Uppsala University
Signals and Systems

SPACE-TIME PROCESSING AND EQUALIZATION FOR WIRELESS COMMUNICATIONS

Erik Lindskog

Dissertation in Signal Processing to be publicly examined in room K23, Magistern, Dag Hammarskjölds väg 31, Uppsala, on June 7, 1999, at 10.15 a.m., for the degree of Doctor of Philosophy. The discussion will be held in Swedish.

ABSTRACT

Erik Lindskog, 1999. Space-time processing and equalization for wireless communications. Uppsala, 318pp. ISBN 91-506-1350-2.

In this thesis several aspects of space-time processing and equalization for wireless communications are treated. We discuss several different methods of improving estimates of space-time channels, such as temporal parametrization, spatial parametrization, reduced rank channel estimation, bootstrap channel estimation, and joint estimation of an FIR channel and an AR noise model. In wireless communication the signal is often subject to intersymbol interference as well as interference from other users. We here discuss space-time decision feedback equalizers and space-time maximum likelihood sequence estimators, which can alleviate the impact of these factors. In case the wireless channel does not experience a large amount of coupled delay and angle spread, sufficient performance may be obtained by an equalizer with a less complex structure. We therefore discuss various reduced complexity equalizers and symbol sequence estimators. We also discuss re-estimating the channel and/or re-tuning the equalizer with a bootstrap method using estimated symbols. With this method we can improve the performance of the channel estimation, the equalization, and the interferer suppression. This method can also be used to suppress asynchronous interferers. When equalizers and symbol detection algorithms are designed based on estimated channels we need to consider how errors in the estimated channels, or errors due to time variations, affect the performance of the equalizer or symbol detector. We show that equalizers tuned based on ordinary least squares estimated channels exhibit a degree of self-robustification, which automatically compensates for potential errors in the channel estimates.

Key-words: Space-time processing, channel estimation, equalization, maximum likelihood sequence estimation, decision feedback equalization, interference suppression, robustness.

Erik Lindskog, Signals and Systems, Uppsala University, PO Box 528, SE-751 20 Uppsala, Sweden.

Uppsala University

Signals and Systems

SPACE-TIME PROCESSING AND EQUALIZATION FOR WIRELESS COMMUNICATIONS

Erik Lindskog

UPPSALA UNIVERSITY 1999

Dissertation for the degree of Doctor of Philosophy
in Signal Processing at Uppsala University, 1999

ABSTRACT

Erik Lindskog, 1999. Space-time processing and equalization for wireless communications. Uppsala, 318pp. ISBN 91-506-1350-2.

In this thesis several aspects of space-time processing and equalization for wireless communications are treated. We discuss several different methods of improving estimates of space-time channels, such as temporal parametrization, spatial parametrization, reduced rank channel estimation, bootstrap channel estimation, and joint estimation of an FIR channel and an AR noise model. In wireless communication the signal is often subject to intersymbol interference as well as interference from other users. We here discuss space-time decision feedback equalizers and space-time maximum likelihood sequence estimators, which can alleviate the impact of these factors. In case the wireless channel does not experience a large amount of coupled delay and angle spread, sufficient performance may be obtained by an equalizer with a less complex structure. We therefore discuss various reduced complexity equalizers and symbol sequence estimators. We also discuss re-estimating the channel and/or re-tuning the equalizer with a bootstrap method using estimated symbols. With this method we can improve the performance of the channel estimation, the equalization, and the interferer suppression. This method can also be used to suppress asynchronous interferers. When equalizers and symbol detection algorithms are designed based on estimated channels we need to consider how errors in the estimated channels, or errors due to time variations, affect the performance of the equalizer or symbol detector. We show that equalizers tuned based on ordinary least squares estimated channels exhibit a degree of self-robustification, which automatically compensates for potential errors in the channel estimates.

Key-words: Space-time processing, channel estimation, equalization, maximum likelihood sequence estimation, decision feedback equalization, interference suppression, robustness.

Erik Lindskog, Signals and Systems, Uppsala University, PO Box 528, SE-751 20 Uppsala, Sweden.

ISBN 91-506-1350-2

Printed in Sweden by Elanders Graphic Systems AB, Angered 1999
Distributed by Signals and Systems, Uppsala University, Uppsala, Sweden

To my parents

Preface

This thesis consists of seven chapters. The first chapter is an overview of space-space time processing for wireless communications and serves as an introduction to the subject. The remaining chapters treat various aspects of channel estimation, equalization and interferer suppression.

Parts of the material in the respective chapters has been published in the following papers:

Chapter 1:

Arogyaswami Paulraj and Erik Lindskog,
"A Taxonomy of space-time processing for wireless networks", IEE Proceedings, Radar, Sonar and Navigation, vol. 145, no. 1, February 1998.

Chapter 2:

Erik Lindskog,
"Array Channel Identification Using Directional of Arrival Parametrization". Proceedings of IEEE International Conference on Universal Personal Communications, Cambridge, Massachusetts, U.S.A., Sept 29 - Oct 2 1996, vol 2, pp 999-1003.

Erik Lindskog,
"Channel Estimation Exploiting Pulse Shaping Information - A Channel Interpolation Approach", UPTEC Report 97138R, Uppsala University, Signals and Systems, 1997, PO Box 528, SE-751 20 Uppsala, Sweden.

Erik Lindskog and Jonas Strandell,
"Multi-user Channel Estimation Exploiting Pulse Shaping Information". Proceedings of European Signal Processing Conference, Rhodos, Greece, Sept 8-11 1998.

Erik Lindskog and Claes Tidestav,
"Reduced rank channel estimation", IEEE Vehicular Technology Conference, Houston, Texas, USA, May 16-20 1999.

The methods in Section 2.3 and 2.5 has also been combined and published in:

Jonas Strandell and Erik Lindskog,
"Channel estimation by maximum likelihood projection onto a parametrized subspace". Proceedings of European Signal Processing Conference, Rhodos, Greece, Sept 8-11 1998,

Chapter 3:

Erik Lindskog, Anders Ahlén and Mikael Sternad,
"Combined Spatial and Temporal Equalization using an Adaptive Antenna Array and a Decision Feedback Equalization Scheme", Proceedings of Int. Conf. on Acoustics, Speech and Signal Processing, Detroit, Michigan, U.S.A., May 8-12 1995, vol. 2, pp 1189-1192.

Erik Lindskog, Anders Ahlén and Mikael Sternad,
"Spatio-Temporal Equalization for Multipath Environments in Mobile Radio Applications". Proceedings of 45h IEEE Vehicular Technology Conference, Rosemont, Illinois, USA, vol. 1, July 26-29 1995, pp 399-403.

Chapter 4:

Erik Lindskog,
"Multi-channel maximum likelihood sequence estimation". Proceedings of the 47th IEEE Vehicular Technology Conference, Phoenix, Arizona, USA, May 5-7 1997, vol. 2, pp 715-719.

Chapter 5:

Erik Lindskog,
"Making SMI-beamforming insensitive to the sampling timing for GSM signals", Proceedings of the Sixth International Symposium on Personal, Indoor and Mobile Radio Communications, Toronto, Canada, September 27-29 1995, vol. 2, pp 664-668

Erik Lindskog and Claes Tidestav,
"Reduced rank equalization", In Proceedings of IEEE International Symposium on Personal, Indoor and Mobile Radio Communication, Boston, Massachusetts, USA, September 8-11 1998, Vol 3, pp 1081-1085.

Chapter 6:

Erik Lindskog,
"Combatting Co-Channel Interference in a TDMA System Using Interference Estimates From Adjacent Frames". Proceedings of 29th Asilomar Conference on Signals, Systems & Computers, Pacific Grove, California, U.S.A., Oct 30 - Nov 1 1995, vol. 1, pp 367-371.

Claes Tidestav and Erik Lindskog,
"Bootstrap Equalization", Proceedings of IEEE International Conference on Universal Personal Communications, Florence, Italy, October 5-9 1998.

Chapter 7:

Mikael Sternad and Anders Ahlén and Erik Lindskog,
"Robust Decision Feedback Equalizers", Proceedings of Int. Conf. on Acoustics, Speech and Signal Processing, Minneapolis, Minnesota, U.S.A., April 1993, vol. 3, pp 555-558.

Erik Lindskog, Mikael Sternad and Anders Ahlén,
"Designing Decision Feedback Equalizers to be Robust with respect to Channel Time Variations", Proceedings of Nordic Radio Symposium seminar, Uppsala, Sweden, November 10-11 1993.

Acknowledgments

First, I wish to express my gratitude to my supervisors Professor Anders Ahlén and Dr. Mikael Sternad for their valuable guidance and support during my time as a graduate student. It has been very stimulating and rewarding to work in their group. Anders always has a positive outlook on things, and Mikael will always find interest in every problem you approach him with. I also want to give my thanks to them for their careful reading of various manuscripts, including this thesis.

I would also like to thank my other co-authors on various papers, Professor Arogyaswami Paulraj, Claes Tidestav, Jonas Strandell, Mattias Wennström, Dr. Tommy Öberg Dr. Anders Rydberg and Yulianto Naserudin for interesting and rewarding co-operation. Furthermore, I would like to thank Mats Cedervall and Boon Ng for sharing some of their insights into the problem of temporal parametrization of a channel.

I am also grateful to Dr Sören Andersson, Dr Ulf Forssén, and Henrik Damm at Ericsson Radio Systems for the experimental data that has been used in some of the evaluations in this thesis.

My thanks also to the Swedish Research Council for Engineering Sciences (TFR), who has been the major financial supporter of this work under contracts 281-95-800 and 281-98-654.

During my time as a graduate student I have belonged to two different departments, System and Control in “House 8” and Signals and Systems in “Magistern”. It has been joyful and rewarding to meet and work with the members of both departments and I would like express my warmest gratitude to all of them. I would also like to thank our shared system administrator, Ove Ewerlid, for valuable help with computer problems and for all the interesting discussions we have had at our favorite “hang-out”. A special mentioning also of my first office mate Kenth Öhrn for the good company and to Fredrik Lingvall for introducing me to climbing in Uppsala.

I would also like to thank Professor John Proakis for the opportunity to visit him and his department at Northeastern University in Boston, USA. I had a great year at Northeastern University. I especially appreciated the group meetings where many interesting presentations were held.

I made many friends in Boston and I would like to thank them all for their friendship. Among these I would especially like to thank: Bjørn Bjerke, my Norwegian office mate who I shared a lot of time with. I now have a much greater understanding of the Norwegian language. José Fridman for his humorous comments and his valuable advice. Arda Aksu for his joyful and friendly attitude. Erozan Kurtas for several interesting discussions. My room mates from Sigourney st, Timoor Sahkaruk, Zoran Mihajlovic and Vladic Davidkovich for their companionship. Zoran Coric, Sasha Djakovic, Alex Pantelic and Thomas and Tanya Thiemann, for the good times they showed me. Håkan and Beth Thyr for showing me some of the "great outdoors" of America: Håkan introducing me to climbing and Beth for sailing with me on the Charles River.

I am also very grateful to Professor Arogyaswami Paulraj for the opportunity to visit him and his excellent research group at Stanford University, USA. I had a very stimulating and rewarding visit. I especially valued the open minded scientific discussions at the weekly group meetings. My thanks also to all the members in the group.

In California I had the privilege to become closer acquainted with my friend William Klippgen from Norway, who I had met briefly in Boston. I want to thank him for housing me when I arrived to California and for the great companionship he offered, then and later. I also would like to thank my special friends, Maja Popovic and Kanna Rajan, for all the things we did together during my stay in California. Further, my thanks to Roger Germundsson for a great Christmas and New Years Eve celebration and for interesting discussions about the construction, operation and repairs of a Corvette sports car.

My greatest thanks to my parents, Inger and Jan, as well as the rest of my family including my aunt Anita, for all their help and support. I also would like to thank my brother Olof for the joy of sharing a boat together and for the adventure of sailing the same boat across the sea to Gotland.

Finally I would like to thank Maider for her patience and love.

Contents

Chapter 1

Space-Time Processing in Wireless Communication

1.1 Introduction

With multiple antennas, received and transmitted signals can be separated not only with temporal processing but also with spatial processing. We call the combination of spatial and temporal processing *space-time processing*. Space-time processing is a tool for improving the overall economy and efficiency of a digital cellular radio system by exploiting the use of multiple antennas. Most current cellular radio modems do not, however, efficiently exploit the spatial dimension offered by multiple antennas. As outlined in this chapter, the spatial domain can be used to reduce co-channel interference, increase diversity gain, improve array gain, and reduce intersymbol interference [1]. These improvements can have significant impact on the overall performance of a wireless network. The aim of this thesis is to develop, explore and investigate signal processing algorithms that combine spatial and temporal processing, to attain results which cannot be obtained by either spatial or temporal processing individually.

[1] Diversity gain is the gain obtained by using multiple signals, or components of a signal, that fade independently. Array gain can be defined as the gain in signal to noise ratio that can be achieved with spatial beamforming. Intersymbol interference is the interference caused by delayed versions of the desired signal.

Space-time processing in a receiver improves the signal to interference ratio through co-channel interference cancellation, mitigates fading through improved receive diversity, offers higher signal to noise ratio through array gain and reduces intersymbol interference through spatial equalization. Likewise, space-time processing in the transmitter reduces co-channel interference generation, improves transmit diversity and in some cases also minimizes intersymbol interference generation.

In this thesis we will consider space-time processing algorithms for channel estimation, equalization and interferer suppression. The use of multiple antennas in a receiver has many advantages but it also has some disadvantages. From a signal processing point of view, the use of multiple antennas has two main disadvantages:

- Increased computational complexity.
- Increased difficulty to accurately tune the receiver algorithms based on short sequences of training data.

We will in this thesis discuss methods that can contribute to alleviate these problems. As an application we are mainly concerned with wireless TDMA communication systems although many of the ideas and algorithms can transfer to other types of systems and channels.

The algorithms, problems and solutions studied in the present thesis are summarized and introduced in the introductions of each of the chapters below.

In Chapter 2 we study algorithms for estimation of space-time channels. We explore the use of a priori information and known structure of the space-time channel, in order to improve estimates of the channel and of the spectrum of the noise plus interference. We also discuss joint estimation of an FIR channel and autoregressive (AR) noise model. Finally we also discuss re-estimating the channel iteratively utilizing detected symbols as an artificially created training sequence.

In Chapter 3 we present and derive different optimal space-time decision feedback equalizers and discuss some tuning options for the same equalizers.

In Chapter 4 we treat the space-time maximum likelihood sequence estimator. Several approaches to space-time maximum likelihood sequence estimation, which are basically equivalent, as presented. There exists two main approaches, the log-likelihood approach and the matched filter approach. We show how one approach can be derived from the other. We also discuss tuning options for the maximum likelihood sequence estimator.

In Chapter 5 we present and discuss different ways of performing reduced complexity space-time equalization. The main idea is that we can often, to some extent, decouple the spatial and the temporal processing. We generalize this concept into *reduced rank equalization* where we show how the the space-time structure of the channel can be exploited in order to design equalizers with a simplified space-time structure.

In Chapter 6 we present a method of improving on the equalization and the interferer suppression by re-tuning the equalizer iteratively, using detected symbols. We call this *bootstrap* equalization. Bootstrapping can significantly improve the suppression of co-channel interferers. We also present a method that improves on the suppression of co-channel interferers that appear *outside* the training sequence, by utilizing information from adjacent frames. This method is also combined with bootstrap equalization and the two methods are found to complement each other.

In Chapter 7 we present a method for designing equalizers that are robust to errors in the channel estimates and/or time variations in the channel. We show that an equalizer can be made robust with respect to decision errors, resulting in shorter error bursts, and that it can be made robust against some time variations in the channel. Finally we show that an indirectly tuned space-time decision feedback equalizer, utilizing a sample matrix estimate of the spatial spectrum of the noise plus interference, is automatically spatially robustified against channel estimation errors caused by strong co-channel interferers.

1.2 Outline of Space-Time Processing Schemes

To place the work of the thesis in a wider context, this introductory chapter will outline and classify how space-time processing enters in and affect different aspects of a wireless communication system. One classification can be based on the *system architecture*, and covers different design choices for the physical layer of the wireless network (Section 1.3). Another classification can be based on *algorithms* and refers to choices of signal processing algorithms and optimization criteria (Section 1.4). Underlying and affecting both these classifications are the *characteristics of the propagation channel* that include angle, delay and Doppler spreads (Section 1.5). See Figure 1.1.

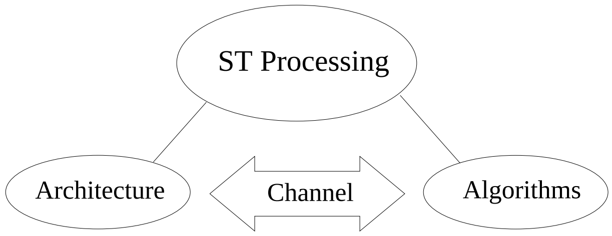


Figure 1.1: Space-time processing classifications according to architecture and algorithms, both affected by the channel properties.

1.3 Architecture Based Classification

An architecture oriented classification can be based on different choices of the physical layer design of the wireless system that is directly affected by space-time processing. We can perform architectural classification along the three directions shown in Figure 1.2. First, in the *Link Structure* we can make choices about where and how space-time processing is to be applied to the network elements. Next, in *Channel Reuse* we make choices for reusing the frequency spectrum. Finally, *Multiple Access* is an important aspect of the physical layer that affects space-time processing.

1.3.1 Link Structure

The link structure refers to all aspects of space-time processing related to the radio links between the base station and the subscriber. The link structure, in turn, can be classified based on the number of antennas at the base and the subscriber unit, and on the use of space-time processing at the receiver and at the transmitter. We discuss these issues below.

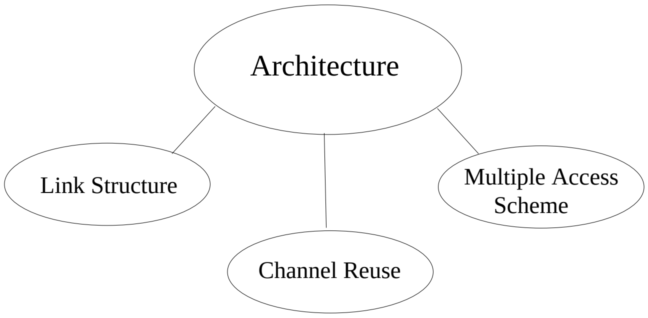


Figure 1.2: Architecture classification.

Space-Time Processing at the Base Station and the Subscriber unit

Space-time processing using multiple antennas can be applied at the base station, the subscriber unit or at both locations. The differences in propagation environment, physical limitations and cost constraints result in different choices of type and number of antennas at the base and the subscriber unit. Base stations can employ multiple antenna elements more easily because the size and cost constraints are less restrictive. The use of multiple antennas is an important source of diversity when the correlation between the antenna elements is not too high. At the subscriber unit, the presence of local scatterers provides adequate decorrelation with a spacing of 0.3 to 0.5 wavelengths between the antennas. At the base stations, were scatterers normally are more distant, a spacing of 5 to 10 wavelengths may be needed to obtain similar decorrelation [51]. For these reasons, the number of antennas, element design, spacing and topology have different drivers at the base and the subscriber unit. Space-time processing at the base station is the primary focus today although space-time processing at the subscriber unit is becoming more feasible. One example of the latter is the use of dual antennas in the handset in the North American PACS standard and in the Japanese standard PDC.

Receive and Transmit Space-Time Processing

Space-time processing can be used either in receive alone, in transmit alone or on both links. The factors that influence these choices are discussed here. The key difference between the two links is the difficulty in determining the transmit channel needed for transmit space-time processing.

Space-time processing performance in receive and transmit can be very different due to the differences in the knowledge of the associated channels. In receive, the channel can be estimated (by non-blind or blind methods) since the signal has traveled through the channel before being observed at the receiver. Also, the interference is present at the receiver input and can therefore be characterized and canceled, see Chapter 2. On the other hand, in transmit, the channel is encountered after the signal leaves the antenna array. The use of space-time processing in transmit therefore requires prior knowledge of the channel. Moreover, interference reduction in transmit requires knowledge of the channels to the co-channel subscribers. Again, these are difficult to estimate. Both these factors makes transmit space-time processing challenging. See [28].

Figure 1.3 shows different link structures obtained when varying the number of antennas used in receive and in transmit. These options can be associated with the downlink (base to subscriber) or the uplink (subscriber to base). Depending on the number of antennas, we can classify the channel as Single Input **(SI)** or Multiple Input **(MI)** for transmit and Single Output **(SO)** or Multiple Output **(MO)** for receive. In the following chapters of this thesis we shall focus on algorithms for the **SIMO** channel case, that is, a single transmit antenna and multiple receive antennas, see Figure 1.3. Some aspects of the methods discussed may however be applicable to other types of channels.

1.3.2 Channel Reuse

Channel Reuse Between Cells

The wireless channel can be shared among users in different ways. The different schemes for sharing of the channel are called multiple access schemes.

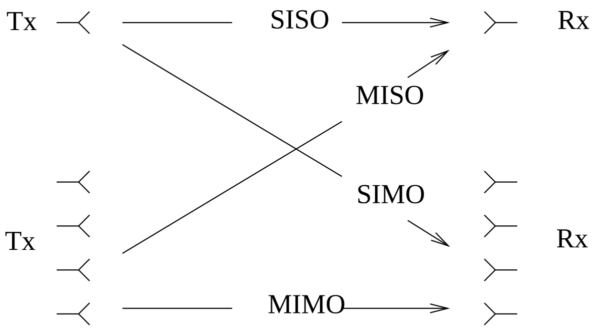


Figure 1.3: Link structures obtained for different numbers of antennas at the transmitter (Tx) and the receiver (Rx).

We can either assign each user a specific frequence band. This method is called *frequence division multiple access* (FDMA). Another alternative is to divide the channel into time slots and assign one time slot to each user. This method is called *time division multiple access* (TDMA). A third method is to separate the users by the use of codes. Each user then modulates his message with a specific code. The codes are deigned to be approximately orthogonal such that the users can be separated. This method is called *code division multiple access* (CDMA). Many practical systems are a combination of two, or even all three, of these methods.

Current TDMA systems employ channel reuse between cells. We expect to see only one desired signal at the base station or the subscriber unit, and interfering co-channel signals from other cells.

TDMA systems typically have a reuse factor[2], K, ranging from three to twelve. Smaller reuse factors offer higher spectral efficiency. The lower limit of the reuse factor depends on the tolerance to co-channel interference. Furthermore, one can use sectorization wherein a cell is divided into a number of equal sectors and the frequencies within the cell are divided among the sectors. The sectors in a cluster then all use different frequencies. Sectorization further reduces the effect of co-channel interference. It is typical to describe a cellular layout as K/L where K refer to the number of cells per

[2]The reuse factor is the number of cells in a cluster, where the channel allocation is repeated for each cluster of cells. Cells within a cluster do not use the same frequency. The reuse factor therefore determines how close a co-channel cell can be located.

cluster and L refer to the number of sectors per cluster.

CDMA systems generally have a reuse factor of one, i.e. the whole spectrum is reused in every cell. The cells can also here be divided into sectors where each sector in general will reuse the whole spectrum.

The wireless link from a subscriber to a base station is called the *uplink*, while the wireless link from a base station to a subscriber is called the *downlink*. Space-time processing can be used in both the uplink and the downlink to reduce co-channel interference. This can allow a decreased reuse factor in TDMA. In the uplink, the base station can use space-time processing to suppress strong co-channel interferers. In the downlink, the base station can have directive transmission to minimize interference to other co-channel users. Space-time processing can also be employed at the subscriber unit to reduce co-channel interference on both links.

Channel Reuse Within Cells

Considering TDMA, it may be possible to use space-time processing to support two or more links on the same channel within a cell. We can approach channel reuse from two view-points, reuse at the base station and reuse at the subscriber unit, see Figure 1.4. These are discussed below.

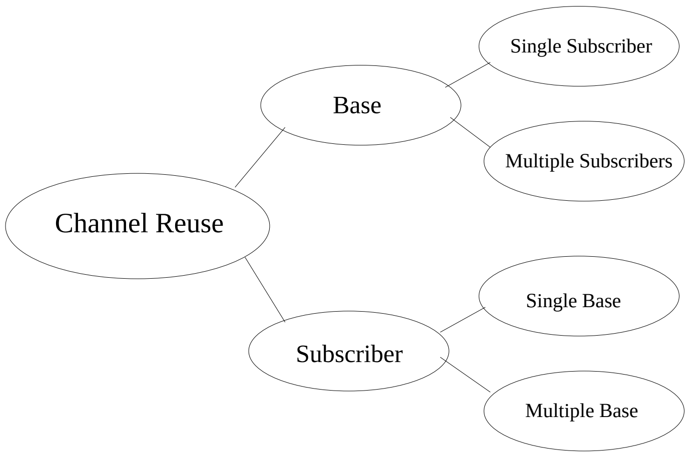


Figure 1.4: Channel reuse classification.

Base Centered: Single vs Multi Subscriber Operation

The case when only one subscriber per channel is supported within a cell is here referred to as the Single Subscriber case (SS). On the other hand, as mentioned above, it is possible to support reuse within a cell wherein we support multiple subscribers within the same cell (or sector) on the same channel. We call this the Multiple Subscriber case (MS). When supporting multiple subscribers per cell and channel in TDMA systems the signals typically have to be separated using space-time processing. This can either be performed with single user detector algorithms as the equalizers and symbol sequence estimators described in this thesis or it can be performed with multi-user detectors, see for example [127][25][107] and [105].

CDMA systems support multiple subscribers on the same frequency channel. The subscribers use different spreading codes and can therefore be separated with time processing alone. However, space-time processing can improve the performance.

The uplink and downlink in a communication system can have different channel reuse factors. We can, for instance, support aggressive reuse in the uplink, since receive space-time processing is easier to implement, and use less aggressive reuse in the downlink where channel estimation problems may limit co-channel interference cancellation. In order to balance the total number of subscribers in both links, asymmetric bandwidth assignment on the two links is then required.

Subscriber Centered: Single Base Source vs Multiple Base Source

Subscriber units normally receive the downlink signals from one base station. However, for subscriber units with multiple antennas it is possible to receive multiple co-channel signals, carrying different information signals, from different antennas at the same base station. Typically a high data rate signal would be split into multiple signals with lower data rates which are transmitted simultaneously from different antennas on the same channel. Space-time processing can be used to separate the co-channel signals and then combine these after demodulation resulting in higher spectral efficiency. The subscriber unit can do the same thing if it is equipped with multiple antennas, possibly using polarization diversity. Assuming the signals are received by the multiple antennas at the base station, they can then be separated using space-time processing. See e.g. [78, 131, 31, 27].

1.3.3 Multiple Access

The choice of multiple access (MA), i.e. how the available spectrum is shared among users, plays a major role in the design of space-time processing methods due to its effect on the characteristics of the co-channel interference.

In TDMA, since the signal is not spread, one or two strong sources of co-channel interference may be present in the reuse between cell configuration [34]. Space-time processing can be used to suppress these few, but strong, interferers. This can be performed in many different ways and several such methods will be discussed in the following chapters of this thesis.

In CDMA, all users share the same channel and are separated by different spreading codes, allowing time domain processing to reduce co-channel interference. Therefore, in CDMA, the space-time processing has to deal with a large number of weak interferers. This difference affects the strategies for co-channel interference suppression.

We can combine link, channel reuse and multiple access approaches to yield a variety of different architectures that employ space-time processing. Example configurations where multiple antennas are used at the base station together with reuse within cells in a TDMA system are shown in Figure 1.5.

1.4 Algorithm Based Classification

Algorithms for space-time processing can be divided into those used for channel estimation and those used for receive and transmit processing, see Figure 1.6. We will discuss algorithms in these groups separately below. The receive algorithms for TDMA and CDMA are treated separately. We will discuss algorithms for channel estimation and for TDMA reception in more detail since these are main applications of this thesis.

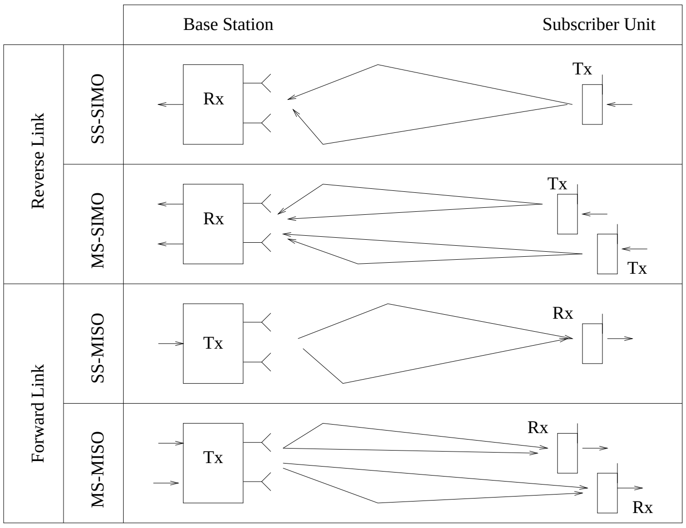


Figure 1.5: Example TDMA configurations with either a single subscriber (SS) or multiple subscribers (MS) transmitting with a single (SI) or multiple antennas (MI) and receiving with a single antenna (SO) or multiple antennas (MO).

1.4.1 Channel Estimation Algorithms

Receive Channel Estimation

In receive channel estimation algorithms we can use non-blind or blind methods. In blind methods, no training signals are available and the underlying structure of the channel and/or the signal modulation format can be used to estimate the channel. Blind methods for channel estimation have been an active area of research. See [81] for a review of these methods. In non-blind methods, training signals are transmitted along with the information signal so as to enable channel estimation by the receiver. In this thesis we will concentrate on non-blind methods using a short training sequence of known symbols, as in a TDMA system[3].

[3]Blind methods have the disadvantage that they typically require long data sequences and have difficulties in adapting to rapidly time-varying channels. Furthermore, training sequences exists in present and future proposed wireless TDMA and CDMA standards. They occupy at the most about 15 percent of the symbols. Therefore there is at the most a gain of 15 percent to be achieved by using blind methods.

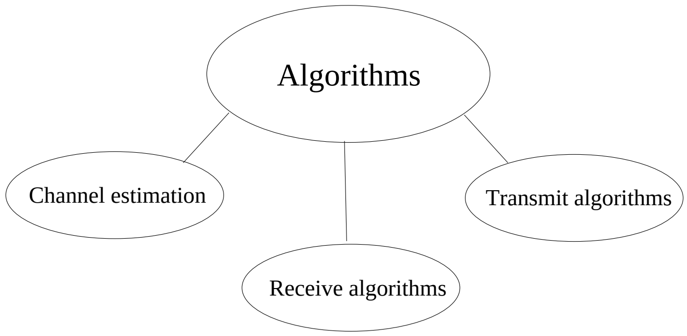


Figure 1.6: Algorithm classification.

In a TDMA system there is typically a short training sequence available that can be used for channel estimation. Most wireless radio channels can be well modeled with a discrete time FIR filter. The most straightforward way of channel estimation is to estimate the taps in the FIR filter with a least squares method. The spectrum of the noise and interference can then, for example, be estimated as sample matrix estimate for different time lags. We can, however, identify at least five different in which these estimates of the channel and noise properties can be improved:

1. Structural constraints and *a priori* information can be utilized to reduce the number of parameters that need to be estimated.

2. The data can be pre-filtered or projected onto a subspace to improve the signal to noise ratio.

3. The number of training data can be increased by utilizing detected symbols during time intervals where the symbol sequence is initially unknown.

4. Simplified parametric models can be estimated for the the noise and interference.

5. Joint estimation of the channel and a spatio-temporal noise model.

Different methods that utilize these ways to improve the estimates are presented in Chapter 2.

In Section 2.3, the known temporal filtering in the transmitter and the receiver is utilized to reduce the number of parameters that need to be identified. This can improve the estimate of the channel, especially if the training sequence is short and the pulse shaping performed in the transmitter (and the receiver) has a long time span. This is, for example, the case in the North American standard IS-136.

In Section 2.5, the channel is modeled with signal paths which are parametrized with their directions of arrival and and respective gains. These parameters are then identified and used to form an improved channel estimate. Here it is noted that the channel estimate can indeed be improved. However, the bit-error-rate (BER) of the equalizer is not improved to the same degree. This has to do with the fact that not all aspects of a channel estimate are be important, see the discussion at the end of this sub-section.

In Section 2.6, the spatial properties of the channel are exploited in a non-parametric fashion. The received signal or an initial estimate of the channel is here projected onto the signal subspace of the received signal. In this way the impact of noise on the channel estimate is reduced. When the received signal is projected onto the signal subspace the dimension of the problem is reduced, which results in lower computational complexity and improved estimates of the spectrum for the noise plus interference.

In Section 2.7, the number of training symbols is artificially increased by utilizing decided data to re-estimate the channel and the noise plus interferer spectrum after an initial equalization. The increase in the amount of training data improves the estimates of both the channel and the spectrum of the noise plus interference. This improves the equalization and the suppression of co-channel interferers.

In Section 2.8 we discuss estimating the spatio-temporal spectrum using the residuals from the channel estimation. We there note, that it may be difficult to form an accurate estimate of the spatio-temporal spectrum of the noise plus interference, when we have many antennas and only a short training sequence. By restricting us to model only the spatial spectrum of the interference, it is however often possible to get a useful estimate. The use of only the spatial spectrum of the noise plus interference in a space-time equalizer will however result in an algorithm that only performs spatial interferer suppression.

In Section 2.9, we discuss joint estimation of FIR channel and a spatio-temporal AR model for the noise plus interference. From the discussion in Section 2.9.2, we can understand why an AR model, even a *low order* AR model, for the noise plus interference can be useful in the design of an equalizer or maximum likelihood sequence estimator. The important observation to make is that the the AR noise model denominator does not have to *model* the noise particularly well, it only has to be able to *suppress* the noise plus interference as a part of a noise whitening filter. Such a method was proposed in [7] where it was used in conjunction with a space-time MLSE.

An observation with regard to channel estimation is that not all aspects of a channel estimate are important for equalization. For example, the estimate of the channel in a spatial dimension that is occupied by a strong co-channel interferer may be of little importance. The reason for this is that the dimensions of signal space occupied by the interferer may anyway be canceled out by the equalizer. In that case, the quality of the channel estimate in this dimension is of little importance. This effect can for example be seen in the simulations of Chapter 2 in Sections 2.5 and 2.6. It is discussed further in Section 7.4 of Chapter 7.

Transmit Channel Estimation

In estimation of the transmit channel, the two main methods are the use of *reciprocity* and *feedback*. In the reciprocity method, we use that fact that the transmit and receive channels at the same frequency and at the same time are identical according to the principle of reciprocity [123]. Since the receive channel can be estimated as described earlier, the transmit channel can therefore sometimes be approximated using this principle.

In frequency division duplexed (FDD) systems, the transmit and receive frequencies are separated by perhaps 4 to 5% of the carrier frequency. The channels are then no longer reciprocal, however, if the angular spread of the signal is small, then the *spatial* signature of the channel will still be approximately reciprocal [84]. If the transmit channel cannot be viewed as reciprocal we may attempt to parametrize the taps in the FIR channel in terms of signal paths with directions of arrival and gains. This estimate of the receive channel can then be converted to an estimate of the transmit channel. A method that can be used to achieve such channel estimates is discussed in Section 2.5 of Chapter 2.5. In practice this may be difficult utilize though. One problem is that we need accurately calibrated antenna arrays.

In time division duplexed (TDD) systems, receive and transmit are separated in time but not in frequence. In principle we can then rely on the reciprocity and use the estimates of the receive channel as estimates of the transmit channel. Note however that for time-varying channels the reciprocity will only be valid if the duplexing time is much shorter than the coherence time. The accuracy of the transmit channel estimation thus depends upon the duplexing technique and the channel characteristics.

Another approach for transmit channel estimation uses feedback. The signal received at the receiver is fed back to the transmitter, allowing the transmitter to estimate the channel, see for example [29]. Alternatively, parameters of the transmit channel identified at the receiver can be fed back to the transmitter. Once again, the performance of the feedback techniques depends on the channel characteristics and the nature of the feedback algorithm. Transmit space-time channel estimation offers special challenges and remains an active area of research. See [81] for more details.

1.4.2 TDMA Receive Algorithms

In TDMA, the main tasks to be performed by a space-time receiver are diversity reception, intersymbol interference equalization, and co-channel interference suppression. We can classify TDMA algorithms into space-time processing that is decoupled or joint in the spatial and temporal domains. Figure 1.7 illustrates this.

Single User Decoupled Space-Time Approach

In space-time processing, we can decouple the space and time processing. This will lead to a spatial beamformer front end followed by a temporal processor (equalizer). The pure spatial processor can be used to reduce co-channel interference while maximizing spatial diversity. The output of

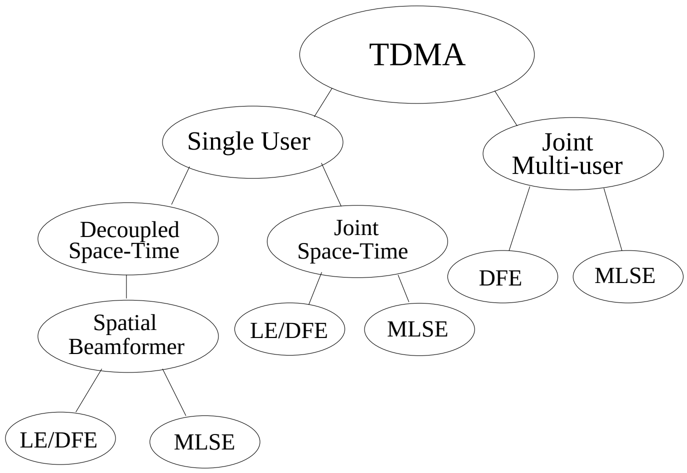


Figure 1.7: TDMA receiver algorithm classification for single-user and multiple user space-time processing.

the spatial processor is fed to a temporal processor for intersymbol interference reduction and recovery of temporal diversity. The spatial processor can range from a fully adaptive beamformer to a simpler switched beam system. The main options for the temporal processor are a linear equalizer (LE), a decision feedback equalizer (DFE), or a maximum likelihood sequence estimator (MLSE). These receivers, and their MISO space-time generalizations are introduced in Chapters 3 and 4. A class of robust temporal equalizers, which include linear equalizers and decision feedback equalizers as special cases, will furthermore be derived and discussed in Section 7.1 of Chapter 7.

The spatial beamformer and the temporal equalizer can either be tuned separately, the beamformer first and then the equalizer, or they can be tuned jointly. In Chapter 5, examples of both these approaches are investigated.

In Section 5.2 the problem of tuning a spatial beamformer independently from a subsequent temporal equalizer is discussed. Since the proper training sequence for the tuning of the spatial beamformer is typically not known, a method of introducing some degrees of freedom in the reference signal is introduced. It is shown that this can improve the performance when there is an uncertainty in the synchronization of the system. However, this method of beamforming will likely improve the performance in other more general situations as well. It will likely be useful in scenarios involving intersymbol interference due to delay spread in the propagation channel.

In Section 5.3 the problem of *jointly* tuning a beamformer and a following temporal equalizer is addressed. By performing a singular value decomposition of a noise whitened channel, beamformers and temporal equalizers can be jointly tuned. The method is applied both to decision feedback equalizers and maximum likelihood sequence estimators. This method can also be applied to the case when the noise plus interference is modeled by a space-time AR model. In this case the "beamformer" actually becomes a space-time MISO filter which performs spatio-temporal whitening of the noise plus interference.

Single User Joint Space-Time Approach

In the presence of coupled angle and delay spreads for the desired signal, i.e. when the spatial and the temporal spreading cannot be decoupled, a joint space-time processing approach has performance advantages. Joint space-time processing is also superior, as compared to decoupled space-time methods, when dealing with delay spread in the co-channel interference. A number of receiver structures have been proposed, broadly divided into space-time linear equalizers, space-time decision feedback equalizers, and space-time maximum likelihood sequence estimators. Several such algorithms are discussed in the following chapters of this thesis.

A problem encountered when tuning the space-time equalizers using a short training sequence is that if we have many antenna elements, then it may be difficult to fully utilize the spatio-temporal spectrum of the noise plus interference. The tuning of the equalizers can then easily become ill-conditioned or singular. One solution to this problem is to concentrate on estimating and utilizing the spatial spectrum of the noise plus interference. This will result in an equalizer that only suppresses interference in the spatial domain. When using many antenna elements this may however be good enough. Another solution can be to estimate a low order AR model for the spatio-temporal spectrum of the noise plus interference. This is advantageous for two reasons. First, an AR model for the noise plus interference is more easily estimated than an moving average (MA) model (which in effect is what is done when estimating the spectrum with sample matrix estimates of different lags) and second, the low order AR model shows up as a low order FIR filter factor in the filters of the optimal equalizers. When the channel is modeled by an FIR filter, and the noise plus interference is modeled by an AR model, then the MMSE optimal DFE will have a structure with only FIR filters. This is a good feature since when the filters of the equalizer contains IIR filters we have to worry the location of their poles[4]. See also [81, 130, 8, 61, 59, 121, 71].

For purely temporal processing with a single antenna receiver the decision feedback equalizer will typically be considerably better than the linear equalizer and the MLSE will typically be better then the decision feedback equalizer. When employing space-time processing and using a relatively large number of receive antennas the difference in performance between the three equalizers will however be less pronounced [18]. The reason for this is that by adding the spatial dimension the channel can more easily be inverted by the linear equalizer. The feedback filter of the space-time DFE and the and the sequence estimation of the MLSE will then not add as much to the performance as in the single antenna, purely temporal, case.

Multi-User Detection

With multiple antennas, joint multi-user detection in TDMA becomes more robust than with one antenna [5]. The spatial dimension helps to separate multi-user signals. The two main choices are a multi-user decision feedback equalizer [23, 25, 107, 110, 108, 106] or a multi-user maximum likelihood sequence estimator, see e.g. [134]. The multi-user decision feedback equalizer has a computational advantage over the MLSE since its complexity grows linearly with the number of users, whereas it increases exponentially in the number of users for the MLSE. An important problem to solve in multi-user detection is the estimation of the channels to the users. If the signals are correlated then there can be an advantage of estimating the channels jointly. However, in this case the number of channel coefficients can easily become too many to handle with a short training sequence. In Section 2.4

[4]For example, a feedback filter of a decision feedback equalizer with poles close to the stability boundary can cause long error propagation events.

[5]If there is multipath propagation in the channel resulting in a delay spread of at least of the order of one symbol interval, then it is, although difficult, possible to do some multi-user detection with only one antenna. If the signalling is binary we can also utilize the fact that we can create two different signals from one antenna by treating the real and the imaginary part separately, see Section 1.5.4

of Chapter 2 temporal parametrization, utilizing the knowledge of the pulse shaping in the transmitter (and possibly also in the receiver), is used to reduce the number of parameters to be estimated. It is shown that this can be valuable in joint multi-user channel estimation.

1.4.3 CDMA Receive Algorithms

We here restrict the discussion to DS-CDMA. In DS-CDMA the main tasks of the receiver are MA interference suppression and detection. A tree diagram of algorithm choices is shown in Figure 1.8.

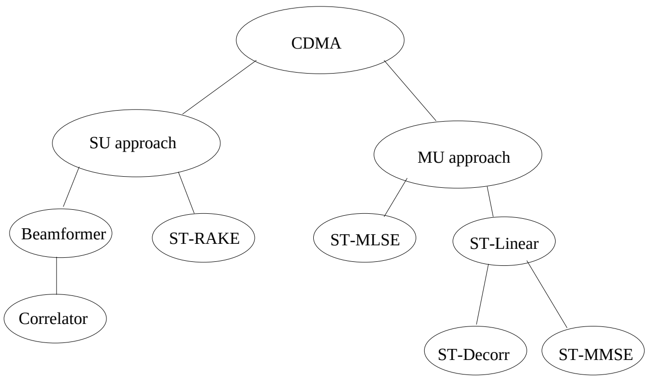


Figure 1.8: CDMA receiver algorithms, for single-user and joint multi-user space-time processing.

There are two main classes of detection schemes for CDMA: the single-user detection approach and the multi-user detection approach. In the single-user approach, only one users signal at a time is recognized and the other signals are treated as noise. In the multi-user detection approach, all users are detected jointly.

Single-user Detection

The space-time processing can here be decoupled into a spatial beamformer followed by a simple decorrelating detector using the spreading code. This is the space-time counterpart of the simple decorrelating receiver. It exploits a single received path (finger) and is useful in environments with no delay spread caused by multipath propagation. See for example [102].

If several paths (fingers) are present, the filtering can be done done jointly in time and space. The natural generalization of the RAKE-detector would be the Space-Time RAKE (ST-RAKE) receiver which is a combination of one beamformer per path followed by a RAKE combiner. This algorithm can also be viewed as a matched filter in both the spatial and the temporal domain. See for example [73].

Multi-user Detection

Although one can conceive spatially and temporally decoupled multi-user detection schemes, we will here only consider joint space-time multi-user MLSE detectors and linear detectors. The space-time multi-user MLSE generalizes from the time-only multi-user MLSE. See [118] and [70]. The MLSE will be optimal if the channels for all users are known. However, as in TDMA, it is computationally complex.

The linear detectors are much less computationally complex. Examples of linear detectors are the space-time decorrelating detector (ST-Decorr) and the space-time MMSE detector (ST-MMSE), see [70] and [72].

1.4.4 Transmit Algorithms

Space-time transmit algorithms use a variety of techniques to maximize diversity, minimize generated co-channel interference and also in some situations pre-equalize the channel for intersymbol interference [81][28]. In general, since intersymbol interference equalization can be implemented at the receiver, transmit space-time processing focuses on diversity gain maximization and co-channel interference reduction. In cases when the transmit channels for the signal and co-channel users are known, the transmit algorithms can implement optimum space-time weighting to maximize diversity gain while minimizing generated co-channel interference. Here, again, both decoupled and joint space-time approaches can be considered, with the latter offering improved performance. In a more likely scenario, the channels of co-channel users are unknown and the signal channel is known only approximately. Transmit algorithms may then use simple beamforming with the beam steered towards the dominant mobiles direction with low side lobes to reduce co-channel interference generation. In the extreme case when no channel knowledge is available, the transmit algorithms reduces to pure diversity maximization schemes. These schemes convert the space diversity of the transmit antennas into other forms of diversity that can be exploited by the receiver. Some examples include phase rolling, delay diversity and space-time coding [114, 113, 112].

1.5 Influence of the Channel on Space-Time Processing

Space-time processing algorithms are profoundly influenced by the channel characteristics[6]. A description of the effects of the channel characteristics and the corresponding mitigation techniques are given in Figure 5.6.

	Effect	Mitigation
Doppler Spread	- Time varying channel - Reduces reciprocity in TDD - Time selective fading	- Channel tracking - Reduce TDD turn around time - Time diversity
Delay Spread	- ISI - Reduces reciprocity in time channel - Frequency selective fading	- Equalization/RAKE - Angle selectivity - Frequency diversity
Angle Spread	- Space selective fading - Reduces reciprocity in space channel	- Space diversity - Reduce frequency spread in FDD

Figure 1.9: Channel characteristics influencing space-time processing.

[6]In space-time processing, the channel is broadly defined to include also the interference channels.

1.5.1 Doppler Spread

Doppler spread, induced by the motion of subscribers or scatterers, has a strong influence on space-time processing algorithms in different dimensions. The Doppler spread is large in macro-cells which serve high mobility subscribers. Also it increases with higher operating frequencies. Doppler spread is also present in low mobility (microcell) or fixed wireless networks due to mobility of scatterers (e.g. traffic). A discussion of Doppler and delay spreads of mobile radio channels can be found in [51].

In a TDMA system, if the time period of a *frame* is small compared to the coherence time of the channel (as in GSM), the channel will be reasonably constant during the frame, and we do not need to track the channel during the frame. On the other hand if the frame duration is comparable to, or longer, than the coherence time of the channel (as in IS-136), the channel changes significantly and we need to track the channel during the frame. Although the channel typically does not vary very much during a frame in GSM it can vary a non-negligible amount if the speed is very high, say on a high speed train, and the carrier frequence is high, say 1800 MHz. In this case some improvements can be achieved by designing an equalizer which will be robust against the anticipated time variations during the frame. How to design such equalizers is discussed in Section 7.3 of Chapter 7.

Fading can sometimes be combatted in the time domain by interleaving and coding. This is however only effective if the coherence time is shorter than the interleaver depth. For slowly time varying channels, other forms of diversity may be necessary to ensure acceptable link quality, e.g. frequency hopping. Also, as mentioned in Section 1.4.1, in time division duplex systems, the reciprocity of the channel is valid only if the channel coherence time is much larger than the duplexing time.

1.5.2 Delay Spread

Delay spread arises from multipath propagation and can be large in macrocell systems with antennas located above the roof tops. It is most pronounced in hilly terrain areas and least pronounced in flat rural terrain installations. Microcells using antennas mounted below the roof tops, tend to have small delay spreads.

Delay spread affects space-time processing algorithms in several ways. In TDMA systems, if the symbol period is much shorter than the delay spread of the channel, we can avoid equalizers (as in PACS and PHP). In contrast, in GSM, the delay spread can be much larger than the symbol period, mandating the use of equalizers. In general, combined space and time processing is more effective for delay spread mitigation than time processing alone.

Likewise, in CDMA, if the delay spread is larger than the chip period, we have inter-chip interference which, however, is usually less insidious than the intersymbol interference in TDMA. Typically, the diversity in paths is exploited by a RAKE receiver.

Delay spread in the channel will increase the number of uncorrelated signals impinging on an antenna array. If there is no delay spread in the channel, then the number of uncorrelated signals will be equal to the number of users, desired and undesired, transmitting to the antenna array. Note however the transmitted signals are assumed to be temporally white. Thus, if the channels from the different users to the antenna array has a delay spread, then the number of uncorrelated signals will be increased. Roughly we can say that each delay spread of a symbol interval adds one uncorrelated signal per user.

1.5.3 Angle Spread

Angle spread arises from multipath arrivals from different directions. It is largest at the subscriber unit, where local scatterers may result in 360 degrees angle spread. At the base station, the angle spread is large in microcells with below roof top antennas. Base stations in macro-cells witness less angle spread, it is the lowest in rural regions and becomes significant in urban and hilly regions.

Angle spread influences a number of space-time processing issues. First, high angle spread increases spatial diversity which should be exploited by space-time processing. Next, the reciprocity of the spatial signature of the channel is reduced if the angle spread is large [84]. The angle spread also reduces the effectiveness of methods parametrizing the signal in directions of arrival since these will be less distinct.

The angle spread and the delay spread will affect the spatio-temporal structure of the channel. The spatio-temporal structure of the channel can be either coupled or can decouple. If we only have angular spread and no delay spread in the propagation channel then the overall channel will have a decoupled spatio-temporal structure. The channel can then be described by a temporal filter, representing the temporal pulse shaping in the transmitter and the receiver followed by a purely spatial SIMO filter representing the spatial spreading of the propagation channel. The spatio-temporal structure of the channel affects the structure and complexity of appropriate space-time equalizer. This is discussed more in detail in Section 5.3 of Chapter 5.

1.5.4 Different Realizations of a Multi-Channel Receiver

A multi-channel receiver can be realized by using multiple antenna elements that are spatially separated. There is however many other ways of realizing a multi-channel receiver.

Polarization diversity is an obvious way to realize multiple channels. Although the two polarizations of the signal may be collected from the same spatial location, they typically have encountered independently fading channels. In most of the methods we use here we can simply view the signals as coming from two independent antenna elements.

Fractionally spaced sampling can be of interest if the symbol sampling frequence is lower than twice the maximum frequence content of the signal. Fractionally spaced sampling can be handled as a multi-channel receive problem. The multiple samples during a symbol interval can be treated as if they were coming from different sensors. The channels will be correlated, but not identical and can thus aid the equalization or the symbol detection. This can be viewed as an analogy to the multiple antenna receiver and many of the techniques applied to spatio-temporal processing can be utilized in together with fractionally spaced sampling.

Partitioning into real and imaginary part: When the modulation format only utilizes one of the dimensions in the complex plane, for example if we have binary phase shift keying modulation, then the real and the imaginary part of the signal can be treated as two different signals. Since the transmitted signal is real (or imaginary) the real and imaginary part of the received signal will effectively be a two-dimensional signal (one real and one imaginary component). Each antenna element thus provides two signals. In Rayleigh fading environments, the real and imaginary parts of the received signal will fade independently.In the simulations presented in this thesis we have in general not exploited these two dimensions of the signal when dealing with modulation formats that only occupy one dimension in the complex plane (the GMSK modulation used in GSM effectively being one such modulation format). It is however easy to incorporate this - simply replace the complex signals with twice as many real-valued signals. For example, an eight sensor receiver can then be realized with only four physical antennas. An example where we *have* performed separate real and imaginary processing can be found in Section 3.4.

Multiple spatially separated antenna elements can be arranged in different ways. We can identify two main configurations. Either the antennas can be closely spaced, say equally spaced, $\lambda/2$ apart, in a linear or circular array. This configuration allows us to identify directions of arrival for the incoming signals and we can transmit in well defined directions. However, if the signal environment does not contain scatterers located close to the receiving antenna, it may be the case that the received signals at all antenna elements fade simultaneously. The close spacing of the antenna elements will then not be optimal from a diversity point of view. Instead the antenna elements can be placed far apart such that they experience independently fading channels. This ensures that we obtain maximum diversity gain. It will, however, then be very difficult or impossible to identify from which directions the signals are arriving. It will likewise be very difficult or impossible to transmit in any well defined direction. From a reception point of view this configuration is most beneficial, unless we want to exploit the path structure of the received signal.

1.6 Some Notes on Special Notation

In this thesis we primarily assume time-invariant channels, unless explicitly stated otherwise. Furthermore, the transmitted symbol sequence, $d(t)$, is in general assumed to be temporally white. The discrete time index is denoted with t.

We will use the delay operator, q^{-1}, and the advance operator, q, defined by

$$q^{-1}x(t) = x(t-1) \qquad qx(t) = x(t+1). \tag{1.1}$$

An FIR filter can thus be expressed with a polynomial $a(q^{-1})$ with complex valued coefficients, as

$$\begin{aligned} y(t) &= a(q^{-1})x(t) = (a_0 + a_1 q^{-1} + \ldots + a_{na} q^{-na})x(t) \\ &= a_0 x(t) + a_1 x(t-1) + \ldots + a_{na} x(t-na). \end{aligned} \tag{1.2}$$

The above filter contains only powers of q^{-1} and is thus causal. In some cases a filter will have terms involving both powers of q as well as powers of q^{-1}. Such a filter will be referred to as double-sided or non-causal. If the filter contains powers of q only, it will be referred to as anti-causal.

Multiple input - single output (MISO) and single input - multiple output (SIMO) filters can be represented with polynomial row and column vectors, respectively, which are denoted by boldface lowercase letters. MIMO filters can be represented by polynomial matrices, denoted by uppercase boldface letters.

A polynomial matrix $\boldsymbol{A}(q^{-1})$ is said to be *stably invertible* if $\det\big(\boldsymbol{A}(z^{-1})\big)$ has all its poles strictly inside the unit circle. Then the transfer function $\boldsymbol{A}^{-1}(q^{-1})$ will be stable.

A polynomial matrix $\boldsymbol{A}(q^{-1})$ is *causally invertible* if its leading matrix coefficient, $\boldsymbol{A}_0$, is non-singular. Then the transfer function, $\boldsymbol{A}(q^{-1})$, will be causal.

The complex conjugate transpose of a filter (SISO,MISO,SIMO or MIMO)

$$\boldsymbol{A}(q^{-1}) = \boldsymbol{A}_0 + \boldsymbol{A}_1 q^{-1} + \ldots + \boldsymbol{A}_{na} q^{-na} \tag{1.3}$$

is defined as

$$(\boldsymbol{A}(q^{-1}))^H \triangleq \boldsymbol{A}^H(q) = \boldsymbol{A}_0^H + \boldsymbol{A}_1^H q + \ldots + \boldsymbol{A}_{na}^H q^{na}. \tag{1.4}$$

This is also generalized to non-causal polynomial matrices.

The notation $\boldsymbol{R_{nn}}(k)$ stands for the lagged covariance

$$\boldsymbol{R_{nn}}(k) = E[\boldsymbol{n}(t)\boldsymbol{n}^H(t-k)] \tag{1.5}$$

and the notation $\boldsymbol{R_{nn}}$ normally means

$$\boldsymbol{R_{nn}} = \boldsymbol{R_{nn}}(0) = E[\boldsymbol{n}(t)\boldsymbol{n}^H(t)]. \tag{1.6}$$

Given an Hermitian and positive semi-definite matrix, $\boldsymbol{R}$, we define the square root, $\boldsymbol{R}^{1/2}$, as the positive semi-definite matrix such that

$$\boldsymbol{R} = \boldsymbol{R}^{1/2}(\boldsymbol{R}^{1/2})^H. \tag{1.7}$$

For simplicity we define the notation

$$\boldsymbol{R}^{H/2} \triangleq (\boldsymbol{R}^{1/2})^H \tag{1.8}$$

such that we can write

$$\boldsymbol{R} = \boldsymbol{R}^{1/2}\boldsymbol{R}^{H/2} \tag{1.9}$$

We will often discuss relative errors with respect to matrices and vectors. As a measure of the relative error in the approximation $\widehat{\boldsymbol{A}}$ of the matrix $\boldsymbol{A}$ we will typically use

$$\text{Relative error in } \widehat{\boldsymbol{A}} \triangleq \frac{\| \widehat{\boldsymbol{A}} - \boldsymbol{A} \|_2}{\| \boldsymbol{A} \|_2} \tag{1.10}$$

where the notation $\| \bullet \|_2$ represents the Frobenius norm, i.e. the square root of the sum of the square of the components.

Chapter 2

Channel Estimation

2.1 Introduction

We will here consider the estimation of wireless communication channels from the transmitted symbols to received sampled signals. These channels include the modulation process, the propagation channel as well as transmitter and receiver filters. We also discuss the modeling of the noise and interference that affect the received signal.

A general baseband model of a wireless communication channel is depicted in Figure 2.1. The symbols, $d(t)$, representing source and channel coded information (in discrete time), are first modulated onto a signal (pulse shaping) and are then transmitted over the propagation channel to the receiver.

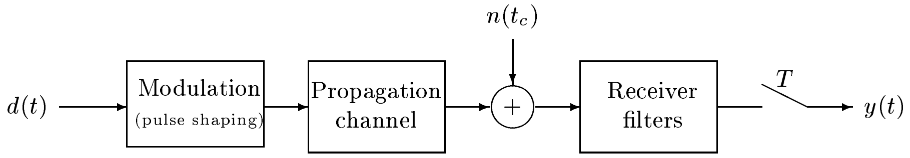


Figure 2.1: Model of a scalar wireless communication channel.

Additive thermal noise and interference is represented by the term $n(t_c)$[1], where t_c denotes continuous time. The received combination of desired signal and noise plus interference is filtered in the receiver prior to sampling. This will in general cause the noise component in the sampled signal to be temporally colored.

The propagation channel includes multipath propagation which causes spatial and temporal spreading of the signal, i.e. the signal arrives from different directions and with different time delays. If the source, the receiver or the environment is moving or changing then the channel will also be time-varying. We will here however mostly be considering time-invariant channel models. This can be justified when considering a TDMA scheme if the frame lengths are short.

The baseband channel consisting of the modulation, a time-invariant propagation channel and the receiver filters, can typically be modeled by a time-invariant discrete-time FIR filter as

$$y(t) = b(q^{-1})d(t) + n(t) \tag{2.1}$$

where t is an integer representing the discrete time, $y(t)$ is the sampled received signal, $b(q^{-1}) = b_0 + b_1 q^{-1} + \ldots + b_{nb} q^{-nb}$ is the FIR filter representing the channel for the desired signal, $d(t)$ is the transmitted discrete symbol sequence and $n(t)$ is the noise plus interference at the sampling instants. The received signal $y(t)$ is typically a complex valued signal with an in-phase (real) and a quadrature (imaginary) component. Likewise the coefficients in the FIR filter, $b(q^{-1})$, are typically complex valued. Depending on what type of modulation is used, the transmitted symbols, $d(t)$, will either be real or complex valued.

In order to model the communication channel with an FIR filter as in (2.1) we need to be able to describe the modulation process with a linear FIR model. The modulation applied in the GSM standard is an example of a modulation that is non-linear. However, as demonstrated in Appendix 2.A.1, it can be approximated with a linear model after some processing.

If we have multiple antennas at the receiver we can collect the received signals in a column vector, $\boldsymbol{y}(t) = [y_1(t)\ y_2(t)\ \ldots\ y_M(t)]^T$, where M is the

[1]The thermal noise is added in several stages of the receiver but we have here chosen to represent all noise effects by a term added prior to any filtering in the receiver.

number of antennas. The received signal can now be modeled as

$$\boldsymbol{y}(t) = \boldsymbol{b}(q^{-1})d(t) + \boldsymbol{n}(t) \tag{2.2}$$

where $\boldsymbol{b}(q^{-1})$ is a polynomial column vector containing the polynomial channels $b_i(q^{-1})$, of degree nb, to the individual antennas

$$\boldsymbol{b}(q^{-1}) = [b_1(q^{-1})\ b_2(q^{-1})\ \ldots\ b_M(q^{-1})]^T. \tag{2.3}$$

The noise plus interference to the different antennas, $n_i(t)$, is represented by the vector $\boldsymbol{n}(t)$

$$\boldsymbol{n}(t) = [n_1(t)\ n_2(t)\ \ldots\ n_M(t)]^T. \tag{2.4}$$

The noise plus interference will in some of the considered cases be modeled as as a sum of interfering co-channel users plus thermal noise. For the case with multiple antennas this can be expressed as

$$\boldsymbol{n}(t) = \sum_{k=1}^{K} \boldsymbol{b}_k(q^{-1})d_k(t) + \boldsymbol{v}(t) \tag{2.5}$$

where K is the number of interfering co-channel users, $\boldsymbol{b}_k(q^{-1})$ represent their respective vector FIR channels, $d_k(t)$ are the corresponding symbol sequences and $\boldsymbol{v}(t)$ is the thermal noise.

Estimation of the wireless channel is of interest for many reasons. A main reason is that knowledge of the channel can be a step in the design and tuning of the detectors. The detector can for example be an equalizer or a maximum likelihood sequence estimator. Another reason for estimating the *uplink* channel from the mobile to the base-station is that this estimate can be used for estimating the *downlink* channel to optimize downlink transmission.

If the frequencies for the uplink and the downlink transmissions are the same, as in a time division duplexing system, then the downlink channel is basically the same as the uplink channel. If the frequencies in the up- and downlink are different, as in a frequency division duplexing system, it may be necessary to parametrize the channel estimate in terms of directions of arrival and their associated path gains. An example of such an approach is presented in Section 2.5.

When using multiple receive antennas, the quality of the channel estimates will be especially important. With many antenna elements the potential improvements in the BER of a model based equalizer by using better channel estimates is larger than when only one antenna is used. The reason for this is that with multiple antennas it is possible to form very deep nulls suppressing interferers as well as high gains amplifying the desired signal. However, in order to achieve the maximum improvement, the coefficients of the space-time equalizer have to be accurately tuned.

The most straightforward approach for estimating a wireless channel, as the one in (2.1), is to directly estimate the coefficients of the FIR model of the channel. The coefficients of this FIR filter can be estimated using a least squares approach as shown in Section 2.2.

The residuals from this channel estimation can be used to estimate the space-time covariance matrix for the noise plus interference, $\boldsymbol{n}(t)$, in (2.2) by forming sample-matrix estimates of the covariance for different time lags, as described in Section 2.8.

Since channel models in general, and spatio-temporal noise models in particular, will have low accuracy when their parameters are estimated using few data, ways to improve upon the basic least squares estimate of the FIR parameters will be explored in this chapter. The estimates of the channel and noise properties can be improved in at least five different ways:

1. Structural constraints and *a priori* information can be utilized to reduce the number of parameters that need to be estimated.
2. The data can be pre-filtered or projected onto a subspace to improve the signal to noise ratio.
3. The number of training data can be increased by utilizing detected symbols during time intervals where the symbol sequence is initially unknown.
4. Simplified parametric models can be estimated for the the noise and interference.
5. Joint estimation of the channel and a spatio-temporal noise model.

We will in this chapter utilize all five principles, alone or in combination. The following main methods will be explored and suggested.

Utilizing known factors in the channel impulse response: When we model the total channel from the transmitted symbols to the received samples as an FIR channel, we have not used of the fact that the pulse shaping performed in the modulation is known. The only part of the channel we really need to estimate is the propagation channel and the part of the transmitter and receiver filters that are unknown. Such an approach has been presented in [75] and for the special case of a GSM channel in [89]. We will in Sections 2.3 and 2.4 consider an approach that is slightly different. In its simplest form, it concatenates an unknown discrete time FIR filter modeling the propagation channel with a known discrete time FIR filter modeling the pulse shaping in the modulation and known parts of the receiver filter. However, discrete-time FIR models of the pulse shaping, etc., cannot be determined completely from their known continuous-time shapes. Instead, the total channel can be approximated as a sum of several unknown discrete-time FIR filters representing the propagation channel, each multiplied by known FIR filters. The known FIR filters are sampled versions of the pulse shaping function, sampled with different specific sampling offsets. By increasing the number of branches, the model can be arbitrarily refined. Fractionally spaced sampling can also be introduced. We now only need to estimate the coefficients of the unknown FIR filter representing the propagation channel.

In Section 2.4 the method of utilizing pulse shaping information presented in Section 2.3 is generalized and applied to multi-user channel estimation. When we want to estimate channels to multiple user jointly it is of great importance to keep the number of parameters small. The temporal parametrization helps in doing this and we can see in the simulations that it pays off. By utilizing the pulse shaping information we can estimate the channels to more users jointly.

Parametrizing the spatial structure of the channel: If the signals arrive to an antenna array from a few directions only, then we may model the channel in terms of a few signal paths parametrized in angles of arrival and gains. Such an approach using a method called CDEML [15] is investigated in Section 2.5. In the scenario studied in Section 2.5.3 we can see that the performance measured in terms of equalizer BER is improved. The largest improvement is however achieved when the correct number of signal paths is assumed. This is of course a potential weakness with the method.

In Section 2.6, a different use of the spatial channel structure is presented. This is a non-parametric method, based on the fact that a model of a typical space-time channel will not utilize all its available degrees of freedom. The channel, as well as the desired signal, lies in a subspace defined by the dominant eigenvectors of the spatial signal covariance matrix. By projecting either the taps of the channel [74] or the received signal samples onto this subspace, the performance of the channel estimation and the subsequent equalization can be improved. Projecting the received signal onto this subspace turns out to be particularly effective. By only retaining the significant components in the subspace, a reduction of the dimension of the problem is furthermore achieved. This results in a better estimated spectrum for the noise plus interference as well as in reduced complexity for all subsequent processing.

It should be noted that a projection of the taps of the channel estimate or of the received signal samples onto the signal subspace will not suppress co-channel interferers, since the interference will be contained in the signal space. The attained improvements are partly due to a removal of noise components, and thus constitutes a form of noise-reducing pre-processing. The performance improvement is also due to the reduced dimension, which will facilitate the estimation of the covariance matrix for the noise plus interference.

Utilizing decided data in the channel and noise estimation: Most of the channel estimation schemes in this chapter are based on short sequences of training data. After an initial equalization and estimation of the transmitted data symbols in a TDMA frame, the channel can however be re-estimated utilizing all, or a part of, the symbols in the frame. Unless there are too many errors in the initial estimates of the data symbols, this new channel estimate will be improved. The procedure can be repeated by, re-estimating the symbols and re-estimating the channel. In Section 2.7 this method, here called *bootstrapping*, is presented. Channel estimation via bootstrapping is a good way to handle co-channel interferers. The artificially created long training sequence reduces the channel estimation errors introduced by the strong co-channel interferers. The estimate of the noise plus interference covariance matrix is also improved, which is important when suppressing co-channel interferers.

Estimation of the spatial covariance structure only: In many cases, it may not be feasible to utilize the full space-time spectra, or the covariance matrices involving many time lags, since their estimates will have poor accuracy when based on few data. One solution to this problem is to only utilize the spatial spectrum of the noise plus interference, i.e. the noise covariance matrix for lag zero. This will result in space-time equalizers that suppress the noise plus interference spatially, see Chapters 3 and 4.

Estimation of a space-time AR model for the noise plus interference: A good way of catching some of the spatio-temporal features the noise and interference spectrum is to model it as a space-time AR process and estimate this noise model jointly with the channel for the desired signal. These estimates can then be used in a space-time equalizer in order to realize space-time supression of the interferers. This was proposed in [7] in conjunction to an MLSE. This method can be very useful when the number of strong interfering signals are more than can be handeled with spatial-only interference suppression. From the discussion in Section 2.9.2, we can understand why an AR model, even a *low order* AR model, for the noise plus interference can be useful in an equalizer or maximum likelihood sequence estimator. The important observation to make is that the the AR noise model denominator does not have to *model* the noise particularly well, it only has to be able to *suppress* the noise plus interference as a part of a noise whitening filter. Furthermore, the AR-model for the noise plus interference is easy to utilize in the design of either a space-time DFE as described in Section 3.2.2 of Chapter 3 or with a space-time MLSE as described in Chapter 4. An example AR modeling of the noise plus interference combined with a space-time DFE can be seen in one of the simulations in Section 3.4. With a large number of antennas antennas or a high order of the AR noise model, the number of parameters can however become large compared to the number of available equations. This can make them potentially difficult to estimate accurately, especially if the SNR is not high enough. It should also be noted that this method may be combined with the Bootstrap channel estimation in Section 2.7. The extra artificial training symbols achieved with the Bootstrap method can potentially help the estimation AR models when using a large number of antennas.

2.2 Directly Parametrized FIR Channel Estimation

If we assume that a known sequence of training symbols, $\{d(t)\}_{t=1}^{N}$, is available, then a straightforward way to estimate the channel between the transmitted symbols and the received samples at one antenna is to model it as a scalar FIR channel as in (2.1) and estimate the coefficients with a least squares method. The received signal, $y(t)$, is thus modeled as

$$y(t) = b(q^{-1})d(t) + n(t) \tag{2.6}$$

where $y(t)$ is the received signal, $d(t)$ is the transmitted symbols, $n(t)$ is additive thermal noise and interference and

$$b(q^{-1}) = b_0 + b_1 q^{-1} + \ldots + b_{nb} q^{-nb} \tag{2.7}$$

is a FIR filter model of the total scalar channel.

Alternatively, in vector notation we can write

$$y(t) = \boldsymbol{b}\boldsymbol{d}(t) + n(t) \tag{2.8}$$

where $\boldsymbol{d}(t)$ is a column vector containing delayed transmitted symbols

$$\boldsymbol{d}(t) = [d(t)\ d(t-1)\ \ldots\ d(t-nb)]^T \tag{2.9}$$

and $\boldsymbol{b}$ is a row vector containing the FIR filter taps

$$\boldsymbol{b} = [b_0\ b_1\ \ldots\ b_{nb}]. \tag{2.10}$$

The error between the true received signal and the model of the received signal, measuring over the training sequence data $\{d(t), y(t)\}$, t=1,2,...,N) is given by

$$J = \sum_{t=nb+1}^{N} |y(t) - \boldsymbol{b}\boldsymbol{d}(t)|^2. \tag{2.11}$$

The parameter row vector, $\boldsymbol{b}$, that minimizes this norm is given by the standard least squares solution [87]

$$\hat{\boldsymbol{b}}_{\mathrm{LS}} = \hat{\boldsymbol{R}}_{\boldsymbol{d}y}^{H} \hat{\boldsymbol{R}}_{\boldsymbol{dd}}^{-1} \tag{2.12}$$

where

$$\hat{\boldsymbol{R}}_{\boldsymbol{d}y} = \frac{1}{N - nb} \sum_{t=nb+1}^{N} \boldsymbol{d}(t) y^H(t) \tag{2.13}$$

and

$$\hat{\boldsymbol{R}}_{\boldsymbol{dd}} = \frac{1}{N - nb} \sum_{t=nb+1}^{N} \boldsymbol{d}(t) \boldsymbol{d}^H(t). \tag{2.14}$$

If we have multiple antennas at the receiver we can collect the received signals in a column vector, $\boldsymbol{y}(t) = [y_1(t)\ y_2(t)\ \ldots\ y_M(t)]^T$, where M is the number of antennas. The received signal can now, as in (2.2), be modeled as

$$\boldsymbol{y}(t) = \boldsymbol{b}(q^{-1}) d(t) + \boldsymbol{n}(t) \tag{2.15}$$

where $\boldsymbol{b}(q^{-1})$ is a polynomial column vector containing the polynomial channels $b_i(q^{-1})$, of degree nb, to the individual antennas

$$\boldsymbol{b}(q^{-1}) = [b_1(q^{-1})\ b_2(q^{-1})\ \ldots\ b_M(q^{-1})]^T. \tag{2.16}$$

The noise plus interference to the different antennas, $n_i(t)$, is represented by the vector $\boldsymbol{n}(t)$

$$\boldsymbol{n}(t) = [n_1(t)\ n_2(t)\ \ldots\ n_M(t)]^T. \tag{2.17}$$

Switching from polynomial notation to matrices and vectors the received signal channel can alternatively be modeled as

$$\boldsymbol{y}(t) = \boldsymbol{B}\boldsymbol{d}(t) + \boldsymbol{n}(t) \tag{2.18}$$

where $\boldsymbol{B}$ is the $M \times (nb+1)$ *channel matrix*

$$\boldsymbol{B} = \begin{bmatrix} \boldsymbol{b}_1^T & \boldsymbol{b}_2^T & \ldots & \boldsymbol{b}_M^T \end{bmatrix}^T \tag{2.19}$$

with the coefficients of the polynomials $b_i(q^{-1})$ in the vectors

$$\boldsymbol{b}_i = [b_{i0}\ \ b_{i1}\ \ \ldots\ \ b_{inb}] \tag{2.20}$$

and $\boldsymbol{d}(t)$ is the vector with delayed symbols in (2.9).

Each row, $\boldsymbol{b}_i$, in the channel matrix thus contains the channel coefficients for the channel to a specific antenna element. The FIR least squares estimate for each row in (2.19) is given by (2.12). By combining these estimates we can se that the FIR least square estimate of the channel matrix $\boldsymbol{B}$ can be written as

$$\hat{\boldsymbol{B}}_{\mathrm{LS}} = \hat{\boldsymbol{R}}_{\boldsymbol{dy}}^{H} \hat{\boldsymbol{R}}_{\boldsymbol{dd}}^{-1} \tag{2.21}$$

where

$$\hat{\boldsymbol{R}}_{\boldsymbol{dy}} = \frac{1}{N - nb} \sum_{t=nb+1}^{N} \boldsymbol{d}(t) \boldsymbol{y}^{H}(t) \tag{2.22}$$

and where $\hat{\boldsymbol{R}}_{\boldsymbol{dd}}$ is given by (2.14).

2.3 Temporal Parametrization

By modeling the total channel between the transmitted symbols and the received samples as a directly parametrized FIR channel one does not utilize the fact that this channel includes the known effects of pulse shaping due to modulation at the transmitter and filtering in the receiver. By utilizing this knowledge, the channel estimate can be improved since the number of parameters to be estimated is reduced.

A commonly used channel estimation method for GSM channels that utilizes pulse shaping information can be found in [89]. Here the received training part of the signal is correlated with the modulated training sequence. The training sequence is constructed so that the correlation approximately will give the unknown part of the channel, i.e. excluding the pulse shaping.

In [75] a method for channel estimation is presented that discretizes the convolution between the continuous-time pulse shaping function and the "unknown" continuous-time channel impulse response. This results in a parametrization of the total channel in terms of parameters for the unknown part of the channel. This parametrization is then used in order to estimate the total channel.

The channel estimation method presented here also derives a parametrization of the total channel in terms of parameters for the unknown part of the channel. The approach to the modeling is however different from that in [75]. The method considered here is based on an approximation which uses a set of pulse shaping functions sampled at different time instants. This can approximately be viewed as an interpolation[2] between sampled versions of the pulse shaping function with different offsets in the sampling instants, similar to the method used in [56] to introduce some degrees of freedom in the reference signal for a sample matrix inversion beamforming algorithm.

The method in [75] is formulated as a method for multiple antennas. A closer study however reveals that the channel estimation to the different antennas decouples and it is thus a purely temporal method. The method considered here is only presented as a purely temporal method. In the simulation example presented in Section 2.3.3, multiple antennas are used, but the channel estimation is applied to each of the antennas independently.

Since the method presented here and the method in [75] are based on similar principles it is believed that they are equivalent in their performance. The method presented here can however adds useful insight into the modeling of the channel.

Other approaches where pulse shaping information is used for channel estimation can also be found, for example, in [43], [49] and [50]. A blind method using pulse shaping information is presented in [20].

