance of the LP detector. It compares the performance of an LP detector with optimal index L and delay D to that of an MMSE detector with the same memory.

Experiment 5-6. We consider random 3-tap ($M = 2$) channels of dimension 6×2 . The elements of each tap are drawn independently from a complex Gaussian distri-

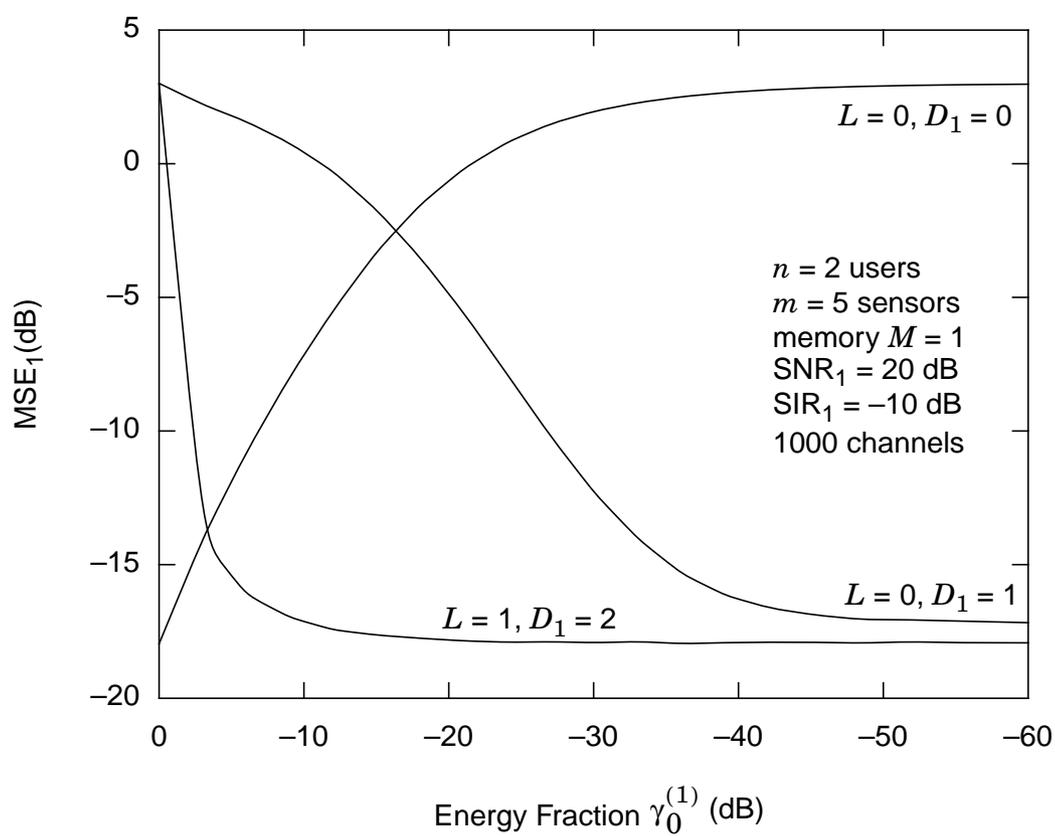


Figure 5-9. Comparison of several LP-based detectors as a function of the energy fraction $\gamma_0^{(1)}$ for 2-tap channels with fixed SNR.

bution and then the columns are scaled to control the SNR of user 1. The SNR of user 2 is always set 10 dB above that of user 1, so that $\text{SIR}_1 = -10$ dB. We consider 10000 channels at each SNR point. For each channel we implement the optimal minimum-order LP detector, *i.e.* considering all possible indices $L \in \{0, 1, M = 2\}$ and all possible delays $D \in \{0, 1, \dots, M+L+N\}$, and using $N = \tilde{N} = 1$ when the corresponding predictor exists. Therefore, counting the spatial predictor, the LP detector contains at most 3 taps. As a benchmark, we also implement the optimal 3-tap MMSE detector. Fig. 5-10 compares the performance of the best LP detector with that of the best MMSE detector. From the horizontal gap between the curves, we see that the LP detector suffers an SNR penalty relative to the MMSE detector. The size of this penalty in high SNR is roughly equivalent to the amount of energy discarded by the temporal predictors. If the energy were evenly distributed among the 3 channel taps, we would expect the LP detector to keep only about 33% of the channel energy. However, for this example, the largest tap on average contains 36% of the total channel energy, so the average penalty is $1 / 0.36 = 4.44$ dB.

Of course, the performance of the optimal LP detector for any particular channel depends on the distribution of channel energy. The performance is best if the channel memory is small or if a significant fraction of the channel energy is concentrated in a single tap.

We now briefly discuss some of the properties of the LP detector. We stress that these properties hold for any index $L \in \{0, 1, \dots, M\}$ and for any order (N, \tilde{N}) satisfying $N \geq \left\lceil \frac{Mn}{m-n} \right\rceil$ and $\tilde{N} \geq \left\lceil \frac{Ln}{m-n} \right\rceil$.

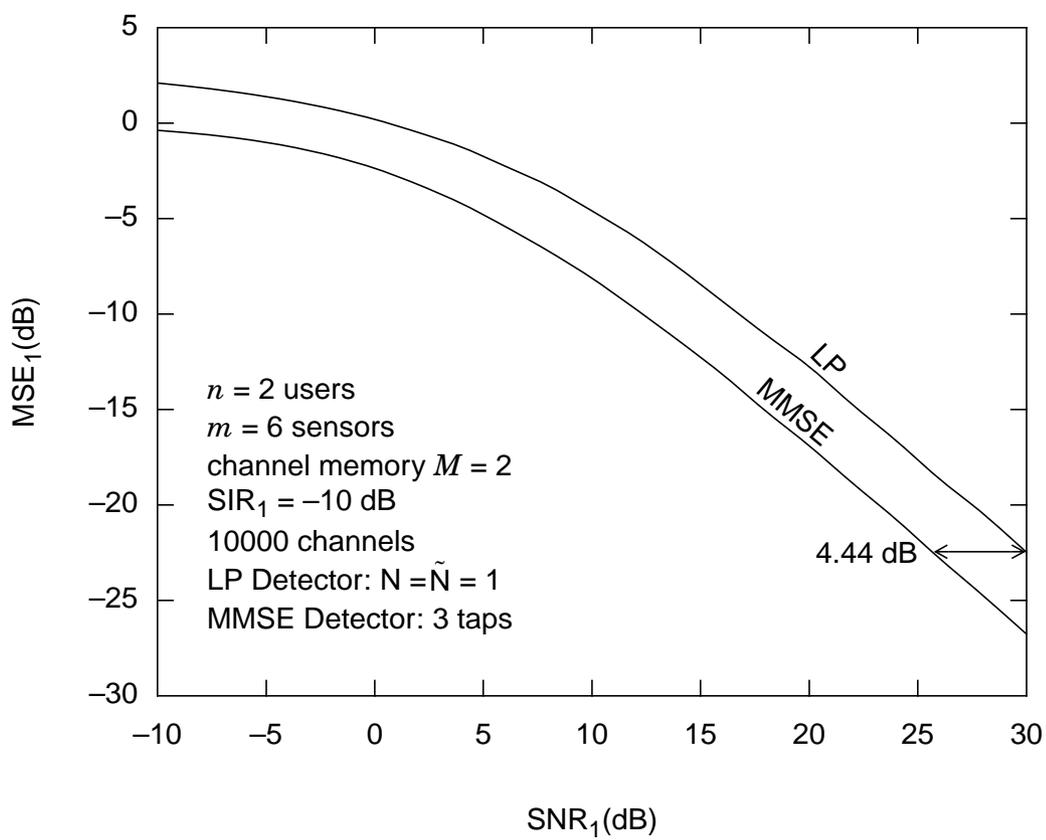


Figure 5-10. A comparison of the optimal LP detector with an MMSE detector with equal memory.

Property 5-1. The LP detector approaches a delayed zero-forcing detector in the limit as noise energy goes to zero.

$$\lim_{\sigma \rightarrow 0^+} \mathbf{C}_{LP}(z) = \mathbf{H}_L^\dagger z^{-L-\tilde{N}} [\mathbf{I} - \tilde{\mathbf{A}}^{-1}(z)][\mathbf{I} - \mathbf{A}^L(z)], \quad (5-64)$$

where $\mathbf{A}^L(z)$ is given by (5-26), and $\tilde{\mathbf{A}}^{-1}(z)$ is given by (5-13), substituting $\mathbf{H}^L(z) = \sum_{i=0}^L \mathbf{H}_i z^{-i}$ for $\mathbf{H}(z)$.

Property 5-1 can be argued without mathematical rigor by considering $\mathbf{C}_{LP}(z)$ for very small positive noise variance σ^2 . In this case, all predictors are approximately equal to their respective minimum-norm solutions. Specifically, (5-49) is approximately (5-26), and (5-54) is approximately (5-13) with $\mathbf{H}(z) = \mathbf{H}^L(z)$. It follows that $\Phi_{\mathbf{b}} \approx \mathbf{H}_L \mathbf{H}_L^*$ (5-61); $\Phi_{\mathbf{b}}$ remains full rank, although possibly poorly conditioned. Hence, \mathbf{M}^{-1} in (5-63) is approximately equal to the minimum-norm monic factor in a generalized Cholesky factorization (4-13) of $\mathbf{H}_L \mathbf{H}_L^*$. (\mathbf{M}^{-1} is well defined with or without noise.) For very small σ , the term \mathbf{D}^{-1} in (5-63) has exactly $m - n$ very large values, yet the product $\mathbf{JQD}^{-1}\mathbf{M}^{-1} \approx \mathbf{H}_L^\dagger$ is well behaved. We remark that $\mathbf{C}_{LP}(z)$ as defined in (5-63) does not exist for $\sigma = 0$ because \mathbf{D}^{-1} is undefined; however, the detector definition can be extended to the zero-noise case by simply substituting \mathbf{D}^\dagger for \mathbf{D}^{-1} .

Property 5-2. The LP detector is optimally near-far resistant.

Optimal near-far resistance is inherited from the zero-forcing detector.

Property 5-3. For $\sigma > 0$, the output of the LP detector is spatially white, but not temporally white. For $\sigma = 0$, the output is spatio-temporally white.

Property 5-3 follows from inspection of (5-61).

Property 5-4. The LP detector is information lossless up to the truncation matrix \mathbf{J} .

Property 5-4 follows because all detector operations, except the truncation matrix \mathbf{J} , are full column rank on the unit circle, and thus invertible. Although in general, inversion would require an infinite number of taps, this property does suggest that other blind detector architectures might be designed using linear prediction as a front end.

5.6 Blind Adaptive Implementations

A block diagram of a blind adaptive LP detector is shown in Fig. 5-11. The predictors are determined adaptively by minimizing their respective prediction errors. The first stage is an $(L+1)$ -step forward temporal predictor of order N : $\hat{\mathbf{P}}^L(z) = \sum_{i=L+1}^{L+N} \hat{\mathbf{P}}_i z^{-i}$. Let $\mathbf{R}_{k-L}^T = [\mathbf{r}_{k-L-1}^T \dots \mathbf{r}_{k-L-N}^T]$, of dimension $mN \times 1$, denote a stacked and delayed (by $L+1$) observation vector, and let $\hat{\mathbf{P}}^L(k) = [\hat{\mathbf{P}}_{L+1} \dots \hat{\mathbf{P}}_{L+N}]$, of dimension $m \times mN$, denote the matrix of prediction coefficients. (We denote dependence on time k parenthetically for the coefficients to avoid confusion with coefficient indices.) The coefficients can be adapted to minimize the variance of the prediction error $\mathbf{e}_k = \mathbf{r}_k - \hat{\mathbf{P}}^L(k)\mathbf{R}_k$ as follows:

$$\hat{\mathbf{P}}^L(k+1) = \hat{\mathbf{P}}^L(k) + \mu_f \mathbf{e}_k \mathbf{R}_k^* . \quad (5-65)$$

Because the update is derived from a cost that is quadratic in $\hat{\mathbf{P}}^L$, convergence to $\hat{\mathbf{P}}^L = [\mathbf{P}_{L+1} \dots \mathbf{P}_{L+N}]$, as defined in (5-49), is guaranteed for sufficiently small step size μ_f .

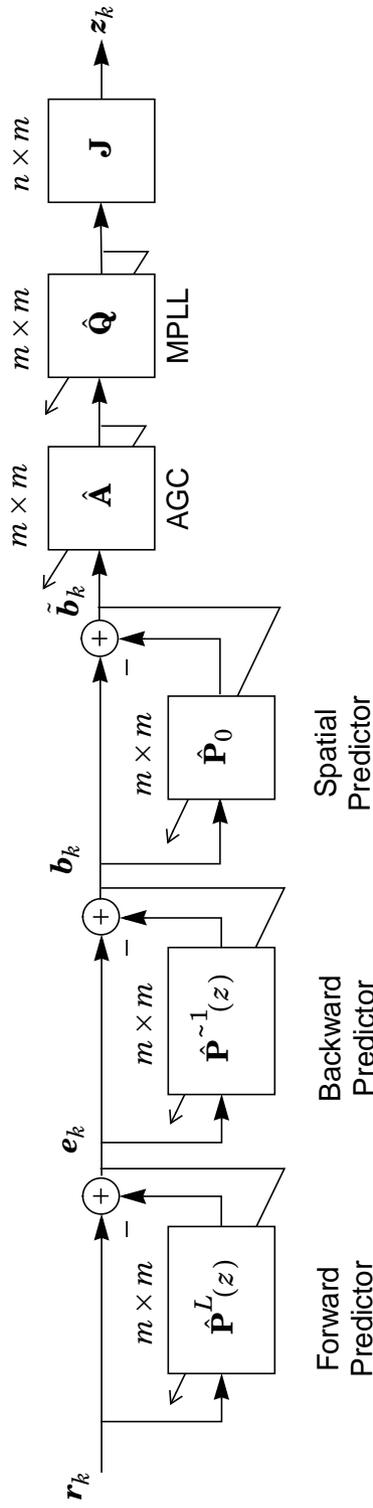


Figure 5-11. A Blind Adaptive Implementation of the LP Detector.

The second stage is a delayed one-step backward temporal predictor of order N : $\hat{\mathbf{P}}^{\sim 1}(z)z^{-N} = \sum_{i=-N}^{-1} \hat{\mathbf{P}}_i z^{-i-N}$. Observe that $\hat{\mathbf{P}}^{\sim 1}(z)z^{-N}$ is causal. Let $\mathbf{E}_k^T = [\mathbf{e}_{k-1}^T \dots \mathbf{e}_{k-\tilde{N}}^T]$ of dimension $mN \times 1$ denote a stacked and delayed (by 1) version of the forward prediction error, and let $\hat{\mathbf{P}}^{\sim 1}(k) = [\hat{\mathbf{P}}_{-\tilde{N}} \dots \hat{\mathbf{P}}_{-1}]$, of dimension $m \times mN$, denote a matrix of backward predictor coefficients. The coefficients can be adapted to minimize the variance of the backward prediction error $\mathbf{b}_k = \mathbf{e}_k - \hat{\mathbf{P}}^{\sim 1}(k)\mathbf{E}_k$ as follows:

$$\hat{\mathbf{P}}^{\sim 1}(k+1) = \hat{\mathbf{P}}^{\sim 1}(k) + \mu_b \mathbf{b}_k \mathbf{E}_k^* . \quad (5-66)$$

After convergence of the forward predictor $\hat{\mathbf{P}}^L$, convergence of $\hat{\mathbf{P}}^{\sim 1}$ to $[\mathbf{P}_{-\tilde{N}} \dots \mathbf{P}_{-1}]$, as defined in (5-54), is guaranteed for a sufficiently small step size μ_b .

The final adaptive stages of the detector are designed to spatially whiten and optimally rotate the backward prediction error \mathbf{b}_k . Let $\hat{\mathbf{P}}_0$ denote a strictly lower-triangular spatial predictor. This spatial predictor can be adapted by using a constrained LMS algorithm designed to minimize the variance of the spatial prediction error $\tilde{\mathbf{b}}_k = \mathbf{b}_k - \hat{\mathbf{P}}_0 \mathbf{b}_k$ as follows:

$$\hat{\mathbf{P}}_0(k+1) = \hat{\mathbf{P}}_0(k) + \mu_s \tilde{\mathbf{b}}_k \mathbf{b}_k^* . \quad (5-67)$$

$$\hat{\mathbf{P}}_0(k+1) = \hat{\mathbf{P}}_0(k+1) \otimes \mathbf{L}, \quad (5-68)$$

where, as in chapter 4, ' \otimes ' denotes a component-wise product, and \mathbf{L} is an $m \times m$ mask that properly constrains $\hat{\mathbf{P}}_0$. If preceding stages have converged, then $\hat{\mathbf{P}}_0$ converges to $\mathbf{I} - \mathbf{M}^{-1}$ for a sufficiently small step size μ_s , where \mathbf{M} is a valid factor in the Cholesky factorization of $\Phi_{\mathbf{b}} = E[\mathbf{b}_k \mathbf{b}_k^*]$.

The last two adaptive stages, namely the AGC and rotator, can be implemented exactly as explained in chapter 4. See section 4.3.1 for details regarding the AGC, and section 4.3.2 for a discussion of the adaptive rotator.

There are several variations of this basic implementation outlined above. For example, if $L = 0$, then the backward predictor does not exist. The update of (5-66) becomes unnecessary, and we can replace (5-67) with the following:

$$\hat{\mathbf{P}}_0(k+1) = \hat{\mathbf{P}}_0(k) + \mu_s \tilde{\mathbf{e}}_k \mathbf{e}_k^*, \quad (5-69)$$

where $\tilde{\mathbf{e}}_k = \mathbf{e}_k - \hat{\mathbf{P}}_0 \mathbf{e}_k$. Similarly, if $L = M$, there is no forward predictor. The update of (5-65) becomes unnecessary, and we can replace (5-66) with the following:

$$\hat{\mathbf{P}}^{\sim 1}(k+1) = \hat{\mathbf{P}}^{\sim 1}(k) + \mu_b \mathbf{b}_k \mathbf{R}_k^*, \quad (5-70)$$

where $\mathbf{R}_k^T = [\mathbf{r}_{k-1}^T \dots \mathbf{r}_{k-N}^T]$, and $\mathbf{b}_k = \mathbf{r}_k - \hat{\mathbf{P}}^{\sim 1}(k) \mathbf{R}_k$. As another variation, we might choose to combine the implementation of the spatial predictor with the backward temporal predictor (or forward temporal predictor if $L = 0$). For example, we could define a backward spatio-temporal predictor as $\hat{\mathbf{P}}^{\sim 1}(z) z^{-N} = \sum_{i=0}^{-N} \hat{\mathbf{P}}_i z^{-i-N}$, where $\hat{\mathbf{P}}_0$ is strictly lower triangular. Let $\tilde{\mathbf{E}}_k^T = [\mathbf{e}_k^T \dots \mathbf{e}_{k-N}^T]$, of dimension $m(N+1) \times 1$, denote a stacked version of the forward spatio-temporal prediction error. (Note that $\tilde{\mathbf{E}}_k$ differs from \mathbf{E}_k in (5-66) by inclusion of the term \mathbf{e}_k .) Let $\hat{\mathbf{P}}^{\sim 1}(k) = [\hat{\mathbf{P}}_{-N} \dots \hat{\mathbf{P}}_{-1} \hat{\mathbf{P}}_0]$, of dimension $m \times m(N+1)$, denote a matrix of prediction coefficients. The coefficients can be adapted to minimize the variance of the spatio-temporal prediction error $\tilde{\mathbf{b}}_k = \mathbf{e}_k - \hat{\mathbf{P}}^{\sim 1}(k) \tilde{\mathbf{E}}_k$ as follows:

$$\hat{\mathbf{P}}^{\sim 1}(k+1) = \hat{\mathbf{P}}^{\sim 1}(k) + \mu_b \tilde{\mathbf{b}}_k \tilde{\mathbf{E}}_k^*. \quad (5-71)$$

$$\hat{\mathbf{P}}_0(k+1) = \hat{\mathbf{P}}_0(k+1) \otimes \mathbf{L}. \quad (5-72)$$

The update of (5-71) thus combines (5-66) and (5-67). If the step size μ_b is sufficiently small, and the forward predictor has properly converged, then $\hat{\mathbf{P}}^{-1}$ is guaranteed to converge to

$$\hat{\mathbf{P}}^{-1} = [\tilde{\mathbf{P}}_{-\tilde{N}} \ \dots \ \tilde{\mathbf{P}}_{-1} \ \tilde{\mathbf{P}}_0] = [\mathbf{M}^{-1} \cdot [\mathbf{P}_{-\tilde{N}} \ \dots \ \mathbf{P}_{-1}] \ | \ (\mathbf{I} - \mathbf{M}^{-1})], \quad (5-73)$$

where \mathbf{M} is a term in a Cholesky factorization of $\Phi_{\mathbf{b}}$ (5-61), and where $[\mathbf{P}_{-\tilde{N}} \ \dots \ \mathbf{P}_{-1}]$ is defined by (5-54). In terms of numerical complexity, the combined spatio-temporal update is virtually the same as separate temporal and spatial updates; however, in terms of program complexity, the combined approach may have an advantage. The advantage of separate updates is that the step sizes μ_b and μ_s in (5-66) and (5-67) can be chosen independently to optimize convergence speed and misadjustment.

5.7 Experimental Results

We now consider additional computer experiments. The first of these compares a fractionally spaced single-user LP-based equalizer (*i.e.*, an LP detector with $n = 1$) with the well-known fractionally spaced constant-modulus algorithm (FS-CMA) [138]. The experiment clearly illustrates two of the primary advantages of basing blind algorithms primarily on second-order statistics —fast convergence and an inherent compatibility with shaped constellations. Recall from the discussion in chapter 2 that classical blind equalization techniques, *i.e.* those based on HOS, are unable to cope with shaped constellations. Unshaped systems transmit all symbols in the alphabet with equal probability; the probability distribution is uniform. In contrast, shaped systems favor symbols that are closer to

the origin; the distribution is nearer to Gaussian. A blind equalizer based on HOS is bound to fail for a Gaussian input distribution, because such distributions are completely characterized by their first and second-order statistics. For example, Leblanc *et al.* [139] have shown that CMA deteriorates as the kurtosis of the channel input approaches that of a Gaussian distribution. And yet, a near-Gaussian distribution is a prerequisite to achieving capacity on an AWGN channel. The following experiment demonstrates this false dichotomy.

Experiment 5-7. LP versus CMA. This experiment compares a $T/2$ -spaced CMA equalizer to a 1×2 LP detector (equalizer) for a system using shaped 64-QAM. The shaping is implemented by quantizing a complex Gaussian random variable to its nearest constellation point.¹⁸ By so doing, points closer to the edge are used more often, while points near the origin are used less often. The degree of shaping is controlled by the variance of the Gaussian random variable, and it is quantified by the kurtosis $\kappa = \frac{E[|x_k|^4]}{E^2[|x_k|^2]}$ of the channel input x_k . For an unshaped 64-QAM constellation, $\kappa = 1.381$, while for a 64-QAM constellation shaped according to the above procedure, κ is larger. The kurtosis of a complex Gaussian random process is $\kappa = 2$. The constellation is scaled to have unit energy so that the modulus parameter used in the CMA update is equal to the kurtosis. We decimate the output of the FS-CMA equalizer with a baud-rate sampler, and then append a first-order PLL to recover the complex scalar $e^{j\theta}$ left unresolved by CMA. (An analogous 2-dimensional PLL, followed by a 1×2 truncation matrix \mathbf{J} , is used as an integral part of the LP equalizer.)

18. Strictly speaking, the complex plane is tiled with 64-QAM constellations, and points chosen from the secondary constellations are mapped isomorphically back to the primary constellation.

The LP index is chosen to be $L = 0$, so there is no backward predictor. Each equalizer spans 4 baud. The channel is randomly generated, virtually noiseless (80 dB SNR), and also spans 4 baud.

We conduct three trials with varying degrees of shaping. For the first trial, there is no shaping; all symbols from the 64-QAM constellation are equally likely, as depicted by the histogram¹⁹ in Fig. 5-12(a). The kurtosis of this unshaped constellation is experimentally determined to be $\kappa = 1.386$ (which is close to the theoretical value of 1.381). Also shown in Fig. 5-12(a) are the outputs of both equalizers at 2500, 5000, and 10000 baud. The upper row corresponds to the LP equalizer, while the lower row corresponds to the CMA equalizer. We see that both equalizers are effective for the unshaped trial, although the LP equalizer converges slightly faster.

The second trial uses moderate shaping with a measured kurtosis of $\kappa = 1.595$. We see in Fig. 5-12(b) that the convergence rate of the LP equalizer is virtually unchanged, while that of the CMA equalizer slows substantially.

The constellation is heavily shaped for the third trial. The kurtosis is measured to be $\kappa = 1.857$, which is close to $\kappa = 2$, the kurtosis of a complex Gaussian random variable.²⁰ In Fig. 5-12(c) we see that the CMA equalizer fails completely even after 10000 baud. In contrast, the LP equalizer is still easily able to recover the shaped constellation. The experiment thus demonstrates the benefits of the LP equalizer and, in particular, the benefits of minimal reliance on HOS.

19. The histograms in this experiment are for the real part of the symbols only. By symmetry, the histogram of the imaginary part of the symbols is virtually identical.

20. In the language of Leblanc *et al.* [137], a distribution with kurtosis equal to that of a Gaussian distribution is *mesokurtic*.

We remark that, for each trial, the step sizes in all updates are optimized for convergence speed. For the LP equalizer, the step sizes are the same across all trials: $\mu_f = 0.5(2^{-k/500})$ and $\mu_s = 0.2(1 + k/600)$ for the temporal and spatial predictors, respectively, $\mu_a = 0.06(2^{-k/700})$ for the AGC, and $\lambda = 0.5(1 + k/800)$ for the rotator. In contrast, for the CMA equalizer, the initial step sizes are decreased for increasing kurtosis in order to maintain stability. We use $\mu = 0.09(2^{-k/700})$, $\mu = 0.08(2^{-k/1200})$, and $\mu = 0.05(2^{-k/1200})$ for trials 1, 2, and 3, respectively.

The previous experiment demonstrates the benefits of statistical decoupling, that is, decoupling use of second and higher-order statistics, and using HOS minimally only at the last step. The LP equalizer decomposes the equalization task into two steps: the first uses only SOS and thus is not affected by the Gaussianity of the transmitted symbols, whereas the second uses HOS implicitly, by relying on knowledge of the finite alphabet. Strictly speaking, the LP equalizer does use HOS, but only at the last step, and only to resolve a memoryless unitary ambiguity. SOS are enough to perform the initial and more demanding task of eliminating the channel memory, and it is this initial task that is oblivious to the Gaussianity of the symbols.

The next experiment demonstrates the effectiveness of the LP detector for an asynchronous multiuser CDMA system. The experiment is a generalization of Experiment 4-4 to channels with memory.

Experiment 5-8. Asynchronous CDMA. We now consider the 2-user asynchronous CDMA system illustrated in Fig. 5-13. Both users transmit QPSK. The spreading code for the i -th user is given by $f_i(t) = \sum_{j=0}^{m-1} c_j^{(i)} p(t - jT/m)$, where the

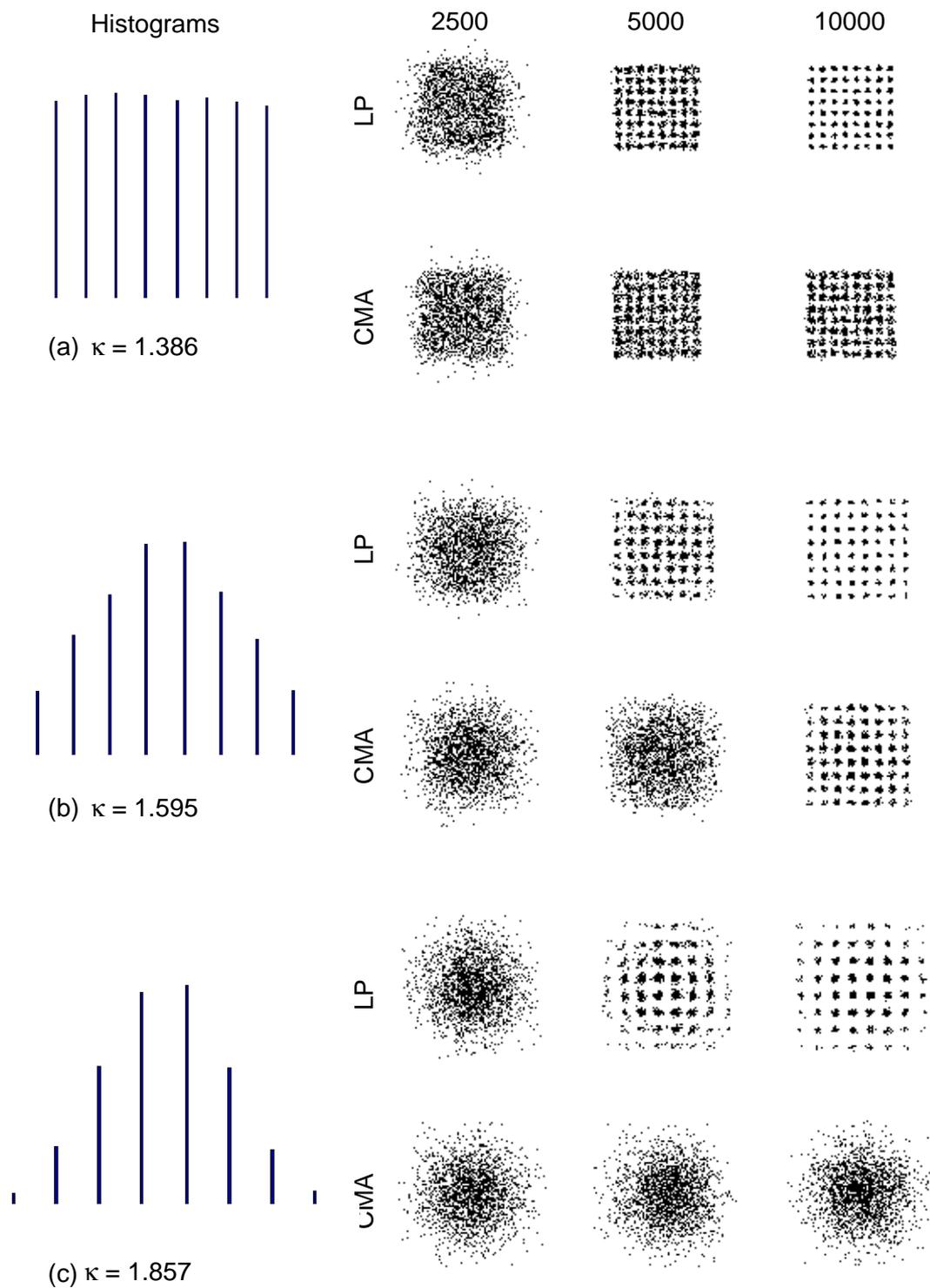


Figure 5-12. Convergence of LP versus CMA: (a) unshaped 64-QAM; (b) moderately shaped; and (c) heavily shaped.

chip-pulse shape $p(t) = \frac{\sin(\pi t m / T)}{(\pi t m / T)}$ is an ideal sinc function with bandwidth equal to half the chip rate m / T . In terms of the chip-rate PAM sequence, this $p(t)$ is the ideal zero-excess-bandwidth pulse shape. The chip sequences $\{c_j^{(i)}, j = 0, \dots, m - 1\}$ are of length $m = 16$ with period equal to one baud interval T . They are randomly generated:

$$\begin{aligned} \{c_j^{(1)}\} &= \{ +1 \ -1 \ +1 \ +1 \ -1 \ -1 \ -1 \ -1 \ -1 \ +1 \ +1 \ +1 \ -1 \ -1 \ -1 \ -1 \}, \\ \{c_j^{(2)}\} &= \{ +1 \ +1 \ +1 \ -1 \ +1 \ +1 \ +1 \ -1 \ +1 \ -1 \ -1 \ -1 \ -1 \ +1 \ -1 \ +1 \}. \end{aligned} \quad (5-74)$$

Their normalized correlation is $\rho = -3/8 = -0.375$. Both CDMA signals are subject to severe channel dispersion modeled by a first-order low-pass filter $h(t)$ with 3-dB bandwidth equal to one-fourth the chip rate, $W = 1/4T_c$. The signals are delayed by $\tau_1 = 0.7 T_c$ and $\tau_2 = 6.2 T_c$, respectively, where $T_c = T/m$ is the chip duration. The receiver front end consists of an anti-aliasing filter, followed by a chip-rate sampler, and a S/P converter. Because the transmitters have zero excess bandwidth, the front-end low-pass filter is identical to the chip-pulse shape $p(t)$. The chip-rate samples are grouped in blocks of $m = 16$ to generate the baud-rate sequence \mathbf{r}_k . The equivalent transfer function $\mathbf{H}(z)$ has dimension 16×2 . The amplitudes A_1 and A_2 and the noise variance σ^2 are selected such that $\text{SNR}_1 = 40$ dB and $\text{SNR}_2 = 35$ dB.

We implement the blind adaptive LP detector with index $L = 0$ and order $N = 1$ using $\mu_f = 0.5(2^{-k/500})$ for the forward predictor, $\mu_s = 0.05(2^{-k/600})$ for the spatial predictor, $\mu_\alpha = 0.1(1 + k/1000)$ for the AGC, and $\lambda = 0.8(2^{-k/700})$ for the rotator. Fig. 5-14 shows the recovered constellations at steady state (18000 to 20000 baud).

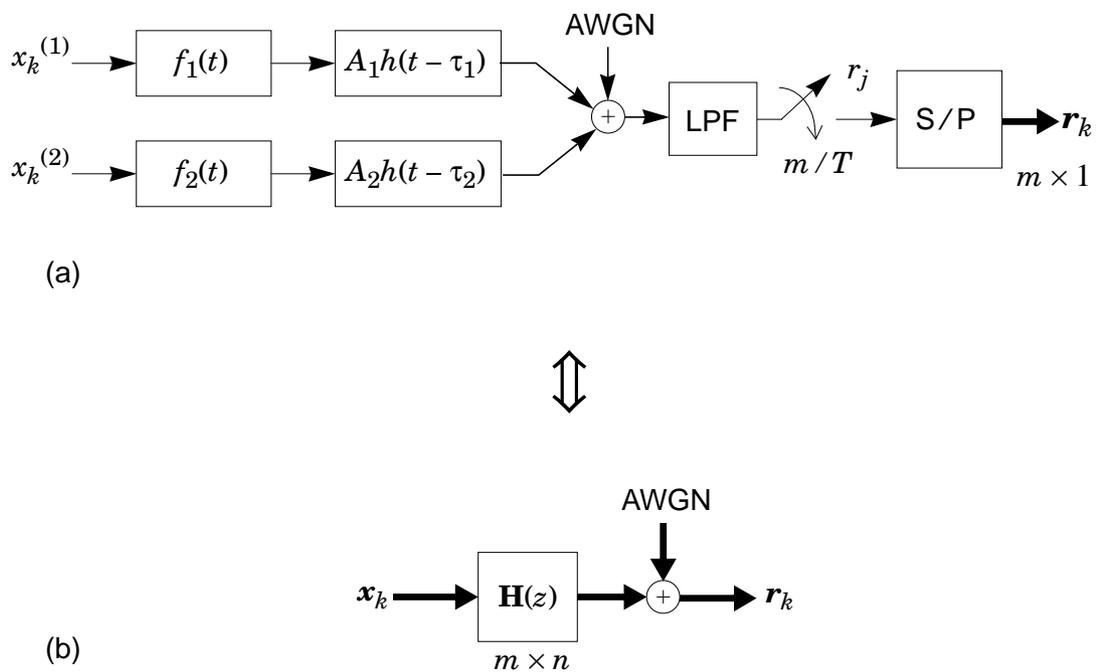


Figure 5-13. (a) A two-user asynchronous CDMA system with a chip-rate sampling receiver; (b) An equivalent MIMO FIR channel.

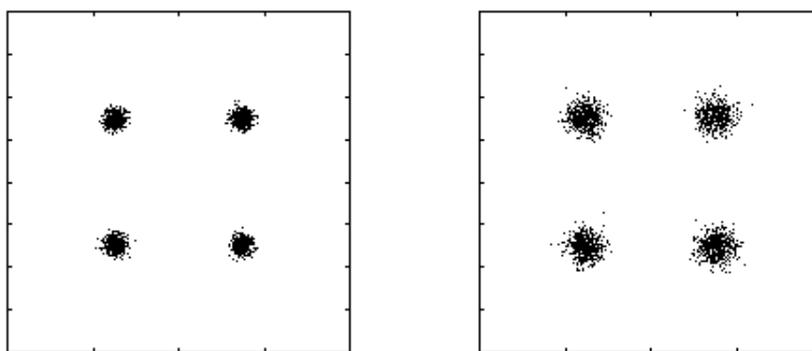


Figure 5-14. Recovered constellations for the asynchronous CDMA system of Experiment 5-8.

5.8 Chapter Summary

We have shown that, under mild assumptions, there exist many equivalent representations of tall FIR channels, *e.g.*, moving average, autoregressive, and others. The existence of a finitely parameterized AR model, in particular, implies that tall channels have an FIR, and therefore stable, left-inverse; thus, in this sense, they are minimum phase. Hence, second-order statistics are sufficient to identify or equalize a tall FIR channel up to an ambiguous unitary ambiguity. The minimum-phase property also immediately suggests linear prediction as an effective way to eliminate or to nearly eliminate channel memory. Using this idea, we have presented a family of blind multiuser detectors that exploit the special properties of tall FIR channels.

The forward LP detector of Definition 5-6 exploits the AR channel model of section 5.1.1. The first stage of the detector is a one-step forward prediction-error filter. Without noise, such a filter effectively converts the channel $\mathbf{H}(z)$ of (1-1) into a memoryless channel \mathbf{H}_0 . Although with noise, this conversion is only approximate, the spatial methods of chapter 4 can then nevertheless be applied to recover the transmitted sequence \mathbf{x}_k . The blind adaptive implementation we have proposed has relatively low complexity, fast convergence, and an inherent insensitivity to source distributions. With a zero-delay constraint, the FLP detector has performance near that of the MMSE detector. However its performance is highly dependent upon the energy in the zero-th tap \mathbf{H}_0 , assuming that it is significant relative to noise.

The generalized (forward-backward) LP detector of Definition 5-7 exploits the ARMA and AP models of sections 5.1.2 and 5.1.3, respectively. The first stage of the LP detector is an $(L+1)$ -step forward prediction-error filter, which roughly converts the

channel of $\mathbf{H}(z) = \mathbf{H}_0 + \dots + \mathbf{H}_M z^{-M}$ of (1-1), which has memory M , into $\mathbf{H}^L(z) = \mathbf{H}_0 + \dots + \mathbf{H}_L z^{-L}$, which has memory L . The second stage of the detector is a one-step backward prediction-error filter, which roughly converts $\mathbf{H}^L(z)$ into a memoryless channel $\mathbf{H}_L z^{-L-\tilde{N}}$. (Recall that \tilde{N} is the order of the backward predictor). The spatial methods of chapter 4 are then applied to invert \mathbf{H}_L . The generalized LP detector is in fact a family of detectors, one for each index L , and the FLP detector can be viewed as the special case for which $L = 0$. Using an index other than $L = 0$ can result in better performance, but at the expense of increased complexity. Ideally the index L should correspond to the tap of greatest energy.

There is one remaining deficiency in the blind implementation of the generalized detector. If the channel is unknown, then the channel energy distribution is also unknown. Therefore, there is no obvious way to choose the optimal index L . One possibility is to try several or all possible indices, either serially or in parallel. Of course, either method has an associated cost. The serial method would increase the recovery time, while the parallel method would increase the complexity of the detector. A clever method for choosing L remains an open issue, although we discuss one promising possibility in the future work section of chapter 7.

APPENDIX 5-1: PROOF OF THEOREM 5-2 AND COROLLARY 5-2.1
(EXISTENCE OF THE ARMA MODEL)

To prove the existence of an ARMA model, we need to show that $[\mathbf{I} - \mathbf{A}^L(z)]\mathbf{H}(z) = \mathbf{H}^L(z)$ has a solution. We can express the equation in block-matrix form as follows:

$$[\mathbf{I} \ \mathbf{0}_{m \times mL} \ -\mathbf{A}_{L+1} \ \dots \ -\mathbf{A}_{L+N}] \mathcal{H}_{N+L+1}^f = [\mathbf{H}_0 \ \dots \ \mathbf{H}_L \ \mathbf{0}_{m \times n(M+N)}]. \quad (5-75)$$

Subtracting $[\mathbf{I}_m \ \mathbf{0}_{m \times (N+L)m}] \mathcal{H}_{N+L+1}^f$ from both sides of (5-75) yields

$$[\mathbf{0}_{m \times m(L+1)} \ -\mathbf{A}_{L+1} \ \dots \ -\mathbf{A}_{N+L}] \mathcal{H}_{N+L+1}^f = -[\mathbf{0}_{m \times m} \ \mathbf{H}_{L+1} \ \dots \ \mathbf{H}_M \ \mathbf{0}_{m \times (N+L)n}] \quad (5-76)$$

$$[\mathbf{A}_{L+1} \ \dots \ \mathbf{A}_{L+N}] \mathcal{H}_N^f = [\mathbf{H}_{L+1} \ \dots \ \mathbf{H}_M \ \mathbf{0}_{m \times n(N+L)}]. \quad (5-77)$$

This is a system of $mn(M+N)$ scalar equations with m^2N unknowns. If (5-2) is satisfied then \mathcal{H}_N^f is full rank for all N [137]; therefore, the equations are linearly independent, and the system has a solution if $N \geq \left\lceil \frac{Mn}{m-n} \right\rceil$. (Theorem 5-2) \square

We now solve for the AR parameters of the ARMA model. By substituting, we see that (5-11) satisfies (5-77):

$$\begin{aligned} [\mathbf{A}_{L+1} \ \dots \ \mathbf{A}_{L+N}] \mathcal{H} &= [\mathbf{H}_{L+1} \ \dots \ \mathbf{H}_M \ \mathbf{0}_{m \times n(N+L)}] \mathcal{H}^\dagger \mathcal{H} + \mathbf{V} \mathcal{H} \\ &= [\mathbf{H}_{L+1} \ \dots \ \mathbf{H}_M \ \mathbf{0}_{m \times n(N+L)}]. \end{aligned} \quad (5-78)$$

The solution is unique if and only if the dimension of $null(\mathcal{H})$ is trivial; *i.e.*, if and only if $mN - n(M+N) = 0$. (Corollary 5-2.1) \square

APPENDIX 5-2: PROOF OF THEOREM 5-3 AND COROLLARY 5-3.1
(EXISTENCE OF THE AP MODEL)

Expressing $[\mathbf{I} - \mathbf{A}^{-1}(z)]z^{-N}\mathbf{H}(z) = \mathbf{H}_M z^{-M-N}$ in block-matrix form yields

$$[-\mathbf{A}_{-N} \dots -\mathbf{A}_{-1} \quad \mathbf{I}] \mathcal{H}_{N+1}^f = [\mathbf{0}_{m \times (M+N)n} \quad \mathbf{H}_M]. \quad (5-79)$$

Subtracting $[\mathbf{0}_{m \times Nm} \quad \mathbf{I}_m] \mathcal{H}_{N+1}^f$ from both sides of (5-79) yields

$$[-\mathbf{A}_{-N} \dots -\mathbf{A}_{-1} \quad \mathbf{0}_{m \times m}] \mathcal{H}_{N+1}^f = -[\mathbf{0}_{m \times Nn} \quad \mathbf{H}_0 \dots \mathbf{H}_{M-1} \quad \mathbf{0}_{m \times m}], \quad (5-80)$$

$$[\mathbf{A}_{-N} \dots \mathbf{A}_{-1}] \mathcal{H}_N^f = [\mathbf{0}_{m \times Nn} \quad \mathbf{H}_0 \dots \mathbf{H}_{M-1}]. \quad (5-81)$$

This is a system of $mn(M+N)$ scalar equations with m^2N unknowns. If (5-2) is satisfied then \mathcal{H}_N^f is full rank for all N [137]; therefore, the equations are linearly independent, and the system has a solution if $N \geq \left\lceil \frac{Mn}{m-n} \right\rceil$. (Theorem 5-3) \square

We can solve for the AP parameters by substituting; we see that (5-13) satisfies (5-81):

$$\begin{aligned} [\mathbf{A}_{-N} \dots \mathbf{A}_{-1}] \mathcal{H} &= [\mathbf{0}_{m \times Nn} \quad \mathbf{H}_0 \dots \mathbf{H}_{M-1}] \mathcal{H}^\dagger \mathcal{H} + \mathbf{V} \mathcal{H} \\ &= [\mathbf{0}_{m \times Nn} \quad \mathbf{H}_0 \dots \mathbf{H}_{M-1}]. \end{aligned} \quad (5-82)$$

The solution is unique if and only if the dimension of $\text{null}(\mathcal{H})$ is trivial; *i.e.*, if and only if $mN - n(M+N) = 0$. (Corollary 5-3.1) \square

**APPENDIX 5-3: PROOF OF THEOREMS 5-8, 5-10, COROLLARIES 5-8.1,
5-10.1 (OPTIMAL L -STEP FORWARD PREDICTOR)**

Applying an $(L+1)$ -step forward predictor of order N to the channel (1-1) output \mathbf{r}_k produces a forward prediction error \mathbf{e}_k , which can be expressed as follows:

$$\mathbf{e}_k = \left[\mathbf{I}_m \mid \mathbf{0}_{m \times mL} \mid \overbrace{\begin{matrix} -\mathbf{P}_{L+1} & -\mathbf{P}_{L+2} & \cdots & -\mathbf{P}_{L+N} \end{matrix}}^{-\mathbf{P}^L} \right] \times$$

$$\left\{ \begin{array}{c} \mathbf{K}^L \\ \left[\begin{array}{c|c} \mathbf{H}_0 \ \mathbf{H}_1 \ \cdots \ \mathbf{H}_L & \mathbf{H}_{L+1} \ \cdots \ \mathbf{H}_M \ \mathbf{0} \ \cdots \ \mathbf{0} \\ \mathbf{0} \ \mathbf{H}_0 \ \cdots \ \mathbf{H}_{L-1} & \mathbf{H}_L \ \cdots \ \mathbf{H}_M \ \mathbf{0} \ \cdots \ \mathbf{0} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{0} & \mathbf{H}_0 & \mathbf{H}_1 \ \cdots \ \mathbf{H}_L & \cdots \ \mathbf{H}_M \ \mathbf{0} \ \cdots \ \mathbf{0} \\ \hline \mathbf{0}_{mN \times n(L+1)} & \mathbf{H}_0 \ \mathbf{H}_1 \ \cdots \ \mathbf{H}_L & \cdots \ \mathbf{H}_M & \mathbf{0} \\ & \mathbf{0} & \ddots & \mathbf{H}_0 \ \mathbf{H}_1 \ \cdots \ \mathbf{H}_L & \cdots & \mathbf{H}_M \end{array} \right] \begin{array}{c} \mathbf{x}_k \\ \vdots \\ \mathbf{x}_{k-L} \\ \hline \mathbf{x}_{k-L-1} \\ \vdots \\ \mathbf{x}_{k-L-M-N} \end{array} \right\} \mathbf{X}_{k-L-1}$$

$$\left. \begin{array}{c} \mathcal{H} = \mathcal{H}_N \\ \mathcal{H}_{N+L+1} \end{array} \right\}$$

$$+ \left\{ \begin{array}{c} \mathbf{n}_k \\ \mathbf{n}_{k-L} \\ \hline \mathbf{n}_{k-L-1} \\ \mathbf{n}_{k-L-N} \end{array} \right\} \mathbf{N}_{k-L-1}$$

$$\mathbf{e}_k = \sum_{i=0}^L \mathbf{H}_i \mathbf{x}_{k-i} + (\mathbf{K}^L - \mathbf{P}^L \mathcal{H}) \mathbf{X}_{k-L-1} + \mathbf{n}_k - \mathbf{P}^L \mathbf{N}_{k-L-1}. \quad (5-83)$$

Using the fact that the signal and noise are white and uncorrelated, we can express the mean-square prediction error $\xi = \text{tr}E[\mathbf{e}_k \mathbf{e}_k^*]$ as follows:

$$\xi = \sum_{i=0}^L \text{tr}(\mathbf{H}_i \mathbf{H}_i^*) + \text{tr}[(\mathbf{K}^L - \mathbf{P}^L \mathcal{H})(\mathbf{K}^L - \mathbf{P}^L \mathcal{H})^* + \sigma^2 \mathbf{I} + \sigma^2 \mathbf{P}^L \mathbf{P}^{L*}]. \quad (5-84)$$

We can neglect terms that are independent of \mathbf{P}^L , so it suffices to minimize

$$J(\mathbf{P}^L) = (1/2) \text{tr}[(\mathbf{K}^L - \mathbf{P}^L \mathcal{H})(\mathbf{K}^L - \mathbf{P}^L \mathcal{H})^* + \sigma^2 \mathbf{P}^L \mathbf{P}^{L*}]. \quad (5-85)$$

Taking the gradient of $J(\mathbf{P}^L)$ and setting it equal to $\mathbf{0}$, we have

$$\nabla J(\mathbf{P}^L) = (\mathbf{P}^L \mathcal{H} - \mathbf{K}^L) \mathcal{H}^* + \sigma^2 \mathbf{P}^L = \mathbf{P}^L (\mathcal{H} \mathcal{H}^* + \sigma^2 \mathbf{I}) - \mathbf{K}^L \mathcal{H}^* = \mathbf{0}. \quad (5-86)$$

Since the Hessian $\nabla^2 J(\mathbf{P}^L) = (\mathcal{H} \mathcal{H}^* + \sigma^2 \mathbf{I}) > 0$ is positive definite, the solution of (5-86) minimizes (5-85). The optimal coefficients are thus given by (5-48), restated here:

$$\mathbf{P}^L = \mathbf{K}^L \mathcal{H}^* (\mathcal{H} \mathcal{H}^* + \sigma^2 \mathbf{I})^{-1}. \quad (5-87)$$

The expression of (5-49) follows from the identity $\mathcal{H}^* (\mathcal{H} \mathcal{H}^* + \sigma^2 \mathbf{I})^{-1} = (\mathcal{H}^* \mathcal{H} + \sigma^2 \mathbf{I})^{-1} \mathcal{H}^*$, which is derived in Appendix 3-1. (Theorem 5-10) \square

The expressions of (5-30) and (5-31) follow from (5-48) and (5-49), respectively, by simply substituting $L = 0$. (Theorem 5-8) \square

The coefficients of the forward cascade $\mathbf{F}^L = [\mathbf{F}_0 \dots \mathbf{F}_{M+L+N}]$ for general L are given by the following convolution representation:

$$\mathbf{F}^L = [\mathbf{I}_m \mid \mathbf{0}_{m \times mL} \mid -\mathbf{P}^L] \mathcal{H}_{N+L+1} = [\mathbf{H}_0 \dots \mathbf{H}_L \mid (\mathbf{K}^L - \mathbf{P}^L \mathcal{H})]. \quad (5-88)$$

Substituting the optimal predictor coefficients (5-49), we obtain

$$\mathbf{F}^L = [\mathbf{H}_0 \dots \mathbf{H}_L \mid (\mathbf{K}^L - \mathbf{K}^L (\mathcal{H}^* \mathcal{H} + \sigma^2 \mathbf{I})^{-1} \mathcal{H}^* \mathcal{H})]. \quad (5-89)$$

Left-factoring $\mathbf{K}^L(\mathcal{H}^* \mathcal{H} + \sigma^2 \mathbf{I})^{-1}$ yields

$$\begin{aligned}
\mathbf{F}^L &= [\mathbf{H}_0 \dots \mathbf{H}_L \mid \mathbf{K}^L(\mathcal{H}^* \mathcal{H} + \sigma^2 \mathbf{I})^{-1}\{(\mathcal{H}^* \mathcal{H} + \sigma^2 \mathbf{I}) - \mathcal{H}^* \mathcal{H}\}] \\
&= [\mathbf{H}_0 \dots \mathbf{H}_L \mid \mathbf{K}^L \sigma^2 (\mathcal{H}^* \mathcal{H} + \sigma^2 \mathbf{I})^{-1}] \\
&= [\mathbf{H}_0 \dots \mathbf{H}_L \mid [\mathbf{H}_{L+1} \dots \mathbf{H}_M \mathbf{0}_{m \times n(L+N)}] \cdot \Psi], \tag{5-90}
\end{aligned}$$

which verifies (5-50). (Corollary 5-10.1) \square

The expression in (5-32) follows from (5-50) by substituting $L = 0$. We need to derive (5-33) to complete the proof of Corollary 5-8.1. Using the expression for the forward prediction error \mathbf{e}_k given by (5-83) with $L = 0$, the forward prediction error covariance is given by

$$\Phi_e = E[\mathbf{e}_k \mathbf{e}_k^*] = \mathbf{H}_0 \mathbf{H}_0^* + \sigma^2 \mathbf{I} + (\mathbf{K} - \mathbf{P} \mathcal{H})(\mathbf{K} - \mathbf{P} \mathcal{H})^* + \sigma^2 \mathbf{P} \mathbf{P}^*. \tag{5-91}$$

Substituting $\mathbf{K} \Psi$ for $(\mathbf{K} - \mathbf{P} \mathcal{H})$ and $\sigma^{-2} \mathbf{K} \Psi \mathcal{H}^*$ for \mathbf{P} in (5-91) yields

$$\begin{aligned}
\Phi_e &= \mathbf{H}_0 \mathbf{H}_0^* + \sigma^2 \mathbf{I} + \mathbf{K} \Psi \Psi \mathbf{K}^* + \sigma^2 \sigma^{-2} \mathbf{K} \Psi \mathcal{H}^* \mathcal{H} \Psi \mathbf{K}^* \sigma^{-2} \\
&= \mathbf{H}_0 \mathbf{H}_0^* + \sigma^2 \mathbf{I} + \mathbf{K} \Psi (\sigma^2 \mathbf{I}) \Psi \mathbf{K}^* \sigma^{-2} + \mathbf{K} \Psi (\mathcal{H}^* \mathcal{H}) \Psi \mathbf{K}^* \sigma^{-2} \\
&= \mathbf{H}_0 \mathbf{H}_0^* + \sigma^2 \mathbf{I} + \mathbf{K} \Psi (\sigma^2 \mathbf{I} + \mathcal{H}^* \mathcal{H}) \Psi \mathbf{K}^* \sigma^{-2} \\
&= \mathbf{H}_0 \mathbf{H}_0^* + \sigma^2 \mathbf{I} + \mathbf{K} \Psi \Psi^{-1} \Psi \mathbf{K}^* \\
&= \mathbf{H}_0 \mathbf{H}_0^* + \sigma^2 \mathbf{I} + \mathbf{K} \Psi \mathbf{K}^*, \tag{5-92}
\end{aligned}$$

which verifies (5-33). (Corollary 5-8.1) \square

APPENDIX 5-4: DERIVATION OF LEMMA 5-4

Let $\mathbf{e}_k = \sum_{i=0}^{M+N} \mathbf{F}_i \mathbf{x}_k + \mathbf{v}_k$, where \mathbf{x}_k and \mathbf{v}_k are independent with covariances \mathbf{I} and Φ_v , respectively. In matrix notation, we can express \mathbf{e}_k as

$$\mathbf{e}_k = \mathbf{F}\mathbf{X}_k - \mathbf{v}_k, \quad (5-93)$$

where $\mathbf{F} = [\mathbf{F}_0 \ \mathbf{F}_1 \ \dots \ \mathbf{F}_{M+N}]$, and $\mathbf{X}_k^T = [\mathbf{x}_k^T \ \mathbf{x}_{k-1}^T \ \dots \ \mathbf{x}_{k-M-N}^T]$. For an $n \times m$ memory-less detector \mathbf{C} , the D -delay MSE sum is given by

$$J(\mathbf{C}) = E[\|\mathbf{C}\mathbf{e}_k - \mathbf{x}_{k-D}\|^2] = E[\|\mathbf{C}\mathbf{F}\mathbf{X}_k - \mathbf{C}\mathbf{v}_k - \mathbf{x}_{k-D}\|^2]. \quad (5-94)$$

If $D \in \{0, 1, \dots, M+N\}$, then we have

$$J(\mathbf{C}) = E[\|(\mathbf{C}\mathbf{F} - \tilde{\mathbf{I}}_D)\mathbf{X}_k - \mathbf{C}\mathbf{v}_k\|^2], \quad (5-95)$$

where $\tilde{\mathbf{I}}_D \equiv [\mathbf{0}_{n \times nD} \ \mathbf{I}_n \ \mathbf{0}_{n \times n(M+N-D)}]$ such that $\mathbf{x}_{k-D} = \tilde{\mathbf{I}}_D \mathbf{X}_k$. Observe that \mathbf{X}_k and \mathbf{v}_k are independent, so

$$J(\mathbf{C}) = \text{tr}[(\mathbf{C}\mathbf{F} - \tilde{\mathbf{I}}_D)(\mathbf{C}\mathbf{F} - \tilde{\mathbf{I}}_D)^* + \mathbf{C}\Phi_v\mathbf{C}^*]. \quad (5-96)$$

The gradient of $J(\mathbf{C})$ is then

$$\nabla J(\mathbf{P}) = 2(\mathbf{C}\mathbf{F} - \tilde{\mathbf{I}}_D)\mathbf{F}^* + 2\mathbf{C}\Phi_v = \mathbf{0}. \quad (5-97)$$

With $\tilde{\mathbf{I}}_D \mathbf{F}^* = \mathbf{F}_D^*$, we have

$$\mathbf{C}_{MMSE} = \mathbf{F}_D^* (\mathbf{F}\mathbf{F}^* + \Phi_v)^{-1} = \mathbf{F}_D^* \Phi_e^{-1}. \quad \square \quad (5-98)$$