2.3.1 Channel Modeling

In continuous-time, a linear communication channel with linear modulation can be modeled by a linear pulse shaping filter, $p(t_c)$, concatenated with a linear filter, $h(t_c)$, representing the propagation channel. In the pulse shaping filter we can include all known linear filtering which is performed both at the receiver and at the transmitter. Any unknown filtering in the transmitter and the receiver can be included in the model of the unknown propagation channel. This continuous-time model is depicted in

[2] Since the approximation can be scaled in absolute magnitude, and more than two sampled pulse shaping functions can be involved, it will however not be interpolation in the true sense but rather a linear combination of the sampled pulse shaping functions. The word "interpolation" will however be used here in this wider sense. This modeling is illustrated below for two different pulse shaping functions used in GSM [24] and IS-136.

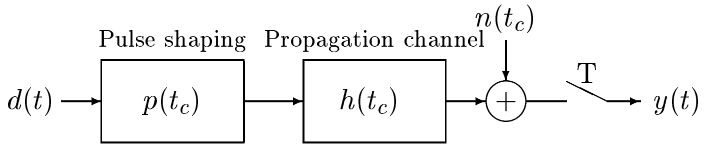


Figure 2.2: Continuous-time channel model where $p(t_c)$ represents all known pulse shaping and filtering at the transmitter and the receiver.

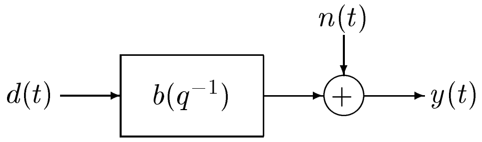


Figure 2.3: Discrete-time FIR channel model describing the symbol-spaced received signal $y(t)$.

Figure 2.2. The resulting discrete-time channel can be modeled with an FIR filter, $b(q^{-1}) = b_0 + b_1 q^{-1} + \ldots + b_{nb} q^{-nb}$, as in Figure 2.3. By using a known training sequence the taps in $b(q^{-1})$ can be estimated with a least squares method. However, by estimating $b(q^{-1})$ directly we do not utilize the a priori knowledge of the pulse shaping filter $p(t_c)$.

A first step towards incorporating the a priori knowledge of the pulse shaping filter into the discrete-time channel model is to sample the impulse response, $p(t_c)$, T-spaced (symbol spaced) and form an FIR filter $p(q^{-1})$ from these samples. The discrete-time model then becomes $p(q^{-1})$ followed by a T-spaced discretization, $h(q^{-1})$, of the propagation channel $h(t_c)$. This model is depicted in Figure 2.4.

The channel estimation is now restricted to estimating the coefficients of the

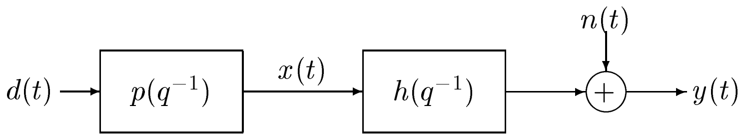


Figure 2.4: Discrete-time channel model with a pulse shaping filter.

.

FIR filter $h(q^{-1})$, using the "modulated" signal $x(t)$ as the input signal. A potential problem with this method is that the model may be too coarse. When designing $p(q^{-1})$ we have to choose where to sample the pulse shaping function $p(t)$. For a pulse resulting from the modulation of a single symbol passing through the system and being sampled at these instants, this will be a perfect model. Such a pulse will be represented by a single tap in the channel $h(q^{-1})$.

However, for a pulse passing through the channel and being sampled at time instants in between the chosen sampling points of $p(t)$, the model will not be perfect. The approximation of such a sampled pulse will essentially be a combination of two shifted versions of the sampled pulse shaping function, $p(q^{-1})$. Apart from a possible scaling factor this can be viewed as an interpolation between the two shifted pulse shaping filters. This will result in a representation with essentially two adjacent taps in the channel $h(q^{-1})$.

If improved accuracy is desired in the model, more than one sampled version of the pulse shaping function $p(t)$ (similarly to [56]) can be used as in Figure 2.5. The modeling of the channel is here divided into two branches with two different sampled versions of the pulse shaping filter, $p_{0.0}(q^{-1})$ and $p_{0.5}(q^{-1})$, with their sampling instants offset by half a symbol interval. The subscript refers to the offset of the filters center tap from the center or peak of the pulse shaping function $p(t)$. See the examples depicted in Figures 2.9 and 2.10, respectively.

Each discrete-time pulse shaping filter is followed by a discrete-time channel filter, $h_1(q^{-1})$ and $h_2(q^{-1})$ respectively. Each pulse passing through the system will now be represented as a single tap in $h_1(q^{-1})$ or $h_2(q^{-1})$ or a combination of two or more adjacent taps. Again this can be viewed as an interpolation among adjacent (possibly more than one) sampled and shifted pulse shaping functions. Note that the interpolation is now performed among T/2-spaced sampled and shifted pulse shaping functions $p_{0.0}(q^{-1})$ and $p_{0.5}(q^{-1})$. This interpolation is thus improved compared to the T-spaced interpolation in Figure 2.4. If an even finer approximation is desired, a larger number of pulse shaping filters, with less spacing between the sampling instants, can be used.

Fractionally spaced sampling, with an oversampling factor of 2, can be represented in the continuous-time model by introducing a positive time shift of -T/2 before an extra sampler, as showed in Figure 2.6. By using more

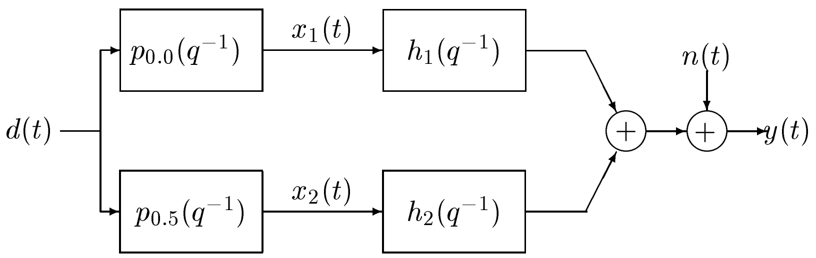


Figure 2.5: An improved discrete-time channel model utilizing multiple pulse shaping filters.

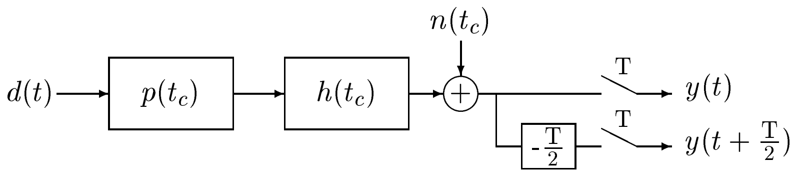


Figure 2.6: Continuous-time channel model with fractionally spaced sampling, where the block $-\frac{T}{2}$ represents a positive time shift of half a symbol.

such branches finer fractionally spaced sampling can be introduced in the model.

The discrete-time model obtained by introducing fractionally spaced sampling in the scheme of Figure 2.5 is displayed in Figure 2.7. The sampled pulse shaping functions $p_{-0.5}(q^{-1})$ and $p_{0.0}(q^{-1})$ for the "T/2-branch" will have their sampling instants offset by -T/2 from $p_{0.0}(q^{-1})$ and $p_{0.5}$ in the "T-branch". It is important to note though that the *same channel filters* $h_1(q^{-1})$ and $h_2(q^{-1})$ can be used in the two branches. The reason for this is that what is a good interpolation in both the "T-branch" between $p_{0.0}(q^{-1})$ and $p_{0.5}(q^{-1})$ will also be a good interpolation between $p_{-0.5}(q^{-1})$ and $p_{0.0}(q^{-1})$ in the "T/2-branch". An important consequence of this is that the number of parameters to be estimated in the fractionally spaced channel model does not increase while the number of equations does due to the extra data points. The estimates of the channel filters $h_1(q^{-1})$ and $h_2(q^{-1})$ can thus be improved with fractionally spaced sampling.

An alternative way of presenting the model in Figure 2.7 is to exchange

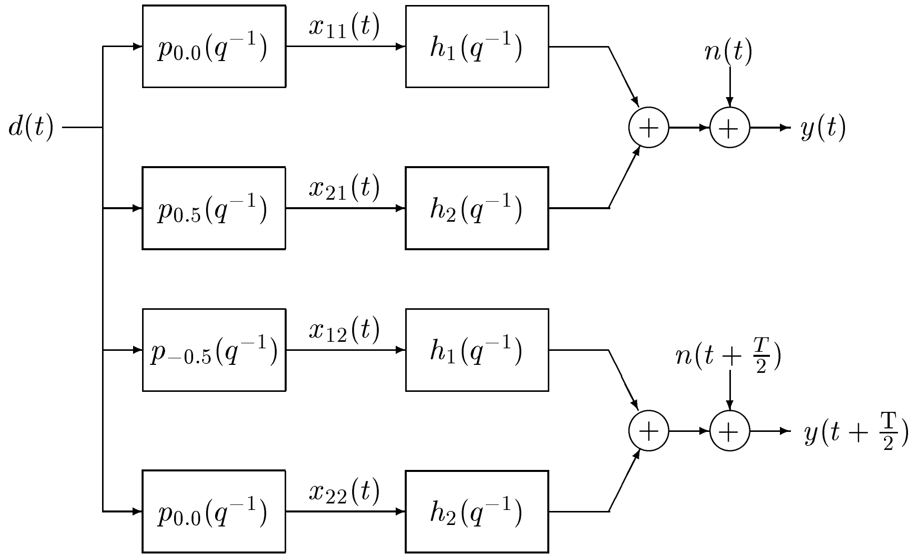


Figure 2.7: Discrete-time channel model with fractionally spaced sampling and multiple pulse shaping filters per sampling branch.

places between the pulse shaping and channel filters as seen in Figure 2.8. In this description it is clear that the same channel filters can be used in the "T-branch" and the "T/2-branch". One can also interpret the taps in the channel filters, $h_1(q^{-1})$ and $h_2(q^{-1})$, as every second tap (even and odd respectively) in a T/2-spaced channel model.

2.3.2 Channel Estimation

For clarity of the presentation the equations below are presented for the case with T/2-spaced fractional sampling and two pulse shaping filters per branch as in Figure 2.7. The equations can however easily be extended to any amount of oversampling and any number of pulse shaping filters per branch.

Let us arrange the fractional samples of the received signal in a row vector

$$\boldsymbol{y}(t) = [y(t) \;\; y(t + \frac{\mathrm{T}}{2})]. \tag{2.23}$$

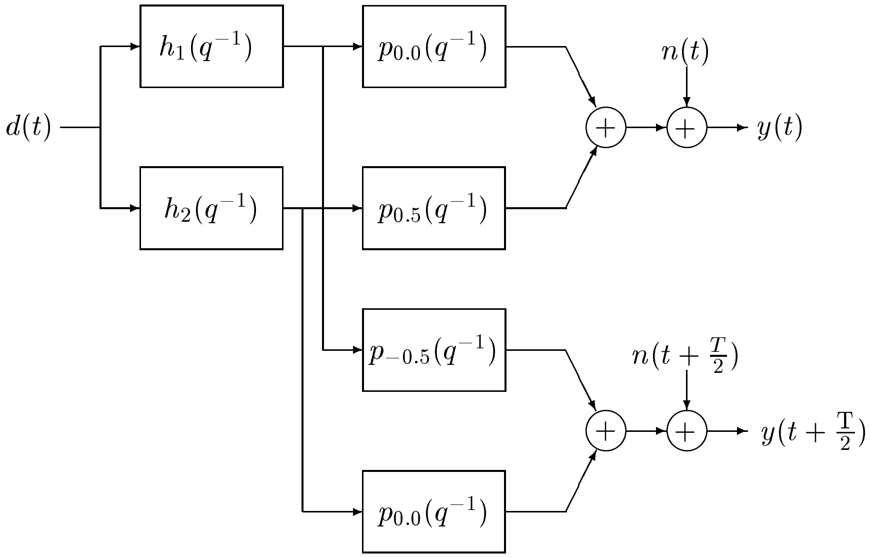


Figure 2.8: Re-ordered structural channel model with fractionally spaced sampling.

This signal vector can now be expressed as

$$\boldsymbol{y}(t) = \boldsymbol{b}(q^{-1})d(t) + \boldsymbol{n}(t) = \boldsymbol{h}(q^{-1})\boldsymbol{P}(q^{-1})d(t) + \boldsymbol{n}(t) \tag{2.24}$$

where $\boldsymbol{b}(q^{-1})$ is a polynomial row vector containing the channel polynomials corresponding to the fractionally spaced samples. Furthermore, the pulse shaping matrix, $\boldsymbol{P}(q^{-1})$, is given by

$$\boldsymbol{P}(q^{-1}) = \begin{bmatrix} p_{0.0}(q^{-1}) & p_{-0.5}(q^{-1}) \\ p_{0.5}(q^{-1}) & p_{0.0}(q^{-1}) \end{bmatrix} \tag{2.25}$$

and the channel vector $\boldsymbol{h}(q^{-1})$ is given by

$$\boldsymbol{h}(q^{-1}) = [h_1(q^{-1}) \;\; h_2(q^{-1})]. \tag{2.26}$$

The vector $\boldsymbol{n}(t) = [n(t) \;\; n(t+\frac{\mathrm{T}}{2})]$ represents the fractionally spaced sampled additive noise.

By using the "modulated" signal, *cf* Figure 2.7

$$\boldsymbol{X}(t) = \boldsymbol{P}(q^{-1})d(t) \tag{2.27}$$

the received signal can now be written as

$$\boldsymbol{y}(t) = \boldsymbol{h}(q^{-1})\boldsymbol{X}(t) + \boldsymbol{n}(t). \tag{2.28}$$

We see that we have a multiple-input multiple-output identification problem for estimating the channel $\boldsymbol{h}(q^{-1})$. This can, as shown below, be solved as a least squares problem.

In order to form a system of equations we vectorize equation (2.28) giving

$$\boldsymbol{y}(t) = \boldsymbol{h}\boldsymbol{\mathcal{X}}(t) + \boldsymbol{n}(t) \tag{2.29}$$

where

$$\boldsymbol{h} = [h_{10} \;\; h_{11} \;\; \ldots \;\; h_{1nh} \;\; h_{20} \;\; h_{21} \;\; \ldots \;\; h_{2nh}] \tag{2.30}$$

and

$$\boldsymbol{\mathcal{X}}(t) = \begin{bmatrix} \bar{\boldsymbol{x}}_{11}(t) & \bar{\boldsymbol{x}}_{12}(t) \\ \bar{\boldsymbol{x}}_{21}(t) & \bar{\boldsymbol{x}}_{22}(t) \end{bmatrix} \tag{2.31}$$

with

$$\bar{\boldsymbol{x}}_{ij}(t) = [x_{ij}(t) \;\; x_{ij}(t-1) \;\; \ldots \;\; x_{ij}(t-nh)]^T. \tag{2.32}$$

The unknown channel parameter vector $\boldsymbol{h}$ can now be estimated with the least squares method as

$$\hat{\boldsymbol{h}} = \hat{\boldsymbol{R}}_{\boldsymbol{y}\boldsymbol{\mathcal{X}}}\hat{\boldsymbol{R}}_{\boldsymbol{\mathcal{X}}\boldsymbol{\mathcal{X}}}^{-1} \tag{2.33}$$

where

$$\hat{\boldsymbol{R}}_{\boldsymbol{y}\boldsymbol{\mathcal{X}}} = \frac{1}{t_{max} - t_{min} + 1} \sum_{t=t_{min}}^{t=t_{max}} \boldsymbol{y}(t)\boldsymbol{\mathcal{X}}^H(t) \tag{2.34}$$

and

$$\hat{\boldsymbol{R}}_{\boldsymbol{\mathcal{X}}\boldsymbol{\mathcal{X}}} = \frac{1}{t_{max} - t_{min} + 1} \sum_{t=t_{min}}^{t=t_{max}} \boldsymbol{\mathcal{X}}(t)\boldsymbol{\mathcal{X}}^H(t). \tag{2.35}$$

An estimate of the total channel $\boldsymbol{b}(q^{-1})$ is then obtained as

$$\hat{\boldsymbol{b}}(q^{-1}) = \hat{\boldsymbol{h}}(q^{-1})\boldsymbol{P}(q^{-1}). \tag{2.36}$$

The time instants t_{min} and t_{max} represent the minimum and maximum time samples of the used training data. Due to the pulse shaping, the taps at the beginning and the end of the channel $\boldsymbol{b}(q^{-1})$ will be small. Extending the training sequence with a couple of zero elements will improve the channel estimate if the SNR is low since more equations are then being used. At high SNR it could however worsen the channel estimate since an incorrect assumption is made when extending the training sequence with zeros. In the simulations presented below, the training sequence was extended with two zero elements on either side.

In order to simplify the presentation, it has here been assumed that all propagation channel polynomials $h_i(q^{-1})$ have the same number of coefficients. This is however not necessary. For example, if the time delays in the propagation channels only span one symbol interval, [0,T], then the q^{-1} term, h_{21}, in $h_2(q^{-1})$ is not necessary since it extends the possible time delays to the interval [T,1.5T]. Only the zero order coefficient, h_{20}, is required and consequently coefficients associated with larger delays should not be included in the polynomial $h_2(q^{-1})$. The parameter vector $\boldsymbol{h}$ and the regressor $\boldsymbol{\mathcal{X}}(t)$ should be reduced accordingly. The removal of superfluous parameters will lead to a better estimation of the remaining parameters since unnecessary degrees of freedom have been removed.

By changing the number of pulse shaping filters per branch and their spacing and removing unnecessary coefficients in the channel filters $h_i(q^{-1})$, an arbitrary refined model, spanning an arbitrary delay spread, can be obtained.

2.3.3 Example

To illustrate the use of pulse shaping information in channel estimation we have here applied it to two different pulse shapes: a derotated linearization of the GMSK modulation with BT=0.3 as in GSM, see Appendix 2.A.1, and a raised cosine pulse with a roll-off factor of 0.35 as in the North American standard IS-136. In both cases binary signaling is assumed. For the GSM signal, a receiver filter modeled as a fourth order Butterworth filter with a bandwidth of 90 kHz, has also been included.

A channel with independently Rayleigh fading taps, of the same average power, with the delays 0, 0.25, 0.5, 0.75 and 1.0 symbol periods has been used as a test channel. The channels were constant during each frame but independently fading between frames. The receiver had two antenna elements.

Five different channel estimation algorithms were used. Three utilizing the pulse shaping information and two using a directly parametrized FIR channel model. The three algorithms utilizing the pulse shaping information used T-,T/2- and T/3-spaced interpolation respectively, spanning delays in the interval [0,T]. The two FIR channel models used 6 and 4 taps with a delay of 1 and 2 taps respectively. The FIR channel taps were estimated with a standard least squares method. All methods used the same number of training data for the channel estimation in the GSM and the IS-136 examples respectively. For the GSM pulse 26 symbols and for the IS-136 pulse, 14 symbols were used in the training sequence[3]. The equalization was performed with an MLSE operating on a channel truncated to a pure delay of one tap followed by 5 taps. This truncation did not affect the resulting BER significantly except possibly for high SNR. The MLSE used was a multi-channel MLSE assuming white Gaussian noise, see for example [59].

In Figures 2.9 and 2.10, the taps of the selected pulse shaping polynomials can be seen for the linearized GSM pulse and IS-136 pulse respectively. For the GSM pulse, the fourth order lowpass Butterworth receiver filter with bandwidth 90 KHz has been included in the pulse shaping function. The dotted line represents T-spaced interpolation between time shifted versions of $p_{0.0}(q^{-1})$ and $p_{-0.5}(q^{-1})$ in the "T-branch" and "T/2-branch" respectively. We can see that due to the smoother GSM pulse, the T-spaced interpolation in Figure 2.9 will be much better for this pulse as compared to the IS-136 pulse in Figure 2.10.

In Figures 2.11 and 2.12, the relative channel estimation errors[4] for the T, T/2 and T/3 spaced interpolation can be seen for the GSM and the IS-136 pulses respectively. Note that the channel error is relatively small for the GSM pulse with T-spaced interpolation. This agrees with the dotted interpolation lines drawn in Figure 2.9. The major part of the improvement in the modeling comes when going from T-spaced interpolation to T/2-spaced interpolation. The improvement between T/2-spaced and T/3-spaced inter-

[3]This corresponds to the number of training symbols in the respective systems.

[4]The relative channel errors are defined as the Frobenious norm of the error in the channel matrix divided by the Frobenious norm of the true channel matrix.

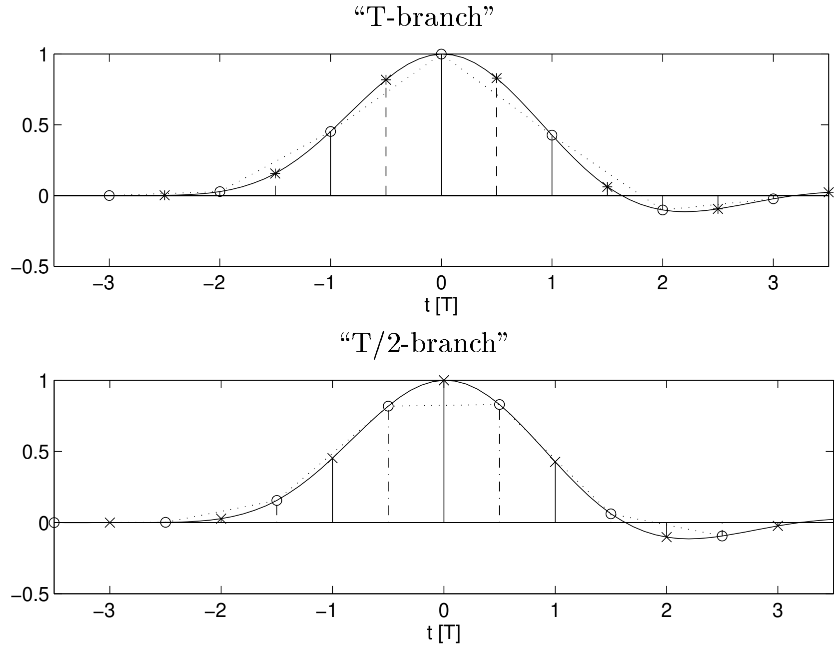


Figure 2.9: Real part of a derotated linearized GSM pulse (including the receiver filter causing the asymmetry), $p(t)$, and real parts of the taps of the derotated pulse shaping filters, $p_{0.0}(q^{-1})$ (o) and $p_{0.5}(q^{-1})$ ($*$) for the “T-branch” and $p_{-0.5}(q^{-1})$ (o) and $p_{0.0}(q^{-1})$ (x) for the “T/2-branch”. The imaginary parts are approximately zero.

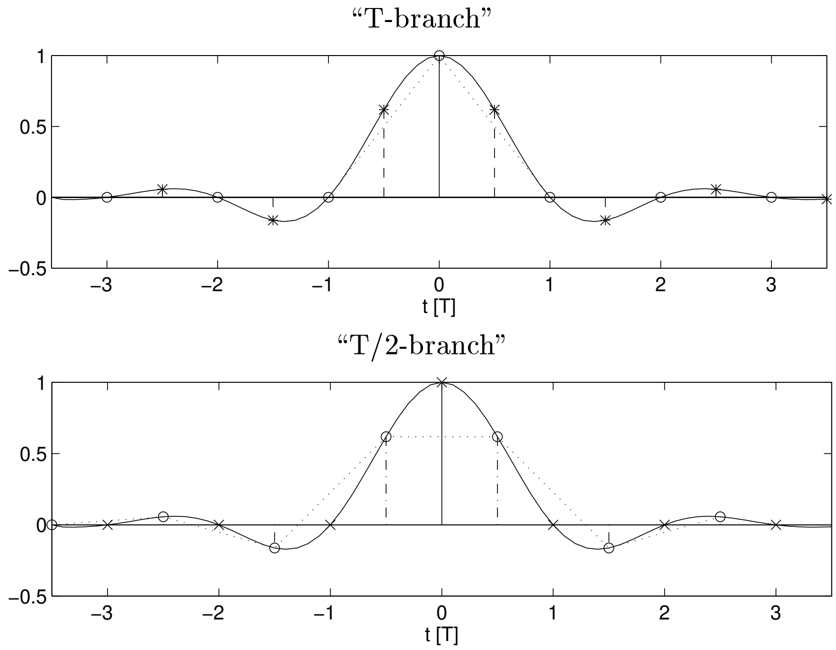


Figure 2.10: IS-136 pulse, $p(t)$, and the taps of the pulse shaping filters, $p_{0.0}(q^{-1})$ (o) and $p_{0.5}(q^{-1})$ $(*)$ for the "T-branch" and $p_{-0.5}(q^{-1})$ (o) and $p_{0.0}(q^{-1})$ (x) for the "T/2-branch".

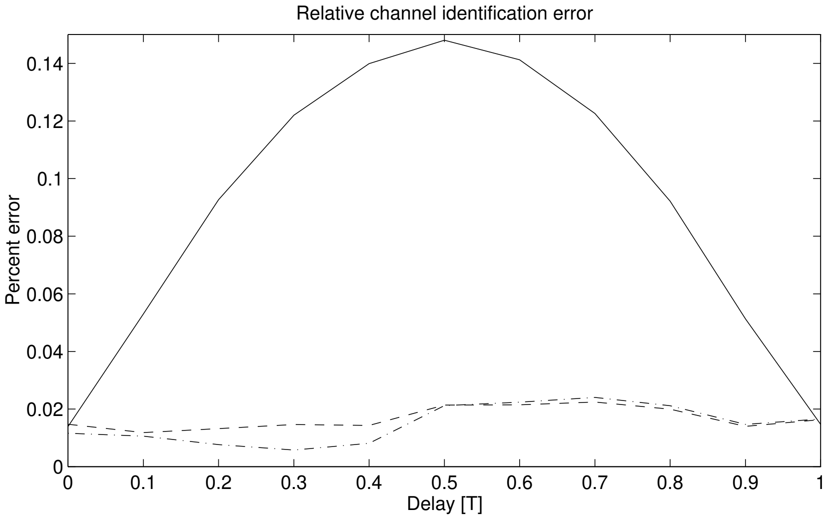


Figure 2.11: Relative approximation error for a single GSM pulse with a delay between 0 and T. T-spaced modeling (solid), T/2-spaced modeling (dashed) and T/3-spaced modeling (dash-dotted).

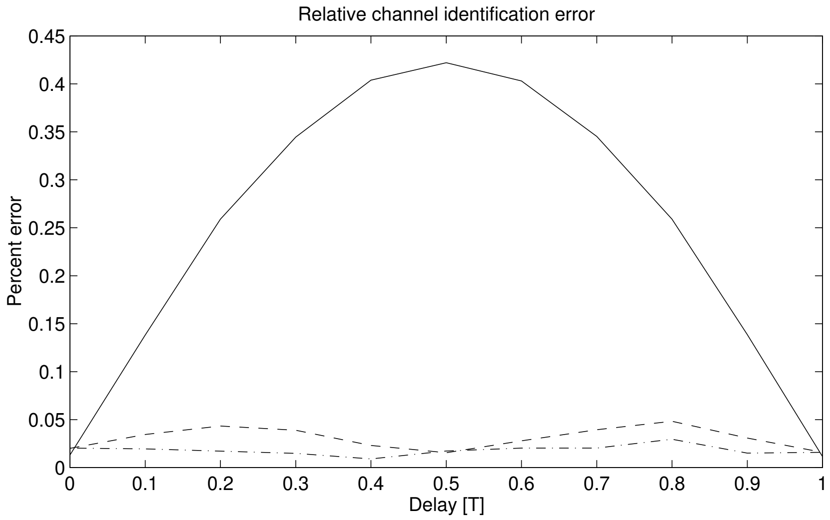


Figure 2.12: Relative approximation error for a single IS-136 pulse with a delay between 0 and T. T-spaced modeling (solid), T/2-spaced modeling (dashed) and T/3-spaced modeling (dash-dotted).

polation is smaller, at least measured in absolute terms. We can conclude that the T-spaced interpolation seems sufficient in GSM and that the T/2-spaced interpolation should suffice for both pulses except when very accurate modeling is required.

An example of how the interpolation is performed can be seen in Figure 2.13 for the IS-136 pulse, without noise. We can see how the coefficients vary as a single pulse with delay varying from 0 to T is being modeled. Although mainly two taps at a time model the pulse, a third tap in the T/2- and T/3-spaced interpolation has a non-negligible amplitude. The sum of the taps is also different from one. Thus, as mentioned earlier, the method is not a true interpolation between pulse shaping filters but rather an optimized linear combination of shifted pulse shaping filters.

In Figures 2.14 and 2.15 the resulting BER of the MLSE and the relative channel errors can be seen for the GSM and IS-136 pulse respectively. In both cases, utilization of the pulse shaping function provides better performance. For the GSM pulse using 26 training symbols, the improvement is however not so large. This can possibly be explained by the longer training sequence. Utilizing the pulse shaping function reduces the number of complex parameters to estimate to 2, 3 and 4 for T-, T/2- and T/3-spaced

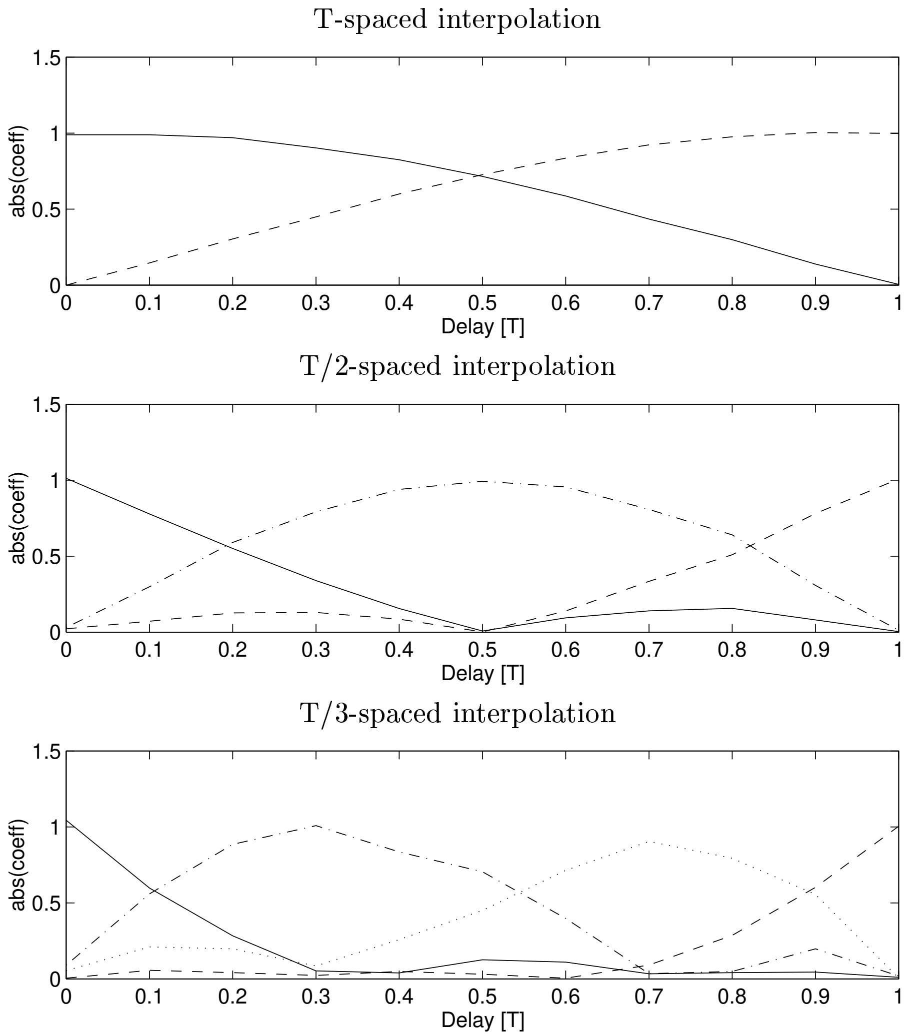


Figure 2.13: Coefficients in the parameter vector $\boldsymbol{h}$ for T, T/2 and T/3 spaced interpolation as the delay of a single pulse varies from 0 to T. Where applicable: coefficient number one (solid), coefficient number two (dashed), coefficient number three (dash-dotted) and coefficient number four (dotted).

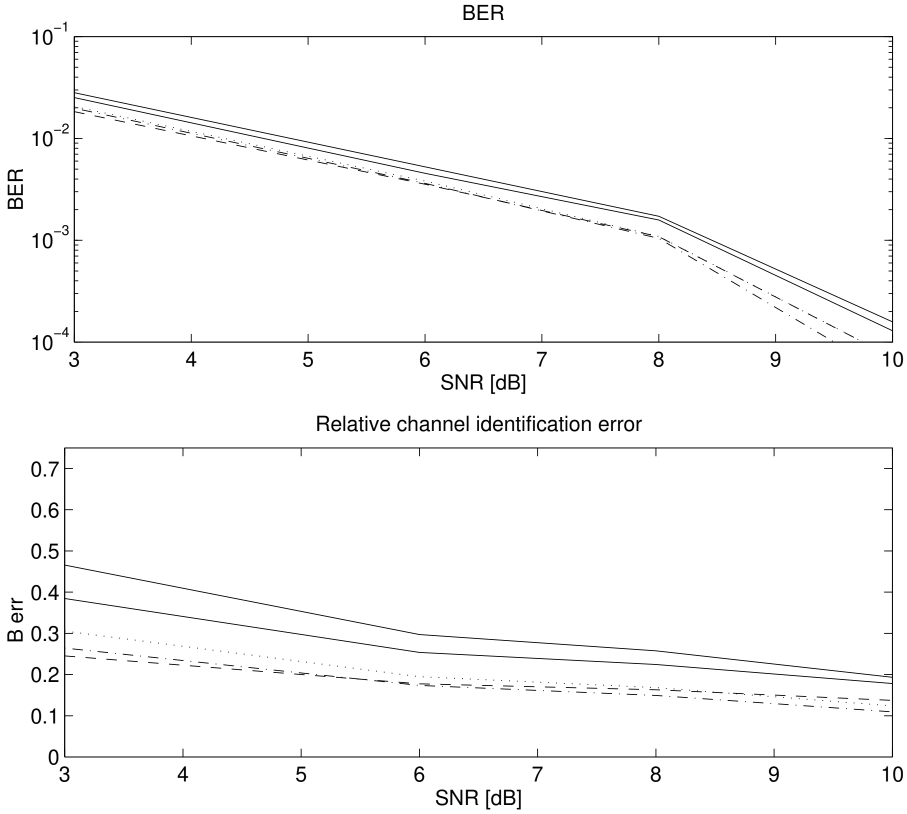


Figure 2.14: BER and relative channel error for the GSM pulse and the chosen test channel. FIR channel estimation with 6 (upper solid) and 4 taps (lower solid), T-spaced interpolation (dashed), T/2-spaced interpolation (dash-dotted) and T/3-spaced interpolation (dotted).

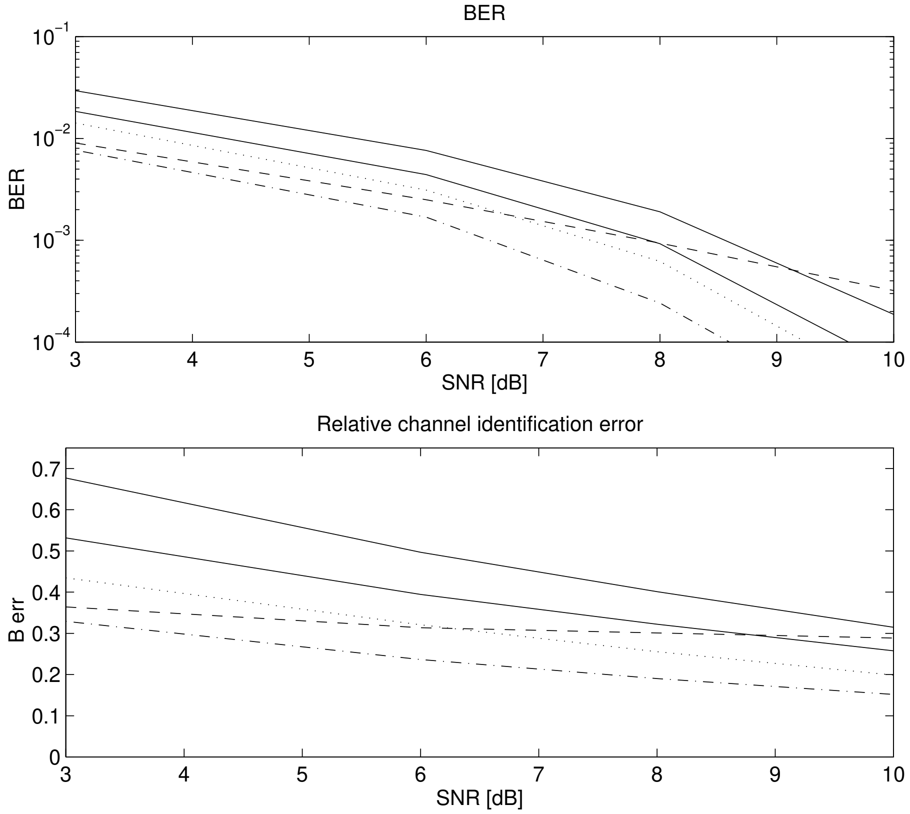


Figure 2.15: BER and relative channel error for the IS-136 pulse and the chosen test channel. FIR channel estimation with 6 (upper solid) and 4 taps (lower solid), T-spaced interpolation (dashed), T/2-spaced interpolation (dash-dotted) and T/3-spaced interpolation (dotted).

interpolation as compared to 12 and 8 complex coefficients for the FIR channel estimation ($T/2$ fractionally spaced). This reduction of the number of parameters will have the largest impact on the accuracy when using shorter training sequences, as for example in the IS-136 case with only 14 training symbols.

Even though the BER is insignificantly improved in the GSM case, there is a non-negligible improvement in the *channel* estimation error, especially for low SNR. Depending on how sensitive the used equalizer is to channel estimation errors, this may pay off also for the GSM case. For example, an equalizer using many antennas is often more dependent on good channel estimates than an equalizer using only a few antenna elements.

For the GSM case, the T-spaced interpolation appears to be reasonably good. Only at the higher SNR is the T/2-spaced interpolation better. The T/3-spaced interpolation is a worse choice, at least for low SNR. This is most likely because the identification becomes ill-conditioned as adjacent pulse shaping filters become more correlated.

A different way of performing channel estimation using information about the pulse shaping function using correlation between the received signal and the modulated transmitted training sequence can be found in [89]. No comparison has however been done with that method.

For the IS-136 pulse, the T-spaced interpolation is too coarse. The true channel can thus not be accurately modeled. This is particularly apparent at high SNR. The T/2-spaced interpolation has the best performance. Again the T/3-spaced interpolation has a worse performance than the T/2-spaced interpolation, suggesting that the model then is ill-conditioned. The improvements in both channel estimation error and detector BER is higher for this case. This is most likely due to the shorter training sequence.

2.3.4 Summary

We can conclude that by utilizing the pulse shaping information we can model the channel well with a reduced number of parameters and, not unexpectedly, the channel estimate improves compared to estimating the taps of the FIR channel model directly.

For the simulations with the GSM pulse, T-spaced interpolation appears to suffice when the performance is measured in BER. The T/2-spaced interpolation has better channel estimates though which could possibly be of value in a different scenario.

For the example simulations with the IS-136 pulse, the T-spaced interpolation is insufficient. The T/2-spaced interpolation performed notably better in both BER and relative channel error.

For both pulse shapes the performance decreased when using the T/3-spaced interpolation. This is suspected to be due to the identification becoming ill-conditioned as the correlation between the pulse shaping filters increase.

2.4 Temporal Parametrization of Multi-User Channels

A wireless communication system were capacity is an issue is often interference limited, i.e. the signal quality is limited because of co-channel interference rather than noise. If the channels to the interfering signals can be estimated jointly with the desired signal, the channel estimates of the desired signal as well as of the interferers can be improved. This can be utilized in the equalization of the signal.

Since in joint multi-user channel estimation, the number of parameters to be estimated grows linearly with the number of users while the number of equations remains constant, it is important to use as few parameters as possible per user. The number of equations is limited since the length of the training sequence is limited.

One way to economize on the number of parameters to estimate, is to utilize knowledge of the pulse shaping in the transmitter and the receiver as described in Section 2.3. Since this method economizes on the number of parameters to be estimated for each user, it will improve the channel estimates in line with the Parsimony principle. In some cases it will also make otherwise impossible joint multi-user channel estimation possible. We illustrate this by studying a multi-user channel estimation example.

2.4.1 Channel Estimation

The temporal parametrization in Section 2.3.2 can be generalized to multiple users. A model for two users with one antenna element, using fractionally spaced sampling (two samples per symbol) is depicted in Figure 2.16. The model of the sampled output from one antenna can be expressed in equations as

$$\boldsymbol{y}(t) = \boldsymbol{h}(q^{-1})\boldsymbol{\mathcal{X}}(t) + \boldsymbol{n}(t) \tag{2.37}$$

where $\boldsymbol{n}(t) = [n(t) \;\; n(t+\frac{T}{2})]$ is a vector containing the noise samples and the unknown propagation channels for the users, $\boldsymbol{h}(q^{-1})$, is given by

$$\boldsymbol{h}(q^{-1}) = [\boldsymbol{h}_1(q^{-1}) \;\; \boldsymbol{h}_2(q^{-1})] \tag{2.38}$$

where

$$\boldsymbol{h}_i(q^{-1}) = \left[h_{i1}(q^{-1}) \;\; h_{i2}(q^{-1})\right] \tag{2.39}$$

and

$$\boldsymbol{X}(t) = \begin{bmatrix} x_{111}(t) & x_{112}(t) \\ x_{121}(t) & x_{122}(t) \\ x_{211}(t) & x_{212}(t) \\ x_{221}(t) & x_{222}(t) \end{bmatrix} = \begin{bmatrix} \boldsymbol{P}(q^{-1})d_1(t) \\ \boldsymbol{P}(q^{-1})d_2(t) \end{bmatrix}. \tag{2.40}$$

The pulse shaping matrix is the same as for the single user case, i.e.

$$\boldsymbol{P}(q^{-1}) = \begin{bmatrix} p_{0.0}(q^{-1}) & p_{-0.5}(q^{-1}) \\ p_{0.5}(q^{-1}) & p_{0.0}(q^{-1}) \end{bmatrix}. \tag{2.41}$$

Compared with the model in (2.23) – (2.28), we now, with two users, have twice as many parameters to estimate but the same number of equations (sampled outputs).

In order to form a system of equations we vectorize equation (2.37) and obtain

$$\boldsymbol{y}(t) = \boldsymbol{h}\boldsymbol{\mathcal{X}}(t) \tag{2.42}$$

where

$$\boldsymbol{h} = [\boldsymbol{h}_1 \;\; \boldsymbol{h}_2] \tag{2.43}$$

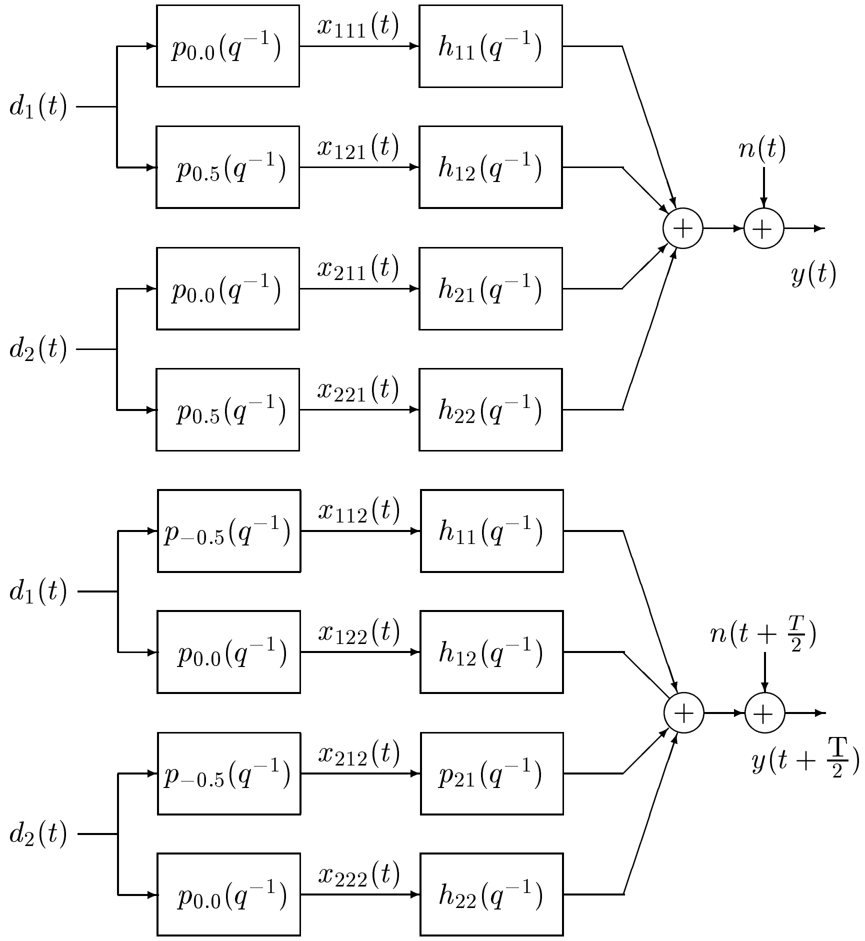


Figure 2.16: Example multi-user channel model for two users sending the messages $d_1(t)$ and $d_2(t)$ respectively to one antenna element, where the received signal is sampled twice per symbol time.

with

$$\boldsymbol{h}_i = [h_{i1} \;\; h_{i2}] \tag{2.44}$$

$$\boldsymbol{h}_{ij} = [h_{ij0} \;\; h_{ij1} \;\; \ldots \;\; h_{ijnh}] \tag{2.45}$$

and

$$\boldsymbol{\mathcal{X}}(t) = \begin{bmatrix} \bar{x}_{111}(t) & \bar{x}_{112}(t) \\ \bar{x}_{121}(t) & \bar{x}_{122}(t) \\ \bar{x}_{211}(t) & \bar{x}_{212}(t) \\ \bar{x}_{221}(t) & \bar{x}_{222}(t) \end{bmatrix} \tag{2.46}$$

where

$$\bar{x}_{ijk}(t) = [x_{ijk}(t) \;\; x_{ijk}(t-1) \;\; \ldots \;\; x_{ijk}(t-nh)]^T. \tag{2.47}$$

The unknown channel $\boldsymbol{h}$ can now be estimated with the least squares method as

$$\hat{\boldsymbol{h}} = \hat{R}_{\boldsymbol{y}\boldsymbol{\mathcal{X}}} \hat{R}^{-1}_{\boldsymbol{\mathcal{X}}\boldsymbol{\mathcal{X}}} \tag{2.48}$$

where

$$\hat{R}_{\boldsymbol{y}\boldsymbol{\mathcal{X}}} = \frac{1}{t_{max} - t_{min} + 1} \sum_{t=t_{min}}^{t=t_{max}} \boldsymbol{y}(t) \boldsymbol{\mathcal{X}}^H(t) \tag{2.49}$$

and

$$\hat{R}_{\boldsymbol{\mathcal{X}}\boldsymbol{\mathcal{X}}} = \frac{1}{t_{max} - t_{min} + 1} \sum_{t=t_{min}}^{t=t_{max}} \boldsymbol{\mathcal{X}}(t) \boldsymbol{\mathcal{X}}^H(t) \tag{2.50}$$

The total channel $\boldsymbol{b}_i(q^{-1})$ to each of the two users can then be estimated with

$$\hat{\boldsymbol{b}}_i(q^{-1}) = \hat{\boldsymbol{h}}_i(q^{-1}) \boldsymbol{P}(q^{-1}). \tag{2.51}$$

The time instants t_{min} and t_{max} represents the minimum and maximum time samples of the training data used. In multi-antenna systems, a channel estimate (2.51) can be computed separately for each antenna output. Similar considerations as in Section 2.3.2 apply, such as the choice of the coarseness of the modeling.

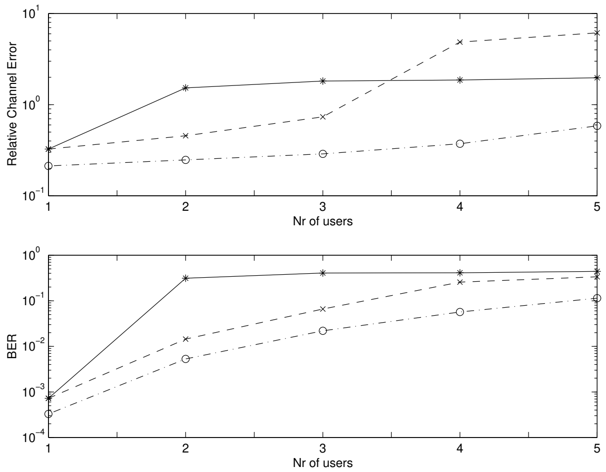


Figure 2.17: Relative channel error and the BER for one user for the multi-user channel estimation example. Single user LS estimation (solid *), Joint muli-user LS (dashed x), Joint multi-user estimation utilizing pulse shaping information (dash-dotted o).

2.4.2 Examples

Since it would be advantageous to use multiple antennas for multi-user detection, we here use a receiver with 4 antennas. The number of users is ranging from one to 4. Each mobile had a channel with equal average power Rayleigh fading taps at delays 0.00, 0.33, 0.67 and 1.00 symbol intervals. The taps for each mobile and antenna element were independently fading. The channels were constant during each frame but independently fading between frames. A raised cosine pulse with roll-off 0.35 was used for the pulse shaping. Other parameters were: 18 training symbols, fractionally spaced sampling (two samples per symbol), T/2-spaced modeling. The SNR was 3dB and all mobiles had equal average relative strength.

After the channels from the mobiles to all antennas had been estimated, a single user multi-channel MLSE with spatial interference cancellation, see Chapter 4, was applied to the received signal, to detect one user. In Figure 2.17 the relative channel estimation errors and the BER for the detected user can be seen for different number of mobiles present.

First we note that the joint LS channel estimation performs better than the single user LS channel estimation for the cases with two and three users. The reason for this is likely the improved modeling that is accomplished by modeling all users instead of treating some of them as noise. For four or more users the joint LS method is however worse. Then the parameters in the joint model are simply too many. For the number of users studied here, the method utilizing the pulse shaping information performs better than both the single and joint LS channel estimation. This is because this method has fewer parameters to estimate per user. For the joint multi-user channel estimation, this becomes especially important as the number of users increases.

With pulse shaping included, the channel spans about 4-6 symbol intervals depending on where we choose to truncate. In the LS methods we chose to use 4 taps and in the pulse shaping method we used 2 taps in each branch of $\mathrm{H}(q^{-1})$. The joint LS method will thus have 8 parameters *per user* to estimate with 36 equations (2×4 and 2×18 because of the fractionally spaced sampling). The pulse shaping method will have 4 parameters (2 per branch) *per user* to estimate with 36 equations. As the number of users increases this difference becomes more important. We can understand this by considering a limiting case when the number of users is so large that the joint LS method has more parameters than equations while this is not the case for the pulse shaping method. In this case the joint LS method will not work, while the pulse shaping method will. This is why we see that the joint LS method degrades in performance faster than the method utilizing the pulse shaping when the number of users increases, *cf* Figure 2.17.

2.4.3 Summary

When estimating the channels to multiple users jointly with a short training sequence it is important to reduce the number of parameters per user. This is because the number of parameters increases linearly with the number of users while the number of equations, determined by the length of the effective length of the training sequence[5], remains constant.

[5]With the effective length of the training sequence we mean the number of training relations that can be formed with the method we are using. This number is usually smaller than the length of the training sequence since we cannot fully exploit the training symbols in the beginning and in the end of the training sequence.

We have demonstrated one way of doing this by means of a method which utilize the pulse shaping information in the transmitter and in the receiver in order to economize on the number of parameters to be estimated. In the multi-user channel estimation example shown we can see that the method improves the channel estimate considerably, especially as the number of users increases.

For the pulse shaping method it is also interesting to note that when performing fractionally spaced sampling the number of channel parameters to estimate does *not* increase. The number of equations *do* however, and this improves the channel estimates as long as the fractionally spaced noise samples are not too much correlated.

From the model errors, in Section 2.3 one can see that T/2-spaced modeling suffices for the GSM and IS-136 pulses studied. For the GSM pulse it seems to suffice with T-spaced modeling (or some spacing between T and T/2, e.g. 2T/3 for example).

It is possible to reduce the number of parameters to be estimated even further. The number of taps that are non-zero in a channel will vary with time. In order to accommodate paths with long delay times one will have to use a channel with many taps if a fixed set of taps is being used.

Instead of using a long channel with many taps, one can try to determine which specific taps should be non-zero at the time of the channel estimation. Since the majority of the channels will only contain a few non-zero taps, this will typically reduce the number of taps considerably. Again this will be valuable for multi-user channel estimation.

2.5 Spatial Parametrization

When using an antenna array it may be possible to utilize the spatial structure of the channel to obtain a model with fewer parameters. For example if the signal arrives from distinct directions this can be utilized in the channel estimation.

Here estimation of the channels to an antenna array using a parametrization in directions of arrival and gains of signal paths is investigated. This channel estimate is compared to the temporal only least squares FIR channel estimation.

When we estimate the FIR channel coefficients to each of the antenna elements *independently* using, for example, the least squares method in (2.21), we do not utilized the spatial structure of the incoming signal. If the signal arrives from a small number of directions, we may obtain improved channel estimates if we parametrize the joint channel to *all* antenna elements in terms of directions of arrival (DOA's) and path gains. This can be realized by using the CDEML algorithm (Coherent DEcoupled Maximum Likelihood Estimation) [15]. A version of the CDEML algorithm is briefly presented and discussed below. The algorithm is also tested on a scenario with multipath propagation and intersymbol interference and compared with the traditional least squares channel FIR channel estimation. The performance of the algorithm is also measured in terms of the bit error rate (BER) of a DFE, designed from the estimated channels.

2.5.1 Least Squares Channel Estimation

The estimation of the channel by identifying the parameters of a vector FIR model, as described in Section 2.2, is a purely temporal method. The estimate of the channel matrix $\boldsymbol{B}$ in (2.21)

$$\hat{\boldsymbol{B}}_{\text{LS}} = \hat{\boldsymbol{R}}_{\boldsymbol{dy}}^{H} \hat{\boldsymbol{R}}_{\boldsymbol{dd}}^{-1} \tag{2.52}$$

is just a compact notation for the separate temporal only FIR least squares channel estimates to the individual antennas.

This channel estimate, as mentioned above, makes no assumption about the number of incoming signal paths. This makes it robust with respect to such assumptions, but if the number of incoming signals per symbol delay is small relative to the number of antennas, an improvement may be possible. A method utilizing this fact is presented next.

2.5.2 Coherent Decoupled Maximum Likelihood Channel Estimation

The method described below is a version of the CDEML algorithm presented in [15]. Let the overall channel model be represented by the vector FIR model in (2.18)

$$\boldsymbol{y}(t) = \boldsymbol{B}\boldsymbol{d}(t) + \boldsymbol{n}(t) \tag{2.53}$$

where $\boldsymbol{d}(t) = [d(t)\ d(t-1)\ \ldots\ d(t-nb)]^T$ and $\boldsymbol{B}$ is the channel matrix (2.19) (2.20) of size $M \times (nb+1)$. The signal $d(t)$ is the transmitted discrete symbol sequence. The noise plus interference, $\boldsymbol{n}(t)$, is here assumed to be circularly symmetric[6], zero-mean and Gaussian with second order moments given by

$$E[\boldsymbol{n}(t)\boldsymbol{n}^H(s)] = \boldsymbol{Q}\delta_{t,s} \qquad E[\boldsymbol{n}(t)\boldsymbol{n}^T(s)] = \mathbf{0}. \tag{2.54}$$

Consider the DOA's for the incoming signals[7]

$$\boldsymbol{\theta} = [\theta_{01}\ \ldots\ \theta_{0,k_1} \ldots\ \theta_{nb,1} \ldots\ \theta_{nb,k_{nb}}]^T \tag{2.55}$$

and the corresponding gains

$$\boldsymbol{\gamma} = [\gamma_{01}\ \cdots\ \gamma_{0,k_1} \cdots\ \gamma_{nb,1} \cdots\ \gamma_{nb,k_{nb}}]^T. \tag{2.56}$$

The indices in the expressions above for $\boldsymbol{\theta}$ and $\boldsymbol{\gamma}$ denote tap number (delay) and path number within each tap, in that order. Each tap is assumed to be affected by k_c signal paths arriving from significantly different directions θ_{c1} to θ_{ck_c}. The matrix $\boldsymbol{B}$ can now be parametrized in terms of these DOA's and gains

$$\boldsymbol{B}(\boldsymbol{\theta}, \boldsymbol{\gamma}) = \boldsymbol{A}(\boldsymbol{\theta})\boldsymbol{\Gamma}(\boldsymbol{\gamma}) \tag{2.57}$$

where

$$\boldsymbol{A}(\boldsymbol{\theta}) = [\boldsymbol{a}(\theta_{01})\ \ldots\ \boldsymbol{a}(\theta_{0,k_1})\ \ldots\ \boldsymbol{a}(\theta_{nb,1})\ \ldots\ \boldsymbol{a}(\theta_{nb,k_{nb}})] \tag{2.58}$$

[6] A complex random variable is said to be circularly symmetric if its probability density function is circularly symmetric around its zero mean. This will be the case if its real and imaginary parts are uncorrelated, equally scaled and have zero mean.

[7] The first number in the subscript of the DOA θ_{ij} refers to the tap number and the second number refers to the number of the DOA within the tap.

and

$$\boldsymbol{\Gamma}(\boldsymbol{\gamma}) = \begin{bmatrix} \gamma_{01} & 0 & \cdots & 0 \\ \vdots & \vdots & \ddots & \\ \gamma_{0k_1} & & & \\ 0 & \gamma_{11} & & \vdots \\ & \vdots & & \\ & \gamma_{1k_2} & & \\ \vdots & & \ddots & 0 \\ & & & \gamma_{nb1} \\ & & & \vdots \\ 0 & & & \gamma_{nbk_{nb}} \end{bmatrix}. \tag{2.59}$$

The vectors, $\boldsymbol{a}(\theta_{ij})$), in (2.58) constitute the array response vectors for a signals arriving from the angles θ_{ij}.

The parameter vectors $\boldsymbol{\theta}$ and $\boldsymbol{\gamma}$, as seen above, are partitioned according to which delay in $\boldsymbol{B}$ they correspond to:

$$\boldsymbol{\gamma} = [\boldsymbol{\gamma}_0^T \ \boldsymbol{\gamma}_1^T \ \ldots \ \boldsymbol{\gamma}_{nb}^T] \ \text{ where } \ \boldsymbol{\gamma}_c = [\gamma_{c1} \ \ldots \ \gamma_{ck_c}]^T \tag{2.60}$$

$$\boldsymbol{\theta} = [\boldsymbol{\theta}_0^T \ \boldsymbol{\theta}_1^T \ \ldots \ \boldsymbol{\theta}_{nb}^T], \ \text{ where } \ \boldsymbol{\theta}_c = [\theta_{c1} \ \ldots \ \theta_{ck_c}]^T. \tag{2.61}$$

It can be shown (see [15]) that a large sample maximum likelihood estimate of $\boldsymbol{\theta}$ and $\boldsymbol{\gamma}$ can be obtained by minimizing

$$F(\boldsymbol{\theta}, \boldsymbol{\gamma}) = tr[\boldsymbol{R_{dd}}(\boldsymbol{B}(\boldsymbol{\theta}, \boldsymbol{\gamma}) - \hat{\boldsymbol{B}}_{\mathrm{LS}})^H \hat{\boldsymbol{Q}}^{-1} (\boldsymbol{B}(\boldsymbol{\theta}, \boldsymbol{\gamma}) - \hat{\boldsymbol{B}}_{\mathrm{LS}})] \tag{2.62}$$

where

$$\boldsymbol{R_{dd}} = \lim_{N \to \infty} \frac{1}{N - nb} \sum_{t=nb+1}^{N} \boldsymbol{d}(t) \boldsymbol{d}^H(t). \tag{2.63}$$

Above, $\hat{\boldsymbol{B}}_{LS}$ is the least squares channel matrix estimate (2.52) and

$$\hat{\boldsymbol{Q}} = \frac{1}{N - nb} \sum_{t=nb+1}^{N} (\boldsymbol{y}(t) - \hat{\boldsymbol{B}}_{\mathrm{LS}} \boldsymbol{d}(t))(\boldsymbol{y}(t) - \hat{\boldsymbol{B}}_{\mathrm{LS}} \boldsymbol{d}(t))^H \tag{2.64}$$

is an estimate of the noise plus interference covariance matrix.

In digital communications the symbol sequence $d(t)$ is assumed to be white. The covariance matrix, $\boldsymbol{R_{dd}}$, will then be diagonal. In this case, the maximum likelihood estimates of the angles, $\boldsymbol{\theta}$, and the gains, $\boldsymbol{\gamma}$, can be found by considering the following minimization for each *column*, $\hat{\boldsymbol{b}}_c$, in $\hat{\boldsymbol{B}}$ *separately*:

$$\{\hat{\boldsymbol{\theta}}_{\boldsymbol{c}}, \hat{\boldsymbol{\gamma}}_{\boldsymbol{c}}\} = \arg\min_{\boldsymbol{\theta}_c, \boldsymbol{\gamma}_c} [\boldsymbol{A}(\boldsymbol{\theta}_c)\boldsymbol{\gamma}_c - \hat{\boldsymbol{b}}_c]^H \hat{\boldsymbol{Q}}^{-1} [\boldsymbol{A}(\boldsymbol{\theta}_c)\boldsymbol{\gamma}_c - \hat{\boldsymbol{b}}_c] \tag{2.65}$$

where $\hat{\boldsymbol{b}}_c$ is the relevant column in $\hat{\boldsymbol{B}}$ which corresponds to a vector tap in the FIR channel $\boldsymbol{b}(q^{-1})$.

The fact that the minimization can be decoupled greatly reduces the computational complexity. From (2.65) we can see that we are looking for $\boldsymbol{\theta}_c$'s and $\boldsymbol{\gamma}_c$'s that minimize the weighted squared norm of the difference between the least squares estimated column $\hat{\boldsymbol{b}}_c$ of $\hat{\boldsymbol{B}}_{\mathrm{LS}}$ and the parametrized estimate $\boldsymbol{A}(\boldsymbol{\theta}_c)\boldsymbol{\gamma}_c$.

Returning to (2.62), we see that when $\boldsymbol{R_{dd}}$ is diagonal, the overall criterion we want to minimize, is simply the weighted squared Frobenius norm of the difference between the parametrized channel matrix, $\boldsymbol{B}(\boldsymbol{\theta}, \boldsymbol{\gamma})$ and the least squares estimated channel matrix $\hat{\boldsymbol{B}}_{LS}$, i.e. $\| \boldsymbol{B}(\boldsymbol{\theta}, \boldsymbol{\gamma}) - \hat{\boldsymbol{B}}_{\mathrm{LS}} \|^2_{\boldsymbol{Q}^{-1}}$.

For a given set of angles $\hat{\boldsymbol{\theta}}_c$, the minimizing gain $\hat{\boldsymbol{\gamma}}_c$ in (2.65) is given by the weighted least squares solution

$$\hat{\boldsymbol{\gamma}}_c(\hat{\boldsymbol{\theta}}_c) = \{\boldsymbol{A}^H(\hat{\boldsymbol{\theta}}_c)\hat{\boldsymbol{Q}}^{-1}\boldsymbol{A}(\hat{\boldsymbol{\theta}}_c)\}^{-1}\boldsymbol{A}^H(\hat{\boldsymbol{\theta}}_c)\hat{\boldsymbol{Q}}^{-1}\hat{\boldsymbol{b}}_c. \tag{2.66}$$

By substituting (2.66) into (2.65) we obtain an expression for the estimated directions of arrival [15]

$$\begin{aligned}\hat{\boldsymbol{\theta}}_c \;=\; & \arg\min_{\boldsymbol{\theta}_c}\{\hat{\boldsymbol{b}}_c^H[\hat{\boldsymbol{Q}}^{-1} - \hat{\boldsymbol{Q}}^{-1}\boldsymbol{A}(\boldsymbol{\theta}_c) \times \\ & (\boldsymbol{A}^H(\boldsymbol{\theta}_c)\hat{\boldsymbol{Q}}^{-1}\boldsymbol{A}(\boldsymbol{\theta}_c))^{-1}\boldsymbol{A}^H(\boldsymbol{\theta}_c)\hat{\boldsymbol{Q}}^{-1}]\hat{\boldsymbol{b}}_c\}. \end{aligned} \tag{2.67}$$

In order to assure convergence to the global minimum, a good initial value is required for $\boldsymbol{\theta}_c$. In [15], it is proposed to initialize $\boldsymbol{\theta}_c$ with the k_c lowest local minima of the function

$$f(\theta) = \hat{\boldsymbol{b}}_c^H \left[\hat{\boldsymbol{Q}}^{-1} - \frac{\hat{\boldsymbol{Q}}^{-1} \boldsymbol{a}(\theta) \boldsymbol{a}^H(\theta) \hat{\boldsymbol{Q}}^{-1}}{\boldsymbol{a}^H(\theta) \hat{\boldsymbol{Q}}^{-1} \boldsymbol{a}(\theta)} \right] \hat{\boldsymbol{b}}_c. \tag{2.68}$$

This is exactly the cost function to be minimized if there was only one signal arriving per symbol delay. As the first term is independent of θ, we can instead look for local *maximas* to the function

$$f_{0,c}(\theta) = \frac{\boldsymbol{a}^H(\theta) \hat{\boldsymbol{Q}}^{-1} \hat{\boldsymbol{b}}_c \hat{\boldsymbol{b}}_c^H \hat{\boldsymbol{Q}}^{-1} \boldsymbol{a}(\theta)}{\boldsymbol{a}^H(\theta) \hat{\boldsymbol{Q}}^{-1} \boldsymbol{a}(\theta)}. \tag{2.69}$$

In the simulations performed in this study, this initialization procedure has been found to have some problems. It could, for example, have difficulties in estimating DOA's of signals that were close to a strong co-channel interferer. The presence of a strong co-channel interferer in the noise plus interference covariance matrix, $\boldsymbol{Q}$, can cause a dip in the function $f_{0,c}(\theta)$.

In an attempt to alleviate this problem, $\boldsymbol{Q}^{-1}$ can be removed from $f_{0,c}(\theta)$. The result is a simple "beamformer" DOA estimator. The initial values of the components of θ_c can thus be chosen as the local maximas of the function

$$f_{1,c}(\theta) = \frac{\boldsymbol{a}^H(\theta) \hat{\boldsymbol{b}}_c \hat{\boldsymbol{b}}_c^H \boldsymbol{a}(\theta)}{\boldsymbol{a}^H(\theta) \boldsymbol{a}(\theta)}. \tag{2.70}$$

However, both of these methods have been found to have considerable difficulties in estimating initial values for the DOA's when coherent sources are present. The peaks for the two functions in (2.69) and (2.70) can then have peaks in directions not corresponding to a DOA. This is because side lobes of the "beamformers" involved may pick up the signals and combine them constructively depending on the particular relative phases of the signals involved. These combined signals can have a stronger amplitude then the signals caught by the main lobes of the "beamformers". A DOA will then be indicated at the wrong direction. A solution to this problem is to constrain the antenna array to a uniform linear array. In this case the non-linear minimization of (2.67) can be replaced with a polynomial root-finding technique similar to the one proposed in [94] and [95].

Once the directions of arrival in $\boldsymbol{\theta}_c$ have been estimated the gains in $\boldsymbol{\gamma}_c$ can be computed using (2.66). The estimated $\hat{\boldsymbol{\theta}}$ and $\hat{\boldsymbol{\gamma}}$ are used to form the improved parametrized channel matrix estimate

$$\hat{\boldsymbol{B}}_{\text{CDEML}} = \boldsymbol{A}(\hat{\boldsymbol{\theta}})\boldsymbol{\Gamma}(\hat{\boldsymbol{\gamma}}). \tag{2.71}$$

2.5.3 Simulation Study

A simulation study has been conducted to evaluate the performance of the algorithm in a scenario with multipath propagation and intersymbol interference.

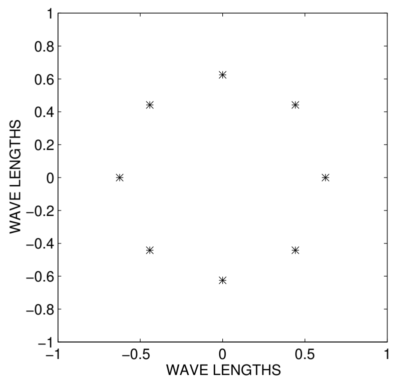
Scenarios

The algorithms were tested using a circular array, see Figure 2.18. The desired signal arrives from the directions 0, 60, -120 and 120 degrees respectively. The respective temporal channels are $1 + 0.5q^{-1}$, $0.5q^{-1} + 0.8q^{-2}$, $0.5q^{-2} + 0.2q^{-3}$ and $0.2q^{-3} + 0.3q^{-4}$. Two-tap channels are chosen in order to simulate imperfect sampling timing or partial response modulation. When co-channel interferers are included they impinge on the antenna array through single tap channels from the directions 90, -60 and -90 degrees, with a total average SIR of 0 dB. The SNR was varied from -3 dB to 6 dB. The scenario is illustrated in Figure 2.19. The number of training symbols used here were 26.

An FIR channel model with 5 taps was used for the channel matrix. Based on this model, a DFE with a feedforward filter of length 4, feedback filters of length 3 and a smoothing lag, ℓ, of 3 was designed and used for estimation of the transmitted symbols.

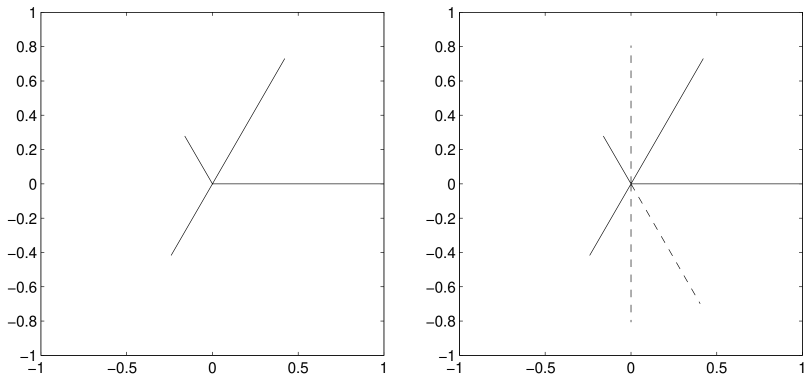
Initial Values for the DOA's

Both the initialization suggested in [15] and the modified initialization suggested in (2.70) have problems with achieving good initial values for the DOA's. In Figure 2.20, examples of the two functions can be seen. By comparing the plots for the functions $f_{0,c}(\theta)$ given by (2.69) and $f_{1,c}(\theta)$ given by (2.70), one can see the result of including $\boldsymbol{Q}^{-1}$ in $f_{0,c}(\theta)$. It can be seen



Circular array

Figure 2.18: Antenna array configuration.



SIR=∞ SIR=0 dB

Figure 2.19: Signal configuration. The solid lines represent the incoming directions of the desired signal and the dashed lines represent co-channel interferers. The left figure depicts the case without co-channel interferers and the right figure depicts the case with co-channel interferers.

that $f_{0,c}(\theta)$ has dips or reduced amplitude at or close to the locations of the co-channel interferers.

Both initializations find the DOA for the signal with delay zero.

In the case with a delay of one symbol interval neither of the initialization algorithms provided a good initial estimate of the DOA's.

For the cases with delays of two and three symbol intervals, the function f_0 has the peak corresponding to the signal arriving from -120 degrees reduced as a result of the presence of the co-channel interferer at the same angle. Note that the co-channel interferer previously at -90 degrees has been moved to -60 degrees in order to illustrate the masking effect. As a result the algorithm using f_0 misses this DOA.

When the signal strength becomes low, as for the delay of four symbol intervals, the peaks corresponding to the true DOA's becomes less distinct and they are therefore more difficult to detect. Both of the algorithms suffer from this effect.

In conclusion it can be said that neither of the functions are ideal for finding initial values for the DOA's. In the experiments performed in conjunction with this study, the DOA estimator which uses the initialization $f_{1,c}(\theta)$ defined in (2.70), have been slightly better for estimating the DOA's. This estimator has therefore been used in the simulations of this study.

If a *uniform linear array* is used, the k_c-dimensional search in (2.67) can be reduced to a polynomial root-finding operation by using a technique similar to that developed in [94] and [95]. Considering the problems of finding good initial values for the DOA's it may be a good idea to restrict the array geometry to the ULA case, and use the root-finding approach. It should be noted, however, that it is not always practical to restrict the array design to the ULA case. One such case of practical interest is when the two polarizations of the signal is measured at each element location.

Delay 0 $f_{0,c}(\theta)$

$f_{1,c}(\theta)$

Delay 2 $f_{0,c}(\theta)$

$f_{1,c}(\theta)$

Delay 3 $f_{0,c}(\theta)$

$f_{1,c}(\theta)$

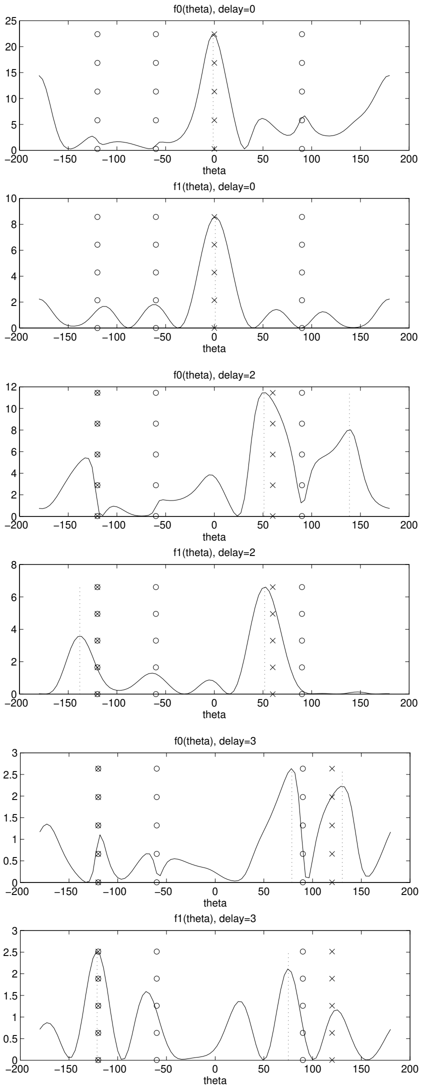


Figure 2.20: Functions, $f_{0,c}(\theta)$ in (2.69) and $f_{1,c}(\theta)$ (2.70) used for finding initial values for the DOA's. The functions are plotted for the delays of 0,2 and 3 symbol intervals in the channel for the ULA case. The minimas chosen as initial values are marked with dotted vertical lines. The true DOA's are marked with crosses. The directions of the co-channel interferers are marked with lines of circles. SIR=0dB and SNR=3dB. Note that the co-channel interferer at -45 degrees has been moved to -60 degrees in order to illustrate the masking effect as discussed in the text.

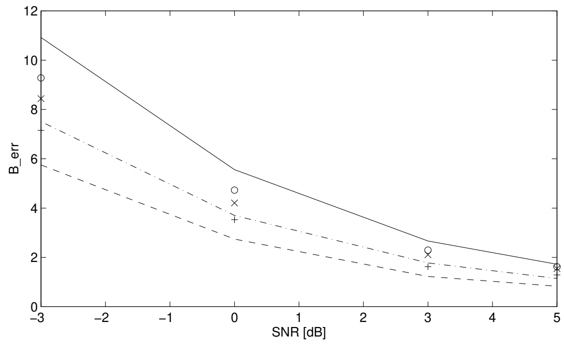
Performance in Terms of Channel Estimation and DFE Bit Error Rate.

The CDEML algorithm was evaluated assuming a different number of DOA's, initializing them either with the true DOA's, or estimating initial DOA's using the expression in (2.70). For the simulation case using "true" DOA's as initial values, if the number of DOA's *assumed* where greater than the *actual* number of DOA's, then the extra DOA's were randomized.

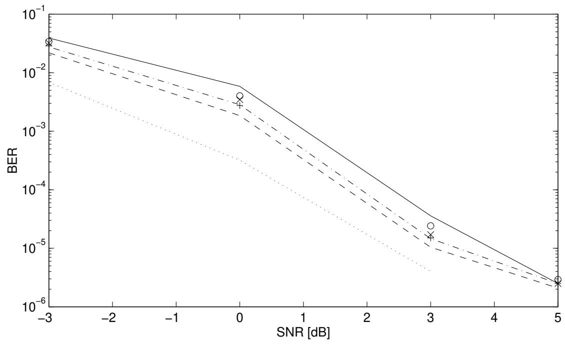
The total mean square error in the estimated channel for different signal-to-noise ratios is illustrated in Figures 2.21 and 2.22. As can be seen, the best performance is achieved by using the CDEML algorithm with the correct number of DOA's, and the true DOA's as initial values. It can also be seen that both overestimating the number of signal paths and using estimated DOA's as initial values, deteriorates the performance of the algorithm. In the scenarios studied, however, all the CDEML versions had better or equal performance, than the direct least squares estimation without any DOA parametrization.