**APPENDIX 5-5: PROOF OF LEMMA 5-5 (THE RELATIONSHIP BETWEEN
THE FLP AND MMSE DETECTORS AT ZERO-DELAY)**

Let $\mathbf{C}(z) = \sum_{i=0}^N \mathbf{C}_i z^{-i}$ be an $n \times m$ detector with $N+1$ taps for the channel of (1-1) with observation \mathbf{r}_k . We can express the output $\mathbf{z}_k = \sum_{i=0}^N \mathbf{C}_i \mathbf{r}_{k-i}$ of this detector in matrix notation as follows:

$$\mathbf{z}_k = \underbrace{\begin{bmatrix} \tilde{\mathbf{C}} \\ \mathbf{C}_0 | \mathbf{C}_1 \dots \mathbf{C}_N \end{bmatrix}}_{\mathbf{C}} \left\{ \underbrace{\begin{bmatrix} \mathbf{H}_0 & \mathbf{H}_1 & \dots & \mathbf{H}_M & \mathbf{0} & \dots & \mathbf{0} \\ \mathbf{0} & \mathbf{H}_0 & \mathbf{H}_1 & \dots & \mathbf{H}_M & & \vdots \\ \vdots & & \ddots & \ddots & & \ddots & \mathbf{0} \\ \mathbf{0} & \dots & \mathbf{0} & \mathbf{H}_0 & \mathbf{H}_1 & \dots & \mathbf{H}_M \end{bmatrix}}_{\mathcal{H}_{N+1}} \underbrace{\begin{bmatrix} \mathbf{x}_k \\ \mathbf{x}_{k-1} \\ \vdots \\ \mathbf{x}_{k-M-N} \end{bmatrix}}_{\mathbf{X}_k} + \underbrace{\begin{bmatrix} \mathbf{n}_k \\ \mathbf{n}_{k-1} \\ \vdots \\ \mathbf{n}_{k-N} \end{bmatrix}}_{\mathbf{N}_k} \right\}$$

$$\mathbf{z}_k = \mathbf{C}(\mathcal{H}_{N+1} \mathbf{X}_k + \mathbf{N}_k). \quad (5-99)$$

The zero-delay detector error $\mathbf{e}_k = \mathbf{z}_k - \mathbf{x}_k$ can then be expressed as

$$\mathbf{e}_k = (\mathbf{C} \mathcal{H}_{N+1} - \tilde{\mathbf{I}}_0) \mathbf{X}_k + \mathbf{C} \mathbf{N}_k, \quad (5-100)$$

where $\tilde{\mathbf{I}}_0 \equiv [\mathbf{I}_n \ \mathbf{0}_{n \times n(M+N)}]$ such that $\mathbf{x}_k = \tilde{\mathbf{I}}_0 \mathbf{X}_k$. The total MSE $tr E[\mathbf{e}_k \mathbf{e}_k^*]$ is then

$$J(\mathbf{C}) = tr(\mathbf{C} \mathcal{H}_{N+1} - \tilde{\mathbf{I}}_0)(\mathbf{C} \mathcal{H}_{N+1} - \tilde{\mathbf{I}}_0)^* + \sigma^2 tr \mathbf{C} \mathbf{C}^*. \quad (5-101)$$

Taking the gradient of $J(\mathbf{C})$ and setting it equal to $\mathbf{0}$ produces

$$\nabla J(\mathbf{C}) = (\mathbf{C} \mathcal{H}_{N+1} - \tilde{\mathbf{I}}_0) \mathcal{H}_{N+1}^* + \sigma^2 \mathbf{C} = \mathbf{0} \quad (5-102)$$

$$\mathbf{C}(\mathcal{H}_{N+1} \mathcal{H}_{N+1}^* + \sigma^2 \mathbf{I}) = \tilde{\mathbf{I}}_0 \mathcal{H}_{N+1}^*. \quad (5-103)$$

We can partition (5-103) as

$$[\mathbf{C}_0 \quad \tilde{\mathbf{C}}] \begin{bmatrix} \mathbf{H}_0 \mathbf{H}_0^* + \mathbf{K} \mathbf{K}^* + \sigma^2 \mathbf{I}_m & \mathbf{K} \mathcal{H}^* \\ \mathcal{H} \mathbf{K}^* & \mathcal{H} \mathcal{H}^* + \sigma^2 \mathbf{I}_{mN} \end{bmatrix} = [\mathbf{H}_0^* \quad \mathbf{0}_{n \times mN}], \quad (5-104)$$

which yields the following system of equations:

$$\mathbf{C}_0 (\mathbf{H}_0 \mathbf{H}_0^* + \mathbf{K} \mathbf{K}^* + \sigma^2 \mathbf{I}) + \tilde{\mathbf{C}} \mathcal{H} \mathbf{K}^* = \mathbf{H}_0^* \quad (5-105)$$

$$\mathbf{C}_0 \mathbf{K} \mathcal{H}^* + \tilde{\mathbf{C}} (\mathcal{H} \mathcal{H}^* + \sigma^2 \mathbf{I}) = \mathbf{0}. \quad (5-106)$$

The solution to (5-106) is

$$\tilde{\mathbf{C}} = -\mathbf{C}_0 \mathbf{K} \mathcal{H}^* (\mathcal{H} \mathcal{H}^* + \sigma^2 \mathbf{I})^{-1}. \quad (5-107)$$

Therefore, the optimal MMSE taps can be written as

$$\begin{aligned} \mathbf{C} = [\mathbf{I} \quad \tilde{\mathbf{C}}] &= \mathbf{C}_0 [\mathbf{I} \mid -\mathbf{K} \mathcal{H}^* (\mathcal{H} \mathcal{H}^* + \sigma^2 \mathbf{I})^{-1}] \\ &= \mathbf{C}_0 [\mathbf{I} \mid -\mathbf{P}]. \end{aligned} \quad (5-108)$$

where $\mathbf{P} = [\mathbf{P}_1 \dots \mathbf{P}_N]$ is given by (5-30). Hence, the one-step forward prediction-error filter is a right-factor of the $(N+1)$ -tap MMSE detector:

$$\mathbf{C}_{MMSE}(z) = \mathbf{C}_0 [\mathbf{I} - \mathbf{P}(z)]. \quad (5-109)$$

The coefficient \mathbf{C}_0 can be derived from (5-105), but this is unnecessary; \mathbf{C}_0 must be the $n \times m$ matrix that minimizes the total MSE given \mathbf{e}_k , the forward prediction error. If there were another matrix \mathbf{D}_0 producing smaller total MSE, then $\mathbf{D}(z) = \mathbf{D}_0 [\mathbf{I} - \mathbf{P}(z)]$ would have smaller total MSE than $\mathbf{C}_{MMSE}(z)$, a contradiction. \square

APPENDIX 5-6: PROOF OF THEOREM 5-9
(PRECURSORS BELOW THE NOISE FLOOR ARE IGNORED.)

We can express the matrix of channel taps as a matrix of column vectors:

$$[\mathbf{H}_0 \mathbf{H}_1 \dots \mathbf{H}_M] = [\mathbf{h}_0^{(1)} \dots \mathbf{h}_0^{(n)} \mid \mathbf{h}_1^{(1)} \dots \mathbf{h}_1^{(n)} \mid \dots \mid \mathbf{h}_M^{(1)} \dots \mathbf{h}_M^{(n)}]. \quad (5-110)$$

We can then express Ψ as follows:

$$\Psi = \sigma^2 \begin{bmatrix} \mathbf{h}_0^{(1)*} \mathbf{h}_0^{(1)} + \sigma^2 & \mathbf{h}_0^{(1)*} \tilde{\mathbf{K}} \\ \tilde{\mathbf{K}}^* \mathbf{h}_0^{(1)} & \tilde{\mathcal{H}}^* \tilde{\mathcal{H}} + \sigma^2 \mathbf{I} \end{bmatrix}^{-1}, \quad (5-111)$$

where $\tilde{\mathbf{K}} = [\mathbf{h}_0^{(2)} \dots \mathbf{h}_0^{(n)} \mid \mathbf{h}_1^{(1)} \dots \mathbf{h}_0^{(n)} \mid \dots \mid \mathbf{h}_M^{(1)} \dots \mathbf{h}_M^{(n)} \mid \mathbf{0}_{m \times Nn}]$, and where $\tilde{\mathcal{H}}$ is an $mN-1 \times n(M+N)-1$ block-Toeplitz matrix given by

$$\tilde{\mathcal{H}} = \begin{bmatrix} \mathbf{h}_0^{(2)} & \dots & \mathbf{h}_0^{(n)} & \mathbf{h}_1^{(1)} & \dots & \mathbf{h}_1^{(n)} & \dots & \mathbf{h}_M^{(n)} & \mathbf{0} & \dots & \mathbf{0} \\ \mathbf{0} & \mathbf{h}_0^{(2)} & \dots & \mathbf{h}_0^{(n)} & \mathbf{h}_1^{(1)} & \dots & \mathbf{h}_1^{(n)} & \dots & \mathbf{h}_M^{(n)} & \dots & \mathbf{0} \\ \vdots & & \ddots & & \ddots & \ddots & & & & \ddots & \vdots \\ \mathbf{0} & \dots & \mathbf{0} & \mathbf{h}_0^{(2)} & \dots & \mathbf{h}_0^{(n)} & \mathbf{h}_1^{(1)} & \dots & \mathbf{h}_1^{(n)} & \dots & \mathbf{h}_M^{(n)} \end{bmatrix}. \quad (5-112)$$

Taking the limit of Ψ as \mathbf{H}_0 goes to zero yields

$$\lim_{\mathbf{H}_0 \rightarrow \mathbf{0}} \Psi = \sigma^2 \begin{bmatrix} \sigma^2 & \mathbf{0} \\ \mathbf{0} & \tilde{\mathcal{H}}^* \tilde{\mathcal{H}} + \sigma^2 \mathbf{I} \end{bmatrix}^{-1} = \begin{bmatrix} 1 & \mathbf{0} \\ \mathbf{0} & \sigma^2 (\tilde{\mathcal{H}}^* \tilde{\mathcal{H}} + \sigma^2 \mathbf{I})^{-1} \end{bmatrix} \equiv \begin{bmatrix} 1 & \mathbf{0} \\ \mathbf{0} & \tilde{\Psi} \end{bmatrix}. \quad (5-113)$$

Therefore, as $\mathbf{h}_0^{(1)}$ becomes vanishingly small, the coefficients of $\mathbf{F}(z)$ become

$$\lim_{\mathbf{h}_0^{(1)} \rightarrow \mathbf{0}} [\mathbf{h}_0^{(1)} \dots \mathbf{h}_0^{(n)} [\mathbf{h}_1^{(1)} \dots \mathbf{h}_M^{(n)} \mathbf{0}_{m \times Nn}] \cdot \Psi] =$$

$$[\mathbf{0}_{m \times 1} \mathbf{h}_0^{(2)} \dots \mathbf{h}_0^{(n)} \mathbf{h}_1^{(1)} [\mathbf{h}_1^{(2)} \dots \mathbf{h}_1^{(n)} \dots \mathbf{h}_M^{(n)} \mathbf{0}_{m \times Nn}] \cdot \tilde{\Psi}]. \quad (5-114)$$

Let $\mathbf{h}^{(i)}(z)$ denote the i -th column of $\mathbf{H}(z)$ so that $\mathbf{H}(z) = [\mathbf{h}^{(1)}(z) \dots \mathbf{h}^{(n)}(z)]$. An equivalent expression for (5-114) is then given by

$$\lim_{\mathbf{h}_0^{(1)} \rightarrow \mathbf{0}} [\mathbf{I} - \mathbf{P}(z)] \mathbf{H}(z) = [\mathbf{I} - \mathbf{P}(z)] [\tilde{\mathbf{h}}^{(1)}(z) \mathbf{h}^{(2)}(z) \dots \mathbf{h}^{(n)}(z)], \quad (5-115)$$

where $\tilde{\mathbf{h}}^{(1)}(z) = \sum_{k=1}^M \mathbf{h}_k^{(1)} z^{-k}$. The predictor behaves as if the channel were $\tilde{\mathbf{H}}(z) = [\tilde{\mathbf{h}}^{(1)}(z) \mathbf{h}^{(2)}(z) \dots \mathbf{h}^{(n)}(z)]$, rather than $\mathbf{H}(z)$.

The same argument can be repeated for the new effective channel $\tilde{\mathbf{H}}(z)$; therefore, by induction, (5-115) generalizes to

$$\lim_{\Gamma_1 \rightarrow \mathbf{0}} [\mathbf{I} - \mathbf{P}(z)] \mathbf{H}(z) = [\mathbf{I} - \mathbf{P}(z)] [\sum_{k=L_1+1}^M \mathbf{h}_k^{(1)} z^{-k} \mid \mathbf{h}^{(2)}(z) \dots \mathbf{h}^{(n)}(z)], \quad (5-116)$$

where $\Gamma_1 = [\mathbf{h}_0^{(1)} \mathbf{h}_1^{(1)} \dots \mathbf{h}_{L_1}^{(1)}]$. Finally, we argue that (5-116) must hold for any user i , not just user 1. We can change the labeling of users by right-multiplying $\mathbf{H}(z)$ by any permutation matrix Π of our choosing, and because $\mathbf{H}(z)\Pi$ and $\mathbf{H}(z)$ have identical second-order output statistics, the behavior of the predictor is the same. It follows that

$$\lim_{\Gamma \rightarrow \mathbf{0}} [\mathbf{I} - \mathbf{P}(z)] \mathbf{H}(z) =$$

$$[\mathbf{I} - \mathbf{P}(z)] [\sum_{k=L_1+1}^M \mathbf{h}_k^{(1)} z^{-k} \mid \dots \mid \sum_{k=L_n+1}^M \mathbf{h}_k^{(n)} z^{-k}], \quad (5-117)$$

where $\Gamma = [\Gamma_1 \Gamma_2 \dots \Gamma_n]$ is the total precursor. Taking the limit of (5-117) as the noise variance goes to zero verifies (5-45). \square

APPENDIX 5-7: PROOF OF THEOREM 5-11
(OPTIMAL BACKWARD PREDICTOR FOLLOWING FORWARD PREDICTOR)

The backward prediction error \mathbf{b}_k can be expressed in matrix notation as follows:

$$\mathbf{b}_k = \left[\begin{array}{ccc|c} \overbrace{-\mathbf{P}_{\tilde{N}} \quad -\mathbf{P}_{\tilde{N}-1} \quad \dots \quad -\mathbf{P}_{\tilde{N}-1}}^{-\mathbf{P}^{(-1)}} & & & \mathbf{I}_m \end{array} \right] \times$$

$$\left\{ \begin{array}{l} \mathcal{F} = \mathcal{F}_{\tilde{N}} \\ \left[\begin{array}{cccc|c} \mathbf{F}_0 & \dots & \mathbf{F}_{\tilde{M}-1} & \mathbf{F}_{\tilde{M}} & \mathbf{0} & \dots & \mathbf{0} \\ \vdots & & \ddots & \ddots & \ddots & & \vdots \\ \mathbf{0} & \dots & \mathbf{F}_0 & \dots & \mathbf{F}_{\tilde{M}-1} & \mathbf{F}_{\tilde{M}} & \mathbf{0} \\ \hline \mathbf{0} & \dots & \mathbf{0} & \mathbf{F}_0 & \dots & \mathbf{F}_{\tilde{M}-1} & \mathbf{F}_{\tilde{M}} \end{array} \right] \begin{array}{c} \mathbf{X}_k \\ \mathbf{x}_k \\ \vdots \\ \mathbf{x}_{k-\tilde{M}-\tilde{N}+2} \\ \mathbf{x}_{k-\tilde{M}-\tilde{N}+1} \\ \hline \mathbf{x}_{k-\tilde{M}-\tilde{N}} \end{array} \\ \mathbf{K}_F \\ \mathcal{E} = \mathcal{E}_{\tilde{N}} \\ \left[\begin{array}{cccc|c} \mathbf{I}_m & \mathbf{0} & -\mathbf{P}_{L+1} & \dots & -\mathbf{P}_{L+N} & \mathbf{0} & \dots & \mathbf{0} \\ \mathbf{0} & \dots & \mathbf{I}_m & \mathbf{0} & -\mathbf{P}_{L+1} & \dots & -\mathbf{P}_{L+N} & \mathbf{0} \\ \hline \mathbf{0} & \dots & \mathbf{0} & \mathbf{I}_m & \dots & \dots & -\mathbf{P}_{L+N-1} & -\mathbf{P}_{L+N} \end{array} \right] \begin{array}{c} \mathbf{N}_k \\ \mathbf{n}_k \\ \vdots \\ \mathbf{n}_{k-L-N-\tilde{N}+2} \\ \mathbf{n}_{k-L-N-\tilde{N}+1} \\ \hline \mathbf{n}_{k-L-N-\tilde{N}} \end{array} \\ \mathbf{K}_E \end{array} \right\} +$$

$$\mathbf{b}_k = -\mathbf{P}^{-1} \mathcal{F} \mathbf{X}_k + \mathbf{K}_F \mathbf{X}_k + \mathbf{F}_{\tilde{M}} \mathbf{x}_{k-\tilde{M}-\tilde{N}}$$

$$-\mathbf{P}^{-1} \mathcal{E} \mathbf{N}_k + \mathbf{K}_E \mathbf{N}_k + \mathbf{P}_{L+N} \mathbf{n}_{k-L-N-\tilde{N}}. \quad (5-118)$$

The signal and noise are white and uncorrelated, so we can express the mean-square backward prediction error $\xi = \text{tr}E[\mathbf{b}_k \mathbf{b}_k^*]$ as follows:

$$\begin{aligned} \xi = \text{tr}[\mathbf{F}_{\tilde{M}} \mathbf{F}_{\tilde{M}}^* + (\mathbf{K}_F - \mathbf{P}^{(-1)}\mathcal{F})(\mathbf{K}_F - \mathbf{P}^{-1}\mathcal{F})^*] \\ + \sigma^2 \text{tr}[\mathbf{P}_{L+N} \mathbf{P}_{L+N}^* + \sigma^2(\mathbf{K}_E - \mathbf{P}^{-1}\mathcal{E})(\mathbf{K}_E - \mathbf{P}^{-1}\mathcal{E})^*]. \end{aligned} \quad (5-119)$$

We can neglect terms that are independent of \mathbf{P}^{-1} , so it suffices to minimize

$$J(\mathbf{P}^{-1}) = \text{tr}[(\mathbf{K}_F - \mathbf{P}^{-1}\mathcal{F})(\mathbf{K}_F - \mathbf{P}^{-1}\mathcal{F})^* + \sigma^2(\mathbf{K}_E - \mathbf{P}^{-1}\mathcal{E})(\mathbf{K}_E - \mathbf{P}^{-1}\mathcal{E})^*]. \quad (5-120)$$

Taking the gradient of $J(\mathbf{P}^{-1})$ and setting it equal to $\mathbf{0}$, we have

$$\begin{aligned} (1/2) \nabla J(\mathbf{P}^{-1}) &= (\mathbf{K}_F - \mathbf{P}^{-1}\mathcal{F})\mathcal{F}^* + \sigma^2(\mathbf{K}_E - \mathbf{P}^{-1}\mathcal{E})\mathcal{E}^* \\ &= \mathbf{P}^{-1}(\mathcal{F}\mathcal{F}^* + \sigma^2\mathcal{E}\mathcal{E}^*) - (\mathbf{K}_F\mathcal{F}^* + \sigma^2\mathbf{K}_E\mathcal{E}^*) = \mathbf{0}. \end{aligned} \quad (5-121)$$

Since $\mathcal{F}\mathcal{F}^* + \sigma^2\mathcal{E}\mathcal{E}^* > 0$, the solution of (5-121) minimizes (5-120). The optimal coefficients are thus given by

$$\mathbf{P}^{-1} = (\mathbf{K}_F\mathcal{F}^* + \sigma^2\mathbf{K}_E\mathcal{E}^*)(\mathcal{F}\mathcal{F}^* + \sigma^2\mathcal{E}\mathcal{E}^*)^{-1}. \quad \square \quad (5-122)$$

APPENDIX 5-8: A BACKWARD LP DETECTOR

Applying a one-step backward predictor of order N to the channel (1-1) output \mathbf{r}_k produces a backward prediction error \mathbf{b}_k , which can be expressed as follows:

$$\mathbf{b}_k = \underbrace{\left[\begin{array}{c|c} -\mathbf{P}_{-\tilde{N}} \dots -\mathbf{P}_{-1} & \mathbf{I} \end{array} \right]}_{\mathbf{P}^{-1}} \times \left(\underbrace{\left[\begin{array}{cccc|c} \mathbf{H}_0 \dots \mathbf{H}_{M-1} \mathbf{H}_M & \mathbf{0} & \dots & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \dots & \dots & \dots & \vdots \\ \vdots & \mathbf{H}_0 & \dots & \mathbf{H}_{M-1} & \mathbf{H}_M & \mathbf{0} \\ \hline \mathbf{0} & \dots & \mathbf{0} & \mathbf{H}_0 & \dots & \mathbf{H}_{M-1} & \mathbf{H}_M \end{array} \right]}_{\mathbf{K}^{-1}} \underbrace{\left[\begin{array}{c} \mathbf{x}_k \\ \vdots \\ \mathbf{x}_{k-M-N+1} \\ \mathbf{x}_{k-M-N} \end{array} \right]}_{\mathbf{X}_k} \right) + \underbrace{\left[\begin{array}{c} \mathbf{n}_k \\ \mathbf{n}_{k-N+1} \\ \mathbf{n}_{k-N} \end{array} \right]}_{\mathbf{N}_k}$$

$$\mathbf{b}_k = (\mathbf{K}^{-1} - \mathbf{P}^{-1}\mathcal{H})\mathbf{X}_k + \mathbf{H}_M\mathbf{x}_{k-M-N} + \mathbf{n}_k - \mathbf{P}^{-1}\mathbf{N}_k. \quad (5-123)$$

Using the fact that the signal and noise are white and uncorrelated, we can express the mean-square prediction error $\xi = \text{tr}E[\mathbf{b}_k\mathbf{b}_k^*]$ as follows:

$$\xi = \text{tr}[(\mathbf{K}^{-1} - \mathbf{P}^{-1}\mathcal{H})(\mathbf{K}^{-1} - \mathbf{P}^{-1}\mathcal{H})^* + \mathbf{H}_M\mathbf{H}_M^* + \sigma^2\mathbf{I} + \sigma^2\mathbf{P}^{-1}(\mathbf{P}^{-1})^*]. \quad (5-124)$$

We can neglect terms that are independent of \mathbf{P}^{-1} , so it suffices to minimize

$$J(\mathbf{P}^{-1}) = (1/2) \text{tr}[(\mathbf{K}^{-1} - \mathbf{P}^{-1}\mathcal{H})(\mathbf{K}^{-1} - \mathbf{P}^{-1}\mathcal{H})^* + \sigma^2\mathbf{P}^{-1}(\mathbf{P}^{-1})^*]. \quad (5-125)$$

Taking the gradient of $J(\mathbf{P}^{-1})$ and setting it equal to $\mathbf{0}$, we have

$$\nabla J(\mathbf{P}^{-1}) = (\mathbf{P}^{-1}\mathcal{H} - \mathbf{K}^{-1})\mathcal{H}^* + \sigma^2\mathbf{P}^{-1} = \mathbf{P}^{-1}(\mathcal{H}\mathcal{H}^* + \sigma^2\mathbf{I}) - \mathbf{K}^{-1}\mathcal{H}^* = \mathbf{0}. \quad (5-126)$$

Since the Hessian $\nabla^2 J(\mathbf{P}^{-1}) = (\mathcal{H}\mathcal{H}^* + \sigma^2\mathbf{I}) > 0$ is positive definite, the solution of (5-126) minimizes (5-125). The optimal coefficients are thus given by

$$\mathbf{P}^{-1} = [\mathbf{P}_{-N} \dots \mathbf{P}_{-1}] = \mathbf{K}^{-1}\mathcal{H}^*(\mathcal{H}\mathcal{H}^* + \sigma^2\mathbf{I})^{-1} \quad (5-127)$$

$$= \mathbf{K}^{-1}(\mathcal{H}^*\mathcal{H} + \sigma^2\mathbf{I})^{-1}\mathcal{H}^*, \quad (5-128)$$

where $\mathbf{K}^{-1} = [\mathbf{0}_{m \times nN} \ \mathbf{H}_0 \ \dots \ \mathbf{H}_{M-1}]$, and where we have used the identity $\mathcal{H}^*(\mathcal{H}\mathcal{H}^* + \sigma^2\mathbf{I})^{-1} = (\mathcal{H}^*\mathcal{H} + \sigma^2\mathbf{I})^{-1}\mathcal{H}^*$, from Appendix 3-1, in (5-128).

The coefficients of the backward cascade are given by the following convolution representation:

$$\mathbf{B} = [\mathbf{B}_0 \dots \mathbf{B}_{M+N}] = [\mathbf{P}^{-1} \mid \mathbf{I}_m] \mathcal{H}_{N+1} = [(\mathbf{K}^{-1} - \mathbf{P}^{-1}\mathcal{H}) \mid \mathbf{H}_M]. \quad (5-129)$$

Substituting the optimal predictor coefficients of (5-128), we obtain

$$\mathbf{B} = [(\mathbf{K}^{-1} - \mathbf{K}^{-1}(\mathcal{H}^*\mathcal{H} + \sigma^2\mathbf{I})^{-1}\mathcal{H}^*\mathcal{H}) \mid \mathbf{H}_M]. \quad (5-130)$$

Left-factoring $\mathbf{K}^{-1}(\mathcal{H}^*\mathcal{H} + \sigma^2\mathbf{I})^{-1}$ yields

$$\begin{aligned} \mathbf{B} &= [\mathbf{K}^{-1}(\mathcal{H}^*\mathcal{H} + \sigma^2\mathbf{I})^{-1}\{(\mathcal{H}^*\mathcal{H} + \sigma^2\mathbf{I}) - \mathcal{H}^*\mathcal{H}\} \mid \mathbf{H}_M] \\ &= [\mathbf{K}^{-1}\sigma^2(\mathcal{H}^*\mathcal{H} + \sigma^2\mathbf{I})^{-1} \mid \mathbf{H}_M] \\ &= [\mathbf{0}_{m \times nN} \ \mathbf{H}_0 \ \dots \ \mathbf{H}_{M-1}] \Psi \mid \mathbf{H}_M]. \end{aligned} \quad (5-131)$$

The backward prediction-error covariance $\Phi_{\mathbf{b}} = E[\mathbf{b}_k \mathbf{b}_k^*]$ is then given by the following:

$$\Phi_{\mathbf{b}} = \mathbf{H}_M \mathbf{H}_M^* + \sigma^2\mathbf{I} + (\mathbf{K}^{-1} - \mathbf{P}^{-1}\mathcal{H})(\mathbf{K}^{-1} - \mathbf{P}^{-1}\mathcal{H})^* + \sigma^2\mathbf{P}^{-1}(\mathbf{P}^{-1})^*. \quad (5-132)$$

Substituting $\mathbf{K}^{-1}\Psi$ for $(\mathbf{K}^{-1} - \mathbf{P}^{-1}\mathcal{H})$ and $\sigma^{-2}\mathbf{K}^{-1}\Psi\mathcal{H}^*$ for \mathbf{P}^{-1} in (5-132) yields

$$\begin{aligned}
\Phi_{\mathbf{b}} &= \mathbf{H}_M \mathbf{H}_M^* + \sigma^2 \mathbf{I} + \mathbf{K}^{-1} \Psi \Psi (\mathbf{K}^{-1})^* + \sigma^2 \sigma^{-2} \mathbf{K}^{-1} \Psi \mathcal{H}^* \mathcal{H} \Psi (\mathbf{K}^{-1})^* \sigma^{-2} \\
&= \mathbf{H}_M \mathbf{H}_M^* + \sigma^2 \mathbf{I} + \mathbf{K}^{-1} \Psi (\sigma^2 \mathbf{I}) \Psi (\mathbf{K}^{-1})^* \sigma^{-2} + \mathbf{K}^{-1} \Psi (\mathcal{H}^* \mathcal{H}) \Psi (\mathbf{K}^{-1})^* \sigma^{-2} \\
&= \mathbf{H}_M \mathbf{H}_M^* + \sigma^2 \mathbf{I} + \mathbf{K}^{-1} \Psi (\sigma^2 \mathbf{I} + \mathcal{H}^* \mathcal{H}) \Psi (\mathbf{K}^{-1})^* \sigma^{-2} \\
&= \mathbf{H}_M \mathbf{H}_M^* + \sigma^2 \mathbf{I} + \mathbf{K}^{-1} \Psi \Psi^{-1} \Psi (\mathbf{K}^{-1})^* \\
\Phi_{\mathbf{b}} &= \mathbf{H}_M \mathbf{H}_M^* + \sigma^2 \mathbf{I} + \mathbf{K}^{-1} \Psi (\mathbf{K}^{-1})^*. \tag{5-133}
\end{aligned}$$

Using the Cholesky factorization of $\Phi_{\mathbf{b}} = \mathbf{M} \mathbf{D}^2 \mathbf{M}$, we can define a prediction-based whitener according to $\mathbf{W} = \mathbf{D}^{-1} \mathbf{M}^{-1}$. We can then derive the optimal rotator as before. We define $\tilde{\mathbf{B}}$ such that its i -th column is the i -th column of \mathbf{B}_{D_i} , for all $i \in \{1, 2, \dots, n\}$ where D_i is the optimal delay for the i -th user. The optimal rotation $\tilde{\mathbf{Q}}$ then satisfies

$$\mathbf{J} \tilde{\mathbf{Q}} = \mathcal{P}^* (\mathbf{D}^{-1} \mathbf{M}^{-1} \tilde{\mathbf{B}}). \tag{5-134}$$

The backward LP detector is formally defined as follows.

Definition 5-8. For the $m \times n$ channel $\mathbf{H}(z)$ of (1-1) with $m > n$ and $\sigma > 0$, the $n \times m$ backward LP detector of order N is uniquely defined as

$$\mathbf{C}_{BLP}(z) = \mathbf{J} \tilde{\mathbf{Q}} \mathbf{D}^{-1} \mathbf{M}^{-1} [\mathbf{I} - \mathbf{P}^{-1}(z)], \tag{5-135}$$

where $\mathbf{P}^{-1}(z)$ is given by (5-128), where $\mathbf{M} \mathbf{D}^2 \mathbf{M}$ is the Cholesky factorization of $\Phi_{\mathbf{b}}$ given by (5-133), and where $\tilde{\mathbf{Q}}$ satisfies (5-134).

CHAPTER 6

STACKED DETECTORS

TALL CHANNELS have another important property which we have yet to exploit. In addition to the MA, AR, ARMA, and AP models of the previous chapter, tall FIR channels have a *memoryless* representation. By stacking a sufficient number of receiver observations $\mathbf{r}_k = \mathbf{H}_0\mathbf{x}_k + \mathbf{H}_1\mathbf{x}_{k-1} + \dots + \mathbf{H}_M\mathbf{x}_{k-M} + \mathbf{n}_k$ (1-1), we can effectively convert a tall FIR channel $\mathbf{H}(z)$ into a tall memoryless block-Toeplitz channel \mathcal{H} . The detectors of chapters 3 and 4, which are defined in terms of a memoryless channel $\mathbf{r}_k = \mathbf{H}\mathbf{x}_k + \mathbf{n}_k$ (3-1), can thus be generalized to channels with memory. We call these generalizations *stacked detectors*. They can be, but are not always, higher in complexity than the prediction-based detectors of chapter 5; however, they offer better performance. Moreover, the adaptive implementations we propose are more fully blind; they do not need to know or estimate the channel energy distribution, nor even need to know, necessarily, the number of users or the particular constellations being used.

This chapter is organized as follows. In section 6.1, we show that if $\mathbf{H}(z)$ (1-1) is tall, we can use stacking to effectively transform it into a memoryless block-Toeplitz channel \mathcal{H} that is also tall. We then introduce detectors based on this block-Toeplitz model. In section 6.2, we define the *stacked MMSE detector*, which has performance equivalent to a conventional N -tap MMSE detector with the delay optimized for each user. In section 6.3,

we define the *stacked ZF detector*. The minimum-norm realization has performance equivalent to the N -tap MMSE zero-forcing detector, that is, the unique detector, among the class of zero-forcing detectors, that minimizes the MSE for all users. In section 6.4, we introduce the *stacked WR detector*, which has performance near to that of the stacked MMSE detector, but which has lower implementation complexity for some channels. In section 6.5, we define the signal and noise subspaces, and the subspace separator, in terms of \mathcal{H} . In section 6.6, we extend the channel diagonalization scheme of chapter 3 to channels with memory. We show that for almost all tall FIR channels, there exists a lossless precoder of finite complexity that can be used to completely eliminate both ISI and MUI in the receiver, without noise enhancement. In section 6.7, we present blind adaptive implementations of the detectors. In section 6.8, we provide simulation results for an antenna-array application with multipath, and an asynchronous-CDMA application.

6.1 A Memoryless Channel Model

We first transform the channel of (1-1) into an equivalent but higher-dimensional channel without memory. Stacking N consecutive receiver observation vectors \mathbf{r}_k from (1-1) yields

$$\underbrace{\begin{bmatrix} \mathbf{r}_k \\ \mathbf{r}_{k-1} \\ \vdots \\ \mathbf{r}_{k-N+1} \end{bmatrix}}_{\mathbf{R}_k} = \underbrace{\begin{bmatrix} \mathbf{H}_0 & \mathbf{H}_1 & \cdots & \mathbf{H}_M & \mathbf{0} & \cdots & \mathbf{0} \\ \mathbf{0} & \mathbf{H}_0 & \mathbf{H}_1 & \cdots & \mathbf{H}_M & & \vdots \\ \vdots & & \ddots & \ddots & & \ddots & \mathbf{0} \\ \mathbf{0} & \cdots & \mathbf{0} & \mathbf{H}_0 & \mathbf{H}_1 & \cdots & \mathbf{H}_M \end{bmatrix}}_{\mathcal{H}} \times \underbrace{\begin{bmatrix} \mathbf{x}_k \\ \mathbf{x}_{k-1} \\ \vdots \\ \mathbf{x}_{k-N-M+1} \end{bmatrix}}_{\mathbf{X}_k} + \underbrace{\begin{bmatrix} \mathbf{n}_k \\ \mathbf{n}_{k-1} \\ \vdots \\ \mathbf{n}_{k-N+1} \end{bmatrix}}_{\mathbf{N}_k}, \quad (6-1)$$

where \mathbf{R}_k , \mathbf{X}_k , and \mathbf{N}_k are stacked versions of the channel output, input, and noise, respectively, and where \mathcal{H} is an $mN \times n(M+N)$ block-Toeplitz channel matrix. The model of (6-1) can be interpreted as a memoryless system with $n(M+N)$ virtual users and mN virtual sensors. If the original channel $\mathbf{H}(z)$ is strictly tall, then there always exists a sufficient stacking depth $N > \frac{Mn}{m-n}$ such that \mathcal{H} is also tall. Equivalent models are illustrated in Fig. 6-1, in which part (a) shows the FIR channel $\mathbf{H}(z)$ of (1-1) followed by a stacking operation in the receiver, and part (b) shows the memoryless block-Toeplitz channel \mathcal{H} of (6-1) preceded by a stacking operation in the transmitter.

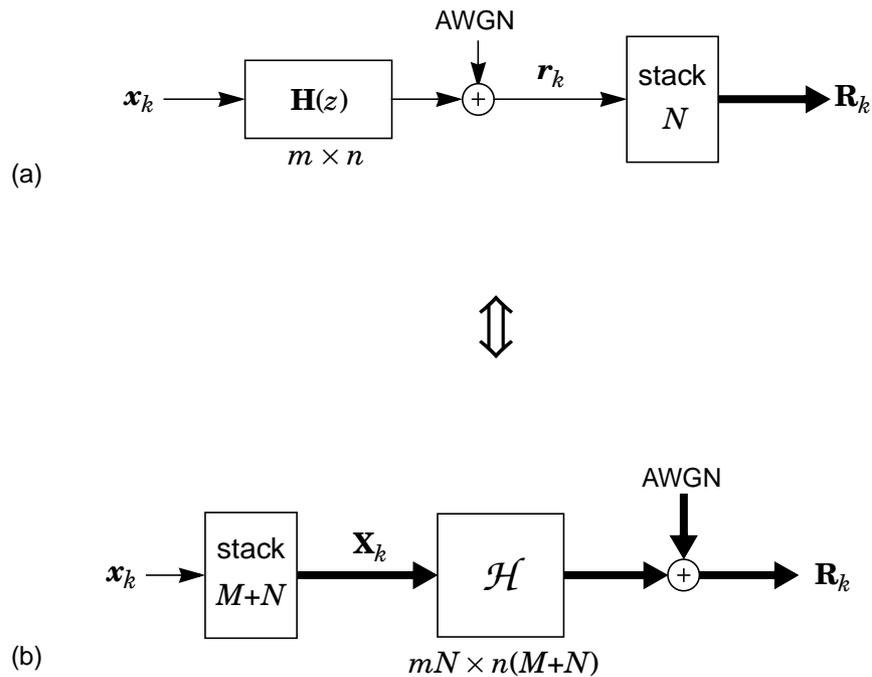


Figure 6-1. Equivalent models: (a) an FIR channel followed by receiver stacking, and (b) a memoryless block-Toeplitz channel preceded by transmitter stacking.

The advantage of (6-1) is that it closely resembles the memoryless model $\mathbf{r}_k = \mathbf{H}\mathbf{x}_k + \mathbf{n}_k$ of (3-1), used throughout chapters 3 and 4. Hence, the concepts and detection strategies developed in those chapters for \mathbf{H} of (3-1) can be applied to \mathcal{H} of (6-1), and therefore implicitly to $\mathbf{H}(z)$ of (1-1). The detectors described in chapters 3 and 4 assume only that \mathbf{H} is full column rank, and that the signal and noise are independent and spatially white. So, likewise, if \mathcal{H} is full column rank, and if the stacked input and noise vectors are independent and spatially white, then the detection strategies of those chapters can be applied to \mathcal{H} . We begin with the assumption that (5-2) holds, implying that \mathcal{H} is full column rank; however, as we show later, this assumption is not essential. We also assume that the signal and noise are independent, zero mean and satisfy the following: $E[\mathbf{x}_k\mathbf{x}_k^*] = \mathbf{I}$, $E[\mathbf{x}_k\mathbf{x}_{k-l}^*] = \mathbf{0}$ for $|l| < M+N$, $E[\mathbf{n}_k\mathbf{n}_k^*] = \sigma^2\mathbf{I}$, where $\sigma > 0$, and $E[\mathbf{n}_k\mathbf{n}_{k-l}^*] = \mathbf{0}$ for $|l| < N$. This signal and noise can be white, but this is not necessary. We make no assumptions regarding the autocorrelation of the signal or noise outside the given range of lags l .

6.2 The Stacked MMSE Detector

Let $\mathbf{Z}_k = C\mathbf{R}_k$ be an estimate of the stacked channel input \mathbf{X}_k , where C has dimension $n(N+M) \times mN$. We define the stacked MMSE detector as the detector that minimizes the total mean-square error between \mathbf{Z}_k and \mathbf{X}_k .

Definition 6-1. The **stacked MMSE detector** C_{MMSE} for (6-1) is the $n(M+N) \times mN$ matrix C that minimizes $E[\|C\mathbf{R}_k - \mathbf{X}_k\|^2]$.

In a manner analogous to Lemma 3-4, the stacked MMSE detector can be expressed in terms of \mathcal{H} as follows.

Lemma 6-1. For the channel \mathcal{H} of (6-1) with $\sigma > 0$, the unique stacked MMSE detector can be expressed in three equivalent ways.

$$C_{MMSE} = \mathcal{H}^* (\mathcal{H} \mathcal{H}^* + \sigma^2 \mathbf{I})^{-1} \quad (6-2)$$

$$= (\mathcal{H}^* \mathcal{H} + \sigma^2 \mathbf{I})^{-1} \mathcal{H}^* \quad (6-3)$$

$$= \mathcal{V} \mathcal{S}^* (\mathcal{S} \mathcal{S}^* + \sigma^2 \mathbf{I})^{-2} \mathcal{U}^*, \quad (6-4)$$

where $\mathcal{H} = \mathcal{U} \mathcal{S} \mathcal{V}^*$ is a channel SVD.

Unlike the channel \mathcal{H} , the detector C_{MMSE} is not block-Toeplitz.

We can relate the stacked MMSE detector to a set of conventional MMSE detectors as follows. Partition C_{MMSE} into $M+N$ block rows of dimension $n \times mN$:

$$C_{MMSE} = \begin{bmatrix} \mathbf{C}_0 \\ \mathbf{C}_1 \\ \vdots \\ \mathbf{C}_{M+N-1} \end{bmatrix}. \quad (6-5)$$

Let \mathbf{C}_D denote the D -th block row of C_{MMSE} , and let $\mathbf{z}_k^{(D)}$ denote the corresponding D -th block row of the output \mathbf{Z}_k , so that

$$\mathbf{Z}_k = \begin{bmatrix} \mathbf{z}_k^{(0)} \\ \mathbf{z}_k^{(1)} \\ \vdots \\ \mathbf{z}_k^{(M+N-1)} \end{bmatrix}. \quad (6-6)$$

Observe that the total mean-square error $E[\|\mathbf{Z}_k - \mathbf{X}_k\|^2]$ can be expressed as a sum:

$$\begin{aligned}
E[\|\mathbf{Z}_k - \mathbf{X}_k\|^2] &= \sum_{D=0}^{M+N-1} E[\|\mathbf{z}_k^{(D)} - \mathbf{x}_{k-D}\|^2] \\
&= \sum_{D=0}^{M+N-1} E[\|\mathbf{C}_D \mathbf{R}_k - \mathbf{x}_{k-D}\|^2]. \tag{6-7}
\end{aligned}$$

Therefore, the D -th block row of the \mathbf{C} that minimizes $E[\|\mathbf{Z}_k - \mathbf{X}_k\|^2]$ is the $n \times mN$ matrix \mathbf{C}_D that minimizes $E[\|\mathbf{C}_D \mathbf{R}_k - \mathbf{x}_{k-D}\|^2]$. So $\mathbf{z}_k^{(D)} = \mathbf{C}_D \mathbf{R}_k = \sum_{j=0}^{N-1} \mathbf{C}_{D,j} \mathbf{r}_{k-j}$, where the coefficients $\mathbf{C}_{D,j}$ have dimension $n \times m$. It follows that $\mathbf{z}_k^{(D)}$ is the output of an N -tap filter with transfer function $\mathbf{C}_D(z) = \sum_{j=0}^{N-1} \mathbf{C}_{D,j} z^{-j}$.

Theorem 6-1. The stacked MMSE detector \mathbf{C}_{MMSE} of Definition 6-1 simultaneously implements $M+N$ different conventional N -tap MMSE multiuser detectors, $\mathbf{C}_0(z)$ through $\mathbf{C}_{M+N-1}(z)$, corresponding to decision delays $D = 0$ through $M+N-1$.

Therefore, if we stack N observation vectors and implement the stacked MMSE detector \mathbf{C}_{MMSE} of Lemma 6-1, then the D -th block row of its output $\mathbf{z}_k^{(D)}$ is the output of an N -tap MMSE detector $\mathbf{C}_D(z)$, minimizing the D -delay MSE of all users $E[\|\mathbf{z}_k^{(D)} - \mathbf{x}_{k-D}\|^2]$.

Of course a decision device must follow \mathbf{C}_{MMSE} . Our approach is to use a bank of $n(N+M)$ independent decision devices, one for each output, and to choose for each user the associated output with smallest mean slicer error, as illustrated in Fig. 6-2. This approach is equivalent to implementing a conventional $n \times m$, N -tap MMSE multiuser detector with the delay optimized for each user.

6.3 The Stacked Zero-Forcing Detector

As we discussed in chapter 3, the MMSE linear multiuser detector does not completely eliminate interference, but rather finds the best compromise between interference and noise. Complete elimination of interference is accomplished with a zero-forcing detector.

Definition 6-2. For the channel \mathcal{H} of (6-1), a **stacked zero-forcing detector** C_{ZF} is an $n(M+N) \times mN$ matrix satisfying $C\mathcal{H} = \mathbf{I}$.

Extending Lemma 3-5 to (6-1), we can express the stacked ZF detector in terms of the block-Toeplitz channel matrix \mathcal{H} and its SVD as follows.

Lemma 6-2. For the channel \mathcal{H} of (6-1), a stacked ZF detector can be expressed in two equivalent ways:

$$C_{ZF} = \mathcal{H}^\dagger + \mathcal{N} \quad (6-8)$$

$$= \mathcal{V}\mathcal{S}^\dagger\mathcal{U}^* + \mathcal{N}, \quad (6-9)$$

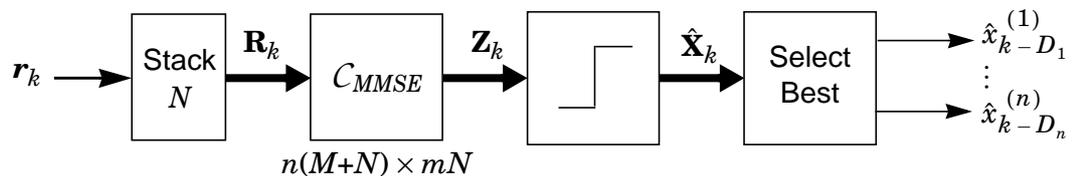


Figure 6-2. A block diagram of the stacked MMSE detector.

where $\mathcal{H} = \mathcal{U}\mathcal{S}\mathcal{V}^*$, and where $\mathcal{N}^* \in \text{null}(\mathcal{H}^*)$. The stacked ZF detector is unique if and only if $mN = n(M+N)$.

We can relate the stacked ZF detector to the conventional ZF detector as follows. Once again partition the detector C_{ZF} into $M+N$ block rows of dimension $n \times mN$, and let \mathbf{C}_D denote the D -th block row of the detector. The D -th block row of the detector output can then be expressed as $\mathbf{z}_k^{(D)} = \mathbf{C}_D \mathbf{R}_k = \sum_{j=0}^{N-1} \mathbf{C}_{D,j} \mathbf{r}_{k-j}$, where the coefficients $\mathbf{C}_{D,j}$ again have dimension $n \times m$. By definition C_{ZF} completely eliminates both ISI and MUI, so $\mathbf{Z}_k = C_{ZF} \mathbf{R}_k = \mathbf{X}_k + C_{ZF} \mathbf{N}_k$. It follows that $\mathbf{z}_k^{(D)} = \mathbf{x}_{k-D} + \mathbf{C}_D \mathbf{N}_k$ is the output of a conventional N -tap D -delay zero-forcing detector: $\mathbf{C}_D(z) = \sum_{j=0}^{N-1} \mathbf{C}_{D,j} z^{-j}$.

Theorem 6-2. The stacked-ZF detector simultaneously implements $M+N$ different conventional delayed zero-forcing multiuser detectors, $\mathbf{C}_0(z)$ through $\mathbf{C}_{M+N-1}(z)$, corresponding to decision delays of $D = 0$ through $D = M+N-1$.

The MSE sum corresponding to $\mathbf{C}_D(z)$ is proportional to the norm of \mathbf{C}_D :

$$\begin{aligned} E[\|\mathbf{z}_k^{(D)} - \mathbf{x}_{k-D}\|^2] &= E\left\|\sum_{j=0}^{N-1} \mathbf{C}_{D,j} \mathbf{r}_{k-j} - \mathbf{x}_{k-D}\right\|^2 \\ &= E\left\|\sum_{j=0}^{N-1} \mathbf{C}_{D,j} \mathbf{n}_{k-j}\right\|^2 \\ &= \sigma^2 \|\mathbf{C}_D\|_F^2. \end{aligned} \tag{6-10}$$

Clearly, $\|\mathbf{C}_D\|_F^2$ is a function of D , so these lower-dimensional ZF detectors do not have the same MSE performance.

There is indeed an optimal N -tap ZF detector that minimizes the MSE of all users subject to a zero-interference constraint.

Definition 6-3. For the channel of (1-1), let $\mathbf{C}(z) = \sum_{j=0}^{N-1} \mathbf{C}_j z^{-j}$ denote an $n \times m$, N -tap detector with output \mathbf{z}_k . Let $\xi = E[\|\mathbf{z}_k - \tilde{\mathbf{x}}_k\|^2]$ denote the total MSE, where $\tilde{\mathbf{x}}_k = [x_{k-D_1}^{(1)}, x_{k-D_2}^{(2)}, \dots, x_{k-D_n}^{(n)}]^T$ for delays $D_i \in \{0 \dots M+N-1\}$, $i \in \{1, \dots, n\}$. The **minimum-MSE zero-forcing detector** $\mathbf{C}_{MZF}(z)$ is the filter that minimizes ξ over all possible delays subject to the constraint that $\mathbf{z}_k = \tilde{\mathbf{x}}_k + \sum_{j=0}^{N-1} \mathbf{C}_j \mathbf{n}_{k-j}$.

It follows from Theorem 6-2 that if we implement the stacked ZF detector of Lemma 6-2, followed by a bank of independent slicers, and once again, choose for each user, the associated output with smallest mean slicer error, we, in effect, implement $\mathbf{C}_{MZF}(z)$ of Definition 6-3. Later, in Experiments 6-1 and 6-2, we compare the MSE performance of the stacked MMSE and ZF detectors. However, we first define one additional stacked detector.

6.4 The Stacked Whiten-Rotate Detector

We can also generalize the whiten-rotate detector of chapter 4 to channels with memory. In the context of (6-1), an $n(N+M) \times mN$ matrix C is said to be a whitener if the covariance of $\mathbf{Z}_k = C \mathbf{R}_k$ is the identity matrix, $C \Phi_{\mathbf{R}} C^* = \mathbf{I}$, where $\Phi_{\mathbf{R}} = E(\mathbf{R}_k \mathbf{R}_k^*)$. We define the stacked whiten-rotate detector as the whitener with minimal MSE.

Definition 6-4. The **canonical stacked whiten-rotate detector** C_{WR} for (6-1) is the $n(N+M) \times mN$ whitener that minimizes the MSE sum $E[\|C \mathbf{R}_k - \mathbf{X}_k\|^2]$.

Using arguments similar to those in section 4.1, it follows that for any $mN \times mN$ whitener \mathcal{W} , satisfying $\mathcal{W}\Phi_{\mathbf{R}}\mathcal{W}^* = \mathbf{I}$, the stacked whiten-rotate detector can be expressed as $\mathcal{J}Q\mathcal{W}$, where

$$Q = \begin{bmatrix} \bar{\mathcal{V}} & \mathbf{0} \\ \mathbf{0} & \mathcal{V}_N \end{bmatrix} \bar{\mathbf{u}}^*, \quad (6-11)$$

where $\bar{\mathbf{u}}$ and $\bar{\mathcal{V}}$ are factors in a SVD of $\mathcal{W}\mathcal{H} = \bar{\mathbf{u}}\bar{\mathcal{S}}\bar{\mathcal{V}}^*$, and where $\mathcal{J} = [\mathbf{I} \ \mathbf{0}]$ is a truncation matrix of dimension $n(N+M) \times mN$.

One such whitener \mathcal{W} , based on linear prediction, is given by

$$\mathcal{W} = \mathcal{D}^{-1}\mathcal{M}^{-1}, \quad (6-12)$$

where $\Phi_{\mathbf{R}} = \mathcal{M}\mathcal{D}^2\mathcal{M}^*$ is a Cholesky factorization. (See Lemma 4-3 and Theorem 4-2.) Although we are free to use any whitener \mathcal{W} , the whitener of (6-12) reveals an interesting relationship between the stacked WR detector and the prediction-based detectors of chapter 5. Partition $\mathcal{J}Q$ into $M+N$ block rows of dimension $n \times mN$, and partition \mathcal{W} of (6-12) into N block rows of dimension $m \times mN$ as follows:

$$\mathcal{J}Q = \begin{bmatrix} \mathbf{Q}_0 \\ \mathbf{Q}_1 \\ \vdots \\ \mathbf{Q}_{M+N-1} \end{bmatrix}, \quad \mathcal{W} = \begin{bmatrix} \mathbf{W}_0 \\ \mathbf{W}_1 \\ \vdots \\ \mathbf{W}_{N-1} \end{bmatrix}. \quad (6-13)$$

Let $\mathbf{V}_k = \mathcal{W}\mathbf{R}_k$ and $\mathbf{Z}_k = \mathcal{J}Q\mathbf{V}_k$ denote the outputs of the whitener and the detector, respectively. Let $\mathbf{v}_k^{(i)}$ denote the i -th block row of \mathbf{V}_k , and let $\mathbf{z}_k^{(D)}$ denote the D -th block row of \mathbf{Z}_k , so that

$$\mathbf{V}_k = \begin{bmatrix} \mathbf{v}_k^{(0)} \\ \mathbf{v}_k^{(1)} \\ \vdots \\ \mathbf{v}_k^{(N-1)} \end{bmatrix}, \quad \mathbf{Z}_k = \begin{bmatrix} \mathbf{z}_k^{(0)} \\ \mathbf{z}_k^{(1)} \\ \vdots \\ \mathbf{z}_k^{(M+N-1)} \end{bmatrix}. \quad (6-14)$$

Observe that

$$\mathbf{v}_k^{(i)} = \mathbf{W}_i \mathbf{R}_k = \sum_{j=0}^{N-1} \mathbf{W}_{i,j} \mathbf{r}_{k-j} = \sum_{j=0}^i \mathbf{W}_{i,j} \mathbf{r}_{k-j}, \quad (6-15)$$

where the coefficients $\mathbf{W}_{i,j}$ have dimension $m \times m$. Because \mathcal{W} is lower triangular and monic, so is $\mathbf{W}_{i,i}$. Moreover, $\mathbf{W}_{i,j} = \mathbf{0}$ for $j > i$; so, the last summation in (6-15) does not include these terms. Thus $\mathbf{v}_k^{(i)}$ is the scaled prediction error from a spatio-temporal prediction-error filter. Specifically, $\mathbf{W}_i(z)$ is the cascade of a one-step backward temporal prediction-error filter of order i , a lower-triangular spatial prediction-error filter, and a diagonal gain:

$$\begin{aligned} \mathbf{W}_i(z) &= \sum_{j=0}^i \mathbf{W}_{i,j} z^{-j} = \mathbf{A}_i (\mathbf{I} - \sum_{j=0}^i \tilde{\mathbf{P}}_{-j}^{(i)} z^{-j}) \\ &= \mathbf{A}_i (\mathbf{I} - \tilde{\mathbf{P}}_0^{(i)}) (\mathbf{I} - \sum_{j=1}^i \mathbf{P}_{-j}^{(i)} z^{-j}), \end{aligned} \quad (6-16)$$

where $\mathbf{A}_i = \text{diag}(\mathbf{W}_{i,i})$, and where $\tilde{\mathbf{P}}_0^{(i)}$ is strictly lower triangular. Let $\mathbf{e}_k^{(i)} = \mathbf{r}_k - \sum_{j=0}^i \tilde{\mathbf{P}}_{-j}^{(i)} \mathbf{r}_{k-j}$ denote the prediction error of the i -th spatio-temporal prediction error filter $\mathbf{I} - \sum_{j=0}^i \tilde{\mathbf{P}}_{-j}^{(i)} z^{-j}$, and define a stacked error as

$$\mathbf{E}_k = \begin{bmatrix} \mathbf{e}_k^{(0)} \\ \mathbf{e}_k^{(1)} \\ \vdots \\ \mathbf{e}_k^{(N-1)} \end{bmatrix}. \quad (6-17)$$

Observe that \mathcal{M}^{-1} minimizes $E[\|\mathbf{E}_k\|^2]$, and thus $E[\|\mathbf{e}_k^{(i)}\|^2]$ for all $i \in \{0 \dots N-1\}$, so it follows that the coefficients $\tilde{\mathbf{P}}_{-j}^{(i)}$ in (6-16) also minimize $\xi = E[\|\mathbf{e}_k^{(i)}\|^2]$. The D -th block row of the detector output $\mathbf{z}_k^{(D)}$, an estimate of \mathbf{x}_{k-D} , is therefore a linear combination of N *optimal* prediction errors:

$$\mathbf{z}_k^{(D)} = \mathbf{Q}_D \mathbf{V}_k = \sum_{i=0}^{N-1} \mathbf{Q}_{D,i} \mathbf{A}_i \mathbf{e}_k^{(i)}, \quad (6-18)$$

where the coefficients $\mathbf{Q}_{D,j}$ have dimension $n \times m$. We thus arrive at the following theorem, which is also represented graphically in Fig. 6-3.

Theorem 6-3. The stacked WR detector simultaneously generates estimates $\mathbf{z}_k^{(D)}$ of \mathbf{x}_{k-D} for all delays $D \in \{0 \dots M+N-1\}$ by using linear combinations of the errors from N optimal spatio-temporal prediction error filters of the form $\mathbf{I} - \sum_{j=0}^i \tilde{\mathbf{P}}_{-j}^{(i)} z^{-j}$, for $i = 0$ through $i = N-1$.

The architecture of Fig. 6-3 is not proposed as an efficient implementation. Nevertheless, Theorem 6-3 has pedagogical importance. It says that spatio-temporal linear prediction-based detection can be used to approach MMSE detection for any delay, not just for delay zero as shown by Lemma 5-5 of the previous chapter. The architecture of Fig. 6-3 illustrates this important connection with the detectors of chapter 5.