It should be kept in mind though, that the scenario investigated is very well suited for a DOA parametrized channel estimation method such as CDEML. This is because many antenna elements are used, and only a few signal paths per delay impinge on the antenna array. The number of degrees of freedom is thus considerably reduced by parametrizing the channel in DOA's, rather than using the full channel matrix representation.

The estimated channel, and the residuals, were used in the tuning of a DFE with only spatial interference suppression as described in Section 3.3.3. The improved channel estimates result in improved bit error rates as seen in the lower diagrams of Figures 2.21 and 2.22. In these figures, the BER for a DFE where the channel taps are tuned with a direct method, as described in Section 3.3.1 using 1000 training symbols, is also presented. Note also the improvement of the channel estimates for the case with a SIR of 0 dB. It appears that, at least in this simulation, the CDEML algorithm handles the presence of co-channel interferers better than the least squares method. The improvement of the channel estimate is however not reflected to the same degree in the BER. This suggests that the CDEML algorithm improves the channel estimate in dimensions that are not so important for the equalization.

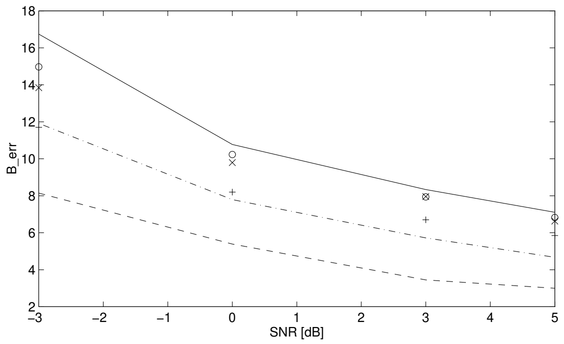


SIR=∞

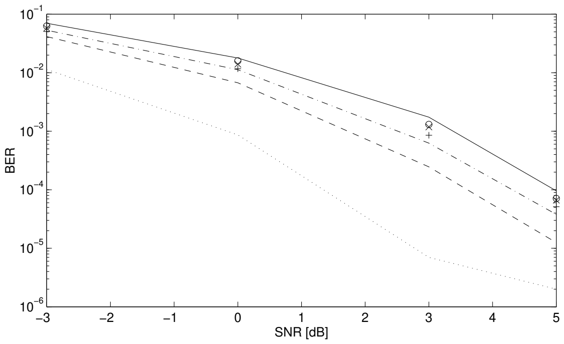


SIR=∞

Figure 2.21: Total mean square error in the channel matrix, B_err, and BER for the equalizer without co-channel interferers. Least squares method (solid line), CDEML with true number of DOA's and true DOA's as initial values (dashed), CDEML with one extra DOA per delay and true DOA's as initial values (dash-dotted), CDEML with true number of DOA's and estimated DOA's as initial values (+), CDEML with one extra DOA per delay and estimated DOA's as initial values (x), CDEML with four DOA's for each delay and estimated DOA's as initial values (o). BER of the equalizer tuned with a direct method using a very long training sequence (dotted).



SIR=0 dB



SIR=0 dB

Figure 2.22: Total mean square error in the channel matrix, B_err, and BER for the equalizer when SIR=0. Least squares method (solid line), CDEML with true number of DOA's and true DOA's as initial values (dashed), CDEML with one extra DOA per delay and true DOA's as initial values (dash-dotted), CDEML with true number of DOA's and estimated DOA's as initial values (+), CDEML with one extra DOA per delay and estimated DOA's as initial values (x), CDEML with four DOA's for each delay and estimated DOA's as initial values (o). BER of the equalizer tuned with a direct method using a very long training sequence (dotted).

2.5.4 Conclusions

It has been shown that when using an antenna array it is possible to improve the channel estimation by parametrizing the channel into directions of arrival of the signal paths, and their gains. The CDEML algorithm performs this in a computationally efficient way since the problem is decoupled into one minimization problem per tap in the channel. In the simulations performed in this study, it was also found that the CDEML algorithm handles the presence of co-channel interferers better than the least squares algorithm.

As seen from the simulations, some improvements can be achieved with the CDEML algorithm by using the correct or almost correct number of paths per delay. When the assumed number of paths is larger then the true number the performance gain is however reduced. It would therefore be of importance to use algorithms that can estimate the number of DOA's present. It has also been shown that it is of importance to have good initial estimates of the DOA's.

Unfortunately, neither the method suggested in [15] nor the simplified version presented in (2.70), does a good job in forming initial estimates of the DOA's. Unless better algorithms are used for finding initial estimates of the DOA's, the best strategy is probably to constrain the array geometry to a uniform linear array, and replace the nonlinear minimization with a polynomial root-finding technique similar to the one proposed in [94] and [95].

Although the CDEML algorithm had better performance than the algorithm based on the directly parametrized least squares FIR channel estimation, it should be noted that there is a non-negligible increase in the complexity and the CDEML algorithm will be less robust against mismodeling of the scenario.

Although there are improvements in the quality of the channel estimates, this improvement is *not* significantly reflected in the BER of the equalizer. This suggests that the CDEML algorithm improves the channel estimates in dimensions that are not so important for the equalization.

This method can be generalized to be used with other initial channel estimates than the least squares estimate of the coefficients in the FIR channel model. In [101] the method has been generalized to be used in combination with the temporal parametrization of Section 2.3. Is is important to note that the field of estimating directions of arrival is very big and many other methods exists for doing this. An overview of the field can be found in [46].

2.6 Reduced Rank Channel Estimation

Instead of attempting to utilize the spatial structure of a wireless space-time channel by parametrizing it in terms of directions of arrival and gains of signal paths as in Section 2.5 we will here take a directly parametrized approach. This will result in a much simpler method which also shows better performance improvements.

To describe the space-time channel from a transmitter to a multi-antenna receiver, several parameters are required. However, in many cases the temporal channels to different antenna elements will be correlated. One such case when this occurs is when a partial response signal is sent through a channel with very little delay spread. Since all intersymbol interference is caused by the modulation, the time dispersion experienced at the different antenna elements will be highly correlated. When this situation occurs, a *reduced rank* representation of the channel may be used.

The reduced rank property of a channel can be exploited in the channel estimation. One way is to estimate the space-time channel with a maximum likelihood method under the constraint that the resulting channel should be low rank [96]. Another method is to exploit the fact that the vector taps of the channel will lie in the subspace spanned by the signal eigenvectors to the spatial data covariance matrix [74]. We will call this subspace the *spatial signal subspace* or just the *signal subspace*. The channel estimate is here first formed as a straightforward least squares channel estimate. The vector taps in this channel estimate are then projected onto the spatial signal subspace. If co-channel interferers are present, the spatial signal subspace will be spanned by all signal components, desired as well as undesired.

We here propose a third method where the received signal samples are projected directly onto the spatial signal subspace defined above. As well as removing components in the noise subspace, this has the advantage that the dimension of the received signal vector is reduced. This turns out to give better performance as well as lower complexity, compared to the other two methods.

2.6.1 Channel Model

Using vector notation in analogy with (2.18), which modeled a single user, we can model the received desired signal from user i as

$$\boldsymbol{s}_i(t) = \boldsymbol{B}_i \boldsymbol{d}_i(t) \tag{2.72}$$

where $\boldsymbol{B}_i$ is the $M \times nb_i$ channel matrix for user i and $\boldsymbol{d}_i$ contains delayed versions of the symbols transmitted by user i

$$\boldsymbol{d}_i(t) = [d_i(t)\ d_i(t-1)\ \ldots\ d_i(t-nb_i)]^T . \tag{2.73}$$

Including co-channel interferers in the channel model, the vector signal $\boldsymbol{y}(t)$ measured at the antenna can thus be described as

$$\boldsymbol{y}(t) = \mathbf{s}_0(t) + \sum_{k=1}^{N_{co}} \mathbf{s}_k(t) + \boldsymbol{n}(t) = \boldsymbol{s}(t) + \boldsymbol{n}(t) \tag{2.74}$$

$$= \boldsymbol{B}_0 \boldsymbol{d}_0(t) + \sum_{k=1}^{N_{co}} \boldsymbol{B}_k \boldsymbol{d}_k(t) + \boldsymbol{n}(t)\ . \tag{2.75}$$

We are interested in the signal from user 0, whereas the signals from users 1 to N_{co} constitute co-channel interference. Furthermore, the term $\boldsymbol{n}(t)$ constitutes noise, which here is assumed to be spatially and temporally white.

In general, the channel matrix for user i, $\boldsymbol{B}_i$, has rank $\max(M, nb_i + 1)$. However, in some cases the channel matrices loose rank. When a channel is represented by a rank $N < \max(M, nb_i + 1)$ channel matrix, we call it a *reduced* or *low* rank channel. We will now outline in what scenarios such channels may occur.

Consider a multipath propagation model with K "paths" for one of the channel matrices $\boldsymbol{B}_j$ in (2.75), here simply called $\boldsymbol{B}$

$$\boldsymbol{B} = \sum_{k=1}^{K} \boldsymbol{a}(\theta_k)\boldsymbol{p}^T(\tau_k). \tag{2.76}$$

Here $\boldsymbol{a}(\theta_k)$ is the array response for the desired signal traveling along path k and arriving from direction θ_k. The column vector $\boldsymbol{p}(\tau_k)$ contains a sampled version of the transmitter pulse shaping function $p(t)$ with the sampling offset determined by the relative path delay τ_k

$$\boldsymbol{p}(\tau_k) \ \stackrel{\triangle}{=} \ \begin{bmatrix} p(\tau_k) & p(\tau_k - T) & \dots & p(\tau_k - nb_0T) \end{bmatrix}^T. \tag{2.77}$$

If there is no delay spread in the channel, i.e. all $\tau_k = \tau, \ \forall k$, then all $\boldsymbol{p}(\tau_k)$ will be equal, say $\boldsymbol{p}(\tau_k) = \boldsymbol{p}_0, \ \forall k$, and the channel matrix can be written as

$$\boldsymbol{B} = \left(\sum_{k=1}^{K} \boldsymbol{a}(\theta_k) \right) \boldsymbol{p}_0^T = \boldsymbol{a}\boldsymbol{p}_0^T \tag{2.78}$$

where the column vector $\boldsymbol{a}$ is given by

$$\boldsymbol{a} = \sum_{k=1}^{K} \boldsymbol{a}(\theta_k). \tag{2.79}$$

The channel matrix thus has rank one. If there is some delay spread but it is small, so that

$$\boldsymbol{p}(\tau_k) \approx \boldsymbol{p}_0, \ \forall k$$

then the channel matrix will have a rank which is approximately one. To obtain a channel matrix with an approximate rank larger than one we thus need a "significant" delay spread.

Let us now assume that the paths can be grouped such that the paths within each group have similar propagation delays. Furthermore, assume that the spatial signatures for the paths in different groups are different. The approximate rank of the channel model would then be determined by the number of such groups with significant energy. If the number of such groups is small, then the channel matrix can be approximated by a low rank matrix.

In the following two sections we present and compare three methods for exploiting, the low rank property of the channel matrices in the channel estimation.

2.6.2 Maximum-Likelihood Reduced-Rank Channel Estimation

Assume that the channel matrix $\boldsymbol{B}$ can be decomposed as

$$\boldsymbol{B} = \boldsymbol{A}\boldsymbol{P} \tag{2.80}$$

where $\boldsymbol{A}$ is an $M \times r$ matrix and $\boldsymbol{P}$ is an $r \times (nb+1)$ matrix. The channel matrix then has a maximum rank of r.

If the noise vector $\boldsymbol{n}(t)$ is temporally white and Gaussian distributed, then the maximum likelihood rank r estimate of $\boldsymbol{B}$ can be found by generalizing the result in [96] to complex-valued signals and models, as

$$\hat{\boldsymbol{B}}_{ML} = \hat{\boldsymbol{R}}_{\boldsymbol{yd}}\hat{\boldsymbol{R}}_{\boldsymbol{dd}}^{-1/2}\hat{\boldsymbol{S}}\hat{\boldsymbol{S}}^{H}\hat{\boldsymbol{R}}_{\boldsymbol{dd}}^{-1/2} \tag{2.81}$$

where

$$\hat{\boldsymbol{R}}_{\boldsymbol{yd}} = \frac{1}{t_{max} - t_{min} + 1}\sum_{t=t_{min}}^{t_{max}} \boldsymbol{y}(t)\boldsymbol{d}^{H}(t) \tag{2.82}$$

$$\hat{\boldsymbol{R}}_{\boldsymbol{dd}} = \frac{1}{t_{max} - t_{min} + 1}\sum_{t=t_{min}}^{t_{max}} \boldsymbol{d}(t)\boldsymbol{d}^{H}(t) \tag{2.83}$$

and t_{min} and t_{max} are the indices of the first and last sample utilized. Furthermore, the matrix $\hat{\boldsymbol{S}}$ is defined as

$$\hat{\boldsymbol{S}} \triangleq \begin{bmatrix}\hat{\boldsymbol{v}}_1 & \dots & \hat{\boldsymbol{v}}_r\end{bmatrix} \tag{2.84}$$

where $\hat{\boldsymbol{v}}_1, \dots, \hat{\boldsymbol{v}}_r$ are the r dominant eigenvectors of the matrix

$$\hat{\boldsymbol{V}} = \hat{\boldsymbol{R}}_{\boldsymbol{dd}}^{-1/2}\hat{\boldsymbol{R}}_{\boldsymbol{yd}}^{H}\hat{\boldsymbol{R}}_{\boldsymbol{yy}}^{-1}\hat{\boldsymbol{R}}_{\boldsymbol{yd}}\hat{\boldsymbol{R}}_{\boldsymbol{dd}}^{-1/2} \tag{2.85}$$

where $\hat{\boldsymbol{R}}_{\boldsymbol{yd}}$ and $\hat{\boldsymbol{R}}_{\boldsymbol{dd}}$ are given above and

$$\hat{\boldsymbol{R}}_{\boldsymbol{yy}} = \frac{1}{t_{max} - t_{min} + 1}\sum_{t=t_{min}}^{t_{max}} \boldsymbol{y}(t)\boldsymbol{y}^{H}(t). \tag{2.86}$$

This estimate will have a rank no larger than r. An advantage with this estimation algorithm is that it does not require the noise to be spatially white since no assumption about the color of the noise is made. Strong spatial color of the noise term induced by strong co-channel interferers can thus be handled. The temporal noise color is however not accounted for since the noise is assumed to be temporally white.

2.6.3 Signal Subspace Projection

The maximum likelihood reduced rank channel estimation makes no assumption about the spatial color of the noise $\boldsymbol{n}(t)$ and thus does not use this information in the channel estimation. If the noise $\boldsymbol{n}(t)$ is spatially white it may however be possible to utilize this fact to improve the performance.

We will now use (2.74) to divide the received signal into a signal part and a noise part. Since the signals $d_i(t)$ are assumed to be uncorrelated with the noise $\boldsymbol{n}(t)$, we can decompose the covariance matrix of the received signal as

$$E[\boldsymbol{y}(t)\boldsymbol{y}^H(t)] = \boldsymbol{R_{yy}} = \boldsymbol{R_{ss}} + \boldsymbol{R_{nn}} \tag{2.87}$$

where

$$\boldsymbol{R_{ss}} = E\left[\boldsymbol{s}(t)\boldsymbol{s}^H(t)\right] \qquad \boldsymbol{R_{nn}} = E\left[\boldsymbol{n}(t)\boldsymbol{n}^H(t)\right]. \tag{2.88}$$

We now make the critical assumption that the noise vector $\boldsymbol{n}(t)$ consists only of white noise with variance σ_n^2:

$$\boldsymbol{R_{nn}} = \sigma_n^2 \boldsymbol{I}.$$

The M-dimensional space containing the received signal vectors can now be divided into two subspaces: the *signal subspace* and the *noise subspace*. The signal subspace is the subspace spanned by the eigenvectors of the signal covariance matrix $\boldsymbol{R_{ss}}$

$$\textit{Signal subspace} = \operatorname{span}(\boldsymbol{v}_1^s, \dots, \boldsymbol{v}_r^s) \tag{2.89}$$

where $\boldsymbol{v}_1^s, \dots, \boldsymbol{v}_r^s$ are the r dominant eigenvectors of the signal covariance matrix. The noise subspace is the orthogonal complement of the signal subspace

$$\textit{Noise subspace} = \textit{Signal subspace}^{\perp}. \tag{2.90}$$

We note that an eigenvector $\boldsymbol{v}_i$ of the signal covariance matrix $\boldsymbol{R_{ss}}$ with eigenvalue λ_i^s is also an eigenvector to the received signal covariance matrix $\boldsymbol{R_{yy}}$ but with eigenvalue $\lambda_i^s + \sigma_n^2$ since

$$\boldsymbol{R_{yy}}\boldsymbol{v}_i^s = \left(\boldsymbol{R_{ss}} + \sigma_n^2\boldsymbol{I}\right)\boldsymbol{v}_i^s = \lambda_i^s\boldsymbol{v}_i^s + \sigma_n^2\boldsymbol{v}_i^s = (\lambda_i^s + \sigma_n^2)\boldsymbol{v}_i^s. \tag{2.91}$$

Furthermore, since the vectors in the noise subspace are orthogonal to the vectors in the signal subspace, they are orthogonal to all columns in the signal covariance matrix. Every vector in the noise subspace is then an eigenvector to $\boldsymbol{R_{yy}}$ with eigenvalue equal to σ_n^2, since for any vector $\boldsymbol{v}^n$ in the noise subspace

$$\boldsymbol{R_{yy}}\boldsymbol{v}^n = (\boldsymbol{R_{ss}} + \sigma_n^2 \boldsymbol{I})\boldsymbol{v}^n = \sigma_n^2 \boldsymbol{v}^n . \tag{2.92}$$

The noise subspace will thus be spanned by the eigenvectors of $\boldsymbol{R_{yy}}$ with eigenvalues equal to σ_n^2 and the signal subspace will be spanned by the eigenvectors of $\boldsymbol{R_{yy}}$ with eigenvalues strictly greater than σ_n^2. A base of vectors spanning the signal subspace can thus be constructed by selecting the eigenvectors of $\boldsymbol{R_{yy}}$ with eigenvalues above the noise level σ_n^2.

Using (2.74) and (2.75), the signal part of the spatial covariance matrix can now be expressed as

$$\boldsymbol{R_{ss}} = \boldsymbol{B}_0 \boldsymbol{B}_0^H + \sum_{i=1}^{N_{co}} \boldsymbol{B}_i \boldsymbol{B}_i^H \ . \tag{2.93}$$

given that all users transmit uncorrelated signal sequences.

When any of the channel matrices has full rank, or when many co-channel interferers are present, the rank of $\boldsymbol{R_{ss}}$ will be full. However, when the channel of the desired user has low rank, and when only a few dominant co-channel interferers with low rank channels are present, $\boldsymbol{R_{ss}}$ may loose rank.

Channel Signal-Subspace Projection

As noted in [74], the standard least squares channel estimate $\hat{\boldsymbol{B}}_{LS}$ (2.21) can be improved by exploiting that all columns of the true channel matrix $\boldsymbol{B}$ lie in the signal subspace (2.89). To utilize this property, we will project the least squares estimate $\hat{\boldsymbol{B}}_{LS}$ onto an estimate of the signal subspace:

$$\hat{\boldsymbol{V}}_s \triangleq \begin{bmatrix} \hat{\boldsymbol{v}}_1^s & \dots & \hat{\boldsymbol{v}}_r^s \end{bmatrix} , \tag{2.94}$$

where $\hat{\boldsymbol{v}}_i^s,\ i = 1, \ldots, r$ are the r largest eigenvectors to the sample covariance matrix

$$\hat{\boldsymbol{R}}_{\boldsymbol{yy}} \triangleq \frac{1}{t_{max} - t_{min} + 1} \sum_{t=t_{min}}^{t_{max}} \boldsymbol{y}(t)\boldsymbol{y}^H(t). \tag{2.95}$$

Note that a known training sequence of transmitted symbols is not needed to estimate $\boldsymbol{R}_{\boldsymbol{yy}}$. The estimate of the signal subspace can thus (in a TDMA system) be estimated based on the whole frame of received samples.

The resulting estimate of the channel will then be given by

$$\hat{\boldsymbol{B}}_{ss} = \hat{\boldsymbol{V}}_s \hat{\boldsymbol{V}}_s^H \hat{\boldsymbol{B}}_{LS}. \tag{2.96}$$

This projection will remove components outside the (estimated) signal subspace. If the signal subspace is reasonably well estimated this will improve the channel estimate as some noise-induced estimation errors will be removed. This method, which was suggested in [74], will here be called *channel signal-subspace projection.*

It may be difficult to determine where the noise level is, i.e. to choose at which eigenvalue level to consider an eigenvector to be a signal eigenvector or a noise eigenvector. In practice one will have to adopt a suitable criterion for determining this level.

It should be pointed out that we are not restricted to using a directly parametrized FIR channel estimate as the initial channel estimate. We can in fact choose from many different channel estimation methods. The fact that the columns of the channel vector lies in the spatial signal subspace is an inherent property of the channel model (2.75) and not of the estimation method.

Data Signal-Subspace Projection

We can also project the received signal onto the estimated signal subspace (2.94) *before* any other processing is performed. We thus form the new data vectors of dimension $(r|1)$

$$\boldsymbol{y}_{ss}(t) = \hat{\boldsymbol{V}}_s^H \boldsymbol{y}(t) \tag{2.97}$$

and estimate the channel and noise plus interference covariance matrix of this new data set. The channel estimate will be exactly the same as for the channel signal subspace projection method (after re-transformation to the full space) but the quality of the estimate of the noise-plus-interference spatial covariance matrix is improved.

Another very important feature of this method is that since the dimension of the projected signal vector, $\boldsymbol{y}_{ss}(t)$ $(r|1)$, is lower than the dimension of the original signal vector $\boldsymbol{y}(t)$ $(M|1)$, all processing including the equalizer tuning and execution is reduced in complexity.

A requirement for both of these methods is of course that the number of antennas, M, is strictly greater than the rank, r, of the channel. The larger the difference between the number of antennas and the rank of the channel, the better. A larger difference means a larger noise subspace. More components of the estimated channel or the received signal created by noise only are then removed by the projection.

It may not be necessary to re-estimate the signal subspace for each time instant the channel is estimated. If we have a fading multipath channel where the gains of the paths vary rapidly while the directions of arrival of the paths vary slowly. Then, as pointed out in [30], the subspace to which the channel belongs, and thus the signal, will only vary slowly. By estimating this subspace we can potentially reuse it when estimating new channels, for instance in following frames in a TDMA system.

A disadvantage with these methods compared to the maximum likelihood reduced rank channel estimation of Section 2.6.2 is that signal subspace projection cannot remove estimation errors due to the co-channel interference. Without using the training sequence we cannot distinguish between the desired signal and the co-channel interferers. The signal subspace will be spanned by the combined set of eigenvectors from the desired signal spatial covariance matrix and the co-channel interferer spatial covariance matrices. However it is not always important to remove the errors in the channel estimate caused by the co-channel interferers. In fact these errors are sometimes automatically compensated for by the equalizer as is discussed in Chapter 7.

2.6.4 Estimation of the Spatial Noise-plus-Interference Covariance Matrix

Another important quantity of the channel for space-time equalization is the spatial noise plus interferer covariance matrix,

$$\boldsymbol{R_{nn}} = E[(\sum_{i=1}^{K} \boldsymbol{B}_i \boldsymbol{d}_i(t) + \boldsymbol{n}(t))(\sum_{i=1}^{K} \boldsymbol{B}_i \boldsymbol{d}_i(t) + \boldsymbol{n}(t))^H], \tag{2.98}$$

where the expectation is taken with respect to $\boldsymbol{d}_i(t)$ and $\boldsymbol{n}(t)$. In all the three methods described above, this matrix is estimated from the residuals of the channel estimation. An improved channel estimate will thus indirectly also result in an improved noise plus interference covariance matrix estimate. Note however, that for the here proposed data signal-subspace projection method, the dimension of the covariance matrix is r, whereas it for the other two methods has dimension M.

2.6.5 Simulations

For the simulations we used a circular array with radius 1.25 λ and ten antenna elements. We simulated a scenario where the desired signal arrives along two paths at angles 0 and 45 degrees. The channels (i.e. pulse shaping) in the two paths were $0.44 + q^{-1} + 0.44q^{-2}$ and $0.44q^{-1} + q^{-2} + 0.44q^{-3}$ respectively. Two different co-channel interferers impinged on the array from directions -45 and 90 degrees, each having the temporal channel $0.44 + q^{-1} + 0.44q^{-2}$. Spatially and temporally white noise was added giving a signal to noise ratio of 3dB. The transmitted signal was BPSK modulated. A multi-channel MLSE, see Chapter 4, was used for the equalization.

We applied the three methods discussed in Sections 2.6.2 and 2.6.3 this to this scenario. For the maximum likelihood reduced rank estimation, a rank two channel was estimated, whereas for the two projection algorithms, the respective quantities were projected onto a rank four estimate of the signal subspace [8].

[8] We only used the data received during the training sequence to estimate the signal subspace to obtain a fair comparison with the maximum likelihood reduced rank channel estimation method.

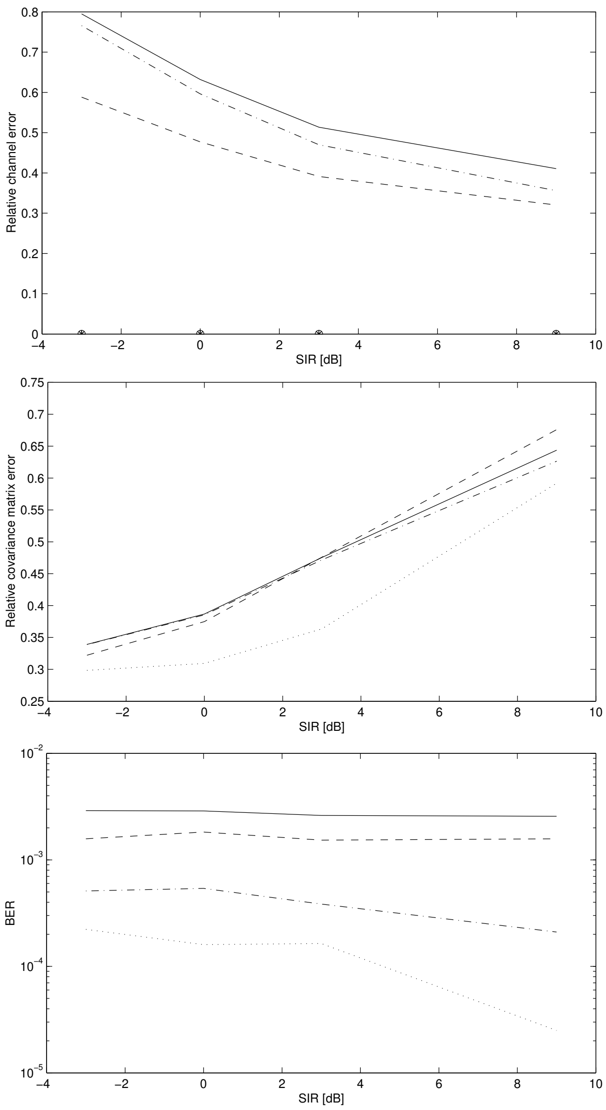


Figure 2.23: Relative channel error, relative error in the spatial covariance matrix for the noise plus interference and BER as a function of SIR with a SNR of 3dB using the FIR least squares (solid), maximum likelihood reduced rank method (dashed), channel signal subspace projection method (dash-dotted) and data signal subspace projection method (dotted). Only the relative covariance matrix error and the BER is shown for the data signal subspace projection method. The relative channel error for this method is the same as for the channel signal subspace projection method.

In Figure 2.23 the relative channel estimation error and relative covariance matrix errors as well as the resulting BER for the equalizer can be seen. Note that the maximum likelihood reduced rank estimation method (MLRR) has the smallest relative channel error. However, the BER of the detector based on the model estimated with MLRR is not the lowest. On the other hand, both the channel signal subspace projection method (CSSP) and the data signal subspace projection method (DSSP) have significant improvements in terms of equalizer BER even though their relative channel estimation errors are larger than for the MLRR method. Note that the DSSP method has the best noise plus interferer spatial covariance matrix estimate and the best BER performance.

2.6.6 Summary

We have here considered different methods of exploiting the reduced rank property of a space-time channel in wireless communications. Three methods have been studied, a maximum likelihood reduced rank channel estimation method which finds the channel of a given rank in a maximum likelihood sense [96], and two signal subspace projection methods which projects either the channel estimate [74] or, as for the here proposed method, the received signal samples onto an estimate of the spatial signal subspace.

Even though the model estimated using the maximum likelihood reduced rank method provides the best accuracy, a detector based on this model does not achieve the lowest BER. The BER from detectors based on the signal subspace projection methods are lower, despite the fact that the accuracy of the estimated channel models is worse. Apparently, the signal subspace projection methods find channel estimates that are better suited for the purpose of space-time equalization.

The here proposed method of projecting the received signal samples directly on the signal subspace gives the best equalizer performance. This improvement is contributed to *a better estimated spatial noise plus interference covariance matrix*. Since this method also has lower complexity than the channel signal subspace projection method, it is clearly preferable.

2.7 Bootstrap Channel Estimation

Consider a TDMA system with the data transmitted in frames containing a short training sequence. When performing space-time equalization we benefit from good estimates of the channel to the desired user and good estimate of the spatial spectrum of the interferers. With a short training sequence it may be difficult to achieve good estimates. However, as will be seen in Figure 3.8, the performance of an equalizer can sometimes be considerably improved by using a longer training sequence. For the indirectly tuned DFE this is due to improved channel and interferer spectrum estimates. It is however generally not desirable to increase the length of the training sequence since this will reduce the efficiency of the information transmission. Also, new equalizers and multi-antenna receivers may be desirable in present standards, where the length of training sequence cannot be altered. The channel estimates and thus the equalizer performance can however be improved by using a bootstrap approach.

The bootstrap approach can briefly be explained as follows:

First an initial estimate of the channel and the interferer spectrum is formed using the training data. Using this channel information an equalizer is computed. This equalizer is then used to obtain estimates of *all* symbols in the frame. Using all the so obtained symbol estimates, a new (hopefully improved) channel estimate is computed. Based on this channel estimate a new equalizer is computed which in turn can be used to form yet new symbol estimates. This procedure can now be repeated as many times as desired.

The bootstrap channel estimating procedure creates an artificial long training sequence. This sequence will, of course, have errors in it which will degrade the performance of the channel estimates. However, if the number of erroneous symbols are not too many the improvement of the longer training sequence will hopefully outweigh the degradation caused by the incorrect artificial training symbols.

Bootstrap channel estimation will of course require the channel to be approximately stationary in order to make use of the longer artificial training sequence.

We here presents simulations illustrating the bootstrap principle. We here use a receiver with 8 antennas. We here used a receiver with 8 antennas. The antennas were considered to be spaced far enough apart such that the channels could be considered uncorrelated. The channel to each antenna consisted of 5 fading taps of equal average power with time delays 0, 0.25, 0.5, 0.75 and 1.00 symbol intervals, respectively. The total channel was created by applying GMSK modulation with a BT product of 0.3, as used in the GSM standard, to each of the taps. A co-channel interferer with the same channel characteristics as the desired signal was added along with Gaussian noise.

The channel was estimated without bootstrapping using 26 training symbols and with bootstrapping with one, two and 5 iterations. The number of extra symbols used in the bootstrapping was 122, giving a total of 148 symbols in the re-estimation of the channel in the bootstrap algorithm. A genie aided version of the channel estimation was also performed. Here it was assumed that all 148 symbols where known and used in the channel estimation.

The relative errors in the estimation of the channel and the noise plus interferer covariance estimates were computed using the Frobenius norm, i.e.

$$\text{Relative channel error} = \frac{\| \hat{\boldsymbol{B}} - \boldsymbol{B} \|_2}{\| \boldsymbol{B} \|_2} \tag{2.99}$$

and

$$\text{Relative covariance error} = \frac{\| \hat{\boldsymbol{R}}_{\boldsymbol{nn}} - \boldsymbol{R}_{\boldsymbol{nn}} \|_2}{\| \boldsymbol{R}_{\boldsymbol{nn}} \|_2} \tag{2.100}$$

where $\hat{\boldsymbol{B}}$ and $\boldsymbol{B}$ are the estimated and true channel matrices and $\hat{\boldsymbol{R}}_{\boldsymbol{nn}}$ and $\boldsymbol{R}_{\boldsymbol{nn}}$ are the estimated and true spatial noise plus interference co-variance matrices, respectively. The notation $\| \bullet \|_2$ represents the Frobenius norm.

In Figures 2.24 and 2.25 we can see how the relative errors of the channel estimate and the estimated covariance matrix of the noise plus interference is reduced by the bootstrapping. In this simulation it is apparent that the major part of the improvement is achieved after the first re-estimation and re-tuning of the equalizer using the initially estimated symbols.

These improvements in estimation of the channel improves the BER of the equalizer, as can be seen in Figure 2.26. In this simulation the bootstrapped

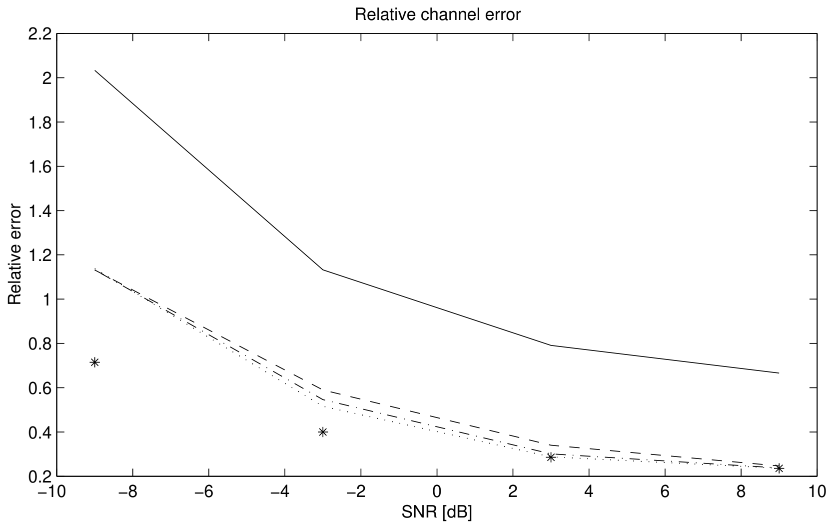


Figure 2.24: Relative channel estimation error as a function of SNR. SIR=0dB. No bootstrapping (solid), bootstrapping with one iteration (dashed), two iterations (dash-dotted), 5 iterations (dotted) and genie aided (*) (using 148 correct symbols).

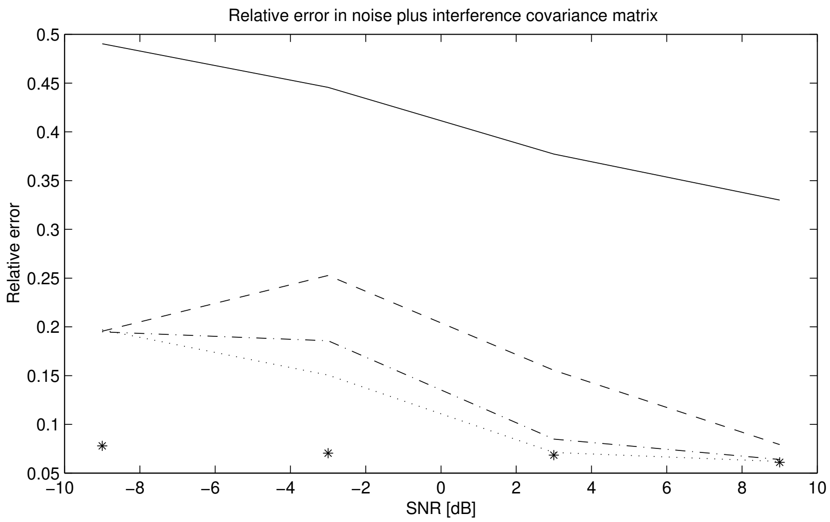


Figure 2.25: Relative error in the spatial covariance matrix as a function of SNR. SIR=0dB. No bootstrapping (solid), bootstrapping with one iteration (dashed), two iterations (dash-dotted), 5 iterations (dotted) and genie aided (*) (using 148 correct symbols).

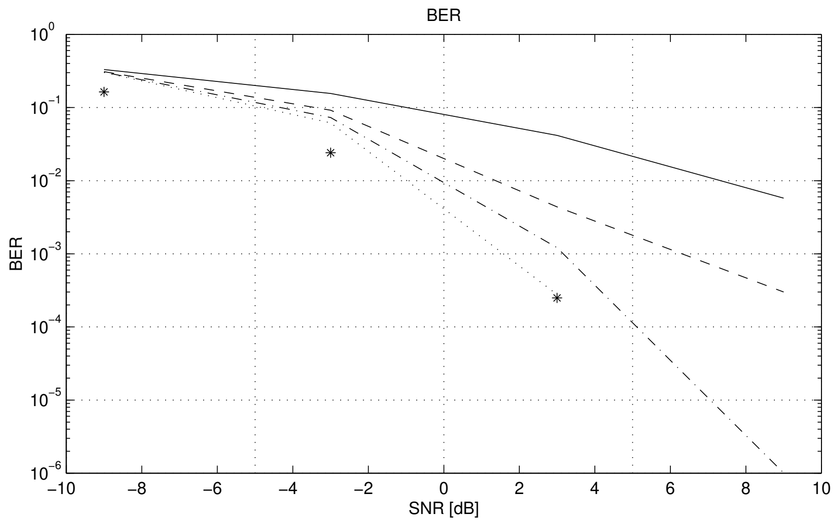


Figure 2.26: BER as a function of SNR. SIR=0dB. No bootstrapping (solid), bootstrapping with one iteration (dashed), two iterations (dash-dotted), 5 iterations (dotted) and genie aided (*) (using 148 correct symbols).

equalizer achieves almost the same performance as the genie-aided equalizer trained using all 148 symbols. This will however not always be the case. We have noted that if the initial equalization results in a too high BER (say larger than about 10-20 percent), then bootstrapping may not be able to improve the performance of the equalizer.

2.8 Estimation of Noise plus Interference MA Spectrum

When tuning an equalizer indirectly we need an estimate of the spectrum of the noise plus interference. Let us consider the model of the noise plus interference term, $\boldsymbol{n}(t)$, in (2.5)

$$\boldsymbol{n}(t) = \sum_{k=1}^{K} \boldsymbol{b}_k(q^{-1}) d_k(t) + \boldsymbol{v}(t). \tag{2.101}$$

If we assume that the thermal noise component, $\boldsymbol{v}(t)$, is temporally white then we can model the noise and interference with an MA process as

$$\boldsymbol{n}(t) = \boldsymbol{M}(q^{-1})\boldsymbol{v}'(t) \tag{2.102}$$

where $\boldsymbol{M}(q^{-1})$ is an $M \times M$ polynomial matrix of order nm

$$\boldsymbol{M}(q^{-1}) = \begin{bmatrix} m_{11}(q^{-1}) & \dots & m_{1M}(q^{-1}) \\ \vdots & \ddots & \vdots \\ m_{M1}(q^{-1}) & \dots & m_{MM}(q^{-1}) \end{bmatrix} \tag{2.103}$$

with

$$m_{ij}(q^{-1}) = m_{ij,0} + m_{ij,1}q^{-1} + \ldots + m_{ij,nm}q^{-nm} \tag{2.104}$$

and $\boldsymbol{v}'(t)$ is the temporally and spatially white innovations vector for the noise process, i.e.

$$E[\boldsymbol{v}'(t_1)\boldsymbol{v}'^H(t_2)] = \delta_{t_1t_2}\boldsymbol{I}. \tag{2.105}$$

The spectrum for the interference plus noise term, $\boldsymbol{n}(t)$, can now be expressed as

$$\boldsymbol{R_{nn}}(z, z^{-1})) = \boldsymbol{M}(z^{-1})\boldsymbol{M}^H(z). \tag{2.106}$$

This spectrum is a complex conjugate symmetric double sided polynomial in z and z^{-1}, i.e.

$$\boldsymbol{R_{nn}}(z, z^{-1})) = \boldsymbol{R}^H_{\boldsymbol{nn}}(nm)z^{nm} + \ldots + \boldsymbol{R_{nn}}(0) + \ldots + \boldsymbol{R_{nn}}(nm)z^{-nm} \tag{2.107}$$

where $\boldsymbol{R_{nn}}(k)$ is the autocovariance for the noise plus interferer spectrum with lag k

$$\boldsymbol{R_{nn}}(k) = E[\boldsymbol{n}(t)\boldsymbol{n}^H(t-k)]. \tag{2.108}$$

Using the residuals from the estimation of the channel for the desired signal

$$\hat{\boldsymbol{n}}(t) = \boldsymbol{y}(t) - \hat{\boldsymbol{b}}(q^{-1})d(t) \tag{2.109}$$

we can form sample matrix estimates of the autocovariance for the noise plus interference for different lags as

$$\hat{\boldsymbol{R}}_{\boldsymbol{nn}}(k) = \frac{1}{t_{max} - t_{min} + 1} \sum_{t=t_{min}}^{t_{max}} \hat{\boldsymbol{n}}(t)\hat{\boldsymbol{n}}^H(t-k) \tag{2.110}$$

where the summation borders t_{max} and t_{min} are determined by how many residuals can be formed from the training data.

A problem with estimating the spectrum of the noise plus interference spectrum is that usually only a short training sequence is available and thus the number of residuals is small. If the receiver has many antenna elements or many diversity branches of other sorts, then the number of coefficients in the spatio-temporal spectrum of the noise plus interference may become prohibitingly high. In this case we can choose to only estimate and use the spatial spectrum of the noise plus interference, i.e. we can estimate $\boldsymbol{R_{nn}}(0)$ as

$$\hat{\boldsymbol{R}}_{\boldsymbol{nn}}(0) = \frac{1}{t_{max} - t_{min} + 1} \sum_{t=t_{min}}^{t_{max}} \hat{\boldsymbol{n}}(t)\hat{\boldsymbol{n}}^H(t). \tag{2.111}$$

When using many antenna elements, the tuning of an equalizer using only the spatial spectrum of the noise plus interference is often more well behaved than what is obtained by using also the temporal spectrum of the interference. The price we pay for this restriction is that our resulting equalizer will only suppress the interference spatially.

With a sufficiently long training sequence it is, of course, possible to take some aspects of the temporal spectrum of the interference into consideration. We can, for example, choose to model the spectrum with only a few non-zero time lags of the autocovariance for the noise plus interference. When we do this it is important that the equalizer has a proper structure for space-time suppression of interference, see Section 3.2.3 in Chapter 3 and see Chapter 4.

2.9 Joint FIR Channel and AR Noise Model Estimation

2.9.1 Joint FIR Channel and AR Noise Model Estimation

As mentioned in Section 2.8, estimating the spatial spectrum using the sample matrix estimates as in (3.90), will often prove difficult unless the training sequence is long. Instead of modeling the noise plus interference with an MA model as described in Section 2.8 we can choose to model the noise plus interference with an AR model [7]. The model for the received vector of signals, $\boldsymbol{y}(t)$, then becomes

$$\boldsymbol{y}(t) = \boldsymbol{b}(q^{-1})d(t) + \boldsymbol{N}^{-1}(q^{-1})\boldsymbol{M}_0\boldsymbol{v}(t) \tag{2.112}$$

where, as in (2.15), $\boldsymbol{b}(q^{-1})$ is a polynomial column vector containing the polynomial channels $b_i(q^{-1})$, of degree nb, to the individual antennas:

$$\boldsymbol{b}(q^{-1}) = [b_1(q^{-1})\ b_2(q^{-1})\ \dots\ b_M(q^{-1})]^T. \tag{2.113}$$

The matrix $\boldsymbol{M}_0$ is a non-singular constant $M \times M$ matrix

$$\boldsymbol{M}_0 = \begin{bmatrix} m_{11} & \dots & m_{1M} \\ \vdots & \ddots & \vdots \\ m_{M1} & \dots & m_{MM} \end{bmatrix} \tag{2.114}$$

and $\boldsymbol{N}(q^{-1})$ is a stably invertible [9] $M \times M$ polynomial matrix

$$\boldsymbol{N}(q^{-1}) = \begin{bmatrix} n_{11}(q^{-1}) & \dots & n_{1M}(q^{-1}) \\ \vdots & \ddots & \vdots \\ n_{M1}(q^{-1}) & \dots & n_{MM}(q^{-1}) \end{bmatrix} \tag{2.115}$$

with the components

$$n_{ij}(q^{-1}) = n_{ij,0} + n_{ij,1}q^{-1} + \dots + n_{ij,nn}q^{-nn}. \tag{2.116}$$

We assume that its leading matrix coefficient of $\boldsymbol{N}(q^{-1})$ is equal to the identity matrix, i.e. $\boldsymbol{N}_0 = \boldsymbol{I}$.

The $M \times 1$ vector $\boldsymbol{v}(t)$ is a temporally and spatially white innovations sequence with covariance matrix equal to unity, i.e.

$$E[\boldsymbol{v}(t_1)\boldsymbol{v}^H(t_2)] = \delta_{t_1 t_2}\boldsymbol{I}. \tag{2.117}$$

Multiplying both sides of (2.112) from the left with $\boldsymbol{N}(q^{-1})$ gives

$$\boldsymbol{N}(q^{-1})\boldsymbol{y}(t) = \boldsymbol{N}(q^{-1})\boldsymbol{b}(q^{-1})d(t) + \boldsymbol{M}_0\boldsymbol{v}(t). \tag{2.118}$$

[9] The determinant of $\boldsymbol{N}(q^{-1})$ has all zeros strictly inside the unit circle.

Let us now introduce the polynomial vector

$$\boldsymbol{b}_{\boldsymbol{N}}(q^{-1}) = \boldsymbol{N}(q^{-1})\boldsymbol{b}(q^{-1}) \tag{2.119}$$

and the polynomial matrix

$$\boldsymbol{N}^1(q^{-1}) = \boldsymbol{N}_1 + \ldots + \boldsymbol{N}_{nn}q^{-(nn-1)}. \tag{2.120}$$

We can now rewrite (2.118) as

$$\boldsymbol{y}(t) = \boldsymbol{b}_{\boldsymbol{N}}(q^{-1})d(t) - \boldsymbol{N}^1(q^{-1})\boldsymbol{y}(t-1) + \boldsymbol{M}_0\boldsymbol{v}(t). \tag{2.121}$$

Equation (2.121) can be handled as an ordinary regression problem where $d(t)$ and $\boldsymbol{y}(t-1)$, and their lagged values, are the regression variables. Note that the error term, $\boldsymbol{M}_0\boldsymbol{v}(t)$, is temporally white and uncorrelated with the regression variables. A standard least squares method can now be applied to compute estimates, $\hat{\boldsymbol{N}}^1(q^{-1})$ and $\hat{\boldsymbol{b}}_{\boldsymbol{N}}(q^{-1})$, $\boldsymbol{N}^1(q^{-1})$ and $\boldsymbol{b}_{\boldsymbol{N}}(q^{-1})$, respectively [87]. The equations for this least squares estimation is outlined in Appendix 2.B. The estimate of the noise model denominator, $\boldsymbol{N}(q^{-1})$, can now be formed as

$$\hat{\boldsymbol{N}}(q^{-1}) = \boldsymbol{I} + q^{-1}\hat{\boldsymbol{N}}^1(q^{-1}) \tag{2.122}$$

The constant matrix $\boldsymbol{M}_0$ can be estimated as the square root of the sample matrix estimate $\hat{\boldsymbol{R}}_{\hat{\boldsymbol{r}}\hat{\boldsymbol{r}}}$ formed as

$$\hat{\boldsymbol{R}}_{\hat{\boldsymbol{r}}\hat{\boldsymbol{r}}} = \frac{1}{t_{max} - t_{min} + 1} \sum_{t=t_{min}}^{t_{max}} \hat{\boldsymbol{r}}(t)\hat{\boldsymbol{r}}^H(t) \tag{2.123}$$

where $\hat{\boldsymbol{r}}(t)$ are the residuals

$$\hat{\boldsymbol{r}}(t) = \hat{\boldsymbol{N}}(q^{-1})\boldsymbol{y}(t) - \hat{\boldsymbol{b}}_{\boldsymbol{N}}(q^{-1})d(t). \tag{2.124}$$

The estimate of the matrix $\boldsymbol{M}_0$ can thus be expressed as

$$\hat{\boldsymbol{M}}_0 = \hat{\boldsymbol{R}}_{\hat{\boldsymbol{r}}\hat{\boldsymbol{r}}}^{1/2}. \tag{2.125}$$

Note however, that often we don't need to actually compute this estimate as we often will be using the product $\boldsymbol{M}_0\boldsymbol{M}_0^H$ which can be estimated directly with $\hat{\boldsymbol{R}}_{\hat{\boldsymbol{r}}\hat{\boldsymbol{r}}}$.

A necessary condition to be able to compute the estimates of $\boldsymbol{N}^1(q^{-1})$ and $\boldsymbol{b_N}(q^{-1})$ is that the number of parameters to estimate is less than the number of equations, i.e. we must require

$$M^2 nn + M(nb + nn + 1) < MN_{eff} \tag{2.126}$$

where N_{eff} is the *effective* length of the training sequence, i.e. the number of training relations we can form by taking the temporal length of the channels into consideration.

However, to obtain a reasonable estimate of the matrix $\boldsymbol{M}_0$, or the matrix $\boldsymbol{M}_0\boldsymbol{M}_0^H$, we need some extra equations. Since we typically will need $\boldsymbol{M}_0$ or $\boldsymbol{M}_0\boldsymbol{M}_0^H$ to be invertible, it will be required to have full rank. If we only have the minimum number of equations required to be able to compute the estimates of $\boldsymbol{N}^1(q^{-1})$ and $\boldsymbol{b_N}(q^{-1})$ as stated in (2.126), then the residuals, $\hat{\boldsymbol{r}}(t)$, will be identical to zero. For each training symbol added (adding M equations) over the minimum required, we add one degree of freedom in the residuals. Each degree of freedom in the residuals add one to the rank of the matrix $\hat{\boldsymbol{R}}_{\hat{\boldsymbol{r}}\hat{\boldsymbol{r}}}$. With M extra training symbols, the estimate of the matrix $\boldsymbol{M}_0\boldsymbol{M}_0^H$, i.e. $\hat{\boldsymbol{R}}_{\hat{\boldsymbol{r}}\hat{\boldsymbol{r}}}$, will in general be full rank. As a result the estimates of $\boldsymbol{M}_0\boldsymbol{M}_0^H$ and $\boldsymbol{M}_0$ will be invertible. We can thus say that we need the effective length of the training sequence to be at least $Mnn+nb+nn+1+M$. With $N_{eff} = N_{tseq} - nb - nn$ we require that the length of the training sequence fulfill

$$N_{tseq} \geq M(nn + 1) + 2 * (nb + nn) + 1. \tag{2.127}$$

2.9.2 Motivation for Joint FIR Channel and AR Noise Model Estimation

One may argue that the use of an AR model for the noise plus interference is not appropriate, as the true model for the noise plus interference more resembles either an MA or possibly an ARMA model. A *high* order AR model may be able to model an MA noise plus interferer spectrum appropriately but not a *low* order AR model. However, we will here show that a low order AR model, in fact, can be appropriate when used in an equalizer or symbol detection algorithm.

When we perform joint estimation of an FIR channel and an AR noise model, we can from (2.118) see that we are adjusting $\boldsymbol{N}(q^{-1})$ and $\boldsymbol{b_N}(q^{-1})$

to minimize the variance in the signal,

$$\boldsymbol{r}(t) = \boldsymbol{M}_0 \boldsymbol{v}(t) = \boldsymbol{N}(q^{-1})\boldsymbol{y}(t) - \boldsymbol{N}(q^{-1})\boldsymbol{b}(q^{-1})d(t), \tag{2.128}$$

under the constraint that the leading coefficient of $\boldsymbol{N}(q^{-1})$ is equal to the identity matrix.

Let us now assume that our received signal, $\boldsymbol{y}(t)$, is described by the general FIR channel model,

$$\boldsymbol{y}(t) = \boldsymbol{b}(q^{-1})d(t) + \boldsymbol{n}(t), \tag{2.129}$$

where the noise plus interference term, $\boldsymbol{n}(t)$, *is not* well modeled by an AR process, as assumed in (2.112). We can for example assume that the noise is well modeled by an moving average process.

Using (2.129) in (2.128) gives,

$$\boldsymbol{r}(t) = \boldsymbol{N}(q^{-1})\left(\boldsymbol{y}(t) - \boldsymbol{b}(q^{-1})d(t)\right) = \boldsymbol{N}(q^{-1})\boldsymbol{n}(t). \tag{2.130}$$

We can thus see that we, in effect, are adjusting $\boldsymbol{N}(q^{-1})$ to minimize the variance in the signal $\boldsymbol{r}(t) = \boldsymbol{N}(q^{-1})\boldsymbol{n}(t)$, under the constraint that the leading coefficient of $\boldsymbol{N}(q^{-1})$ is equal to the identity matrix.

A closer examination of the space-time equalizers and space-time maximum likelihood sequence estimators, that will be presented in Chapters 3 and 4, show that they all contain a noise whitening filter as a first processing step. Ideally the noise whitening filter should whiten the noise. However, if the noise whitening filter *cannot* whiten the noise, it should at least be designed to suppress the noise as much as possible. With the AR model of the noise plus interference in (2.112), the noise whitening filter becomes,

$$\boldsymbol{M}_0^{-1}\boldsymbol{N}(q^{-1}). \tag{2.131}$$

Thus, apart from the constant matrix factor, $\boldsymbol{M}_0^{-1}$, the polynomial MIMO filter, $\boldsymbol{N}(q^{-1})$, is the noise whitening filter[10].

From (2.130), we see that $\boldsymbol{N}(q^{-1})$ really is adjusted to suppress the noise as much as possible. The resulting estimate, $\hat{\boldsymbol{N}}(q^{-1})$, will thus likely serve well as a front end processing in a space-time DFE or a space-time MLSE.

[10] Without the factor, $\boldsymbol{M}_0^{-1}$, the filter $\boldsymbol{N}(q^{-1})$ will perform space-time noise whitening up to a remaining spatial color, which is handled by $\boldsymbol{M}_0^{-1}$

An important observation to make, is that the resulting polynomial filter, $\hat{\boldsymbol{N}}(q^{-1})$, does not have to *model* the noise well, it only has to *suppress* it. Suppressing the noise is easier than modeling it, and therefore a low order AR model for the noise, although crude, can be useful.

2.9.3 Reduced Complexity AR Noise Modeling

If we are using a large number of antennas the number of parameters may be large compared to the number of available training symbols. To alleviate this problem we may consider reducing the number of parameters in the model. This can be done in many different ways. The obvious solution is, of course, to decrease the model order of the AR filter. If we reduce it to zero we get a pure spatial model of the noise with no temporal color. We can however consider some other options.

One option is to restrict the denominator polynomial matrix $\boldsymbol{N}(q^{-1})$ in the model (2.112) to be diagonal, i.e. let

$$\boldsymbol{N}(q^{-1}) = \boldsymbol{N}_D(q^{-1}) = \begin{bmatrix} n_1(q^{-1}) & & 0 \\ & \ddots & \\ 0 & & n_M(q^{-1}) \end{bmatrix} \tag{2.132}$$

with the diagonal elements

$$n_i(q^{-1}) = 1 + n_{i1}q^{-1} + \ldots + n_{inn}q^{-nn}. \tag{2.133}$$

The noise plus interference term at each antenna element then has its own autoregressive filter $n_i(q^{-1})$. This noise model will, of course, not be as general as when we allow $\boldsymbol{N}(q^{-1})$ to be a full matrix, but it will be able to catch some of the spatio-temporal properties of the noise plus interferer spectrum. The number of unknowns is here considerably reduced as $\boldsymbol{N}_D(q^{-1})$ only has Mnn number of unknowns as compared to M^2nn for the full polynomial matrix $\boldsymbol{N}(q^{-1})$. The minimum effective length of the training sequence is thus here given by $nn + nb + nn + 1 + M$.

The next step in reducing the number of parameters in the model (2.112) while keeping some of the temporal modeling is to require all diagonal elements of $\boldsymbol{N}_D(q^{-1})$ to be equal, i.e. we require

$$\boldsymbol{N}(q^{-1}) = \begin{bmatrix} n(q^{-1}) & & 0 \\ & \ddots & \\ 0 & & n(q^{-1}) \end{bmatrix} \tag{2.134}$$

where $n(q^{-1})$ is the common denominator polynomial for all antenna elements

$$n(q^{-1}) = 1 + n_1 q^{-1} + \ldots + n_{nn} q^{-nn}. \tag{2.135}$$

The estimation of this model is however more complicated as it will involve constrained optimization, constraining the denominators for the noise models at the different antenna elements to be equal. In this model of the noise the spatial color and the temporal color decouple completely, i.e. we have the same spatial color for each time tap in the model and we have the same temporal color for each antenna element. An equalizer designed using this type of noise model would perform decoupled spatial and temporal suppression of the noise plus interference.

2.A Appendix

2.A.1 Linearization of the Modulation in GSM

The propagation in the wireless channel is typically well modeled by an FIR filter. However, all modulation schemes are not linear. In order to model the whole communication channel with an FIR filter as in (2.1) we need to describe the modulation process with a linear FIR model. The modulation applied in the GSM standard is an example of a modulation that is non-linear. However, as we will see here, it can be approximated with a linear model after some processing.

2.A.2 Modulation in GSM

The modulation used in the GSM system is Gaussian minimum shift keying (GMSK) with a bandwidth-time product (BT product) of 0.3 [24]. The modulated baseband signal can in continuous time, t_c, be expressed as as

$$s(t_c) = \exp(i\pi h \sum_n a(n)\phi(t_c - nT)) \tag{2.136}$$

with

$$a(n) = d(n)d(n-1) \tag{2.137}$$

where $d(\cdot) = \pm 1$ are the transmitted binary symbols and $h = 1/2$ is the modulation index used. The function $\phi(t_c)$ is defined as

$$\phi(t_c) \triangleq \int_{-\infty}^{t_c} g(\tau)d\tau \tag{2.138}$$

where

$$g(\tau) = f(\tau) \star \frac{1}{T}\mathrm{rect}(\tau/T) \tag{2.139}$$

with

$$\mathrm{rect}(x) = 1 \quad \text{for } |x| < 1/2 \text{ and zero otherwise} \tag{2.140}$$

and

$$f(t) = \frac{1}{\sqrt{2\pi}\sigma T}\exp(-{t_c}^2/(2\sigma^2 T^2)) \tag{2.141}$$

where

$$\sigma = \frac{\sqrt{\ln(2)}}{2\pi BT} \tag{2.142}$$

where $BT = 0.3$ is the bandwidth time product.

2.A.3 Linearization without Receiver Filter

As shown in [12], using the techniques in [48], a signal modulated as in the GSM standard, shown in (2.136), using $BT = 0.3$ can be well approximated by a linear filtering operation through a pulse-shaping filter $p(\cdot)$:

$$s(t_c) \approx \hat{s}(t_c) = \sum_n i^n d(n) p(t_c - nT) \tag{2.143}$$

where

$$p(t_c) = \Pi_{l=0}^{L-1} sin(\frac{\pi}{2}\psi(t_c + lT)) \tag{2.144}$$

and

$$\psi(t_c) = \begin{cases} \phi_{st}(t_c) & t \leq LT \\ 1 - \phi_{st}(t_c - LT) & t > LT \end{cases} . \tag{2.145}$$

The function $\phi_{st}(t_c)$ is the phase shift function $\phi(t_c)$ in (2.138) shifted and truncated such that it is zero for $t_c < 0$ and constant for $t_c > LT$.

By multiplying the received signal with $i^{-t_c/T}$ we can form the *rotated* received signal

$$s_r(t_c) = i^{-t_c/T} s(t_c). \tag{2.146}$$

The rotated received signal, $s_r(t_c)$, can be approximated as

$$s_r(t_c) \approx \hat{s}_r(t_c) = \sum_n d(n) i^{-(t_c - nT)/T} p(t_c - nT). \tag{2.147}$$

The GMSK modulated signal can thus be approximated as pulse amplitude modulation with the pulse

$$p_r(t_c) = i^{-t_c/T} p(t_c). \tag{2.148}$$

Using this approximation one should be able to achieve a good description of the channel between the transmitted symbols $d(t)$ and the received rotated signal, $s_r(t_c)$, as seen in Figure 2.27, with an FIR filter.

In a sampled system we can thus express the sampled modulated signal as

$$s_r(t) = p_r(q^{-1})d(t) \tag{2.149}$$

where

$$p_r(q^{-1}) = p_r(0) + p_r(T)q^{-1} + p_r(2T)q^{-2} + \ldots + p_r(LT)q^{-L}. \tag{2.150}$$

In order to simplify the notation we will however at this point drop the r subscripts and just write

$$s(t) = p(q^{-1})d(t) \tag{2.151}$$

and simply assume that the rotation has been performed on $s(t)$ and the coefficients of $p(q^{-1})$. The coefficients of $p(q^{-1}) = p_0 + p_1 q^{-1} + \ldots + p_L q^{-L}$ will thus be

$$p_i = p_r(iT) \quad , i = 0, 1, \ldots , L. \tag{2.152}$$

In Table 2.1, FIR filters approximating the channel $p(q^{-1})$ have been tabulated for different offsets in the sampling instant. These channels have been computed by creating a signal modulated as in the GSM standard (with BT=0.3) and fitting filters[11] to the channels between the transmitted binary signal and the rotated received signal.

2.A.4 Linearization with a Receiver Filter

With a receiver filter, the channel between the transmitted symbols and the received samples will have a somewhat longer impulse response. To show

[11] The FIR filters are computed by estimating the channel with a least squares algorithm using many data points.

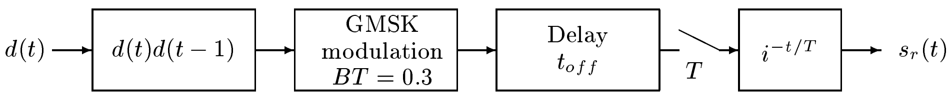


Figure 2.27: Model of the channel between the transmitted symbols and the rotated received GMSK modulated samples.

$p_{t_{off}}(q^{-1})$ coefficients			
t_{off}	p_0	p_1	p_2
-0.5	0.04∠1.59	0.71∠0.00	0.71∠-1.57
-0.4	0.05∠1.65	0.78∠0.00	0.62∠-1.58
-0.3	0.09∠1.55	0.84∠0.00	0.52∠-1.57
-0.2	0.13∠1.64	0.88∠0.00	0.43∠-1.59
-0.1	0.19∠1.53	0.92∠0.00	0.34∠-1.55
-0.0	0.26∠1.60	0.93∠0.00	0.26∠-1.60
0.1	0.34∠1.60	0.90∠0.00	0.19∠-1.63
0.2	0.43∠1.57	0.89∠0.00	0.13∠-1.59
0.3	0.52∠1.57	0.83∠0.00	0.09∠-1.59
0.4	0.62∠1.55	0.78∠0.00	0.05∠-1.44
0.5	0.71∠1.57	0.71∠0.00	0.04∠-1.55

Table 2.1: Discrete-time channel approximation of sampled GSM modulation. The sampling offset relative to the center of the symbol, t_{off}, is in units of a symbol interval. The channels are $p(q^{-1}) = p_0 + p_1 q^{-1} + p_2 q^{-2}$, $(q^{-1}d(t) = d(t-1))$. The phase of the channel taps has been rotated such that the middle tap has zero phase.

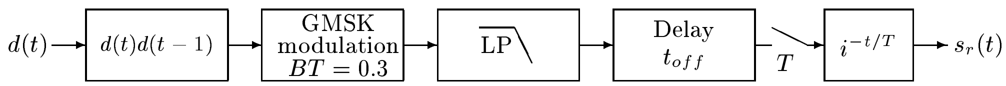


Figure 2.28: Model of the channel between the transmitted symbols and the received samples.

	$p_{t_{off}}(q^{-1})$ coefficients				
t_{off} [T]	p_0	p_1	p_2	p_3	p_4
-0.5	$0.01\angle{-1.97}$	$0.15\angle{+1.53}$	$0.81\angle{-0.01}$	$0.84\angle{-1.57}$	$0.07\angle{-3.05}$
-0.4	$0.01\angle{-2.24}$	$0.20\angle{+1.53}$	$0.87\angle{-0.01}$	$0.77\angle{-1.57}$	$0.02\angle{-2.88}$
-0.3	$0.01\angle{-2.62}$	$0.25\angle{+1.54}$	$0.92\angle{-0.01}$	$0.69\angle{-1.57}$	$0.02\angle{-0.26}$
-0.2	$0.01\angle{-2.92}$	$0.31\angle{+1.54}$	$0.96\angle{+0.00}$	$0.61\angle{-1.57}$	$0.05\angle{-0.08}$
-0.1	$0.02\angle{-3.08}$	$0.37\angle{+1.55}$	$0.99\angle{+0.00}$	$0.52\angle{-1.56}$	$0.08\angle{-0.05}$
0.0	$0.03\angle{+3.12}$	$0.44\angle{+1.55}$	$1.00\angle{+0.00}$	$0.44\angle{-1.56}$	$0.10\angle{-0.03}$
0.1	$0.04\angle{+3.09}$	$0.52\angle{+1.55}$	$1.00\angle{+0.00}$	$0.35\angle{-1.56}$	$0.11\angle{-0.01}$
0.2	$0.06\angle{+3.09}$	$0.59\angle{+1.56}$	$0.98\angle{+0.00}$	$0.28\angle{-1.55}$	$0.11\angle{+0.00}$
0.3	$0.08\angle{+3.09}$	$0.67\angle{+1.56}$	$0.95\angle{+0.00}$	$0.20\angle{-1.54}$	$0.11\angle{+0.01}$
0.4	$0.11\angle{+3.09}$	$0.74\angle{+1.56}$	$0.90\angle{+0.01}$	$0.13\angle{-1.53}$	$0.10\angle{+0.03}$
0.5	$0.15\angle{+3.10}$	$0.81\angle{+1.56}$	$0.84\angle{+0.01}$	$0.07\angle{-1.50}$	$0.09\angle{+0.04}$

Table 2.2: Discrete-time channels approximating the channel between the transmitted symbols and the received samples portrayed in Figure 2.28. The sampling offset relative to the center of the symbol, t_{off}, is in units of a symbol interval. The channels are $p_{t_{off}}(q^{-1}) = p_0 + p_1 q^{-1} + p_2 q^{-2} + p_3 q^{-3} + p_4 q^{-4}$, $(q^{-1}d(t) = d(t-1))$.

this effect a fourth order Butterworth lowpass filter with a bandwidth of 90 kHz has here been used to model a receiver filter. The symbol rate used was 270833 kbit/s ($T = 3.69\mu$s) as in GSM. An illustration of the channel between the transmitted symbols $d(t)$ and the received rotated samples can be seen in Figure 2.28.

In Table 2.2 the resulting approximating channels are displayed. We can observe that the effective impulse response length has become slightly increased.

2.B Least Squares for FIR Channel and AR Noise Estimation

We here outline the equations for the least squares method applied to (2.121).

The coefficients of the estimate,

$$\hat{\boldsymbol{N}}(q^{-1}) = \boldsymbol{I} + \hat{\boldsymbol{N}}_1 q^{-1} + \ldots \hat{\boldsymbol{N}}_{nn} q^{-nn} \tag{2.153}$$

and the estimate

$$\hat{\boldsymbol{b}}_{\boldsymbol{N}}(q^{-1}) = \hat{\boldsymbol{b}}_{\boldsymbol{N},0} + \ldots \hat{\boldsymbol{b}}_{\boldsymbol{N},nb+nn} q^{-nb-nn} \tag{2.154}$$

can be computed as

$$\begin{bmatrix} \hat{\boldsymbol{N}}^1 & \hat{\boldsymbol{B}}_{\boldsymbol{N}} \end{bmatrix} = \hat{\boldsymbol{R}}_{\boldsymbol{yx}} \hat{\boldsymbol{R}}_{\boldsymbol{xx}}^{-1} \tag{2.155}$$

where

$$\hat{\boldsymbol{R}}_{\boldsymbol{yx}} = \frac{1}{N_{tseq} - nb - nn} \sum_{t=nb+nn++1}^{N_{tseq}} \boldsymbol{y}(t)\boldsymbol{x}(t)^H \tag{2.156}$$

$$\hat{\boldsymbol{R}}_{\boldsymbol{xx}} = \frac{1}{N_{tseq} - nb - nn} \sum_{t=nb+nn+1}^{N_{tseq}} \boldsymbol{x}(t)\boldsymbol{x}^H(t) \tag{2.157}$$

with

$$\boldsymbol{x}(t) = \begin{bmatrix} -\boldsymbol{y}^T(t-1) & \ldots & -\boldsymbol{y}^T(t-nn) & \boldsymbol{d}^T(t) \end{bmatrix}^T \tag{2.158}$$

$$\boldsymbol{d}(t) = \begin{bmatrix} d(t) & \ldots & d(t-nb-nn) \end{bmatrix}^T \tag{2.159}$$

The coefficients in $\hat{\boldsymbol{N}}(q^{-1})$, except for the unit leading coefficient, should be extracted from $\hat{\boldsymbol{N}}^1$ according to

$$\hat{\boldsymbol{N}}^1 = \begin{bmatrix} \hat{\boldsymbol{N}}_1 & \ldots & \hat{\boldsymbol{N}}_{nn} \end{bmatrix} \tag{2.160}$$

and the coefficients in $\hat{\boldsymbol{b}}_{\boldsymbol{N}}(q^{-1})$ should be extracted from $\hat{\boldsymbol{B}}_{\boldsymbol{N}}$ according to

$$\hat{\boldsymbol{B}}_{\boldsymbol{N}} = \begin{bmatrix} \hat{\boldsymbol{b}}_{\boldsymbol{N},0} & \ldots & \hat{\boldsymbol{b}}_{\boldsymbol{N},nn} \end{bmatrix} \tag{2.161}$$

Chapter 3

Space-Time Decision Feedback Equalization

3.1 Introduction

In a wireless communication channel there can often be a considerable delay spread. This delay can be larger than the symbol time. In this case the signaling over the channel experiences intersymbol interference. The received signal samples, $y(t)$, can then often be modeled as in (2.1) with an FIR channel model

$$\begin{aligned} y(t) &= b(q^{-1})d(t) + n(t) \\ &= (b_0 + b_1 q^{-1} + \ldots + b_{nb} q^{-nb})d(t) + n(t) \\ &= b_0 d(t) + b_1 d(t-1) + \ldots + b_{nb} d(t-nb) + n(t) \end{aligned} \tag{3.1}$$

where $d(t)$ are the transmitted symbols, $n(t)$ is a term representing the noise plus interference and $\{b_i\}$ are the coefficients of the FIR channel model

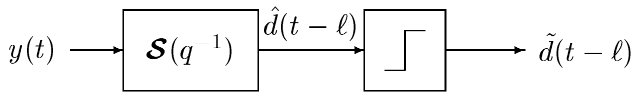


Figure 3.1: Linear time-only equalizer.

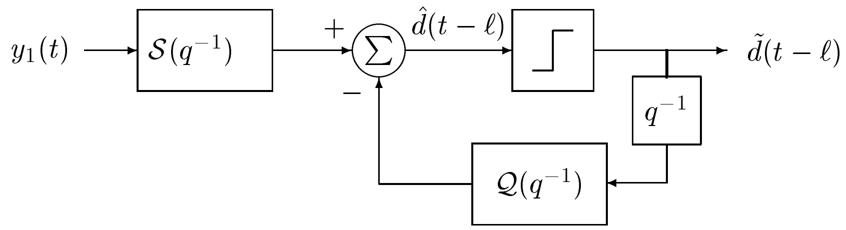


Figure 3.2: Time-only decision feedback equalizer.

$b(q^{-1})$.

In order to detect the transmitted signal, we need to combat the intersymbol interference resulting from the delay spread, and the noise plus interference in the term $n(t)$. There are three major choices of detectors that can be employed to solve this problem: the linear equalizer, the decision feedback equalizer and the maximum likelihood sequence estimator. Of course, there exists other possible detectors as well, but the three mentioned above are the most "popular" ones. We will therefore focus our attention on these algorithms.

First we consider the *linear equalizer*, as depicted in Figure 3.1 which forms an estimate of the transmitted symbol by linear filtering of the received signal

$$\hat{d}(t-\ell) = \boldsymbol{\mathcal{S}}(q^{-1})\boldsymbol{y}(t) \tag{3.2}$$

where $\boldsymbol{\mathcal{S}}(q^{-1})$ is a general stable and causal linear filter. The final estimate of the symbol is achieved by feeding $\hat{d}(t-\ell)$ through a decision device and forming a discrete estimate of the transmitted discrete symbol. The parameter ℓ is the *smoothing lag* of the equalizer. A larger smoothing lag improves the performance of the equalizer, at least up to some point.

The linear temporal equalizer has a considerable drawback. In essence it attempts to invert the channel to equalize the channel and detect the transmitted symbols. This can be difficult if the channel is non-minimum phase as the equalizer filter then would need poles outside the unit circle and thus would become unstable. At the same time the equalizer should avoid to amplify received noise. Increasing the smoothing lag will help the equalization but some channels will still be difficult to equalize.

Instead of using a linear equalizer we can use a *decision feedback equalizer* (DFE). In a decision feedback equalizer we add a linear filter, $\mathcal{Q}(q^{-1})$, which is processing previously decided symbols from the output of the equalizer as depicted in Figure 3.2. The *feedforward filter* $\boldsymbol{\mathcal{S}}(q^{-1})$ will then be aided by the *feedback filter* $\mathcal{Q}(q^{-1})$ such that the feedforward filter no longer needs to *invert* the channel. As long as the majority of the symbols are correctly decided this will facilitate the channel equalization and the symbol detection. The performance of the temporal decision feedback equalizer will often be considerably better than the temporal linear equalizer.

The third type of algorithm we consider is a maximum likelihood sequence estimator (MLSE) [26, 116]. A maximum likelihood sequence estimator processes the received data and searches for the most likely symbol *sequence* to have been transmitted. The maximum likelihood sequence estimator has an advantage over the linear equalizer and the decision feedback equalizer in that it detects symbol sequences rather than individual symbols as for the linear and the decision feedback equalizer. If the channel is perfectly known and the noise is white and Gaussian then the maximum likelihood sequence estimator finds the optimal estimate of the transmitted symbol sequence. The maximum likelihood sequence estimator generalized to a space-time MLSE will be discussed in Chapter 4.

The equalizers described above only operates on a single input signal and therefore only performs temporal processing. However, if we have multiple received signals from multiple antennas we can perform space-time processing. The multiple signals do not necessarily need to come from spatially separated antennas. They can be from some other sources of diversity, for example, from antennas with different polarization. All this processing can be treated in the frame-work of space-time processing. We can, of course, generalize all of the above equalizers to work with multiple input signals.

An interesting fact to note is that when we generalize these equalizers to work with multiple input signals, their difference in performance tends to be reduced. The reason for this is that the linear equalizer, the decision feedback equalizer as well as the maximum likelihood sequence estimator, all has a linear filter up front operating on the received signal. Basically what happens is that the linear filter can more easily invert the channel or produce the desired signal when we are performing space-time processing compared to when we are performing temporal processing only. As a result the performance of the different equalizers will be more comparable.

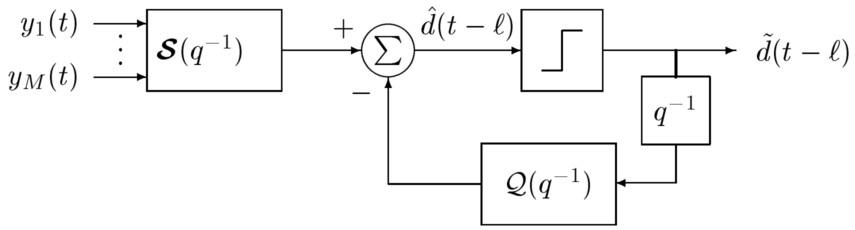


Figure 3.3: General space-time DFE with linear filters. The feedforward filter is here a linear MISO filter, represented with the rational MISO transfer function operator $\boldsymbol{\mathcal{S}}(q^{-1})$. Likewise the scalar linear feedback filter is represented with a rational SISO transfer function operator $\mathcal{Q}(q^{-1})$.