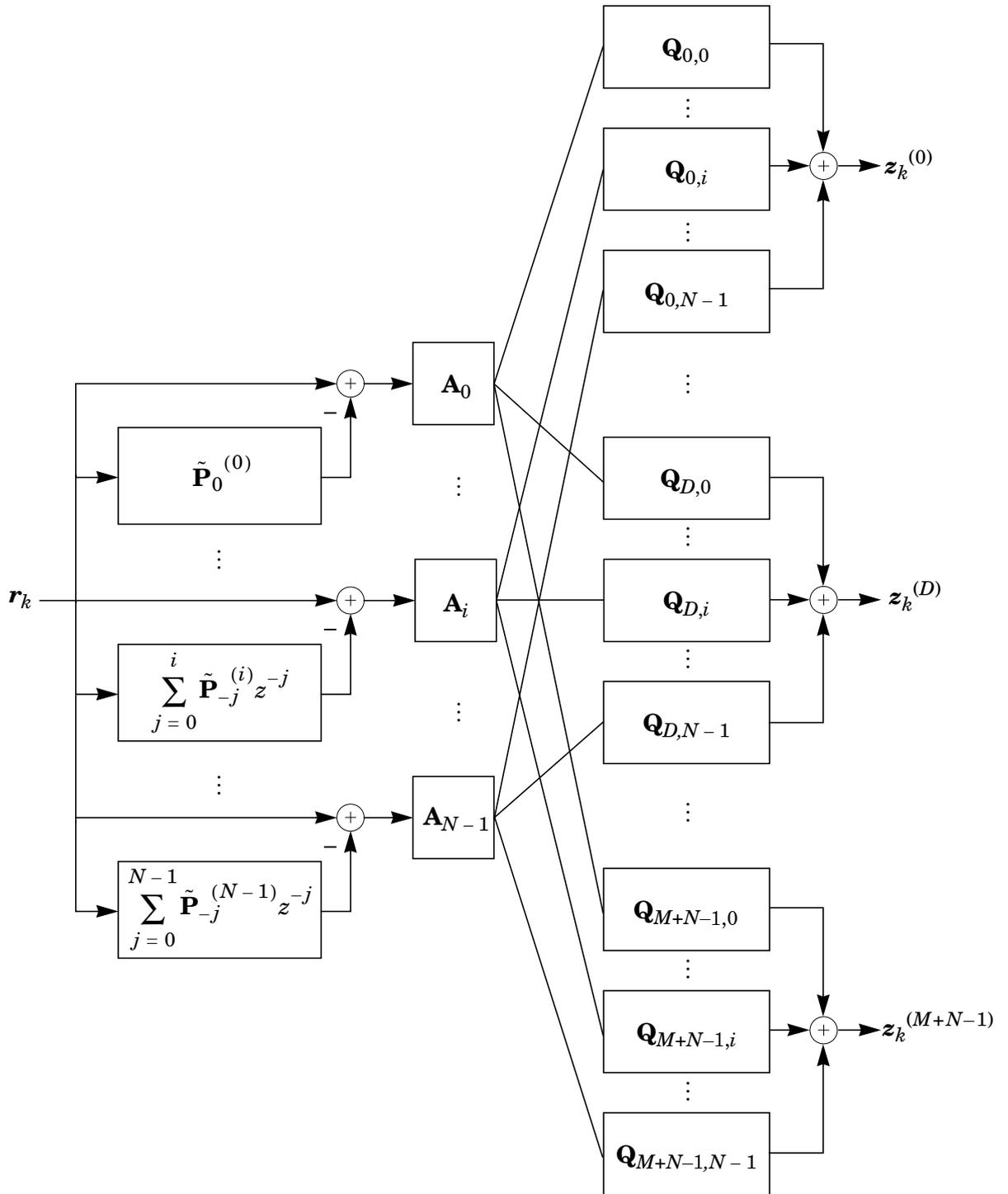


Figure 6-3. An interpretation of the stacked WR detector showing its relationship to the spatio-temporal prediction-based detectors.

Most of the theoretical development of chapter 4 pertaining to the whiten-rotate detector \mathbf{C}_{WR} of Definition 4-1, is equally valid for the stacked whiten-rotate detector C_{WR} of Definition 6-4. We need only to make appropriate substitutions: \mathbf{X}_k for \mathbf{x}_k , \mathcal{H} for \mathbf{H} , etc. Lemma 4-1, for example, which expresses the WR detector in terms of a channel SVD holds also for the stacked WR detector. The properties of the WR detector hold for the stacked WR detector as well. Results in chapter 4 that apply to a particular user i , such as Lemma 4-2 or Property 4-4, apply to a particular virtual user in the context of (6-1), where user i at delay D corresponds to virtual user $i + nD$.

It is beneficial to reformulate Lemma 4-2, in particular, to express the MSE of the stacked detectors in terms of real users and delays. Let \mathbf{Z}_k denote the output of a stacked detector, either MMSE, ZF, or WR, such that $Z_k^{(i+nD)}$ is an estimate of $x_{k-D}^{(i)}$, and define $\text{MSE}_{i,D} = E[|Z_k^{(i+nD)} - x_{k-D}^{(i)}|^2]$ as the D -delay MSE for user i .

Lemma 6-3. The D -delay MSE for user i , denoted $\text{MSE}_{i,D}$, of the stacked MMSE, ZF, and WR detectors, respectively, can be expressed as

$$\text{MSE}_{D,i}^{MMSE} = \sigma^2 v_{i+nD}^* (\tilde{\mathcal{S}}^2 + \sigma^2 \mathbf{I})^{-1} v_{i+nD}, \quad (6-19)$$

$$\text{MSE}_{D,i}^{ZF} = \sigma^2 v_{i+nD}^* \tilde{\mathcal{S}}^{-1} v_{i+nD}, \quad (6-20)$$

$$\text{MSE}_{D,i}^{WR} = 2 v_{i+nD}^* [\mathbf{I} - (\tilde{\mathcal{S}}^2 + \sigma^2 \mathbf{I})^{-1/2} \tilde{\mathcal{S}}] v_{i+nD}, \quad (6-21)$$

where v_j is the j -th column of \mathcal{V}^* , from $\mathcal{H} = \mathcal{U}\mathcal{S}\mathcal{V}^*$, and where $\tilde{\mathcal{S}} = \mathcal{J}\mathcal{S}$.

In the following experiments, we use Lemma 6-3 to compare the performance of the stacked detectors in terms of MSE. We pay particular attention to the performance difference between the stacked WR and MMSE detectors.

Experiment 6-1. Consider a system with $n = 2$ users and $m = 10$ sensors having channel memory $M = 5$. Using a stacking depth of $N = 2$, we compare the mean-square errors of the stacked MMSE, ZF, and WR detectors for both users at all delays 0 through $M+N-1 = 8$. We consider only a single, randomly generated channel, where the elements of the each tap are drawn independently from a zero-mean, unit-variance, Gaussian distribution, and are then scaled such that $\text{SNR}_1 = 20$ dB and $\text{SNR}_2 = 10$ dB. Fig. 6-4 shows that the WR detector suffers less than a 1-dB penalty relative to the MMSE detector, regardless of delay or user. It also shows that, for all detector types, the optimal delay for each user can be dramatically different.

Experiment 6-2. Consider now a system with $n = 3$ users and $m = 30$ sensors having channel memory $M = 3$. Using a stacking depth of $N = 6$, we compare the mean-square errors of the stacked detectors for each user at all delays 0 through $M+N-1 = 8$. We consider an ensemble of 10 random channels, generated as in the previous experiment except that the SNRs for users 1 through 3 were set at 20 dB, 10 dB, and 0 dB, respectively. In Fig. 6-5, we see that on average the optimal delay is $D = 4$, independent of detector type or user. We find that, at this delay, the average MSE penalties incurred by using the stacked WR detector instead of the stacked MMSE detector are 0.0343 dB, 0.1332 dB, and 0.7513 dB for users 1, 2, and 3, respectively — less than 1 dB in all cases.

The stacking depth N should be chosen to balance performance and complexity. At a minimum, we must have $N > \frac{Mn}{m-n}$ in order for \mathcal{H} to be tall. However, larger values of N result in better performance, especially at low SNR. In Experiment 6-1 we use the min-

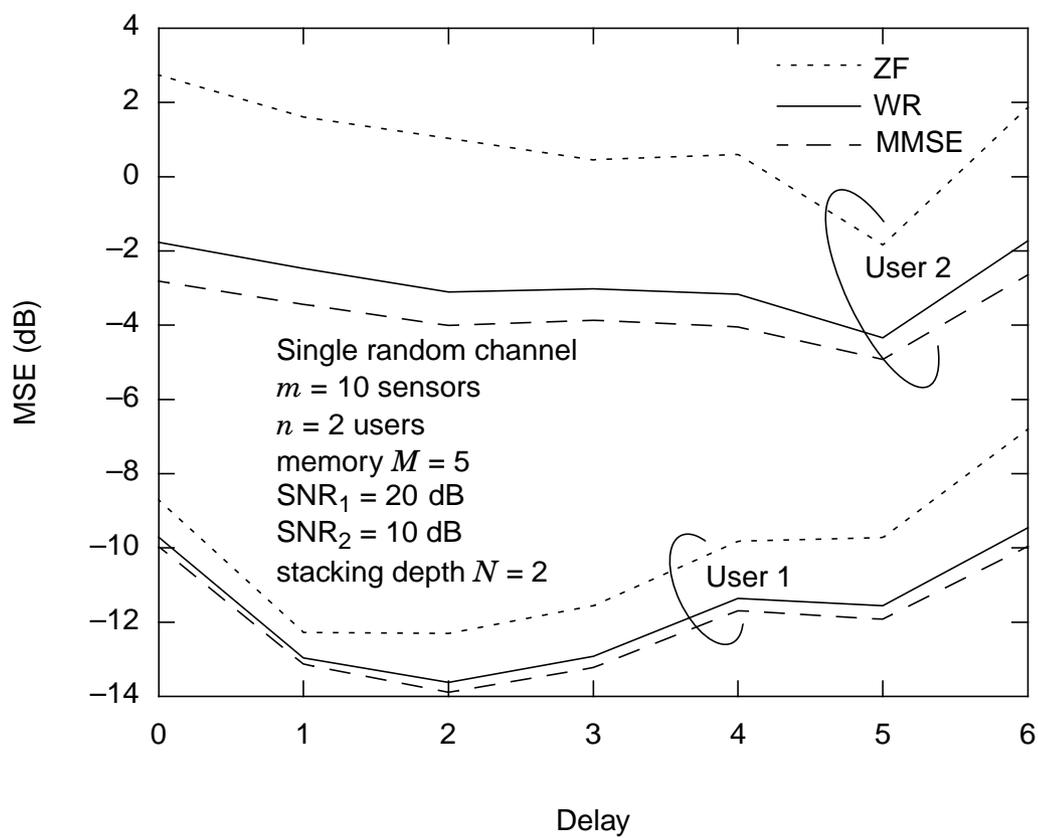


Figure 6-4. Comparison of the stacked detectors for a single random channel.

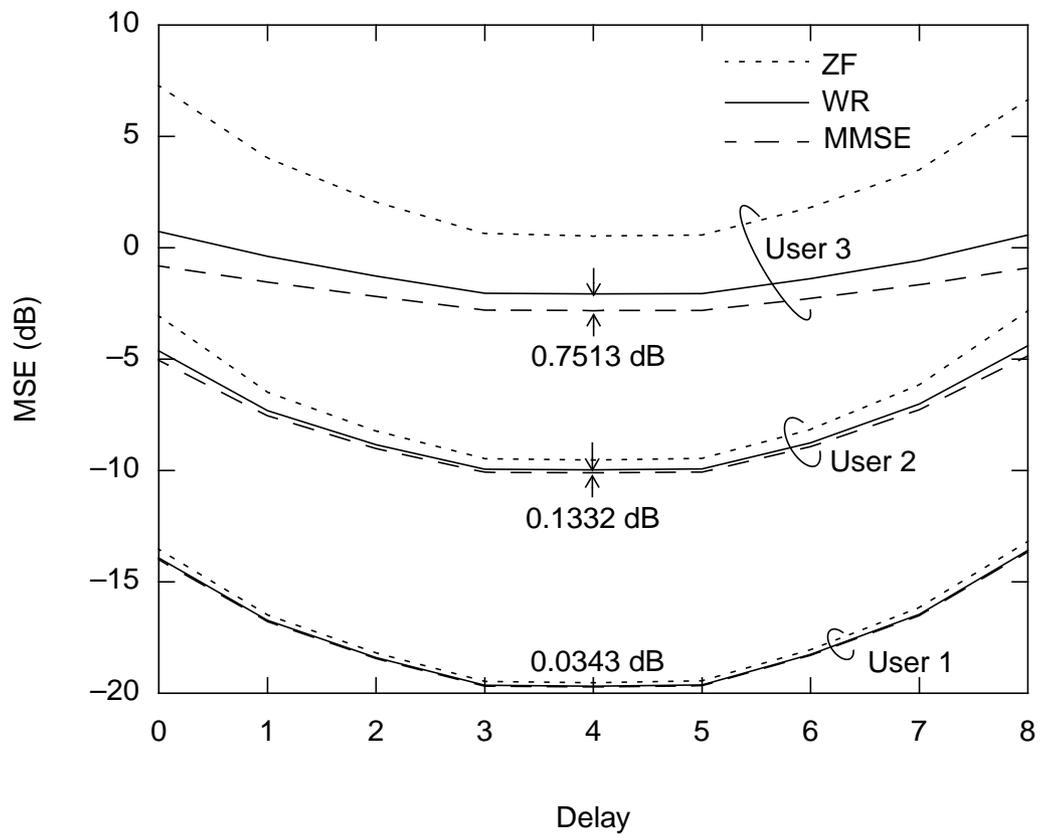


Figure 6-5. Comparison of the stacked detectors: MSE versus delay.

imum allowable stacking depth of $N = 2$, whereas in Experiment 6-2, we use a depth of $N = 6$, substantially larger than the necessary minimum of 1. Using larger values of N produces better performance, but eventually with diminishing returns. It is evident from Fig. 6-4 that a stacking depth of $N = 6$ is sufficient for Experiment 6-2, because the optimal MSE for user i is roughly SNR_i^{-1} . In Fig. 6-5, however, the optimal MSE for user i is significantly greater than SNR_i^{-1} , indicating that a larger stacking depth would have been beneficial in this case. The following experiment illustrates the performance of the stacked detectors as a function of the stacking depth N .

Experiment 6-3. We again use random 4-tap ($M = 3$) channels of dimension 30×3 , generated as in Experiment 6-2, except that we vary the stacking depth from $N = 1$ to $N = 10$. We conduct 10 trials for each value of N and record the MSE at the optimal delay for each user and for each detector type. Observe that Fig. 6-6 verifies that use of larger values of N improves performance, but with diminishing returns for $N > 5$.

In the next experiment, we analyze more closely the MSE penalty of the stacked WR detector relative to the stacked MMSE detector. Hows the ensemble average of these penalties. Since the optimal delay is not always 4, the resulting penalty is slightly smaller than those at delay 4, reported in Experiment 6-2.

Experiment 6-4. We again use random 4-tap ($M = 3$) channels of dimension 30×3 , generated as in Experiment 6-2, except that we vary the SNRs of the users; SNR_1 ranges from 0 to 20 dB, while SNR_2 and SNR_3 are 10 dB and 20 dB greater than SNR_1 . We conduct 10 trials at each SNR point and measure the MSE penalty of the stacked WR detector relative to the stacked MMSE for each user at the optimal

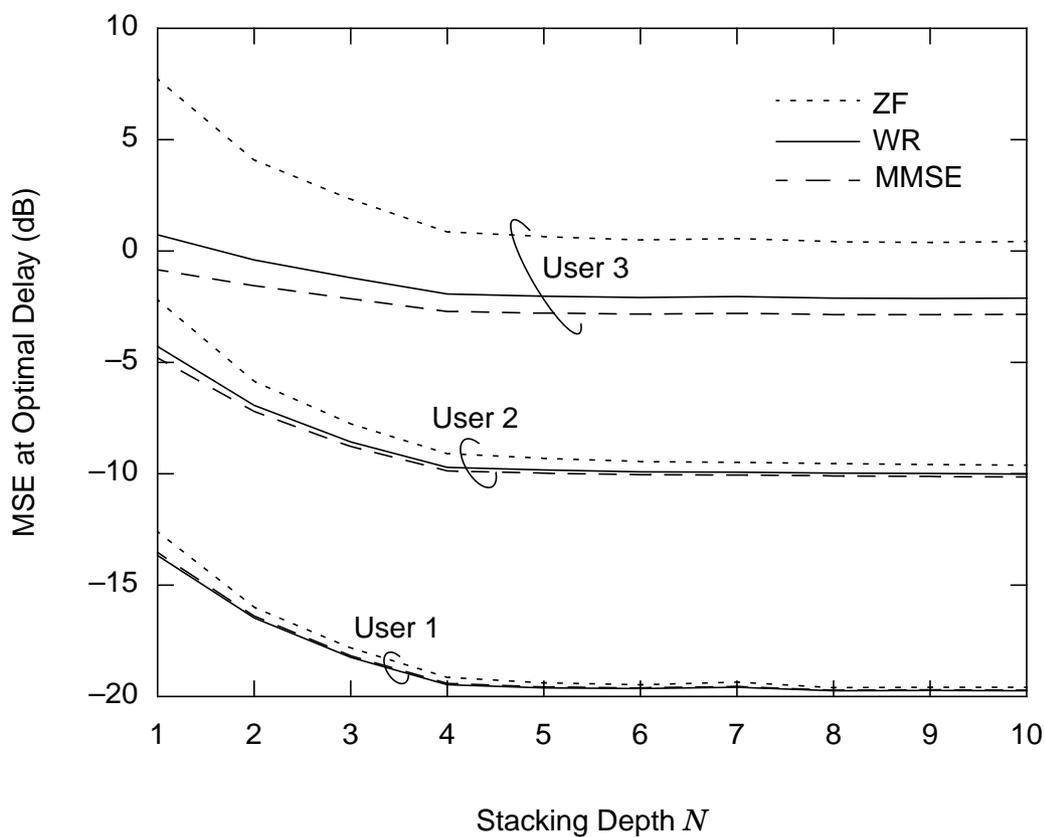


Figure 6-6. Performance of the stacked detectors as a function of the stacking depth N .

delay. Fig. 6-7 shows the ensemble average of these penalties. Since the optimal delay is not always 4, the resulting penalty is slightly smaller than those at delay 4, reported in Experiment 6-2.

We conclude that the stacked WR detector, followed by a bank of independent slicers, performs nearly as well as the stacked MMSE detector, especially for very tall channels or high SNRs, and thus is nearly equivalent to a conventional MMSE detector with the delay optimized for each user.

6.5 Signal and Noise Subspaces

We now define the signal and noise subspaces, and the subspace separator in the context of (6-1). We need only substitute \mathcal{H} for \mathbf{H} in Definition 3-1.

Definition 6-5. For the block-Toeplitz channel \mathcal{H} of (6-1), the **signal subspace** is the range or column span of \mathcal{H} : $\mathbf{S} = \text{range}(\mathcal{H})$; the **noise subspace** is the left null space of \mathcal{H} : $\mathbf{N} = \mathbf{S}^\perp = \text{null}(\mathcal{H}^*)$.

The dimensionality of the signal subspace is equal to the rank of \mathcal{H} , which is $n(M+N)$; the dimensionality of the noise subspace is $mN - n(M+N)$. A subspace-separation matrix is easily defined as in Definition 3-1, but with $n(M+N)$ and mN playing the roles of n and m , respectively.

Definition 6-6. For the $mN \times n(M+N)$ channel \mathcal{H} of (6-1), an $mN \times mN$ unitary matrix Θ is a **subspace-separation matrix** if and only if the last $mN - n(M+N)$ rows of $\Theta\mathcal{H}$ are identically zero.

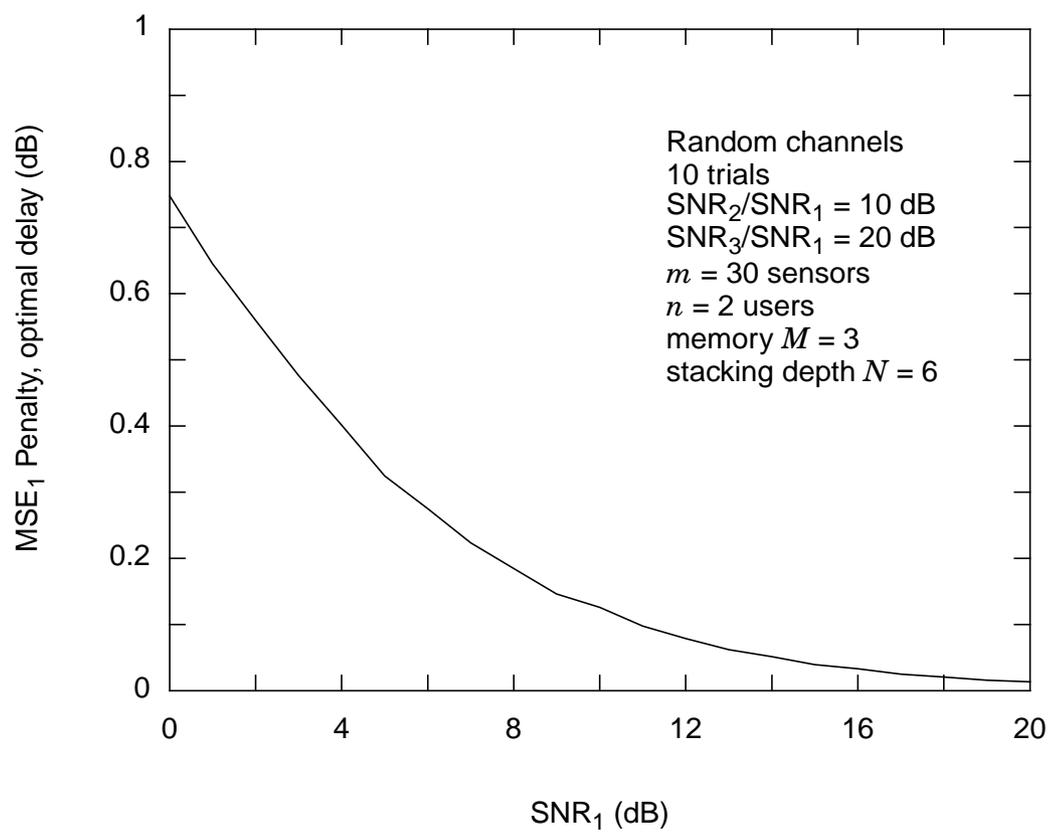


Figure 6-7. Average MSE penalty of the stacked WR detector relative to the stacked MMSE detector.

Lemma 3-3, which expresses the mathematical form of a subspace-separation matrix in terms of the left factor of a channel SVD, is easily extended.

Lemma 6-4. For \mathcal{H} of (6-1), a unitary subspace-separation matrix must be of the form:

$$\Theta = \begin{bmatrix} \mathcal{U}_S & \mathbf{0} \\ \mathbf{0} & \mathcal{U}_N \end{bmatrix} \mathcal{U}^*, \quad (6-22)$$

where \mathcal{U}_S and \mathcal{U}_N are arbitrary unitary matrices of dimension $n(M+N)$ and $mN - n(M+N)$, respectively, and where \mathcal{U} is the left factor of any channel SVD $\mathcal{H} = \mathcal{U}\mathcal{S}\mathcal{V}^*$.

Following the development of section 3.1, we see that if Θ satisfies (6-22), then the last $mN - n(M+N)$ components of $\mathbf{Y}_k = \Theta \mathbf{R}_k$ contain no signal energy:

$$\mathbf{Y}_k = \begin{bmatrix} \tilde{\mathcal{H}} \\ \mathbf{0} \end{bmatrix} \mathbf{X}_k + \Theta \mathbf{N}_k, \quad (6-23)$$

where $\tilde{\mathcal{H}} = \mathcal{U}_S \tilde{\mathcal{S}} \mathcal{V}^*$ is of dimension $n(M+N) \times n(M+N)$. Discarding these noise-only components effectively produces a square channel model:

$$\tilde{\mathbf{R}}_k = \mathcal{J} \mathbf{Y}_k = \tilde{\mathcal{H}} \mathbf{X}_k + \tilde{\mathbf{N}}_k, \quad (6-24)$$

where $E[\tilde{\mathbf{N}}_k \tilde{\mathbf{N}}_k^*] = \sigma^2 \mathbf{I}$, and where $\tilde{\mathcal{H}}$, unlike \mathcal{H} , is not block-Toeplitz. The $n(M+N)$ components of $\tilde{\mathbf{R}}_k$ are sufficient for estimating \mathbf{X}_k . In particular, Theorem 4-1 extends, which we restate here in the context of (6-1).

Theorem 6-4. The cascade of a signal-subspace projector $\mathcal{J}\Theta$ and a stacked detector $\tilde{\mathcal{C}}$, either MMSE, ZF, or WR, designed for the reduced channel $\tilde{\mathcal{H}}$, precisely implements the stacked detector \mathcal{C} designed for the original block-Toeplitz channel \mathcal{H} .

Proof: The proof is identical to that of Theorem 4-1.

6.6 Channel Diagonalization and Lossless Precoding

We can also extend the channel diagonalization idea of chapter 3 to channels with memory. Consider again, for example, a single-user array-to-array communication problem with n transmit antennas and m receive antennas, where $m > n$. Such a system can be modeled by $\mathbf{r}_k = \mathbf{H}_0\mathbf{x}_k + \mathbf{H}_1\mathbf{x}_{k-1} + \dots + \mathbf{H}_M\mathbf{x}_{k-M} + \mathbf{n}_k$ of (1-1). The i -th component $r_k^{(i)}$ of \mathbf{r}_k corresponds to the sequence received at the i -th receive antenna, and the j -th component $x_k^{(j)}$ of \mathbf{x}_k corresponds to the sequence transmitted by the j -th transmit antenna. With the total power, averaged over all transmit antennas, constrained according to $\sum_{i=1}^n E[|x_k^{(i)}|^2] \leq P$, we once again, as in section 3.6, address the question of how to achieve Shannon capacity. Given $\mathbf{H}(z) = \mathbf{H}_0 + \mathbf{H}_1z^{-1} + \dots + \mathbf{H}_Mz^{-M}$, we choose $N \geq \frac{Mn}{m-n}$, such that \mathcal{H} of (6-1) is square or tall. The SVD of $\mathcal{H} = \mathcal{U}\mathcal{S}\mathcal{V}^*$ can then be used to design both a transmitter precoder \mathcal{V} and a receiver front-end filter \mathcal{U}^* such that the overall system is diagonal:

$$\mathcal{S} = \mathcal{U}^* \mathcal{H} \mathcal{V} \quad (6-25)$$

The system is illustrated in Fig. 6-8. In the transmitter, we first form a precoded vector according to $\mathbf{X}_j = \mathcal{V}\mathbf{W}_j$, where \mathbf{W}_j is an $n(M+N) \times 1$ vector of symbol sequences.

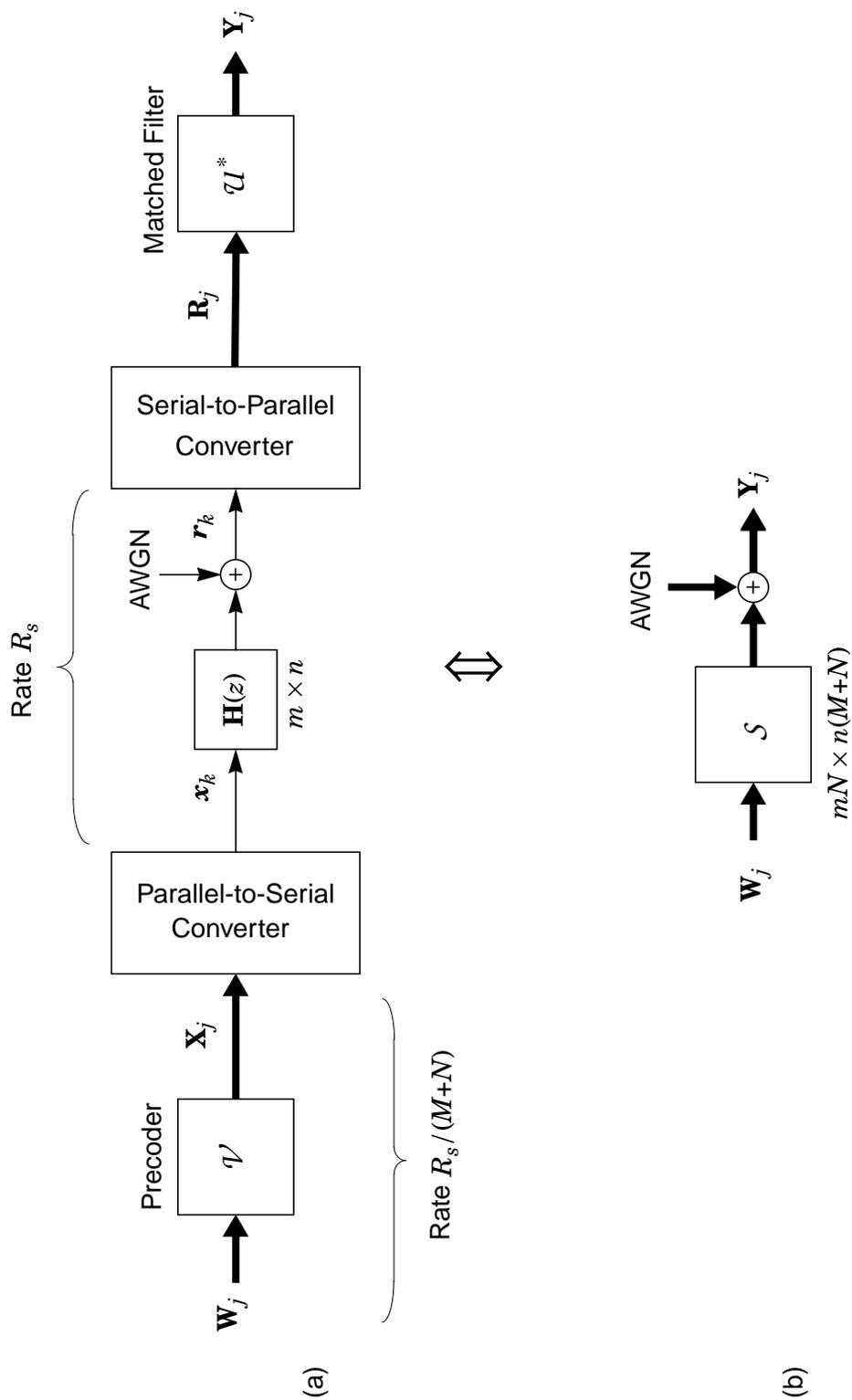


Figure 6-8. (a) A channel-diagonalization technique for tall FIR channels; (b) Equivalent diagonal channel model.

The norm-preserving property of \mathcal{V} ensures that \mathbf{W}_j and \mathbf{X}_j satisfy the same power constraint: $E[\|\mathbf{W}_j\|^2] = E[\|\mathbf{X}_j\|^2] = \sum_{i=1}^n E[|x_k^{(i)}|^2] \leq P$. We pass \mathbf{X}_j through a parallel-to-serial (P/S) converter to produce \mathbf{x}_k and transmit it across the channel. We denote the symbol rate as $R_s = 1/T$, so the block rate, *i.e.* the rate at which the blocks \mathbf{W}_j are transmitted, is $R_s/(M+N)$. The information bit rate is $R_b = bR_s/(M+N)$, where b is the average total number of information bits conveyed by each block \mathbf{W}_j . In the receiver, we pass \mathbf{r}_k through a serial-to-parallel (S/P) converter to form \mathbf{R}_j . We then rotate \mathbf{R}_j according to $\mathbf{Y}_j = \mathcal{U}^* \mathbf{R}_j$. The front-end rotation can be interpreted as a matched filter. Provided that the P/S and S/P converters are phase synchronized, the effective channel from \mathbf{W}_j to \mathbf{Y}_j is diagonal:

$$\mathbf{Y}_j = \mathcal{S} \mathbf{W}_j + \tilde{\mathbf{N}}_j, \quad (6-26)$$

where $E[\tilde{\mathbf{N}}_j \tilde{\mathbf{N}}_j^*] = \sigma^2 \mathbf{I}_{mN}$. We can thus approach capacity by properly distributing power and information among the subchannels [140].

Certainly, \mathcal{U}^* and \mathcal{V} can be estimated adaptively; however, the truly interesting aspect of this idea lies not in the method of adapting these filters, but in their existence. Because these filters are unitary, they are invertible, and thus *information lossless*. Moreover, they have *finite complexity*. For the special case of $n = 1$, our precoding technique can be viewed as a generalization of vector precoding, an idea proposed by Kasturia [141], which is similar to discrete multitone modulation (DMT) [23]. What distinguishes our scheme is that it is designed expressly for tall channels derived either through oversampling or through receiver sensor diversity, and it is therefore *both* lossless and of finite complexity. In contrast, the technique of Kasturia for baud-spaced channels, and the more well-known Tomlinson-Harashima [142, 143] technique, are either lossy [144] or of infi-

nite complexity. Furthermore, our scheme easily accommodates multichannel scenarios (with $n > 1$) such as array-to-array communication.

6.7 Adaptive Implementations

The adaptive techniques for subspace separation, singular-value decomposition, and spatial whitening, described in the context of memoryless channels (3-1), can be extended to channels with memory (6-1) by making only minor modifications. We can use the adaptive algorithms described in section 3.4 or section 4.3 to blindly implement the stacked MMSE, ZF, or WR detector by simply replacing \mathbf{r}_k with \mathbf{R}_k , n with $\eta = n(M+N)$, and m with mN . We summarize the modified algorithms below.

6.7.1 An Adaptive Stacked MMSE or ZF Detector

As suggested by (6-4) and (6-7), the stacked MMSE and ZF detectors can be implemented by using a rotate-scale-rotate architecture, in a manner analogous to the implementation of the MMSE detector of Definition 3-3 for memoryless channels. The first step is to rotate the stacked observation vector \mathbf{R}_k with an $mN \times mN$ unitary matrix $\hat{\mathcal{U}}$, adapted according to

$$\hat{\mathcal{U}}_{k+1} = \hat{\mathcal{U}}_k \mathcal{R}^\lambda(\mathcal{G} \hat{\mathcal{U}}_k \mathbf{R}_k \rightarrow \hat{\mathcal{U}}_k \mathbf{R}_k), \quad (6-27)$$

where \mathcal{G} is diagonal with strictly decreasing elements satisfying $g_1 > g_2 > \dots > g_{mN}$. If Conjecture 3-2 holds, then the recursion converges to $\hat{\mathcal{U}} = \mathcal{U}$, a valid left factor in a channel SVD $\mathcal{H} = \mathcal{U}\mathcal{S}\mathcal{V}^*$, at which point \mathcal{V} is uniquely specified.

The second step is to scale the output of the rotation, $\mathbf{Y}_k = \hat{\mathcal{U}}_k^* \mathbf{R}_k$, by a diagonal $\eta \times mN$ matrix \mathcal{D} defined as $\mathcal{D}_{MMSE} = \mathcal{S}^* (\mathcal{S}\mathcal{S}^* + \sigma^2 \mathbf{I})^{-2}$ for the stacked MMSE detector

or as $\mathcal{D}_{ZF} = \mathcal{S}^\dagger$ for the stacked ZF detector. The dimension of \mathcal{D} is $\eta \times mN$, where η is the column rank of \mathcal{H} , or equivalently, the dimension of the signal subspace. Hence, in order to estimate \mathcal{D} , the receiver must first estimate η .

We can blindly estimate η , along with \mathcal{S} and σ^2 , and thus \mathcal{D} , by estimating the power in the components of \mathbf{Y}_k with equations analogous to (3-29) through (3-31). If the recursion of (6-27) has converged, then the autocorrelation of $\mathbf{Y}_k = \mathcal{U}_k^* \mathbf{R}_k$ is given by

$$\Phi_{\mathbf{Y}} = \mathcal{S}\mathcal{S}^* + \sigma^2\mathbf{I}. \quad (6-28)$$

The power in the components of \mathbf{Y}_k are the thus eigenvalues of $\Phi_{\mathbf{R}}$:

$$E[|Y_k^{(i)}|^2] = \begin{cases} s_i^2 + \sigma^2 & i \in \{1, \dots, \eta\} \\ \sigma^2 & i \in \{\eta+1, \dots, m\} \end{cases} \quad (6-29)$$

The eigenvalues can be estimated recursively:

$$\hat{\epsilon}_k^{(i)} = \alpha \hat{\epsilon}_{k-1}^{(i)} + (1 - \alpha) |Y_k^{(i)}|^2, \quad (6-30)$$

where $0 < \alpha < 1$ is a smoothing factor. A threshold can then be used to estimate η , the number of significant eigenvalues. (The choice of a threshold is discussed further in section 6.8.) Given η , we can estimate the singular values s_i of \mathcal{H} and the noise variance σ^2 as follows:

$$\hat{\sigma}_k^2 = \frac{1}{mN - \eta} \sum_{i=\eta+1}^{mN} \hat{\epsilon}_k^{(i)}, \quad (6-31)$$

$$\hat{s}_i(k) = (\hat{\epsilon}_k^{(i)} - \hat{\sigma}_k^2)^{1/2}. \quad (6-32)$$

Certainly, this procedure is prone to inaccuracies; first, an inaccurate estimate of \mathcal{U} would violate (6-28); second, noise limits the accuracy of the eigenvalue estimates in (6-30); and third, some of the singular values may be quite small, and hence it may be unclear which eigenvalues correspond to noise alone and which correspond to noise plus signal. We investigate the sensitivity of the algorithms to inaccuracies in the estimate of η in section 6.8. We show that the proposed detectors are in fact very robust in this respect.

The final step in implementing the detector is to rotate the output of the gain stage, $\mathbf{W}_k = \hat{\mathcal{D}}\mathbf{Y}_k$, by a $\eta \times \eta$ unitary matrix $\hat{\mathcal{V}}^*$ adapted according to the MPLL recursion:

$$\hat{\mathcal{V}}_{k+1} = \mathcal{R}^{\lambda}(\hat{\mathbf{X}}_k \rightarrow \mathbf{Z}_k)^* \hat{\mathcal{V}}_k, \quad (6-33)$$

where $\mathbf{Z}_k = \hat{\mathcal{V}}_k \mathbf{W}_k$, and where $\hat{\mathbf{X}}_k = \mathbf{dec}(\mathbf{Z}_k)$ is a quantized decision vector. The recursion should converge to \mathcal{V} up to a permutation ambiguity:

$$\hat{\mathcal{V}} = \mathcal{K}\mathcal{V}. \quad (6-34)$$

We address the implications of the permutation in section 6.7.4.

6.7.2 An Adaptive Stacked WR Detector

As suggested by (6-12), the stacked WR detector can be implemented using spatial prediction in a manner analogous to that of section 4.3. We can predict \mathbf{R}_k according to

$$\hat{\mathbf{R}}_k = \mathcal{P}\mathbf{R}_k, \quad (6-35)$$

where \mathcal{P} is an $mN \times mN$ strictly lower-triangular matrix of prediction coefficients. An estimate of \mathcal{P} can be adapted according to the least-mean-square algorithm:

$$\hat{\mathcal{P}}_{k+1} = (\hat{\mathcal{P}}_k + \mu \mathbf{E}_k \mathbf{R}_k^*) \otimes \mathcal{L}, \quad (6-36)$$

where $\mathbf{E}_k = \mathbf{R}_k - \hat{\mathcal{P}}_k \mathbf{R}_k$ is the prediction error, where μ is a step size, where ‘ \otimes ’ denotes a component-wise product, and \mathcal{L} is a mask that constrains $\hat{\mathcal{P}}_k$ to be strictly lower triangular. For sufficiently small step-size μ , the predictor converges to \mathcal{M}^{-1} of (6-12). Following the predictor by a bank of AGCs, adapted according to (4-16), completes the implementation of the spatial whitener.

To implement the final stage, the output of the AGC bank is rotated by an $mN \times mN$ unitary matrix \hat{Q} adapted according to

$$\hat{Q}_{k+1} = \mathcal{R}^\lambda(\hat{\mathbf{X}}_k \rightarrow \mathbf{Y}_k)^* \hat{Q}_k, \quad (6-37)$$

where $\mathbf{Y}_k = \hat{Q}_k \mathbf{E}_k$, and where

$$\hat{\mathbf{X}}_k = \begin{bmatrix} \mathit{dec}(\cdot) & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{bmatrix} \mathbf{Y}_k. \quad (6-38)$$

For $i \leq \eta$, $\hat{X}_k^{(i)} = \mathit{dec}_i(Y_k^{(i)})$ is the point in the constellation of user i closest to $Y_k^{(i)}$, but for $i > \eta$, $\hat{X}_k^{(i)}$ is set to zero. The recursion of (6-37) should converge to Q of (6-11) up to an ambiguous permutation matrix:

$$\hat{Q} = \begin{bmatrix} \mathcal{K} & \mathbf{0} \\ \mathbf{0} & \mathbf{I}_{mN-\eta} \end{bmatrix} Q. \quad (6-39)$$

6.7.3 An Adaptive Subspace Separator

A subspace separator can be used as the front end of either detector without loss of any signal information. The subspace separator is implemented by a $mN \times mN$ unitary matrix $\hat{\Theta}$ adapted according to

$$\hat{\Theta}_{k+1} = \mathcal{R}^\lambda(\mathcal{G} \hat{\Theta}_k \mathbf{R}_k \rightarrow \hat{\Theta}_k \mathbf{R}_k)^* \hat{\Theta}_k, \quad (6-40)$$

where \mathcal{G} is diagonal with elements satisfying $g_1 \dots g_\eta = 1$ and $g_{\eta+1} \dots g_{mN} = 0$. The recursion should converge to a subspace separator of the form given by (6-22). The last $mN - \eta$ components of $\hat{\Theta}$ can be discarded to form the reduced channel given by (6-24).

6.7.4 Selection of the Detector Outputs

The last rotation stage in any of the proposed adaptive stacked detectors converges with a permutation ambiguity. We say the decisions are *correct* if the slicer output is given by $\hat{\mathbf{X}}_k = \mathcal{K}\mathbf{X}_k$, for some complex permutation matrix \mathcal{K} . Components of \mathbf{X}_k , *i.e.* virtual users, can thus be relabeled or (assuming QAM constellations) arbitrarily rotated by a multiple of 90° . The final step is to resolve the ambiguity \mathcal{K} ; this is the function of the “select best” block in Fig. 6-2. This step entails first determining an estimate \hat{n} of the number of users, and then assigning of the \hat{n} MPLL outputs to the \hat{n} identified users. We outline several techniques below.

One method is to correlate the slicer output vector $\hat{\mathbf{X}}_k$ with itself at lag 1; *i.e.* estimate $E[\hat{\mathbf{X}}_k \hat{\mathbf{X}}_{k-1}^*]$. The stacked channel input \mathbf{X}_k has a shifting property that can be exploited:

$$E[\mathbf{X}_k \mathbf{X}_{k-1}^*] = \begin{bmatrix} \mathbf{0} & & & & \mathbf{0} \\ \mathbf{I}_m & \mathbf{0} & & & \\ & \mathbf{I}_m & \mathbf{0} & & \\ & & \ddots & \ddots & \\ \mathbf{0} & & & \mathbf{I}_m & \mathbf{0} \end{bmatrix}. \quad (6-41)$$

In practice, large entries in $E[\hat{\mathbf{X}}_k \hat{\mathbf{X}}_{k-1}^*]$ indicate which outputs belong to the same user. The best output for each user can then be selected from among the outputs associated with

that user. As with any blind multiuser detector, there remains an arbitrary relabeling and phase rotation of the \hat{n} users. See Appendix 6-1 for an example of this technique.

Another method is to correlate scaler slicer outputs at different time lags; uncorrelated outputs correspond to constellations of distinct users. More precisely, we first select the detector output with the smallest mean slicer error, thereby recovering one user. We then consider the detector output with the next smallest mean slicer error. If this output is uncorrelated with the previous selection at all time lags $L \in \{-M-N+1 \dots M+N-1\}$, then we select it too, thereby recovering a second user; otherwise we reject it. We continue this process, rejecting all outputs correlated to previous selections, until we have recovered all users.

A third technique is to consider the structure of the stacked detector itself. Rather than correlate detector outputs at various time lags, rows of the stacked detector can be correlated at various shifts. Rows corresponding to distinct users are nearly orthogonal at all shifts, whereas rows corresponding to identical users are highly correlated.

6.8 Experimental Results

We now present results from computer experiments that demonstrate the proposed blind adaptive implementation of the stacked MMSE detector. The first experiment of this section is designed to characterize the performance of the algorithm if the estimate of the signal subspace dimension η is inaccurate. Our approach is to implement the stacked MMSE detector as outlined in section 6.7.1, but to intentionally override the threshold test and to use a dimension estimate that is either too small or too large. As we demonstrate, the proposed algorithm is very robust to inaccuracies in the estimate.

Experiment 6-5. A Linear Array with Multipath. Consider a system with 2 transmitters and a receiver using a linear array of 10 antenna elements with half-wavelength spacing. The antenna array receives energy from each user along two paths: a line-of-sight (LOS) path and a reflected path. Each path is characterized by its amplitude A , its propagation delay τ , and the angle of incidence θ . For user 1, the parameters are $(A, \tau, \theta) = (0.25, -0.2, 10^\circ)$ and $(A, \tau, \theta) = (0.2, 2.3, 85^\circ)$ for the LOS path and reflected path, respectively. For user 2, the parameters are $(0.22, -0.1, 25^\circ)$ and $(0.032, 2.1, 90^\circ)$. Both users transmit with zero excess bandwidth, which implies that the channel memory is infinite. To distinguish the users, user 1 transmits QPSK, and user 2 transmits BPSK. For the described scenario, $\text{SNR}_1 = 20$ dB and $\text{SNR}_2 = 17$ dB. (The receiver parameters are as follows: $\lambda_1 = 1 / (1 + (k/800)^2)$ in (6-27), $\alpha = 0.99$ in (6-30), and $\lambda_2 = 1 / (1 + (k/900))$ in (6-37).) The stacking depth is $N = 4$, so the stacked observation has dimension 40.

The recursion of (6-30) produces the eigenvalue estimates illustrated in Fig. 6-9. The middle threshold, indicated by the dashed line, produces an signal subspace dimension estimate of $\hat{\eta} = 12$. However, the threshold might arguably be set higher or lower, as shown by the dotted lines, producing estimates of that range from $\hat{\eta} = 10$ to 14. Some of these represent an overestimate of the channel rank, and some, an underestimation.

We conduct 5 trials, overriding the threshold test and forcing the estimate to be $\hat{\eta} = 10, 11, 12, 13,$ and 14. The detector outputs at steady state (10000 baud) for each case are shown in Fig. 6-10. The constellations in the first column are the detector outputs for $\hat{\eta} = 10$. Although $\eta > 10$, the proposed algorithm is still able to produce at least one mildly clean constellation for each user. We see the cleanest

constellations when $\hat{\eta} = 12$. For $\hat{\eta} = 13$ and 14, we can see that the effect of overestimating the signal subspace dimension is to simply produce extra noisy constellations, which can be ignored. Regardless of the dimension estimate, there is at least one clean constellation for each user.

The previous experiment demonstrates that the proposed algorithm is not sensitive to the estimate of the signal subspace dimension. A similar experiment for an asynchronous-CDMA application, detailed in [145], confirms these results as well. Moreover, the proposed algorithm does not even need to know the number of users n , and it needs only limited knowledge about the constellations of the users. The slicer assumes that both users are transmitting QPSK. We can in fact use QPSK or quadrant decisions in (6-38) for practically all QAM constellations. Remarkably, the recursion of (6-37) still converges. So, the

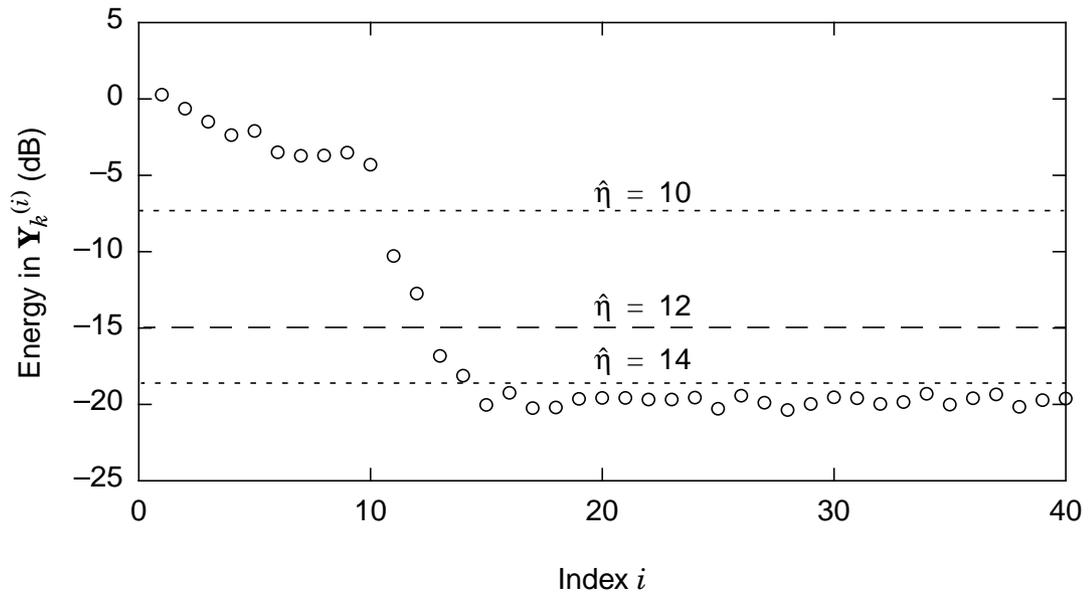


Figure 6-9. Estimates of the eigenvalues of $\Phi_{\mathbf{R}}$ in Experiment 6-5 produced by the recursion of (6-30) at steady state.

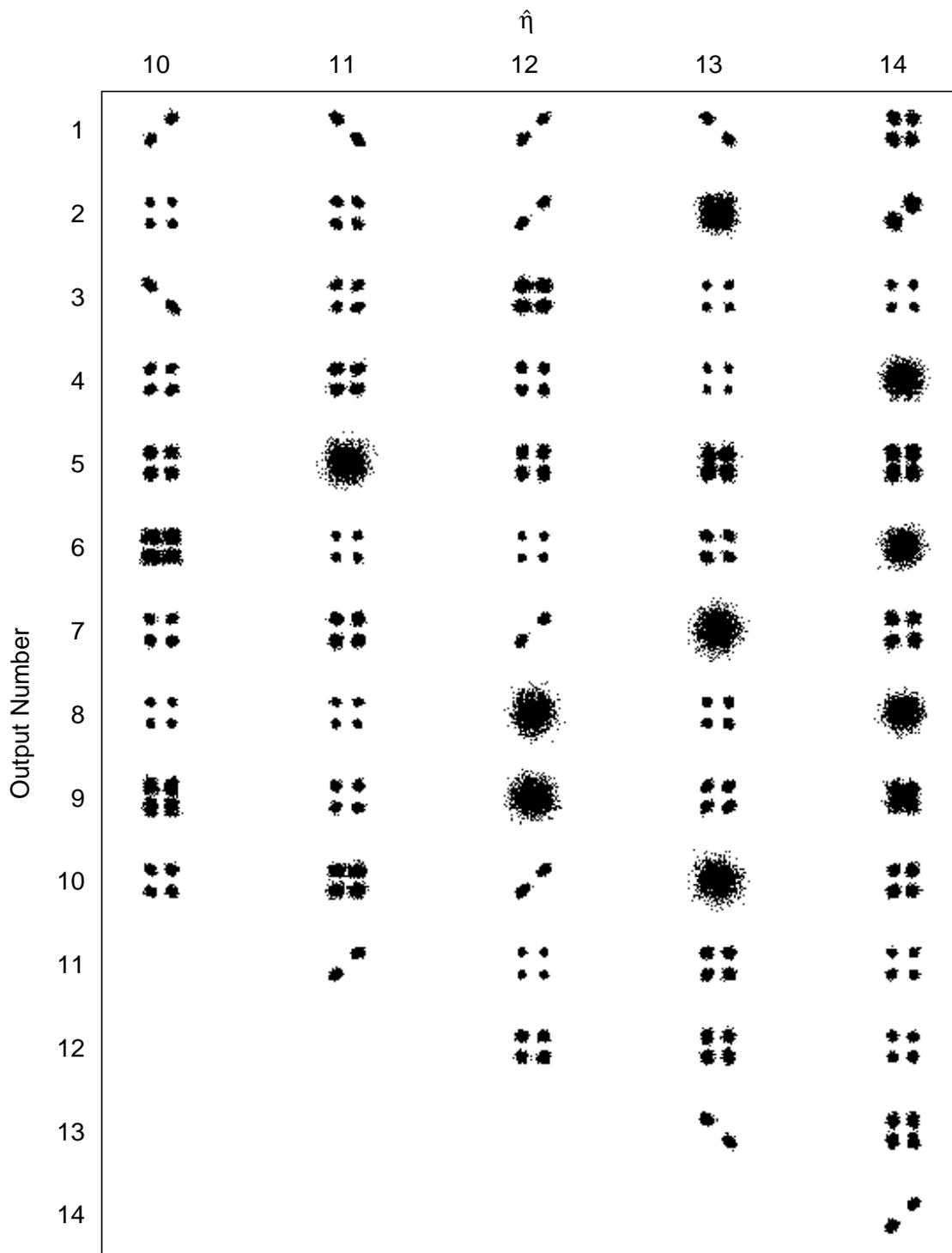


Figure 6-10. Outputs of the stacked MMSE detector as a function of the signal subspace dimension estimate $\hat{\eta}$ for Experiment 6-5.

proposed detectors are very blind indeed. The next experiment, in which one user transmits QPSK, and the other transmits 16-QAM, demonstrates this idea.

Experiment 6-6. Asynchronous CDMA. We now consider another 2-user asynchronous CDMA system. The system is identical to that of Experiment 5-8 except that the amplitudes A_1 and A_2 and the noise variance σ^2 are selected such that $\text{SNR}_1 = 25$ dB and $\text{SNR}_2 = 20$ dB. We implement the blind adaptive stacked MMSE detector with the following parameters: $\lambda_1 = 1/(1 + (k/500)^2)$ in (6-27), $\alpha = 0.99$ in (6-30), and $\lambda_2 = 1/(1 + (k/1000))$ in (6-37). The constellations for users 1 and 2 are 16-QAM and QPSK, respectively. Fig. 6-11 shows the best outputs of the detector at steady state (18000 to 20000 baud) for stacking depths of $N = 1$ and 2. We see that even a stacking depth of 1, *i.e.* no stacking, is sufficient to produce good results. A stacking depth of 2 produces only slight improvement.

We remark that, for CDMA applications, the stacked MMSE detector often requires no stacking ($N = 1$), in which case it has both better performance and lower complexity than the prediction-based detectors of chapter 5.

6.9 Chapter Summary

We have shown that by using a simple stacking procedure, we can effectively convert the tall $m \times n$ FIR channel $\mathbf{H}(z) = \mathbf{H}_0 + \mathbf{H}_1 z^{-1} + \dots + \mathbf{H}_M z^{-M}$ of (1-1) into the tall $mN \times n(M+N)$ memoryless block-Toeplitz channel \mathcal{H} of (6-1). The algorithms for adaptive subspace separation, singular-value decomposition, and spatial whitening can then be

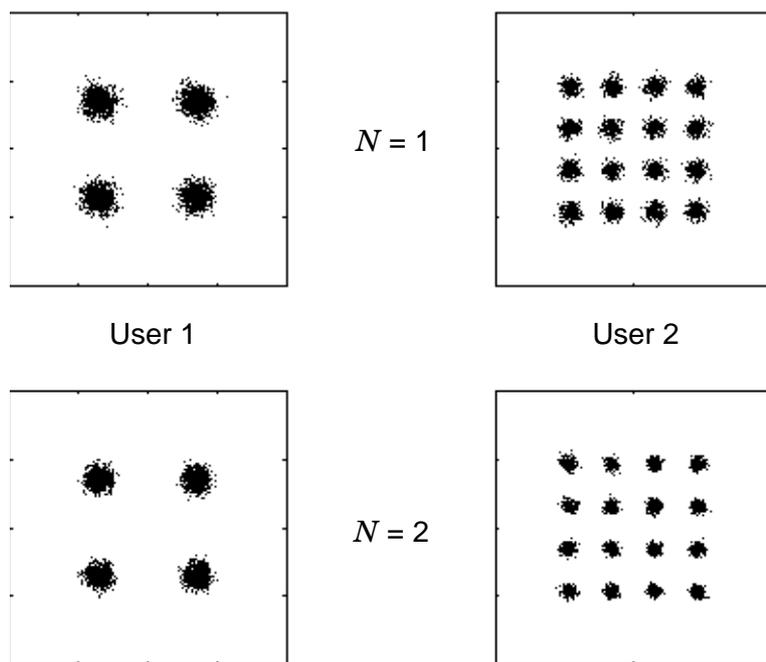


Figure 6-11. Best outputs of stacked MMSE detector for the asynchronous-CDMA application of Experiment 6-6.

easily extended to \mathcal{H} and thus implicitly to $\mathbf{H}(z)$. We have defined stacked MMSE, ZF, and WR detectors in terms of \mathcal{H} .

The stacked MMSE detector, according to Theorem 6-1, implements $M+N$ conventional MMSE multiuser detectors, one corresponding to each decision delay 0 through $M+N-1$. Following the detector by a bank of independent slicers, and choosing for each user the associated output with the smallest mean slicer error, effectively implements an N -tap MMSE multiuser detector with the delay optimized for each user. Although more complex, in general, than the detectors based on linear prediction, the stacked MMSE detector has better performance in noise.

The stacked ZF detector, according to Theorem 6-2, implements $M+N$ conventional delayed ZF detectors, one corresponding to each delay 0 through $M+N-1$. Following the minimum-norm detector by a bank of independent slicers, and choosing for each user the output with the smallest mean slicer error, effectively implements an N -tap multiuser detector that, among the class of ZF detectors, minimizes the MSE of each user.

The stacked WR detector, according to Theorem 6-3, estimates the channel input at each delay 0 through $M+N-1$ by using linear combinations of prediction errors. It can be interpreted as a prediction-based approximation to the stacked MMSE detector. It has nearly the same performance, but can have lower complexity when $mN \approx n(M+N)$.

We have detailed blind adaptive implementations of these detectors, and have shown that the stacked MMSE detector, in particular, is an excellent candidate for CDMA applications. It is robust to inaccuracies in the estimate of the signal subspace dimension. It needs only minimal information regarding the number and constellations of the users. It has better performance than the prediction-based detectors of chapter 5, and can have

lower complexity. Moreover, the blind algorithms we have proposed in this chapter, like those of chapter 5, and preceding chapters, exploit primarily second-order statistics and are thus insensitive to channel input distributions.

As a bonus, we have also developed, for tall FIR channels, a novel space-time (or time only) precoder that is both lossless and of finite complexity. It can be used to completely eliminate ISI and MUI in the receiver without noise enhancement, thereby facilitating transmission approaching capacity.