In this chapter we will consider *the space-time decision feedback equalizer*. The linear equalizer can be considered a special case of the decision feedback equalizer, i.e. without the feedback filter. A general *space-time* DFE with linear filters can be seen in Figure 3.3. The M input signals from, say, M antenna elements[1], are filtered in the feedforward filter $\boldsymbol{\mathcal{S}}(q^{-1})$. Previously decided symbols are filtered in the feedback filter $\mathcal{Q}(q^{-1})$ and its output is subtracted from that of the feedforward filter. The resulting signal is fed to a decision device to form an estimate, $\hat{d}(t-\ell)$ of the transmitted discrete symbol $d(t-\ell)$, where ℓ represents the smoothing lag of the equalizer. The transmitted symbol sequence is assumed to be temporally white. Since the space-time DFE utilizes signals from M antenna elements, we can perform space-time filtering in the feedforward filter. As compared to the scalar case, this supplies the space-time DFE with extra signal dimensions that can be used for signal enhancement and interference suppression.

This chapter considers optimal tuning solutions for space-time DFE's for different channels and constraints. First in Section 3.2.1, the general MMSE optimal space-time DFE with linear filters for an ARMA channel with ARMA noise is presented. This estimator, derived assuming correct past decisions, is a generalization of the scalar MMSE optimal DFE presented in [90]. If we restrict the channel model to have diagonal denominator polynomial matrices, then it is also a special case to the multi-user MIMO-DFE presented in [106]. In general, the space-time DFE has a space-time feedforward IIR filter and a scalar IIR feedback filter. The rational filters, present in the optimal design, can sometimes cause complications. The IIR filters

[1]The M signals can also arise from using other types of diversity branches.

optimized assuming previous correct decisions may have poles close to the stability limit. If this is the case for the feedback filter, then an erroneous decision can cause a long error burst. This effect can be seen in Section 7.2 of Chapter 7 where robustness against decision errors are studied.

Most wireless communication channels can be well modeled by FIR filters. This is therefore an important special case to consider. Furthermore, if we restrict the interference and noise model to being an AR-model, then the resulting DFE with optimal structure will contain only FIR filters. From the discussion in Section 2.9.2, we can understand why an AR model, even a *low order* AR model, for the noise plus interference can be useful when used in a decision feedback equalizer. The important observation to make is that the the AR noise model denominator does not have to *model* the noise particularly well, it only has to be able to *suppress* the noise plus interference as a part of a noise whitening filter. The use of an AR model for the noise plus interference was proposed in [7] in conjunction to a maximum likelihood sequence estimator. It can however also be used together with a decision feedback equalizer. An advantage of using an AR noise model together with a decision feedback equalizer is that the increase in complexity for the execution of the DFE with the AR noise model is linear in the order of the AR noise model used, while the increase in execution complexity for the MLSE is exponential in the order of the AR noise model. For this reason it is more feasible with higher order AR noise models together with the DFE than together with an MLSE. The MMSE optimal space-time DFE for an FIR channel with AR-noise is treated separately, and presented in Section 3.2.2. An example AR modeling of the noise plus interference combined with a space-time decision feedback equalizer can be seen in one of the simulations in Section 3.4.

Yet another alternative is to restrict the space-time DFE to have FIR filters of specified orders in the feedforward and the feedback filters and optimize the coefficient values, given the structure. The MMSE optimal space-time DFE with FIR filters of a given order is therefore presented in Section 3.2.3. The method can use the general space-time spectrum of the noise plus interference.

Although, the space-time DFE's introduced above are optimal under the given conditions, they are somewhat unpractical, since their tuning requires the solution of a large system of linear equations.

The MMF-DFE presented in Section 3.2.4 has the space-time processing concentrated to a multi-dimensional matched filter followed by a scalar temporal DFE [83]. The MMF-DFE is a generalization to colored noise of the results presented in [8][9]. If we restrict us to consider only the spatial covariance of the noise plus interference or an AR model of the noise plus interference, then the multi-dimensional filter is very easy to compute, and so is the scalar DFE. In the simulations of Section 3.4 the MMF-DFE achieves about the same performance as the other DFE's above. The difference in complexity is especially large when the number of antennas is large, making the MMF-DFE preferable over the others for this case. A drawback with the MMF-DFE is however that it will introduce a larger decision delay. This will be of significant importance only when time-varying channels need to be tracked.

In Section 3.3 different tuning options for the above mentioned DFE's are presented. It is noted that when using many antenna elements it may be difficult to make use of the full spatio-temporal spectrum of the noise plus interference, due to the difficulty of accurately estimating it. As mentioned in Section 2.3, one solution to this problem is to utilize only the spatial spectrum of the noise plus interference. Interference suppression is then performed in the spatial domain only [61]. This is useful when we have a large number of antenna elements. It should however be noted that the number of antennas can be large in more than one way. First, it can be large in the sense that the number of parameters to be tuned in the equalizers becomes large compared to the length of the training sequence. Second, it can be large compared to the number of uncorrelated interfering signals present, making spatial suppression of the interferers effective. In the simulations performed here in Section3.4, the number of antennas was large in both these senses. The conclusions therefore really has to be restricted to the specific case studied.

Another solution to the problem of handling the spatio-temporal color of the noise plus interference can be, as mentioned above, to estimate a spatio-temporal AR model for the noise plus interference. This is described in Section 2.9.

3.2 Optimal Space-Time Decision Feedback Equalizers

3.2.1 Optimal Space-Time DFE for ARMA Channels with ARMA Noise

A general linear time-invariant space-time channel can be described by an ARMA model for both the desired signal and the noise plus interference. We will here discuss the MMSE optimal space-time DFE for such a channel. This is a generalization of the MMSE optimal DFE for the scalar case [90]. If we restrict the channel model to have diagonal denominators, then it is also a special case to the multi-user MIMO-DFE presented in [106].

Let the received signal vector $\boldsymbol{y}(t)$, of dimensions $M \times 1$, be modeled as

$$\boldsymbol{y}(t) = \boldsymbol{A}^{-1}(q^{-1})\boldsymbol{b}(q^{-1})d(t) + \boldsymbol{N}^{-1}(q^{-1})\boldsymbol{M}(q^{-1})\boldsymbol{v}(t). \tag{3.3}$$

Here $\boldsymbol{A}(q^{-1})$ and $\boldsymbol{N}(q^{-1})$ are assumed to be stably and causally invertible $M \times M$ polynomial matrices of orders na and nn respectively. The matrix $\boldsymbol{M}(q^{-1})$ is assumed to be a stably and causally invertible $M \times M$ polynomial matrix of order nm and $\boldsymbol{b}(q^{-1})$ is an $M \times 1$ polynomial column vector of order nb. The $M \times 1$ vector $\boldsymbol{v}(t)$ is an innovations sequence for the model of the noise plus interference term, $\boldsymbol{N}^{-1}(q^{-1})\boldsymbol{M}(q^{-1})\boldsymbol{v}(t)$.

The sequence of transmitted symbols $\{d(t)\}$ are assumed to be temporally white with zero mean. Furthermore, we assume the transmitted symbols to be scaled to have unit variance, i.e.

$$E[d(t)d^H(t)] = 1. \tag{3.4}$$

The polynomial matrices $\boldsymbol{A}(q^{-1})$, $\boldsymbol{N}(q^{-1})$ and $\boldsymbol{M}(q^{-1})$ are thus assumed to be of the forms

$$\boldsymbol{A}(q^{-1}) = \begin{bmatrix} a_{11}(q^{-1}) & \dots & a_{1M}(q^{-1}) \\ \vdots & \ddots & \vdots \\ a_{M1}(q^{-1}) & \dots & a_{MM}(q^{-1}) \end{bmatrix} \tag{3.5}$$

$$\boldsymbol{N}(q^{-1}) = \begin{bmatrix} n_{11}(q^{-1}) & \dots & n_{1M}(q^{-1}) \\ \vdots & \ddots & \vdots \\ n_{M1}(q^{-1}) & \dots & n_{MM}(q^{-1}) \end{bmatrix} \tag{3.6}$$

and

$$\boldsymbol{M}(q^{-1}) = \begin{bmatrix} m_{11}(q^{-1}) & \dots & m_{1M}(q^{-1}) \\ \vdots & \ddots & \vdots \\ m_{M1}(q^{-1}) & \dots & m_{MM}(q^{-1}) \end{bmatrix} \tag{3.7}$$

with the respective elements

$$a_{ij}(q^{-1}) = a_{ij,0} + a_{ij,1}q^{-1} + \dots + a_{ij,na}q^{-na} \ , \tag{3.8}$$

$$n_{ij}(q^{-1}) = n_{ij,0} + n_{ij,1}q^{-1} + \dots + n_{ij,nn}q^{-nn} \tag{3.9}$$

and

$$m_{ij}(q^{-1}) = m_{ij,0} + m_{ij,1}q^{-1} + \dots + m_{ij,nn}q^{-nm}. \tag{3.10}$$

The requirement that the matrices $\boldsymbol{A}(q^{-1})$, $\boldsymbol{N}(q^{-1})$ and $\boldsymbol{M}(q^{-1})$ are stably invertible is equivalent to require that their determinants have zeros strictly inside the unit circle, while causality is guaranteed if the leading the leading matrix coefficient of the matrix polynomials $\boldsymbol{A}(q^{-1})$ and $\boldsymbol{N}(q^{-1})$ are non-singular. For simplicity we will assume that the leading coefficient matrices of $\boldsymbol{A}(q^{-1})$ and $\boldsymbol{N}(q^{-1})$ are equal to the unit matrix.

The channel for the desired signal is a polynomial column vector

$$\boldsymbol{b}(q^{-1}) = \begin{bmatrix} b_1(q^{-1}) \\ \vdots \\ b_M(q^{-1}) \end{bmatrix} \tag{3.11}$$

with

$$b_i(q^{-1}) = b_{i,0} + b_{i,1}q^{-1} + \dots + b_{i,nb}q^{-nb}. \tag{3.12}$$

We have here assumed that all elements of the polynomial matrices and vectors have the same degree. Different degrees can however be obtained by setting trailing polynomial coefficients to zero.

The $M \times 1$ vector $\boldsymbol{v}(t)$ is a temporally and spatially white innovations sequence with the covariance matrix equal to unity, i.e.

$$E[\boldsymbol{v}(t_1)\boldsymbol{v}^H(t_2)] = \delta_{t_1 t_2}\boldsymbol{I}. \tag{3.13}$$

The estimated symbol prior to the decision device in the DFE can be expressed as

$$\hat{d}(t-\ell) = \boldsymbol{\mathcal{S}}(q^{-1})\boldsymbol{y}(t) - \mathcal{Q}(q^{-1})\tilde{d}(t-\ell-1) \tag{3.14}$$

where $\boldsymbol{\mathcal{S}}(q^{-1})$ and $\mathcal{Q}(q^{-1})$ are stable and causal rational filters of dimensions $1 \times M$ and 1×1 respectively, see Figure 3.3. The parameter ℓ, known as the decision delay, or the smoothing lag, is chosen by the user. The signal $\tilde{d}(t-\ell-1)$ denotes previously decided symbols. The estimate $\hat{d}(t-\ell)$ is fed into the decision device to produce a hard estimate $\tilde{d}(t-\ell)$ of the transmitted discrete symbol $d(t-\ell)$. To maintain a linear filtering problem we adopt the commonly used assumption that all previous decisions fed into the feedback filter are correct.

In Appendix 3.A.1, the derivation of the resulting MMSE optimal DFE is carried out in detail. We will here summarize the result. Assuming $\boldsymbol{A}(q^{-1})$ and $\boldsymbol{M}(q^{-1})$ to be stably and causally invertible, the rational feedforward filters which minimizes the criterion

$$J = E\left[|d(t-\ell) - \hat{d}(t-\ell)|^2\right] \tag{3.15}$$

is given by

$$\boldsymbol{\mathcal{S}}(q^{-1}) = \boldsymbol{s}_0(q^{-1})\boldsymbol{M}^{-1}(q^{-1})\boldsymbol{N}(q^{-1}) \tag{3.16}$$

while the corresponding optimal rational feedback filter is given by

$$\mathcal{Q}(q^{-1}) = \frac{Q(q^{-1})}{a(q^{-1})m(q^{-1})} \tag{3.17}$$

with

$$a(q^{-1}) = \det\left(\boldsymbol{A}(q^{-1})\right) \tag{3.18}$$

$$m(q^{-1}) = \det\left(\boldsymbol{M}(q^{-1})\right). \tag{3.19}$$

The degree of the polynomial $\boldsymbol{s}_0(q^{-1})$ is equal to the smoothing lag, ℓ and the degree of the polynomial $Q(q^{-1})$ is equal to $\max[nb+nn+(M-1)(nm+na), M(nm+na)]-1$.

The coefficients of $\boldsymbol{s}_0(q^{-1})$ can be computed by solving the system of equations

$$\begin{bmatrix} -\boldsymbol{u}_0^H & & 0 & p_0^H & & 0 \\ \vdots & \ddots & & \vdots & \ddots & \\ -\boldsymbol{u}_\ell^H & \cdots & -\boldsymbol{u}_0^H & p_\ell^H & \cdots & p_0^H \\ -\boldsymbol{H}_0 & \cdots & -\boldsymbol{H}_\ell & \boldsymbol{c}_0 & \cdots & \boldsymbol{c}_\ell \\ & \ddots & \vdots & & \ddots & \vdots \\ 0 & & -\boldsymbol{H}_0 & 0 & & \boldsymbol{c}_0 \end{bmatrix} \begin{bmatrix} \boldsymbol{s}_{0,0}^H \\ \vdots \\ \boldsymbol{s}_{0,\ell}^H \\ l_{1_\ell} \\ \vdots \\ l_{10} \end{bmatrix} = \begin{bmatrix} 0 \\ \vdots \\ p_0^H \\ 0 \\ \vdots \\ 0 \end{bmatrix} \tag{3.20}$$

where

$$\begin{aligned} \boldsymbol{H}(q^{-1}) &= \boldsymbol{H}_0 + \boldsymbol{H}_1 q^{-1} + \ldots + \boldsymbol{H}_{nh} q^{-nh} \\ &\triangleq a(q^{-1})\mathrm{adj}\left(\boldsymbol{N}(q^{-1})\right)\boldsymbol{M}(q^{-1}) \end{aligned} \tag{3.21}$$

$$\boldsymbol{u}(q^{-1}) = \mathrm{adj}\left(\boldsymbol{M}(q^{-1})\right)\boldsymbol{N}(q^{-1})\mathrm{adj}\left(\boldsymbol{A}(q^{-1})\right)\boldsymbol{b}(q^{-1}) \tag{3.22}$$

$$p(q^{-1}) \triangleq a(q^{-1})m(q^{-1}) \tag{3.23}$$

and

$$\boldsymbol{c}(q^{-1}) = \boldsymbol{c}_0 + \boldsymbol{c}_1 q^{-1} + \ldots + \boldsymbol{c}_{nc} q^{-nc} = n(q^{-1})\mathrm{adj}\left(\boldsymbol{A}(q^{-1})\right)\boldsymbol{b}(q^{-1}). \tag{3.24}$$

Here $\boldsymbol{H}_i = 0$ if $i > nh$, $\boldsymbol{c}_i = 0$ if $i > nc$, $\boldsymbol{u}_i = 0$ if $i > nu$ and $p_i = 0$ if $i > np$.

The coefficients of the scalar feedback filter polynomial, $Q(q^{-1})$ of order $np = \max[nb+nn+(M-1)(nm+na), M(nm+na)]$, can be computed as

$$\begin{bmatrix} Q_0^H \\ \vdots \\ Q_{nq}^H \end{bmatrix} = \begin{bmatrix} -\boldsymbol{u}_{\ell+1}^H & \cdots & -\boldsymbol{u}_1^H & p_{\ell+1}^H & \cdots & p_1^H \\ \vdots & & \vdots & \vdots & & \vdots \\ -\boldsymbol{u}_{\ell+1+nq}^H & \cdots & -\boldsymbol{u}_{1+nq}^H & p_{\ell+1+nq}^H & \cdots & p_{1+nq}^H \end{bmatrix} \begin{bmatrix} \boldsymbol{s}_{0,0}^H \\ \vdots \\ \boldsymbol{s}_{0,\ell}^H \\ l_{1_\ell} \\ \vdots \\ l_{11} \\ l_{10}-1 \end{bmatrix}. \tag{3.25}$$

Again $\boldsymbol{u}_i = 0$ if $i > nu$ and $p_i = 0$ if $i > np$.

3.2.2 Optimal Space-Time DFE for FIR Channels with AR Noise

The space-time DFE in Section 3.2.1 is often unnecessarily complex. A typical wireless communication channel can be modeled by an FIR channel. This is therefore an important special case to consider. It is also practical to have decision feedback equalizers with FIR filters in the feedforward and the feedback filters, rather than IIR-filters as is the case for the space-time for DFE for ARMA channels with ARMA noise[2]. We will here show that if we have a space-time FIR-model for the channel of the desired signal and a space-time AR-model for the noise and interference, then the MMSE-optimal DFE will have only FIR-filters in the feedforward and feedback filters. The filters can also be computed in a simple manner.

If the noise consists mainly of co-channel interferers, then an AR model will typically not be physically motivated. A sum of interferers each propagating through an FIR channel would be better described by a moving average (MA) model. From the discussion in Section 2.9.2, we can however understand why an AR model, even a *low order* AR model, for the noise plus interference can be useful when used in a decision feedback equalizer. The important observation to make is that the the AR noise model denominator does not have to model the noise particularly well, it only has to be able to suppress the noise as a part of a noise whitening filter. The use of an AR model for the noise was proposed in [7] in conjunction with a space-time MLSE. When we use an AR model for the noise in an MLSE, the memory length of the Viterbi algorithm used will be increased. This increases the complexity of Viterbi algorithm by a factor K^{nn}, where K is the number of symbols in the alphabet and nn is the order of the AR model for the noise plus interference. Thus the complexity for the MLSE using an AR noise model increases exponentially with the order of the AR noise model. When an AR model for the noise plus interference is used together with a space-time decision feedback equalizer, on the other hand, the increase in complexity is only linear in the order of the AR noise model. As a result we can allow AR noise models of a higher order for the space-time DFE.

[2] A reason for this is that if we have an IIR filter as feedback filter, then an erroneous symbol decision can cause a long error burst if the filter has poles close to the unit circle. If we use FIR filters only, then the number of symbols an erroneous decision directly can affect will be limited by the length of the FIR feedback filter.

The $M \times 1$ received signal vector $\boldsymbol{y}(t)$ is here modeled as

$$\begin{aligned}\boldsymbol{y}(t) &= \boldsymbol{b}(q^{-1})d(t) + \boldsymbol{N}^{-1}(q^{-1})\boldsymbol{M}_0\boldsymbol{v}(t) \\ &= \boldsymbol{b}_0 d(t) + \cdots + \boldsymbol{b}_{nb} d(t-nb) + \boldsymbol{N}^{-1}(q^{-1})\boldsymbol{M}_0\boldsymbol{v}(t) \qquad (3.26)\end{aligned}$$

where the $M \times 1$ polynomial column vector $\boldsymbol{b}(q^{-1})$ is the channel for the desired signal

$$\boldsymbol{b}(q^{-1}) = \begin{bmatrix} b_1(q^{-1}) \\ \vdots \\ b_M(q^{-1}) \end{bmatrix} \qquad (3.27)$$

with

$$b_i(q^{-1}) = b_{i,0} + b_{i,1}q^{-1} + \ldots + b_{i,nb}q^{-nb}. \qquad (3.28)$$

The noise model numerator, $\boldsymbol{M}_0$, is an $M \times M$ constant matrix

$$\boldsymbol{M}_0 = \begin{bmatrix} m_{11} & \ldots & m_{1M} \\ \vdots & \ddots & \vdots \\ m_{M1} & \ldots & m_{MM} \end{bmatrix} \qquad (3.29)$$

which is assumed to be nonsingular. For the noise model denominator, $\boldsymbol{N}(q^{-1})$, the transmitted symbols, $d(t)$, and the noise sequence, $\boldsymbol{v}(t)$, we adopt the same assumptions as in Section 3.2.1.

Given the channel model (3.26), with $\boldsymbol{M}_0$ invertible, and assuming correct past decisions, $\tilde{d}(t-\ell-1)$, to be fed into the feedback filter, then as shown in detail in Appendix 3.A.2, the MMSE optimal DFE with linear filters is given by

$$\hat{d}(t-\ell) = \boldsymbol{s}(q^{-1})y(t) - Q(q^{-1})\tilde{d}(t-\ell-1) \qquad (3.30)$$

where the feedforward FIR filter, $\boldsymbol{s}(q^{-1})$, is given by the row vector

$$\boldsymbol{s}(q^{-1}) = \boldsymbol{s}_0(q^{-1})\boldsymbol{M}_0^{-1}\boldsymbol{N}(q^{-1}). \qquad (3.31)$$

See Figure 3.4.

The coefficients of the polynomial vector $\boldsymbol{s}_0(q^{-1})$ of degree ℓ can be computed by solving the system of equations

$$(\boldsymbol{\mathcal{B}}'\boldsymbol{\mathcal{B}}'^H + \boldsymbol{I})\boldsymbol{s}_0^H = \begin{bmatrix} \boldsymbol{b}'_\ell \\ \vdots \\ \boldsymbol{b}'_0 \end{bmatrix} \qquad (3.32)$$

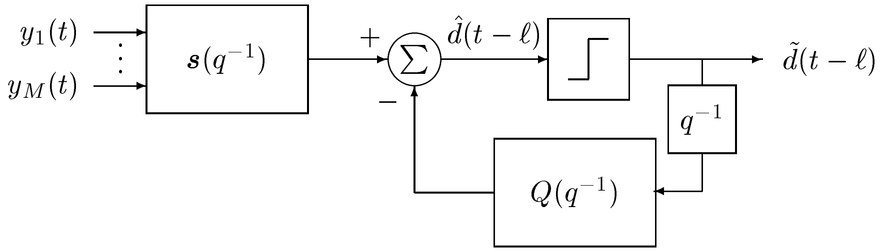


Figure 3.4: The structure of the space-time DFE for an FIR channel with AR noise. The feedforward filter, $\boldsymbol{s}(q^{-1})$, is a MISO FIR filter of order $\ell+nn$ and the feedback filter, $Q(q^{-1})$, is a scalar FIR filter of order $nb+nn-1$, where ℓ is the smoothing lag of the equalizer, nn is the order of the space-time AR noise model and nb is the order of the space-time FIR channel for the desired signal.

where

$$\boldsymbol{\mathcal{B}}' = \begin{bmatrix} \boldsymbol{b}'_0 & \cdots & \boldsymbol{b}'_\ell \\ & \ddots & \vdots \\ 0 & & \boldsymbol{b}'_0 \end{bmatrix} \tag{3.33}$$

$$\boldsymbol{s}_0 = \begin{bmatrix} \boldsymbol{s}_{0,0} & \ldots & \boldsymbol{s}_{0,\ell} \end{bmatrix} \tag{3.34}$$

where $\boldsymbol{b}'_i$ are the vector taps of the noise whitened channel

$$\boldsymbol{b}'(q^{-1}) = \boldsymbol{b}'_0 + \boldsymbol{b}'_1 q^{-1} + \ldots + \boldsymbol{b}'_{nb'} q^{-nb'} = \boldsymbol{M}_0^{-1} \boldsymbol{N}(q^{-1}) \boldsymbol{b}(q^{-1}) \tag{3.35}$$

with $\boldsymbol{b}'_i = 0$ if $i > nb' = nb + nn$, and $\boldsymbol{s}_{0,k}$ are the vector taps in the polynomial

$$\boldsymbol{s}_0(q^{-1}) = \boldsymbol{s}_{0,0} + \boldsymbol{s}_{0,1} q^{-1} + \ldots + \boldsymbol{s}_{0,\ell} q^{-\ell}. \tag{3.36}$$

The matrix $(\boldsymbol{\mathcal{B}}'\boldsymbol{\mathcal{B}}'^H + \boldsymbol{I})$ will be nonsingular since $\boldsymbol{I}$ obviously is a full rank matrix and $\boldsymbol{\mathcal{B}}'\boldsymbol{\mathcal{B}}'^H$ is a positive semi-definite matrix, making $(\boldsymbol{\mathcal{B}}'\boldsymbol{\mathcal{B}}'^H + \boldsymbol{I})$ a full rank (and thus invertible) matrix.

The coefficients of the feedback polynomial, $Q(q^{-1})$ of order $nq = nb+nn-1$, can be computed as

$$\begin{bmatrix} Q_0^H \\ \vdots \\ Q_{nq}^H \end{bmatrix} = \begin{bmatrix} \boldsymbol{b}'^H_{\ell+1} & \cdots & \boldsymbol{b}'^H_1 \\ \vdots & & \vdots \\ \boldsymbol{b}'^H_{\ell+nb+nn} & \cdots & \boldsymbol{b}'^H_{nb+nn} \end{bmatrix} \begin{bmatrix} \boldsymbol{s}^H_{0,0} \\ \vdots \\ \boldsymbol{s}^H_{0,\ell} \end{bmatrix} \tag{3.37}$$

where $\boldsymbol{b}'_i = 0$ if $i > nb' = nb + nn$.

We thus see that with the channel model of (3.26) with an FIR model for the channel of the desired user and an AR model for the noise plus interference, the optimal space-time DFE takes on a particular simple form with FIR filters in the feedforward and feedback filters.

We have here assumed that the denominator in the AR model for the noise plus interference is a full matrix, with the zeros of its determinant inside the unit circle. We can however, of course, restrict the denominator to be a diagonal matrix. This can be advantageous since, as discussed in Section 2.9, a diagonal AR noise model can be easier to estimate. The space-time DFE using a diagonal denominator in the AR-model will be able to perform some coupled space-time suppression of interference, but it will not be able to suppress interferers in the space-time domain as general as the space-time DFE using a full denominator matrix.

If we further restrict all elements of the diagonal denominator matrix to be equal, then the spatio-temporal model of the noise spectrum decouples into a separate spatial model, $\boldsymbol{M}_0$, and a common temporal model, $n(q^{-1})$, being the common denominator polynomial in the AR model. The resulting DFE can then only do *decoupled* space-time suppression of interferers.

3.2.3 Optimal Fixed Order Space-Time FIR-DFE for a FIR Channel with Colored Noise

Let us here consider the case of FIR channels with spatially and temporally colored noise plus interference. This will be the typical scenario when we have co-channel interferers experiencing intersymbol interference impinging on the antenna array. As we saw in Section 3.2.1, a DFE using FIR-filters is in general not optimal. There are however some advantages with the use of FIR-filters in the DFE. IIR-filters can have poles very close to the unit circle, resulting in very long impulse responses in the filters. This can for example result in long error propagation events caused by a long impulse response from an incorrect decision fed back through an IIR feedback filter. We will therefore here *constrain* the DFE to contain only FIR filters with specific degrees and tune the coefficient to minimize the MMSE prior to the decision device.

We here model the $M \times 1$ received signal vector $\boldsymbol{y}(t)$ using polynomial notation as

$$\begin{aligned} \boldsymbol{y}(t) &= \boldsymbol{b}(q^{-1})d(t) + \boldsymbol{n}(t) \\ &= (\boldsymbol{b}_0 + \boldsymbol{b}_1 q^{-1} + \ldots + \boldsymbol{b}_{nb} q^{-nb})d(t) + \boldsymbol{n}(t) \end{aligned} \tag{3.38}$$

where the spatially and temporally colored vector of noise plus interference samples $\boldsymbol{n}(t)$ has the matrix-valued covariance function

$$\boldsymbol{R}_{\boldsymbol{nn}}(k) \triangleq E[\boldsymbol{n}(t)\boldsymbol{n}^H(t-k)]. \tag{3.39}$$

Note that the noise plus interference term $\boldsymbol{n}(t)$ is assumed to be uncorrelated to the transmitted symbols $d(t)$. Further the transmitted symbols, $d(t)$, are as usual assumed white with unit variance.

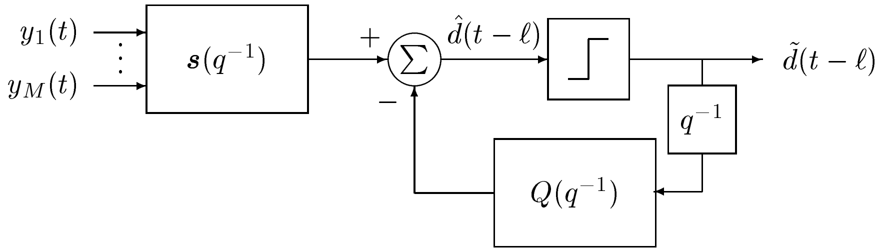


Figure 3.5: The structure of a space-time DFE with fixed order FIR filters. The orders of the MISO FIR feedforward filter, $\boldsymbol{s}(q^{-1})$, and the scalar FIR feedback filter, $Q(q^{-1})$, can in principle be chosen arbitrarily.

The space-time DFE with fixed order FIR filters in the feedforward and feedback filters is depicted in Figure 3.5. The estimated symbol prior to the decision device can thus be written as

$$\hat{d}(t-\ell) = \boldsymbol{s}(q^{-1})\boldsymbol{y}(t) - Q(q^{-1})\tilde{d}(t-\ell-1). \tag{3.40}$$

This estimate is then used as input to the decision device to produce a hard estimate $\tilde{d}(t-\ell)$ of the transmitted discrete symbol $d(t-\ell)$.

We now have to choose the smoothing lag and the orders of the feedforward and the feedback filter. With perfect channel knowledge a larger smoothing lag generally results in better performance. However, as a rule of thumb it is adequate to select the smoothing lag such that most of the received energy originating from a transmitted symbol has passed through the feedforward filter of the DFE before a decision is made.

As a guidance for the orders of the feedforward and the feedback filters we can study the orders of the FIR DFE in Section 3.2.2 operating on a FIR channel with AR noise. The filters in this DFE have the orders

$$\deg\, \boldsymbol{s}(q^{-1}) = \ell + nn \tag{3.41}$$

$$\deg\, Q(q^{-1}) = nb + nn - 1 \tag{3.42}$$

where nn is the order of the AR model for the noise. Here we do not assume that a parametric AR model for the noise plus interference is available, but we can use the order nn as a tuning knob. We will therefore assume that we select a number $nn \geq 0$ and use the orders from equations (3.41) and (3.42). With this choice of filter orders we will automatically get $\ell \leq ns$, which according to the structure of the optimal DFE is a sound constraint.

It is possible to have a smoothing lag, ℓ that is larger than the order of the feedforward filter. The structures of the optimal DFE however suggests that this is not meaningful. We will therefore assume that $\ell \leq ns$.

As usual we assume that all previous decisions affecting the current symbol estimate are *correct*, i.e. $\tilde{d}(t-\ell-k) = d(t-\ell-k)$, for $k = 1, \ldots, nq+1$.

We now want to adjust the equalizer coefficients to minimize the MSE criterion

$$J = E[|d(t-\ell) - \hat{d}(t-\ell)|^2]. \tag{3.43}$$

The coefficients that minimizes the criterion (3.43) can be computed by means of the Wiener-Hopf equations [35]. Define the parameter vector

$$\boldsymbol{\theta} \stackrel{\triangle}{=} \begin{bmatrix} \boldsymbol{s} & \boldsymbol{Q} \end{bmatrix} \tag{3.44}$$

where

$$\boldsymbol{s} = \begin{bmatrix} \boldsymbol{s}_0 & \boldsymbol{s}_1 & \ldots & \boldsymbol{s}_{ns} \end{bmatrix} \tag{3.45}$$

and

$$\boldsymbol{Q} = \begin{bmatrix} Q_0 & Q_2 & \ldots & Q_{nq} \end{bmatrix} \tag{3.46}$$

with $\boldsymbol{s}_i$ being the row vector coefficient of order i to the feedforward filter $\boldsymbol{s}(q^{-1}) = \boldsymbol{s}_0 + \boldsymbol{s}_1 q^{-1} + \ldots + \boldsymbol{s}_{ns} q^{-ns}$ and Q_i being the coefficients of the feedback filter.

Furthermore, define a regression vector

$$\boldsymbol{\varphi}(t) \triangleq \begin{bmatrix} \bar{\boldsymbol{y}}(t) \\ -\boldsymbol{d}_Q(t) \end{bmatrix} \tag{3.47}$$

where

$$\bar{\boldsymbol{y}} \triangleq \begin{bmatrix} \boldsymbol{y}(t) \\ \boldsymbol{y}(t-1) \\ \vdots \\ \boldsymbol{y}(t-ns) \end{bmatrix} \quad \boldsymbol{d}_Q(t) \triangleq \begin{bmatrix} d(t-\ell-1) \\ d(t-\ell-2) \\ \cdots \\ d(t-\ell-nq-1) \end{bmatrix}. \tag{3.48}$$

The filtering of the DFE in (3.40) can now, in vector form, be expressed as

$$\hat{d}(t-\ell) = \boldsymbol{\theta}\boldsymbol{\varphi}(t). \tag{3.49}$$

The parameter vector that minimizes (3.43) is now given by the solution to the Wiener-Hopf equations

$$\boldsymbol{R}_{\varphi\varphi}\boldsymbol{\theta}^H = \boldsymbol{R}_{\varphi d_\ell} \tag{3.50}$$

where

$$\boldsymbol{R}_{\varphi\varphi} \triangleq E[\boldsymbol{\varphi}(t)\boldsymbol{\varphi}^H(t)] \quad \boldsymbol{R}_{\varphi d_\ell} \triangleq E[\boldsymbol{\varphi}(t)d^H(t-\ell)]. \tag{3.51}$$

The covariance matrices $\boldsymbol{R}_{\varphi\varphi}$ and $\boldsymbol{R}_{\varphi d_\ell}$ can be expressed in the coefficients of the channel $\boldsymbol{b}(q^{-1})$ and the autocovariance function for the noise plus interference, $\boldsymbol{R}_{\boldsymbol{nn}}(k)$, see Appendix 3.A.3.

Consider now forming estimates $\hat{\boldsymbol{R}}_{\varphi\varphi}$ and $\hat{\boldsymbol{R}}_{\varphi d_\ell}$ of the matrices $\boldsymbol{R}_{\varphi\varphi}$ and $\boldsymbol{R}_{\varphi d_\ell}$ in (3.50). We can form such estimates either directly from the available training data, as will be described in Section 3.3.1, or we can form estimates by using an estimate of the channel, $\hat{\boldsymbol{b}}(q^{-1})$, and estimates of the spatio-temporal covariances, $\hat{\boldsymbol{R}}_{\boldsymbol{nn}}(k)$, for the noise plus interference in the equations in Appendix 3.A.3.

Assuming $\hat{\boldsymbol{R}}_{\varphi\varphi}$ is nonsingular, we can then compute an estimate of $\boldsymbol{\theta}$ as

$$\hat{\boldsymbol{\theta}}^H = \hat{\boldsymbol{R}}_{\varphi\varphi}^{-1}\hat{\boldsymbol{R}}_{\varphi d_\ell}. \tag{3.52}$$

From $\hat{\boldsymbol{\theta}}$ we can then extract the coefficients of the feedforward and the feedback filter according to (3.44).

The solution to (3.50) can however be expressed in a slightly more compact form as shown in [110]. We will here show this in a new and simple way. Let us return to the expression in (3.40) for the signal, $\hat{d}(t-\ell)$, prior to the decision device

$$\hat{d}(t-\ell) = \boldsymbol{s}(q^{-1})\boldsymbol{y}(t) - Q(q^{-1})\tilde{d}(t-\ell-1). \tag{3.53}$$

In matrix notation we can write $\hat{d}(t-\ell)$ as

$$\hat{d}(t-\ell) = \boldsymbol{s}\bar{\boldsymbol{y}}(t) - \boldsymbol{Q}\boldsymbol{d}_Q(t) \tag{3.54}$$

where $\boldsymbol{s}$ is defined in (3.45), $\boldsymbol{Q}$ is defined in (3.46) and $\bar{\boldsymbol{y}}(t)$ and $\boldsymbol{d}_Q(t)$ is defined in (3.48). The feedback filter is assumed to filter correct symbols delayed by $\ell+1$ time samples. The feedback filter will then, provided it is long enough, cancel all effects of symbols $d(t)$ with a delay larger than ℓ. This is the role of the feedback filter. As a result, as also noted in [54] for the scalar case, we can replace the problem of tuning the feedforward *and* feedback filters coefficients in (3.54) with the problem of tuning *only* the feedforward filter coefficients in

$$\hat{d}(t-\ell) = \boldsymbol{s}\bar{\boldsymbol{y}}_\ell(t) \tag{3.55}$$

where the signal vector $\bar{\boldsymbol{y}}_\ell(t)$ is the same signal as $\bar{\boldsymbol{y}}(t)$, except that we have removed all dependencies on symbols $d(t)$ delayed more than ℓ samples. The signal vector $\bar{\boldsymbol{y}}_\ell(t)$ can thus be expressed as

$$\bar{\boldsymbol{y}}_\ell(t) = \boldsymbol{\mathcal{B}}\boldsymbol{d}_\ell(t) + \boldsymbol{n}(t) \tag{3.56}$$

where

$$\boldsymbol{\mathcal{B}} = \begin{bmatrix} \boldsymbol{b}_0 & \dots & \boldsymbol{b}_\ell \\ \vdots & \ddots & \vdots \\ 0 & \dots & \boldsymbol{b}_0 \\ 0 & \dots & 0 \\ \vdots & & \vdots \\ 0 & \dots & 0 \end{bmatrix} \tag{3.57}$$

with $\boldsymbol{b}_k$ being the taps in the FIR filter channel (3.38) and

$$\boldsymbol{d}_\ell(t) = \begin{bmatrix} d(t) \\ \vdots \\ d(t-\ell) \end{bmatrix}. \tag{3.58}$$

Note that $\bar{\boldsymbol{y}}_\ell(t)$ does not depend on symbols $d(\tau)$ for $\tau < t - \ell$.

The coefficient vector for the MMSE optimal feedforward filter, $\boldsymbol{s}$, can thus be found as the solution to the Wiener-Hopf equations

$$\boldsymbol{R}_{\bar{\boldsymbol{y}}_\ell \bar{\boldsymbol{y}}_\ell} \boldsymbol{s}^H = \boldsymbol{R}_{\bar{\boldsymbol{y}}_\ell d_\ell} \tag{3.59}$$

where

$$\boldsymbol{R}_{\bar{\boldsymbol{y}}_\ell \bar{\boldsymbol{y}}_\ell} = E[\bar{\boldsymbol{y}}(t)\bar{\boldsymbol{y}}^H(t)] \tag{3.60}$$

and

$$\boldsymbol{R}_{\bar{\boldsymbol{y}}_\ell d_\ell} = E[\bar{\boldsymbol{y}}(t) d^H(t-\ell)]. \tag{3.61}$$

Using (3.56) and the assumption that $d(t)$ is white and uncorrelated with $\boldsymbol{n}(t)$, we can rewrite the system of equations (3.59) for the feedforward filter coefficients, $\boldsymbol{s}$ defined in (3.45), as

$$(\boldsymbol{\mathcal{B}}\boldsymbol{\mathcal{B}}^H + \boldsymbol{\mathcal{R}}_{\bar{\boldsymbol{n}}\bar{\boldsymbol{n}}})\boldsymbol{s}^H = \boldsymbol{\mathcal{B}}_\ell \tag{3.62}$$

where $\boldsymbol{\mathcal{B}}$ is defined in (3.57) and

$$\boldsymbol{\mathcal{R}}_{\bar{\boldsymbol{n}}\bar{\boldsymbol{n}}} = \begin{bmatrix} \boldsymbol{R}_{\boldsymbol{nn}}(0) & \ldots & \boldsymbol{R}_{\boldsymbol{nn}}(ns) \\ \vdots & \ddots & \vdots \\ \boldsymbol{R}_{\boldsymbol{nn}}(-ns) & \ldots & \boldsymbol{R}_{\boldsymbol{nn}}(0) \end{bmatrix} \quad \boldsymbol{\mathcal{B}}_\ell = \begin{bmatrix} \boldsymbol{b}_\ell \\ \boldsymbol{b}_{\ell-1} \\ \vdots \\ \boldsymbol{b}_0 \\ 0 \\ \vdots \\ 0 \end{bmatrix}. \tag{3.63}$$

The $M \times M$ matrices $\boldsymbol{R}_{\boldsymbol{nn}}(k)$ in $\boldsymbol{\mathcal{R}}_{\bar{\boldsymbol{n}}\bar{\boldsymbol{n}}}$ are evaluations of the autocovariance function of the noise plus interference defined in (2.108).

Given the coefficients of feedforward filter $\boldsymbol{s} = \begin{bmatrix}\boldsymbol{s}_0 & \dots & \boldsymbol{s}_{ns}\end{bmatrix}$, the coefficients of feedback filter, can using (3.37) with $nn = 0$, be computed as

$$\begin{bmatrix} Q_0^H \\ \vdots \\ Q_{nq}^H \end{bmatrix} = \begin{bmatrix} \boldsymbol{b}_{\ell+1}^H & \cdots & \boldsymbol{b}_1^H \\ \vdots & & \vdots \\ \boldsymbol{b}_{\ell+nb}^H & \cdots & \boldsymbol{b}_{nb}^H \end{bmatrix} \begin{bmatrix} \boldsymbol{s}_1^H \\ \vdots \\ \boldsymbol{s}_\ell^H \end{bmatrix} \tag{3.64}$$

where $\boldsymbol{b}_i = 0$ if $i > nb$.

3.2.4 Multidimensional Matched Filter DFE

Another option for implementation of a DFE is the multidimensional matched filter DFE (MMF-DFE). Here we utilize the fact that if we could allow an infinite smoothing lag, then the optimal DFE would consist of a multidimensional matched filter followed by a scalar DFE with an infinite smoothing lag [83]. The multidimensional matched filter can take spatially and temporally colored noise into consideration. In reality we cannot accept an infinite smoothing lag in the scalar DFE. With a finite smoothing lag the MMF-DFE is theoretically suboptimal. However, as we will see in the simulations in Section 3.4, in practice it *may* work just as well as the space-time DFE's in Sections 3.2.1 – 3.2.3.

Consider the received signal (2.2) at the antenna array

$$\boldsymbol{y}(t) = \boldsymbol{b}(q^{-1})d(t) + \boldsymbol{n}(t). \tag{3.65}$$

We assume that the noise plus interference term, $\boldsymbol{n}(t)$, can be spatially and temporally colored and we define a double-sided transfer operator matrix, $\boldsymbol{R_{nn}}(q, q^{-1})$, from its spatio-temporal covariance matrix function, $\boldsymbol{R_{nn}}(k)$, as[3]

$$\boldsymbol{R_{nn}}(q, q^{-1}) = \sum_{k=-\infty}^{\infty} \boldsymbol{R_{nn}}(k) q^{-k} = \sum_{k=-\infty}^{\infty} E[\boldsymbol{n}(t)\boldsymbol{n}^H(t-k)]q^{-k}. \tag{3.66}$$

For simplicity we will sometimes refer to the matrix $\boldsymbol{R_{nn}}(q, q^{-1})$ as the "spatio-temporal covariance matrix".

[3] A more traditional notation would be to replace q with the variable z. $\boldsymbol{R_{nn}}(z, z^{-1})$ would then represent the spatio-temporal *spectrum* for $\boldsymbol{n}(t)$. Using the q operator however simplifies the presentation.

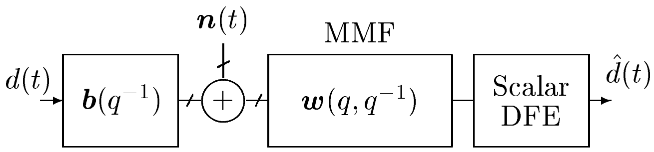


Figure 3.6: MMF-DFE

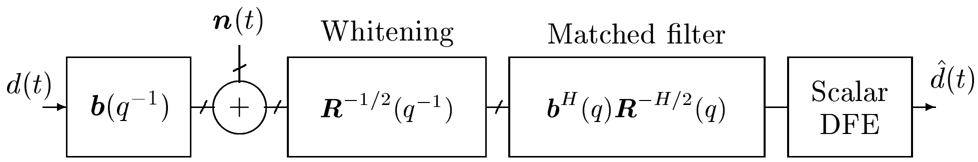


Figure 3.7: MMF-DFE with the MMF decomposed into a whitening filter and a matched filter.

Given this channel, and allowing an infinite smoothing lag, one can show that the MSE optimal DFE consists of a multidimensional matched filter

$$\boldsymbol{w}(q, q^{-1}) = \boldsymbol{b}^H(q)\boldsymbol{R}^{-1}(q, q^{-1}). \tag{3.67}$$

followed by a scalar DFE with a possibly infinite smoothing lag. In [83] this result is shown in continuous time. The structure of the MMF-DFE is depicted in Figure 3.6. It should be noted that if the smoothing lag is *finite*, then the MMSE optimal DFE cannot in general be decoupled into a matched filter followed by a DFE [90, 106].

Let us perform a spectral factorization on $\boldsymbol{R}_{\boldsymbol{nn}}(q, q^{-1})$, i.e. let

$$\boldsymbol{R}_{\boldsymbol{nn}}(q, q^{-1}) = \boldsymbol{R}_{\boldsymbol{nn}}^{1/2}(q^{-1})\boldsymbol{R}_{\boldsymbol{nn}}^{H/2}(q) \tag{3.68}$$

where $\boldsymbol{R}_{\boldsymbol{nn}}^{1/2}(q^{-1})$ is stable and causal [47, 40, 36]. We can now decompose the multidimensional matched filter into a noise whitening filter followed by a filter matched to the channel including the noise whitening filter, as shown in Figure 3.7.

Note that if the noise process, $\boldsymbol{n}(t)$, is modeled by an AR-process

$$\boldsymbol{n}(t) = \boldsymbol{N}^{-1}(q^{-1})\boldsymbol{M}_0\boldsymbol{v}(t) \tag{3.69}$$

where $\boldsymbol{N}^{-1}(q^{-1})$ is an $M \times M$ polynomial matrix, $\boldsymbol{M}_0$ is a constant matrix, and $\boldsymbol{v}(t)$ is a $m \times 1$ white innovation process, then the multi-dimensional matched filter takes on the simple form of a row vector of non-causal FIR-filters

$$\boldsymbol{w}(q, q^{-1}) = \boldsymbol{b}^H(q)\boldsymbol{N}^H(q)\boldsymbol{M}_0^{-H}\boldsymbol{M}_0^{-1}\boldsymbol{N}(q^{-1}). \tag{3.70}$$

Note that the MMF-filter have an impulse response with finite extent, both forward and backward in time. An AR model for the noise is thus useful model since it translates into an FIR multi-dimensional matched filter (when the channel for the desired signal is an FIR channel).

A potential disadvantage with this approach is that the multidimensional matched filter increases the length of the intersymbol interference before the scalar DFE is applied. This can make a channel that is difficult to equalize even more demanding. Also, as mentioned in the beginning of this subsection, in theory it requires an infinite smoothing lag. This is of course not desirable and in practice we have to design the scalar DFE with a finite smoothing lag. Furthermore we need to truncate the MMF-filter if the noise-whitened channel has an impulse response with infinite duration. However, as mentioned in the beginning, we will see in the simulations of Sections 3.2.1 – 3.2.3, that an MMF combined with a properly tuned scalar DFE may in practice perform just as well as the other space-time DFE's studied. To answer this question in more depth it is however necessary to perform more extensive studies covering a wider range of possible channels and scenarios.

3.3 Some Tuning Options

To implement the optimal fixed order FIR-DFE as described in Section 3.2.3 we need channel and noise plus interference covariance estimates of high quality. Since these estimates are not normally available a priori, several different tuning options can be considered. We will here discuss some of these options.

3.3.1 Directly Tuned Decision Feedback Equalizer (D-DFE)

As we saw in Section 3.2.2, when the channel for the desired signal is an FIR-channel and the noise plus interference is temporally white (or can be modeled by an autoregressive noise model), the MMSE optimum structure for the DFE has FIR filters in both the feedforward and the feedback branches. Even if this is not the case we can restrict the DFE to contain only FIR filters of predetermined orders, *cf* Section 3.2.3.

We will here restrict the DFE to a FIR filter structure, but instead of using an indirect approach for the tuning as in Section 3.2.3, we will consider tuning of the FIR filter coefficients of the DFE directly from data.

The estimate of the symbol prior to the decision device is given by

$$\hat{d}(t-\ell) = \boldsymbol{s}(q^{-1})y(t) - Q(q^{-1})\tilde{d}(t-\ell-1) \tag{3.71}$$

where $\boldsymbol{s}(q^{-1})$ is a $M \times 1$ polynomial row vector of order ns and $Q(q^{-1})$ is a polynomial of order nq, as in Figure 3.5.

As in Section 3.2.3, we want to minimize the MSE criterion

$$J = E[|d(t-\ell) - \hat{d}(t-\ell)|^2] \ . \tag{3.72}$$

with respect to the coefficients of $\boldsymbol{s}(q^{-1})$ and $Q(q^{-1})$. The DFE parameters, $\boldsymbol{\theta}$, defined in (3.44) that minimizes J are, as in (3.50), given by the solution to the system of equations

$$\boldsymbol{R}_{\varphi\varphi}\boldsymbol{\theta}^H = \boldsymbol{R}_{\varphi d_\ell} \tag{3.73}$$

We can now replace $\boldsymbol{R}_{\varphi\varphi}$ and $\boldsymbol{R}_{\varphi d_\ell}$ with the sample matrix estimates

$$\hat{\boldsymbol{R}}_{\varphi\varphi} = \frac{1}{(t_{max} - t_{min} + 1)}\Big(\sum_{t=t_{min}}^{t_{max}} \boldsymbol{\varphi}(t)\boldsymbol{\varphi}^H(t)\Big) + \sigma_a^2 \boldsymbol{I}_{\bar{\boldsymbol{y}}\bar{\boldsymbol{y}}} \tag{3.74}$$

and

$$\hat{\boldsymbol{R}}_{\varphi d_\ell} = \frac{1}{(t_{min} - t_{max} + 1)} \sum_{t=t_{min}}^{t_{max}} \boldsymbol{\varphi}(t) d^H(t-\ell) \tag{3.75}$$

where $\boldsymbol{\varphi}$ is the regression vector defined in (3.47). The matrix $I_{\bar{\boldsymbol{y}}\bar{\boldsymbol{y}}}$ has an $M(ns+1) \times M(ns+1)$ identity matrix in its upper left corner and zeros elsewhere. The term $\sigma_a^2 I_{\bar{\boldsymbol{y}}\bar{\boldsymbol{y}}}$ is added to regularize the matrix $\hat{\boldsymbol{R}}_{\boldsymbol{\varphi}\boldsymbol{\varphi}}$ in case it is singular or ill-conditioned. Asymptotically, with an infinite number of data (and some noise), σ_a^2 should be set to zero. However, with a finite amount of data and/or a limited computational accuracy, the solution can be improved by a good choice of σ_a^2. The matrix $\hat{\boldsymbol{R}}_{\boldsymbol{\varphi}\boldsymbol{\varphi}}$ can, of course, not be inverted without this regularization if the number of training relations, $t_{max} - t_{min} + 1$, are fewer than the number of parameters.

The equalizer coefficients in $\boldsymbol{\theta}$ can now be estimated as

$$\hat{\boldsymbol{\theta}}^H = \hat{\boldsymbol{R}}_{\boldsymbol{\varphi}\boldsymbol{\varphi}}^{-1} \hat{\boldsymbol{R}}_{\boldsymbol{\varphi} d_\ell}. \tag{3.76}$$

In the estimates of $\boldsymbol{R}_{\boldsymbol{\varphi}\boldsymbol{\varphi}}$ and $\boldsymbol{R}_{\boldsymbol{\varphi} d_\ell}$ in (3.74) and (3.75), the summation indices t_{min} and t_{max} are chosen such that the training data is optimally utilized. It can sometimes be advantageous to add zeros at the beginning and at the end of the training sequence to increase the number of possible terms in the sample matrix estimates.

The directly tuned DFE has the advantage that it will always attempt to achieve the best equalization performance possible given the specified structure and the available number of parameters. The direct tuning of the DFE also minimizes a more relevant criterion than the indirectly tuned DFE does [11]. The directly tuned space-time DFE, however, has the disadvantage that it can have a large number of parameters to tune on a small number of training symbols when many receiver antennas are used. Since all parameters are tuned jointly, the system of equations can easily become under-determined. In this case we need to regularize it to obtain a meaningful solution. This will be apparent in the simulations below.

3.3.2 Indirectly Tuned DFE (I-DFE)

As in the previous section we will here assume that we are constraining the DFE to have only FIR filters in the feedforward and feedback filters. The structure is thus the same as in the previous section and as in Figure 3.5. However, instead of tuning the coefficients using the data directly we will here tune the equalizer coefficients via estimates of the channel and the covariance of the noise plus interference.

Consider the equation for the coefficients of the feedforward filter in (3.62)

$$(\boldsymbol{\mathcal{B}}\boldsymbol{\mathcal{B}}^H + \boldsymbol{\mathcal{R}}_{\bar{n}\bar{n}})\boldsymbol{s}^H = \boldsymbol{\mathcal{B}}_\ell. \tag{3.77}$$

The matrices $\boldsymbol{\mathcal{B}}$ and $\boldsymbol{\mathcal{B}}_\ell$ in (3.57) and (3.63) can be formed from the channel estimates $\hat{\boldsymbol{b}}(q^{-1})$ as

$$\hat{\boldsymbol{\mathcal{B}}} = \begin{bmatrix} \hat{\boldsymbol{b}}_0 & \dots & \hat{\boldsymbol{b}}_\ell \\ \vdots & \ddots & \vdots \\ 0 & \dots & \hat{\boldsymbol{b}}_0 \\ 0 & \dots & 0 \\ \vdots & & \vdots \\ 0 & \dots & 0 \end{bmatrix} \qquad \hat{\boldsymbol{\mathcal{B}}}_\ell = \begin{bmatrix} \hat{\boldsymbol{b}}_\ell \\ \hat{\boldsymbol{b}}_{\ell-1} \\ \vdots \\ \hat{\boldsymbol{b}}_0 \\ 0 \\ \vdots \\ 0 \end{bmatrix} \tag{3.78}$$

where the channel estimate, $\hat{\boldsymbol{b}}(q^{-1})$, can, for example, be obtained with any of the methods presented in Chapter 2.

As an estimate of the spatio-temporal covariance matrix,

$$\boldsymbol{\mathcal{R}}_{\bar{n}\bar{n}} = \begin{bmatrix} \boldsymbol{R}_{\boldsymbol{nn}}(0) & \dots & \boldsymbol{R}_{\boldsymbol{nn}}(ns) \\ \vdots & \ddots & \vdots \\ \boldsymbol{R}_{\boldsymbol{nn}}(-ns) & \dots & \boldsymbol{R}_{\boldsymbol{nn}}(0) \end{bmatrix} \tag{3.79}$$

we here propose the use of a regularized sample matrix estimate

$$\hat{\boldsymbol{\mathcal{R}}}_{\bar{n}\bar{n}} = \frac{1}{t_{max} - t_{min} + 1} \sum_{t=t_{min}}^{t_{max}} \hat{\bar{\boldsymbol{n}}}(t)\hat{\bar{\boldsymbol{n}}}^H(t) + \sigma_a^2 I \tag{3.80}$$

where

$$\hat{\bar{\boldsymbol{n}}} = \left[\hat{\boldsymbol{n}}^T(t) \;\; \hat{\boldsymbol{n}}^T(t-1) \;\; \dots \;\; \hat{\boldsymbol{n}}^T(t-ns)\right]^T \tag{3.81}$$

and where $\hat{\boldsymbol{n}}(t-k)$ are lagged residuals from the FIR channel estimation

$$\hat{\boldsymbol{n}}(t) = \boldsymbol{y}(t) - \hat{\boldsymbol{b}}(q^{-1})d(t). \tag{3.82}$$

The term $\sigma_a^2 I$ is the regularization term adding σ_a^2 to the diagonal elements of the spatio-temporal noise plus interference covariance matrix.

If the number of parameters in the feedforward filter of the spatio-temporal DFE is large compared to the number of training symbols available, then the matrix

$$(\hat{\mathcal{B}}\hat{\mathcal{B}}^H + \hat{\mathcal{R}}_{\bar{n}\bar{n}}) \tag{3.83}$$

can be ill-conditioned or singular. The rank of the term $\hat{\mathcal{B}}\hat{\mathcal{B}}^H$ can never be larger than $\ell + 1$ since this is the number of columns in $\hat{\mathcal{B}}$. With a large number of antennas it is likely that ℓ is much smaller than the number of parameters in the feedforward filter, $M(ns+1)$, and thus in this case the term $\hat{\mathcal{B}}\hat{\mathcal{B}}^H$ only contributes marginally to the rank of the matrix $\hat{\mathcal{B}}\hat{\mathcal{B}}^H + \hat{\mathcal{R}}_{\bar{n}\bar{n}}$. Without the diagonal term, $\sigma_a^2 I$, the rank of $\hat{\mathcal{R}}_{\bar{n}\bar{n}}$ will be limited by, i.e. less than, the number of terms in the sum in (3.80). The number of terms is here typically determined by the effective length of the training sequence[4] Thus, if the number of parameters in $\boldsymbol{s}(q^{-1})$, i.e. Mns, is comparable to, or larger than, the effective length of the training sequence, and we don't add the term $\sigma_a^2 \boldsymbol{I}$, then the matrix $(\hat{\mathcal{B}}\hat{\mathcal{B}}^H + \hat{\mathcal{R}}_{\bar{n}\bar{n}})$ can, or will, be ill-conditioned and even singular. By adding the diagonal term, $\sigma_a^2 I$, we can improve the condition number for the matrix and compute a meaningful solution. This solution will, similar to the D-DFE in Section 3.3.1, suffer from having a large number of coefficients to tune. However, it performs better than the D-DFE in the simulations below. The reason for this is probably that the "signal part", the $\hat{\mathcal{B}}\hat{\mathcal{B}}^H$ part of the matrix $(\hat{\mathcal{B}}\hat{\mathcal{B}}^H + \hat{\mathcal{R}}_{\bar{n}\bar{n}})$, is given by a model that is defined by a relatively low number of FIR channel parameters.

3.3.3 Indirectly Tuned DFE with Spatial-Only Interference Cancellation (IS-DFE)

The problem with the ill-conditioned estimate of the spatio-temporal noise plus interference covariance matrix, $\hat{\mathcal{R}}_{\bar{n}\bar{n}}$, can be avoided by replacing it

[4] With the effective length of the training sequence we refer to the number of training relations that can be formed with the method we are using; here $t_{max} - t_{min}$.

with a spatial-only noise plus interference covariance matrix

$$\hat{\boldsymbol{\mathcal{R}}}_{\boldsymbol{nn}} = \begin{bmatrix} \hat{\boldsymbol{R}}_{\boldsymbol{nn}} & & \mathbf{0} \\ & \ddots & \\ \mathbf{0} & & \hat{\boldsymbol{R}}_{\boldsymbol{nn}} \end{bmatrix} \tag{3.84}$$

where $\hat{\boldsymbol{R}}_{\boldsymbol{nn}}$ is an estimate of the *spatial* interference plus noise covariance matrix, i.e. the covariance matrix function for lag zero

$$\hat{\boldsymbol{R}}_{\boldsymbol{nn}} = \frac{1}{t_{max} - t_{min} + 1} \sum_{t=t_{min}}^{t_{max}} \hat{\boldsymbol{n}}(t)\hat{\boldsymbol{n}}^H(t) \tag{3.85}$$

where $\hat{\boldsymbol{n}}(t)$ is computed as in (3.82).

The spatial only covariance matrix estimate $\hat{\boldsymbol{\mathcal{R}}}_{\boldsymbol{nn}}$ will be much better conditioned. As long as the effective length of the training sequence[5] symbols used is large compared to the number of antennas, regularization will typically not be required. This estimate will, of course, not contain any information about the temporal correlations of the noise plus interferers. However, if the desired signal and the interferers can be separated spatially, then this approach may provide adequate performance.

3.3.4 Indirectly Tuned DFE with an AR Noise Model (AR-DFE)

For the AR-DFE we will estimate an FIR channel with an AR model for the noise plus interference, as described in Section 2.9, and design an MMSE optimal equalizer using this model as described in Section 3.2.2.

Let us first, using the method described in Section 2.9, jointly compute an estimate, $\hat{\boldsymbol{b}}_{\boldsymbol{N}}(q^{-1})$, of the channel, $\boldsymbol{b}_{\boldsymbol{N}}(q^{-1}) = \boldsymbol{N}(q^{-1})\boldsymbol{b}(q^{-1})$, and an estimate, $\hat{\boldsymbol{N}}(q^{-1})$ of the noise whitening filter $\boldsymbol{N}(q^{-1})$ in (2.118). Using the residuals, $\boldsymbol{r}(t)$, from this estimation we can form an estimate of the matrix $\boldsymbol{M}_0$ in (2.118) as

$$\hat{\boldsymbol{M}}_0 = \left(\frac{1}{t_{max} - t_{min} + 1} \sum_{t=t_{min}}^{t_{max}} \boldsymbol{r}(t)\boldsymbol{r}^H(t) \right)^{-1/2} . \tag{3.86}$$

[5]The number of training relations that can be formed with the method we are using; here $t_{max} - t_{min}$.

We can now form an estimate of the noise whitened channel

$$\boldsymbol{b}'(q^{-1}) = \boldsymbol{M}_0^{-1}\boldsymbol{N}(q^{-1})\boldsymbol{b}(q^{-1}) \tag{3.87}$$

as

$$\hat{\boldsymbol{b}}'(q^{-1}) = \hat{\boldsymbol{M}}_0^{-1}\hat{\boldsymbol{b}}_{\boldsymbol{N}}(q^{-1}). \tag{3.88}$$

Given this estimate we can compute an estimate $\hat{\boldsymbol{s}}_0(q^{-1})$ of the polynomial $\boldsymbol{s}_0(q^{-1})$ by solving (3.32), using the the coefficients of $\hat{\boldsymbol{b}}'(q^{-1})$ in place of the coefficients of $\boldsymbol{b}(q^{-1})$. The estimated feedforward filter can then be formed as

$$\hat{\boldsymbol{s}}(q^{-1}) = \hat{\boldsymbol{s}}_0(q^{-1})\hat{\boldsymbol{M}}_0^{-1}\hat{\boldsymbol{N}}(q^{-1}) \tag{3.89}$$

Finally the coefficients of the estimated feedback filter, $\hat{Q}(q^{-1})$, can be computed by using the estimates of the polynomial $\boldsymbol{s}_0(q^{-1})$ and the estimate of the noise whitened channel $\boldsymbol{b}'(q^{-1})$ in (3.37).

This DFE will perform some spatio-temporal interference suppression. Depending on the order of the AR noise model it will perform different amounts of temporal filtering. The DFE can, without a significant increase in complexity, handle an AR noise model of relatively high degree. However, depending on the number of antennas and the amount of training data available, the order of the AR noise model may be limited by the channel estimation procedure.

3.3.5 Indirectly Tuned MMF-DFE with Spatial-Only Interference Cancellation (IS-MMF-DFE)

The IS-MMF-DFE[6] is an indirectly tuned MMF-DFE utilizing only the spatial spectrum of the noise plus interference. An estimate, $\hat{\boldsymbol{b}}(q^{-1})$, of the channel for the desired signal can be computed with, for example, one of the methods described in Chapter 2. The spatial noise plus interference spectrum is estimated as a sample matrix estimate, as in Section 2.8, formed from the residuals, $\hat{\boldsymbol{n}}(t) = \boldsymbol{y}(t) - \hat{\boldsymbol{b}}(q^{-1})d(t)$ from the channel estimation

$$\hat{\boldsymbol{R}}_{\boldsymbol{nn}} = \hat{\boldsymbol{R}}_{\boldsymbol{nn}}(0) = \frac{1}{t_{max} - t_{min} + 1} \sum_{t=t_{min}}^{t_{max}} \hat{\boldsymbol{n}}(t)\hat{\boldsymbol{n}}^H(t). \tag{3.90}$$

[6]The "IS" stands for Indirect tuning with Spatial-only interference cancellation.

Using these estimates, an estimated multidimensional matched filter can be formed as

$$\hat{\boldsymbol{w}}(q,q^{-1}) = \hat{\boldsymbol{b}}^H(q)\hat{\boldsymbol{R}}_{\boldsymbol{nn}}^{-1} \tag{3.91}$$

The output from this MMF-filter will then be processed by a scalar DFE.

This scalar DFE can, for example, be tuned either with a direct method using the output from the MMF during the training sequence or indirectly via estimates of the total channel, from the transmitted symbols to the signal after the MMF. This new channel can be obtained either by re-estimating the channels after the MMF or by simply computing the new channel from the initial channel and covariance estimates and the computed MMF.

When tuning the scalar DFE after the MMF, it is a good idea to utilize the temporal spectrum of the noise after the MMF. With direct tuning the noise spectrum is automatically taken into account. When using the indirect methods, the noise spectrum can be re-estimated, as a sample matrix estimate of the temporal noise covariance matrix involving the relative time lags, formed from the channel estimation residuals. It can also be formed by combining the already estimated spatial noise plus interference covariance matrix, prior to the MMF, with the coloring of the MMF filter itself.

The processing in the IS-MMF-DFE can be summarized as:

1. Estimate the channels. Denote the vector of channel estimates $\hat{\boldsymbol{b}}(q^{-1})$.
2. Estimate the spatial noise plus interference covariance matrix, for example as a sample matrix estimate formed from the residuals $\hat{\boldsymbol{n}}(t)$, defined by (3.82), as
$$\hat{\boldsymbol{R}}_{\boldsymbol{nn}} = \frac{1}{t_{max} - t_{min} + 1} \sum_{t=t_{min}}^{t_{max}} \hat{\boldsymbol{n}}(t)\hat{\boldsymbol{n}}^H(t). \tag{3.92}$$
3. Form the MMF filter as $\hat{\boldsymbol{w}}(q,q^{-1}) = \hat{\boldsymbol{b}}^H(q)\hat{\boldsymbol{R}}_{\boldsymbol{nn}}^{-1}$.
4. Compute a scalar indirect FIR-DFE, with a suitable smoothing lag, as in Section 3.2.3 (equations (3.62) and (3.64)) using the channel from the transmitted symbols to the MMF output,

$\hat{\boldsymbol{b}}^H(q)\hat{\boldsymbol{R}}_{\boldsymbol{nn}}^{-1}\hat{\boldsymbol{b}}(q^{-1})$, and the noise spectrum, $\boldsymbol{w}(q,q^{-1})\hat{\boldsymbol{R}}_{\boldsymbol{nn}}\boldsymbol{w}^H(q,q^{-1})$. The channel and noise spectrum will, of course, have to be corrected with the appropriate delays so that the channel is causal. The smoothing lag should preferably be chosen large enough to cover this delay, plus the major part of the energy caused by a symbol transmitted over the channel. Alternatively the temporal spectrum of the noise plus interference can be re-estimated by forming a sample matrix estimate of the temporal covariance matrix from the residuals, as briefly described above.

An advantage with the IS-MMF-DFE approach over the IS-DFE approach of Section 3.3.3 is that it is less computationally intensive to tune. In the IS-DFE we have to invert the fairly large $M(ns+1) \times M(ns+1)$ matrix, $(\boldsymbol{\mathcal{B}}\boldsymbol{\mathcal{B}}^H + \boldsymbol{\mathcal{R}}_{\hat{\boldsymbol{n}}\hat{\boldsymbol{n}}})$, whereas for the IS-MMF-DFE only the smaller $M \times M$ matrix $\hat{\boldsymbol{R}}_{\boldsymbol{nn}}$ has to be inverted.

3.3.6 Indirectly Tuned MMF-DFE with an AR Noise Model (AR-MMF-DFE)

As for the AR-DFE in Section 3.3.4 we will here estimate an FIR channel with an AR model for the noise plus interference, as described in Section 2.9.

We can thus, using the method described in Section 2.9, jointly compute an estimate, $\hat{\boldsymbol{b}}_{\boldsymbol{N}}(q^{-1})$, of the channel, $\boldsymbol{b}_{\boldsymbol{N}} = \boldsymbol{N}(q^{-1})\boldsymbol{b}(q^{-1})$, and an estimate, $\hat{\boldsymbol{N}}(q^{-1})$ of the noise whitening filter $\boldsymbol{N}(q^{-1})$ in (2.118). Using the residuals, $\boldsymbol{r}(t)$, from this estimation we can form an estimate of the matrix $\boldsymbol{M}_0$ in (2.118) as

$$\hat{\boldsymbol{M}}_0 = \left(\frac{1}{t_{max} - t_{min} + 1} \sum_{t=t_{min}}^{t_{max}} \boldsymbol{r}(t)\boldsymbol{r}^H(t) \right)^{-1/2}. \tag{3.93}$$

Using these estimates we form an estimate of the noise whitened channel as

$$\hat{\boldsymbol{b}}'(q^{-1}) = \hat{\boldsymbol{M}}_0^{-1}\hat{\boldsymbol{b}}_{\boldsymbol{N}}(q^{-1}). \tag{3.94}$$

An estimate of the multidimensional matched filter can now be formed as

$$\hat{\boldsymbol{w}}(q,q^{-1}) = \hat{\boldsymbol{b}}'^H(q)\hat{\boldsymbol{M}}_0^{-1}\hat{\boldsymbol{N}}(q^{-1}). \tag{3.95}$$

The output from this MMF-filter will then be processed by a scalar DFE.

This scalar DFE can be tuned more or less exactly as the scalar DFE used in the IS-MMF-DFE of Section 3.3.5.

The processing in the AR-MMF-DFE can be summarized as:

1. Form the estimates $\hat{\boldsymbol{b}}_{\boldsymbol{N}}(q^{-1})$ and $\hat{\boldsymbol{N}}(q^{-1})$ of the corresponding entities in (2.118) using the method of Section 2.9.

2. Form the estimate $\hat{\boldsymbol{M}}_0$ using (3.93).

3. Form an estimate of the noise whitened channel as $\hat{\boldsymbol{b}}'(q^{-1}) = \hat{\boldsymbol{M}}_0^{-1}\hat{\boldsymbol{b}}_{\boldsymbol{N}}(q^{-1})$.

4. Form the MMF filter as $\hat{\boldsymbol{w}}(q, q^{-1}) = \hat{\boldsymbol{b}}'^{H}(q)\hat{\boldsymbol{M}}_0^{-1}\hat{\boldsymbol{N}}(q^{-1})$

5. Compute a scalar indirect FIR-DFE, with a suitable smoothing lag, as in Section 3.2.3 (equations (3.62) and (3.64)) using the channel from the transmitted symbols to the MMF output, $\hat{\boldsymbol{b}}'^{H}(q)\hat{\boldsymbol{b}}'(q^{-1})$, and the noise spectrum, $\hat{\boldsymbol{b}}'^{H}(q)\hat{\boldsymbol{b}}'(q^{-1})$. As for the IS-MMF-DFE, the channel and noise spectrum will, of course, have to be corrected with the appropriate delays so that the channel is causal. The smoothing lag should preferably be chosen large enough to cover this delay, plus the major part of the energy caused by a symbol transmitted over the channel. Alternatively the temporal spectrum of the noise plus interference can be re-estimated by forming a sample matrix estimate of the temporal covarance matrix from the residuals, as briefly described above.

Again, an advantage with the AR-MMF-DFE approach over the AR-DFE approach of Section 3.3.4 is that it is less computationally intensive to tune. The reason for this is that when tuning the AR-DFE we have to invert the fairly large $M(ns+1) \times M(ns+1)$ matrix, $(\boldsymbol{\mathcal{B}}'\boldsymbol{\mathcal{B}}'^{H} + I)$, in (3.34).

3.4 Simulations

An Eight Element Antenna Array

The performance of the D-, I-, IS-DFE and the MMF-DFE will here be illustrated by some simulations.

A circular array consisting of eight antennas was used. The desired signal is impinging on the array from the directions $\alpha = 0$, 30 -60 and 180 degrees, through the channels $B(q^{-1}) = 1 + 0.5q^{-1}$, $0.5q^{-1} + 0.8q^{-2}$, $0.5q^{-2} + 0.2q^{-3}$ and $0.2q^{-3} + 0.3q^{-4}$ respectively. Three co-channel interferers are impinging on the array from the directions $\alpha_{co} = 135$, -30 and 235 degrees respectively with the channels $\gamma(0.5 - 0.5q^{-1})$, $\gamma(0.7q^{-1} - 0.4q^{-2})$ and $\gamma(-0.5q^{-2} + 0.5q^{-3})$. The constant γ was varied to achieve different signal to interference ratios. Independent white noise was also added to the simulated received signal to vary the signal to noise ratio. Furthermore, the length of the training sequence was varied to study how it affected the performance of the equalizers. The transmitted symbols, $d(t)$ had the values $d(t) = \pm 1$.

For the indirect algorithms, utilizing a channel estimate, the channel was estimated with a simple least squares method as described in Section 2.2. When required the space-time or space-only spectrum of the noise plus interference was estimated as a sample matrix estimate formed from the residuals of the channel estimation.

The smoothing lag and the order of the feedforward filter for the D-, I- and IS-DFE's were chosen to $\ell = ns = 3$ and the order of the feedback filter was chosen to $nq = 3$.

The MMF-DFE was tuned as suggested for the IS-MMF-DFE in Section 3.3.5. The MMF-filter had 5 taps and used a spatial-only estimate of the noise plus interference spectrum. The scalar DFE following the MMF-filter had 9 taps in the feedforward filter (with a smoothing lag of 8) and 6 taps in the feedback filter. The scalar DFE was tuned utilizing the transformed channel after the MMF-filter and the estimated temporal noise plus interference spectrum, all as described in Section 3.3.5.

The regularizing constant σ_a^2 for the D- and I-DFE was set equal to the variance of the noise. This value was found to work well in these examples. Note, however, that for training sequences that are long compared to the number of jointly tuned parameters, this value should most likely be decreased. In a real system we would of course have to use an estimate of the noise variance since the true noise variance would not be known.

In Figure 3.8 we can see the resulting BER for the different equalizers as a function of the length of the training sequence. The equalizers attempting to utilize the full required space-time spectrum of the noise plus interference, the D- and the I-DFE, perform poorly for short training sequences. The IS-DFE and the IS-MMF-DFE here have almost equal performance and they are superior to the D- and I-DFE's , except for very long training sequences. The reason for this difference in performance is here that the equations for computing the parameters of the IS-DFE and the IS-MMF-DFE are much more well conditioned and reliable as a result of not attempting to utilize the full estimated space-time spectrum of the noise plus interference.

In Figure 3.9 we can see the resulting BER for the different equalizers as a function of the signal to interference ratio. The training sequence was here 26 symbols long and the signal to noise ratio was 7 dB. Again we see that the IS-MMF-DFE and the IS-DFE have an almost equal performance. We can note here that both the IS-DFE and the IS-MMF-DFE is successful in suppressing the strong interference. The D- and I-DFE's are less successful as the power of the interference increases. However, note that at high SIR the D- and I-DFE algorithms are gaining on the IS-DFE and the IS-MMF-DFE. When the strength of the interferer is reduced the regularization in the D- and I-DFE's is becoming more adequate.

A "One Element" Antenna Array

We here illustrate the performance of different versions of the AR-DFE of Section 3.3.4 with a simulation example. Apart from using different orders of the AR noise model we also show the effect of processing the real and imaginary parts of the signal separately. Finally we apply bootstrapping, as explained in Section 2.7.

We use a scenario with only two received signals. These two signals can for example be two polarizations of the same signal. In this case we would have

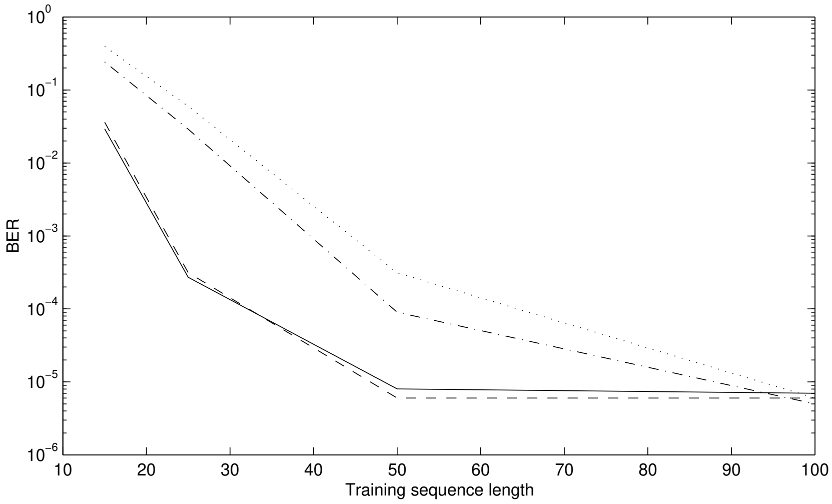


Figure 3.8: BER as a function of the training sequence length. SNR: 7dB, SIR=-18dB. IS-DFE (solid), IS-MMF-DFE (dashed), I-DFE (dash-dotted) and D-DFE (dotted).

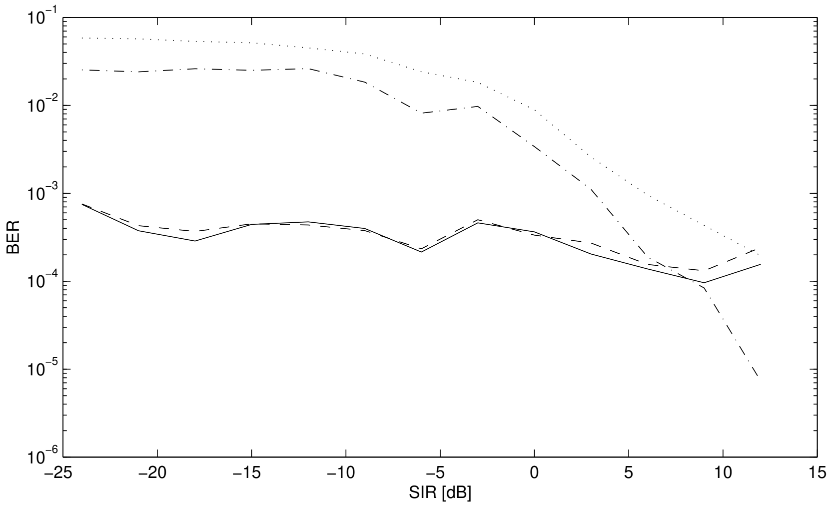


Figure 3.9: BER as a function of SIR. SNR: 7dB. Number of training symbols: 26. IS-DFE (solid), IS-MMF-DFE (dashed), I-DFE (dash-dotted) and D-DFE (dotted).

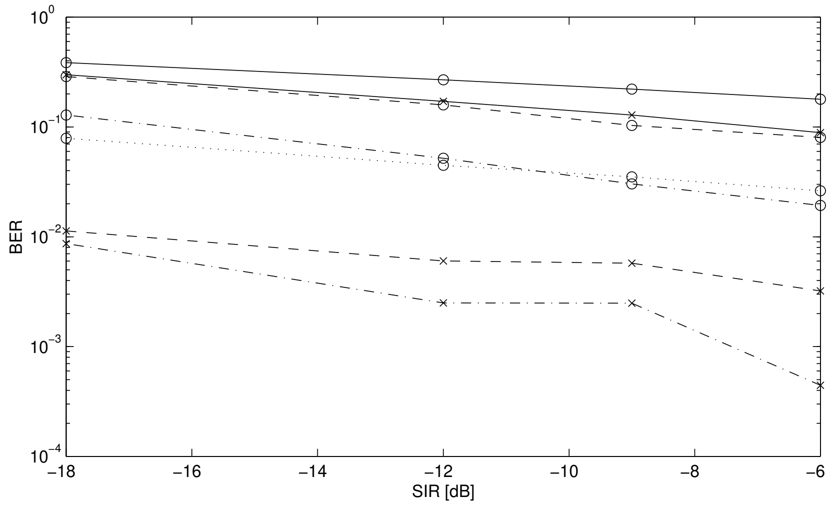


Figure 3.10: BER as a function of SIR for different versions of the AR-DFE. SNR: 20dB. Number of training symbols: 26. AR-DFE with $nn = 0$ (solid-o), $nn = 1$ (dashed-o), $nn = 2$ (dash-dotted-o),$nn = 3$ (dotted-o), AR-DFE processing the real and imaginary part of the signals separately with $nn = 0$ (solid-x), $nn = 1$ (dashed-x), and $nn = 1$ with bootstrapping (dash-dotted-x).

only one physical antenna but with two polarizations. We assume that these two signals fade independently.

We have one user and one co-channel interferer. Both the desired signal and the co-channel interferer has a channel with 5 independently Rayleigh fading taps with the average relative powers -3dB, 0dB, -2dB, -6dB and -8dB and the time delays 0, 0.2, 0.5, 1.6 and 2.3 symbol intervals respectively. The average signal to noise ratio is 20 dB. The transmitted binary symbols are modulated with GMSK modulation with a bandwidth-time product of 0.3 (as in GSM).

All algorithms used a channel model with 5 taps and a smoothing lag of 4 symbol intervals. We compared the following algorithms: AR-DFE with the order of the AR noise model $nn = 0$, $nn = 1$ $nn = 2$ and $nn = 3$, AR-DFE processing the real and imaginary parts of the signal separately with $nn = 0$, $nn = 1$ and the latter AR-DFE with two stages of bootstrapping.

In Figure 3.10 we can see the resulting BER for the different equalizers as a function of the signal to interference ratio. None of the algorithms, not processing the real and imaginary parts of the signal separately, has a good performance for this scenario. The AR-DFE with $nn = 1$ processing the real and imaginary parts separately however has a very good performance and bootstrapping of the same algorithm improves on the performance somewhat. The AR-DFE processing the real and imaginary signals independently has an advantage since it in effect has twice as many signals. This is why it can cancel the co-channel interference and achieve a good BER. It should however be noted that the signal to interference ratio in this scenario is high, 20dB, but it illustrates the principle.

3.5 Discussion

In the simulations presented in Section 3.4 we saw that the IS-MMF-DFE and the IS-DFE achieved very similar performance. However, when using many antenna elements, the IS-MMF-DFE will require much less computations to tune. The main reason for this is that when computing the IS-DFE, as in Section 3.2.3, we need to solve the system of equations (3.62). The dimension of this system of equations is given by the number of antennas times the smoothing lag plus one, i.e. $M(ns+1)$. When tuning the IS-MMF-DFE, however, we only need to invert the spatial covariance matrix for the noise plus interference, of dimension M. A drawback with the MMF-DFE is that the time delay of this equalizer typically will be larger than the time delay of the IS-DFE. The reason for this is that the MMF-DFE introduces extra intersymbol interference via the MMF-filter which then has to be removed by a scalar DFE using long filters. If the channel is time-invariant within a data frame and no continuous adaptation of the equalizer has to be performed, this is of minor importance. As long as the performance for the MMF-DFE is comparable to a DFE optimized for a finite smoothing lag, the MMF-DFE will thus be preferred when the number of antennas is large.

Neither the IS-DFE nor the IS-MMF-DFE utilizes the temporal spectrum of the noise and interference. In a scenario as the one studied, with a relatively large number of antennas, attempting to make *full* use of the temporal spectrum of the noise plus interference in the specific way realized by the D- and I-DFE's does appear not to be meaningful. It should however be noted that the number of antennas can be "large" in more than one way. First, it can be large in the sense that the number of parameters to be tuned in the equalizers becomes large compared to the length of the training sequence. Second, it can be large compared to the number of uncorrelated interferers present, making spatial suppression of the interferers effective. In the simulations performed in this section, the number of antennas was large in both these senses. The situation can, for example, be different if we have a very large number of uncorrelated interfering signals such that spatial interference suppression is not possible, even though we have many antennas.

We can however consider other ways to make use of the temporal dimension of the noise plus interferer spectrum. The AR-DFE of Section 3.3.4 utilizes a space-time AR model of interference plus noise. This DFE will therefore be able to do spatio-temporal interference suppression. Compared to using an AR noise model together with maximum likelihood sequence estimator [7], the decision feedback equalizer has the advantage that the complexity only increases *linearly* with the order of the AR-filter while the complexity increases *exponentially* for the maximum likelihood sequence estimator. It is therefore more feasible to use higher order AR models together with a decision feedback equalizer than with a maximum likelihood sequence estimator. With a large number of antennas, the number of parameters in the model can however become large compared to the number of available equations. This can make them potentially difficult to estimate accurately, especially if the signal to noise ratio is not high enough.

In the simulation example with two received signals, see Figure 3.10, we can see an example of the use of an AR model for the noise plus interference, resulting in space-time suppression of the interferers. We also demonstrate that when we have a binary modulated signal, we can potentially improve the performance by processing the real and imaginary parts of the signal separately.

We can see in Figure 3.9 the BER for the IS-DFE and the MMF-DFE are fairly insensitive to the strength of the co-channel interferers. The reason for this is that spatial subspace occupied by the interference is disjoint from the spatial subspace occupied by the desired signal. The spatial interference suppression can thus completely cancel out the interference without impairing the desired signal significantly. This will be discussed in more detail in Section 7.4 of Chapter 7.

3.A Appendix

3.A.1 Derivation of the Optimal Space-Time DFE.

To simplify the derivation of optimal equalizer coefficients we shall adopt the usual assumption that all previous decisions affecting the current symbol estimate are *correct*, i.e. we let $\tilde{d}(t-\ell-1) = d(t-\ell-1)$.

We now minimize the criterion

$$J = E\left[|d(t-\ell) - \hat{d}(t-\ell)|^2\right] \triangleq E\left[|z(t)|^2\right] \tag{3.96}$$

where

$$\begin{aligned} z(t) &= d(t-\ell) - \hat{d}(t-\ell) \\ &= \left(q^{-\ell} - \boldsymbol{\mathcal{S}}(q^{-1})\boldsymbol{A}^{-1}(q^{-1})\boldsymbol{b}(q^{-1}) + q^{-\ell-1}\mathcal{Q}(q^{-1})\right)d(t) \\ &\quad -\boldsymbol{\mathcal{S}}(q^{-1})\boldsymbol{N}^{-1}(q^{-1})\boldsymbol{M}(q^{-1})\boldsymbol{v}(t) \end{aligned} \tag{3.97}$$

and where the expectation $E[\cdot]$ is taken with respect to the realization of the noise, $\boldsymbol{v}(t)$, and the symbol sequence, $d(t)$. The derivation that follows utilizes the techniques for MSE minimization using polynomial equations introduced in [3].

In order to minimize the criterion (3.96), $z(t)$ has to be orthogonal to any stably and causally filtered version of the input signals $\boldsymbol{y}(t)$ and $d(t-\ell-1)$. The signal $z(t)$ thus has to be orthogonal to

$$e_1(t) = \boldsymbol{g}_1(q^{-1})\boldsymbol{y}(t) \tag{3.98}$$

and to

$$e_2(t) = g_2(q^{-1})d(t-\ell-1) \tag{3.99}$$

where $\boldsymbol{g}_1(q^{-1})$, of dimension $1 \times M$, and $g_2(q^{-1})$ are arbitrary stable and causal filters.

Orthogonality with respect to $e_1(t)$ gives[7]

$$
\begin{aligned}
&E[z(t)e_1^H(t)] = \\
&\quad E\big[\big\{\big(q^{-\ell} - \boldsymbol{\mathcal{S}}\boldsymbol{A}^{-1}\boldsymbol{b} + q^{-\ell-1}\mathcal{Q}\big)d(t) - \boldsymbol{\mathcal{S}}\boldsymbol{N}^{-1}\boldsymbol{M}\boldsymbol{v}(t)\big\} \\
&\quad \times\big\{\boldsymbol{g}_1(\boldsymbol{A}^{-1}\boldsymbol{b}d(t) + \boldsymbol{N}^{-1}\boldsymbol{M}\boldsymbol{v}(t))\big\}^H\big] = 0.
\end{aligned}
\tag{3.100}
$$

Evaluating the expectation with respect to the signals $d(t)$ and $\boldsymbol{v}(t)$, using Parsevals formula [87], we can rewrite (3.100) as

$$
\begin{aligned}
&E[z(t)e_1^H(t)] = \\
&= \frac{1}{2\pi i}\oint_{|z|=1}\Big[\big(z^{-\ell} - \boldsymbol{\mathcal{S}}\boldsymbol{A}^{-1}\boldsymbol{b} + z^{-\ell-1}\mathcal{Q}\big)\boldsymbol{b}^H\boldsymbol{A}^{-H} - \\
&\quad \boldsymbol{\mathcal{S}}\boldsymbol{N}^{-1}\boldsymbol{M}\boldsymbol{M}^H\boldsymbol{N}^{-H}\Big]\boldsymbol{g}_1^H\frac{dz}{z} = 0.
\end{aligned}
\tag{3.101}
$$

We will now write the inverses $\boldsymbol{A}^{-H}$ and $\boldsymbol{N}^{-H}$ as

$$
\boldsymbol{A}^{-H} = a^{-H}\widetilde{\boldsymbol{A}}^H
\tag{3.102}
$$

$$
\boldsymbol{N}^{-H} = n^{-H}\widetilde{\boldsymbol{N}}^H
\tag{3.103}
$$

where $a = \det(\boldsymbol{A})$, $n = \det(\boldsymbol{N})$, $\widetilde{\boldsymbol{A}} = \mathrm{adj}(\boldsymbol{A})$ and $\widetilde{\boldsymbol{N}} = \mathrm{adj}(\boldsymbol{N})$, respectively. Inserting (3.102) and (3.103) into (3.101) gives

$$
\begin{aligned}
&E[z(t)e_1^H(t)] = \\
&= \frac{1}{2\pi i}\oint_{|z|=1}\Big[\big(z^{-\ell} - \boldsymbol{\mathcal{S}}\boldsymbol{A}^{-1}\boldsymbol{b} + z^{-\ell-1}\mathcal{Q}\big)\boldsymbol{b}^H\widetilde{\boldsymbol{A}}^H n^H - \\
&\quad \boldsymbol{\mathcal{S}}\boldsymbol{N}^{-1}\boldsymbol{M}\boldsymbol{M}^H\widetilde{\boldsymbol{N}}^H a^H\Big]a^{-H}n^{-H}\boldsymbol{g}_1^H\frac{dz}{z} = 0.
\end{aligned}
\tag{3.104}
$$

In order for (3.104) to be zero for all admissible $\boldsymbol{g}_1$, we must require that the integrand is analytic inside the unit circle, i.e. it should have no poles inside the unit circle. The factor $a^{-H}(z)n^{-H}(z)\boldsymbol{g}_1^H(z)$ has no poles inside the unit circle since $a^{-1}(z^{-1})$, $n^{-1}(z^{-1})$ and $\boldsymbol{g}_1(z^{-1})$ are all stable.

[7] In the derivation below we will, to simplify, sometimes omit the argument q^{-1} or q in the rational functions and the polynomials.

Therefore we require that a stable rational column vector $\boldsymbol{l}_2 = \boldsymbol{l}_2(z^{-1})$ exists such that all poles of $\boldsymbol{l}_2^H(z)$ are outside $|z| = 1$ and

$$\left(z^{-\ell} - \boldsymbol{\mathcal{S}}\boldsymbol{A}^{-1}\boldsymbol{b} + z^{-\ell-1}\mathcal{Q}\right)\boldsymbol{b}^H\widetilde{\boldsymbol{A}}^H n^H - \boldsymbol{\mathcal{S}}\boldsymbol{N}^{-1}\boldsymbol{M}\boldsymbol{M}^H\widetilde{\boldsymbol{N}}^H a^H = z\boldsymbol{l}_2^H. \quad (3.105)$$

Similarly, orthogonality with respect to $e_2(t) = g_2(q^{-1})d(t-\ell-1)$ gives

$$\begin{aligned} &E[z(t)e_2^H(t)] = \\ &= E\big[\big\{(q^{-\ell} - \boldsymbol{\mathcal{S}}\boldsymbol{A}^{-1}\boldsymbol{b} + q^{-\ell-1}\mathcal{Q})d(t) - \boldsymbol{\mathcal{S}}\boldsymbol{N}^{-1}\boldsymbol{M}\boldsymbol{v}(t)\big\} \\ &\times\big\{g_2 q^{-\ell-1}d(t)\big\}^H\big] = 0. \end{aligned} \quad (3.106)$$

Evaluating the expectation with respect to the signals $d(t)$ and $\boldsymbol{v}(t)$, using Parsevals formula, we obtain

$$E[z(t)e_2^H(t)] = \frac{1}{2\pi i}\oint_{|z|=1}(z - z^{\ell+1}\boldsymbol{\mathcal{S}}\boldsymbol{A}^{-1}\boldsymbol{b} + \mathcal{Q})g_2^H\frac{dz}{z} = 0. \quad (3.107)$$

In order for (3.107) to be zero for all admissible $g_2 = g_2(z^{-1})$, we must require

$$z - z^{\ell+1}\boldsymbol{\mathcal{S}}\boldsymbol{A}^{-1}\boldsymbol{b} + \mathcal{Q} = zl_1^H \quad (3.108)$$

where $l_1^H = l_1^H(z)$ is a rational function, with all poles outside $|z| = 1$, to be determined. We observe that $l_1^H(z)$ will be a polynomial since the left-hand side of (3.108) has no poles outside the unit circle, because $\boldsymbol{A}(q^{-1})$ is stable and $\boldsymbol{\mathcal{S}}(q^{-1})$ is required to be stable.

Multiplying (3.108) with $z^{-\ell-1}\boldsymbol{b}^H\widetilde{\boldsymbol{A}}^H n^H$ from the right and subtracting it from equation (3.105) gives

$$z^{-\ell}l_1^H\boldsymbol{b}^H\widetilde{\boldsymbol{A}}^H n^H - \boldsymbol{\mathcal{S}}\boldsymbol{N}^{-1}\boldsymbol{M}\boldsymbol{M}^H\widetilde{\boldsymbol{N}}^H a^H = z\boldsymbol{l}_2^H. \quad (3.109)$$

Multiplying (3.109) with z^ℓ gives

$$l_1^H\boldsymbol{b}^H\widetilde{\boldsymbol{A}}^H n^H - z^\ell\boldsymbol{\mathcal{S}}\boldsymbol{N}^{-1}\boldsymbol{M}\boldsymbol{M}^H\widetilde{\boldsymbol{N}}^H a^H = z^{\ell+1}\boldsymbol{l}_2^H. \quad (3.110)$$

Now, since $\boldsymbol{\mathcal{S}}$ and $\boldsymbol{N}^{-1}$, are stable rational matrices in z^{-1}, they will have poles inside the unit circle. Furthermore, since $\boldsymbol{M}$, is a polynomial matrix in z^{-1} it will contribute poles at the origin. Since the right hand side is required to have no poles inside the unit circle, the factor $\boldsymbol{N}^{-1}\boldsymbol{M}$ must therefore be canceled by $\boldsymbol{\mathcal{S}}$. Thus, $\boldsymbol{\mathcal{S}}$ must have the structure

$$\boldsymbol{\mathcal{S}} = \boldsymbol{s}_0 \boldsymbol{M}^{-1}\boldsymbol{N}. \tag{3.111}$$

We will below show that $\boldsymbol{s}_0$ must be a *polynomial* row vector.

By inserting (3.111) into (3.110) we obtain

$$l_1^H \boldsymbol{b}^H \widetilde{\boldsymbol{A}}^H n^H - z^\ell \boldsymbol{s}_0 \boldsymbol{M}^H \widetilde{\boldsymbol{N}}^H a^H = z^{\ell+1}\boldsymbol{l}_2^H. \tag{3.112}$$

Since $\boldsymbol{s}_0$ is the only factor that can have poles inside the unit circle it must in fact have no poles, other than in the origin, since these could not be matched by any other factor in (3.112), i.e. it must be a polynomial in z^{-1}. Furthermore, since $\boldsymbol{l}_2^H$ is the only remaining potentially rational factor in (3.112), it too must be a polynomial matrix.

Similar to (3.102) and (3.103) we can write the inverse $\boldsymbol{M}^{-1}$ as

$$\boldsymbol{M}^{-1} = m^{-1}\widetilde{\boldsymbol{M}} \tag{3.113}$$

where $m = \det(\boldsymbol{M})$ and $\widetilde{\boldsymbol{M}} = \mathrm{adj}(\boldsymbol{M})$.

Using $\boldsymbol{\mathcal{S}} = \boldsymbol{s}_0\boldsymbol{M}^{-1}\boldsymbol{N}$ in (3.108) and inserting (3.102) and (3.113) gives

$$z - z^{\ell+1} m^{-1} a^{-1} \boldsymbol{s}_0 \widetilde{\boldsymbol{M}} \boldsymbol{N} \widetilde{\boldsymbol{A}} \boldsymbol{b} + \mathcal{Q} = z l_1^H. \tag{3.114}$$

By defining

$$p = p(z^{-1}) \stackrel{\triangle}{=} a(z^{-1})m(z^{-1}) = \det\left(\boldsymbol{A}(z^{-1})\right)\det\left(\boldsymbol{M}(z^{-1})\right) \tag{3.115}$$

we can write (3.114) as

$$zp - z^{\ell+1}\boldsymbol{s}_0 \widetilde{\boldsymbol{M}} \boldsymbol{N} \widetilde{\boldsymbol{A}} \boldsymbol{b} + \mathcal{Q}p = z l_1^H p. \tag{3.116}$$

The right hand side of (3.116) is required to have no poles inside $0 < |z| < 1$. Since $\mathcal{Q}p$ is the only term in (3.116) that could have poles in this region, it must in fact have no poles there, and must thus be a polynomial in z^{-1}. We thus require that

$$\mathcal{Q} = p^{-1}Q \tag{3.117}$$

where Q is a polynomial in z^{-1} yet to be determined.

By inserting (3.117) into (3.116) we obtain

$$zp - z^{\ell+1}\boldsymbol{s}_0\widetilde{\boldsymbol{M}}\boldsymbol{N}\widetilde{\boldsymbol{A}}\boldsymbol{b} + Q = zl_1^H p. \tag{3.118}$$

To summarize, we thus have to solve the two coupled Diophantine equations (3.112) and (3.116)

$$l_1^H \boldsymbol{b}^H \widetilde{\boldsymbol{A}}^H n^H - z^{\ell}\boldsymbol{s}_0\boldsymbol{M}^H\widetilde{\boldsymbol{N}}^H a^H = z^{\ell+1}\boldsymbol{l}_2^H \tag{3.119}$$

$$zp - z^{\ell+1}\boldsymbol{s}_0\widetilde{\boldsymbol{M}}\boldsymbol{N}\widetilde{\boldsymbol{A}}\boldsymbol{b} + Q = zl_1^H p \tag{3.120}$$

with respect to the polynomials and the polynomial vectors l_1, $\boldsymbol{s}_0$, $\boldsymbol{l}_2$ and Q.

Since the maximal positive power of z on the left hand side of (3.120) should match the power on the right hand side, we can conclude that

$$\deg\, l_1(z^{-1}) = \ell.$$

Likewise, by inspecting the negative powers of z in (3.119) we can note that we must have

$$\deg\, \boldsymbol{s}_0(z^{-1}) = \ell.$$

Furthermore, using that $\deg\, \boldsymbol{s}_0(z^{-1}) = \ell$ in (3.120) and investigating the negative powers of z gives

$$\deg\, Q(z^{-1}) = \max[nb + nn + (M-1)(nm+na), M(nm+na)] - 1.$$

Finally, using that $\deg\, l_1^H(z) = \ell$ in (3.119) and considering the positive powers of z gives

$$\deg\, \boldsymbol{l}_2(z^{-1}) = \max[nb + (M-1)na + Mnn, nm + (M-1)nn + Mna] - 1.$$

To summarize, and exchanging z for q, we have the following orders of the polynomials

$$\begin{aligned}
\deg\, \boldsymbol{s}_0(q^{-1}) &= \ell \\
\deg\, l_1(q^{-1}) &= \ell \\
\deg\, Q(q^{-1}) &= \max[nb + nn + (M-1)(nm+na), M(nm+na)] - 1 \\
\deg\, \boldsymbol{l}_2(q^{-1}) &= \max[nb + (M-1)na + Mnn, nm + (M-1)nn + Mna] - 1.
\end{aligned}$$

We will now show that the coefficients to the polynomials $\boldsymbol{s}_0(q^{-1})$, $\boldsymbol{l}_1(q^{-1})$, $Q(q^{-1})$ and $\boldsymbol{l}_2(q^{-1})$ can be found by solving a linear system of equations. Let us first take the complex conjugate transpose of equations (3.119) and (3.120) and rearrange them to obtain

$$z^{-\ell} a\widetilde{\boldsymbol{N}}\boldsymbol{M}\boldsymbol{s}_0^H - n\widetilde{\boldsymbol{A}}\boldsymbol{b}l_1 = -z^{-\ell-1}\boldsymbol{l}_2 \tag{3.121}$$

$$z^{-\ell}\boldsymbol{b}^H\widetilde{\boldsymbol{A}}^H\boldsymbol{N}^H\widetilde{\boldsymbol{M}}^H\boldsymbol{s}_0^H + p^H l_1 = zQ^H + p^H. \tag{3.122}$$

Each power of z or z^{-1} in (3.121) gives an equation in the coefficients of s_0^H and l_1. By combining these equations into a system of equations we obtain

$$\left[\begin{array}{cccccc} \boldsymbol{H}_{nl_2+1} & \cdots & \boldsymbol{H}_{nl_2+1+\ell} & -\boldsymbol{c}_{nl_2+1} & \cdots & -\boldsymbol{c}_{nl_2+1+\ell} \\ \vdots & & \vdots & \vdots & & \vdots \\ \boldsymbol{H}_1 & \cdots & \boldsymbol{H}_{\ell+1} & -\boldsymbol{c}_1 & \cdots & -\boldsymbol{c}_{\ell+1} \\ \hline \boldsymbol{H}_0 & \cdots & \boldsymbol{H}_\ell & -\boldsymbol{c}_0 & \cdots & -\boldsymbol{c}_\ell \\ & \ddots & \vdots & & \ddots & \vdots \\ 0 & & \boldsymbol{H}_0 & 0 & & -\boldsymbol{c}_0 \end{array}\right] \left[\begin{array}{c} \boldsymbol{s}_{0,0}^H \\ \vdots \\ \boldsymbol{s}_{0,\ell}^H \\ l_{1_\ell} \\ \vdots \\ l_{10} \end{array}\right] = \left[\begin{array}{c} -\boldsymbol{l}_{2nl_2} \\ \vdots \\ -\boldsymbol{l}_{20} \\ \hline 0 \\ \vdots \\ 0 \end{array}\right] \tag{3.123}$$

where

$$\boldsymbol{H}(z^{-1}) = \boldsymbol{H}_0 + \boldsymbol{H}_1 z^{-1} + \ldots + \boldsymbol{H}_{nh} z^{-nh} \triangleq a\widetilde{\boldsymbol{N}}\boldsymbol{M} \tag{3.124}$$

$$\boldsymbol{c}(z^{-1}) = \boldsymbol{c}_0 + \boldsymbol{c}_1 z^{-1} + \ldots + \boldsymbol{c}_{nc} z^{-nc} \triangleq n\widetilde{\boldsymbol{A}}\boldsymbol{b} \tag{3.125}$$

and $\boldsymbol{H}_i = 0$ if $i > nh = Mna + (M-1)nn + nm$ and $\boldsymbol{c}_i = 0$ if $i > nc = Mnn + (M-1)na + nb$. The unknown parameters in the equation are $\boldsymbol{s}_{0,k}$, the vector taps in the polynomial

$$\boldsymbol{s}_0(q^{-1}) = \boldsymbol{s}_{0,0} + \boldsymbol{s}_{0,1}q^{-1} + \ldots + \boldsymbol{s}_{0,\ell}q^{-\ell} \tag{3.126}$$

and l_{1k}, the scalar coefficients of the polynomial

$$l_1(q^{-1}) = l_{10} + l_{01}q^{-1} + \ldots + l_{0\ell}q^{-\ell}. \tag{3.127}$$

Similarly, equation (3.122) gives the system of equations

$$\left[\begin{array}{cccccc} \boldsymbol{u}_0^H & & 0 & p_0^H & & 0 \\ \vdots & \ddots & & \vdots & \ddots & \\ \boldsymbol{u}_\ell^H & \cdots & \boldsymbol{u}_0^H & p_\ell^H & \cdots & p_0^H \\ \hline \boldsymbol{u}_{\ell+1}^H & \cdots & \boldsymbol{u}_1^H & p_{\ell+1}^H & \cdots & p_1^H \\ \vdots & & \vdots & \vdots & & \vdots \\ \boldsymbol{u}_{\ell+nf}^H & \cdots & \boldsymbol{u}_{nf}^H & p_{\ell+nf}^H & \cdots & p_{nf}^H \end{array}\right] \left[\begin{array}{c} \boldsymbol{s}_{0,0}^H \\ \vdots \\ \boldsymbol{s}_{0,\ell}^H \\ l_{1_\ell} \\ \vdots \\ l_{10} \end{array}\right] = \left[\begin{array}{c} 0 \\ \vdots \\ p_0^H \\ \hline f_1^H \\ \vdots \\ f_{nf}^H \end{array}\right] \tag{3.128}$$

where

$$f(z^{-1}) \triangleq 1 + f_1 z^{-1} + \ldots + f_{nf} z^{-nf} = p(z^{-1}) + z^{-1} Q(z^{-1}) \tag{3.129}$$

and

$$\boldsymbol{u}(z^{-1}) = \boldsymbol{u}_0 + \boldsymbol{u}_1 q^{-1} + \ldots + \boldsymbol{u}_{nu} q^{-nu} = \widetilde{\boldsymbol{M}} \boldsymbol{N} \widetilde{\boldsymbol{A}} \boldsymbol{b}. \tag{3.130}$$

Again $\boldsymbol{u}_i = 0$ if $i > nu = (M-1)(nm+na) + nn + nb$ and $p_i = 0$ if $i > np = M^2 nanm$.

Equations (3.123) and (3.128) are coupled and cannot be solved separately. However, by combining the top $\ell + 1$ equations from (3.128) with the lower half of the equations from (3.123) we obtain the system of $(M+1)(\ell+1)$ equations in the $M(\ell+1)$ coefficients of $\boldsymbol{s}_0(z^{-1})$ and the $\ell + 1$ coefficients of $l_1(z^{-1})$

$$\begin{bmatrix} \boldsymbol{u}_0^H & & 0 & p_0^H & & 0 \\ \vdots & \cdots & & \vdots & \ddots & \\ \boldsymbol{u}_\ell^H & \cdots & \boldsymbol{u}_0^H & p_\ell^H & \cdots & p_0^H \\ \boldsymbol{H}_0 & \cdots & \boldsymbol{H}_\ell & -\boldsymbol{c}_0 & \cdots & -\boldsymbol{c}_\ell \\ & \ddots & \vdots & & \ddots & \vdots \\ 0 & & \boldsymbol{H}_0 & 0 & & -\boldsymbol{c}_0 \end{bmatrix} \begin{bmatrix} \boldsymbol{s}_{0,0}^H \\ \vdots \\ \boldsymbol{s}_{0,\ell}^H \\ l_{1_\ell} \\ \vdots \\ l_{10} \end{bmatrix} = \begin{bmatrix} 0 \\ \vdots \\ p_0^H \\ 0 \\ \vdots \\ 0 \end{bmatrix}. \tag{3.131}$$

As long as the left hand side matrix is invertible, we can obtain a unique solution for the coefficients of the polynomial vector $\boldsymbol{s}_0(z^{-1})$ and for the coefficients of the polynomial $l_1(z^{-1})$. We can note that the polynomial

$$Q(z^{-1}) = z(f(z^{-1}) - p(z^{-1})) \tag{3.132}$$

will be causal since $f(z^{-1})$ is monic and $p(z^{-1})$ is monic by assumption as $\det\left(\boldsymbol{A}(z^{-1})\right)$ and $\det\left(\boldsymbol{M}(z^{-1})\right)$ both are monic polynomials. As the coefficients of $f(z^{-1})$ can be computed using the lower set of equations in (3.128), we see that we can compute the coefficients of the polynomial $Q(z^{-1}) = Q_0 + Q_1 z^{-1} + \ldots + Q_{nq} z^{-nq}$ of order $nq = \max[nb + nn + (M -$

$1)(nm+na), M(nm+na)] - 1$ as

$$\begin{bmatrix} Q_0^H \\ \vdots \\ Q_{nq}^H \end{bmatrix} = \begin{bmatrix} \boldsymbol{u}_{\ell+1}^H & \cdots & \boldsymbol{u}_1^H & p_{\ell+1}^H & \cdots & p_1^H \\ \vdots & & \vdots & \vdots & & \vdots \\ \boldsymbol{u}_{\ell+1+nq}^H & \cdots & \boldsymbol{u}_{1+nq}^H & p_{\ell+1+nq}^H & \cdots & p_{1+nq}^H \end{bmatrix} \begin{bmatrix} \boldsymbol{s}_{0,0}^H \\ \vdots \\ \boldsymbol{s}_{0,\ell}^H \\ l_{1_\ell} \\ \vdots \\ l_{11} \\ l_{10} - 1 \end{bmatrix} \tag{3.133}$$

where $\boldsymbol{u}_i = 0$ if $i > nu$ and $p_i = 0$ if $i > np$.

The rational feedforward filter is now, using (3.111), given by

$$\boldsymbol{\mathcal{S}}(q^{-1}) = \boldsymbol{s}_0(q^{-1})\boldsymbol{M}^{-1}(q^{-1})\boldsymbol{N}(q^{-1}) \tag{3.134}$$

and the rational feedback filter is, using (3.117), given by

$$\mathcal{Q}(q^{-1}) = \frac{Q(q^{-1})}{a(q^{-1})m(q^{-1})}. \tag{3.135}$$

3.A.2 Derivation of the Optimal Space-Time DFE for AR Noise.

If we assume, as shown in Figure 3.4, that the MMSE-optimal DFE has FIR filters in the feedforward and the feedback branches, then the estimated symbol prior to the decision device can be written as

$$\hat{d}(t-\ell) = \boldsymbol{s}(q^{-1})y(t) - Q(q^{-1})\tilde{d}(t-\ell-1). \tag{3.136}$$

Although this assumption is true, we can for the time being consider the feedforward filter $\boldsymbol{s}(q^{-1})$ and the feedback filter $Q(q^{-1})$ to be *potentially rational* filters.

To make the derivation of optimal equalizer coefficients feasible we shall as usual adopt the assumption that all previous decisions affecting the current symbol estimate are *correct*.

As in Appendix 3.A.1 we want to minimize the criterion

$$J = E\left[|d(t-\ell) - \hat{d}(t-\ell|t)|^2\right] = E\left[|z(t)|^2\right] \tag{3.137}$$

where

$$\begin{aligned} z(t) &= d(t-\ell) - \hat{d}(t-\ell|t) \\ &= \left(q^{-\ell} - \boldsymbol{s}\boldsymbol{b} + q^{-\ell-1}Q\right)d(t) - \boldsymbol{s}\boldsymbol{N}^{-1}\boldsymbol{M}_0\boldsymbol{v}(t). \end{aligned} \tag{3.138}$$

The expectation $E[\cdot]$ is with respect to realizations of the noise, $\boldsymbol{n}(t)$, and the symbol sequence, $d(t)$. The parameter ℓ is as before the smoothing lag.

In order to minimize the criterion (3.137), $z(t)$ has to be orthogonal to any filtered version of the input signals $\boldsymbol{y}(t)$ and $d(t-\ell-1)$. The signal $z(t)$ thus has to be orthogonal to

$$e_1(t) = \boldsymbol{g}_1\boldsymbol{y}(t) \tag{3.139}$$

and to

$$e_2(t) = g_2 d(t-\ell-1) \tag{3.140}$$

where $\boldsymbol{g}_1 = \boldsymbol{g}_1(q^{-1})$ and $g_2 = g_2(q^{-1})$ are arbitrary stable and causal *IIR* filters. Note that $\boldsymbol{g}_1$ is a row vector IIR filter.

Orthogonality with respect to $e_1(t)$ gives

$$
\begin{aligned}
&E[z(t)e_1^H(t)] = \\
&\quad E\big[\{(q^{-\ell} - \boldsymbol{s}\boldsymbol{b} + q^{-\ell-1}Q)d(t) - \boldsymbol{s}\boldsymbol{N}^{-1}\boldsymbol{M}_0\boldsymbol{v}(t)\} \\
&\quad \times\{\boldsymbol{g}_1(\boldsymbol{b}d(t) + \boldsymbol{N}^{-1}\boldsymbol{M}_0\boldsymbol{v}(t))\}^H\big] = 0.
\end{aligned}
\tag{3.141}
$$

Evaluating the expectation with respect to the signals $d(t)$ and $\boldsymbol{v}(t)$ gives

$$
\begin{aligned}
&E[z(t)e_1^H(t)] = \\
&= \frac{1}{2\pi i}\oint_{|z|=1}\Big[(z^{-\ell} - \boldsymbol{s}\boldsymbol{b} + z^{-\ell-1}Q)\boldsymbol{b}^H - \\
&\qquad \boldsymbol{s}\boldsymbol{N}^{-1}\boldsymbol{M}_0\boldsymbol{M}_0^H\boldsymbol{N}^{-H}\Big]\boldsymbol{g}_1^H\frac{dz}{z} = 0.
\end{aligned}
\tag{3.142}
$$

Separating out $\boldsymbol{M}_0^H\boldsymbol{N}^{-H}$ from the parenthesis gives

$$
\begin{aligned}
&E[z(t)e_1(t)] = \\
&= \frac{1}{2\pi i}\oint_{|z|=1}\Big[(z^{-\ell} - \boldsymbol{s}\boldsymbol{b} + z^{-\ell-1}Q)\boldsymbol{b}^H\boldsymbol{N}^H\boldsymbol{M}^{-H} - \\
&\qquad \boldsymbol{s}\boldsymbol{N}^{-1}\boldsymbol{M}_0\Big]\boldsymbol{M}_0^H\boldsymbol{N}^{-H}\boldsymbol{g}_1^H\frac{dz}{z} = 0.
\end{aligned}
\tag{3.143}
$$

For this equation to be zero for all admissible $\boldsymbol{g}_1$, we must require that the integrand is analytic inside the unit circle, i.e. it should have no poles inside the unit circle. The factor $\boldsymbol{M}_0^H\boldsymbol{N}^{-H}\boldsymbol{g}_1^H$ has no poles inside the unit circle since $\boldsymbol{N}^{-1}$ and $\boldsymbol{g}_1$ are stable. Therefore we require that a stable rational column vector $\boldsymbol{l}_2 = \boldsymbol{l}_2(z^{-1})$ exists such that all poles of $\boldsymbol{l}_2^H(z)$ are located outside the unit circle and

$$
\big(z^{-\ell} - \boldsymbol{s}\boldsymbol{b} + z^{-\ell-1}Q\big)\boldsymbol{b}^H\boldsymbol{N}^H\boldsymbol{M}_0^{-H} - \boldsymbol{s}\boldsymbol{N}^{-1}\boldsymbol{M}_0 = z\boldsymbol{l}_2^H.
\tag{3.144}
$$

Similarly, orthogonality with respect to $e_2(t) = g_2(q^{-1})d(t-\ell-1)$ gives

$$
\begin{aligned}
&E[z(t)e_2^H(t)] = \\
&= E\big[\{(q^{-\ell} - \boldsymbol{s}\boldsymbol{b} + q^{-\ell-1}Q)d(t) - \boldsymbol{s}\boldsymbol{N}^{-1}\boldsymbol{M}_0\boldsymbol{v}(t)\} \\
&\times\{g_2q^{-\ell-1}d(t)\}^H\big] = 0.
\end{aligned}
\tag{3.145}
$$

Evaluating the expectation with respect to the signals $d(t)$ and $\boldsymbol{v}(t)$, using Parsevals formula, we obtain

$$
\frac{1}{2\pi i}\oint_{|z|=1}(z - z^{\ell+1}\boldsymbol{s}\boldsymbol{b} + Q)g_2^H\frac{dz}{z} = 0.
\tag{3.146}
$$

In order for this equation to be zero for all admissible $g_2 = g_2(z^{-1})$, we must require

$$z - z^{\ell+1}\boldsymbol{s}\boldsymbol{b} + Q = z l_1^H \tag{3.147}$$

where $l_1^H = l_1^H(z)$ is a rational function, with all poles outside the unit circle, to be determined. We can however conclude that l_1 must be a polynomial since the left-hand side has no poles outside the unit circle.

Multiplying (3.147) with $z^{-\ell-1}\boldsymbol{b}^H\boldsymbol{N}^H\boldsymbol{M}_0^{-H}$ from the right and subtracting it from equation (3.144) gives

$$z^{-\ell} l_1^H \boldsymbol{b}^H\boldsymbol{N}^H\boldsymbol{M}_0^{-H} - \boldsymbol{s}\boldsymbol{N}^{-1}\boldsymbol{M}_0 = z\boldsymbol{l}_2^H. \tag{3.148}$$

Multiplying (3.148) with z^ℓ gives

$$l_1^H \boldsymbol{b}^H\boldsymbol{N}^H\boldsymbol{M}_0^{-H} - z^\ell\boldsymbol{s}\boldsymbol{N}^{-1}\boldsymbol{M}_0 = z^{\ell+1}\boldsymbol{l}_2^H. \tag{3.149}$$

Now, $\boldsymbol{s}$ and $\boldsymbol{N}^{-1}$ (being polynomial or stable rational matrices in z^{-1}) are the only factors in the left hand side that may have poles inside the unit circle. Since the right hand side is required to have no poles inside the unit circle, the factor $\boldsymbol{N}^{-1}$ must be canceled by $\boldsymbol{s}$. We select to also include the constant matrix factor $\boldsymbol{M}_0^{-1}$ in $\boldsymbol{s}$. The row vector $\boldsymbol{s}$ will thus have the structure

$$\boldsymbol{s} = \boldsymbol{s}_0\boldsymbol{M}_0^{-1}\boldsymbol{N}. \tag{3.150}$$

We will below show that $\boldsymbol{s}_0$ is a *polynomial* row vector.

By inserting (3.150) into (3.149) we obtain

$$l_1^H \boldsymbol{b}^H\boldsymbol{N}^H\boldsymbol{M}_0^{-H} - z^\ell\boldsymbol{s}_0 = z^{\ell+1}\boldsymbol{l}_2^H. \tag{3.151}$$

Since $\boldsymbol{s}_0$ is the only factor that can have poles inside the unit circle it must in fact have no poles at all, except at the origin, since they cannot be matched by some other factor in (3.151). With $\boldsymbol{s}_0$ being a *polynomial* matrix in z^{-1}, we can from (3.150) see that the potentially rational feedforward filter $\boldsymbol{s}(q^{-1})$ is indeed a *polynomial*, as predicted. Furthermore, since $\boldsymbol{l}_2^H$ is the only remaining potentially rational factor in (3.151), it must also be a *polynomial* matrix.

Using $\boldsymbol{s} = \boldsymbol{s}_0\boldsymbol{M}_0\boldsymbol{N}$ in (3.147), we obtain

$$z - z^{\ell+1}\boldsymbol{s}_0\boldsymbol{M}_0^{-1}\boldsymbol{N}\boldsymbol{b} + Q = z l_1^H. \tag{3.152}$$

Since, $Q(q^{-1})$, our potentially rational feedback filter, is the only factor that can contain poles in $0 < |z| < 1$ and cannot have poles in outside the unit circle, it too must indeed be a *polynomial*, as predicted at the outset. Furthermore, since l_1^H is the only factor that can have poles outside the unit circle, and nowhere else, it must be a *polynomial*.

To summarize, we have to solve the two coupled Diophantine equations (3.151) and (3.152)

$$l_1^H \boldsymbol{b}^H \boldsymbol{N}^H \boldsymbol{M}_0^{-H} - z^{\ell} \boldsymbol{s}_0 = z^{\ell+1} \boldsymbol{l}_2^H \tag{3.153}$$

$$z - z^{\ell+1} \boldsymbol{s}_0 \boldsymbol{M}_0^{-1} \boldsymbol{N} \boldsymbol{b} + Q = z l_1^H \tag{3.154}$$

with respect to the polynomials and polynomial vectors l_1, $\boldsymbol{s}_0$, $\boldsymbol{l}_2$ and Q, respectively.

From the maximal positive power of z in (3.154) it can be concluded that

$$\deg\, l_1(z^{-1}) = \ell. \tag{3.155}$$

Likewise, by noting that there will exist no negative powers of z in (3.153) when $\ell \geq 0$, it can be concluded that

$$\deg\, \boldsymbol{s}_0(z^{-1}) = \ell. \tag{3.156}$$

Furthermore, using $\deg\, \boldsymbol{s}_0(z^{-1}) = \ell$ in (3.154) and investigating the negative powers of z gives

$$\deg\, Q(z^{-1}) = nb + nn - 1. \tag{3.157}$$

Finally, using that $\deg\, l_1^H(z) = \ell$ in (3.153) and considering the positive powers of z gives

$$\deg\, \boldsymbol{l}_2(z^{-1}) = nb + nn - 1. \tag{3.158}$$

We will now show that the coefficients to the polynomials $\boldsymbol{s}_0(q^{-1})$, $\boldsymbol{l}_1(q^{-1})$, $Q(q^{-1})$ and $\boldsymbol{l}_2(q^{-1})$ can be found by solving a system of equations. Let us first take the complex conjugate of equations (3.153) and (3.154) and rearrange them to obtain

$$z^{-\ell} \boldsymbol{s}_0^H - \boldsymbol{M}_0^{-1} \boldsymbol{N} \boldsymbol{b} l_1 = -z^{-\ell-1} \boldsymbol{l}_2 \tag{3.159}$$

$$z^{-\ell}\boldsymbol{b}^H\boldsymbol{N}^H\boldsymbol{M}_0^{-H}\boldsymbol{s}_0^H + l_1 = zQ^H + 1. \tag{3.160}$$

Each power of z or z^{-1} in (3.159) gives an equation in the coefficients. By combining these equations into a system of equations we obtain

$$\left[\begin{array}{ccccc} 0 & \cdots & 0 & -\boldsymbol{b}'_{nl_2+1} & \cdots & -\boldsymbol{b}'_{nl_2+1+\ell} \\ \vdots & & \vdots & \vdots & & \vdots \\ 0 & \cdots & 0 & -\boldsymbol{b}'_1 & \cdots & -\boldsymbol{b}'_{\ell+1} \\ \hline \boldsymbol{I} & & 0 & -\boldsymbol{b}'_0 & \cdots & -\boldsymbol{b}'_\ell \\ & \ddots & & & \ddots & \vdots \\ 0 & & \boldsymbol{I} & 0 & & -\boldsymbol{b}'_0 \end{array}\right] \left[\begin{array}{c} \boldsymbol{s}_{0,0}^H \\ \vdots \\ \boldsymbol{s}_{0,\ell}^H \\ l_{1_\ell} \\ \vdots \\ l_{10} \end{array}\right] = \left[\begin{array}{c} -\boldsymbol{l}_{2nl_2} \\ \vdots \\ -\boldsymbol{l}_{20} \\ \hline 0 \\ \vdots \\ 0 \end{array}\right] \tag{3.161}$$

where $\boldsymbol{b}'_i$ are the vector taps of the noise whitened channel

$$\boldsymbol{b}'(z^{-1}) = \boldsymbol{b}'_0 + \boldsymbol{b}'_1 z^{-1} + \ldots + \boldsymbol{b}'_{nb'} z^{-nb'} = \boldsymbol{M}_0^{-1}\boldsymbol{N}\boldsymbol{b} \tag{3.162}$$

with $\boldsymbol{b}'_i = 0$ if $i > nb' = nb + nn$. Furthermore $\boldsymbol{s}_{0,k}$ are the vector taps in the polynomial

$$\boldsymbol{s}_0(q^{-1}) = \boldsymbol{s}_{0,0} + \boldsymbol{s}_{0,1}q^{-1} + \ldots + \boldsymbol{s}_{0,\ell}q^{-\ell} \tag{3.163}$$

and l_{1k} are the scalar coefficients of the polynomial

$$l_1(q^{-1}) = l_{10} + l_{01}q^{-1} + \ldots + l_{0\ell}q^{-\ell}. \tag{3.164}$$

Similarly, equation (3.160) gives the system of equations

$$\left[\begin{array}{cccccc} \boldsymbol{b}'^H_0 & & 0 & 1 & & 0 \\ \vdots & \ddots & & & \ddots & \\ \boldsymbol{b}'^H_\ell & \cdots & \boldsymbol{b}'^H_0 & 0 & & 1 \\ \hline \boldsymbol{b}'^H_{\ell+1} & \cdots & \boldsymbol{b}'^H_1 & 0 & \cdots & 0 \\ \vdots & & \vdots & \vdots & & \vdots \\ \boldsymbol{b}'^H_{nf} & \cdots & \boldsymbol{b}'^H_{nf-\ell} & 0 & \cdots & 0 \end{array}\right] \left[\begin{array}{c} \boldsymbol{s}_{0,0}^H \\ \vdots \\ \boldsymbol{s}_{0,\ell}^H \\ l_{1_\ell} \\ \vdots \\ l_{10} \end{array}\right] = \left[\begin{array}{c} 0 \\ \vdots \\ 1 \\ \hline f_1^H \\ \vdots \\ f_{nf}^H \end{array}\right] \tag{3.165}$$

where the monic polynomial $f(z^{-1})$ of degree $nf = nb + nn$ is given by

$$f(z^{-1}) = 1 + f_1 z^{-1} + \ldots + f_{nf}z^{-nf} = 1 + z^{-1}Q(z^{-1}). \tag{3.166}$$

Again $\boldsymbol{b}'_i = 0$ if $i > nb'$.

Equations (3.161) and (3.165) cannot be solved separately. However, by combining the top $\ell+1$ equations from (3.165) with the lower half of the equations from (3.161) we obtain the system of equations

$$\begin{bmatrix} \boldsymbol{b'}_0^H & & 0 & 1 & & 0 \\ \vdots & \cdots & & & \ddots & \\ \boldsymbol{b'}_\ell^H & \cdots & \boldsymbol{b'}_0^H & 0 & & 1 \\ \boldsymbol{I} & & 0 & -\boldsymbol{b'}_0 & \cdots & -\boldsymbol{b'}_\ell \\ & \ddots & \vdots & & \ddots & \vdots \\ 0 & & \boldsymbol{I} & 0 & & -\boldsymbol{b'}_0 \end{bmatrix} \begin{bmatrix} \boldsymbol{s}_{0,0}^H \\ \vdots \\ \boldsymbol{s}_{0,\ell}^H \\ l_{1_\ell} \\ \vdots \\ l_{10} \end{bmatrix} = \begin{bmatrix} 0 \\ \vdots \\ 1 \\ 0 \\ \vdots \\ 0 \end{bmatrix}. \tag{3.167}$$

By introducing the notation

$$\boldsymbol{\mathcal{B}}' = \begin{bmatrix} \boldsymbol{b'}_0 & \cdots & \boldsymbol{b'}_\ell \\ & \ddots & \vdots \\ 0 & & \boldsymbol{b'}_0 \end{bmatrix} \tag{3.168}$$

$$\boldsymbol{s}_0 = \begin{bmatrix} \boldsymbol{s}_{0,0} & \ldots & \boldsymbol{s}_{0,\ell} \end{bmatrix} \tag{3.169}$$

and

$$\boldsymbol{l}_1 = \begin{bmatrix} l_{1\ell} \\ \vdots \\ l_{10} \end{bmatrix} \tag{3.170}$$

we can write (3.167) as

$$\begin{bmatrix} \boldsymbol{\mathcal{B}}'^H & \boldsymbol{I}_{(\ell+1)\times(\ell+1)} \\ \boldsymbol{I}_{M(\ell+1)\times M(\ell+1)} & -\boldsymbol{\mathcal{B}}' \end{bmatrix} \begin{bmatrix} \boldsymbol{s}_0^H \\ \boldsymbol{l}_1 \end{bmatrix} = \left[\begin{array}{c} 0 \\ \vdots \\ 1 \\ \hline 0 \\ \vdots \\ 0 \end{array}\right]. \tag{3.171}$$

We can now separate (3.167) into a system of two matrix equations as

$$\boldsymbol{\mathcal{B}}'^H \boldsymbol{s}_0^H + \boldsymbol{l}_1 = \begin{bmatrix} 0 \\ \vdots \\ 1 \end{bmatrix} \tag{3.172}$$

$$\boldsymbol{s}_0^H - \boldsymbol{\mathcal{B}}'\boldsymbol{l}_1 = \begin{bmatrix} 0 \\ \vdots \\ 0 \end{bmatrix}. \tag{3.173}$$

Equation (3.173) now gives

$$\boldsymbol{\mathcal{B}}'\boldsymbol{l}_1 = \boldsymbol{s}_0^H. \tag{3.174}$$

Let us now multiply (3.172) with $\boldsymbol{\mathcal{B}}'$ from the left to obtain

$$\boldsymbol{\mathcal{B}}'\boldsymbol{\mathcal{B}}'^H\boldsymbol{s}_0^H + \boldsymbol{\mathcal{B}}'\boldsymbol{l}_1 = \boldsymbol{\mathcal{B}}'\begin{bmatrix} 0 \\ \vdots \\ 1 \end{bmatrix} = \begin{bmatrix} \boldsymbol{b}'_\ell \\ \vdots \\ \boldsymbol{b}'_0 \end{bmatrix}. \tag{3.175}$$

Inserting (3.174) into (3.175) gives

$$(\boldsymbol{\mathcal{B}}'\boldsymbol{\mathcal{B}}'^H + \boldsymbol{I})\boldsymbol{s}_0^H = \begin{bmatrix} \boldsymbol{b}'_\ell \\ \vdots \\ \boldsymbol{b}'_0 \end{bmatrix}. \tag{3.176}$$

As long as $(\boldsymbol{\mathcal{B}}'\boldsymbol{\mathcal{B}}'^H + \boldsymbol{I})$ is invertible we can compute

$$\boldsymbol{s}_0^H = (\boldsymbol{\mathcal{B}}'\boldsymbol{\mathcal{B}}'^H + \boldsymbol{I})^{-1}\begin{bmatrix} \boldsymbol{b}'_\ell \\ \vdots \\ \boldsymbol{b}'_0 \end{bmatrix} \tag{3.177}$$

and extract the coefficients for the feedforward filter from $\boldsymbol{s}^H$ according to (3.169). Noting that

$$Q(z^{-1}) = z(f(z^{-1}) - 1) \tag{3.178}$$

and that the coefficients of the polynomial $f(z^{-1})$ can be computed using the lower set of equations in (3.165), we see that we can compute the coefficients of the feedback polynomial $Q(z^{-1}) = Q_0 + Q_1 z^{-1} + \ldots + Q_{nq} z^{-nq}$ of order $nq = nb + nn - 1$ as

$$\begin{bmatrix} Q_0^H \\ \vdots \\ Q_{nq} \end{bmatrix} = \begin{bmatrix} \boldsymbol{b}'^H_{\ell+1} & \cdots & \boldsymbol{b}'^H_1 \\ \vdots & & \vdots \\ \boldsymbol{b}'^H_{\ell+nb+nn} & \cdots & \boldsymbol{b}'^H_{nb+nn} \end{bmatrix}\begin{bmatrix} \boldsymbol{s}_{0,0}^H \\ \vdots \\ \boldsymbol{s}_{0,\ell}^H \end{bmatrix} \tag{3.179}$$

where $\boldsymbol{b}'_i = 0$ if $i > nb' = nb + nn$.

Our polynomial feedforward filter is now, using (3.150), given by

$$\boldsymbol{s}(q^{-1}) = \boldsymbol{s}_0(q^{-1})\boldsymbol{M}_0^{-1}\boldsymbol{N}(q^{-1}). \tag{3.180}$$

3.A.3 Fixed Order FIR-DFE Wiener Equations in Channel Parameters

The covariance matrices $\boldsymbol{R}_{\varphi\varphi}$ and $\boldsymbol{R}_{\varphi d_\ell}$ in (3.51) can be decomposed as

$$\boldsymbol{R}_{\varphi\varphi} = \begin{bmatrix} \boldsymbol{R}_{\bar{\boldsymbol{y}}\bar{\boldsymbol{y}}} & -\boldsymbol{R}_{\bar{\boldsymbol{y}}\boldsymbol{d}_Q} \\ -\boldsymbol{R}_{\boldsymbol{d}_Q\bar{\boldsymbol{y}}} & \boldsymbol{I} \end{bmatrix} \tag{3.181}$$

where

$$\boldsymbol{R}_{\bar{\boldsymbol{y}}\bar{\boldsymbol{y}}} \triangleq E[\bar{\boldsymbol{y}}\bar{\boldsymbol{y}}^H] \quad \boldsymbol{R}_{\bar{\boldsymbol{y}}\boldsymbol{d}_Q} \triangleq E[\bar{\boldsymbol{y}}\boldsymbol{d}_Q^H] \quad \boldsymbol{R}_{\boldsymbol{d}_Q\bar{\boldsymbol{y}}} \triangleq E[\boldsymbol{d}_Q\bar{\boldsymbol{y}}^H] = \boldsymbol{R}_{\bar{\boldsymbol{y}}\boldsymbol{d}_Q}^H \tag{3.182}$$

and

$$\boldsymbol{R}_{\varphi d_\ell} = \begin{bmatrix} \boldsymbol{R}_{\bar{\boldsymbol{y}}d_\ell} \\ -\boldsymbol{R}_{\boldsymbol{d}_Q d_\ell} \end{bmatrix} \tag{3.183}$$

where

$$\boldsymbol{R}_{\bar{\boldsymbol{y}}d_\ell} \triangleq E[\bar{\boldsymbol{y}}d^H(t-\ell)] \quad \boldsymbol{R}_{\boldsymbol{d}_Q d_\ell} \triangleq E[\boldsymbol{d}_Q d^(t-\ell)] = 0. \tag{3.184}$$

By using (3.38), and the assumption that $d(t)$ and $\boldsymbol{n}(t)$ are uncorrelated, we can note that

$$\boldsymbol{R}_{\bar{\boldsymbol{y}}\bar{\boldsymbol{y}}} = \boldsymbol{R}_{\bar{\boldsymbol{y}}_s\bar{\boldsymbol{y}}_s} + \boldsymbol{\mathcal{R}}_{\bar{\boldsymbol{n}}\bar{\boldsymbol{n}}} \tag{3.185}$$

where $\boldsymbol{R}_{\bar{\boldsymbol{y}}_s\bar{\boldsymbol{y}}_s}$ is the covariance of the desired signal part of $\bar{\boldsymbol{y}}(t)$ in (3.48)

$$\boldsymbol{R}_{\bar{\boldsymbol{y}}_s\bar{\boldsymbol{y}}_s} = \begin{bmatrix} \sum_{k=0}^{nb} \boldsymbol{b}_k\boldsymbol{b}_k^H & \cdots & \sum_{k=ns}^{nb} \boldsymbol{b}_k\boldsymbol{b}_{k-ns}^H \\ \vdots & \ddots & \vdots \\ \sum_{k=0}^{nb-ns} \boldsymbol{b}_k\boldsymbol{b}_{k+ns}^H & \cdots & \sum_{k=0}^{nb} \boldsymbol{b}_k\boldsymbol{b}_k^H \end{bmatrix} \tag{3.186}$$

where a sum is zero if the starting index is greater than the end index in the sum. The vectors $\boldsymbol{b}_k$ are the vector taps in the channel $\boldsymbol{b}(q^{-1})$ in (3.38). The matrix $\boldsymbol{\mathcal{R}}_{\bar{\boldsymbol{n}}\bar{\boldsymbol{n}}}$ is the covariance of the noise plus interference part of $\bar{\boldsymbol{n}}(t)$ in (3.48)

$$\boldsymbol{\mathcal{R}}_{\bar{\boldsymbol{n}}\bar{\boldsymbol{n}}} = \begin{bmatrix} \boldsymbol{R}_{\boldsymbol{nn}}(0) & \ldots & \boldsymbol{R}_{\boldsymbol{nn}}(ns) \\ \vdots & \ddots & \vdots \\ \boldsymbol{R}_{\boldsymbol{nn}}(-ns) & \ldots & \boldsymbol{R}_{\boldsymbol{nn}}(0) \end{bmatrix}. \tag{3.187}$$

The $M \times M$ matrices $\boldsymbol{R}_{\boldsymbol{nn}}(k)$ here are evaluations of the autocovariance function of the noise plus interference defined in (2.108). Furthermore, $\boldsymbol{R}_{\bar{\boldsymbol{y}}\boldsymbol{d}_Q}$ and $\boldsymbol{R}_{\bar{\boldsymbol{y}}d_\ell}$ in (3.184) can be expressed as

$$\boldsymbol{R}_{\bar{\boldsymbol{y}}\boldsymbol{d}_Q} = \begin{bmatrix} \boldsymbol{b}_{\ell+1} & \cdots & \boldsymbol{b}_{\ell+1+nq} \\ \vdots & \ddots & \vdots \\ \boldsymbol{b}_{\ell+1-ns} & \cdots & \boldsymbol{b}_{\ell+1+nq-ns} \end{bmatrix} \tag{3.188}$$

$$\boldsymbol{R}_{\bar{\boldsymbol{y}}d_\ell} = \begin{bmatrix} \boldsymbol{b}_\ell \\ \boldsymbol{b}_{\ell-1} \\ \vdots \\ \boldsymbol{b}_0 \\ 0 \\ \vdots \\ 0 \end{bmatrix}. \tag{3.189}$$

Chapter 4

Space-Time ML Sequence Estimation

4.1 Introduction

The use of maximum likelihood sequence estimation (MLSE) [26, 116] will in general provide higher performance than decision feedback equalization. If the true channels are known the MLSE finds the transmitted *sequence* that maximizes the probability of the actual received signal. This can be compared with the DFE that after a linear filtering makes decisions on a *symbol by symbol* basis. When we perform space-time filtering this advantage of the MLSE over the DFE is however not always as large as in the exclusively temporal case. The reason for this is that both the MLSE and the DFE contains linear filters as front-ends. The linear filtering can more easily retrieve the desired signal when performing space-time processing as opposed to when performing scalar temporal processing. The potential for additional improvement due to the subsequent processing is therefore reduced. Sometimes the DFE can even have better performance than the MLSE. It can happen when the information about the channel is incomplete.

We will here consider the generalization of the scalar MLSE to the space-time MLSE. The scalar maximum likelihood sequence estimation [26, 116] can be generalized to work with an array of antennas. We call such an MLSE a *multi-channel MLSE.*

Several different, but basically equivalent, realizations of the multi-channel MLSE can be found in the literature [71, 121, 41, 45, 13, 6, 100]. Three approaches will be presented here, the log-likelihood [13, 41, 45], the noise whitening approach [6] and the multi-dimensional matched filtering approach [71, 121, 100], are presented here. Although these approaches are equivalent in terms of their performance, one of them, the multi-dimensional matched filter approach, is superior with respect to complexity when using more than one antenna.

Handling of spatially colored noise and interference is readily performed with the multi-channel MLSE. The temporal color of the noise and interference can also be included in the formalism, at the price of an increased memory length in the Viterbi algorithm. This will however increase the complexity of the processing. By using the temporal spectrum of the noise and interference the space-time MLSE can however perform spatio-temporal cancellation of interferers.

In Section 4.4, different tuning options for the filters in the MMF-MLSE is discussed. Several options can be conceived for the tuning. We will, however, only compare two options: Direct tuning of the filters and indirect tuning utilizing an estimate of the channel and an estimate of the spatial spectrum of the noise plus interference covariance matrix. The direct tuning of the filters of the MMF-DFE is not well conditioned when using many antenna elements due to the large number of coefficients in the filters, It thus requires a regularization. In the simulations conducted, the performance of the regularized directly tuned MMF-MLSE is however still worse then the indirectly tuned MMF-DFE. The indirect tuning is here much more well conditioned, which results in a better performance.

The indirectly tuned MMF-MLSE used in the simulations only utilizes the spatial spectrum of the interference. As a result it will only suppress co-channel interferers spatially. When using a relatively large number of antennas, in comparison to the length of the training sequence, it can be difficult to achieve good estimates of the spatio-temporal spectrum of the noise plus interference. We can however consider performing joint FIR channel and AR noise model estimation as described in Section 2.9 of Chapter 2. Such a model for the noise and interference will allow for spatio-temporal suppression of co-channel interferers [7]. Another alternative is to use a hybrid MLSE as in [53], or almost equivalent, a reduced rank tuned MLSE as described in Section 5.4 of Chapter 5. In Section 5.4, we show that these two methods are almost equivalent and therefore the hybrid MLSE of [53] can be viewed as a rank one MLSE, see Chapter 5. It appears, however, that the methods in [7] and [53], as well as the reduced rank tuned MLSE in Section 2.9, all require relatively large signal to noise ratios.

4.2 Different Implementations of the Multi-Channel MLSE

4.2.1 Channel Description

As in (2.15) we will model the received signals at the M antenna elements,

$$\boldsymbol{y}(t) = [y_1(t) \quad y_2(t) \quad \dots \quad y_M(t)]^T \tag{4.1}$$

as

$$\boldsymbol{y}(t) = \boldsymbol{b}(q^{-1})d(t) + \boldsymbol{n}(t) \tag{4.2}$$

where $\boldsymbol{b}(q^{-1}) = [b_1(q^{-1}) \quad b_2(q^{-1}) \quad \dots \quad b_M(q^{-1})]^T$ represents the causal FIR channels to the antenna elements for the transmitted scalar symbol sequence $d(t)$ and the vector $\boldsymbol{n}(t)$ represents the noise plus interference. For simplicity we will assume all M channel polynomials $b_i(q^{-1})$ to have the order nb.

4.2.2 Log-Likelihood Metric and Noise Whitening Approach

Consider the spectrum of the noise plus interference $\boldsymbol{n}(t)$, where we have replaced z with q

$$\boldsymbol{R_{nn}}(q, q^{-1}) = \sum_{m=-\infty}^{\infty} E[\boldsymbol{n}(t)\boldsymbol{n}^H(t-m)]q^{-m} \tag{4.3}$$

We will call this the noise plus interference spectrum operator. We can factor this spectrum operator as

$$\boldsymbol{R_{nn}}(q, q^{-1}) = \boldsymbol{R}_{\boldsymbol{nn}}^{1/2}(q^{-1})\boldsymbol{R}_{\boldsymbol{nn}}^{H/2}(q) \tag{4.4}$$

where $\boldsymbol{R}_{\boldsymbol{nn}}^{1/2}(q^{-1})$ is the causal part of the spectral factorization of the noise plus interferer spectrum operator [47, 3, 40, 36] and where $\boldsymbol{R}_{\boldsymbol{nn}}^{H/2}(q)$ is the complex conjugate transpose of $\boldsymbol{R}_{\boldsymbol{nn}}^{1/2}(q^{-1})$.

We can now form the whitened noise plus interference

$$\boldsymbol{n}'(t) = \boldsymbol{R}_{\boldsymbol{nn}}^{-1/2}(q^{-1})\boldsymbol{n}(t). \tag{4.5}$$

A straightforward method of deriving the MLSE is to select the sequence $\{d(t)\}_{t=1}^{N}$ that maximizes the probability of the received sequence[1], $\{\boldsymbol{y}(t)\}_{t=1}^{N}$. If the noise and interference is, for simplicity, assumed Gaussian, then this probability can be expressed as:

$$\begin{aligned}
&P\left(\{\boldsymbol{y}(t)\}_{t=1}^{N} | \{d(t)\}_{t=1}^{N}\right) \propto \\
&\quad exp(-\boldsymbol{n}'(t)^{H}\boldsymbol{n}'(t)) = \\
&\quad exp\Bigg(- \sum_{t=1}^{N} \left[[\boldsymbol{R}_{\boldsymbol{nn}}^{-1/2}(q^{-1}) \left\{ \boldsymbol{y}(t) - \boldsymbol{b}(q^{-1})d(t) \right\} \right]^{H} \\
&\quad \times \left[\boldsymbol{R}_{\boldsymbol{nn}}^{-1/2}(q^{-1}) \left\{ \boldsymbol{y}(t) - \boldsymbol{b}(q^{-1})d(t) \right\} \right] \Bigg).
\end{aligned} \tag{4.6}$$

Maximizing the probability in (4.6), is equivalent to minimizing the log-likelihood metric[2]

$$\begin{aligned}
\mu_{LL}(N) = \sum_{t=1}^{N} &\left[\boldsymbol{R}_{\boldsymbol{nn}}^{-1/2}(q^{-1}) \left\{ \boldsymbol{y}(t) - \boldsymbol{b}(q^{-1})d(t) \right\} \right]^{H} \\
&\times \left[\boldsymbol{R}_{\boldsymbol{nn}}^{-1/2}(q^{-1}) \left\{ \boldsymbol{y}(t) - \boldsymbol{b}(q^{-1})d(t) \right\} \right].
\end{aligned} \tag{4.7}$$

By forming the filtered signal,

$$\boldsymbol{y}'(t) = \boldsymbol{R}_{\boldsymbol{nn}}^{-\frac{1}{2}}(q^{-1})\boldsymbol{y}(t) \tag{4.8}$$

and the potentially *rational* noise whitened channel

$$\boldsymbol{b}'(q^{-1}) = \boldsymbol{R}_{\boldsymbol{nn}}^{-\frac{1}{2}}(q^{-1})\boldsymbol{b}(q^{-1}) \tag{4.9}$$

[1]In this discussion we are neglecting the effects near the beginning and the end of the data records.

[2]Since e^{-x} is monotonically decreasing.

the log-likelihood metric (4.7) can be expressed as

$$\mu_{LL}(N) = \sum_{t=1}^{N} \left[\boldsymbol{y}'(t) - \boldsymbol{b}'(q^{-1})d(t)\right]^H \left[\boldsymbol{y}'(t) - \boldsymbol{b}'(q^{-1})d(t)\right]. \qquad (4.10)$$

The metric (4.10) can be computed recursively, for $t = 1, \dots, N$, as

$$\mu_{LL}(t) = \mu_{LL}(t-1) + \left[\boldsymbol{y}'(t) - \boldsymbol{b}'(q^{-1})d(t)\right]^H \left[\boldsymbol{y}'(t) - \boldsymbol{b}'(q^{-1})d(t)\right] \qquad (4.11)$$

The sequence of transmitted symbols $\{d(t)\}$ that minimizes this metric for a given sequence of received signals $\{\boldsymbol{y}(t)\}$ and channel model $\boldsymbol{b}'(q^{-1})$ can be determined by using the Viterbi algorithm [122]. In the Viterbi algorithm we simply replace the standard scalar metric computation with the above vector formulation. Approaches similar to this can, for example, be found in [41], [13] and [45]. Since the Viterbi algorithm here works with a vector input, $\boldsymbol{y}(t)$, we call it a *vector Viterbi*. The block diagram for an MLSE implemented using this log-likelihood approach is depicted in Figure 4.1a.

If desired, we can move the whitening filter outside the vector Viterbi algorithm as shown in Figure 4.1b. We call this the *noise whitening approach*.

Note that if the noise process, $\boldsymbol{n}(t)$, is modeled by an AR-process

$$\boldsymbol{n}(t) = \boldsymbol{N}^{-1}(q^{-1})\boldsymbol{M}_0\boldsymbol{v}(t) \qquad (4.12)$$

where $\boldsymbol{N}^{-1}(q^{-1})$ is an $M \times M$ polynomial matrix, $\boldsymbol{M}_0$ is a constant matrix, and $\boldsymbol{v}(t)$ is an $M \times 1$ white innovation process, then the noise whitening filter $\boldsymbol{R}_{\boldsymbol{nn}}^{-1/2}(q^{-1})$ is simply given by

$$\boldsymbol{R}_{\boldsymbol{nn}}^{-1/2}(q^{-1}) = \boldsymbol{M}_0^{-1}\boldsymbol{N}(q^{-1}). \qquad (4.13)$$

4.2.3 Multi-Dimensional Matched Filter Approach

Another formulation of the multi-channel MLSE is to formulate it in terms of a multi-dimensional matched filter as in [71], [121] and [97]. Here the vector Viterbi is replaced by a multi-dimensional matched filter followed by a scalar Viterbi. An advantage with this approach is that the computational complexity is reduced when many channels (antennas) are used. These

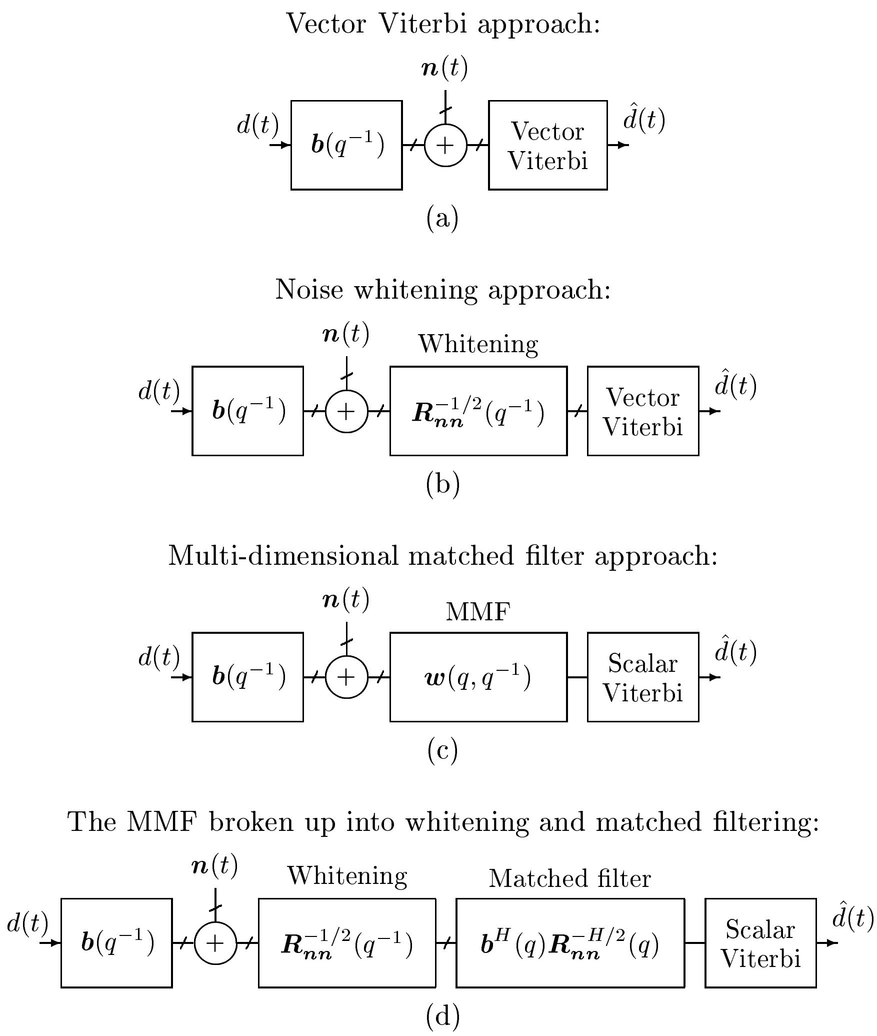


Figure 4.1: Different implementations of the multi-channel MLSE.

formulations can be seen as generalizations of the treatment by Ungerboeck in [116] to the case with multiple channels.

Using $\boldsymbol{y}'(t)$ and $\boldsymbol{b}'(q^{-1})$ (with coefficients $\boldsymbol{b}'_k$) from equations (4.8) and (4.9), the matched filter version can be derived from the log-likelihood metric in (4.7). In Appendix 4.A.1 it is shown that

$$
\begin{aligned}
\mu_{LL}(N) = & \sum_{t=1}^{N} \left[\boldsymbol{R}_{\boldsymbol{nn}}^{-1/2}(q^{-1}) \left\{ \boldsymbol{y}(t) - \boldsymbol{b}(q^{-1})d(t) \right\} \right]^H \\
& \times \left[\boldsymbol{R}_{\boldsymbol{nn}}^{-1/2}(q^{-1}) \left\{ \boldsymbol{y}(t) - \boldsymbol{b}(q^{-1})d(t) \right\} \right] \\
& = -2\mathrm{Re} \left\{ \sum_{n=1}^{N} z^H(n)d(n) \right\} + \sum_{n=1}^{N} \sum_{m=1}^{N} d^H(n)\gamma_{n-m}d(m) \\
& + f(\boldsymbol{y}(\cdot)) + e_{\text{border}}
\end{aligned}
$$

where the term e_{border} is a correction term that only depends on the values of $\boldsymbol{y}(t)$ and $d(t)$ for t close to 1 or N. The scalar signal $z(t)$ is defined by

$$
\begin{aligned}
z(t) & \stackrel{\triangle}{=} \boldsymbol{b}'^H(q)\boldsymbol{y}'(t) = \boldsymbol{b}^H(q)\boldsymbol{R}_{\boldsymbol{nn}}^{-H/2}(q)\boldsymbol{R}_{\boldsymbol{nn}}^{-1/2}(q^{-1})\boldsymbol{y}(t) \\
& = \boldsymbol{b}^H(q)\boldsymbol{R}_{\boldsymbol{nn}}^{-1}(q, q^{-1})\boldsymbol{y}(t)
\end{aligned}
$$

where we have used (4.4) in the last equality.