APPENDIX 6-1:
AN ALGORITHM FOR RESOLVING THE PERMUTATION \mathcal{K}

The last rotation stage for any of the proposed stacked detectors converges with a permutation ambiguity \mathcal{K} , such that, when decisions are *correct*, the slicer output is $\hat{\mathbf{X}}_k = \mathcal{K}\mathbf{X}_k$. In this context, \mathcal{K} permutes *virtual* users, not just actual users. To resolve \mathcal{K} , we must ensure that the temporal ordering of actual users is preserved, and that their relative complex rotation is zero. Mathematically, an acceptable permutation must be of the form:

$$\tilde{\mathcal{K}} = \begin{bmatrix} \mathbf{K} & & \mathbf{0} \\ & \mathbf{K} & \\ \mathbf{0} & \ddots & \\ & & \mathbf{K} \end{bmatrix}, \quad (6-42)$$

where \mathbf{K} is an arbitrary complex permutation matrix of dimension $n \times n$. Fortunately, we have additional information that can be used to ensure that \mathcal{K} satisfies (6-42).

Unlike \mathbf{x}_k , the sequence \mathbf{X}_k is not temporally white, because the vector \mathbf{X}_k , for a given time k , is related to the vector \mathbf{X}_{k-1} by a shift. More precisely, the autocorrelation of \mathbf{X}_k at lag 1 is a *shifting matrix*:

$$E[\mathbf{X}_k \mathbf{X}_{k-1}^*] = \begin{bmatrix} \mathbf{0} & & \mathbf{0} \\ \mathbf{I}_n & \mathbf{0} & \\ & \mathbf{I}_n & \mathbf{0} \\ \mathbf{0} & \ddots & \ddots \\ & & \mathbf{I}_n & \mathbf{0} \end{bmatrix} \equiv \mathcal{T}. \quad (6-43)$$

The term “shifting matrix” originates from the property that pre-multiplying any vector by \mathcal{T} effectively shifts the components of that vector down by n . (The last n components are

discarded at the bottom, and n zeros are shifted in at the top.) Observe that for any arbitrary ambiguity \mathcal{K} , we can left factor ambiguities that are detrimental:

$$\mathcal{K} = \mathcal{K}_0 \tilde{\mathcal{K}}, \quad (6-44)$$

where $\tilde{\mathcal{K}}$ satisfies (6-42), and is thus harmless, but where \mathcal{K}_0 does not satisfy (6-42). If we assume that the decisions are correct, so that $\hat{\mathbf{X}}_k = \mathcal{K} \mathbf{X}_k$, then the autocorrelation of the output $\hat{\mathbf{X}}_k$ at lag 1 is given by

$$E[\hat{\mathbf{X}}_k \hat{\mathbf{X}}_{k-1}^*] = \mathcal{K}^T \mathcal{K}^* = \mathcal{K}_0 \tilde{\mathcal{K}}^T \tilde{\mathcal{K}}^* \mathcal{K}_0^* = \mathcal{K}_0^T \mathcal{K}_0^*. \quad (6-45)$$

The detrimental ambiguity \mathcal{K}_0 can be removed as follows. First, estimate $\Phi = E[\hat{\mathbf{X}}_k \hat{\mathbf{X}}_{k-1}^*]$, by sample averaging and quantizing the elements of the average to the set $\{\pm 1, \pm j, 0\}$. This estimate can be deemed valid by observing the size of the quantization error, and the structure of the resulting matrix. Any valid estimate $\hat{\Phi}$ can then be manipulated to create a shifting matrix, by using a series of row exchanges \mathcal{B}_i :

$$\mathcal{B}_L \dots \mathcal{B}_2 \mathcal{B}_1 \hat{\Phi} \mathcal{B}_1^* \mathcal{B}_2^* \dots \mathcal{B}_L^* = \mathcal{I} \quad (6-46)$$

The product $\mathcal{B}_L \dots \mathcal{B}_2 \mathcal{B}_1 = \mathcal{K}_0^*$ removes the detrimental ambiguities. Consider the following example.

Example 6-1. With $n = 2$ and $n(M+N) = 4$, suppose that the stacked detector converges with the following permutation:

$$\hat{\mathbf{X}}_k = \mathcal{K} \mathbf{X}_k = \begin{bmatrix} 0 & 1 & 0 & 0 \\ 0 & 0 & -j & 0 \\ -1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \mathbf{X}_k. \quad (6-47)$$

The resulting slicer output is $\hat{\mathbf{X}}_k = [x_k^{(2)}, -jx_{k-1}^{(1)}, -x_k^{(1)}, x_{k-1}^{(2)}]^T$. Without knowledge of the permutation we do not know which components belong to which users.

For the given \mathcal{K} , the estimate $\hat{\Phi}$ of $E[\hat{\mathbf{X}}_k \hat{\mathbf{X}}_{k-1}^*]$ should be

$$\hat{\Phi} = \mathcal{K} \mathcal{T} \mathcal{K}^* = \begin{bmatrix} 0 & 1 & 0 & 0 \\ 0 & 0 & -j & 0 \\ -1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \end{bmatrix} \begin{bmatrix} 0 & 0 & -1 & 0 \\ 1 & 0 & 0 & 0 \\ 0 & j & 0 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} = \begin{bmatrix} 0 & 0 & 0 & 0 \\ 0 & 0 & j & 0 \\ 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 \end{bmatrix}. \quad (6-48)$$

Given only the matrix to the right of the equality in (6-48), we now look for a series of row exchanges \mathcal{B}_i that produces a shifting matrix. By inspection, the first operation \mathcal{B}_1 should swap rows 1 and 2 of $\hat{\Phi}$ in order to force row 2 to be all zeros:

$$\mathcal{B}_1 \hat{\Phi} \mathcal{B}_1^* = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} 0 & 0 & 0 & 0 \\ 0 & 0 & j & 0 \\ 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 \end{bmatrix} \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} = \begin{bmatrix} 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & j & 0 & 0 \\ 1 & 0 & 0 & 0 \end{bmatrix}. \quad (6-49)$$

From the result, we see that the next operation \mathcal{B}_2 should multiply row 3 by $-j$, and then swap rows 3 and 4:

$$\mathcal{B}_2(\mathcal{B}_1 \hat{\Phi} \mathcal{B}_1^*) \mathcal{B}_2^* = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & -j & 0 \end{bmatrix} \begin{bmatrix} 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & j & 0 & 0 \\ 1 & 0 & 0 & 0 \end{bmatrix} \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & j \\ 0 & 0 & 1 & 0 \end{bmatrix} = \begin{bmatrix} 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \end{bmatrix} = \mathcal{T}. \quad (6-50)$$

Obtaining \mathcal{T} as a result terminates the algorithm. Observe that multiplying the output of the MPLL by $\mathcal{B}_2 \mathcal{B}_1 = \mathcal{K}_0^*$ removes the detrimental ambiguities:

$$(\mathcal{B}_2\mathcal{B}_1)\mathcal{K}\mathbf{X}_k = \mathcal{K}_0^* \mathcal{K}\mathbf{X}_k =$$

$$\begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & -j & 0 & 0 \end{bmatrix} \begin{bmatrix} 0 & 1 & 0 & 0 \\ 0 & 0 & -j & 0 \\ -1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \mathbf{X}_k = \begin{bmatrix} 0 & 1 & 0 & 0 \\ -1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & -1 & 0 \end{bmatrix} \mathbf{X}_k = \tilde{\mathcal{K}} \mathbf{X}_k. \quad (6-51)$$

The residual permutation $\tilde{\mathcal{K}}$, which is of the form given by (6-42), is harmless because it merely exchanges the user labels and negates one user.

In practice, once \mathcal{K}_0^* is estimated, we can correct the rotator of (6-37) according to $Q_k = \mathcal{K}_0^* Q_k$. We then can start a new running estimate of Φ , and periodically verify that the new estimate continues to be a shifting matrix: $\hat{\Phi} = \mathcal{T}$.

CHAPTER 7

CONCLUSIONS AND FUTURE WORK

7.1 Conclusions

We have developed several new algorithms for blind multiuser detection and equalization based on a philosophy of minimal reliance on higher-order statistics. These algorithms use adaptive linear prediction and subspace or singular-value decomposition to exploit primarily the second-order statistics of the receiver observation, an approach which offers significant advantages over HOS-based or batch-oriented methods. The proposed detectors have good performance, low complexity, fast convergence, and an innate compatibility with shaped signal constellations. We have demonstrated the effectiveness of these algorithms in a wide variety of contexts, including multisensor receivers, code-division multiple-access systems, and fractionally spaced equalizers.

In chapter 3, we have proposed subspace-based detectors for memoryless channels. We have proposed a blind algorithm for adaptively separating the signal and noise subspaces. Because the subspace separator is information-lossless, it can be used as a universal detector front end to reduce the complexity of subsequent processing. We have generalized the subspace separator to further decompose the receive space thereby leading to an algorithm for adaptive singular-value decomposition of the channel. We have pro-

posed fully blind implementations of the MMSE and ZF multiuser detectors, based on the adaptive SVD. We have also proposed a channel diagonalization algorithm, which facilitates transmission approaching capacity in single-user multi-channel contexts such as array-to-array communication.

In chapter 4, we have proposed the canonical whiten-rotate detector for memoryless channels. The WR detector is the unique spatial whitener with minimal MSE. It is information lossless, optimally near-far resistant, and it has near-MMSE performance. We have proposed a blind adaptive implementation based on spatial linear prediction and a simple modification to the MPLL algorithm. We have also proposed a project-first WR architecture that uses the subspace separator in its front end. The two structures are mathematically equivalent, but the project-first approach has lower complexity, and faster convergence, when the dimension of the receiver observation is exceedingly large. We have demonstrated the WR detector for both wideband synchronous-CDMA and narrow-band linear-array applications.

In chapter 5, we have presented a family of blind multiuser detectors that combine spatial and temporal prediction to exploit the unique properties of tall FIR channels. We have shown that tall channels can have many equivalent representations, *e.g.*, moving average, autoregressive, and others. The existence of a finitely parameterized AR model, in particular, implies an FIR, and therefore stable, left-inverse. In a sense, almost all tall FIR channels are minimum phase, a fact which suggests the use of linear prediction. We have proposed the forward LP detector (Definition 5-6), the first stage of which is a one-step forward temporal prediction-error filter. The detector exploits the AR channel model of section 5.1.1 to roughly convert the channel $\mathbf{H}(z)$ of (1-1) into a memoryless

channel \mathbf{H}_0 . Although with noise, this description is only approximate, the spatial methods of chapter 4 can then nevertheless be applied to recover the transmitted sequence. With a zero-delay constraint, the FLP detector performance is near that of the MMSE detector. However its performance in general is highly dependent upon the size of the zero-th tap \mathbf{H}_0 , provided that this tap is significant relative to noise.

Therefore, we have also proposed a more general forward-backward LP detector (Definition 5-7). The first stage of the LP detector is an $(L+1)$ -step forward prediction-error filter; the second stage is a one-step backward prediction-error filter. Together these filters exploit the ARMA and AP models of sections 5.1.2 and 5.1.3, respectively, to roughly convert the channel $\mathbf{H}(z)$ of (1-1) into a memoryless channel $\mathbf{H}_L z^{-L-\tilde{N}}$. (Recall that \tilde{N} is the order of the backward predictor.) The spatial methods of chapter 4 are then applied to invert \mathbf{H}_L . The performance of the LP detector is roughly proportional to the energy in \mathbf{H}_L ; therefore the index L should correspond to the tap of greatest energy. In section 5.6, we have detailed blind adaptive implementations of the detectors and have demonstrated applications, including an adaptive fractionally spaced equalizer (FSE) for single-user systems using highly shaped signal constellations and a blind multiuser detector for asynchronous CDMA systems.

In chapter 6, we have extended the algorithms for adaptive subspace separation, singular-value decomposition, and spatial whitening to channels with memory. We have shown that by using a simple stacking procedure, we can effectively convert the tall $m \times n$ FIR channel $\mathbf{H}(z)$ of (1-1) into the tall $mN \times n(M+N)$ memoryless block-Toeplitz channel \mathcal{H} of (6-1). The algorithms designed for \mathbf{H} of (3-1), then readily extend to \mathcal{H} . We have defined stacked MMSE, ZF, and WR detectors in terms of \mathcal{H} . The stacked MMSE detector

implements $M+N$ conventional MMSE multiuser detectors, and thus can be used to implement an N -tap MMSE multiuser detector with the delay optimized for each user. Although more complex, in general, than the LP-based detectors of chapter 5, the stacked MMSE detector has better performance in noise. The stacked ZF detector can be used to implement an N -tap multiuser detector that, among the class of ZF detectors, minimizes the MSE of each user. The stacked WR detector is a spatio-temporal prediction-based approximation to the stacked MMSE detector. It has nearly the same performance, but can have lower complexity when $mN \approx n(M+N)$. We have detailed blind adaptive implementations of these detectors.

The stacked MMSE detector is particularly appealing. It is robust to inaccuracies in the estimate of the channel order, needs only minimal information regarding the number and constellations of the users, has better performance than the prediction-based detectors of chapter 5, and can have lower complexity in some applications, most notably CDMA.

We have also developed a novel space-time precoder that is simultaneously lossless and of finite complexity. It can be used to eliminate interference in the receiver without noise enhancement.

7.2 Future Research

We have reserved the final section of the thesis to discuss remaining open issues, and to suggest ideas for future research. Open issues include a rigorous proof of convergence for the subspace separator and SVD algorithms, a method for optimizing this convergence, and a method for blind estimation of the index L used in the general forward-backward LP detector. As ideas for future research, we suggest alternative LP-based detector architec-

tures as well as techniques to speed convergence of all detectors for application to fading or other time-varying channels. We also propose an algorithm for blind channel identification based on adaptive correlation matching, which may be of interest to researchers in this field.

7.2.1 SVD Convergence

All simulation results suggest that the subspace-separation (3-20) and complete subspace-decomposition (3-25) algorithms do indeed converge to the desired solutions for sufficiently small loop gain λ ; however, there is no rigorous proof of convergence. Appendix 3-2 presents a heuristic argument but proves only that the inner product of $\mathbf{G}\mathbf{y}$ and \mathbf{y} , where \mathbf{y} is the rotator output, is maximized when

$$J = \sum_i \sum_{j>i} (g_i - g_j)(d_i - d_j) |t_{i,j}|^2 \quad (7-1)$$

is minimized. (Recall from Appendix 3-2 that g_i and d_i are diagonal elements of \mathbf{G} and $\mathbf{D} = \mathbf{S}\mathbf{S}^*$, respectively, for some channel SVD $\mathbf{H} = \mathbf{U}\mathbf{S}\mathbf{V}^*$, and that $t_{i,j}$ are elements of $\mathbf{T} = \mathbf{Q}\mathbf{U}$, where \mathbf{Q} is the rotator.) Although (7-1) is aesthetically pleasing, a more rigorous proof of convergence is still needed and could be the subject of future research.

It is also unclear how the elements g_i should be chosen to optimize convergence speed, although we have given this topic some consideration. For the sake of discussion, assume that the elements d_i are all distinct. (The channel must therefore be square.) In this case, we would ideally like to force all off-diagonal elements $t_{i,j}$ of \mathbf{T} to zero at the same rate. A reasonable approach is to try to make the coefficients of $|t_{i,j}|^2$ in (7-1) constant for all $i \neq j$:

$$(g_i - g_j)(d_i - d_j) = c, \quad (7-2)$$

or equivalently, try to make

$$g_i - g_j = \frac{c}{d_i - d_j} \equiv c_{i,j} \quad (7-3)$$

where c and $c_{i,j}$ are positive real constants. We can express (7-3) as

$$\underbrace{\begin{bmatrix} c_{1,2} \\ \vdots \\ c_{i,j} \\ \vdots \\ c_{m-1,m} \end{bmatrix}}_{\mathbf{c}} = \underbrace{\begin{bmatrix} 1 & -1 & 0 & \dots & 0 \\ & & \ddots & & \\ & & & 1 & \dots & -1 \\ & & & & \ddots & \\ 0 & \dots & 0 & 1 & -1 \end{bmatrix}}_{\mathbf{A}} \underbrace{\begin{bmatrix} g_1 \\ \vdots \\ g_m \end{bmatrix}}_{\mathbf{g}}, \quad (7-4)$$

which is a system of $m(m-1)/2$ equations with m unknowns. For $m > 2$, the system is overdetermined, so there is no general solution. (For $m > 3$, $m(m-1)/2 > m$, and for $m = 3$, $\text{rank}(\mathbf{A}) = 2$.) Furthermore, the least-squares solution $\mathbf{g} = \mathbf{A}^\dagger \mathbf{c}$ does not in general satisfy the constraint on \mathbf{g} (that the elements g_i must be decreasing). Nevertheless, we can see from (7-3) that the element g_i should have a roughly *reciprocal* relationship to the elements d_i . The truly optimal choice for the elements g_i remains an open issue.

7.2.2 Blind Estimation of the LP Index

The primary shortcoming in the blind implementation of the generalized LP detector is how to determine the optimal index L . The brute-force technique of testing multiple indices, either serially or in parallel, substantially increases the recovery time (to steady state) or the complexity of the detector. This increase, in either case, is by a factor of $M+1$ if all possible indices are tested. In this section, we outline a promising alternative for

low-noise environments that at most roughly doubles the recovery time, and only slightly increases the detector complexity.

We first implement the blind adaptive LP detector with index $L = 0$. After convergence, we can use the detector itself to estimate the channel taps. If the noise variance is small, then at steady state, the covariance of the temporal prediction error is approximately given by

$$\Phi_e = E[\mathbf{e}_k \mathbf{e}_k^*] \approx \mathbf{H}_0 \mathbf{H}_0^* . \quad (7-5)$$

Let $\mathbf{M} \mathbf{D}^2 \mathbf{M}^*$ denote the unique minimum-norm generalized Cholesky factorization of $\mathbf{H}_0 \mathbf{H}_0^*$ (where \mathbf{M} is monic, lower triangular, and of dimension $m \times m$, and where \mathbf{D} is diagonal and of dimension $m \times n$). We can use the LP detector's spatial predictor $\hat{\mathbf{P}}_0$ to estimate \mathbf{M} according to $\hat{\mathbf{M}} = (\mathbf{I} - \hat{\mathbf{P}}_0)^{-1}$. The fact that $\mathbf{I} - \hat{\mathbf{P}}_0$ is lower triangular simplifies the calculation of its inverse. Similarly, we can use the LP detector's AGC $\hat{\mathbf{A}}$ to estimate \mathbf{D} . For low noise, exactly $m - n$ of the diagonal terms in the AGC inverse $\hat{\mathbf{A}}^{-1}$ are very small. Let \mathbf{J} denote the $n \times m$ matrix that discards these small terms, so that $\hat{\mathbf{D}} = \mathbf{J} \hat{\mathbf{A}}^{-1}$. An estimate of \mathbf{H}_0 is then given by $\hat{\mathbf{H}}_0 = \hat{\mathbf{M}} \hat{\mathbf{D}} \approx \mathbf{H}_0 \mathbf{U}$, where \mathbf{U} is an arbitrary $(n \times n)$ unitary matrix. We can use the temporal predictor to estimate the other channel taps. We first estimate the AR parameters as $[\hat{\mathbf{A}}_1 \dots \hat{\mathbf{A}}_N] = -[\hat{\mathbf{P}}_1 \dots \hat{\mathbf{P}}_N]$. Recall that the AR and MA parameters are related as follows:

$$[\mathbf{A}_1 \dots \mathbf{A}_N] \mathcal{H} = [\mathbf{H}_1 \dots \mathbf{H}_M \mathbf{0}_{m \times Nn}] . \quad (7-6)$$

Therefore, using $[\hat{\mathbf{A}}_1 \dots \hat{\mathbf{A}}_N]$ and $\hat{\mathbf{H}}_0$, we can recursively estimate the other channel taps as follows:

$$\hat{\mathbf{H}}_i = \sum_{j=0}^{i-1} \hat{\mathbf{A}}_{i-j} \hat{\mathbf{H}}_j \quad \forall i \in \{1, \dots, M\}. \quad (7-7)$$

These estimates, like $\hat{\mathbf{H}}_0$, have an arbitrary unitary ambiguity: $\hat{\mathbf{H}}_i \approx \mathbf{H}_i \mathbf{U}$. However, \mathbf{U} is norm preserving and thus has no effect on the estimate of the channel energy distribution: $tr(\hat{\mathbf{H}}_i \hat{\mathbf{H}}_i^*) \approx tr(\mathbf{H}_i \mathbf{H}_i)$. We simply choose the index L corresponding to the tap of greatest energy $tr(\mathbf{H}_L \mathbf{H}_L)$, and then re-adapt the LP detector with this new index.

Observe that \mathbf{U} also has no effect on the estimate of the energy distribution corresponding to a particular user j : $\|\hat{\mathbf{h}}_i^{(j)}\|^2 \approx \|\mathbf{h}_i^{(j)}\|^2 \quad \forall i, j$. Therefore, we might alternatively choose to implement n separate LP detectors, one for each user $j \in \{1, \dots, n\}$, and choose an index L_j for each corresponding to the tap having greatest energy $\|\mathbf{h}_{L_j}^{(j)}\|^2$ in column j . This approach optimizes the performance for each user, but at the expense of additional complexity.

The proposed algorithm is subject to certain inaccuracies. First, the initial estimates $[\hat{\mathbf{A}}_1 \dots \hat{\mathbf{A}}_N]$ and $\hat{\mathbf{H}}_0$ assume negligible noise. Second, the recursion of (7-7) accumulates errors as i increases. However, there are variations of this basic approach that may improve the accuracy. An estimate of the noise variance might be used somehow to improve the accuracy of the initial estimates of $[\hat{\mathbf{A}}_1 \dots \hat{\mathbf{A}}_N]$ and $\hat{\mathbf{H}}_0$. We might also estimate the channel taps in some other order. For example, we might start with an LP detector of index M , and then, using $[\hat{\mathbf{A}}_1 \dots \hat{\mathbf{A}}_N]$ and $\hat{\mathbf{H}}_M$, recursively estimate the preceding channels taps. We could in fact start with an LP detector of any index L , estimate \mathbf{H}_L , and then devise a recursion to estimate the other channel taps from \mathbf{H}_L . Future research might address the viability of these and other techniques for blindly estimating the LP index. (See, for example, correlation matching of section 7.2.5.)

7.2.3 Alternative LP-Based Architectures

Property 5-4, which states that the LP detector is information lossless up to the $n \times m$ truncation matrix $\mathbf{J} = [\mathbf{I} \ \mathbf{0}]$, suggests yet another detector architecture. We could modify the LP detector of Definition 5-7 (5-63), replacing the memoryless term \mathbf{J} by an $n \times m$ filter $\mathbf{J}(z)$ of sufficient memory such that the resulting detector $\mathbf{C}(z)$ is asymptotically MMSE for some set of user delays:

$$\mathbf{C}(z) = \mathbf{J}(z) \tilde{\mathbf{Q}} \mathbf{D}^{-1} \mathbf{M}^{-1} z^{-\tilde{N}} [\mathbf{I} - \mathbf{P}^{(-1)}(z)] [\mathbf{I} - \mathbf{P}^L(z)], \quad (7-8)$$

where all terms, except $\mathbf{J}(z)$, are as defined in Definition 5-7. Blind adaptive estimation of $\mathbf{C}(z)$ is straightforward. We first initialize one tap of an estimate $\hat{\mathbf{J}}(z)$ to \mathbf{J} , while initializing its other taps to zero. Then, holding $\hat{\mathbf{J}}(z)$ fixed, we adaptively estimate the other terms in (7-8) as we do those in $\mathbf{C}_{LP}(z)$ of (5-63). After convergence of these terms, the eye should be open sufficiently for decision-directed adaptation $\hat{\mathbf{J}}(z)$. The algorithm thus provides a seamless transition from $\mathbf{C}_{LP}(z)$ to an MMSE detector $\mathbf{C}(z)$.

There may also exist other viable LP-based architectures. For example, consider *two-sided* prediction. We could replace the series cascade of temporal prediction-error filters $[\mathbf{I} - \mathbf{P}^{-1}(z)] [\mathbf{I} - \mathbf{P}^L(z)]$ in $\mathbf{C}_{LP}(z)$ of (5-63) with an error filter of the form:

$$\mathbf{E}(z) = \mathbf{I} - \sum_{i=-1}^{-\tilde{N}} \mathbf{P}_i z^{-i} - \sum_{i=1}^N \mathbf{P}_i z^{-i}. \quad (7-9)$$

We can think of (7-9) as the *parallel* combination of two predictors, one forward and one backward, both one-step. There may indeed exist predictor coefficients such that the cascade $\mathbf{E}(z)\mathbf{H}(z)$ is memoryless. Existence of these coefficients would imply that the observation \mathbf{r}_k of (1-1) could be expressed as

$$\mathbf{r}_k = \sum_{i=1}^N \mathbf{P}_i \mathbf{r}_{k-i} + \sum_{i=-1}^{-\tilde{N}} \mathbf{P}_i \mathbf{r}_{k-i} + \mathbf{H}_L \mathbf{x}_k. \quad (7-10)$$

Unfortunately, we cannot solve for the coefficients $\{\mathbf{P}_i\}$ in terms of the block-Toeplitz matrix \mathcal{H} as we can for the standard AR model. (See Corollary 5-1.1 (5-8) and Theorem 5-5 (5-23).) Hence, the necessary and sufficient conditions for the existence of this model, in terms of the channel $\mathbf{H}(z)$, are not known. Nevertheless, further research in this vein might produce LP-based detectors with lower complexity or better performance than those we have already proposed.

7.2.4 Fading Channels

Future work might also seek to speed the convergence of the proposed algorithms for application to fading or other rapidly time-varying channels. All of the adaptive algorithms we have described are designed in the spirit of the LMS algorithm. We use an instantaneous estimate of the gradient in the updates of the predictors and AGC, and a rank-two instantaneous error in the rotator updates. We could of course adopt a philosophy more in the spirit of the recursive least-squares algorithms. Doing so would necessarily increase the numerical complexity of the algorithms, but it would also speed convergence, and extend the utility of the proposed algorithm to rapidly time-varying channels.

7.2.5 Correlation Matching

In this section, we describe a technique for blindly estimating the channel with indirect application to the multiuser detection problem. A tall channel can be identified up to a unitary ambiguity from the second-order statistics of its output. We propose an adaptive scheme based on a correlation matching principle. Consider the memoryless channel \mathbf{H} of

(3-1). Suppose the receiver has an initial estimate of the channel $\hat{\mathbf{H}}$ and its noise variance $\hat{\sigma}^2$. If the corresponding channel output autocorrelation estimate $\hat{\mathbf{H}}\hat{\mathbf{H}}^* + \hat{\sigma}^2\mathbf{I}$ matches that of the actual channel $\mathbf{H}\mathbf{H}^* + \sigma^2\mathbf{I}$, then the channel estimate is accurate to within a unitary ambiguity:

$$\hat{\mathbf{H}}\hat{\mathbf{H}}^* + \hat{\sigma}^2\mathbf{I} = \mathbf{H}\mathbf{H}^* + \sigma^2\mathbf{I} \Leftrightarrow \hat{\sigma}^2 = \sigma^2 \text{ and } \hat{\mathbf{H}} = \mathbf{H}\mathbf{U}. \quad (7-11)$$

The squared Frobenius distance of these correlation matrices serves as a quadratic cost function to develop an adaptive algorithm for estimating the channel. With $\mathbf{H}\mathbf{H}^* + \sigma^2\mathbf{I} = \mathbf{E}[\mathbf{r}_k\mathbf{r}_k^*]$, a deterministic update is given by

$$\hat{\mathbf{H}}_{k+1} = \hat{\mathbf{H}}_k - \mu(\hat{\mathbf{H}}_k\hat{\mathbf{H}}_k^* + \hat{\sigma}^2\mathbf{I} - \mathbf{E}[\mathbf{r}_k\mathbf{r}_k^*])\hat{\mathbf{H}}_k. \quad (7-12)$$

We may drop the expectation operator to produce a stochastic version. The estimate of the noise variance must be provided by other means, such as the subspace separator. Without loss of generality, we can restrict $\hat{\mathbf{H}}$ to be lower triangular to reduce numerical complexity. As described this algorithm has indirect application to the blind multiuser detection problem in that it can be used to implement an MMSE, decision-feedback, or maximum-likelihood detector for memoryless channels. Furthermore, the idea easily extends to channels with memory by replacing \mathbf{H} and $\hat{\mathbf{H}}$ with \mathcal{H} and $\hat{\mathcal{H}}$ as in chapter 6. Correlation matching could also be useful for estimating the optimal index L for the LP detector. Future research might seek to fully develop these ideas.

B I B L I O G R A P H Y

- [1] N. Abramson, ed., *Multiple Access Communications: Foundations for Emerging Technologies*, IEEE Press, New York, 1993.
- [2] EIA/TIA IS-95, "Mobile Station-Base Station Compatibility Standard for Dual-Mode Wideband Spread Spectrum Cellular System.
- [3] A. J. Viterbi, *CDMA: Principles of Spread Spectrum Communication*, Addison-Wesley, Reading, MA, 1995.
- [4] G. L. Stüber, *Principles of Mobile Communication*, Kluwer, Boston, 1996.
- [5] ITU-T Recommendation V.32, "A Family of 2-Wire, Duplex Modems Operating at Data Signalling Rates of up to 9600 bits/s for Use on the General Switched Telephone Network and on Leased Telephone-Type Circuits," Version 3, 1993.
- [6] J. Salz, "Digital Transmission Over Cross-Coupled Linear Channels," *AT&T Technical Journal*, vol. 64, no. 6, pp. 1147-1159, July-August 1985.
- [7] M. L. Honig, P. Crespo, and K. Steiglitz, "Suppression of Near- and Far-End Crosstalk by Linear Pre- and Post-Filtering," *IEEE Journal on Selected Areas in Communications*, vol. 10, no. 3, pp. 614-629, April 1992.
- [8] P. A. Voois and J. M Cioffi, "Multichannel Signal Processing for Multiple-Head Digital Magnetic Recording," *IEEE Transactions on Magnetics*, vol. 30, no. 6, pp. 5100-5114, November 1994.
- [9] S. Verdú, "Recent Progress in Multiuser Detection," in [1], pp. 164-175.