By introducing the *multi-dimensional matched filter*

$$
\boldsymbol{w}(q, q^{-1}) = \boldsymbol{b}^H(q)\boldsymbol{R}_{\boldsymbol{nn}}^{-1}(q, q^{-1}) \tag{4.14}
$$

we can write $z(t)$ as

$$
z(t) = \boldsymbol{w}(q, q^{-1})\boldsymbol{y}(t). \tag{4.15}
$$

The coefficients γ_k are the coefficients of the double sided *metric polynomial*

$$
\gamma(q, q^{-1}) = \sum_{p=-\infty}^{\infty} {\boldsymbol{b}'_{p-n}}^H \boldsymbol{b}'_{p-m} q^{n-m} = \boldsymbol{b}^H(q)\boldsymbol{R}_{\boldsymbol{nn}}^{-1}(q, q^{-1})\boldsymbol{b}(q^{-1}) \tag{4.16}
$$

with the coefficients numbered as

$$
\gamma(q, q^{-1}) = \gamma_{-n\gamma}q^{n\gamma} + \ldots + \gamma_0 + \ldots + \gamma_{n\gamma}q^{-n\gamma}. \tag{4.17}
$$

The term $f(\boldsymbol{y}(\cdot))$ does not depend on the candidate symbols, $d(\cdot)$, and can therefore be neglected in the maximization of μ_{LL}. Furthermore, we will neglect the border correction term e_{border} since it will only have effect on the estimated symbols close to the borders. Thus, neglecting the "border effects" and changing the sign such that the *matched filter metric* is to be *maximized* gives

$$\mu_{MF}(N) = 2\mathrm{Re}\left\{\sum_{n=1}^{N} z^H(n)d(n)\right\} - \sum_{n=1}^{N}\sum_{m=1}^{N} d^H(n)\gamma_{n-m}d(m). \tag{4.18}$$

Since $\gamma(q, q^{-1})$ is complex conjugate symmetric, this metric can be recursively computed as

$$\mu_{MF}(t) = \mu_{MF}(t-1) + \mathrm{Re}\left\{d^H(t)(2z(t) - \gamma_0 d(t) - 2\sum_{m=1}^{n\gamma} \gamma_m d(t-m))\right\}. \tag{4.19}$$

Note also that the memory length in the Viterbi algorithm using the matched filter metric in (4.18), is the same as for the Viterbi algorithm using the log-likelihood metric in (4.10).

If we have an AR model for the noise $\boldsymbol{n}(t)$, as in (4.12), then the multi-dimensional matched filter takes on the simple form of a row vector of non-causal FIR-filters

$$\boldsymbol{w}(q, q^{-1}) = \boldsymbol{b}^H(q)\boldsymbol{N}^H(q)\boldsymbol{M}_0^{-H}\boldsymbol{M}_0^{-1}\boldsymbol{N}(q^{-1}) \tag{4.20}$$

and, similarly, the metric polynomial has the simple form

$$\gamma(q, q^{-1}) = \boldsymbol{b}^H(q)\boldsymbol{N}^H(q)\boldsymbol{M}_0^{-H}\boldsymbol{M}_0^{-1}\boldsymbol{N}(q^{-1})\boldsymbol{b}(q^{-1}). \tag{4.21}$$

Note that both the MMF-filter and the metric polynomials have impulse responses with finite extent, both forward and backward in time.

Note that when we take the temporal color of the noise plus interference into consideration, then the length of the metric polynomial $\gamma(q, q^{-1})$ is increased. Specifically when we have an AR model of order nn for the noise plus interference, then the order $n\gamma$ for the double-sided metric polynomial is increased by nn. From (4.19) we can see that this will increase the memory length of the Viterbi algorithm. This will in turn increase the execution complexity of the Viterbi algorithm by a factor L^{nn} where L is the number of symbols in the alphabet [82]. We can thus see that the complexity of the Viterbi algorithm increases exponentially with the order of the AR noise model. If the order nn is not too large, the increase in complexity may be acceptable.

Due to the matched filtering, *only a scalar Viterbi algorithm is required.* This is the case since the Viterbi algorithm operates on the scalar signal $z(t)$. As a result the complexity of the metric computation in the Viterbi algorithm is reduced.

A block diagram of the MLSE using the multi-dimensional matched filter approach can be seen in Figure 4.1c. Note also that the multi-dimensional matched filter can be broken up into a noise whitening part, $\boldsymbol{R}_{\boldsymbol{nn}}^{-1/2}(q^{-1})$, and a filter matched to the overall resulting SIMO channel, $\boldsymbol{b}^H(q)\boldsymbol{R}_{\boldsymbol{nn}}^{-H/2}(q)$. This is illustrated in Figure 4.1d.

When we have a noise that is temporally colored by anything but an AR process, a problem arises. The filter $\boldsymbol{R}_{\boldsymbol{nn}}^{-1}(q, q^{-1})$ will then have a double sided infinite impulse response. This results in an infinite memory length in the Viterbi algorithm. In such situations, either the metric only or both the metric and the MMF has to be truncated. The same problem will arise in the log-likelihood and noise whitening implementations.

In the derivation of the matched-filter metric we neglected the effect of symbols and signals close to the beginning and the end of the symbol sequence. In effect this was also done when deriving the log-likelihood metric. To get proper estimates of the symbols close to the edges of the sequence, we have to initialize the filters properly.

4.3 Computational Complexity

We here consider the number of operations required when using the log-likelihood and the matched filter metrics, respectively.

Let us assume that the channel $\boldsymbol{b}'(q)$ is approximated by an FIR filter with $nb + nr + 1$ taps. This would be the case if we used $2nr + 1$ taps in the noise plus interference spectrum operator inverse, $\boldsymbol{R}_{\boldsymbol{nn}}^{-1}(q, q^{-1})$ and $nb+1$ taps in the channels. If L is the number of symbols in the alphabet, a Viterbi algorithm then requires, $L^{nb+nr+1}$ metric updates for each detected symbol.

If the channel is assumed stationary, then the computational complexity of the log-likelihood metric and the whitening filter approach (Section 4.2.2), measured by the number of complex multiplications and additions per symbol detected, will be

$$C_{LL} \approx M^2(nr+1) + M(nb+2)L^{(nb+nr+1)} \tag{4.22}$$

where M is the number of channels.

Because the matched-filter metric operates a scalar Viterbi-algorithm, the complexity for this approach, outlined in Section 4.2.3, will be

$$C_{MF} \approx M(2nr+nb+1) + (nb+2)L^{(nb+nr+1)} \tag{4.23}$$

where the first term refers to the complexity in executing the multi-dimensional matched filter and the second term represents the complexity in computing the metric. It can be seen from these expressions, that if the number of antennas is more than one, then the matched-filter metric has a considerable advantage over the log-likelihood metric.

4.4 Tuning the Multi-Dimensional Matched Filter

The multi-dimensional matched filter, $\boldsymbol{w}(q, q^{-1})$, and the metric polynomial, $\gamma(q, q^{-1})$, can be tuned in a few different ways.

4.4.1 Direct MMSE Tuning

We assume here that a short training sequence of the transmitted signal $d(t)$ is known and we want to tune the multi-dimensional matched filter, $\boldsymbol{w}(q, q^{-1})$, and the metric polynomial $\gamma(q, q^{-1})$ using this training sequence and the corresponding received signal samples $\boldsymbol{y}(t)$.

In [71] and [121], generalizations of the direct tuning approach developed in [116] are presented. The coefficients of a feedforward filter, $\boldsymbol{w}(q, q^{-1})$,

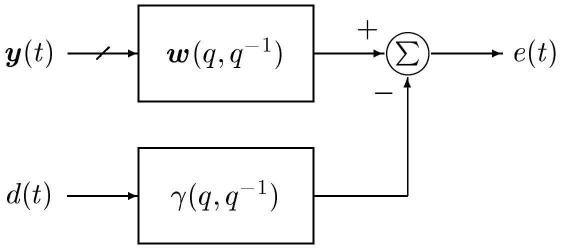


Figure 4.2: MMF filter tuning.

and the coefficients of a non-causal feedback filter, $\gamma(q, q^{-1})$, are tuned to minimize the mean square error (MSE) of the error signal,

$$e(t) = \boldsymbol{w}(q, q^{-1})\boldsymbol{y}(t) - \gamma(q, q^{-1})d(t). \tag{4.24}$$

See Figure 4.2. The polynomial row vector

$$\boldsymbol{w}(q, q^{-1}) = [w_1(q, q^{-1}) \ \ \ldots \ \ w_M(q, q^{-1})] \tag{4.25}$$

is a MISO FIR filter and $\gamma(q, q^{-1})$ is a double sided, complex conjugate symmetric[3], non-causal FIR-filter with the middle coefficient, γ_0, constrained to be equal to one. That is,

$$\gamma(q, q^{-1}) = \gamma_{-n\gamma} q^{+n\gamma} + \ \ \ldots \ \ + 1 + \ \ \ldots \ \ + \gamma_{n\gamma} q^{-n\gamma} \tag{4.26}$$

with $\gamma_{-k} = \gamma_k^H$. By this minimization, noise whitening and matched filtering will be performed by $\boldsymbol{w}(q, q^{-1})$, while $\gamma(q, q^{-1})$ will contain the overall impulse response [121].

We can from this see that the multidimensional matched filter, as all matched filters, maximizes the peak-to-noise ration (PNR), i.e. it maximizes the peak in the impulse response of channel for the desired signal over the noise variance. In other words, it maximizes

$$\frac{\gamma_0}{E[e(t)e^H(t)]} \ . \tag{4.27}$$

[3]If one so desires one can relax the requirement that the metric polynomial should be complex conjugate symmetric.

This is equivalent to minimizing the variance in the error signal $e(t)$

$$E[e(t)e^H(t)] \tag{4.28}$$

while constraining γ_0 to one (or any other value > 0).

It is natural to choose the structure of the feedforward filter, $\boldsymbol{w}(q, q^{-1})$, consistent with an ideal MMF with a truncated noise plus interference spectrum operator, $\boldsymbol{b}^H(q)\tilde{\boldsymbol{R}}_{\boldsymbol{nn}}^{-1}(q, q^{-1})$. The spectrum operator, $\tilde{\boldsymbol{R}}_{\boldsymbol{nn}}^{-1}(q, q^{-1})$, here represents a truncated version of $\boldsymbol{R}_{\boldsymbol{nn}}^{-1}(q, q^{-1})$. The filter $\boldsymbol{w}(q, q^{-1})$ will thus be non-causal or anti-causal, since $\boldsymbol{b}^H(q)$ is anti-causal and $\tilde{\boldsymbol{R}}_{\boldsymbol{nn}}^{-1}(q, q^{-1})$ is either a matrix constant or a double-sided polynomial matrix. It is also natural to choose the number of coefficients in $\gamma(q, q^{-1})$ consistent with the structure chosen for $\boldsymbol{w}(q, q^{-1})$.

The estimates $\hat{\boldsymbol{w}}(q, q^{-1})$ and $\hat{\gamma}(q, q^{-1})$ can be found either adaptively or by solving a system of equations formed directly from the training data. Convergence to the ideal solution will result in an error signal which is white, with minimal variance.

When the true filter orders are used and the training sequence is long enough, the MMF will be given by the estimate $\hat{\boldsymbol{w}}(q, q^{-1})$, up to a multiplicative constant, and the corresponding metric to be used in the Viterbi algorithm will be given by the $\hat{\gamma}(q, q^{-1})$, also up to a multiplicative constant [121].

A problem arises if the available training sequence is short. If for instance the number of training symbols is smaller than the number of coefficients in the filters, then the coefficients cannot be determined uniquely. A regularization of the equations can then be introduced. By adding artificial noise into the system of equations, a solution can be computed, but it will in general be inferior to the true matched filter. However, if the number of antennas is small, say $M = 2$, then a short training sequence can suffice to properly tune the MLSE with a direct method.

By adjusting the number of coefficients in $\boldsymbol{w}(q, q^{-1})$ and $\gamma(q, q^{-1})$, the ability to combat a temporally colored interference can be obtained. Adding more coefficients increases the temporal noise whitening ability as well as the matched filtering capability of the filter at the price of more degrees of freedom and a longer memory in the metric.

4.4.2 Indirect MMF Tuning

In the indirect approach the channel, $\boldsymbol{b}(q^{-1})$, is first estimated. This can be done with one of the methods described in Chapter 2. An advantage with this method, compared to the direct method, is that we can take advantage of *a priori* knowledge when estimating the channel.

The noise plus interference spectrum operator, $\boldsymbol{R_{nn}}(q, q^{-1})$, can, for example, be estimated using the residuals from the identification procedure. If we have a relatively large number of antennas and the amount of training data is small, we may choose to only estimate the spatial spectrum of the noise plus interference, i.e. the coefficient matrix for lag zero in $\boldsymbol{R_{nn}}(q, q^{-1})$. A good option is to model the noise plus interference as an AR process and use this in the spectrum estimation [7]. Modeling the noise plus interference spectrum as an AR process helps to catch some useful temporal aspects of the noise plus spectrum and it fits well into the MLSE algorithm. The resulting MMF filter and the metric will then be FIR filters with finite lengths as in (4.20) and in (4.21). As the number of antennas or the order of the AR noise model is increased, the number of parameters in the model can however become large compared to the number of available equations. This can make them potentially difficult to estimate accurately, especially if the SNR is not high enough.

If the spectrum operator is invertible, estimates of the MMF, $\boldsymbol{w}(q, q^{-1})$, and of the metric polynomial, $\gamma(q, q^{-1})$, can then be formed as

$$\hat{\boldsymbol{w}}(q, q^{-1}) = \hat{\boldsymbol{b}}^H(q)\hat{\boldsymbol{R}}_{\boldsymbol{nn}}^{-1}(q, q^{-1}) \tag{4.29}$$

$$\hat{\gamma}(q, q^{-1}) = \hat{\boldsymbol{b}}^H(q)\hat{\boldsymbol{R}}_{\boldsymbol{nn}}^{-1}(q, q^{-1})\hat{\boldsymbol{b}}(q^{-1}) \tag{4.30}$$

where the "hat" marks quantities derived from the estimated channel, $\hat{\boldsymbol{b}}(q^{-1})$, or the estimated noise spectrum operator, $\hat{\boldsymbol{R}}_{\boldsymbol{nn}}(q, q^{-1})$. When using an AR model for the noise the MMF-filter and the metric polynomial take on the simple forms in (4.20) and in (4.21). If the filter $\hat{\boldsymbol{R}}_{\boldsymbol{nn}}^{-1}(q, q^{-1})$ is a double sided IIR filter, it will have to be truncated. Otherwise the metric in the Viterbi would have infinite memory.

If we use joint FIR channel and AR noise model estimation as described in Section 2.9, we can from the least squares estimates, $\hat{\boldsymbol{N}}(q^{-1})$ and $\hat{\boldsymbol{b}}_{\boldsymbol{N}}(q^{-1})$, of $\boldsymbol{N}(q^{-1})$ and $\boldsymbol{b}_{\boldsymbol{N}}(q^{-1})$ in (2.121) and the estimate $\hat{\boldsymbol{R}}_{\hat{\boldsymbol{r}}\hat{\boldsymbol{r}}}$ from (2.123), form the multidimensional matched filter $\hat{\boldsymbol{w}}(q, q^{-1})$ and the metric polynomial

$\hat{\gamma}(q, q^{-1})$ as

$$\hat{\boldsymbol{w}}(q, q^{-1}) = \hat{\boldsymbol{b}}_{\boldsymbol{N}}^{H}(q)\hat{\boldsymbol{R}}_{\hat{\boldsymbol{r}}\hat{\boldsymbol{r}}}^{-1}\hat{\boldsymbol{N}}(q^{-1}) \tag{4.31}$$

and the metric polynomial

$$\hat{\gamma}(q, q^{-1}) = \hat{\boldsymbol{b}}_{\boldsymbol{N}}^{H}(q)\hat{\boldsymbol{R}}_{\hat{\boldsymbol{r}}\hat{\boldsymbol{r}}}^{-1}\hat{\boldsymbol{b}}_{\boldsymbol{N}}(q^{-1}) \tag{4.32}$$

An interesting question to study, is if the indirect methods can handle a case with very low signal-to-interference ratio (SIR). It could be suggested, that very poor SIR would make estimation of the channels to the individual antenna elements non-feasible. Although the quality of the estimated channels may be compromised, the simulations for the scenario presented here do not show that the indirect methods suffer much from this. This issue will be discussed further in Chapter 7.

4.4.3 Indirect MMSE Tuning

An alternative indirect way of tuning the multi-dimensional matched filter is to perform the minimization of the MSE of the error signal $e(t)$ in (4.24), but instead of forming the systems of equations directly from data, we form them from the estimated channel, $\hat{\boldsymbol{b}}(q^{-1})$, and the estimated noise spectrum operator, $\hat{\boldsymbol{R}}_{\boldsymbol{nn}}(q, q^{-1})$. The number of matrix coefficients of $\hat{\boldsymbol{R}}_{\boldsymbol{nn}}(q, q^{-1})$ used, and the structure and length of the filters $\boldsymbol{w}(q, q^{-1})$ and $\gamma(q, q^{-1})$, affect the temporal noise whitening and matched filtering capabilities. By constraining the filter structures, the memory length in the Viterbi algorithm can be controlled. This will affect the complexity of the Viterbi algorithm.

This indirect version of the MMSE tuning will have an advantage over the direct version. If we cannot obtain a good estimate of the space-time covariance matrix of the noise plus interference, we may here restrict the algorithm to use only the estimate of the spatial covariance of the noise plus interference or for example an estimated AR spectrum. This can result in better performance.

If the same structure is used and the same estimate of the channels and the noise plus interference spectrum is used, then this indirect method and the ordinary indirect MMF tuning discussed in Section 4.4.2 are equivalent.

The methods differ only in the way they handle *temporally* colored noise plus interference. The indirect MMSE tuning of the MMF has a potential advantage, over the ordinary indirect MMF tuning, in that it for a given filter structure, and a given noise plus interference spectrum used, finds the filter coefficients that perform a compromise between noise whitening and matched filtering. The performance of this scheme has however not been investigated here.

4.5 Simulations

The purpose of the simulations presented here is to study how the indirect methods can handle poor signal to interference ratios and to demonstrate that the direct method has problems with short training sequences.

The algorithms compared are the indirect method (either one), when using only the spatial color of the noise plus interference, and two versions of the direct method. One without regularization and one with artificial noise with a variance equal to the real noise variance, added to the diagonal of the system of equations. In all cases 5 taps in $\boldsymbol{w}(q, q^{-1})$ and 9 taps in $\gamma(q, q^{-1})$, were used. The channels for the indirect method were estimated with the standard least squares method with 5 taps.

The algorithms were tested using a circular array with eight antennas equally spaced along a circle with a radius of 0.5 wave lengths. The desired signals arrives from the directions 0,30,-60 and 180 degrees, respectively. The respective channels are $1 + 0.5q^{-1}$, $0.5q^{-1} + 0.8q^{-2}$, $0.5q^{-2} + 0.2q^{-3}$ and $0.2q^{-3} + 0.3q^{-4}$. Two-tap channels are chosen to simulate imperfect sampling timing or partial response modulation. Binary symbols, $d(t) = \pm 1$, are used. Co-channel interferers impinge on the antenna array, also through two tap channels, from the directions -30, 135 and -135 degrees respectively. White Gaussian noise is added.

In Figure 4.3, the BER for the different algorithms can be seen as a function of the SIR. From Figure 4.3, we can note that the indirect method does not suffer significantly in this scenario from the poor SIR. The direct methods, of course, perform poorly here since the training sequence was only 26 symbols long, leaving it with more parameters to tune than number of equations.

In Figure 4.4, the BER is presented as a function of the length of the training sequence used. When the training sequence length increases the performance of the direct methods approaches that of the indirect method. The indirect method performs better for the short training sequencers since it here focuses on only the spatial color of the noise plus interference.

It should be noted that when performing direct tuning, the number of equations is equal to the effective length of the training sequence, while when performing indirect tuning we have MN_{eff}, equations, where M is the number of antennas and N_{eff} is the effective length of the training sequence.

The number of parameters to tune when performing direct tuning is $M(nw+1)+2n\gamma$, where nw is the order of the multidimensional matched filter and $n\gamma$ is the order of the double sided metric polynomial. When performing indirect tuning, we have $M(nb+nn+1)$ parameters to estimate if we perform joint FIR channel and AR noise model estimation, plus we need some extra equations in order to estimate the spatial color of the residuals. The number of parameters for the direct method and the number parameters for the indirect method is often comparable, but the number of equations is not. The indirect method has many more equations when we have multiple antennas.

To make things worse for the direct method, we can note that the effective length of the training sequence for the direct method is $N_{tseq} - 2 * n\gamma$ while it is $N_{tseq} - nb$ for the direct method. Since $n\gamma$ is of the order of nb, the effective length of the training sequence for the direct method is typically *shorter* than for the indirect method.

4.6 Summary

The log-likelihood metric, the noise whitening and the matched filter approach, are all equivalent in terms of performance, at least if we neglect possible differences at the edges of the symbol sequences. The matched filter approach is however superior in terms of computational complexity when more than one antenna is used. The metric computation in the Viterbi part of the algorithm is reduced by a factor equal to the number of antennas, when compared to the log-likelihood metric approach.

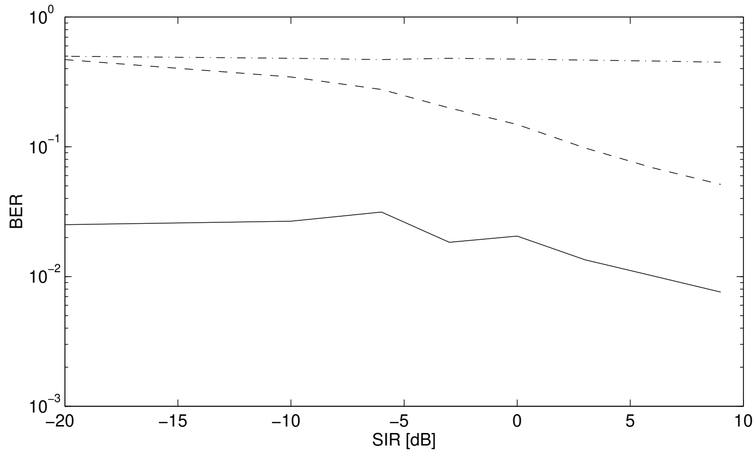


Figure 4.3: BER as a function of the SIR. SNR=2dB. Training sequence length=26. Indirect tuning using only spatial noise color (solid). Direct tuning with regularization (dashed) and without (dash-dotted).

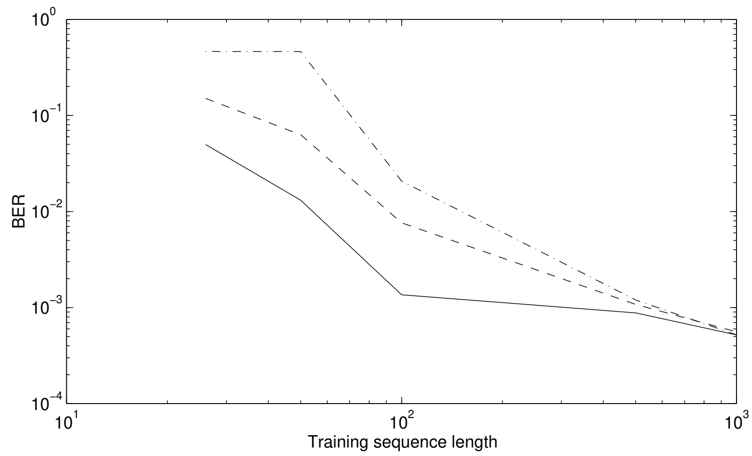


Figure 4.4: BER as a function of the training sequence length. SIR and SNR is 0 dB. Indirect tuning using only spatial noise color (solid). Direct tuning with regularization (dashed) and without (dash-dotted).

When using many antennas and when, as for most TDMA systems, only a short training sequence is available for the tuning, the multi-dimensional matched filter (MMF) can be tuned with an indirect scheme. This can then improve the performance over the direct tuning presented here. A reason for this is that in the indirect method we can choose to use only the spatial spectrum of the noise plus interference. This makes the tuning more well conditioned.

However, if we are using only a small number of antenna elements, say $M = 2$, then the direct method performing spatio-temporal interference suppression can have an advantage over indirect tuning utilizing only the spatial spectrum of the interference.

A way to achieving spatio-temporal interference suppression for an indirectly tuned MLSE is to estimate a *low order AR model* for the noise plus interference [7]. Although the AR model is somewhat inappropriate, since the true model for the noise plus interference resembles a moving average process, the AR model has advantages. From the discussion in Section 2.9.2, we can understand *why* an AR model, even a *low order* AR model, for the noise plus interference can be useful when used in a maximum likelihood sequence estimator. The important observation to make is that the the AR noise model denominator does not have to *model* the noise particularly well, it only has to be able to *suppress* the noise plus interference as a part of a noise whitening filter. Also, with an AR model for noise plus interference and an FIR model for the channel of the desired signal, the MMF is a realizable FIR filter. When we use an AR model for the noise, the memory length of the Viterbi algorithm is however increased. The execution complexity of the Viterbi algorithm therefore increases exponentially with the AR noise model order. If we keep the AR model order small the increase in complexity may however be acceptable. With a large number of antennas or a high order of the AR noise model, the number of parameters in the model can however become large compared to the number of available equations. This can make them potentially difficult to estimate accurately, especially if the SNR is not high enough.

In the simulations shown here, the indirect method is not especially sensitive to poor signal to interference ratios, contrary to what could have been expected. The reason for this is that the indirectly tune MSLE, as well as the indirectly tuned DFE, is self-robustifying to some extent. This phenomenon will be discussed further in Section 7.

The indirect MMSE tuning outlined in Section 4.4.3, has an interesting variability in its structure that allows tradeoffs between complexity and performance. The MMF's ability to temporally whiten the noise and perform matched filtering of the signal can be varied, and the memory length of the subsequent Viterbi-algorithm can be controlled. The utility of this feature in, for example a TDMA system with a short training sequence, remains to be assessed.

4.A Appendix

4.A.1 Deriving the MF Metric from the LL Metric

Using $\boldsymbol{y}'(t)$ and $\boldsymbol{b}'(q^{-1})$ (with coefficients $\boldsymbol{b}'_k$) from equations (4.8) and (4.9), the matched filter version can be derived from the log-likelihood metric in (4.7). We have

$$\mu_{LL}(N) = \sum_{t=1}^{N} \left[\boldsymbol{R}_{\boldsymbol{nn}}^{-1/2}(q^{-1})\left\{\boldsymbol{y}(t) - \boldsymbol{b}(q^{-1})d(t)\right\}\right]^H$$
$$\times \left[\boldsymbol{R}_{\boldsymbol{nn}}^{-1/2}(q^{-1})\left\{\boldsymbol{y}(t) - \boldsymbol{b}(q^{-1})d(t)\right\}\right]$$
$$= \sum_{t=1}^{N} \left[\boldsymbol{R}_{\boldsymbol{nn}}^{-1/2}(q^{-1})\boldsymbol{y}(t)\right]^H \left[\boldsymbol{R}_{\boldsymbol{nn}}^{-1/2}(q^{-1})\boldsymbol{y}(t)\right]$$
$$-2\mathrm{Re}\left\{\sum_{t=1}^{N} \left[\boldsymbol{R}_{\boldsymbol{nn}}^{-1/2}(q^{-1})\boldsymbol{y}(t)\right]^H \boldsymbol{R}_{\boldsymbol{nn}}^{-1/2}(q^{-1})\boldsymbol{b}(q^{-1})d(t)\right\}$$
$$+\sum_{t=1}^{N} \left[\boldsymbol{R}_{\boldsymbol{nn}}^{-1/2}(q^{-1})\boldsymbol{b}(q^{-1})d(t)\right]^H \left[\boldsymbol{R}_{\boldsymbol{nn}}^{-1/2}(q^{-1})\boldsymbol{b}(q^{-1})d(t)\right]$$
$$= f(\boldsymbol{y}(\cdot)) - 2\mathrm{Re}\left\{\sum_{t=1}^{N}\sum_{k=0}^{nb} \boldsymbol{y}'^H(t)\boldsymbol{b}'_k d(t-k)\right\}$$
$$+\sum_{t=1}^{N}\sum_{k}\sum_{l}[\boldsymbol{b}'_k d(t-k)]^H[\boldsymbol{b}'_l d(t-l)]$$

where

$$f(\boldsymbol{y}(\cdot)) \triangleq \sum_{t=1}^{N} \boldsymbol{y}'^H(t)\boldsymbol{y}'(t). \tag{4.33}$$

By replacing t with the new summation variable $n = t - k$ we obtain

$$\mu_{LL}(N) = f(\boldsymbol{y}(\cdot)) - 2\mathrm{Re}\left\{\sum_{k=0}^{nb}\sum_{n=1-k}^{N-k}[\boldsymbol{b}'^H_k\boldsymbol{y}'(k+n)]^H d(n)\right\}$$
$$+\sum_{k=0}^{nb}\sum_{l=0}^{nb}\sum_{n=1-k}^{N-k}[\boldsymbol{b}'_k d(n)]^H[\boldsymbol{b}'_l d(k+n-l)].$$

By replacing summation over l in the last term with summation over the new index $m = k + n - l$ we can write

$$\mu_{LL}(N) = f(\boldsymbol{y}(\cdot)) - 2\mathrm{Re}\left\{\sum_{k=0}^{nb}\sum_{n=1-k}^{N-k}[\boldsymbol{b}_k'^H\boldsymbol{y}'(k+n)]^H d(n)\right\}$$
$$+\sum_{k=0}^{nb}\sum_{n=1-k}^{N-k}\sum_{m=k+n}^{k+n-nb}[\boldsymbol{b}_k' d(n)]^H[\boldsymbol{b}_{k+n-m}' d(m)].$$

We can now assume that $d(t) = 0$ for $t \notin [1, N]$. This is not necessary but it will simplify the derivation and later we will everything happening close to the borders of the interval. This assumption combined with the fact that $\boldsymbol{b}_i = 0$ for $i \notin [0, nb]$ gives

$$\mu_{LL}(N) = f(\boldsymbol{y}(\cdot)) - 2\mathrm{Re}\left\{\sum_{k=0}^{nb}\sum_{n=1}^{N-k}[\boldsymbol{b}_k'^H\boldsymbol{y}'(k+n)]^H d(n)\right\}$$
$$+\sum_{k=-\infty}^{\infty}\sum_{n=-\infty}^{N-k}\sum_{m=k+n}^{k+n-nb}[\boldsymbol{b}_k' d(n)]^H[\boldsymbol{b}_{k+n-m}' d(m)].$$

By representing terms with index n close to N collectively with the term e_{border}, a term we later will neglect, we obtain

$$\mu_{LL}(N) = f(\boldsymbol{y}(\cdot)) - 2\mathrm{Re}\left\{\sum_{k=0}^{nb}\sum_{n=1}^{N}[\boldsymbol{b}_k'^H\boldsymbol{y}'(k+n)]^H d(n)\right\}$$
$$+\sum_{k=-\infty}^{\infty}\sum_{n=-\infty}^{N}\sum_{m=k+n}^{k+n-nb}[\boldsymbol{b}_k' d(n)]^H[\boldsymbol{b}_{k+n-m}' d(m)] + e_{\text{border}}.$$

Finally replacing the summation over k in the last term with a summation over $p = k + n$ gives

$$\mu_{LL}(N) = f(\boldsymbol{y}(\cdot)) - 2\mathrm{Re}\left\{\sum_{k=0}^{nb}\sum_{n=1}^{N}[\boldsymbol{b}_k'^H\boldsymbol{y}'(k+n)]^H d(n)\right\}$$
$$+\sum_{p=-\infty}^{\infty}\sum_{n=-\infty}^{N}\sum_{m=p}^{p-nb}[\boldsymbol{b}_{p-n}' d(n)]^H[\boldsymbol{b}_{p-m}' d(m)] + e_{\text{border}}$$
$$= f(\boldsymbol{y}(\cdot)) - 2\mathrm{Re}\left\{\sum_{k=0}^{nb}\sum_{n=1}^{N}[\boldsymbol{b}_k'^H\boldsymbol{y}'(k+n)]^H d(n)\right\}$$
$$+\sum_{p=-\infty}^{\infty}\sum_{n=-\infty}^{N}\sum_{m=-\infty}^{\infty}[\boldsymbol{b}_{p-n}' d(n)]^H[\boldsymbol{b}_{p-m}' d(m)] + e_{\text{border}}.$$

In the last equality we have used the fact that $\boldsymbol{b}_i = 0$ when $i < 0p$ and $i > nb$. By again exploiting the assumption that $d(t) = 0$ for $t \notin [1, N]$ we can change the summation limits in the second term

$$
\begin{aligned}
\mu_{LL}(N) &= f(\boldsymbol{y}(\cdot)) - 2\mathrm{Re}\left\{\sum_{n=1}^{N}\sum_{k=0}^{nb}[\boldsymbol{b}_k'^{H}\boldsymbol{y}'(n+k)]^H d(n)\right\} \\
&+\sum_{n=1}^{N}\sum_{m=1}^{N} d^H(n)\left[\sum_{p=-\infty}^{\infty}\boldsymbol{b}_{p-n}'^{H}\boldsymbol{b}_{p-m}'\right]d(m) + e_{\text{border}} \\
&= -2\mathrm{Re}\left\{\sum_{n=1}^{N} z^H(n)d(n)\right\} + \sum_{n=1}^{N}\sum_{m=1}^{N} d^H(n)\gamma_{n-m}d(m) \\
&+ f(\boldsymbol{y}(\cdot)) + e_{\text{border}}
\end{aligned} \tag{4.34}
$$

where the scalar signal $z(t)$ is defined by

$$
\begin{aligned}
z(t) &\triangleq \boldsymbol{b}'^{H}(q)\boldsymbol{y}'(t) = \boldsymbol{b}^H(q)\boldsymbol{R}_{\boldsymbol{nn}}^{-H/2}(q)\boldsymbol{R}_{\boldsymbol{nn}}^{-1/2}(q^{-1})\boldsymbol{y}(t) \\
&= \boldsymbol{b}^H(q)\boldsymbol{R}_{\boldsymbol{nn}}^{-1}(q, q^{-1})\boldsymbol{y}(t)
\end{aligned}
$$

where we have used (4.4) in the last equality.

By introducing the *multi-dimensional matched filter*

$$
\boldsymbol{w}(q, q^{-1}) = \boldsymbol{b}^H(q)\boldsymbol{R}_{\boldsymbol{nn}}^{-1}(q, q^{-1}) \tag{4.35}
$$

we can write $z(t)$ as

$$
z(t) = \boldsymbol{w}(q, q^{-1})\boldsymbol{y}(t). \tag{4.36}
$$

The coefficients γ_k ($k = n - m$) in (4.34) are the coefficients of the double sided *metric polynomial*

$$
\gamma(q, q^{-1}) = \sum_{n-m=-\infty}^{\infty}\sum_{p=-\infty}^{\infty}\boldsymbol{b}_{p-n}'^{H}\boldsymbol{b}_{p-m}' q^{n-m} = \boldsymbol{b}^H(q)\boldsymbol{R}_{\boldsymbol{nn}}^{-1}(q, q^{-1})\boldsymbol{b}(q^{-1}) \tag{4.37}
$$

with the coefficients numbered as

$$
\gamma(q, q^{-1}) = \gamma_{-n\gamma}q^{n\gamma} + \ldots + \gamma_0 + \ldots + \gamma_{n\gamma}q^{-n\gamma}. \tag{4.38}
$$

The term $f(\boldsymbol{y}(\cdot))$ does not depend on the candidate symbols, $d(\cdot)$, and can therefore be neglected in the maximization of μ_{LL}. The term e_{border} is a correction term that only depends on the values of $\boldsymbol{y}(t)$ and $d(t)$ for t close to 1 or N. We will neglect possible border effect since they will only affect the estimated symbols close to the borders. Thus, neglecting the "border effects" and changing the sign such that the *matched filter metric* is to be *maximized* gives

$$\mu_{MF}(N) = 2\mathrm{Re}\left\{\sum_{n=1}^{N} z^H(n)d(n)\right\} - \sum_{n=1}^{N}\sum_{m=1}^{N} d^H(n)\gamma_{n-m}d(m). \tag{4.39}$$

Chapter 5

Reduced Complexity Space-Time Equalization

5.1 Introduction

The channel a wireless radio signal passes through can often be described by an FIR model as in (2.15)

$$s(t) = b(q^{-1})d(t) \tag{5.1}$$

where $d(t)$ are the transmitted discrete symbols, $b(q^{-1})$ is the FIR channel and $s(t)$ is the received desired signal (noise free). When using M receive antennas, this signal can be described by an $M \times 1$ column vector $\boldsymbol{s}(t)$ modeled by a vector FIR channel, similar to (2.18)

$$\boldsymbol{s}(t) = \boldsymbol{b}(q^{-1})d(t) \tag{5.2}$$

or in matrix form

$$\boldsymbol{s}(t) = \boldsymbol{B}\boldsymbol{d}(t) \tag{5.3}$$

where $\boldsymbol{B}$ an $M \times (nb+1)$ *channel matrix*, $nb+1$ being the length of the channel, and $\boldsymbol{d}(t) = [d(t)\ d(t-1)\ \ldots\ d(t-nb)]^T$ is a column vector with delayed transmitted symbols. The vector $\boldsymbol{s}(t)$ is an $M \times 1$ column vector containing the desired signal.

The channel matrix, $\boldsymbol{B}$, has $M \times (nb+1)$ coefficients. These coefficients represent $M \times (nb+1)$ degrees of freedom. In realistic channel matrices for wireless channels it is however likely that all these degrees of freedom will not be exploited. The spatial and temporal spreading may have a structure such that there are correlations between the coefficients in the channel matrix. This structure can be exploited when estimating channels, for example as in Section 2.6 in Chapter 2. This structure however also affect the structure of the equalizers. It may not always be necessary to use full fledged spatio-temporal equalizers capable of exploiting all available space-time degrees of freedom as in Chapters 3 and 4. In this chapter we will therefore discuss some equalizers with reduced complexity that do not exploit all available degrees of freedom.

An example of a channel with a simple space-time structure is when the transmitter sends a partial response signal through a propagation channel with negligible delay spread, but with arbitrary angular spread. Even though the propagation channel has no delay spread, the received signal will suffer from intersymbol interference due to the partial response modulation. The temporal impulse response to each antenna will then be the same except for scalings by complex constants. The spatial and temporal spreading of the total channel can thus be decoupled.

The simplest way of performing space-time processing is of course to separate the spatial and the temporal processing. We can have a spatial beamformer followed by a temporal equalizer and tune them independently. One problem to be solved is then how to tune the spatial beamformer in a good way. One may tune the beamformer with the sample matrix inversion method where the beamformer is tuned to minimize the MSE with respect to a reference signal. The selection of a good reference signal is however not trivial since the final performance we are interested in is the BER of the whole equalizer.

Even if there is only a direct wave impinging on the receiver antenna, the proper reference signal to use would depend on the time of travel for the signal from the transmitter to the receiver. Obviously we would have to synchronize properly on a symbol level. However, the timing also needs to be adjusted to a fraction of a symbol if we are to be able to select the appropriate reference signal to use. This is since the sampled version of the transmitted continuous waveform will differ depending on when in time the sampling is performed.

In Section 5.2.2 a method using a "variable" reference signal is presented. Here the reference signal is not fixed but is also tuned along with the beamformer. This is here performed by adding variable components to the reference signal. As we will see in the simulations in Section 5.2.2, this results in a method that is insensitive to synchronization errors within a symbol interval.

Although the performance of this error is evaluated in terms of its sensitivity to a synchronization error in a scenario with a single direct wave impinging on the antenna array, it will also be able to adjust to some intersymbol interference. This can be a valuable feature of the proposed algorithm, especially if the algorithm is extended such that it can handle larger delay spreads.

In Section 5.3 a more general approach to reduced complexity space-time equalization is presented. Here we note the fact that modeling of many space-time channels do not really require all available degrees of freedom in (5.3). Assume for example that we have a channel consisting of some temporal pulse shaping in the transmitter combined with a propagation channel without any significant delay spread. The channel matrix $\boldsymbol{B}$ in (5.3), describing the channel will then only be rank one. The rank of the channel matrix will in fact be approximately equal to the number of groups of signals, with significant different time delays, arriving. Of course only groups of signals with significant energy content are counted. For many channels, it is possible that the rank of the channel matrix therefore will be approximately one or maybe two.

If the channel matrix has reduced rank, then this can be exploited in the design of the equalizers. In Section 5.3 the low rank property of a channel matrix is exploited in order to find simpler space-time structures of a space-time maximum likelihood sequence estimator and a space-time decision feedback equalizer. This results in a maximum likelihood sequence estimator and a decision feedback equalizer which have simpler space-time structures and are less complex. It is also possible to utilize the reduced rank space-time structure of the channel and/or the equalizer in order to facilitate the tuning of the equalizer. This is described in Section 5.4. A solution very similar to the one we arrive to in Section 5.4 was however earlier proposed in [53].

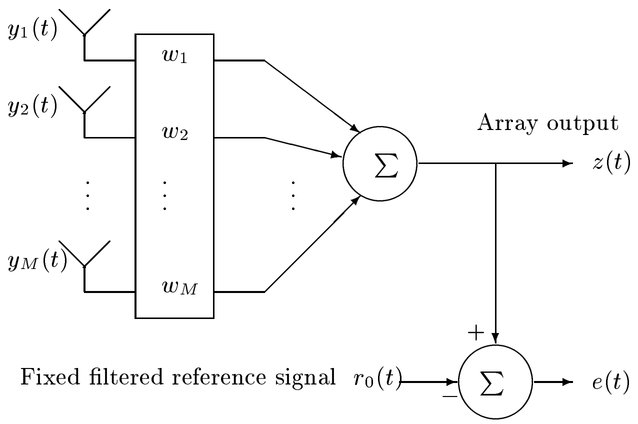


Figure 5.1: Configuration for the Sample Matrix Inversion Method with fixed reference signal filter.

5.2 Spatial Beamforming

The simplest form of space-time processing for equalization is to use a single spatial beamformer followed by a scalar temporal equalizer. The temporal equalizer can be tuned as a scalar version of the equalizers presented in Chapters 3 and 4. We will here consider the tuning of the spatial beamformer. In particular we will present and discuss some aspects of tuning based on the sample matrix inversion method using a training sequence.

5.2.1 SMI Beamforming

If a reference signal describing the desired output from the adaptive antenna is available, then the weights for the adaptive antenna can be tuned with the sample matrix inversion method (SMI). See for example [19]. The reference signal can, for example, be a known part of the transmitted signal.

If we consider a signal with a modulation as in the GSM system, the reference signal can, for example, be generated by filtering the training sequence of the burst through the filter described by the modulation impulse response for no timing offset, tabulated as $t_{off} = 0.0$, $B_{0.0}(q^{-1})$, in Table 2.2. The weights in the beamformer are then adjusted to make the received signal as close as possible to the reference signal, in an MSE sense, as in the standard SMI-method. See [19] and Figure 5.1. Generation of a reference signal with a fixed filter is for instance used in [76].

We will here apply the sample matrix inversion method to a scenario where the transmitted signal is assumed to be modulated with GMSK modulation with a bandwidth-time product of 0.3, as in GSM. We will approximate this modulation with the linear filters presented in Appendix 2.A.1.

Define the vector $\boldsymbol{w}$, consisting of the M beamformer weights

$$\boldsymbol{w} \stackrel{\triangle}{=} [w_1 \, w_2 \ldots w_M] \tag{5.4}$$

and the vector, $\boldsymbol{y}(t)$, consisting of the received signals from the antenna elements, after sampling,

$$\boldsymbol{y}(t) = [y_1(t) \, y_2(t) \ldots y_M(t)]^T. \tag{5.5}$$

The beamformer weights, $\boldsymbol{w}$, are adjusted to minimize[1] the criterion

$$J = \sum_{t=5}^{N} |e(t)|^2 = \sum_{t=5}^{N} |z(t) - r_0(t)|^2 \tag{5.6}$$

where N is the number of symbols in the training sequence, $z(t)$ is the beamformer output expressed as

$$z(t) = \boldsymbol{w} x(t) \tag{5.7}$$

and $r_0(t)$ is the reference signal computed as the known training symbols of $d(t)$ filtered through the pulse shaping filter $p_{0.0}(q^{-1})$

$$\begin{aligned} r_0(t) &= p_{0.0}(q^{-1})d(t) = p_0 d(t) + p_1 d(t-1) + \\ &\quad p_2 d(t-2) + p_3 d(t-3) + p_4 d(t-4). \end{aligned} \tag{5.8}$$

The filter coefficients p_i, i=0,1,..,4, in the generation of the reference signal correspond to the coefficients for $t_{off} = 0.0$ in Table 2.1 .

[1]We sum from $t = 5$ to $t = N$ since $r_0(t)$, the output of the five tap modulation filter (5.8), is not well defined for other values of t.

The minimum of (5.6) is attained by the parameter vector

$$\hat{\boldsymbol{w}}^H = \hat{\boldsymbol{R}}_{\boldsymbol{yy}}^{-1} \hat{\boldsymbol{R}}_{\boldsymbol{y}r} \tag{5.9}$$

where the matrices, $\hat{\boldsymbol{R}}_{xx}$ and $\hat{\boldsymbol{R}}_{\boldsymbol{y}r}$, constitute sample covariance matrices computed from the training sequence data as

$$\hat{\boldsymbol{R}}_{\boldsymbol{yy}} = \frac{1}{N-4} \sum_{t=5}^{N} \boldsymbol{y}(t)\boldsymbol{y}^H(t) \tag{5.10}$$

$$\hat{\boldsymbol{R}}_{\boldsymbol{y}r} = \frac{1}{N-4} \sum_{t=5}^{N} \boldsymbol{y}(t)r^H(t). \tag{5.11}$$

A drawback with this approach is that we have assumed perfect knowledge of the “time-of-arrival” of the desired signal and a good knowledge of the proper training sequence to use. This may not always be the case due to synchronization problems or multipath propagation causing some delay spread. In the following section we therefore present a method which reduces the need for an accurate knowledge of the appropriate training sequence.

5.2.2 SMI Beamforming with Variable Reference Signal

Let us assume that the sampling instant is not synchronized within the symbol interval, implying that the location of the sampling instant within the symbol interval is unknown. In this case the modulation and sampling can be approximated by a set of discrete time channels, parametrized by the location of the sampling instant within the symbol interval.

If one of these channels is selected in order to create the reference signal and the true channel is a different one, then there will be a discrepancy between the reference signal selected for the antenna array to receive optimally, and the actual samples received. When working with short training sequences, this can cause a degradation in the signal to interference and noise ratio (SINR) after the beamformer. In order to avoid such degradation a modified weight adaptation scheme is proposed here. The general idea is to introduce some degrees of freedom in the reference signal.

Although we don't investigate it here, this will also allow for some delay spread in the propagation channel. When delay spread is introduced the received signal waveform will almost always differ from the transmitted waveform as it is made up by a sum of versions of the transmitted waveform with different time delays. In this case we obviously see the need for a variable reference signal allowing for the variability in the received waveform.

Introducing a Variable Reference Signal

In this proposed algorithm, the fixed part of the reference signal filter is the same as for the SMI algorithm in the previous section. Two basis functions with adjustable gains are however added to the reference signal. The aim is to improve the ability to lock on to a received signal sampled at a relative sampling offset different from zero, for instance at $t_{off} = \pm 0.5$. The variable components added, $r_1(t)$ and $r_2(t)$, have here been selected as the training sequence filtered through the *difference* between the channels at $t_{off} = 0.0$ and $t_{off} = \pm 0.5$, i.e.

$$r_1(t) = (p_{-0.5}(q^{-1}) - p_{0.0}(q^{-1}))d(t) \tag{5.12}$$

$$r_2(t) = (p_{0.5}(q^{-1}) - p_{0.0}(q^{-1}))d(t). \tag{5.13}$$

For the GSM case, the pulse shaping polynomials $p_{0.0}(q^{-1})$, $p_{0.5}(q^{-1})$ and $p_{-0.5}(q^{-1})$ can be found in Table 2.1. Other alternatives can also be considered. The general idea is to introduce additional degrees of freedom in the reference signal to attain better modeling of the partly unknown reference signal. See Figure 5.2.

In addition to the weights in the adaptive beamformer, the two coefficients, c_1 and c_2, for the variable components, $r_1(t)$ and $r_2(t)$, in the reference signal are also adjusted, in order to minimize the error between the beamformer output and the reference signal.

The modification of the SMI-algorithm is now straightforward. Introduce the modified parameter and regressor vectors, $\boldsymbol{\theta}$ and $\boldsymbol{\varphi}(t)$

$$\boldsymbol{\theta} = [w_1\, w_2 \ldots w_M\, c_1\, c_2] \tag{5.14}$$

$$\boldsymbol{\varphi}(t) = [y_1(t)\, y_2(t) \ldots y_M(t)\, r_1(t)\, r_2(t)]^T. \tag{5.15}$$

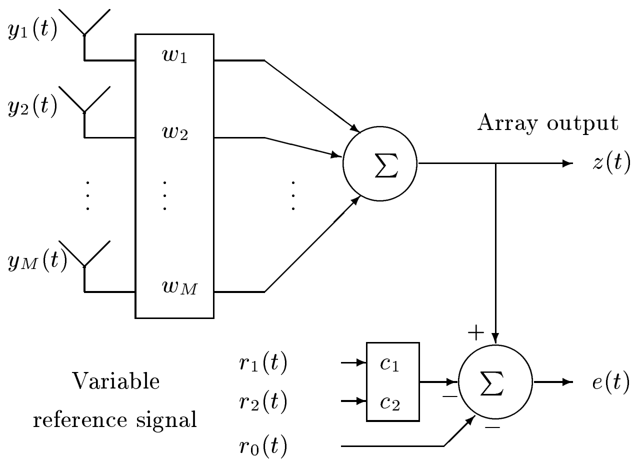


Figure 5.2: Configuration for the Sample Matrix Inversion Method with variable reference signal filter.

The objective is now to adjust the modified parameter vector, $\boldsymbol{\theta}$, to minimize the criterion

$$V(\boldsymbol{\theta}) = \sum_{t=5}^{N} |e(t)|^2 = \sum_{t=5}^{N} |\boldsymbol{\theta}\boldsymbol{\varphi}(t) - r_0(t)|^2 \tag{5.16}$$

where $r_0(t)$ is the reference signal computed as

$$r_0(t) = p_{0.0}(q^{-1})d(t) \tag{5.17}$$

with $p_{0.0}(q^{-1})$ being the pulse shaping filter for zero sampling offset. The minimum of (5.16) is attained by the parameter vector

$$\hat{\boldsymbol{\theta}}^H = \hat{\boldsymbol{R}}_{\varphi\varphi}^{-1}\hat{\boldsymbol{R}}_{\varphi r} \tag{5.18}$$

where the matrices $\hat{\boldsymbol{R}}_{\varphi\varphi}$ and $\hat{\boldsymbol{R}}_{\varphi r}$ are computed as

$$\hat{\boldsymbol{R}}_{\varphi\varphi} = \frac{1}{N-4}\sum_{t=5}^{N} \boldsymbol{\varphi}(t)\boldsymbol{\varphi}^H(t) \tag{5.19}$$

$$\hat{\boldsymbol{R}}_{\varphi r} = \frac{1}{N-4}\sum_{t=5}^{N} \boldsymbol{\varphi}(t)r_0^H(t). \tag{5.20}$$

Performance Evaluation

In order to compare the SMI algorithm with a fixed reference signal and the SMI algorithm with a variable reference signal, the resulting SINR in the signals after the beamformers was computed. The algorithms were also evaluated by applying a maximum likelihood sequence estimator[2] (MLSE) after the beamformer, and computing the resulting bit error rate (BER).

Simulation Settings

A circular antenna array consisting of eight antenna elements ($M = 8$), as shown in Figure 5.3, was used in the simulations.

The desired signal is impinging on the array from the direction $\alpha = 0$ degrees, see Figure 5.4. Co-channel interferers are impinging on the array from the directions $\alpha_{co} = 135$, -30 and -125 degrees, all having a constant channel $b_{co}(q^{-1}) = b_{co}$. The constant b_{co} was selected such that the SIR, averaged over all the antenna elements, became 0 dB. Independent white Gaussian noise giving a SNR of 3 dB, averaged over the antenna elements, was also added.

The relative sampling offset, t_{off}, was varied, resulting in different channels for the desired signal as described in Table 2.1. The beamformers were tuned based on data from a training sequence consisting of 26 binary symbols, as in a GSM system. The channel used in the MLSE was estimated as a three tap FIR-channel, using the same training data. The SINR and the BER was evaluated over 500 and 5000 symbols respectively. This experiment was repeated 100 times for different realizations of both the noise and interference.

[2]The MLSE used a least squares estimate of the channel that was formed by estimating the channel between the transmitted symbols and the received samples, after the beamformer. Instead of estimating the channel after the beamformer one can choose to use the channel estimate $\hat{b}(q^{-1}) = p_{0.0}(q^{-1}) + c_1(p_{-0.5}(q^{-1}) + p_{0.0}(q^{-1})) + c_2(p_{0.5}(q^{-1}) - p_{0.0}(q^{-1}))$ in the MLSE. This could save some complexity in the algorithm.

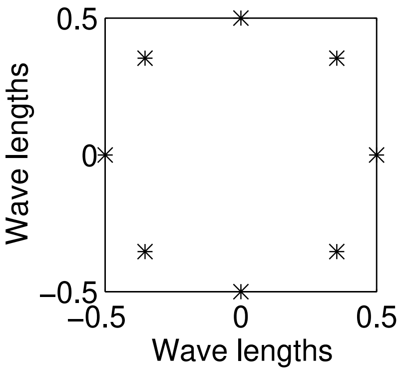


Figure 5.3: Antenna configuration

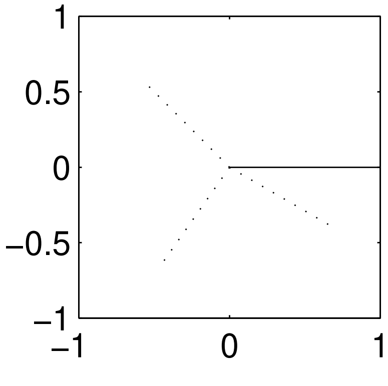


Figure 5.4: Desired signal (solid) and co-channel interferers (dotted).

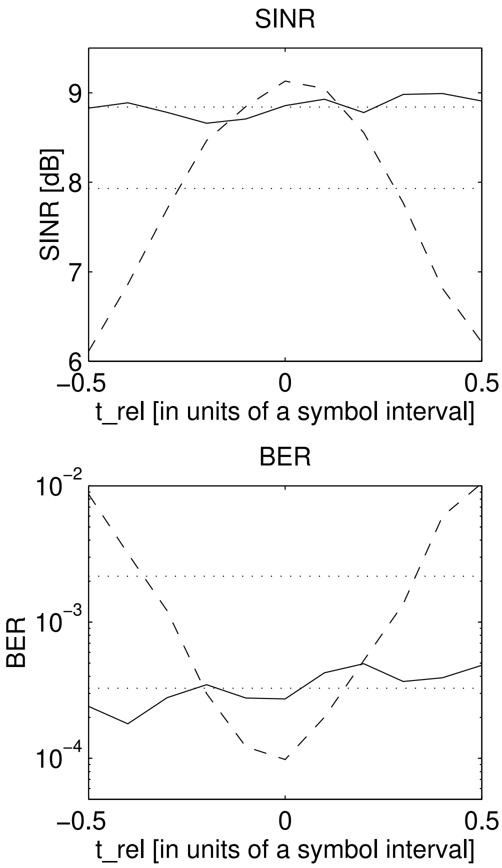


Figure 5.5: SINR after the beamformers and BER for the MLSE as a function of the relative sampling offset, t_{off}. SMI with variable reference signal filter (solid) and SMI with fixed reference signal filter (dashed). The dotted lines show the mean of the SINR's and the BER's, averaged uniformly the symbol interval

Results

The SINR after the beamformers (*cf* Figures 5.1 and 5.2) and the BER for the MLSE, are depicted in Figure 5.5 as a function of the relative sampling offset t_{off}. From the two diagrams it can be seen that the performance of the SMI beamformer with fixed reference signal filter degrades considerably when the relative sampling offset differs from $t_{off} = 0.0$, which it was designed for. The SMI beamformer with variable reference signal filter is however almost insensitive to the sampling instant within the symbol interval. The performance has also been averaged uniformly over one symbol interval. The resulting mean SINR and BER is better for the SMI method with the variable reference signal filter, see the dotted lines in in Figure 5.5. However, if the deviation from $t_{off} = 0.0$ is small, then the SMI method with fixed reference signal filter performs better.

The reason why the beamformer with the fixed reference signal filter has a performance degradation when t_{off} differs from zero, is that the incorrectly modeled channel causes the algorithm to treat part of the desired signal as a disturbance. The algorithm thus decreases the gain slightly in the direction of the desired signal, resulting in a lower SINR.

It should be noted that for longer training sequences the difference between the algorithms is not that pronounced. The differences are amplified by the use of a single short training sequence.

The results presented above also depends the signal to noise ratio. Adding more noise will make the difference between the algorithms smaller. The mismodeling of the reference signal will then be masked by the effects of the noise.

Multipath propagation has not been included in this study. In the presence of multipath propagation, the received signal will differ even more from any reference signal generated from a fixed filter. Introducing a variable reference signal should therefore also be beneficial also in this case. If we want to accommodate a delay spread that is larger that the one symbol interval covered here we will have to add more components to the reference signal. The exact construction of the added basis functions could however be different from the construction presented here.

5.3 Reduced Rank Equalization

We will here study how the structure of the channel for the desired signal affects the structure of the equalizer and how the structure of the channel can be utilized in the design of the equalizer.

As discussed in the beginning of this chapter it is likely that many space-time channels will not be occupying all degrees of freedom in the channel matrix $\boldsymbol{B}$ in (5.3). The channel matrix, $\boldsymbol{B}$, will in such cases not have full rank. Subsequently we call such channels *reduced rank channels.*

The motivating simple example presented in the Section 5.1, with a transmitter sending a partial response signal through a propagation channel with negligible delay spread, but with arbitrary angular spread, is an example of a reduced rank channel. Even though the propagation channel has no delay spread, the received signal will suffer from intersymbol interference due to the partial response modulation. It is easy to see that in this case the channel matrix will have rank one since there is no *coupled* angular and delay spread. All taps will have the same spatial signature.

An equalizer designed from a channel with reduced rank will also have reduced rank in the sense that it will not exploit all spatio-temporal dimensions in the space-time filtering. The reduced rank equalizers have the full rank space-time filtering replaced by a set of beamformers followed by temporal filters. The rank one equalizer will thus have a single beamformer followed by temporal filtering.

An equalizer for a reduced rank channel will have a simpler structure than a general space-time equalizer. The execution of a reduced rank equalizer will be less complex than the full rank counterpart and it will in general be less complex to tune. The most reduction in complexity of the structure, execution and tuning will, of course, be achieved with a rank one equalizer.

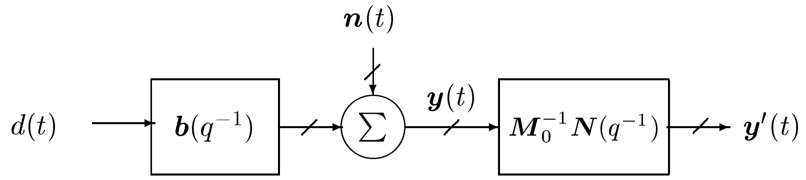


Figure 5.6: Channel with spatially colored noise and noise whitening filter.

5.3.1 Reduced Rank Channel Approximation

Channel Model

We will in this section use a channel model as in (2.112) with an FIR model for the channel of the desired signal and an AR model for the noise plus interference. The received vector of signal samples is thus modeled as

$$\boldsymbol{y}(t) = \boldsymbol{b}(q^{-1})d(t) + \boldsymbol{n}(t) \tag{5.21}$$

where $\boldsymbol{y}(t)$ is the received sequence of vector signal samples, $d(t)$ is the sequence of transmitted symbols and $\boldsymbol{b}(q^{-1})$ is the vector FIR channel for the desired signal defined in (2.113). The noise plus interference $\boldsymbol{n}(t)$ is modeled with an AR model as

$$\boldsymbol{n}(t) = \boldsymbol{N}^{-1}(q^{-1})\boldsymbol{M}_0\boldsymbol{v}(t) \tag{5.22}$$

where $\boldsymbol{N}(q^{-1})$ is a full stably invertible [3] polynomial matrix as in (2.115), $\boldsymbol{M}_0$ a non-singular constant numerator matrix as in (2.114) and $\boldsymbol{v}(t)$ is the spatially and temporarily white innovations sequence. The reason for using this model is, as we have seen in Chapters 3 and 4, that the equalizers designed using this channel model will only contain FIR filters. This will be useful for the complexity reductions treated in this section.

By "noise whitening" the output of the channel $\boldsymbol{y}(t)$ with the noise whitening filter, $\boldsymbol{M}_0^{-1}\boldsymbol{N}(q^{-1})$, as in Figure 5.6, we obtain the "noise whitened" channel

[3] The determinant of $\boldsymbol{N}(q^{-1})$ has all zeros strictly inside the unit circle, and the leading coefficient is non-singular.

model

$$\begin{aligned}
\boldsymbol{y}'(t) &= \boldsymbol{M}_0^{-1}\boldsymbol{N}(q^{-1})\left(\boldsymbol{b}(q^{-1})d(t) + \boldsymbol{n}(t)\right) \\
&= \boldsymbol{b}'(q^{-1})d(t) + \boldsymbol{v}(t) \\
&= \left(\boldsymbol{b}'_0 + \ldots + \boldsymbol{b}'_{nb'}q^{-nb'}\right) d(t) + \boldsymbol{\nu}(t) \qquad (5.23) \\
& \qquad (5.24)
\end{aligned}$$

where

$$\boldsymbol{y}'(t) = \boldsymbol{M}_0^{-1}\boldsymbol{N}(q^{-1})\boldsymbol{y}(t) \tag{5.25}$$

and where the noise whitened channel polynomial

$$\boldsymbol{b}'(q^{-1}) = \boldsymbol{M}_0^{-1}\boldsymbol{N}(q^{-1})\boldsymbol{b}(q^{-1}) \tag{5.26}$$

has the new order $nb' = nb + nn$ where nn is the order of the denominator polynomial matrix $\boldsymbol{N}(q^{-1})$ of the noise model. Also note that the innovations sequence $\boldsymbol{\nu}(t)$ is temporally and spatially white.

The noise whitened channel, $\boldsymbol{b}'(q^{-1})$, will better than $\boldsymbol{b}(q^{-1})$ describe the properties of the channel of importance for the equalization as it has taken the spectrum of the noise plus interference into consideration. We will therefore work with this channel later in this section.

Reduced Rank Model

Let us first express the channel $\boldsymbol{b}(q^{-1})$ in (5.21) with a *channel matrix*

$$\boldsymbol{B} = \left[\begin{array}{cccc} \boldsymbol{b}_0 & \boldsymbol{b}_1 & \ldots & \boldsymbol{b}_{nb} \end{array}\right] \tag{5.27}$$

of dimensions $M \times (nb+1)$. By using a singular value decomposition, we can decompose the channel matrix as

$$\boldsymbol{B} = \boldsymbol{U}\boldsymbol{V}^H \tag{5.28}$$

where the columns of $\boldsymbol{U}$ are the left singular vectors $\boldsymbol{u}_k$ of $\boldsymbol{B}$ and the columns of $\boldsymbol{V}$ are the right singular vectors $\boldsymbol{v}_k$ of $\boldsymbol{B}$ *with singular values included*, i.e.

$$\boldsymbol{U} = \left[\begin{array}{cccc} \boldsymbol{u}_1 & \boldsymbol{u}_2 & \ldots & \boldsymbol{u}_{K_b} \end{array}\right] \tag{5.29}$$

$$\boldsymbol{V} = \begin{bmatrix} \boldsymbol{v}_1 & \boldsymbol{v}_2 & \ldots & \boldsymbol{v}_{K_b} \end{bmatrix} \tag{5.30}$$

where $K_b = \min(M, nb+1)$.

Equivalently, in polynomial notation, the channel can be expressed as the sum of K_b components, each representing a rank one channel:

$$\boldsymbol{b}(q^{-1}) = \sum_{k=1}^{K_b} \boldsymbol{u}_k v_k(q^{-1}). \tag{5.31}$$

The orthogonal spatial signatures are generated by the vectors $\boldsymbol{u}_k$ and the temporal structure of the components of the channel is produced by the polynomials

$$v_k(q^{-1}) = v_{k0}^* + v_{k1}^* q^{-1} + \ldots + v_{knb}^* q^{-nb} \tag{5.32}$$

where the polynomial coefficients, v_{ki}, are the components of the right singular vectors

$$\boldsymbol{v}_k = \begin{bmatrix} v_{k0} & v_{k1} & \ldots & v_{knb} \end{bmatrix}^T. \tag{5.33}$$

This decomposition is depicted in Figure 5.7. The branches are ordered in descending order with respect to the power of the branches.

Note that since $\boldsymbol{U}$ contains left singular vectors,

$$\boldsymbol{U}^H \boldsymbol{U} = \boldsymbol{I} \tag{5.34}$$

i.e. the K_b beamspreaders are orthonormal.

Depending on the singular values, we may be able to approximate the model in (5.31) well with a sum with fewer terms, keeping only the terms with significant singular values. Let us assume that the channel $\boldsymbol{b}(q^{-1})$ can be approximated with the sum

$$\boldsymbol{b}(q^{-1}) \approx \sum_{k=1}^{K_{br}} \boldsymbol{u}_k v_k(q^{-1}) \tag{5.35}$$

where K_{br} is a relatively small number.

In matrix form we can then approximate the channel matrix $\boldsymbol{B}$ as

$$\boldsymbol{B} \approx \boldsymbol{U}_r \boldsymbol{V}_r^H \tag{5.36}$$

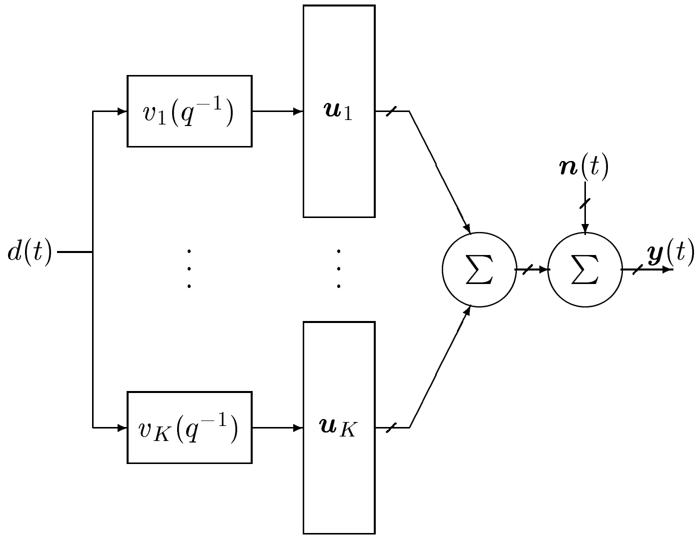


Figure 5.7: Orthogonal decomposition of channel.

where

$$\boldsymbol{U}_r = \left[\begin{array}{cccc} \boldsymbol{u}_1 & \boldsymbol{u}_2 & \ldots & \boldsymbol{u}_{K_{br}} \end{array} \right] \tag{5.37}$$

$$\boldsymbol{V}_r = \left[\begin{array}{cccc} \boldsymbol{v}_1 & \boldsymbol{v}_2 & \ldots & \boldsymbol{v}_{K_{br}} \end{array} \right] \tag{5.38}$$

where K_{br} is the *rank* of the channel. The rank of a wireless channel is determined by the propagation environment. If, for example, the delay spread in the channel is a negligible, then the channel can be well described by one component ($K_{br} = 1$) irrespective of the angular spread in the channel. Likewise, if the angular spread is small, then the channel will be well described by a rank one model. As the propagation environment becomes more complicated with coupled delay and angular spread more principal components are required and the rank of the channel will increase. It is however not unlikely that many practical communication channels can be well approximated with a low rank model. Maybe many channel can be approximated with only a rank one or a rank two model.

The simplest channel model would be the rank one model

$$\boldsymbol{b}_1(q^{-1}) \approx \boldsymbol{u}_1 v_1(q^{-1}) \tag{5.39}$$

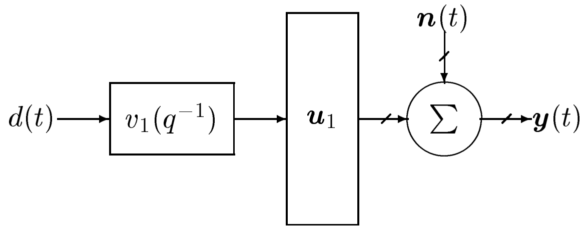


Figure 5.8: Rank one approximation of channel.

as depicted in Figure 5.8.

Let us now instead express the *noise whitened channel* $\boldsymbol{b}'(q^{-1})$ in (5.24) with a *noise whitened channel matrix*

$$\boldsymbol{B}' = \left[\begin{array}{cccc} \boldsymbol{b}'_0 & \boldsymbol{b}'_1 & \ldots & \boldsymbol{b}'_{nb'} \end{array} \right] \tag{5.40}$$

of dimensions $M \times (nb + nn + 1)$. As for the channel matrix $\boldsymbol{B}$ in (5.27), we can by using a singular value decomposition, decompose the noise whitened channel matrix as

$$\boldsymbol{B}' = \boldsymbol{U}'\boldsymbol{V}'^H \tag{5.41}$$

where the columns of $\boldsymbol{U}'$ are the left singular vectors $\boldsymbol{u}'_k$ of $\boldsymbol{B}'$ and the columns of $\boldsymbol{V}'$ are the right singular vectors $\boldsymbol{v}'_k$ of $\boldsymbol{B}'$ *with singular values included*, i.e.

$$\boldsymbol{U}' = \left[\begin{array}{cccc} \boldsymbol{u}'_1 & \boldsymbol{u}'_2 & \ldots & \boldsymbol{u}'_{K_{b'}} \end{array} \right] \tag{5.42}$$

$$\boldsymbol{V}' = \left[\begin{array}{cccc} \boldsymbol{v}'_1 & \boldsymbol{v}'_2 & \ldots & \boldsymbol{v}'_{K_{b'}} \end{array} \right] \tag{5.43}$$

where $K_{b'} = \min(M, nb + nn + 1)$.

Equivalently, in polynomial notation, the noise whitened channel can be expressed as

$$\boldsymbol{b}'(q^{-1}) = \sum_{k=1}^{K_{b'}} \boldsymbol{u}'_k v'_k(q^{-1}) \tag{5.44}$$

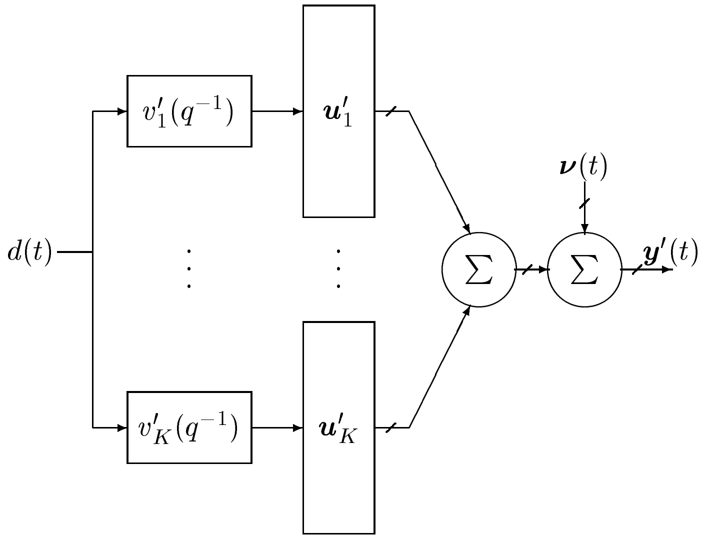


Figure 5.9: Orthogonal decomposition of *whitened* channel.

where the orthogonal spatial signatures are generated by the vectors $\boldsymbol{u}'_k$ and the temporal structure of the components in the channel is produced by the polynomials

$$v'_k(q^{-1}) = {v'}^*_{k0} + {v'}^*_{k1}q^{-1} + \ldots + {v'}^*_{knb'}q^{-nb'} \tag{5.45}$$

where the polynomial coefficients, v'_{ki}, are the components of the right singular vectors

$$\boldsymbol{v}'_k = \begin{bmatrix} v'_{k0} & v'_{k1} & \ldots & v'_{knb'} \end{bmatrix}^T . \tag{5.46}$$

This decomposition is depicted in Figure 5.9. The branches are ordered in descending order with respect to the power of the branches.

Note again that since $\boldsymbol{U}'$ contains left singular vectors,

$$\boldsymbol{U}'^H\boldsymbol{U}' = \boldsymbol{I} \tag{5.47}$$

i.e. the noise whitened beamspreaders are orthonormal.

Let us now consider the noise whitened channel in (5.44). We may wonder if the noise whitened channel can be modeled by a sum like (5.44) but with fewer terms. Let us assume that the channel $\boldsymbol{b}(q^{-1})$ can be approximated with a low rank model as in (5.35). The noise whitened channel can then be approximated as

$$\boldsymbol{b}'(q^{-1}) = \boldsymbol{M}_0^{-1}\boldsymbol{N}(q^{-1})\boldsymbol{b}(q^{-1}) \approx \boldsymbol{M}_0^{-1}\boldsymbol{N}(q^{-1})\sum_{k=1}^{K_{br}}\boldsymbol{u}_k v_k(q^{-1}). \tag{5.48}$$

By writing the denominator matrix $\boldsymbol{N}(q^{-1})$ as

$$\boldsymbol{N}(q^{-1}) = \boldsymbol{N}_0 + \ldots + \boldsymbol{N}_{nn}q^{-nn} \tag{5.49}$$

we can express $\boldsymbol{b}'(q^{-1})$ in (5.48) as

$$\begin{aligned}\boldsymbol{b}'(q^{-1}) &\approx \sum_{k=1}^{K_{br}}\boldsymbol{M}_0^{-1}(\boldsymbol{N}_0 + \ldots + \boldsymbol{N}_{nn}q^{-nn})\boldsymbol{u}_k v_k(q^{-1}) \\ &= \sum_{k=1}^{K_{br}}(\boldsymbol{M}_0^{-1}\boldsymbol{N}_0\boldsymbol{u}_k + \ldots + \boldsymbol{M}_0^{-1}\boldsymbol{N}_{nn}\boldsymbol{u}_k q^{-nn})v_k(q^{-1}).\end{aligned}$$

This new channel contains $K_{br}(1+nn)$ potential spatial signatures, namely $\boldsymbol{M}_0^{-1}\boldsymbol{N}_0\boldsymbol{u}_k$ to $\boldsymbol{M}_0^{-1}\boldsymbol{N}_{nn}\boldsymbol{u}_k$ for $k = 1, \ldots, K_{br}$. Therefore the rank of this channel cannot be higher than $K_{br}(1+nn)$. We thus see that if the channel $\boldsymbol{b}(q^{-1})$ can be reasonably well approximated with a rank K_{br} model, then the noise whitened channel $\boldsymbol{b}'(q^{-1})$ can be reasonably well approximated with a model with rank $K_{b'r} \leq K_{br}(1+nn)$

$$\boldsymbol{b}'(q^{-1}) = \sum_{k=1}^{K_{b'r}}\boldsymbol{u}'_k v'_k(q^{-1}). \tag{5.50}$$

Note that if the noise whitening filter is purely spatial, i.e. $\boldsymbol{N}(q^{-1}) = \boldsymbol{I}$, then noise whitened channel $\boldsymbol{b}(q^{-1})$ has the same rank as the original channel $\boldsymbol{b}(q^{-1})$. Note also that even though we may need K_{br} or $K_{b'r}$ terms in the models in (5.35) and (5.50) we can still make model with fewer terms. These model will be less accurate but will be simpler. What model we use will depend on what is more important, accuracy or simplicity.

The noise whitened channel matrix can thus here be approximated with

$$\boldsymbol{B}' \approx \boldsymbol{U}'_r\boldsymbol{V}'^H_r \tag{5.51}$$

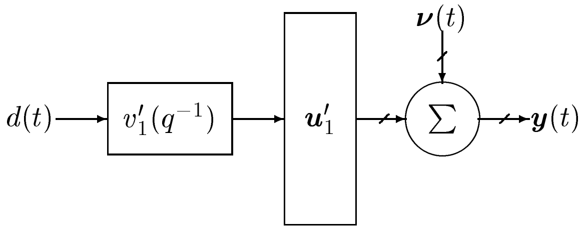


Figure 5.10: Rank one approximation of noise whitened channel.

where

$$\boldsymbol{U}_r' = \begin{bmatrix} \boldsymbol{u}_1' & \boldsymbol{u}_2' & \dots & \boldsymbol{u}'_{K_{b'r}} \end{bmatrix} \tag{5.52}$$

$$\boldsymbol{V}_r' = \begin{bmatrix} \boldsymbol{v}_1' & \boldsymbol{v}_2' & \dots & \boldsymbol{v}'_{K_{b'r}} \end{bmatrix} \tag{5.53}$$

and where $K_{b'r}$ is the *rank* of the noise whitened channel model.

As for the original channel, the simplest model for the nose whitened channel would be the rank one model

$$\boldsymbol{b}_1'(q^{-1}) \approx \boldsymbol{u}_1' v_1'(q^{-1}) \tag{5.54}$$

as depicted in Figure 5.10.

Remark: Instead of using a singular value decomposition, the rank reduced model can be computed with the power method[4] [14]. We show here the method preferred for the case when the number of antennas, M, is larger than the number of taps in the noise whitened channel, $nb+nn+1$ (or $nb+1$ if applied to the original channel). First form the matrix

$$\boldsymbol{A} = \boldsymbol{B'}^H \boldsymbol{B}'. \tag{5.55}$$

Then find the eigenvector corresponding to the largest eigenvalue of this matrix using the power method. In our simulations we found that only a few iterations were required. If the resulting eigenvector is normalized to have its norm equal to the square root of the eigenvalue we will obtain the

[4]The power method is an iterative method for finding the dominating eigenvalue and eigenvector of a matrix.

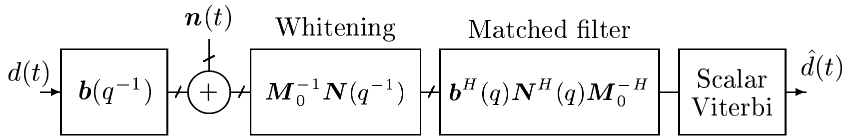


Figure 5.11: Spatio-temporal MLSE with whitening filter and spatio-temporal matched filter.

vector $\boldsymbol{v}'_1$, i.e. the right singular vector of $\boldsymbol{B}'$ corresponding to the largest singular value, with the singular value included. The corresponding left singular vector can be computed as

$$\boldsymbol{u}'_1 = \boldsymbol{B}'\boldsymbol{v}'_1. \tag{5.56}$$

To obtain additional left and right singular vectors the process can, if necessary, be repeated after the component $\boldsymbol{u}'_1\boldsymbol{v}'^H_1$ has been subtracted from $\boldsymbol{B}'^H\boldsymbol{B}'$.

5.3.2 Reduced Rank MLSE

Spatio-Temporal MLSE

As shown in Figure 5.11, the multi-channel MLSE can be separated into a noise whitening filter and a matched filter, the latter matched to the noise whitened channel, followed by a scalar Viterbi. For details see Chapter 4. The matched filter, operating on the noise whitened signal, can then as in (4.14) be expressed as

$$z(t) = \boldsymbol{b}^H(q)\boldsymbol{N}^H(q)\boldsymbol{M}_0^{-H}\boldsymbol{y}'(t) \tag{5.57}$$

where

$$\boldsymbol{b}^H(q)\boldsymbol{N}^H(q)\boldsymbol{M}_0^{-H} \tag{5.58}$$

is the spatio-temporal matched filter.

The signal, $z(t)$, is then processed in a *scalar* Viterbi algorithm in order to find the symbol sequence that maximizes the in (4.19) recursively defined

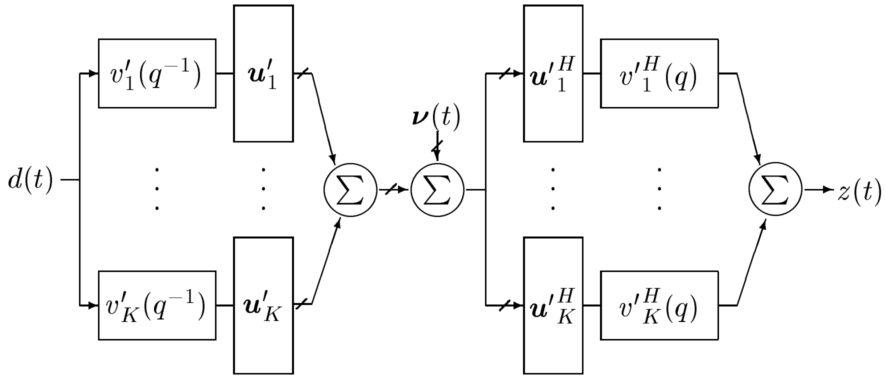


Figure 5.12: Orthogonal decomposition of the noise whitened channel and the matched filter.

matched filter metric

$$\begin{aligned}\mu_{MF}(t) &= \mu_{MF}(t-1) + \mathrm{Re}\Big\{ d^H(t)(2z(t) - \gamma_0 d(t) \\ &\quad -2\sum_{m=1}^{n\gamma} \gamma_m d(t-m))\Big\}\end{aligned} \tag{5.59}$$

where coefficients γ_k are the coefficients of the double sided *metric* polynomial $\boldsymbol{\gamma}(q, q^{-1})$ defined in (4.16) and (4.17)

$$\gamma(q, q^{-1}) = \boldsymbol{b}^H(q)\boldsymbol{N}^H(q)\boldsymbol{M}_0^{-H}\boldsymbol{M}_0^{-1}\boldsymbol{N}(q^{-1})\boldsymbol{b}(q^{-1}). \tag{5.60}$$

Rank Reduction

By using the noise whitened channel, orthogonally decomposed into its principal components we also obtain an orthogonal decomposition of the spatio-temporal matched filter $\boldsymbol{b}^H(q)\boldsymbol{N}^H(q)\boldsymbol{M}_0^{-H}$ in (5.58). The decomposition of the noise whitened channel and the corresponding spatio-temporal matched filter can be seen in Figure 5.12.

The full rank spatio-temporal matched filter $\boldsymbol{b}^H(q)\boldsymbol{N}^H(q)\boldsymbol{M}_0^{-H}$ will here generate a signal, $z(t)$, with maximal *peak to noise ratio* (PNR) defined as

$$\mathrm{PNR} \stackrel{\triangle}{=} \gamma_0^2/\sigma^2 \tag{5.61}$$

where γ_0 is the real middle coefficient in the overall channel, $\boldsymbol{\gamma}(q, q^{-1})$, after the noise whitening filter the matched filter, and σ^2 is the noise variance after the same filters.

This is the optimal filter for the multi-channel MLSE for the considered channel. If we want to design a spatio-temporal matched filter of a prespecified reduced rank, $K_r < K_{b'}$, we can choose it to, given the rank constraint, maximize the PNR prior to the Viterbi algorithm. Since the noise before the filter is white due to the whitening filter, *the filter of rank K_r that gives the maximal peak to noise ratio is simply the filter consisting of the first K_r branches* in the decomposition of the spatio-temporal matched filter. This filter will then only process the K_r components of the received signal $\boldsymbol{y}(t)$ which contribute the largest signal power. Other components will be filtered out since their channel beamspreading filters, $\{\boldsymbol{u}'^H_i\}_{K_r+1}^{K}$, are orthogonal to the receiver beamformers used, $\{\boldsymbol{u}'^H_i\}_1^{K_r}$, due to the orthogonal decomposition.

We also have to include the noise whitening filter in the reduced rank multidimensional matched filter. The PNR optimal rank K_r MLSE will here thus consist of the multidimensional matched filtering

$$z(t) = \sum_{k=1}^{K_r} v'^H_k(q)\boldsymbol{u}'^H_k \boldsymbol{M}_0^{-1}\boldsymbol{N}(q^{-1})\boldsymbol{y}(t). \tag{5.62}$$

followed by a Viterbi algorithm using the metric defined by (5.59) and the metric polynomial

$$\boldsymbol{\gamma}_{K_r}(q, q^{-1}) \triangleq \sum_{k=1}^{K_r} v'^H_k(q) v'_k(q^{-1}). \tag{5.63}$$

The metric polynomial, $\gamma(q, q^{-1})_{K_r}$, represents the channel from the transmitted symbols to the received signal samples. Only the K_r components used in (5.63) will enter into $\gamma(q, q^{-1})_{K_r}$. The other components are canceled by the beamformers. We call the resulting MLSE, depicted in Figure 5.13, a *reduced rank MLSE*. Note that if the noise whitening filter is purely spatial, i.e. $\boldsymbol{N}(q^{-1}) = \boldsymbol{I}$, then the spatial noise whitening, $\boldsymbol{M}_0^{-1}$ can be incorporated in the spatial beamformers, i.e. we obtain the new *noise whitening beamformers* $\tilde{\boldsymbol{u}}'^H_k = \boldsymbol{u}'^H_k \boldsymbol{M}_0^{-1}$.

The simplest reduced rank MLSE will be the rank one MLSE consisting of a noise whitening filter and beamformer followed by a scalar MLSE as in

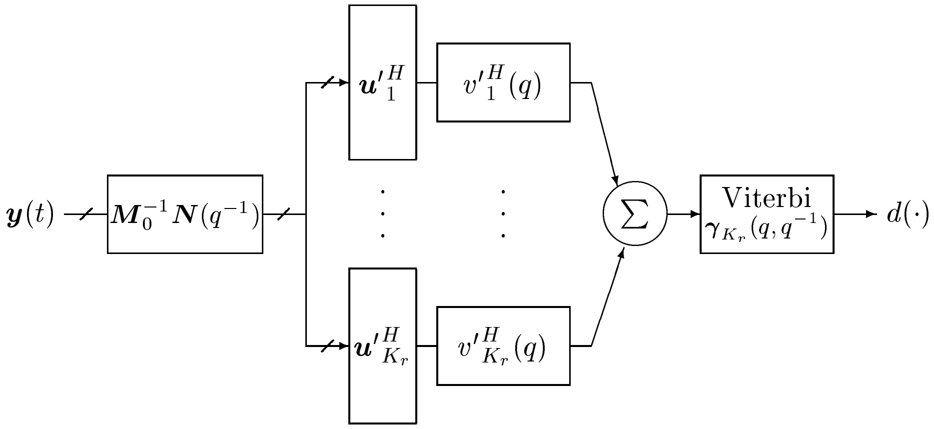


Figure 5.13: Reduced rank MLSE.

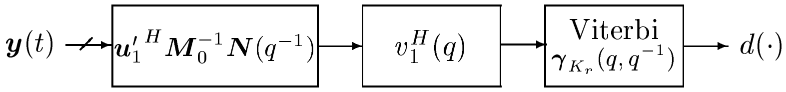


Figure 5.14: Rank one MLSE.

shown in Figure 5.14. Note again that if the noise whitening filter is purely spatial, i.e. $\boldsymbol{N}(q^{-1}) = \boldsymbol{I}$, then the reduced rank MLSE of rank one consists simply of a *noise whitening beamformer*, $\tilde{\boldsymbol{u}}_1' = \boldsymbol{u'}_1^H \boldsymbol{M}_0^{-1}$, followed by the temporal filter $v'^H_1(q)$.

If the noise whitened channel has the same rank as the reduced rank MLSE we want to design, the reduced rank MLSE will of course be optimal since the exact matched filter then is realized.

It is important to note that the reduced rank MLSEs computed by (5.62) and (5.63) will in general not be optimal in a BER sense. One can construct examples showing this. Since the reduced rank MLSE's maximizes the same PNR criterion as for the full rank MLSE, they may however be a good candidates. It is a more difficult problem to find the truly optimal reduced rank MLSE.

It is also important to keep in mind that a requirement for the analysis here is that the true channel really can be modeled with the channel model in (5.21), and the noise model in (5.22). When this is not the case, the analysis here will only hold approximatively.

Remark 1: It may seem strange that the reduced rank MLSE is not optimal even though it optimizes the same PNR criterion as the full rank MLSE. However, one has to keep in mind that the filter for the reduced rank MLSE optimizes the PNR criterion under a constraint, namely the constraint of reduced rank. It is therefore not necessary that this filter should also provide optimal BER when used in a Viterbi algorithm.

Remark 2: When we reduce the rank of the channel and use the reduced rank channel to design the MLSE we may wonder if we need to treat the unmodeled signal as noise and rewhiten with the new noise spectrum. However, this is not necessary since those components of the channel will be spatially orthogonal to the channel components used and will therefore not pass through any of the beamformers in the reduced rank MLSE.

5.3.3 Reduced Rank DFE

Spatio-Temporal DFE

Reduced rank equalization can also be applied in a DFE. Either one can apply the rank reduction to an MMF-DFE or one can apply the rank reduction to a DFE for a FIR channel with AR noise as in Section 3.2.2 as well as to the MMF-DFE in Section 3.2.4. Rank reduction applied to the MMF-DFE will be the direct counterpart to rank reduction applied to the spatio-temporal MLSE described above. The only difference is that the scalar Viterbi is replaced by a scalar DFE. The MMF-DFE has an advantage over the DFE in Section 3.2.2 when the number of antennas or input signals are large. However, in the case of a reduced rank DFE, the number of input signals to the DFE, i.e. the reduced number of beamformers, will typically be small. In this case it is probably preferable to not use the MMF-DFE. The advantage in the tuning complexity will be small, if any, and it will have a higher execution complexity and a larger processing delay.

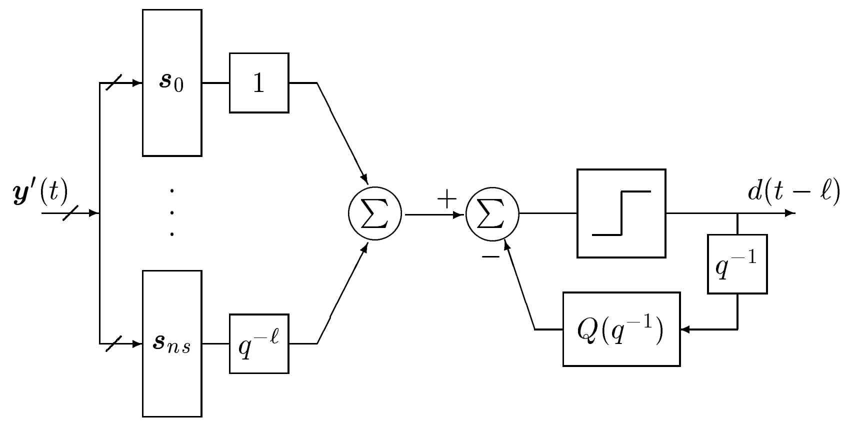


Figure 5.15: Structure of the general MISO FIR decision feedback equalizer, with M antenna elements.

The application of the reduced rank concept to the DFE of Section 3.2.2 requires some further explanation. Consider the representation of a space-time DFE as presented in Figure 5.15. The received signal samples $\boldsymbol{y}(t)$ are filtered through the feedforward filter, $\boldsymbol{s}(q^{-1}) = \boldsymbol{s}_0 + \boldsymbol{s}_1 + \ldots + \boldsymbol{s}_{ns}q^{-ns}$ and previously decided symbols are filtered through the feedback filter $Q(q^{-1}) = Q_0 + Q_1 q^{-1} + \ldots + Q_{nq}q^{-nq}$ to form an estimate $\hat{d}(t-\ell)$ of the symbol $d(t-\ell)$.

When tuning the equalizer we assume that all previously decided symbols fed into the feedback filter are correct. Thus we seek the equalizer which minimizes the criterion

$$J = E[|\hat{d}(t-\ell) - d(t-\ell)|^2]. \tag{5.64}$$

We further assume that the noise has been whitened by the noise whitening filter $\boldsymbol{M}_0^{-1}\boldsymbol{N}(q^{-1})$. The optimal feedforward filter coefficients can then, as in Section 3.2.2, be computed as

$$\boldsymbol{s}(q^{-1}) = \boldsymbol{s}_0(q^{-1})\boldsymbol{M}_0^{-1}\boldsymbol{N}(q^{-1}). \tag{5.65}$$

The coefficients of the polynomial row vector $\boldsymbol{s}_0(q^{-1})$ of degree ℓ can be computed by solving the system of equations

$$(\boldsymbol{\mathcal{B}}'\boldsymbol{\mathcal{B}}'^H + \boldsymbol{I})\boldsymbol{s}_0^H = \begin{bmatrix} \boldsymbol{b}'_\ell \\ \vdots \\ \boldsymbol{b}'_0 \end{bmatrix} \tag{5.66}$$

where

$$\boldsymbol{\mathcal{B}}' = \begin{bmatrix} \boldsymbol{b}'_0 & \cdots & \boldsymbol{b}'_\ell \\ & \ddots & \vdots \\ 0 & & \boldsymbol{b}'_0 \end{bmatrix} \tag{5.67}$$

$$\boldsymbol{s}_0 = \begin{bmatrix} \boldsymbol{s}_{0,0} & \dots & \boldsymbol{s}_{0,\ell} \end{bmatrix} \tag{5.68}$$

and where $\boldsymbol{b}'_i$ are the vector taps of the noise whitened channel

$$\boldsymbol{b}'(z^{-1}) = \boldsymbol{b}'_0 + \boldsymbol{b}'_1 z^{-1} + \ldots + \boldsymbol{b}'_{nb'} z^{-nb'} = \boldsymbol{M}_0^{-1} \boldsymbol{N}(q^{-1}) \boldsymbol{b}(q^{-1}) \tag{5.69}$$

with $\boldsymbol{b}'_i = 0$ if $i > nb' = nb + nn$. The vectors $\boldsymbol{s}_{0,k}$ are the vector taps in the polynomial

$$\boldsymbol{s}_0(q^{-1}) = \boldsymbol{s}_{0,0} + \boldsymbol{s}_{0,1} q^{-1} + \ldots + \boldsymbol{s}_{0,\ell} q^{-\ell}. \tag{5.70}$$

It is of course required that $(\boldsymbol{\mathcal{B}}' \boldsymbol{\mathcal{B}}'^H + \boldsymbol{I})$ is invertible.

The coefficients of the feedback polynomial, $Q(q^{-1})$ of order $nq = nb + nn - 1$, can as in (3.37) be computed as

$$\begin{bmatrix} Q_0^H \\ \vdots \\ Q_{nq}^H \end{bmatrix} = \begin{bmatrix} \boldsymbol{b}'^H_{\ell+1} & \cdots & \boldsymbol{b}'^H_1 \\ \vdots & & \vdots \\ \boldsymbol{b}'^H_{\ell+nb+nn} & \cdots & \boldsymbol{b}'^H_{nb+nn} \end{bmatrix} \begin{bmatrix} \boldsymbol{s}_{0,0}^H \\ \vdots \\ \boldsymbol{s}_{0,\ell}^H \end{bmatrix} \tag{5.71}$$

where $\boldsymbol{b}'_i = 0$ if $i > nb' = nb + nn$.

Rank Reduction

Using the reduced rank decomposition of the noise whitened channel matrix in (5.51), a reduced rank DFE can be computed. We define the matrices

$$\begin{aligned}
\boldsymbol{\mathcal{U}}'_r &\triangleq \begin{bmatrix} \boldsymbol{U}'_r & & 0 \\ & \ddots & \\ 0 & & \boldsymbol{U}'_r \end{bmatrix} \\
\boldsymbol{\mathcal{V}}'_r &\triangleq \begin{bmatrix} \boldsymbol{v}'_{\text{row } 1} & & 0 \\ \vdots & \ddots & \\ \boldsymbol{v}'_{\text{row } m} & \cdots & \boldsymbol{v}'_{\text{row } 1} \end{bmatrix} \\
\boldsymbol{\mathcal{V}}'_{r,\ell} &\triangleq \begin{bmatrix} \boldsymbol{v}'_{\text{row } \ell} & \cdots & \boldsymbol{v}'_{\text{row } 1} \end{bmatrix}
\end{aligned}$$

where

$$\boldsymbol{v}'_{\text{row } k} \triangleq \begin{cases} \text{row } k \text{ in } \boldsymbol{V}'_r & \text{if } k \leq nb+1 \\ 0 & \text{if } k > nb+1 \end{cases} .$$

The matrices $\boldsymbol{U}'_r$ and $\boldsymbol{V}'_r$ are the left and right singular vectors of the reduced rank channel model in (5.51). We can now write

$$\boldsymbol{\mathcal{B}} = \boldsymbol{\mathcal{U}}'_r \boldsymbol{\mathcal{V}}'^{H}_r \tag{5.73}$$

and

$$\begin{bmatrix} \boldsymbol{b}'_\ell \\ \vdots \\ \boldsymbol{b}'_0 \end{bmatrix} = \boldsymbol{\mathcal{U}}'_r \boldsymbol{\mathcal{V}}'^{H}_{r,\ell} \tag{5.74}$$

Inserting (5.73) and (5.74) into (5.66) gives us

$$(\boldsymbol{\mathcal{U}}'_r \boldsymbol{\mathcal{V}}'^{H}_r \boldsymbol{\mathcal{V}}'_r \boldsymbol{\mathcal{U}}'^{H}_r + \boldsymbol{I})\boldsymbol{s}_0^H = \boldsymbol{\mathcal{U}}'_r \boldsymbol{\mathcal{V}}'^{H}_{r,\ell}. \tag{5.75}$$

We now change to the basis consisting of the columns of $[\boldsymbol{\mathcal{U}}'_r \;\; \boldsymbol{\mathcal{U}}'_{r,\perp}]$, where $\boldsymbol{\mathcal{U}}'_{r,\perp}$ is a matrix whose columns span the orthogonal complement to the span of $\boldsymbol{\mathcal{U}}'_r$. Noting that $\boldsymbol{\mathcal{U}}'^{H}_r \boldsymbol{\mathcal{U}}'_r = \mathrm{I}$ and $\boldsymbol{\mathcal{U}}'^{H}_{r,\perp} \boldsymbol{\mathcal{U}}'_r = 0$, equation (5.75) can be reduced to the two equations

$$(\boldsymbol{\mathcal{V}}'^{H}_r \boldsymbol{\mathcal{V}}'_r + \boldsymbol{I})\boldsymbol{\mathcal{U}}'^{H}_r \boldsymbol{s}_0^H = \boldsymbol{\mathcal{V}}'^{H}_{r,\ell} \tag{5.76a}$$

$$\boldsymbol{\mathcal{U}}'^{H}_{r,\perp} \boldsymbol{s}_0^H = 0 \; . \tag{5.76b}$$

To solve (5.76a) we can introduce the product

$$\boldsymbol{g}^H = \begin{bmatrix} \boldsymbol{g}_0 & \ldots & \boldsymbol{g}_\ell \end{bmatrix}^H = \boldsymbol{\mathcal{U}}'^{H}_r \boldsymbol{s}_0^H$$

where $\boldsymbol{g}_i$ are $1 \times K_r$ column vectors, is an unknown vector to be estimated and rewrite equation (5.76a) as

$$(\boldsymbol{\mathcal{V}}'^{H}_r \boldsymbol{\mathcal{V}}'_r + \boldsymbol{I})\boldsymbol{g}^H = \boldsymbol{\mathcal{V}}'^{H}_{r,\ell}.$$

After having solved for $\boldsymbol{g}$ in this equation we can note that, since $\boldsymbol{\mathcal{U}}'^{H}_r \boldsymbol{\mathcal{U}}'_r = \boldsymbol{I}$ and $\boldsymbol{\mathcal{U}}'^{H}_{r,\perp} \boldsymbol{\mathcal{U}}'_r = 0$, the parameter vector

$$\boldsymbol{s}_0 = \boldsymbol{g}\boldsymbol{\mathcal{U}}'^{H}_r \tag{5.77}$$

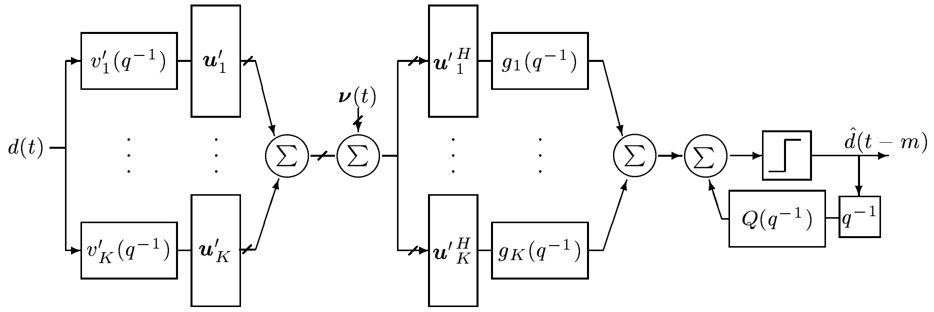


Figure 5.16: Orthogonal decomposition of whitened channel and the feed-forward filter in the DFE.

solves both (5.76a) and (5.76b).

We can now reformulate (5.77) to obtain an an alternative expression for $\boldsymbol{s}_0(q^{-1})$. Due to the special structure of $\boldsymbol{\mathcal{U}}'_r$, we obtain

$$\boldsymbol{s}_0(q^{-1}) = \sum_{k=1}^{K_r} g_k(q^{-1})\boldsymbol{u}'^H_k$$

where we have defined

$$g_k(q^{-1}) \triangleq g_{0k} + g_{1k}q^{-1} + \cdots + g_{\ell k}q^{-\ell}$$

with g_{kj} being element j in $\boldsymbol{g}_k$.

We have thus computed the feedforward filter, of course here excluding the noise whitening filter $\boldsymbol{M}_0^{-1}\boldsymbol{N}(q^{-1})$, which is assumed to proceed the equalizer. The noise whitened channel and the DFE can now be orthogonally decomposed as shown in Figure 5.16.

Let us now consider the construction of a "near optimal" DFE with a lower rank than the rank of the noise whitened channel. This is not a trivial task but we can motivate an *ad hoc* solution. If we want to construct a DFE of rank $K_r < K_{b'}$ we can keep the K_r first branches of the DFE. The beamformers in these branches will collect the signals from the K_r strongest principal components of the channel $\boldsymbol{b}'(q^{-1})$. Since the innovations noise, $\boldsymbol{\nu}(t)$, is white we will in this way have collected the largest amount of signal energy compared to the collected amount of noise energy.

Given the assumed channel, it is likely to believe that the reduced rank DFE constructed in this way is a reasonable reduced rank choice. One can however easily find examples showing that it is not optimal (in an MMSE sense prior to the decision device), for a given reduced rank.

It is important to note that, *unlike in the corresponding case for the MLSE*, the temporal filters here have to be *retuned*, working with the corresponding reduced rank channel. In the MLSE case we could simply remove branches from the full rank solution. As for the multi-dimensional matched filter in (5.62), the noise whitening filter has to be included in the final feedforward filter of the DFE. The feedforward filter with the noise whitening filter included will be given by

$$\boldsymbol{s}(q^{-1}) = \sum_{k=1}^{K_r} g_k(q^{-1})\boldsymbol{u'}_k^H \boldsymbol{M}_0^{-1}\boldsymbol{N}(q^{-1}) \tag{5.78}$$

where

$$g_k(q^{-1}) = g_{0k} + g_{1k}q^{-1} + \ldots + g_{\ell k}q^{-\ell}. \tag{5.79}$$

Given the coefficients of the feedforward filter, the coefficients of the feedback polynomial, $Q(q^{-1})$ of order $nq = nb + nn - 1$, can be computed as

$$\begin{bmatrix} Q_0^H \\ \vdots \\ Q_{nq}^H \end{bmatrix} = \begin{bmatrix} \boldsymbol{b}_{\ell+1}^H & \cdots & \boldsymbol{b}_{\ell+1-ns}^H \\ \vdots & & \vdots \\ \boldsymbol{b}_{\ell+nb+nn}^H & \cdots & \boldsymbol{b'}_{\ell+nb+nn-ns}^H \end{bmatrix} \begin{bmatrix} \boldsymbol{s}_0^H \\ \vdots \\ \boldsymbol{s}_{ns}^H \end{bmatrix} \tag{5.80}$$

where $\boldsymbol{b}_i = 0$ if $i > nb$ or if $i < 0$.

The resulting reduced rank DFE is depicted in Figure 5.17. As for the reduced rank MLSE, we can note that if the noise whitening filter is purely spatial, i.e. $\boldsymbol{N}(q^{-1}) = \boldsymbol{I}$, then the spatial noise whitening, $\boldsymbol{M}_0^{-1}$ can be incorporated in the spatial beamformers, i.e. we get the new *noise whitening beamformers* $\tilde{\boldsymbol{u}}'^H_k = \boldsymbol{u'}_k^H \boldsymbol{M}_0^{-1}$.

Also as for the MLSE, the simplest reduced rank DFE will be the rank one DFE consisting of a noise whitening filter and beamformer followed by a scalar DFE as is shown in Figure 5.18. Note again that if the noise whitening filter is purely spatial, i.e. $\boldsymbol{N}(q^{-1}) = \boldsymbol{I}$ then the reduced rank DFE of rank one consists simply of a *noise whitening beamformer*, $\tilde{\boldsymbol{u}}'_1 = \boldsymbol{u'}_1^H \boldsymbol{M}_0^{-1}$, followed by a scalar DFE.

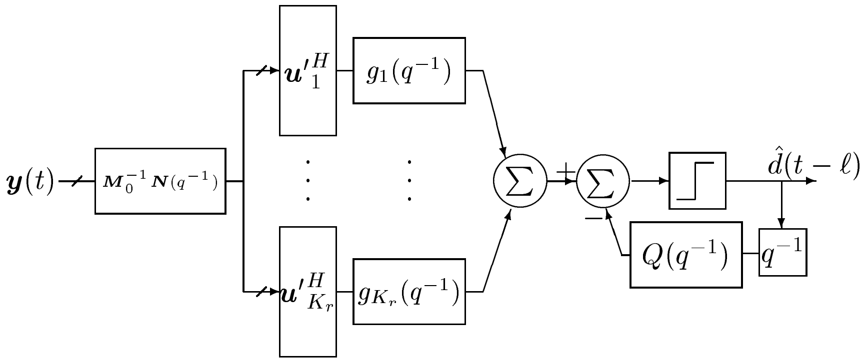


Figure 5.17: Reduced rank DFE.

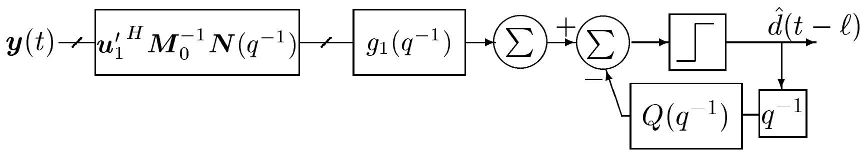


Figure 5.18: Rank one DFE.

Again, it is important to keep in mind that a requirement for the analysis here is that the true channel really can be modeled with the channel model in (5.21), and the noise model in (5.22). When this is not the case, the analysis here will only hold approximatively.

Remark: As for the MLSE, it is not necessary to rewhiten the channel by treating the unused components of the received signal as noise since these components are orthogonal to the used beamformers and thus will not affect the equalization.

5.3.4 Complexity

The execution of the reduced rank filters will typically require less computations than the execution of the corresponding full rank filters. The complexity of the execution of the full rank (FR) and the reduced rank (RR) multidimensional matched filter (MMF) for the space-time MLSE and the MMF-DFE is approximately

$$C_{\text{FR, MMF exe}} \approx N_d M(nb + 2nn + 1)\ cu \tag{5.81}$$
$$C_{\text{RR, MMF exe}} \approx N_d K_r(M + nb + nn + 1) + N_d M nn\ cu. \tag{5.82}$$

Here N_d is the number of symbols equalized, M is the number of antenna elements and $nb+1$ is the channel length and nn is the order of the AR model for the noise plus interference. The *complexity unit, cu*, is the complexity of one complex multiplication and one complex addition.

The corresponding complexities for the execution of the feedforward filter in the FIR-DFE is

$$C_{\text{FR, S exe}} \approx N_d M(\ell + nn + 1)\ cu \tag{5.83}$$
$$C_{\text{RR, S exe}} \approx N_d K_r(M + \ell + 1) + N_d M nn\ cu. \tag{5.84}$$

where ℓ is the decision delay in the feedforward filter in the FIR-DFE.

The largest savings in complexity is achieved when the noise model is purely spatial and we use a rank one equalizer instead of the full rank equalizer. When going from the full rank equalizer to the rank one equalizer, the product between the number of antennas and the channel length or feedforward filter length is replaced with their sum. The saving will thus be the largest when the number of antennas, M, and the length of the feedforward filter, here $\ell + 1$, both are large.

5.3.5 Experiments on Measured Data

To investigate the performance of reduced rank equalization, we applied full and reduced rank versions of the MLSE and the DFE to a set of uplink measurements.

The Measurements

The measurements were performed on an antenna array testbed constructed by Ericsson Radio Systems AB and Ericsson Microwave Systems AB [5]. The testbed implemented the air interface of a DCS-1800 base station. The array had eight antenna outputs. The measurements were performed in downtown Düsseldorf, Germany.

In the measurements one mobile and one interferer were used. Each of them were mounted in a car, which was driving at approximately 30 km/h. The transmit powers of the mobile and the interferer were adjusted so that the scenario would be interference limited, i.e. the performance of the algorithms would be limited by the interference and not by noise.

Algorithms

We used only the spatial spectrum of the noise plus interference in this experiment. Both the spatio-temporal MLSE and the spatio-temporal DFE require the estimation of the multipath channel and the spatial covariance matrix of the noise. We estimated the channel using the off-line least squares method. The spatial noise covariance matrix was computed from the residuals of the channel identification. The estimated channel had five taps. The equalizers utilized the signals from the eight antennas present in the testbed and the DFE had a decision delay of four symbols intervals.

Results

We applied rank 1 and rank 2 versions of the MLSE and the DFE to the experimental data from the array antenna, and compared their performances to those of their full rank (rank 5) counterparts. The results are shown in Figure 5.19.

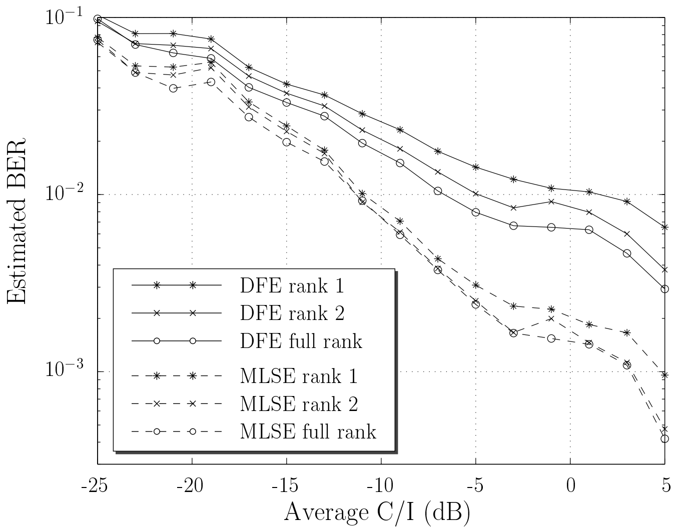


Figure 5.19: Performance of the full and the reduced rank equalizers.

For the MLSE we see that the rank one version performs almost as well as the full rank MLSE. The rank two version has a BER *very* close to the full rank MLSE.

For the DFE the rank one version has some loss in performance. The rank two version conforms better with the performance of the full rank DFE, showing only a small loss in performance.

By applying the rank reduction, we conclude that the complexity of the structure and the complexity in the execution of spatio-temporal equalizers can be reduced. The experimental study presented here demonstrates that for practical wireless communication channels, reduced rank equalizers may provide adequate performance.

5.3.6 Conclusions

By utilizing the spatio-temporal structure of a channel, rank reduced models of the channel can be constructed. By using such rank reduced models, the complexity of the structure and the complexity of the execution of spatio-temporal equalizers can be reduced. The experimental study presented here demonstrates that for practical wireless communication channels, reduced rank equalizers may provide adequate performance.

5.4 Reduced Rank Tuning

In Section 5.3 we derived a reduced rank MLSE from a reduced rank model of the channel. We can however utilize the concept of reduced rank modeling in a different way. We will here utilize a rank one model of a noise whitened channel and derive an alternate way of tuning a rank one MLSE. The solution will turn out to be very similar to the *hybrid* MLSE proposed earlier in [53]. However, with the formalism developed here, we believe that the resulting method can be more easily related to other space-time MLSE solutions.

Let us return the the joint estimation of an FIR channel and an AR noise model in Section 2.9.

To start with we consider the channel model in (2.112)

$$\boldsymbol{y}(t) = \boldsymbol{b}(q^{-1})d(t) + \boldsymbol{N}^{-1}(q^{-1})\boldsymbol{M}_0\boldsymbol{\nu}(t) \tag{5.85}$$

where the entities are defined as in Section 2.9 except that we have replaced $\boldsymbol{\nu}(t)$ with $\boldsymbol{\nu}(t)$ to denote the innovations sequence in the noise model term. Multiplying both sides by $\boldsymbol{N}(q^{-1})$ gives as in (2.118)

$$\boldsymbol{N}(q^{-1})\boldsymbol{y}(t) = \boldsymbol{N}(q^{-1})\boldsymbol{b}(q^{-1})d(t) + \boldsymbol{M}_0\boldsymbol{\nu}(t). \tag{5.86}$$

Let us now assume that we want to design an MLSE with an MMF filter

$$\boldsymbol{w}(q, q^{-1}) = \boldsymbol{b}^H(q)\boldsymbol{N}^H(q)\boldsymbol{M}_0^{-H}\boldsymbol{M}_0^{-1}\boldsymbol{N}(q^{-1}) \tag{5.87}$$

and a metric polynomial

$$\gamma(q, q^{-1}) = \boldsymbol{b}^H(q)\boldsymbol{N}^H(q)\boldsymbol{M}_0^{-H}\boldsymbol{M}_0^{-1}\boldsymbol{N}(q^{-1})\boldsymbol{b}(q^{-1}) \tag{5.88}$$

as in (4.20) and (4.21).

We can now note that as an alternative to estimating $\boldsymbol{N}(q^{-1})$, $\boldsymbol{b}_{\boldsymbol{N}}(q^{-1}) = \boldsymbol{b}(q^{-1})\boldsymbol{N}(q^{-1})$ and $\boldsymbol{M}_0$, as in Section 2.9 of Chapter 2, we can instead choose to estimate the only the products of $\boldsymbol{N}(q^{-1})$ and $\boldsymbol{M}_0^{-1}$ with $\boldsymbol{b}(q^{-1})$ in the MMF, $\boldsymbol{w}(q, q^{-1})$, in (5.87) and the noise whitened channel

$$\boldsymbol{b}'(q^{-1}) = \boldsymbol{N}(q^{-1})\boldsymbol{M}_0^{-1}\boldsymbol{b}(q^{-1}) \tag{5.89}$$

from which the metric polynomial, $\gamma(q, q^{-1})$, can be formed as

$$\gamma(q, q^{-1}) = \boldsymbol{b}'^H(q)\boldsymbol{b}'(q^{-1}). \tag{5.90}$$

Let us first multiply (5.86) with $\boldsymbol{M}_0^{-1}$ from the left giving

$$\boldsymbol{M}_0^{-1}\boldsymbol{N}(q^{-1})\boldsymbol{y}(t) = \boldsymbol{M}_0^{-1}\boldsymbol{N}(q^{-1})\boldsymbol{b}(q^{-1})d(t) + \boldsymbol{\nu}(t). \tag{5.91}$$

Using (5.89), we can rewrite this equation as

$$\boldsymbol{M}_0^{-1}\boldsymbol{N}(q^{-1})\boldsymbol{y}(t) = \boldsymbol{b}'(q^{-1})d(t) + \boldsymbol{\nu}(t). \tag{5.92}$$

We will now model the noise whitened channel, $\boldsymbol{b}'(q^{-1})$, with a *rank one* channel model as in (5.54), i.e. we assume that

$$\boldsymbol{b}'(q^{-1}) = \boldsymbol{u}'v'(q^{-1}). \tag{5.93}$$

where we have dropped the subscript "1" on the vector $\boldsymbol{u}'$ and the polynomial $v'(q^{-1})$. Substituting this into (5.92) gives

$$\boldsymbol{M}_0^{-1}\boldsymbol{N}(q^{-1})\boldsymbol{y}(t) = \boldsymbol{u}'v'(q^{-1})d(t) + \boldsymbol{\nu}(t). \tag{5.94}$$

Let us now multiply both sides of (5.94) with the vector $\boldsymbol{u}'^H$ giving

$$\boldsymbol{c}(q^{-1})\boldsymbol{y}(t) = v'(q^{-1})d(t) + \nu_{\boldsymbol{u}}(t) \tag{5.95}$$

where we have defined

$$\boldsymbol{c}(q^{-1}) = \boldsymbol{u}'^H\boldsymbol{M}_0^{-1}\boldsymbol{N}(q^{-1}) \tag{5.96}$$

and

$$\nu_{\boldsymbol{u}}(t) = \boldsymbol{u}'^H \boldsymbol{\nu}(t). \tag{5.97}$$

We can now, using (5.93) and (5.96) in (5.87), note that the multidimensional matched filter for the channel in question can be formed as

$$\begin{aligned} \boldsymbol{w}(q, q^{-1}) &= \boldsymbol{b}^H(q)\boldsymbol{N}^H(q)\boldsymbol{M}_0^{-H}\boldsymbol{M}_0^{-1}\boldsymbol{N}(q^{-1}) \\ &= \boldsymbol{b}'^H(q)\boldsymbol{M}_0^{-1}\boldsymbol{N}(q^{-1}) \\ &= v'^H(q)\boldsymbol{u}'^H\boldsymbol{M}_0^{-1}\boldsymbol{N}_0(q^{-1}) \\ &= v'^H(q)\boldsymbol{c}(q^{-1}). \end{aligned} \tag{5.98}$$

Similarly we can note that we can form the metric polynomial $\gamma(q, q^{-1})$ in (5.90) as

$$\begin{aligned} \gamma(q, q^{-1}) &= \boldsymbol{b}'^H(q)\boldsymbol{b}'(q^{-1}) \\ &= v'^H\boldsymbol{u}'^H\boldsymbol{u}'v'(q^{-1}) \\ &= v'^H v'(q^{-1}) \end{aligned} \tag{5.99}$$

where we in the last equality have utilized the fact that the norm of the vector $\boldsymbol{u}'$ is equal to one. We can thus see that we can choose to estimate the scalar polynomial $v'(q^{-1})$ of order $nb' = nb + nn$ in (5.93) and the $1 \times M$ polynomial row vector $\boldsymbol{c}(q^{-1})$ of order nn in (5.96).

Remark: The number of coefficients to be estimated with is thus $nb + nn + 1 + M(nn + 1)$. With the effective training sequence length, $N_{eff} = N_{tseq} - nb - nn$, we can require that

$$N_{tseq} \geq M(nn + 1) + 2 * (nb + nn) + 1. \tag{5.100}$$

which is exactly the same requirement as in (2.127) for the minimum training sequence length for the joint FIR channel and AR noise model estimation in Section 2.9 of Chapter 2.

There are different ways that we can estimate the polynomials $\boldsymbol{c}(q^{-1})$ and $v'(q^{-1})$. To investigate how we can attack this problem we can consider how the multidimensional matched filter and the corresponding metric polynomial are tuned in Section 4.4.1 of Chapter 4. See for example Figure 4.2. It is stated there that we should minimize the variance

$$J_e = E[e(t)e^H(t)] \tag{5.101}$$

of the error signal

$$e(t) = \boldsymbol{w}(q, q^{-1})\boldsymbol{y}(t) - \gamma(q, q^{-1})d(t) \tag{5.102}$$

under the constraint that $\gamma_0 = 1$. If we consider (5.99) we can note that would mean that the norm[5], $\parallel v'(q^{-1}) \parallel = \gamma_0$, of the polynomial, $v'(q^{-1})$, would have to be equal to one. This does of course not agree with our reduced rank model (5.93) where only $\boldsymbol{u}'$ is of unit norm. However, the magnitude of the polynomial $v'(q^{-1})$ is not of importance since it can always be compensated for with a multiplicative constant in $\boldsymbol{c}(q^{-1})$. We can therefore with out loss of generality use the constraint

$$\parallel v'(q^{-1}) \parallel = 1. \tag{5.103}$$

If we substitute (5.98) and (5.99) into (5.102) we get

$$\begin{aligned} e(t) &= {v'}^H(q)\boldsymbol{c}(q^{-1})\boldsymbol{y}(t) - {v'}^H(q)v'(q^{-1})d(t) \\ &= {v'}^H(q)\left(\boldsymbol{c}(q^{-1})\boldsymbol{y}(t) - v'(q^{-1})d(t)\right) \\ &= {v'}^H(q)\nu_{\boldsymbol{u}}(t). \end{aligned} \tag{5.104}$$

Substituting (5.104) into (5.101) gives us

$$J_e = E[e(t)e^H(t)] = E[v'(q^{-1})\nu_{\boldsymbol{u}}(t)(v'(q^{-1})\nu_{\boldsymbol{u}}(t))^H]. \tag{5.105}$$

However, since $\nu(t)$ is temporally white, and the norm of $v'(q^{-1})$ is constrained to 1, we have

$$J_e = E[\nu_{\boldsymbol{u}}(t)\nu_{\boldsymbol{u}}^H(t)]. \tag{5.106}$$

We can now rewrite (5.95) in matrix notation as

$$\boldsymbol{c}\bar{\boldsymbol{y}}(t) = \boldsymbol{v}'\boldsymbol{d}(t) + \nu_{\boldsymbol{u}}(t) \tag{5.107}$$

where

$$\boldsymbol{c} = \begin{bmatrix} \boldsymbol{c}_0 & \dots & \boldsymbol{c}_{nn} \end{bmatrix} \tag{5.108}$$

[5] Here the norm $\parallel v'(q^{-1}) \parallel$ is the square root of the sum of the squared magnitudes of the coefficients of the polynomial $v'(q^{-1})$.

$$\bar{\boldsymbol{y}}(t) = \begin{bmatrix} \boldsymbol{y}(t) \\ \vdots \\ \boldsymbol{y}(t-nn) \end{bmatrix} \tag{5.109}$$

$$\boldsymbol{v}' = \begin{bmatrix} v_0' & \dots v_{nb'}' \end{bmatrix} \tag{5.110}$$

$$\boldsymbol{d}(t) = \begin{bmatrix} d(t) \\ \vdots \\ d(t-nb-nn) \end{bmatrix} \tag{5.111}$$

where $\boldsymbol{c}_i$, $i = 0, \dots, nn$, are the row vector coefficients of the row vector polynomial $\boldsymbol{c}(q^{-1})$ of order nn and v_i', $i = 0, \dots, nb'$, are the scalar coefficients of the polynomial $v'(q^{-1})$ of order $nb' = nb + nn$.

The measure J_e in (5.106) can thus be expressed as

$$\begin{aligned} J_e &= E[\nu_{\boldsymbol{u}}(t)\nu_{\boldsymbol{u}}^H(t)] \\ &= E[(\boldsymbol{c}\bar{\boldsymbol{y}}(t) - \boldsymbol{v}'\boldsymbol{d}(t))(\boldsymbol{c}\bar{\boldsymbol{y}}(t) - \boldsymbol{v}'\boldsymbol{d}(t))^H] \\ &= \boldsymbol{c}\boldsymbol{R}_{\bar{\boldsymbol{y}}\bar{\boldsymbol{y}}}\boldsymbol{c}^H - \boldsymbol{c}\boldsymbol{R}_{\bar{\boldsymbol{y}}\boldsymbol{d}}\boldsymbol{v}'^H - \boldsymbol{v}'\boldsymbol{R}_{\boldsymbol{d}\bar{\boldsymbol{y}}}\boldsymbol{c}^H + \boldsymbol{v}'\boldsymbol{R}_{\boldsymbol{d}\boldsymbol{d}}\boldsymbol{v}'^H \end{aligned} \tag{5.112}$$

where

$$\boldsymbol{R}_{\bar{\boldsymbol{y}}\bar{\boldsymbol{y}}} = E[\bar{\boldsymbol{y}}(t)\bar{\boldsymbol{y}}^H(t)] \tag{5.113}$$

$$\boldsymbol{R}_{\bar{\boldsymbol{y}}\boldsymbol{d}} = \boldsymbol{R}_{\boldsymbol{d}\bar{\boldsymbol{y}}}^H = E[\bar{\boldsymbol{y}}(t)\boldsymbol{d}^H(t)] \tag{5.114}$$

and

$$\boldsymbol{R}_{\boldsymbol{d}\boldsymbol{d}} = E[\boldsymbol{d}(t)\boldsymbol{d}^H(t)]. \tag{5.115}$$

We now want to minimize (5.112) with respect to $\boldsymbol{c}(q^{-1})$ and $\boldsymbol{v}'(q^{-1})$, under the constraint

$$\boldsymbol{v}'\boldsymbol{v}'^H = 1. \tag{5.116}$$

To do this we first minimize (5.112), assuming $\boldsymbol{v}'$ to be fixed. Differentiating J_e in (5.112) with respect to $\boldsymbol{c}$ and requiring the derivatives to be zero gives, as shown in Appendix 5.A.1, the solution for $\boldsymbol{c}$

$$\boldsymbol{c} = \boldsymbol{v}'\boldsymbol{R}_{\boldsymbol{d}\bar{\boldsymbol{y}}}\boldsymbol{R}_{\bar{\boldsymbol{y}}\bar{\boldsymbol{y}}}^{-1}. \tag{5.117}$$

Substituting (5.117) into (5.112) now gives

$$J_\nu = \boldsymbol{v}'(\boldsymbol{R_{dd}} - \boldsymbol{R_{d\bar{y}}}\boldsymbol{R_{\bar{y}\bar{y}}^{-1}}\boldsymbol{R_{\bar{y}d}})\boldsymbol{v}'^H \tag{5.118}$$

which should be minimized under the constraint $\boldsymbol{v}'\boldsymbol{v}'^H = 1$. The solution to this is simply given by the eigenvector of the matrix

$$\Lambda = \boldsymbol{R_{dd}} - \boldsymbol{R_{d\bar{y}}}\boldsymbol{R_{\bar{y}\bar{y}}^{-1}}\boldsymbol{R_{\bar{y}d}} \tag{5.119}$$

with minimum eigenvalue, and norm equal to one.

By replacing the covariance matrices in (5.117) and (5.119) with the corresponding sample matrix estimates

$$\hat{\boldsymbol{R}}_{\bar{\boldsymbol{y}}\bar{\boldsymbol{y}}} \triangleq \frac{1}{t_{max} - t_{min} + 1} \sum_{t_{min}}^{t_{max}} \bar{\boldsymbol{y}}(t)\bar{\boldsymbol{y}}^H(t) \tag{5.120}$$

$$\hat{\boldsymbol{R}}_{\boldsymbol{d}\bar{\boldsymbol{y}}} = \hat{\boldsymbol{R}}_{\bar{\boldsymbol{y}}\boldsymbol{d}}^H \triangleq \frac{1}{t_{max} - t_{min} + 1} \sum_{t_{min}}^{t_{max}} \boldsymbol{d}(t)\bar{\boldsymbol{y}}^H(t) \tag{5.121}$$

$$\hat{\boldsymbol{R}}_{\boldsymbol{dd}} \triangleq \frac{1}{t_{max} - t_{min} + 1} \sum_{t_{min}}^{t_{max}} \boldsymbol{d}(t)\boldsymbol{d}^H(t) \tag{5.122}$$

we get the estimates $\hat{\boldsymbol{v}}'$ as the eigenvector to the matrix

$$\hat{\Lambda} = \hat{\boldsymbol{R}}_{\boldsymbol{dd}} - \hat{\boldsymbol{R}}_{\boldsymbol{d}\bar{\boldsymbol{y}}}\hat{\boldsymbol{R}}_{\bar{\boldsymbol{y}}\bar{\boldsymbol{y}}}^{-1}\hat{\boldsymbol{R}}_{\bar{\boldsymbol{y}}\boldsymbol{d}} \tag{5.123}$$

with norm one, and we get the estimate $\hat{\boldsymbol{c}}$ as

$$\hat{\boldsymbol{c}} = \hat{\boldsymbol{v}}'\hat{\boldsymbol{R}}_{\bar{\boldsymbol{d}}\boldsymbol{y}}\hat{\boldsymbol{R}}_{\bar{\boldsymbol{y}}\bar{\boldsymbol{y}}}^{-1}. \tag{5.124}$$

The polynomials $\boldsymbol{c}(q^{-1})$ and $v'(q^{-1})$ and the multidimensional matched filter, $\boldsymbol{w}(q, q^{-1})$ as well as the metric polynomial $\gamma(q, q^{-1})$ can now be computed via (5.108), (5.110), (5.98) and (5.99).

Although not completely identical, a closer study shows this space-time MLSE to be very similar to the space-time MLSE proposed earlier in [53]. There, a two-stage *hybrid* space-time MLSE is proposed, consisting of a space-time linear MISO filter followed by a scalar MLSE. The space-time

MISO filter is tuned to maximize the SINR at its output. This is solved as an optimization problem where the MISO filter coefficients, corresponding to $\boldsymbol{c}(q^{-1})$, and the resulting channel from the transmitted symbol sequence to the output of the MISO filter, corresponding to $\boldsymbol{v}(q^{-1})$, is optimized jointly to maximize the SINR in the signal. It is argued that it is favorable to separate the suppression of co-channel interferers and the handling of the intersymbol interference into two separate parts, a space-time MISO filter and a scalar MLSE. In light of the derivation shown here, this can however be viewed in a different perspective. We can here see that the linear MISO filter, $\boldsymbol{c}(q^{-1})$, can be viewed as *an integral part of a rank one space-time MLSE*, as described here.

Too see that the two methods are almost equivalent we can translate the tuning proposed in [53] into the notation used here, giving the $\hat{\boldsymbol{v}}$ as the eigenvector with *maximum* eigenvalue of the matrix

$$\hat{\Lambda}_2 = (\hat{\boldsymbol{R}}_{\boldsymbol{dd}} - \hat{\boldsymbol{R}}_{\boldsymbol{d}\bar{\boldsymbol{y}}}\hat{\boldsymbol{R}}_{\bar{\boldsymbol{y}}\bar{\boldsymbol{y}}}^{-1}\hat{\boldsymbol{R}}_{\bar{\boldsymbol{y}}\boldsymbol{d}})^{-1}\hat{\boldsymbol{R}}_{\boldsymbol{dd}} \tag{5.125}$$

and the vector $\hat{\boldsymbol{c}}$, as for the method derived here, given by (5.124). The matrix, $\hat{\boldsymbol{R}}_{\boldsymbol{dd}}$, to the right in (5.125) will however, for a training sequence that is chosen to be approximately white, be close to the identity matrix. We can thus with good approximation drop this matrix multiplication from $\hat{\Lambda}_2$ resulting in the new matrix

$$\hat{\Lambda}_2' = (\hat{\boldsymbol{R}}_{\boldsymbol{dd}} - \hat{\boldsymbol{R}}_{\boldsymbol{d}\bar{\boldsymbol{y}}}\hat{\boldsymbol{R}}_{\bar{\boldsymbol{y}}\bar{\boldsymbol{y}}}^{-1}\hat{\boldsymbol{R}}_{\bar{\boldsymbol{y}}\boldsymbol{d}})^{-1}. \tag{5.126}$$

We can now see that

$$\Lambda = (\hat{\Lambda}_2')^{-1}. \tag{5.127}$$

Finding the eigenvector with *minimum* eigenvalue to the matrix $\hat{\Lambda}$ is therefore the same as looking for the eigenvector with the *maximum* eigenvalue to the matrix $\hat{\Lambda}_2'$. Thus, apart from the multiplication with $\hat{\boldsymbol{R}}_{\boldsymbol{dd}}$ in (5.125), the two methods are identical. However, as mentioned above, if the training sequence is approximately white, this matrix will be close to the unit matrix and it turns out we can leave it out without any significant loss in performance.

We can also consider the criteria that the reduced rank tuning and the hybrid MLSE attempts to optimize. The reduced rank tuning strives to maximize the peak to noise ratio in the matched filtered signal $z(t) =$

$\boldsymbol{w}(q,q^{-1})\boldsymbol{y}(t)$. This is achieved by minimizing the variance in the signal $e(t) = \boldsymbol{w}(q,q^{-1})\boldsymbol{y}(t) - \gamma(q,q^{-1})d(t)$, under the constraint that the middle coefficient in $\gamma(q,q^{-1})$, γ_0, should be equal to one [121]. The hybrid MLSE on the other hand strives to maximize the signal to interference and noise ratio at the output of the MISO filter. Maximizing the signal to interference and noise ratio after the MISO filter, $\boldsymbol{c}(q^{-1})$, is however equivalent to minimizing the variance of the signal $\nu_{\boldsymbol{u}}(t) = \boldsymbol{c}(q^{-1})\boldsymbol{y}(t) - v(q^{-1})d(t)$ in (5.97), under the constraint that the norm $\| v(q^{-1}) \|$ should be equal to one. This is exactly the criteria we arrive to for the reduced rank tuning, see (5.106), and therefore the two methods optimize the same criteria. The difference in the methods comes from different realizations of the optimization.

When the parameter nn is zero, then rank one space-time MLSE's discussed here performs only spatial suppression of the interference. When we have $nn > 1$, the rank one MLSE's performs spatio-temporal suppression of the interference which can be advantageous if we have many strong uncorrelated interfering signals and a relatively small number of antennas. In this case the spatio-temporal interference suppression can be powerful, allowing very low signal to interference ratios.

If we have too many antennas and use $nn > 0$ we may, however, have trouble with the tuning of the increasing number of parameters. As the number of parameter increases it also seems that these methods require a relatively high SNR.

5.4.1 Simulations

We will here present a simulation example to study the performance of the reduced rank tuning and the hybrid MLSE. We also compare their performance to a full rank MLSE and a two reduced rank MLSE's computed as in Section 5.3.

We will here view the antenna elements as if they were spaced far enough apart in a Rayleigh fading environment, such that the channels at the individual antenna elements fade independently[6]. The antenna array had six

[6]This is a worst case scenario for reduced rank methods. If the signals arrive from more distinct directions the channel is more likely to be well modeled by a reduced rank model.

antenna elements. The signal-to-noise ratio was 5 dB.

We have one desired user and one co-channel interferer. The channel for the desired user as well as the channel for the co-channel interferer have two independently Rayleigh fading taps, spaced a symbol interval apart, with equal mean power. As modulation we use the pulse shaping of GMSK modulation with a bandwidth-time product of 0.3 as described in Appendix 2.A.1 of Chapter 2. The channel for the desired signal will here in general be a rank two channel, due two the two taps in the channel, with in general different spatial signatures.

We compare compare four different algorithms, a full rank MLSE designed from least squares estimates of the FIR channel and utilizing only the spatial spectrum of the noise plus interference, a rank one and a rank two version of the same equalizer, as described in Section 5.3.2, both using only the spatial spectrum of the noise plus interference, the reduced rank tuning described in this section with $nn = 0$ and a hybrid MLSE [53] with a 1-tap MISO filter. All MLSE's used the same model order of the channel, $nb = 4$, for the desired signal.

In Figure 5.20 the BER for the different algorithms is plotted as a function of the signal to interference ratio. We can make the following observations for this simulation. The reduced rank tuned MLSE and the hybrid MLSE have as expected a very similar performance. However, the rank one MLSE tuned which is tuned using a rank one model of the channel has a slightly better performance than both these algorithms. The full rank MLSE has a better performance than the rank one MLSE. This is not so surprising since the channel is a rank two channel. The rank two MLSE has the best performance. This may seem surprising, however, it only shows that the rank two MLSE benefits from the fact that the channel truly is a rank two channel.

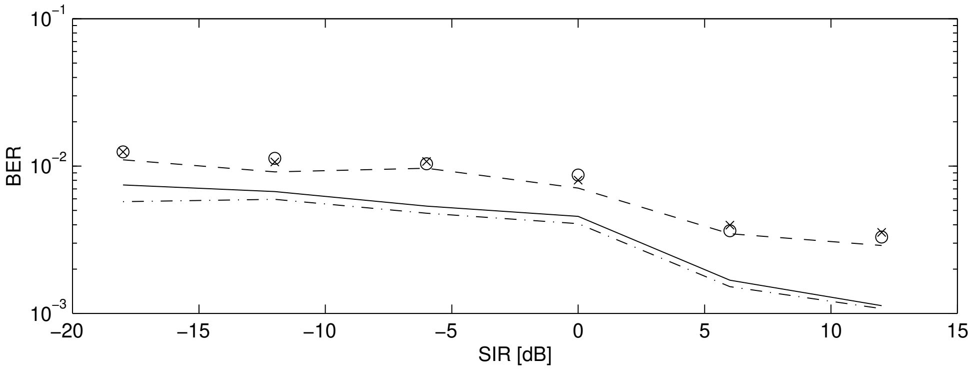


Figure 5.20: 6 antennas and SNR=5 dB. BER as a function of SINR. All algorithms do only spatial interference suppression. Full rank solution (solid), Rank one solution (dashed), Rank two solution (dash-dotted), Reduced rank tuning with $nn = 0$ (x) and the hybrid MLSE with a 1-tap MISO filter (o).

5.A Appendix

5.A.1 Solving for $\boldsymbol{c}$ given $\boldsymbol{v}$

Let us consider $\boldsymbol{v}$ to be constant and solve for the vector $\boldsymbol{c}$ that minimizes J_e

$$J_e = J_e(\boldsymbol{c}) = \boldsymbol{c}\boldsymbol{R}_{\bar{\boldsymbol{y}}\bar{\boldsymbol{y}}}\boldsymbol{c}^H - \boldsymbol{c}\boldsymbol{R}_{\bar{\boldsymbol{y}}\boldsymbol{d}}\boldsymbol{v}'^H - \boldsymbol{v}'\boldsymbol{R}_{\boldsymbol{d}\bar{\boldsymbol{y}}}\boldsymbol{c}_v^H + \boldsymbol{v}'\boldsymbol{R}_{\boldsymbol{dd}}\boldsymbol{v}'^H. \tag{5.128}$$

We are thus looking for a vector $\boldsymbol{c}_0$ such that

$$J_e(\boldsymbol{c}_0 + \delta\boldsymbol{c}) - J_e(\boldsymbol{c}_0) \tag{5.129}$$

is equal to zero for all "infinitely small" vectors $\delta\boldsymbol{c}$. Neglecting terms that are quadratic in $\delta\boldsymbol{c}$ gives us

$$J_e(\boldsymbol{c}_0 + \delta\boldsymbol{c}) - J_e(\boldsymbol{c}_0) = \delta\boldsymbol{c}(\boldsymbol{R}_{\bar{\boldsymbol{y}}\bar{\boldsymbol{y}}}\boldsymbol{c}_v^H - \boldsymbol{R}_{\bar{\boldsymbol{y}}\boldsymbol{d}}\boldsymbol{v}'^H) + (\boldsymbol{c}\boldsymbol{R}_{\bar{\boldsymbol{y}}\bar{\boldsymbol{y}}} - \boldsymbol{v}'\boldsymbol{R}_{\boldsymbol{d}\bar{\boldsymbol{y}}})\delta\boldsymbol{c}_v^H.$$

Requiring this expression to be zero for all vector $\delta\boldsymbol{c}$ gives us the solution for $\boldsymbol{c}_0$ as

$$\boldsymbol{c}_0 = \boldsymbol{v}'\boldsymbol{R}_{\boldsymbol{d}\bar{\boldsymbol{y}}}\boldsymbol{R}_{\bar{\boldsymbol{y}}\bar{\boldsymbol{y}}}^{-1} \tag{5.130}$$

assuming $\boldsymbol{R}_{\bar{\boldsymbol{y}}\bar{\boldsymbol{y}}}^{-1}$ to be invertible.

Chapter 6

Bootstrap Equalization and Interferer Suppression

When we consider co-channel interferer suppression in a TDMA framework we should consider how the interference can be suppressed over the entire frame. For example, it may be the case that an interferer is not present during the training sequence of a frame but appears at some other place in the frame. An equalizer or adaptive antenna tuned using only the training sequence data will then potentially have a seriously degraded performance when subject to this interference. We would therefore like to broaden our view and look at the data outside the training sequence and also the data in adjacent frames, to obtain a more complete picture of the interference environment.

In the sections below we present two different methods of exploiting the data outside the training sequence. First, in Section 6.1, we consider bootstrap equalization, where we utilize the data in the frame outside the training sequence by re-tuning our equalizer using previously estimated symbols [109]. Then, in Section 6.2.1, we utilize the noise plus interference spectrum of adjacent frames in order to suppress co-channel interferers appearing outside the training sequence of the frame currently being processed [56]. Finally, in Section 6.2.2, the two methods are combined and seen to complement each other. First the initial symbols are estimated utilizing the noise plus interferer spectrum of the adjacent frames. Then bootstrapping is applied by re-tuning the equalzier using the initially estimated symbols. With an improved initial symbol estimate, by utilizing noise plus interference spectrum estimates from the adjacent frames, the risk that the initial BER will be too high for the bootstrap equalization algorithm to handle is reduced.

Other presentations discussing the problem of suppressing co-channel interferers outside the training sequence can for example be found in [44] and in [42].

6.1 Bootstrap Co-Channel Interference Cancellation

As seen in Section 2.7 bootstrapping, i.e. re-estimating the channel using estimated symbols, can greatly improve the quality of the channel estimates, including the estimate of the noise plus interference spectrum. This will, of course, also improve the detection of the symbols in an equalizer using the channel and noise plus interference spectrum estimates. If we are primarily interested in the equalization it may not be necessary to estimate the channel. An alternative is to use a direct approach when tuning the equalizer. Re-tuning of such an equalizer utilizing detected symbols will also improve the performance.

A general formulation for bootstrap equalization of a TDMA frame is:

1: Tune the equalizer, direct or indirectly, using the initial training data sequence.

2: Equalize and estimate the data symbols of the frame.

3: Retune the equalizer, direct or indirectly, using all data including the estimated data symbols from step 2.

4: Repeat steps 2 and 3 if desired to improve the performance.

Step 1 and 3 in this algorithm may or may not include estimation of the channel depending on if we are using an indirect or a direct method to tune the equalizer.

As an example we will here consider tuning of the space-time equalizer via an indirect scheme. We then require good estimates of the channel for the desired signal and the spectrum of the noise plus interference. Based on a short training sequence, the quality of the estimates may give far from optimal performance of the equalizer. By increasing the length of the training sequence, a higher accuracy of the channel and noise plus interference spectrum estimates will be obtained, which in turn can lead to a considerable improvement of the space-time equalizer. See for example [61].

The reason why the performance of space-time equalizers improve greatly with the length of the training sequence is that the extra training symbols allow for better positioning of deep nulls in the directions of the interference. However, the length of training sequences cannot be modified in existing standards. Their lengths will also be limited in future systems since training/pilot symbols should only consume a minor part of the available bandwidth. Bootstrap equalization can, by using estimated symbols, attain a performance improvement similar what would be obtained by adding extra training data.

Equalizers which use only a few antennas (say one or two) are less powerful in suppressing co-channel interferers and do not benefit in the same way from improvements in the estimates of the channel and the noise plus interferer spectrum as equalizers with more antennas do. As we will see in an example in Section 6.1.1, bootstrap equalization will not produce the same improvement for such a case.

6.1.1 Experiments on Measured Data

To investigate the interferer suppression capability of the bootstrap algorithm, the method was applied to a set of real world uplink measurements. As equalizers we used a space-time MLSE and a space-time DFE, both with spatial-only co-channel interference suppression.

The Measurements

The measurements were collected from an antenna array testbed constructed by Ericsson Radio Systems AB and Ericsson Microwave Systems AB [5]. The testbed used the air interface of a DCS-1800 base station.

The array had eight antenna elements. A conventional single sector antenna with two polarization diversity branches was also included in the measurement setup.

In the measurements one mobile and one interferer were used. Each of them was mounted on a car, which was driving at approximately 30 km/h. The transmit powers of the mobile and the interferer were adjusted so that the scenario would be interference limited, i.e. the performance of the algorithms would be limited by the interference and not by the noise.

A space-time MLSE and a space-time DFE, both indirectly tuned and utilizing the spatial spectrum of the interference were applied to the data recorded at the array antenna and at the single sector antenna.

Channel Estimation and Equalizer Parameters

Both the spatio-temporal MLSE and the spatio-temporal DFE require the estimation of the multipath channel and the spatial covariance of the noise. We estimated the channel using an FIR model with five taps and the least squares method described in Section 2.2. The spatial noise covariance matrix was computed from the residuals of the channel identification as described in Section 2.8.

The MLSE, as well as the DFE, were both based on the 5-tap channel estimate. For the DFE we chose a smoothing lag of 4 and a feedback filter with 4 taps. The DFE was of the type described in Section 3.3.3 using only the spatial spectrum of the noise plus interference. Similarly the MLSE only utilized the spatial spectrum of the noise plus interference. It was tuned as described in Section 4.4.2.

Results

The array antenna: We applied the proposed bootstrap versions of the MLSE and the DFE to the experimental data from the array antenna. From Figure 6.1, we see that it is indeed worthwhile to use the detected symbols to improve the estimates of the channel and the noise spectrum, especially for high C/I: At 5 dB C/I, the second pass is 14 dB better than the first pass for the MLSE and 17 dB for the DFE.

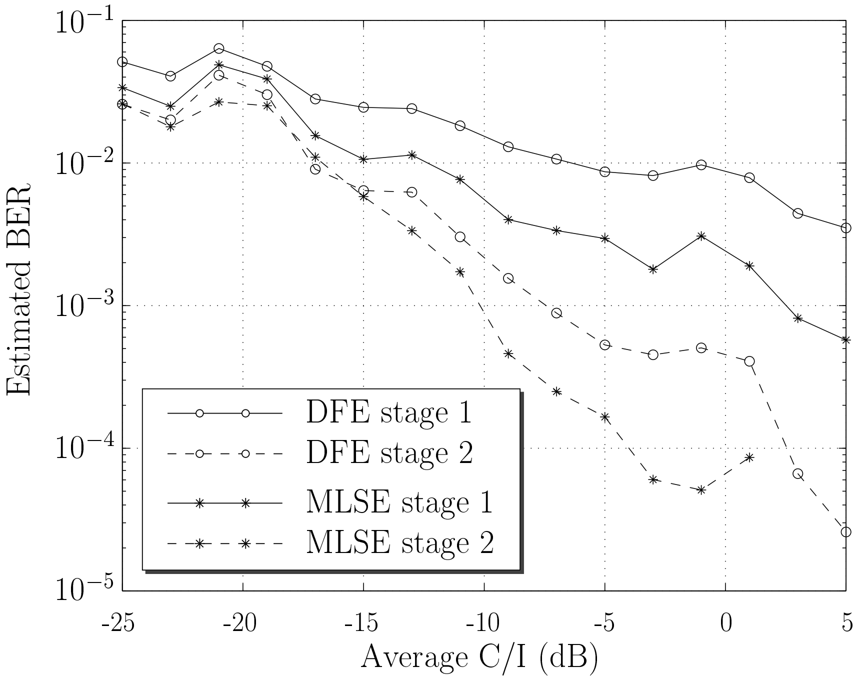


Figure 6.1: Results from the array antenna. Stage 1 refers to the equalizer designed using the training sequence only, whereas stage 2 refers to the equalizer designed using the training sequence and all the detected symbols from stage 1.

The sector antenna: We also applied bootstrapped versions of the MLSE and the DFE to the experimental data from the sector antenna with the two polarization diversity branches. These results are depicted in Figure 6.2.

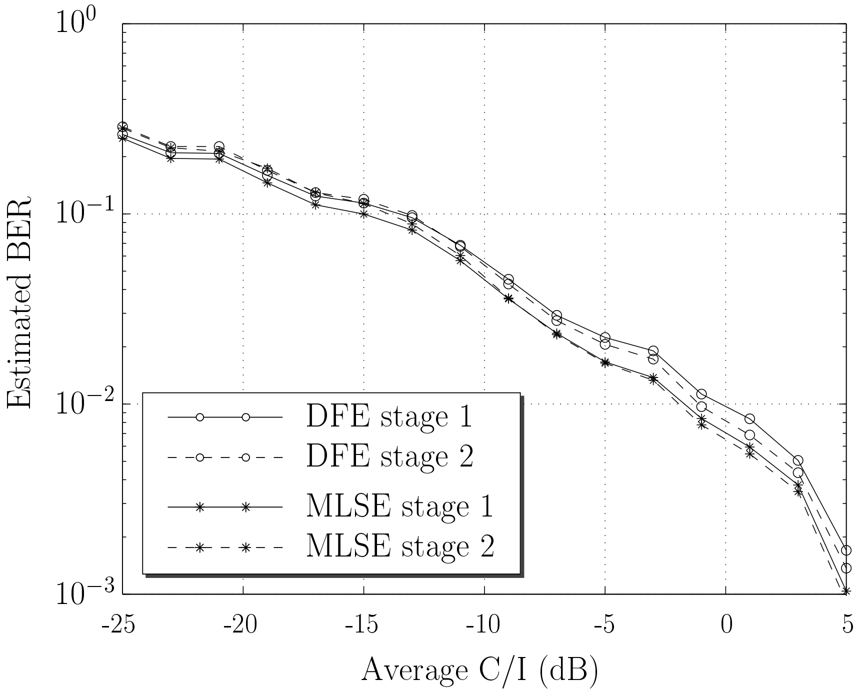


Figure 6.2: Results from the sector antenna. Stage 1 refers to the equalizers designed using the training sequence, whereas stage 2 refers to the equalizers designed using all detected symbols from stage 1 as well as the training sequence.

If we compare Figure 6.2 with Figure 6.1, we see that the gain obtained by using the detected symbols of pass 1 for the sector antenna is only marginal. The reason for this is that the antenna array introduces extra degrees of freedom, which enables powerful space-time processing. However, to make use of these extra degrees of freedom, accurate estimation of the channel and the noise spectrum is essential. For the two-branch diversity antenna, the performance of the space-time processing is not limited by the accuracy of the channel and the noise spectrum estimates, and bootstrapping does not reduce the BER significantly.

Additional stages: Figures 6.1 and 6.2 the results of a single stage of bootstrapping is presented. By iterating the algorithm it is possible to improve the performance even further. Figure 6.3 illustrates how the performance can be improved by using several stages in the antenna array case. In this example it however appears that the major part of the improvement is achieved in stage 2, i.e. after the first retuning stage. Additional stages do, however, give some further improvement. The curves marked with "∞" are lower ideal boundaries for the bootstrapping algorithms. These curves show the performance when the equalizers where tuned with a full frame of correct symbols. We can call it the *genie-aided* performance. Even if we iterate the algorithm through many stages, we will likely not get the performance of the "∞"-curve. For example, some channels will likely have so bad initial equalization results that their BER never will converge, thus the average BER will never reach all the way down to the genie-aided performance.

6.1.2 Discussion

The bootstrap algorithm can typically be applied in conjunction with any type of equalizer and any type of tuning. When using bootstrapping, we can therefore make use of the equalizer and tuning scheme that is most suitable to our specific application. Bootstrapping is particularly useful when the training sequence is too short too obtain a good tuning of the equalizer.

In the example showed in this section, we have seen that when applied to an equalizer utilizing an antenna array, bootstrapping can produce significant improvements in the suppression of the interferers. The retuning here improves the estimate of the channel for the desired signal and, maybe even

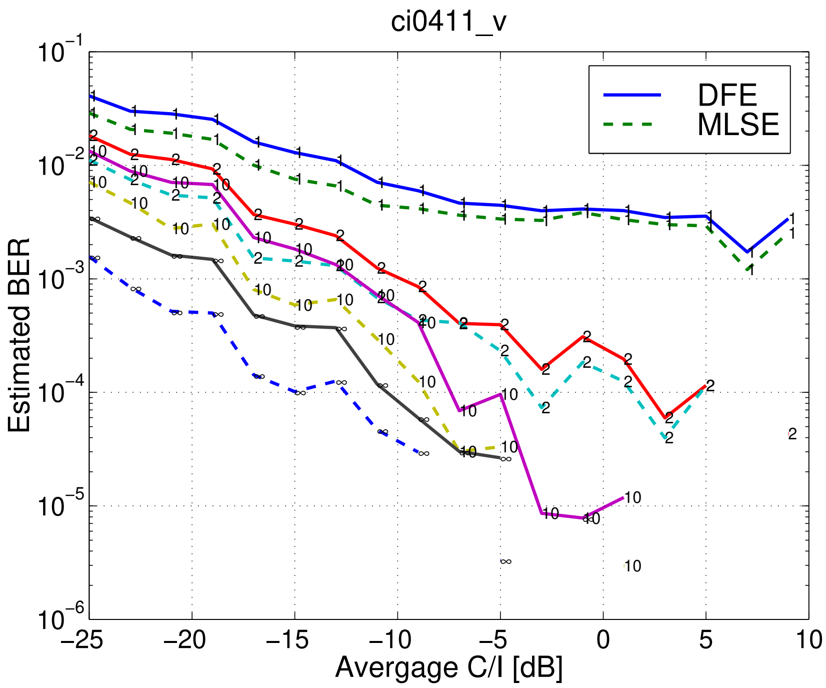


Figure 6.3: Results from the array antenna. DFE (solid) and MLSE (dashed). The numbers by the curves represent the number of iterations performed. The curves marked with "∞" are the result when tuning the equalizers using a full frame of correct symbols.

more importantly, the spectrum of the noise and interference. This allows the equalizer to more efficiently cancel the interference.

6.2 Suppression of Asynchronous Interferers

In a cellular TDMA system it may be the case that co-channel interferers that are not present during the training sequence appear during the data sequence. An equalizer tuned using the training data may then have very poor suppression of such an *asynchronous* interferer. In fact, the interferers may even be amplified.

We will here present two schemes for suppression of asynchronous interferers in a TDMA scheme.

6.2.1 Conservative Initial Detection

If a co-channel interferer is present in a TDMA system during the data sequence but not during the training sequence of a frame, then it will be present during the training sequence of an *adjacent frame*. Information about the interferer plus noise spectrum of adjacent frames can then be utilized in order to suppress such co-channel interferers. The effect on the performance of such a scheme is here illustrated with an example GSM scenario.