- [10] R. Lupas and S. Verdú, "Linear Multiuser Detectors for Synchronous Code-Division Multiple-Access Channels," *IEEE Transactions on Information Theory*, vol. 35, no. 1, pp. 123-136, January 1989.
- [11] S. Verdú, *Multiuser Detection*, Cambridge University Press, Cambridge, UK, 1998.
- [12] S. Haykin, ed., *Blind Deconvolution*, Prentice-Hall, Englewood Cliffs, NJ, 1994.
- [13] M. Honig, U. Madhow, and S. Verdú, "Blind Adaptive Multiuser Detection," *IEEE Transactions on Information Theory*, vol. 41, no. 4, pp. 944-960, July 1995.
- [14] X. Wang and H. V. Poor, "Blind Adaptive Interference Suppression for CDMA Communication Based on Eigenspace Tracking," *Conference on Information Sciences and Systems*, pp. 468-473, Johns Hopkins University, Baltimore, 1997.
- [15] A. Gorokhov, P. Loubaton, and E. Moulines, "Second Order Blind Equalization in Multiple Input Multiple Output FIR Systems: A Weighted Least Squares Approach," *ICASSP*, vol. 5, pp. 2417-2420, Atlanta, 1996.
- [16] W. A Gardner, "A New Method of Channel Identification," *IEEE Transactions on Communications*, vol. 39, no 6, pp. 813-817, June 1991.
- [17] L. Tong, G. Xu, and T. Kailath, "Blind Identification and Equalization Based on Second-Order Statistics: A Time Domain Approach," *IEEE Transactions on Information Theory*, vol. 40, no. 2, pp. 340-349, March 1994.
- [18] R. T. Causey, "A Survey of the Hybrid Fiber-Coax Channel and Proposed Communication Techniques," Ph.D. Qualifying Examination Report, Georgia Institute of Technology, Atlanta, December 1995.
- [19] S. U. H. Qureshi, "Adaptive Equalization," *Proceedings of the IEEE*, vol. 73, no. 9, pp. 1349-1387, September 1985.
- [20] R. D. Gitlin and H. C. Meadors, Jr., "Center-Tap Tracking Algorithms for Timing Recovery," *AT&T Technical Journal*, vol. 66, no. 6, pp. 63-78, November-December 1989.

- [21] G. Ungerboeck, "Fractional Tap-Spacing Equalizer and Consequences for Clock Recovery in Data Modems," *IEEE Transactions of Communications*, vol. COM-24, no. 8, pp. 856-864, August 1976.
- [22] W. A. Gardner, ed., *Cyclostationarity in Communications and Signal Processing*, IEEE Press, New York, 1994.
- [23] J. M. Cioffi, "A Multicarrier Primer," ANSI T1E1.4 Contribution, November 1991.
- [24] J. A. C. Bingham, "Multicarrier Modulation for Data Transmission: An Idea Whose Time Has Come," *IEEE Communications Magazine*, vol. 28, no. 5, pp. 5-14, May 1990.
- [25] S. B. Weinstein and P. M. Ebert, "Data Transmission by Frequency-Division Multiplexing," *IEEE Transactions on Communications*, vol. 19, no. 10, pp. 628-634, October 1971.
- [26] N. J. Fliege, "Orthogonal Multiple Carrier Data Transmission," *European Transactions on Telecommunications*, vol. 3, no. 3, pp. 255-264, May 1992.
- [27] ANSI T1.413-1998, "Telecommunications — Network and Customer Installation Interfaces — Asymmetrical Digital Subscriber Line (ADSL) Metallic Interface," 1998.
- [28] ETS 300 401, "Radio Broadcast Systems: Digital Audio Broadcasting to Mobile, Portable, and Fixed Receivers," March 1994.
- [29] ETS 300 744, "Radio Broadcast Systems for Television, Sound and Data Services: Framing Structure, Channel Coding, and Modulation for Digital Terrestrial Television," March 1997.
- [30] M. Kavehrad and J. Salz, "Cross-Polarization Cancellation and Equalization in Digital Transmission Over Dually Polarized Multipath Fading Channels," *AT&T Technical Journal*, vol. 64, no. 10, pp. 2211-2245, December 1985.

- [31] G. J. Foschini, "Layered Space-Time Architecture for Wireless Communication in a Fading Environment When Using Multiple Antennas," *Bell Labs Technical Journal*, vol. 1. no. 2, pp. 41-59, Autumn 1996.
- [32] A. V. Oppenheim and A. S. Willsky, *Signals and Systems*, Prentice-Hall, Englewood Cliffs, NJ, 1983.
- [33] A. V. Oppenheim and R. W. Schaffer, *Discrete-Time Signal Processing*, Prentice-hall, Englewood Cliffs, NJ, 1989.
- [34] T. Kailath, *Linear Systems*, Prentice-Hall, Englewood Cliffs, NJ, 1980.
- [35] C. A. Desoer and J. D. Schulman, "Zeros and Poles of Matrix Transfer Functions and Their Dynamical Interpretation," *IEEE Transactions on Circuits and Systems*, vol. CAS-21, pp. 3-8, 1974.
- [36] A. G. J. MacFarlane and N. Karcanias, "Poles and Zeros of Linear Multivariable Systems: A Survey of the Algebraic, Geometric and Complex Variable Theory," *International Journal of Control*, vol. 24, pp. 34-74, 1976.
- [37] A. C. Pugh, "Transmission and System Zeros," *International Journal of Control*, vol. 26, pp. 315-324, August 1977.
- [38] P. Dent, B. Gudmundson, and M. Ewerbring, "CDMA-IC: A Novel Code Division Multiple Access Scheme Based on Interference Cancellation," *PIMRC*, pp. 98-102, Boston, 1992.
- [39] P. Patel and J. Holtzman, "Analysis of a DS/CDMA Successive Interference Cancellation Scheme Using Correlations," *Globecom*, vol. 1, pp. 76-80, Houston, 1993.
- [40] D. Divsalar, M. Simon, and D. Raphaeli, "Improved Parallel Interference Cancellation for CDMA," *IEEE Transactions on Communications*, vol. 46, no. 2, pp. 258-263, February 1998.

- [41] M. Varanasi and B. Aazhang, "Multistage Detection in Asynchronous Code-Division Multiple-Access Communication," *IEEE Transactions on Communications*, vol. 38, no. 4, pp. 509-519, April 1990.
- [42] Y. Yoon, R. Kohno, and H. Imai, "A Spread-Spectrum Multi-access System with Co-channel Interference Cancellation Over Multipath Fading Channels," *IEEE Journal on Selected Areas in Communications*, vol. 11, no. 7, pp. 1067-1075, September 1993.
- [43] M. Abdulrahman, A. Sheikh, and D. Falconer, "Decision Feedback Equalization for CDMA in Indoor Wireless Communication," *IEEE Journal on Selected Areas in Communications*, vol. 12, no. 4, pp. 698-706, May 1994.
- [44] A. Duel-Hallen, "Decorrelating Decision-Feedback Multiuser Detector for Synchronous CDMA," *IEEE Transactions on Communications*, vol. 41, no. 2, pp. 285-290, February 1993.
- [45] A. Duel-Hallen, "A Family of Multiuser Decision-Feedback Detectors for Asynchronous Code-Division Multiple-Access Channels," *IEEE Transactions on Communications*, vol. 43, no. 2-4, pp. 421-434, February-April 1995.
- [46] J. Yang and S. Roy, "Joint Transmitter/Receiver Optimization for Multi-input Multi-output Systems with Decision Feedback," *IEEE Transactions on Information Theory*, vol. 40, no. 5, pp. 1334-1347, September 1994.
- [47] W. van Etten, "Maximum Likelihood Receiver for Multiple Channel Transmission Systems," *IEEE Transactions on Communications*, vol. 24, no. 2, pp. 276-283, February 1976.
- [48] J. R. Barry and A. Batra, "Co-Channel Demodulation Using Multi-Input Multi-Output Equalization," *CRASP Annual Report*, July 30, 1996.
- [49] D. Shnidman, "A Generalized Nyquist Criterion and an Optimum Linear Receiver for a Pulse Modulation System," *Bell System Technical Journal*, vol. 46, pp. 2163-2177, November 1967.

- [50] E.A. Lee and D. G. Messerschmitt, *Digital Communication*, Second Edition, Kluwer, Boston, 1994.
- [51] J. G. Proakis, *Digital Communications*, Third Edition, McGraw-Hill, New York, 1995.
- [52] D. D. Falconer, M. Abdulrahman, N. W. K. Lo, and B. R. Peterson, "Advances in Equalization and Diversity for Portable Wireless Systems," *Digital Signal Processing*, vol. 3, no. 3, pp. 148-162, July 1993.
- [53] R. Lupas and S. Verdú, "Near-Far Resistance of Multiuser Detectors in Asynchronous Channels," *IEEE Transaction on Communications*, vol. 38, no. 4, pp. 496-508, April 1990.
- [54] J. J. Bussgang, "Crosscorrelation Functions of Amplitude-Distorted Gaussian Signals," MIT Technical Report No. 216, March 1952.
- [55] S. Bellini, "Bussgang Techniques for Blind Deconvolution and Equalization," in [3], pp. 8-59.
- [56] S. Bellini, "Bussgang Techniques for Blind Equalization," *Globecom*, vol. 3, pp. 1634-1640, Houston, 1986.
- [57] J. E. Mazo, "Analysis of Decision-Directed Equalizer Convergence," *The Bell System Technical Journal*, vol. 59, no. 10, pp. 1857-1886, December 1980.
- [58] B. Widrow and S. D. Stearns, *Adaptive Signal Processing*, Prentice-Hall, Englewood Cliffs, NJ, 1985.
- [59] S. Haykin, *Adaptive Filter Theory*, Second Edition, Prentice-Hall, Englewood Cliffs, NJ, 1991.
- [60] Y. Sato, "A Method of Self-Recovering Equalization for Multilevel Amplitude Modulation Systems," *IEEE Transactions on Communications*, vol. COM-23, no. 6, pp. 679-682, June 1975.

- [61] Y. Sato, H. Oda, and S. Hashimoto, "Blind Suppression of Time Dependency and its Extension to Multi-Dimensional Equalization," *Globecom*, vol. 3, pp. 1652-1656, Houston, 1986.
- [62] D. N. Godard, "Self-Recovering Equalization and Carrier Tracking in Two-Dimensional Data Communication Systems," *IEEE Transactions on Communications*, vol. COM-28, no. 11, pp. 1867-1875, November 1980.
- [63] J. R. Treichler and B. G. Agee, "A New Approach to Multipath Correction of Constant Modulus Signals," *IEEE Transactions on Acoustics, Speech, and Signal Processing*, vol. ASSP-31, no. 2, pp. 459-472, April 1983.
- [64] O. Shalvi and E. Weinstein, "New Criteria for Blind Deconvolution of Nonminimum Phase Systems (Channels)," *IEEE Transactions on Information Theory*, vol. 36, no. 2, pp. 312-321, March 1990.
- [65] C. L. Nikias and A. P. Petropulu, *Higher-Order Spectra Analysis: A Nonlinear Signal Processing Framework*, Prentice-Hall, Englewood Cliffs, NJ, 1993.
- [66] J. R. Treichler, V. Wolff, and C. R. Johnson, Jr., "Observed Misconvergence in the Constant Modulus Adaptive Algorithm," *Asilomar Conference on Signals, Systems, and Computers*, vol. 2, pp. 663-667, Pacific Grove, CA, 1991.
- [67] Z. Ding, C. R. Johnson, Jr., and R. A. Kennedy, "On the (Non)existence of Undesirable Equilibria of Godard Blind Equalizers," *IEEE Transactions on Signal Processing*, vol. 40, no. 10, pp. 2425-2432, October 1992.
- [68] G. J. Foschini, "Equalizing without Altering or Detecting Data," *AT&T Technical Journal*, vol. 64, no. 8, pp. 1885-1911, October 1985.
- [69] A. Benveniste, et. al., "Robust Identification of a Nonminimum Phase System: Blind Adjustment of a Linear Equalizer in Data Communications," *IEEE Transactions on Automatic Control*, vol. AC-25, no. 3, pp. 385-399, June 1980.
- [70] A. Benveniste and M. Goursat, "Blind Equalizers," *IEEE Transactions on Communications*, vol. COM-32, no. 8, pp 871-883, August 1984.

- [71] G. Picchi and G. Prati, "Blind Equalization and Carrier Recovery Using a 'Stop-and-Go' Decision-Directed Algorithm," *IEEE Transactions on Communications*, vol. COM-35, no. 9, pp. 877-887, September 1987.
- [72] Y. Chen, C.L. Nikias, and J. G. Proakis, "CRIMNO: Criterion with Memory Non-linearity for Blind Equalization," *Asilomar Conference on Signals, Systems, and Computers*, pp. 694-698, Pacific Grove, CA, 1991.
- [73] P. Tsakalides and C. L. Nikias, "A New Criterion for Blind Deconvolution of Colored Input Signals," *Asilomar Conference on Signals, Systems, and Computers*, vol. 1, pp. 746-750, Pacific Grove, CA, 1993.
- [74] F.-C. Zheng, S. McLaughlin, and B. Mulgrew, "Blind Equalization of Multilevel PAM Data for Nonminimum Phase Channels via Second- and Fourth-Order Cumulants," *Signal Processing*, vol. 31, no. 3, pp. 313-327, April 1993.
- [75] B. Porat and B. Friedlander, "Blind Adaptive Equalization of Digital Communication Channels Using High-Order Moments," *ICASSP*, vol. 2, pp. 1372-1375, Glasgow, Scotland, 1989.
- [76] B. Porat and B. Friedlander, "Blind Equalization of Digital Communication Channels Using High-Order Moments," *IEEE Transactions on Signal Processing*, vol. 39, no. 2, pp. 522-526, February 1991.
- [77] D. Hatzinakos, "Blind Equalization Using Decision Feedback Prediction and Tricepstrum Principles," *Signal Processing*, vol. 36, no. 3, pp. 261-276, April 1994.
- [78] Y. Chen and C. L. Nikias, "Adaptive Blind Equalization" *Conference on Moments and Signal Processing*, September 1992.
- [79] N. Seshadri, "Joint Data and Channel Estimation Using Fast Blind Trellis Search Techniques," *Globecom*, vol. 3, pp. 1659-1663, San Diego, December 1990.
- [80] M. Ghosh and C. L. Weber, "Maximum Likelihood Blind Equalization," *SPIE Conference*, vol. 1565, pp. 188-195, San Diego, CA, 1991.

- [81] C. F. Wong and T. L. Fine, "Adaptive Blind Equalization Using Artificial Neural Networks," *IEEE International Conference on Neural Networks*, vol. 4, pp. 1974-1979, Washington, DC, 1996.
- [82] T. Chen and R. Chen, "Neural Network Approach to Blind Identification of Stochastic and Deterministic Signals," *Asilomar Conference on Signals, Systems, and Computers*, vol. 2, pp. 892-896, Pacific Grove, CA, 1994.
- [83] D. Donoho, "On Minimum Entropy Deconvolution," in *Applied Time-Series Analysis II*, D. Findley, ed., Academic Press, pp. 565-608, 1981.
- [84] A. Papoulis, *Probability, Random Variables, and Stochastic Processes*, Third Edition, McGraw-Hill, New York, 1991.
- [85] L. Tong *et al.*, "Blind Identification and Equalization Based on Second-Order Statistics: A Frequency Domain Approach," *IEEE Transactions on Information Theory*, vol. 41, no. 1, pp. 329-334, January 1995.
- [86] L. Tong, G. Xu, and T. Kailath, "Necessary and Sufficient Conditions for Channel Identification Based on Second-Order Statistics," *International Symposium on Information Theory*, p. 188, San Antonio, 1993.
- [87] E. Moulines *et al.*, "Subspace Methods for the Blind Identification of Multichannel FIR Filters," *ICASSP*, vol. 4, pp. 573-576, Adelaide, Australia, 1994.
- [88] E. Moulines *et al.*, "Subspace Methods for the Blind Identification of Multichannel FIR Filters," *IEEE Transactions on Signal Processing*, vol. 43, no. 2, pp. 516-525, February 1995.
- [89] R. O. Schmidt, "Multiple Emitter Location and Signal Parameter Estimation," *IEEE Transactions on Antennas and Propagation*, vol. AP-34, no. 3, pp. 276-280, March 1986.
- [90] K. Abed-Meraim, *et al.*, "Prediction Error Methods for Time-Domain Blind Identification of Multichannel FIR Filters," *ICASSP*, vol. 3, pp. 1968-1971, Detroit 1995.

- [91] K. Abed-Meraim, E. Moulines, and P. Loubaton, "Prediction Error Method for Second-Order Blind Identification," *IEEE Transactions on Signal Processing*, vol. 45, no. 3, pp. 694-705, March 1997.
- [92] Z. Ding, "A Blind Channel Identification Algorithm Based on Matrix Outer-Product," *ICC*, vol. 2, pp. 852-856, Dallas, 1996.
- [93] H. Liu and G. Xu, "A Deterministic Approach to Blind Symbol Estimation," *IEEE Signal Processing Letters*, vol. 1, no. 12, pp. 205-207, December 1994.
- [94] D. Slock, "Blind Fractionally Spaced Equalization, Perfect Reconstruction Filter Banks and Multichannel Linear Prediction," *ICASSP*, vol. 4, pp. 585-588, Adelaide, Australia, 1994.
- [95] D. Slock and C. Papadias, "Blind Fractionally Spaced Equalization Based on Cyclostationarity," *VTC*, pp. 1286-1290, Stockholm, 1994.
- [96] D. Slock and C. Papadias, "Further Results on Blind Identification and Equalization of Multiple FIR Channels," *ICASSP*, vol.3, pp. 1964-1967, Detroit, 1995.
- [97] P. P. Vaidyanathan, *Multirate Systems and Filterbanks*, Prentice-Hall, Englewood Cliffs, NJ, 1993.
- [98] G. B. Giannakis and S. D. Halford, "Blind Fractionally Spaced Equalization of Noisy FIR Channels: Direct and Adaptive Solutions," *IEEE Transactions on Signal Processing*, vol. 45, no. 9, pp. 2277-2292, September 1997.
- [99] D. H. Johnson and D. E. Dudgeon, *Array Signal Processing: Concepts and Techniques*, Prentice-Hall, Englewood Cliffs, NJ, 1993.
- [100] L. J. Griffiths and M. J. Rude, "The P-Vector Algorithm: A Linearly Constrained Point of View," *Asilomar Conference on Signals, Systems and Computers*, pp. 457-461, Pacific Grove, CA, 1986.
- [101] R. Roy and T. Kailath, "ESPRIT — Estimation of Signal Parameter via Rotational Invariance Techniques," *IEEE Transactions on Acoustics, Speech, and Signal Processing*, vol. 37, no. 7, pp. 984-995, July 1989.

- [102] J.-F. Cardoso, "Source Separation using Higher Order Moments," *ICASSP*, vol. 4, pp. 2109-2112, Glasgow, Scotland, 1989.
- [103] J.-F. Cardoso and A. Souloumiac, "Blind Beamforming for non-Gaussian Signals", *IEE Proceedings F (Radar and Signal Processing)*, vol. 140, no. 6, pp. 362-370, December 1993.
- [104] J.-F. Cardoso, "Eigen-Structure of the Fourth-Order Cumulant Tensor with Application to the Blind Source Separation Problem", *ICASSP*, vol. 5, pp. 2655-2658, Albuquerque, 1990.
- [105] G. B. Giannakis, Y. Inouye, and J. M. Mendel, "Cumulant Based Information of Multichannel Moving-Average Models," *IEEE Transactions on Automatic Control*, vol. 34, no.7, pp. 783-787, July 1989.
- [106] A. Swami, G. B. Giannakis, and S. Shamsunder, "Multichannel ARMA Processes," *IEEE Transactions on Signal Processing*, vol. 42, no. 4, pp. 898-913, April 1994.
- [107] J. T. Tugnait, "On Blind MIMO Channel Estimation and Blind Signal Separation in Unknown Additive Noise," Proceedings of the *IEEE Signal Processing Workshop on Signal Processing Advances in Wireless Communications (SPAWC)*, pp. 53-60, Paris, 1997.
- [108] P. Comon, "Independent Component Analysis, A New Concept?" *Signal Processing*, vol. 36, no. 3, pp. 287-314, April 1994.
- [109] P. Comon, "Contrasts for Multichannel Blind Deconvolution," *IEEE Signal Processing Letters*, vol. 3, no. 7, pp. 209-211, July 1996.
- [110] E. Moreau and J.-C. Pesquet, "Generalized Contrasts for Multichannel Blind Deconvolution of Linear Systems," *IEEE Signal Processing Letters*, vol. 4, no. 6, pp. 182-183, June 1997.

- [111] J. R. Treichler and M. G. Larimore, "New Processing Techniques Based on the Constant Modulus Adaptive Algorithm," *IEEE Transactions on Acoustics, Speech and Signal Processing*, vol. 33, no. 2, pp. 420-431, April 1985.
- [112] A. Batra and J. R. Barry, "Blind Cancellation of Co-Channel Interference," *Globe-com*, vol. 1, pp. 157-162, Singapore, 1995.
- [113] H. Oda, Y. Sato, "A Method of Multi-Dimensional Blind Equalization," *IEEE International Symposium on Information Theory*, p. 327, San Antonio, 1993.
- [114] L. Castedo, C. J. Escudero, and A. Dapena, "A Blind Signal Separation Method for Multiuser Communications," *IEEE Transactions on Signal Processing*, vol. 45, no. 5, pp. 1343-1348, May 1997.
- [115] C. B. Papadias and A. J. Paulraj, "A Constant Modulus Algorithm for Multiuser Signal Separation in Presence of Delay Spread Using Antenna Arrays," *IEEE Signal Processing Letters*, vol. 4, no. 6, pp. 178-181, June 1997.
- [116] J.-F. Cardoso and B. H. Laheld, "Equivariant Adaptive Source Separation", *IEEE Transactions on Signal Processing*, vol. 44, no.12, pp. 3017-3030, December 1996.
- [117] J. R. Barry and A. Batra, "A Multidimensional Phase-Locked Loop for Blind Equalization of Multi-Input Multi-Output Channels," *ICC*, vol. 3, pp. 1307-1312, Dallas, 1996.
- [118] A. Gorokhov, P. Loubaton, "Blind Identification of MIMO-FIR Systems: A Generalized Linear Prediction Approach," *Signal Processing*, vol. 73, no. 1-2, pp. 105-124, 1998.
- [119] S. Icart and R. Gautier, "Blind Separation of Convolutional Mixtures Using Second and Fourth-Order Moments," *ICASSP*, vol. 5, pp. 3108-3021, Atlanta, 1996.
- [120] M. L. Honig, "A Comparison of Subspace Adaptive Filtering Techniques for DS-CDMA Interference Suppression," vol. 2, pp. 836-840, *Milcom*, Monterey, CA., 1997.

- [121] R. A. Horn and C. R. Johnson, *Matrix Analysis*, Cambridge University Press, Cambridge, UK, 1985.
- [122] B. Yang, "Projection Approximation Subspace Tracking," *IEEE Transaction on Signal Processing*, vol. 43, no. 1, January 1995.
- [123] R. D. DeGroat, "Noniterative Subspace Tracking," *IEEE Transactions on Signal Processing*, vol. 40, no. 3, pp. 571-577, March 1992.
- [124] G. W. Stewart, "An Updating Algorithm for Subspace Tracking," *IEEE Transactions on Signal Processing*, vol. 40, no. 6, pp. 1535-1541, June 1992.
- [125] C. H. Bischof, "On Updating Signal Subspaces," *IEEE Transactions on Signal Processing*, vol. 40, no. 1, pp. 96-105, January 1992.
- [126] Z. Xie, R. T. Short, and C. K. Rushforth, "A Family of Suboptimum Detectors for Coherent Multiuser Communications," *IEEE Journal of Selected Areas in Communications*, vol. 8, no. 4, pp. 683-690, May 1990.
- [127] U. Madhow and M. L. Honig, "MMSE Interference Suppression for Direct-Sequence Spread-Spectrum CDMA," *IEEE Transactions on Communications*, vol. 42, no. 12, pp. 3178-3188, December 1994.
- [128] L. H. Brandenburg and A. D. Wyner, "Capacity of the Gaussian Channel with Memory: The Multivariate Case," *The Bell System Technical Journal*, vol. 53, no. 5, pp. 745-778, May-June 1974.
- [129] K. Abed Meraim, *et al.*, "Asymptotic Performance of Second-Order Blind Source Separation," *ICASSP*, vol. 4, pp. 277-280, Adelaide, Australia, 1994.
- [130] M. Gaeta and J. L. Lacoume, "Source Separation without a Priori Knowledge: the Maximum Likelihood Solution," *European Signal Processing Conference (EUSIPCO)*, vol. 1, pp. 621-624, Barcelona, 1990.
- [131] L. Tong, Y. Inouye, and R. Liu, "Waveform Preserving Blind Estimation of Multiple Independent Sources," *IEEE Transactions on Signal Processing*, vol. 41, no. 7, pp. 2461-2470, July 1993.

- [132] R. T. Causey and J. R. Barry, "Blind Multiuser Detection Using Linear Prediction," *IEEE Journal on Selected Areas in Communications*, vol. 16, no. 9, pp. 1702-1710, December 1998.
- [133] R. T. Causey and J. R. Barry, "An Adaptive Project-Whiten-Rotate Blind Multiuser Detector," *IEEE Globecom's Seventh Communication Theory Mini-Conference*, pp. 125-130, Sydney, November 1998.
- [134] G. H. Golub and C. F. Van Loan, *Matrix Computations*, Third Edition, Johns Hopkins University Press, Baltimore, 1996.
- [135] N. Delfousse and P. Loubaton, "Adaptive Separation of Independent Source: a Deflation Approach," *ICASSP*, vol. 4, pp. 41-44, Adelaide, Australia, 1994.
- [136] M. H. Hayes, *Statistical Digital Signal Processing and Modeling*, Wiley, New York, 1996.
- [137] G. D. Forney, "Minimal Bases of Rational Vector Spaces, with Applications to Multivariable Linear Systems," *SIAM Journal on Control*, vol. 13, no. 3, pp. 493-520, May 1975.
- [138] C. R. Johnson *et al.*, "Blind Equalization Using the Constant Modulus Criterion: A Review," *Proceedings of the IEEE*, vol. 86, no. 10, pp. 1927-1950, October 1998.
- [139] J. P. LeBlanc, I. Fijalkow, and C. R. Johnson, Jr., "Fractionally Spaced Constant Modulus Algorithm Blind Equalizer Error Surface Characterization: Effects of Source Distributions," *ICC*, vol. 5, pp. 2244-2947, Atlanta, 1996.
- [140] I. Kalet, "The Multitone Channel," *IEEE Transactions on Communications*, vol. 37, no. 2, pp. 119-124, February 1989.
- [141] S. Kasturia, J. T. Aslanis, and J. M. Cioffi, "Vector Coding for Partial Response Channels," *IEEE Transactions on Information Theory*, vol. 36, no. 4, pp. 741-762, July 1990.
- [142] M. Tomlinson, "New Automatic Equaliser Employing Modulo Arithmetic," *Electronics Letters*, vol. 7, nos. 5/6, pp. 138-139, March 25, 1971.

- [143] H. Harashima and H. Miyakawa, "Matched-Transmission Technique for Channels with Intersymbol Interference," *IEEE Transactions on Communications*, vol. COM-20, no. 4, pp. 774-780, August 1972.
- [144] J. R. Barry, E. A. Lee, and D. G. Messerschmitt, "Capacity Penalty due to Ideal Zero-Forcing Decision-Feedback Equalization," *IEEE Transactions on Information Theory*, vol. 42, no. 4, pp. 1062-1071, July 1996.
- [145] R. T. Causey and J. R. Barry, "A Fully Blind MMSE Multiuser Detector," to appear, *IEEE Wireless Communications and Networking Conference*, New Orleans, September 1999.

VITA

Richard Todd Causey was born in Birmingham, Alabama on August 28, 1963 and grew up in nearby Tuscaloosa. He received the B.S. degree summa cum laude in 1985 from the University of Alabama, and the M.S. and Ph.D. degrees in 1993 and 1999, respectively, from the Georgia Institute of Technology, all in electrical engineering. As design engineer at Motorola from 1985 to 1988, he was involved in the design of high-speed modems for the public telephone network. With ADTRAN from 1988 to 1992, he helped to develop alternatives to conventional voice-band modems including ISDN and digital subscriber-line (DSL) products. His research interests include multiuser communication theory, wireless, and broadband systems.