In the GSM system data is transmitted in frames, with the training sequence in the middle. In an asynchronous system, co-channel interferers from surrounding cells may thus appear in time as shown in Figure 6.4. For example, co-channel interferer 2 is here present during the data sequence of the "current" frame but not during the training sequence of the current frame. A space-time equalizer tuned to suppress only the co-channel interferers present during the training sequence may thus have a severely degraded performance for the data part of the "current" frame.

The idea is to use spatial spectral information from the adjacent time frame to suppress such interferers. Below we will apply this approach to a simple space-time equalizer consisting of an indirectly tuned beamformer, followed

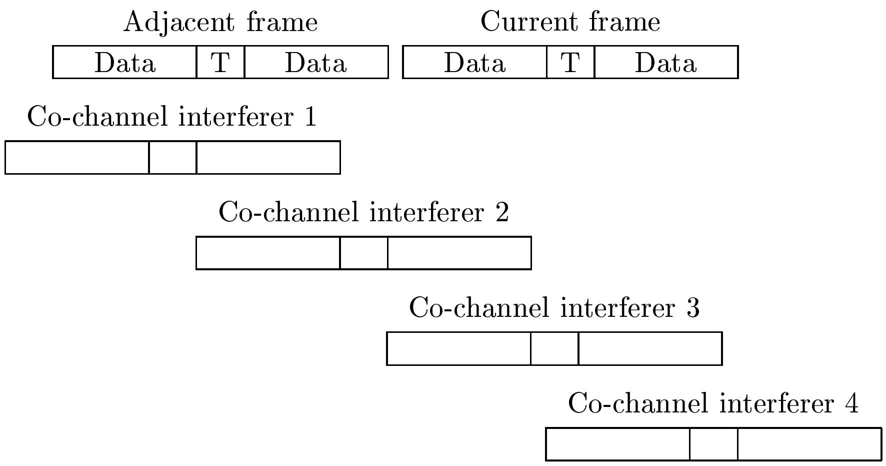


Figure 6.4: Adjacent time frames and example of locations of co-channel interferers in GSM. The training sequence is denoted by "T" and the data sequences are denoted by "Data". Note that every co-channel interferer has to be present during at least one training sequence.

by a scalar MLSE.

Algorithm Outline

We here propose an indirect method for the tuning of the weights in the antenna array. First the FIR channels, $b_i(q^{-1})$, i=1,2,...,M, from the transmitted sequence to each of the antenna elements, see Figure 6.5, are estimated. This estimation is performed based on the data received during the training sequence.

The estimated channels can be used for estimating the spatial spectrum of the desired signal while the residuals from the identification procedure can be used to estimate the spatial spectrum of the interference plus noise.

If a co-channel interferer is not present during the current training sequence it will be present during an adjacent training sequence, see Figure 6.4. When tuning the beamformer we can then do two different computations, one for the part of the data before the training sequence and one for the part of

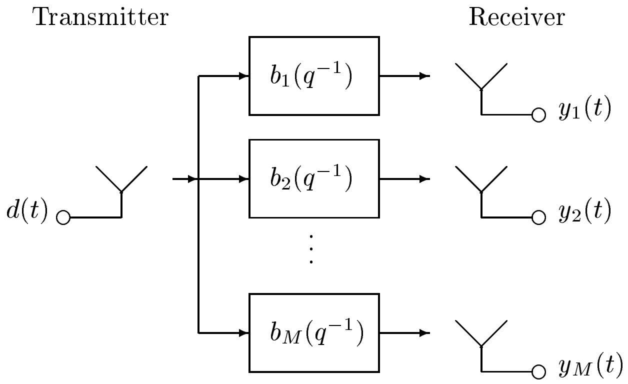


Figure 6.5: Discrete-time baseband channels, $b_i(q^{-1})$, from transmitted symbols, $d(t)$, to received signals $y_i(t)$.

the data after the training sequence. For each tuning, we can use both the spatial spectrum of the interference computed during the current training sequence and the spatial spectrum of the interference computed during the adjacent training sequence. In this way, all co-channel interferers present during each data sequence part, will be accounted for.

Antenna arrays are for practical reasons mainly of interest at the base stations. This scheme is therefore mainly considered for the base station receiver. In this case, the spatial spectrum of the interference plus noise during the adjacent training sequence will either be known from a previous channel estimation procedure (receiving from a different mobile however) or will anyway have to be computed for the subsequent training sequence.

If the scheme is applied at the mobile, then the mobile will have to listen in on the adjacent time frames and estimate their channels and noise plus interference spectrum.

Algorithm Details

The algorithm utilized is similar to the standard sample matrix inversion method presented in Section 5.2.1. However, instead of estimating the covariance matrix of the received signal by means of the sample covariance

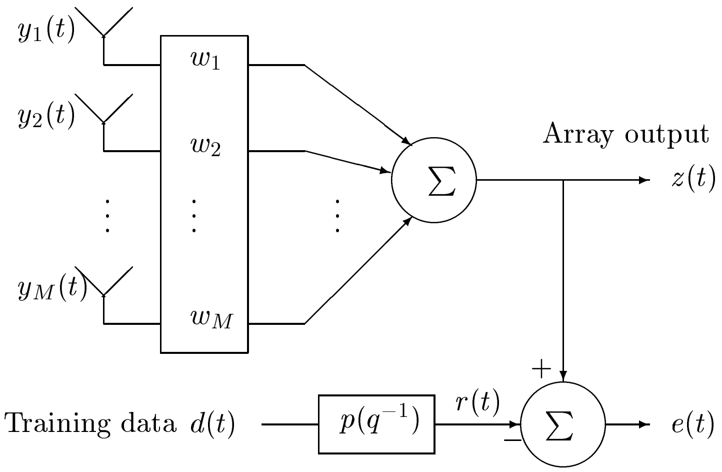


Figure 6.6: The antenna array and the weight coefficients to be tuned.

matrix, we will estimate it based on estimated channels and the sample covariance matrices of the residuals for the current and the adjacent training sequences.

First, the channels

$$b_i(q^{-1}) = b_{i,0} + b_{i,1}q^{-1} + \ldots + b_{i,nb} + q^{-nb} \quad , i = 1, 2, ..., M \tag{6.1}$$

to each of the antenna elements, are estimated with the standard least squares method:

$$\hat{\boldsymbol{b}}_i = (\boldsymbol{\Phi}^H\boldsymbol{\Phi})^{-1}\boldsymbol{\Phi}^H\boldsymbol{D} \quad , i = 1, 2, .., M \tag{6.2}$$

where

$$\hat{\boldsymbol{b}}_i = \begin{bmatrix} \hat{b}_{i,0} \\ \hat{b}_{i,1} \\ \vdots \\ \hat{b}_{i,nb} \end{bmatrix} \tag{6.3}$$

and

$$\boldsymbol{\Phi} = \begin{bmatrix} y_i(nb) & y_i(nb-1) & \dots & y_i(0) \\ y_i(nb+1) & y_i(nb) & \dots & y_i(1) \\ \vdots & \vdots & \vdots & \vdots \\ y(N-1) & y_i(N-2) & \dots & y_i(N-nb) \end{bmatrix}$$

$$\boldsymbol{D} = \begin{bmatrix} d_0 \\ d_1 \\ \vdots \\ d(N-1) \end{bmatrix}. \tag{6.4}$$

The parameter N is the number of symbols in the training sequence. Here $N = 26$ is used.

The antenna array with the beamformer coefficients are depicted in Figure 6.6. In a second step, the beamformer weights, w_i, i=1,2,...,M, are tuned to optimally receive the reference signal

$$r(t) = p(q^{-1})d(t) \tag{6.5}$$

which is an estimate of the transmitted signal. The filter $p(q^{-1})$ models the GMSK modulation used in GSM, see Section 2.A.1. Here, $p(q^{-1}) = 0.44i + 1.00q^{-1} - 0.44iq^{-2}$ was used. This filter is a simple model of the channel between transmitted symbols and received samples. The model includes the GMSK modulation and a model of a receiver filter consisting of a 4th order Butterworth lowpass baseband filter as in Section 2.A.4[1].

The beamformer weights are now selected as an approximation of the Wiener solution

$$\boldsymbol{w} = (\hat{\boldsymbol{R}}_{\boldsymbol{yy}}^{-1}\hat{\boldsymbol{R}}_{\boldsymbol{y}r})^* \tag{6.6}$$

[1]The correct model of the received samples after the GMSK modulation will vary depending on the chosen sampling instant. The chosen model represents only one particular sampling instant. Multipath propagation is not included in the model represented by $p(q^{-1})$, i.e. the propagation channel is modeled as a perfect channel with unit impulse response.

where $(\cdot)^*$ denotes elementwise complex conjugation, and

$$\boldsymbol{w} = \begin{bmatrix} w_1 \\ \vdots \\ w_M \end{bmatrix}. \tag{6.7}$$

The matrices $\hat{\boldsymbol{R}}_{\boldsymbol{yy}}$ and $\hat{\boldsymbol{R}}_{\boldsymbol{y}r}$ are estimates of the covariance matrices

$$\boldsymbol{R}_{\boldsymbol{yy}} = E[\boldsymbol{y}(t)\boldsymbol{y}^H(t)] \tag{6.8}$$

$$\boldsymbol{R}_{\boldsymbol{y}r} = E[\boldsymbol{y}(t)r^H(t)] \tag{6.9}$$

where

$$\boldsymbol{y}(t) = \begin{bmatrix} y_1(t) \\ \vdots \\ y_M(t) \end{bmatrix} \tag{6.10}$$

is the input to the array and $r(t)$ is the reference signal.

The covariance matrix estimate $\hat{\boldsymbol{R}}_{\boldsymbol{y}r}$ is formed, based on the estimated channels, as

$$\hat{\boldsymbol{R}}_{\boldsymbol{y}r} = \hat{\boldsymbol{B}}\boldsymbol{p}^H \tag{6.11}$$

where

$$\hat{\boldsymbol{B}} = \begin{bmatrix} \hat{b}_{1,0} & \hat{b}_{1,1} & \dots & \hat{b}_{1,nb} \\ \hat{b}_{2,0} & \hat{b}_{2,1} & \dots & \hat{b}_{2,nb} \\ \vdots & \vdots & \vdots & \vdots \\ \hat{b}_{M,0} & \hat{b}_{M,1} & \dots & \hat{b}_{M,nb} \end{bmatrix} \tag{6.12}$$

and

$$\boldsymbol{p} = [0.44\mathrm{i}\ \ 1.00\ \ -0.44\mathrm{i}\ \ 0\ \ \dots\ \ 0]. \tag{6.13}$$

An estimate of the covariance matrix estimate $\hat{\boldsymbol{R}}_{\boldsymbol{yy}}$ is obtained as a sum of the desired signal part, $\hat{\boldsymbol{R}}_{\boldsymbol{ss}}$, and an interference plus noise part, $\hat{\boldsymbol{R}}_{\boldsymbol{nn}}$,

$$\hat{\boldsymbol{R}}_{\boldsymbol{yy}} = \hat{\boldsymbol{R}}_{\boldsymbol{ss}} + \hat{\boldsymbol{R}}_{\boldsymbol{nn}}. \tag{6.14}$$

The part $\hat{\boldsymbol{R}}_{\boldsymbol{ss}}$ is formed from the estimated channels as

$$\hat{\boldsymbol{R}}_{\boldsymbol{ss}} = \hat{\boldsymbol{B}}\hat{\boldsymbol{B}}^H \tag{6.15}$$

while $\hat{\boldsymbol{R}}_{\boldsymbol{nn}}$ is formed from the residuals as

$$\begin{aligned}\hat{\boldsymbol{R}}_{\boldsymbol{nn}} = \frac{3}{4N}\sum_{t_c=1}^{N} \hat{\boldsymbol{n}}_c(t_c)\hat{\boldsymbol{n}}_c^H(t_c) + \\ \frac{1}{4N}\sum_{t_a=1}^{N} \hat{\boldsymbol{n}}_a(t_a)\hat{\boldsymbol{n}}_a^H(t_a).\end{aligned} \tag{6.16}$$

Above, $\hat{\boldsymbol{n}}_c(t)$ represents the residuals for the current training sequence whereas $\hat{\boldsymbol{n}}_a(t)$ constitutes the residuals for the adjacent training sequence. The time indices t_c and t_a belong to the current and the adjacent training sequences respectively, while the parameter N is the number of symbols of the training sequence.

Note that only one of the sums in equation (6.16) has to be computed for each new training sequence. If an interferer is present during the current training sequence, it is more likely to be present during the current data sequence part in question, than if it was present during the adjacent training sequence. The contributions from the different training sequences are therefore weighted differently. Other weighting factors than the factors 3/4 and 1/4 suggested above can certainly be considered.

For comparison, we use a version of the algorithm that does not take the interference plus noise spectrum of the adjacent time frame into account. Here, $\hat{\boldsymbol{R}}_{\boldsymbol{nn}}$ is simply formed as

$$\hat{\boldsymbol{R}}_{\boldsymbol{nn}} = \frac{1}{N}\sum_{t_c=1}^{N} \hat{\boldsymbol{n}}_c(t_c)\hat{\boldsymbol{n}}_c^H(t_c). \tag{6.17}$$

Simulations

An example scenario is chosen to illustrate the behavior of the algorithms. The desired signal and co-channel interferers present during the training sequence is shown in Figure 6.8. The received desired signal and co-channel

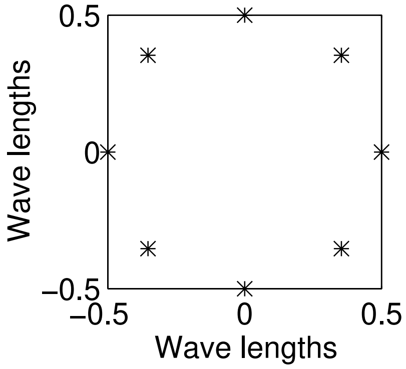


Figure 6.7: Antenna configuration.

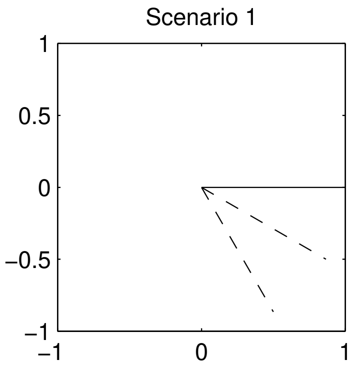


Figure 6.8: Desired signal (solid) and co-channel interferers present during the training sequence (dashed). The angle of incident of the adjacent frame interferer (dotted) is varied between -180 and 180 degrees.

interferers are modeled by filtering binary symbols, ± 1, through the three tap FIR filter $M(q^{-1}) = 0.44i + 1.00q^{-1} - 044iq^{-2}$. This filter is, as described in the previous section, a model of the channel between the transmitted symbols and the received samples. The propagation channel is assumed to be equal to unity. In reality, the true channel would vary depending on the sampling instant and possible intersymbol interference due to multipath propagation. This problem can, for example, be handled with the method described in [56].

The desired signal impinges onto the antenna from 0 degrees and two equal strength co-channel interferers impinge from -30 and -60 degrees respectively. A third co-channel interferer is used in the simulations. This co-channel interferer is thought of as not being present during the training sequence of the desired signal in question. It will either be present or not be present during the desired signals data sequence. It is however always present during the training sequence of the adjacent frame. This co-channel interferer will be referred to as the adjacent frame interferer.

In the simulations, the SNR was 3 dB whereas the SIR was 0 dB, not counting the adjacent frame interferer. This co-channel interferer had a power twice of that of the individual interferers that were present during the current training sequence. All simulations were conducted without fading.

Two performance measures were evaluated for the algorithms. One was the BER for an MLSE after the beamformer, working with a three tap FIR channel. Another performance measure was the signal to interference and noise ratio (SINR) after the beamformer.

The algorithms were tested for two different cases:

- In the first case, the adjacent frame interferer was present during the whole current half of the data sequence part, thus causing interference.
- In the second case, the adjacent frame interferer was *not* present during the current data sequence part.

This second case has to be considered since it is undesired that the performance of the algorithm using the adjacent interferer plus noise spectrum should degraded too much in this case.

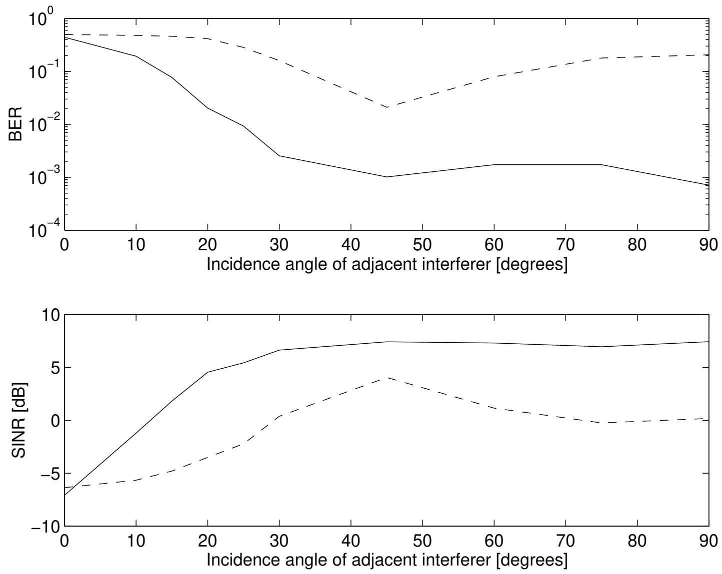


Figure 6.9: Adjacent frame interferer present during the data sequence. BER for the MLSE and SINR after the beamformer. The algorithm using adjacent frame interference plus noise (solid) and the algorithm not using adjacent frame interference plus noise (dashed).

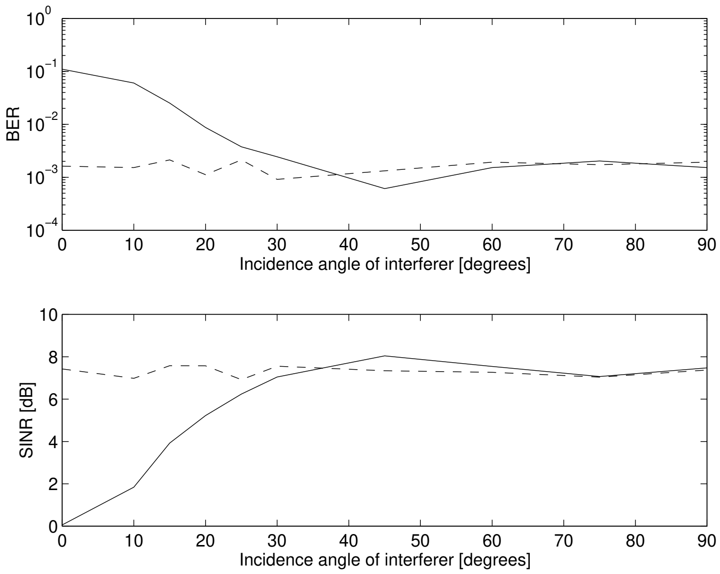


Figure 6.10: Adjacent frame interferer *not* present during the data sequence. BER after the MLSE and SINR after the beamformer for the algorithm using adjacent frame interference plus noise (solid), and the algorithm *not* using the adjacent frame interference plus noise (dashed)

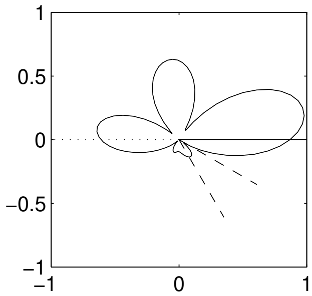


Figure 6.11: Antenna gain when the adjacent interferer is not accounted for. Solid line - Desired signal, Dashed lines - Co-channel interferers and Dotted line - Adjacent frame interferer.

In Figures 6.9 and 6.10 and we can see the resulting BER and SINR performance as a function of the angle of incidence for the adjacent frame interferer. In Figure 6.9 we can see that the SINR as well as the BER is considerably improved as a result of utilizing the adjacent frame interferer plus noise spectrum when the adjacent frame interferer *is* present during the current data sequence being equalized. On the other hand, when the adjacent frame interferer *is not* present during the data sequence being equalized, we can in Figure 6.10 se that the method utilizing the adjacent frame noise plus interferer spectrum has a performance degradation when the adjacent frame interferer impinges from almost the same direction (0 degrees) as the desired signal. This is because this algorithm unnecessarily decreases the gain in the 0 degree direction.

Examples of the antenna gain patterns for the two algorithms can be seen in Figures 6.11 and 6.12, respectively. Figure 6.11 shows a bad case for the algorithm that does not take the adjacent interferer into account. The adjacent frame interferer is amplified by the antenna array, causing a low SINR. In Figure 6.12 the effect of taking the adjacent noise plus interferer spectrum into account is clearly demonstrated. The adjacent frame interferer (dotted line) is now nulled out, resulting in a much better SINR.

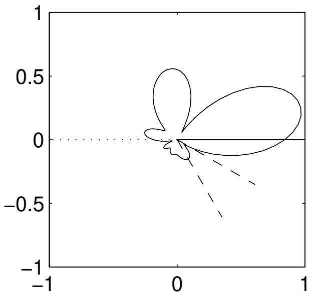


Figure 6.12: Antenna gain when the adjacent interferer is accounted for. Solid line - Desired signal, Dashed lines - Co-channel interferers and Dotted line - Adjacent frame interferer.

Results and Conclusion

As can be seen in Figure 6.9, for the scenario considered, using the interference plus noise spectrum from the adjacent time frame is advantageous if the interferer actually is present during the data sequence part of interest.

If, however, the interferer is *not present* during the data sequence part of interest, as in Figure 6.10, then the proposed algorithm suffers from a performance degradation when the accounted for, but not present, interferer and the desired signal arrives from almost the same angle. The reason for this is that the gain in the direction of the *desired* signal will then be somewhat reduced. However, if the angle between the impinging directions of the interferer and the desired signal is large, then we see from Figure 6.10 that one does not loose any performance by taking the absent interferer into consideration. We thus have to ask the question if the potential gains are larger than the potential losses. This is not a trivial question to answer as it depends on many factors. However, when using an antenna array with many antenna elements the sector where there is a potential loss will become more narrow since the angular resolution of the array increases. The regions of potential gain, on the other hand, will not necessarily decrease in the same way. It therefore seems likely that, as long as we have enough antennas, the potential gains will outweigh the potential losses. We will return to this question in the next section.

The general idea presented here can be applied to other beamforming schemes as long as they separate between signal and interference plus noise spectra. For instance, one can use beamforming with the maximum SNR method. See for example [10].

The method of using adjacent frames has been outlined in a simple form, which constitutes a starting point for improvements and generalizations. For example, when working on the first half of the data in the frame, the adjacent noise spectrum could be formed using the whole last part of the data sequence of the adjacent frame, not only the training sequence. The previously estimated data symbols would then be used instead of the training sequence. A further generalization in this direction is to combine the method with bootstrap equalization. This is discussed in the following section.

6.2.2 Bootstrap Equalization Utilizing Adjacent Frames

Bootstrapping can to some extent handle the problem of compensating for an interferer that is present during the data part of the frame but not present during the training sequence. As long as the resulting BER from the initial equalization is not too high, bootstrapping can save the situation. However, if the initial BER is too high, then the estimated symbols will contain too many errors. The bootstrapping algorithm will then not be able to improve the BER.

The method of utilizing the noise plus interferer spectrum from the adjacent frames can however be utilized as a conservative initial estimate of the symbols. (It is conservative as it may be too cautious since it compensates for interferers in two frames.) Bootstrapping can now be applied using this initial estimate. Since the initial estimate to some extent compensates for all possible asynchronous interferers, the initial BER will hopefully not be too large for the bootstrap equalization.

Also, by using bootstrapping, some of the drawbacks of utilizing the noise plus interferer spectrum from adjacent frames as exemplified in the previous section can be reduced.

Simulation Scenario

The simulation scenario is the same here as in Section 6.2, apart from a few alterations. Instead of using a beamformer followed by a scalar MLSE, we here use an IS-DFE as presented in Section 3.3.3. The channel was estimated by identifying the coefficients of an FIR model with a least squares method as described in Section 2.2. Three taps was used in the FIR channel model. For the DFE, four taps were used in the feedforward filter and two taps were used in the feedback filter. The signal strengths for the particular simulations presented here was SNR=6 dB and SIR=-6 dB. The asynchronous co-channel interferer had the same strength as one of the two equal strength synchronous interferers. The asynchronous interferer is not included in the signal to interference ratio of -6 dB. No fading was considered. Otherwise the scenario was the same as in the simulation of Section 6.2.

Results and Conclusions

First, consider the case where the interferer that is *present* during the training sequence of the adjacent frame, is also present during the data sequence of the current frame that is to be equalized. In Figure 6.13 we can see that both the method of utilizing the adjacent frame noise and interferer spectrum and the method of combining it with bootstrapping give good performance. The use of using the noise and interferer spectrum estimated from the training sequence of the current frame, does not give good performance. Bootstrapping alone does not help in this case either since the BER from the initial equalization is too high.

Now consider the case where the interferer that is present during the training sequence of the adjacent frame is *not* present during the data sequence of the current frame that is to be equalized. In this case we can expect that the methods utilizing the noise plus interferer spectrum of the adjacent frame will have some degradation in the performance. In Figure 6.14 we can see that this indeed is the case. As in Section 6.2.1, this degradation mainly occurs when the direction of the desired signal and the direction of the interferer in the training sequence of the adjacent frame are close. When bootstrapping is added, this degradation is reduced. Of course, the method which does not utilize the noise plus interferer spectrum of the adjacent

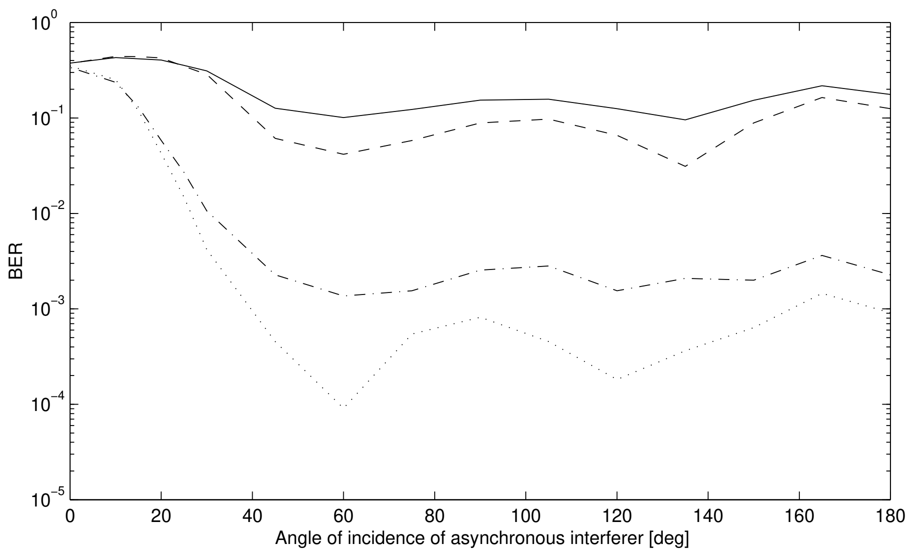


Figure 6.13: BER when the interferer of the adjacent training sequence is also present during the current frame data sequence. Tuning using the training sequence (solid), using training sequence and bootstrapping (dashed), using adjacent frame interference estimate (dash-dotted) and using adjacent frame interference estimate and bootstrapping (dotted).

frames has the best performance. The reason why bootstrapping cannot completely remove the drawback of utilizing the adjacent frame noise plus interference spectrum is that for some realizations, the initial BER will be too high for the bootstrapping algorithm to handle. However, all in all, in this scenario it appears to be a good idea to use the combination of the adjacent frame noise plus interferer spectrum and bootstrapping.

To summarize, in general, the use of the noise plus interference spectrum of adjacent frames will be a tradeoff between suppression of asynchronous interferers and possibly reduced performance due to the use of an erroneous noise plus interference spectrum. By combining the method with bootstrap equalization the negative effect of using a somewhat erroneous noise plus interferer spectrum for the initial equalizer can be reduced. In general the negative effects will not be completely removed. However, the simulations performed in this study suggest that the benefits outweigh the drawbacks.

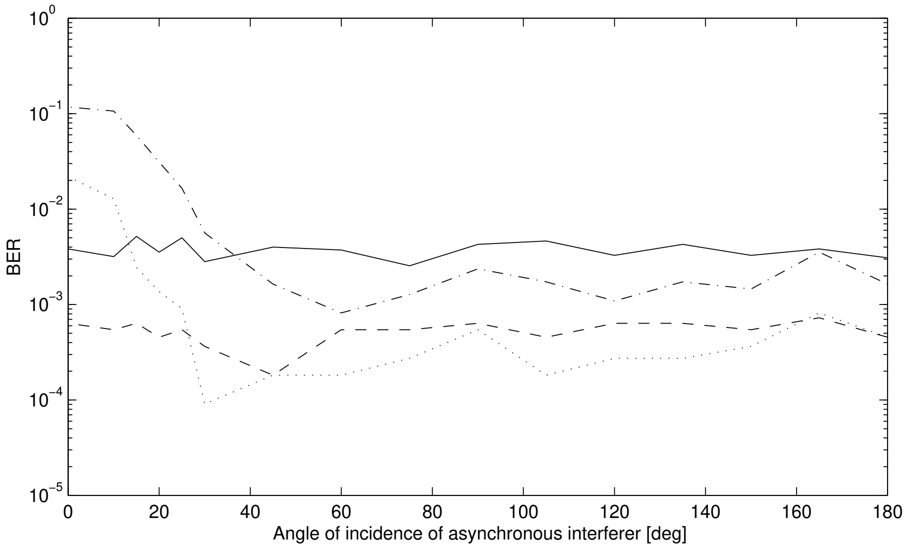


Figure 6.14: BER when the interferer of in the adjacent training sequence is *not* present during the current frame data sequence. Tuning using the training sequence (solid), using training sequence and bootstrapping (dashed), using adjacent frame interference estimate (dash-dotted) and using adjacent frame interference estimate and bootstrapping (dotted).

Chapter 7

Robust Equalization

When perfect estimates of the channel and the noise spectrum are available, we can compute the optimal decision feedback equalizer as in Chapter 3. In reality, however, this is a very unlikely situation and the channel and the noise spectrum will differ from their estimates and they may also vary with time. An equalizer designed based on an estimated channel and noise spectrum may then perform acceptable in some cases but may be inadequate in others. As we will show here, it is possible to use information about the uncertainty and/or the variability of the estimates of the channel and the noise plus interferer spectrum to design an equalizer that is robust against such uncertainty and/or variability. Such an equalizer will be able to accommodate the anticipated channel and noise spectrum estimation errors or time variations, at the expense of some performance degradation in the nominal case. This is attained by minimizing an MSE criterion that is averaged with respect to the uncertainties in the channel and noise models [91, 37, 38].

In Section 7.1 we discuss this averaged design of a scalar decision feedback equalizer that is robust with respect to uncertainties in the channel, in the noise spectrum and also in the quality of decided symbols entering the feedback filter. The robust DFE presented here is general in the sense that it can handle all linear channels and linear noise models [92]. We obtain design equations for a large class of SISO IIR equalizers. DFE's based on nominal models, linear equalizers based on nominal models and robustified DFE's and linear equalizers are special cases within this class.

One of the features of the DFE in Section 7.1 is the possibility to improve the robustness with respect to decision errors. One of the drawbacks with a decision feedback equalizer is that it can suffer from error propagation. An incorrectly estimated symbol that is fed back through the feedback filter can in turn cause the next symbol to be incorrectly estimated, and so on. The DFE can therefore potentially suffer from long error propagation events. Long error propagation events may not be desirable since they may make decoding more difficult and may cause, for instance, a channel tracking algorithm to loose track. In Section 7.2 we investigate the use of robustness against decision errors by studying an example. In this example we can see that by designing the DFE to be robust against decision errors, the lengths of error propagation events can be made much shorter. This is however achieved at the expense of having more numerous error propagation events. It seems that while this method can reduce the average and maximum duration of the error propagation events, it cannot in general substantially reduce the overall BER.

In wireless communication, the channel is often subject to fading, i.e. time variations in the channel. Let us assume a TDMA system transmitting information in frames. If the channel changes significantly over the duration of a TDMA frame, then we are required to track the channel variations and frequently re-design the equalizer, or adapt the equalizer coefficients directly. If, however, the channel variations are smaller but still significant, then it may be possible to design an equalizer that will work well for the range of channels dynamics that may occur during the remaining TDMA frame. In Section 7.3 we use the formalism introduced in Section 7.1 to design an equalizer that is robust with respect to fading in the wireless channel. We investigate the performance of such an equalizer by applying it to a GSM-like TDMA frame at 1800 Mhz. We tune the robust DFE based on an estimate of the channel formed from the training data in the middle of the TDMA frame, and an estimate of the expected channel variations during the frame. For comparison we use a nominal equalizer tuned using only the channel estimate. The robust DFE provides the largest average performance gain over the nominal DFE when the channel variations are large and the noise level is low. In these cases the nominal DFE will be very sensitive and have a degradation in the performance which will make the robust DFE more attractive [62].

In Chapters 3 and 4 we saw that it can be possible to tune space time equalizers indirectly using channel estimates, even though there are strong co-channel interferers present. In the presence of strong co-channel interferers the error in the channel estimate can be very large. However, when these channel estimates are used to design an equalizer, they will often still produce a useful design. In Section 7.4 we investigate why this can be the case. We show that an indirectly tuned space-time DFE, utilizing the estimated spatial spectrum of the noise plus interference is automatically somewhat robustified against co-channel interferers. This is at least true if the number of uncorrelated interfering signals are few enough, such that it is possible to cancel them with spatial-only interferer suppression.

The mechanism for this is that the interferers will cause large spatial model errors in the channel estimates in the directions of the interferers. However, the effect of these interferers on the estimated noise covariance matrix will automatically reduce the spatial gain of the DFE forward filter in precisely these directions. The multi-dimensional filter in a space-time MLSE has many similarities to the feedforward filter of the space-time DFE and is robustified in a similar way. This surprising and advantageous effect of using estimated noise covariance matrices in indirectly designed DFE's and MLSE's will be denoted *self-robustification*.

7.1 Robust Scalar Decision Feedback Equalization

Here design equations are presented for robust and realizable single-input single-output decision feedback equalizers, for IIR channels with colored noise. The design equations are derived based on a probabilistic measure of the model uncertainty: The MMSE, averaged over a whole class of possible true systems, is minimized. A second type of robustification, which affects and reduces the error propagation due to the feedback, is also introduced. The resulting design equations define a large class of equalizers, with DFE's and linear equalizers based on nominal models being special cases.

Design equations which take these two problems into account are presented in Section 7.1.2 below.

7.1.1 Model and Filter Structure

The robustification is based on a stochastic representation of the mismodeling and of decision errors. This seems natural, since estimation of model errors caused by Gaussian noise during the estimation process are of a probabilistic nature.

Let the received, discrete–time, complex baseband signal $y(t)$ be described by a set of models

$$y(t) = \left(\frac{b_0(q^{-1})}{a_0(q^{-1})} + \frac{\Delta b(q^{-1})}{a_1(q^{-1})} \right) d(t-k) + n(t). \tag{7.1}$$

The transmitted symbols $\{d(t)\}$ are assumed to be zero mean and white, with variance $E|d(t)|^2 = \sigma_d^2$.

Similarly, the noise $n(t)$ is described by a set of possible noise models

$$n(t) = \left(\frac{m_0(q^{-1})}{n_0(q^{-1})} + \frac{\Delta m(q^{-1})}{n_1(q^{-1})} \right) v(t). \tag{7.2}$$

where $v(t)$ is zero mean and white, with (uncertain) standard deviation σ_v. In (7.1) and (7.2)

$$b_0(q^{-1})/a_0(q^{-1}) \quad m_0(q^{-1})/n_0(q^{-1}) \tag{7.3}$$

represent time–invariant stable and known *nominal models*, while

$$\Delta b(q^{-1})/a_1(q^{-1}) \quad \Delta m(q^{-1})/n_1(q^{-1}) \tag{7.4}$$

are members of a model error class where the coefficients of the numerator polynomials, e.g.

$$\Delta b(q^{-1}) = \Delta b_0 + \Delta b_1 q^{-1} + \ldots + \Delta b_{\delta b} q^{-\delta b} \tag{7.5}$$

are regarded as (time-independent) stochastic variables, with zero means and known covariance matrices. The stable denominators $a_1(q^{-1})$ and $n_1(q^{-1})$ are assumed to be fixed. Introduce $\bar{E}(\cdot)$ as an average over the coefficients of the stochastic polynomial coefficients. We assume that the coefficients of $\Delta b(q^{-1})$ and $\Delta m(q^{-1})$ have mean value zero, i.e.

$$\bar{E}[\Delta b(q^{-1})] = 0 \quad \bar{E}[\Delta m(q^{-1}] = 0. \tag{7.6}$$

In [91, 38, 37], these representations are shown to be suitable for describing a wide range of model errors. They are related to the stochastic embedding approach introduced in [33, 32].

Consider as an example, an uncertain FIR channel with white noise. It can be expressed as

$$y(t) = \left(b_0(q^{-1}) + \Delta b(q^{-1})\right) d(t-k) + n(t).$$

If $\hat{y}(t) = b_0(q^{-1})d(t-k)$ has been estimated by the least squares method and the order of $b_0(q^{-1})$ is adequate, then the LS covariance matrix can serve directly as an estimate of the covariances[1] $\bar{E}(\Delta b_i \Delta b_j^H)$. We have here introduced a set of possible true channels represented by stochastic variables. The actual (single) time channel will be located at an unknown position. The variance $E[\Delta b_i \Delta b_j]$ can be seen as a rough estimate of the time-dependant distance between the nominal estimated model b_0 and this unknown true system.

Let us introduce a scalar IIR decision feedback equalizer

$$\hat{d}(t-\ell|t) = \frac{s(q^{-1})}{r(q^{-1})}y(t) - \frac{Q(q^{-1})}{p(q^{-1})}\tilde{d}(t-\ell-1) \tag{7.7}$$

where ℓ is the smoothing lag (decision delay) and $\tilde{d}(t)$ is decided data. The denominator polynomials $r(q^{-1})$ and $p(q^{-1})$ are required to be monic (leading unit coefficient) and stable.

In addition to robustifying the design with respect to model uncertainty, we will here also introduce a technique that, to some extent, can alleviate the problem of error propagation in the DFE feedback loop. Errors in the decided data $\tilde{d}(t)$ will be treated as uncertainty, and represented by an *additive white noise* $\kappa(t)$, which is assumed to be uncorrelated[2] with both $d(t-j)$ and $v(t-j)$ for all j. The decided data can thus be expressed as

$$\tilde{d}(t) = d(t) + \kappa(t) \tag{7.8}$$

[1] If the estimates are biased due to an incorrect error model structure, or if the channel differs from that obtained based on the training sequence due to time variations, then the covariance $\bar{E}(\Delta b_i \Delta b_j^H)$ has to be adjusted so that these phenomena are taken into consideration.

[2] This is, of course, a simplification. In reality, the error $\kappa(t)$ is non–stationary since decision errors tend to occur in bursts. There may also exist correlations to past noise samples, in particular to those that caused the error. These nonlinear and time–varying effects are neglected here, to obtain a tuning parameter which is simple to use.

where $\kappa(t)$ has zero mean and variance

$$E|\kappa(t)|^2 = \eta\sigma_d^2. \tag{7.9}$$

The problem of optimizing (7.7) can then be treated with tools for linear quadratic design, since the nonlinear decision element is removed from the signal path to the error $d - \hat{d}$, see Figure 7.1.

The scale factor $\eta \geq 0$ in (7.9) can be used to trade off error propagation against theoretical performance, with $\eta = 0$ representing a belief in error–free decided data whereas a non-zero η represents a situation with decision errors, the larger η the more decision errors are anticipated. The use of a small positive value of η can give a lower bit error rate in cases with severe error propagation compared to a design using $\eta = 0$[3].

7.1.2 Filter Design Equations

The performance of an equalizer of the form (7.7) is now to be optimized with respect to the whole model error class. We thus tune the equalizer to minimize the *averaged MSE criterion*

$$J = \bar{E}E|d(t-\ell) - \hat{d}(t-\ell|t)|^2 \tag{7.10}$$

with respect to the parameters of (7.7) under the constraint that the filters in (7.7) are causal and stable. Here E represents expectation over d, n and κ while $\bar{E}$ is the previously introduced expectation over the model error distribution in (7.1) and (7.2). This type of criterion has been used in connection to other filtering problems, e.g. by Chung and Bélanger [16]. Note that not only the *range* of uncertainties, but also their *likelihood* is taken into account by (7.10); common model deviations will have a greater impact on an estimator design than do very rare "worst cases". Compared to the use of a minimax design, the conservativeness is thus reduced [37, 93]. In communication problems the average performance is in general more important than the performance for the worst case. Due to the presence of coding and interleaving, rare bad situations can be handled, but a high average uncoded bit error rate will degrade the total system performance.

[3]For a related suggestion, using a linear combination of a zero forcing linear equalizer and a zero forcing DFE, see [39].

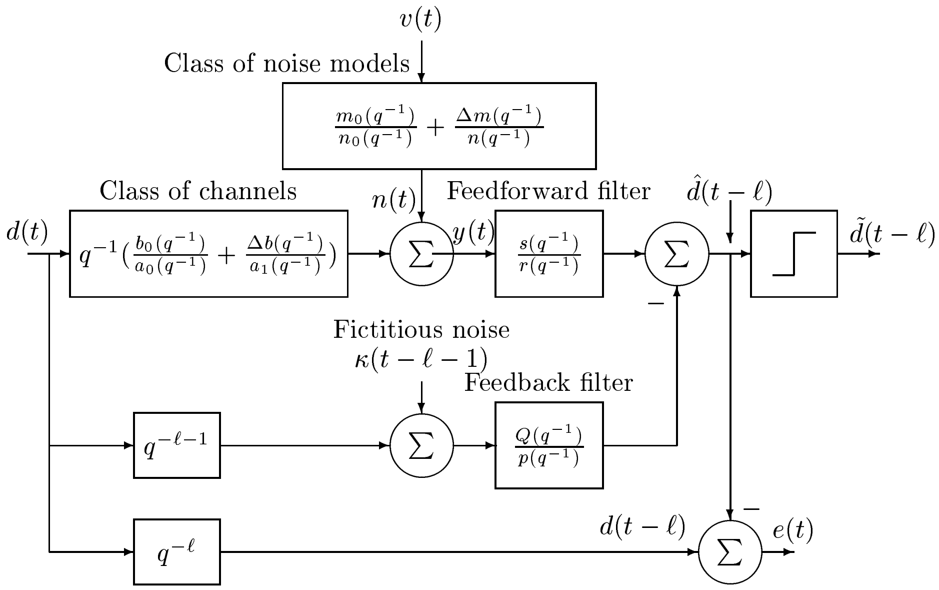


Figure 7.1: A general robust decision feedback equalizer. The feedforward and the feedback filters are optimized by minimizing the power of the error $e(t)$ averaged over the class of possible channels and possible noise disturbance models. Erroneous decisions are modeled by adding a fictitious white noise to previous decided symbols in the design.

In Appendix 7.A.1, design equations are derived for minimizing the criterion (7.10) with respect to the filter coefficients of the DFE (7.7), for an ensemble of systems (7.1), (7.2), assuming previous decisions to be described by (7.8).

As in previous chapters, for polynomials

$$p(q^{-1}) = p_0 + p_1 q^{-1} + \ldots + p_{np} q^{-np} \tag{7.11}$$

we use the convention that

$$p^H(q) \triangleq p_0^H + p_1^H q + \ldots + p_{np}^H q^{np}. \tag{7.12}$$

Introduce polynomials $h(q^{-1})$, $a(q^{-1})$ and $n(q^{-1})$ as

$$b(q^{-1}) \triangleq b_0(q^{-1})a_1(q^{-1}) + \Delta b(q^{-1})a_0(q^{-1}) \tag{7.13}$$

$$h(q^{-1}) \triangleq b_0(q^{-1})a_1(q^{-1})n_0(q^{-1})n_1(q^{-1}) \tag{7.14}$$

$$a(q^{-1}) \triangleq a_0(q^{-1})a_1(q^{-1}) \tag{7.15}$$

$$n(q^{-1}) \triangleq n_0(q^{-1})n_1(q^{-1}) \tag{7.16}$$

$$m(q^{-1}) \triangleq m_0(q^{-1})n_1(q^{-1}) + \Delta m(q^{-1})n_0(q^{-1}) \tag{7.17}$$

and define the double–sided polynomials $\tilde{b}\tilde{b}^H(q,q^{-1})$ and $\tilde{m}\tilde{m}^H(q,q^{-1})$ by

$$\begin{aligned}\tilde{b}\tilde{b}^H(q,q^{-1}) &\triangleq \bar{E}[b(q^{-1})b^H(q)] = b_0(q^{-1})b_0^H(q)a_1(q^{-1})a_1^H(q) \\ &\quad +\bar{E}[\Delta b(q^{-1})\Delta b(q)^H]a_0(q^{-1})a_0^H(q) \end{aligned}\tag{7.18}$$

$$\begin{aligned}\tilde{m}\tilde{m}^H(q,q^{-1}) &\triangleq \bar{E}[m(q^{-1})m^H(q)] = m_0(q^{-1})m_0^H(q)n_1(q^{-1})n_1^H(q) \\ &\quad +\bar{E}[\Delta m(q^{-1})\Delta m^H(q)]n_0(q^{-1})n_0^H(q). \end{aligned}\tag{7.19}$$

Finally, define the variance ratio

$$\rho \triangleq \bar{E}(\sigma_v^2)/\sigma_d^2.$$

Now, let the scalar γ and the stable and monic polynomial $\beta(q^{-1})$ be the solution to the *averaged spectral factorization*

$$\begin{aligned}\gamma\beta(q^{-1})\beta^H(q) &= n(q^{-1})n^H(q)a_0(q^{-1})a_0^H(q)\bar{E}[\Delta b(q^{-1})\Delta b^H(q)] \\ &\quad +\eta n(q^{-1})n^H(q)\tilde{b}\tilde{b}^H(q,q^{-1}) \\ &\quad +(1+\eta)\rho a(q^{-1})a^H(q)\tilde{m}\tilde{m}^H(q,q^{-1}). \end{aligned}\tag{7.20}$$

Let $Q(q^{-1}), s_1(q^{-1}), l_1^H(q)$ and $l_2^H(q)$ be the (unique) solution to the *coupled polynomial equations*

$$\begin{aligned}\beta(q^{-1}) + q^{-1}(1+\eta)Q(q^{-1}) &= q^{\ell-k}h(q^{-1})s_1(q^{-1}) \\ &\quad +\beta(q^{-1})l_1^H(q) \qquad (7.21)\\ -q^{-\ell+k}\eta h^H(q) + q(1+\eta)l_2(q) &= -\gamma\beta^H(q)s_1(q^{-1}) + \\ &\quad q^{-\ell+k}h^H(q)l_1^H(q). \qquad (7.22)\end{aligned}$$

The optimal filters in the equalizer (7.7), which minimize (7.10), are then given by $Q(q^{-1})$ from (7.21) and by

$$\begin{aligned}s(q^{-1}) &= s_1(q^{-1})n(q^{-1})a(q^{-1})\\ r(q^{-1}) &= \beta(q^{-1})\\ p(q^{-1}) &= \beta(q^{-1}). \qquad (7.23)\end{aligned}$$

This result can be derived by adding a variation $\nu(t) = \tau_1(q^{-1})y(t) + \tau_2(q^{-1})\tilde{d}(t-\ell-1)$, with $\tau_1(q^{-1})$ and $\tau_2(q^{-1})$ being arbitrary, but rational, causal and stable transfer function, to the estimate $\hat{d}(t-\ell|t)$. Optimality corresponds to orthogonality of the mean estimation error with respect to this variation. The equations (7.21) and (7.22) arise from requiring orthogonality with respect to the two terms $\tau_1(q^{-1})y(t)$ and $\tau_2(q^{-1})\tilde{d}(t)$, separately. A detailed derivation can be found in Appendix 7.A.1, where a generalization to correlated symbol sequences $d(t)$ is also discussed.

The coupled equations (7.21) and (7.22) can be solved in precisely the same way as the corresponding equations in Appendix 3.A.1: Convert them to systems of linear equations in the coefficients. Then, a new system of linear equations is created by combining all equations with known left-hand sides. This system is solved with respect to the coefficients of $s_1(q^{-1})$ and $l_1^H(q)$. Subsequently, $Q(q^{-1})$ is obtained from (7.21). The polynomial degrees, which are obtained by requiring the powers of q and q^{-1} to be matched on both sides of (7.21) and (7.22), are given by

$$\deg s_1(q^{-1}) = \deg l_1(q^{-1}) = \ell - k \qquad (7.24)$$

$$\deg Q(q^{-1}) = \deg l_2(q^{-1}) = \max\{\deg h(q^{-1}), \deg\beta(q^{-1})\} - 1. \qquad (7.25)$$

Solving the spectral factorization (7.20) is straightforward. With a given right–hand side, it is just an ordinary polynomial (FIR) spectral factorization, for which there exist efficient iterative algorithms [47, 40, 36]. One alternative is to factor the right-hand side and form the polynomial $\beta(q^{-1})$ from all factors with zeros inside $|z| < 1$.

The averaged factors in (7.20) can be evaluated as follows. For a stochastic polynomial

$$\Delta p(q^{-1}) \tag{7.26}$$

of degree δp with zero mean, let the Hermitian parameter covariance matrix be defined as

$$\boldsymbol{P}_{\Delta P} = \begin{bmatrix} \bar{E}|\Delta p_0|^2 & \dots & \bar{E}(\Delta p_0 \Delta p_{\delta p}^H) \\ \vdots & \ddots & \vdots \\ \bar{E}(\Delta p_{\delta p} \Delta p_0^H) & \dots & \bar{E}|\Delta p_{\delta p}|^2 \end{bmatrix} . \tag{7.27}$$

Denote the sum of the diagonal elements g_0, the sum of the elements in the i'th super–diagonal g_i, the sum of elements in the i'th subdiagonal g_{-i}. Note that $g_{-i} = g_i^H$. Then it becomes evident, by direct multiplication of $\Delta p(q^{-1})\Delta p^H(q)$, and taking expectations, that

$$\bar{E}[\Delta p(q^{-1})\Delta p^H(q)] = g_{dp}^H q^{-dp} + \dots + g_1^H q^{-1} + g_0 + g_1 q + \dots + g_{dp} q^{dp}. \tag{7.28}$$

Above, $dp \leq \delta p$, with $dp = 0$ for uncorrelated coefficients.

Note that, apart from the nominal model and the variance ratios ρ and η, only second order moments of the model error distributions need to be known. If this information is not available, then $a_1(q^{-1})$, $n_1(q^{-1})$, ρ, η and the covariance matrices of $\Delta b(q^{-1})$ and $\Delta m(q^{-1})$ can still be used as “robustness tuning knobs”. In the case of no model uncertainty, we then set $\Delta b(q^{-1}) = \Delta m(q^{-1}) = 0$ and $a_1(q^{-1}) = n_1(q^{-1}) = 1$ in (7.18)–(7.23). An increase of the covariance matrix elements of $\Delta b(q^{-1})$ or $\Delta m(q^{-1})$ will result in a more cautious feedback and feedforward filters, with lower gains and lower and broader spectral peaks.

7.1.3 The Class of Equalizers

The equations (7.7)–(7.23) define a class of robust equalizers, with nominally designed linear equalizers and DFE's as special cases:

- If $\eta = 0$ (perfect previous decisions assumed), and with no model uncertainty, then the *IIR DFE* discussed in [90] is obtained[4]. In this case only, a solution of the spectral factorization (7.20) is *not* required. We directly obtain $\beta(q^{-1}) = a_0(q^{-1})m_0(q^{-1})$.

- If $\eta \to \infty$ (decided data are very unreliable), $\| Q(q^{-1}) \| \to 0$. Then, (7.20) and (7.22) reduce to the design equations for a *robust linear equalizer* $s(q^{-1})/r(q^{-1})$, derived in [91]. (Divide (7.20)–(7.22) by η and set $r \stackrel{\Delta}{=} \gamma/\eta$, which is finite.)

- When $\eta \to \infty$ and no model uncertainty is assumed, we obtain the ordinary linear recursive equalizer, discussed in e.g. [2],[90].

The special case of the design equations for a robust DFE assuming perfect decisions ($\eta = 0$), applies to an uncertain and slowly time-varying scalar FIR channel will be discussed in more detail in Section 7.3. The generalization to a MISO DFE design for (time-invariant) uncertain FIR vector channels is derived and discussed in Section 7.4.

The properties of a large number of linear equalizers ($\eta \to \infty$), based on randomly selected 5–tap FIR channel models, are summarized in Section 4.5 of [91]. In about 1/5 of those cases, the nominal channels had pronounced nulls. Equalizers designed without taking any model uncertainty into account then had pronounced spectral peaks. Their performance was very sensitive to model errors. Robust design eliminated the sensitivity, at the price of only a slight increase of the MSE obtained when the model is correct.

[4] With $\Delta b(q^{-1}) = \Delta m(q^{-1}) = 0, a_1(q^{-1}) = n_1(q^{-1}) = 1$, we get $\tilde{b}\tilde{b}^H(q,q^{-1}) = b_0(q^{-1})b_0^H(q)$, $\tilde{m}\tilde{m}^H(q,q^{-1}) = m_0(q^{-1})m_0^H(q)$, $\gamma = \rho$, $\beta(q^{-1}) = a_0(q^{-1})m_0(q^{-1})$ and $h(q^{-1}) = b_0(q^{-1})n_0(q^{-1})$. This case corresponds to the solution in Section 3.2.1 with only one sensor ($M = 1$).

7.2 Robustness Against Decision Errors

The utility of the parameter η is here investigated in an example. We investigate if a proper use of this parameter may reduce the effects of error propagation.

Based on the model, without uncertainty,

$$y(t) = \frac{b_0(q^{-1})}{a_0(q^{-1})}d(t) + n(t) \quad = \quad \frac{(1+0.95q^{-1})}{(1-0.70q^{-1})^3}d(t) + n(t)$$

we designed DFE's for a SNR of 28dB which corresponds to $\rho = 0.553$. The smoothing lag was $\ell = 5$ and the data sequence was binary PAM ($d(t) = \pm 1$). DFE's were designed for different values of the parameter η. Here we have $h(q^{-1}) = b_0(q^{-1})$, $\tilde{b}\tilde{b}^H(q, q^{-1}) = b_0(q^{-1})b_0^H(q)$, $\tilde{m}\tilde{m}^H(q, q^{-1}) = 1$, and thus (7.20) reduces to

$$\gamma\beta(q^{-1})\beta^H(q) = \eta b_0(q^{-1})b_0^H(q) + (1+\eta)\rho a_0(q^{-1})a_0^H(q).$$

The bit error rate (BER) was estimated from simulations using 500000 symbols for each design. The diagram below shows the BER as a function of η. Also shown is the BER without error propagation, i.e. the result we would obtain if correct past decisions could be substituted for $\tilde{d}(t-\ell-1)$ in (7.7). In this difficult problem, a pure DFE has severe problems with error bursts, while a linear equalizer (LE) gives a large BER. However, the BER can be reduced by a factor of 3.5 by using $\eta = 0.001$ instead of $\eta = 0$ (pure DFE).

An error propagation event is in this example defined as the length of sequences of $\tilde{d}(t)$ which contain < 7 consecutive correct decisions. It is evident from the table below, that the use of a small $\eta \geq 0$ reduces the length of error propagation events substantially. If a coding scheme is used which is sensitive to long consecutive sequences of errors, this property is valuable in itself.

It should be noted, however, that although the number of long error propagation events is reduced, the number of short error propagation events is increased. Therefore it is not possible to say if there for arbitrary selected channels in general exists an $\eta \neq 0$ which gives a lower BER than what is obtained by using $\eta = 0$.

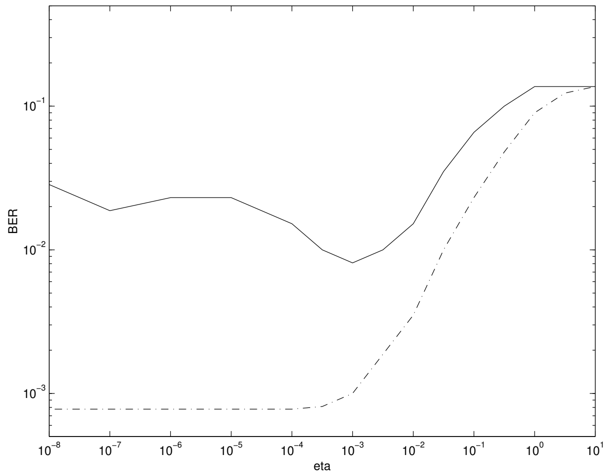


Figure 7.2: BER as function of assumed relative power η of the decision error noise. BER with error propagation (solid) and without error propagation (dashed).

	Number of error propagation events				
Length:	1-2	3-20	21-50	51-100	>100
$\eta = 0$:	328	197	75	66	133
$\eta = 10^{-3}$:	323	232	77	24	–
$\eta = 0.1$:	6100	4469	171	–	–

Table 7.1: Number of error propagation events of different lengths for different values of the relative power η of the noise assumed in the feedback loop.

7.3 Robustness Against Fading

One of the causes of uncertainty in a channel are time-variations that we are not able to track, or choose not to track. We can however design an equalizer which optimizes the performance based on statistical assumptions on the anticipated time variations in the channel.

We will here discuss the design of such a robust equalizer. As an example, we will consider a GSM-type case. In GSM, the channel varies slowly over the duration of a frame. It is therefore feasible to use a fixed equalizer, tuned to the training data, for the equalization of the whole frame. However, when the mobile travels at high speeds, the channel variations can have a non-negligible impact on the performance of the equalizer. In these cases it can be advantageous to take the time variations into consideration when tuning the equalizer.

We will here utilize the design equations for a robust decision feedback equalizer, derived in the previous section, for this purpose. It will be demonstrated how the statistical Rayleigh fading model can be used to adjust models of the channel uncertainty due to a moving transmitter or receiver. An additional contribution to the model uncertainty is caused by the channel estimation error due to noise and interference. We will here restrict the presentation to the special case of FIR channels with uncertain channel coefficients and white noise.

7.3.1 Model and Filter Structure

We will consider the following received, discrete-time, complex baseband signal

$$y(t) = \left(b_0(q^{-1}) + \Delta b(q^{-1})\right) d(t) + n(t) \tag{7.29}$$

where the transmitted symbols $\{d(t)\}$ are assumed to be zero mean and white, with $E|d(t)|^2 = \sigma_d^2$. The noise $n(t)$ is assumed to be zero mean, with variance $E|n(t)|^2 = \sigma_n^2$.

The nominal model of the transmission channel is described by an FIR filter

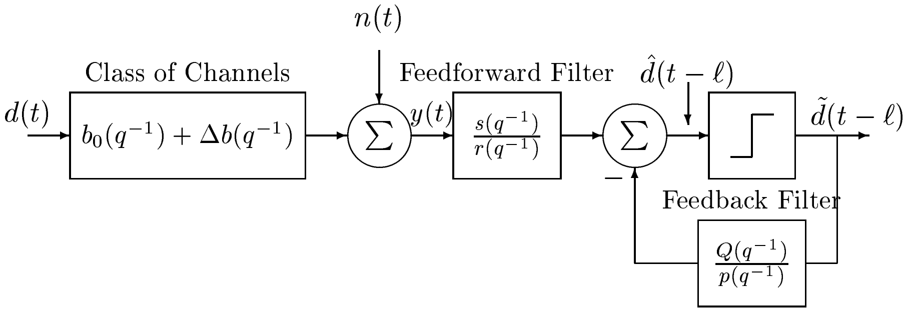


Figure 7.3: Channel with error model and IIR decision feedback equalizer. The coefficients of the polynomial $\Delta b(q^{-1})$ are stochastic variables.

with complex coefficients:

$$b_0(q^{-1}) = b_0 + b_1 q^{-1} + ... + b_{nb} q^{-nb}. \tag{7.30}$$

The set of possible model errors is represented by the "error model"

$$\Delta b(q^{-1}) = \Delta b_0 + \Delta b_1 q^{-1} + ... + \Delta b_{nb} q^{-nb} \tag{7.31}$$

As before, the coefficients Δb_i are stochastic variables with zero mean and known correlations $\bar{E}[\Delta b_i \Delta b_j^H]$. This channel model is a special case of the more general uncertain channel models in Section 7.1, using $k = 0$ and $a(q^{-1}) = 1$.

7.3.2 Filter Design Equations

The robust DFE here is a special case of the general robust DFE presented in Section 7.1. Thus, we assume the general IIR structure

$$\hat{d}(t-\ell|t) = \frac{s(q^{-1})}{r(q^{-1})} y(t) - \frac{Q(q^{-1})}{p(q^{-1})} \tilde{d}(t-\ell-1) \tag{7.32}$$

of the DFE.

The robust DFE that minimizes the averaged MSE criterion

$$J = \bar{E}E|d(t-\ell) - \hat{d}(t-\ell|t)|^2 \tag{7.33}$$

can be computed as follows.

Let us here define

$$\rho \triangleq \sigma_n^2/\sigma_d^2.$$

Let the scalar γ and the stable and monic polynomial $\beta(q^{-1})$ be the solution to the *averaged spectral factorization*

$$\gamma\beta(q^{-1})\beta^H(q) = \bar{E}[\Delta b(q^{-1})\Delta b^H(q)] + \rho \tag{7.34}$$

which is a special case of (7.20) for $a(q^{-1}) = n(q^{-1}) = 1$, $\tilde{m}\tilde{m}^H(q, q^{-1}) = 1$ and $\eta = 0$.

Let $\{Q(q^{-1}), s(q^{-1}), l_1^H(q), l_2^H(q)\}$ be the (unique) solution to the corresponding special cases of (7.21) and (7.21) which constitute the *coupled polynomial equations*

$$\beta(q^{-1}) + q^{-1}Q(q^{-1}) = q^\ell b_0(q^{-1})s(q^{-1}) + \beta(q^{-1})l_1^H(q) \tag{7.35}$$

$$ql_2^H(q) = -\gamma\beta^H(q)s(q^{-1}) + q^{-\ell}b_0^H(q)l_1^H(q) \tag{7.36}$$

with polynomial degrees

$$\deg s = \deg l_1(q^{-1}) = \ell$$

$$\deg Q(q^{-1}) = \deg l_2(q^{-1}) = \max\{\deg b_0(q^{-1}), \deg \beta(q^{-1})\} - 1.$$

The equalizer (7.32), which minimizes (7.33), is now given by $Q(q^{-1})$ and $s(q^{-1})$ from (7.35) and (7.36) and by using the averaged spectral factors as denominator polynomials

$$r(q^{-1}) = p(q^{-1}) = \beta(q^{-1}) \tag{7.37}$$

Along the lines presented for the DFE's in Appendix 3.A.1 and 3.A.2, the coupled equations (7.35) and (7.36) can be solved by converting them to a system of linear equations in the coefficients of s and l_1, respectively.

Let us define

$$\boldsymbol{B} \triangleq \begin{pmatrix} b_0 & 0 & ... & 0 \\ b_1 & b_0 & ... & 0 \\ ... & ... & ... & ... \\ b_\ell & b_{\ell-1} & ... & b_0 \end{pmatrix} \tag{7.38}$$

$$\boldsymbol{\beta} \triangleq \begin{pmatrix} 1 & 0 & ... & 0 \\ \beta_1 & 1 & ... & 0 \\ ... & ... & ... & ... \\ \beta_\ell & \beta_{\ell-1} & ... & 1 \end{pmatrix} \tag{7.39}$$

$$\boldsymbol{S}^T \triangleq \begin{pmatrix} s_0 & s_1 & ... & s_\ell & l_{1,\ell}^H & l_{1,\ell-1}^H & ... & l_{1,0}^H \end{pmatrix} \tag{7.40}$$

$$\boldsymbol{C}^T \triangleq \begin{pmatrix} 0 & 0 & ... & 1 & 0 & 0 & ... & 0 \end{pmatrix} \tag{7.41}$$

where the "1" in $\boldsymbol{C}$ appears in element nr $\ell + 1$. The channel coefficients in the matrix $\boldsymbol{B}$ are here the coefficients of the "nominal" channel model $b_0(q^{-1})$ which here will be the same as the estimated channel. The coefficients in the matrix $\boldsymbol{\beta}$ are the coefficients in the averaged spectral factor $\beta(q^{-1})$ from equation (7.34) with the right hand side evaluated as indicated by (7.27) and (7.28). It can then be shown from (7.35) and (7.36), along the same lines as in Appendix 3.A.1 and 3.A.2, that

$$\boldsymbol{S} = \begin{pmatrix} \boldsymbol{B} & \boldsymbol{\beta} \\ \gamma * \boldsymbol{\beta}^H & -\boldsymbol{B}^H \end{pmatrix}^{-1} \boldsymbol{C}. \tag{7.42}$$

From $\boldsymbol{S}$ we extract $s(q^{-1})$ (and $l_1(q^{-1})$) and from equation (7.35) $Q(q^{-1})$ can be computed as

$$Q(q^{-1}) = q^{\ell+1} b_0(q^{-1}) s(q^{-1}) + q\beta(q^{-1})(l_1^H(q) - 1). \tag{7.43}$$

As in Section 7.1.2, the averaged term, $\bar{E}[\Delta b(q^{-1})\Delta b^H(q)]$, in (7.34) can be evaluated as follows. For a stochastic polynomial $\Delta b(q^{-1})$, of degree nb, let the Hermitian parameter covariance matrix be

$$\boldsymbol{P}_{\Delta b} = \begin{bmatrix} \bar{E}|\Delta b_0|^2 & \dots & \bar{E}(\Delta b_0 \Delta b_{nb}^H) \\ \vdots & \ddots & \vdots \\ \bar{E}(\Delta b_{nb} \Delta b_0^H) & \dots & \bar{E}|\Delta b_{nb}|^2 \end{bmatrix}. \tag{7.44}$$

Denote the sum of the diagonal elements g_0, the sum of the elements in the i'th super–diagonal g_i, the sum of elements in the i'th subdiagonal g_{-i}. Note that $g_{-i} = g_i^H$. Then it becomes evident, by direct multiplication of $\Delta b(q^{-1})\Delta b^H(q)$, and taking expectations, that

$$\bar{E}[\Delta b(q^{-1})\Delta b^H(q)] = g_{db}^H q^{-db} + \ldots + g_1^H q^{-1} + g_0 + g_1 q + \ldots + g_{db} q^{db}. \tag{7.45}$$

Above, $db \leq nb$, with $db = 0$ for uncorrelated coefficients.

If the coefficients in $\Delta b(q^{-1})$ are uncorrelated, then $\bar{E}[\Delta b(q^{-1})\Delta b^H(q)]$ is a constant (not a polynomial) and from equation (7.34) we see that the robust equalizer is achieved by adding the sum of the variances of the coefficients of $\Delta b(q^{-1})$ to the variance ratio ρ. Only in this case will the MMSE-optimal DFE for an FIR channel have FIR feedforward and feedback filters, since (7.34) reduces to $\beta(q^{-1}) = 1$ while $\gamma = \sum \bar{E}|\Delta b_i| + \rho$. Thus, if the uncertainties of the channel coefficients are uncorrelated, the robust equalizer is obtained just by increasing the noise power for which the equalizer is designed. If the uncertainties or model errors in different coefficients are correlated, then the effect of the uncertainty on an MSE-optimal DFE design is equivalent to the effect of a specific colored interference, $n(t) = \beta(q^{-1})$ with $E[|e(t)|^2] = \gamma$. A more general discussion of this interpretation of the effect of model errors on averaged MSE filter designs as equivalent noise can be found in Section 4.8 of [37].

7.3.3 Example

Consider a GSM-channel with three independently Rayleigh fading received signal paths having average relative powers 0 dB, -1.8 dB and -4.8 dB respectively. The three rays are assumed to arrive so that they are separated by one symbol interval. The carrier frequency used is 1800 MHz. The GSM-system uses a partial response modulation stretching over 3-4 symbol intervals, cf Appendix 2.A.1. This results in an overall channel with a total of 5-6 taps.

Fading due to movement of the mobile causes the channel to be slightly time varying over the duration of a burst. An example of such a slow variation is depicted in Figure 7.4.

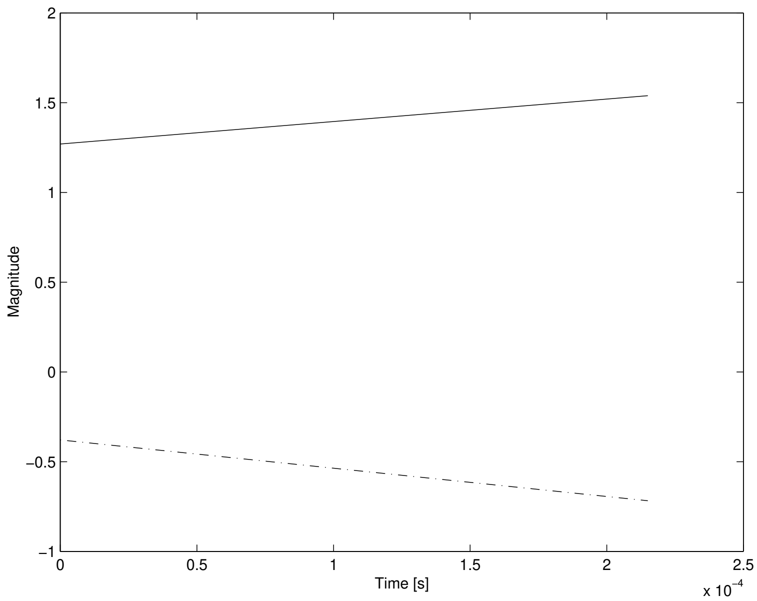


Figure 7.4: Example of the variation of a channel tap at carrier frequency 1800 MHz and a mobile speed of 200 km/h (for example a high-speed train). The time span is 58 symbols of with a symbol period of 3.7μs. Solid line: real part, dashed line: imaginary part.

This time variation can be taken into account when designing a robust DFE. The channel estimated during the training sequence is used as the nominal model $b_0(q^{-1})$. The time variation is viewed as one part of the stochastic uncertainty $\Delta b(q^{-1})$. Another part of the $\Delta b(q^{-1})$ is the uncertainty due to noise in the estimate of the channel, based on the training sequence.

In order to design the equalizer, an estimate of the correlation matrix $\boldsymbol{P}_{\Delta b}$ is required. As indicated above, $\boldsymbol{P}_{\Delta b}$ is here assumed to consists of two parts

$$\boldsymbol{P}_{\Delta b} = \boldsymbol{P}_{\text{Time variation}} + \boldsymbol{P}_{\text{Estimation uncertainty}}. \tag{7.46}$$

The covariance matrix for the uncertainties due to the time variation can be estimated in the following way. We can for a given realization of the time varying channel, define a matrix, $\boldsymbol{P}_{\text{Realization}}$, which quantifies the variability of the channel taps relative to their values at the center of a given time frame at time $T_f/2$, where T_f is the total length of the frame that is to be equalized. This matrix can be expressed as

$$\boldsymbol{P}_{\text{Realization}} = \frac{1}{T_f}\int_{t=0}^{T_f}\left(\boldsymbol{b}(t) - \boldsymbol{b}(\frac{T_f}{2})\right)\left(\boldsymbol{b}(t) - \boldsymbol{b}(\frac{T_f}{2})\right)^H dt \tag{7.47}$$

where $\boldsymbol{b}(t)$ is the vector representation of the coefficients in the polynomial $b(q^{-1})$ at time t,

$$\boldsymbol{b}(t) = \begin{bmatrix} b_0(t) \\ b_1(t) \\ \vdots \\ b_{nb}(t) \end{bmatrix}. \tag{7.48}$$

We have here, as in GSM, assumed that the training sequence is in the middle of the frame. The expression (7.47) would be the natural choice of covariance of the uncertainties $\Delta b(q^{-1})$ if we would design an equalizer to have a robust performance over *this specific* realization of the time varying channel. We are, however, not interested in robustness over a single realization of the time variation but rather over a collection of time varying channels in a Rayleigh fading scenario. We thus take the expectation of $\boldsymbol{P}_{\text{Realization}}$ with respect to the different Rayleigh fading realizations with Rayleigh fading statistics [126], giving

$$\begin{aligned}\boldsymbol{P}_{\text{Time variation}} &= E[\boldsymbol{P}_{\text{Realization}}] \\ &= E\left[\frac{1}{T_f}\int_{t=0}^{T_f}\left(\boldsymbol{b}(t) - \boldsymbol{b}(\frac{T_f}{2})\right)\left(\boldsymbol{b}(t) - \boldsymbol{b}(\frac{T_f}{2})\right)^H\right].\end{aligned} \tag{7.49}$$

To demonstrate the principle for estimating (7.49) on line in a simple case, let us now assume that the fading of the different taps are uncorrelated[5]. The autocorrelation function for a single Rayleigh fading tap is given by

$$E[b_i(t)b_i^*(t-\tau)] = E[b_i(t)b_i^*(t)]J_0(2\pi f_c \frac{v}{c_0}\tau) \tag{7.50}$$

where f_c is the carrier frequency, c_0 is the speed of light and v is the speed of the mobile, which can be estimated [54]. Using (7.50) in (7.49) we can then derive an expression for the correlation matrix;

$$\begin{aligned}
&\boldsymbol{P}_{\text{Time variation}} = \\
&= E\left[\frac{1}{T_f}\int_0^{T_f}\left(\boldsymbol{b}(t)\boldsymbol{b}^H(t) + \boldsymbol{b}(\frac{T_f}{2})\boldsymbol{b}^H(\frac{T_f}{2}) - \boldsymbol{b}(t)\boldsymbol{b}^H(\frac{T_f}{2}) - \boldsymbol{b}(\frac{T_f}{2})\boldsymbol{b}^H(t)\right)\right] dt \\
&\qquad = \frac{1}{T_f}\int_0^{T_f}\left(E[\boldsymbol{b}(t)\boldsymbol{b}^H(t)] + E[\boldsymbol{b}(\frac{T_f}{2})\boldsymbol{b}^H(\frac{T_f}{2})]\right) dt \\
&\qquad - \frac{1}{T_f}\int_0^{T_f}\left(E[\boldsymbol{b}(t)\boldsymbol{b}^H(\frac{T_f}{2})] + E[\boldsymbol{b}(\frac{T_f}{2})\boldsymbol{b}^H(t)]\right) dt \\
&\qquad = 2\boldsymbol{P}_b\left(1 - \frac{2}{T_f}\int_0^{T_f/2} J_0(2\pi f_c \frac{v}{c_0}\tau)d\tau\right)
\end{aligned} \tag{7.51}$$

where $\boldsymbol{P}_b = E[\boldsymbol{b}(t)\boldsymbol{b}^H(t)] = E[\boldsymbol{b}(\frac{T_f}{2})\boldsymbol{b}^H(\frac{T_f}{2})]$.

The correlation matrix for the channel coefficients of the channel $\boldsymbol{P}_b$ can be estimated on-line by averaging over a number of bursts, and the speed of the mobile, v, can be estimated separately. We thus have a way to estimate $\boldsymbol{P}_{\text{Time variation}}$. When accounting for the correlation of the taps due to the pulse shaping in the modulation, we have to include this dependency into the above equation.

In this example the channel was determined by estimating the coefficients in an FIR model with a least squares algorithm as in Section 2.2. The noise is assumed to be uncorrelated with the training data, and the training sequence consists of 26 training symbols. The second part of (7.46),

[5] This will in general not be entirely true since the taps are correlated due to the pulse shaping in the modulation. This assumption will however simplify our exposition since only the autocorrelation function for a single Rayleigh fading tap then has to be considered in the design.

$\boldsymbol{P}_{\text{Estimation uncertainty}}$, can then be computed directly and exactly as in [87] Section 4.2, i.e.

$$\boldsymbol{P}_{\text{Estimation uncertainty}} = \sigma_n^2 (\boldsymbol{D}^H \boldsymbol{D})^{-1} \tag{7.52}$$

where σ_n^2 is the noise variance and $\boldsymbol{D}$ is a matrix formed from the regressors, i.e. the training symbols

$$\boldsymbol{D} = \begin{bmatrix} \boldsymbol{d}^T(t_{min}) \\ \boldsymbol{d}^T(t_{min}+1) \\ \vdots \\ \boldsymbol{d}^T(t_{max}) \end{bmatrix} \tag{7.53}$$

with

$$\boldsymbol{d}^T(t) = \begin{bmatrix} d(t) & d(t-1) & \dots & d(t-nb) \end{bmatrix}. \tag{7.54}$$

In this example, a six tap FIR model was used (i.e. $nb = 5$). The time indices t_{min} and t_{max} are as usual chosen such that all used channel outputs, $b(q^{-1})d(t)$, are well defined. The noise variance, σ_n^2 can be estimated separately from the residuals. $\boldsymbol{P}_{\Delta b}$ can now be computed by using (7.51) and (7.52) in (7.46). With the resulting $\boldsymbol{P}_{\Delta b}$, the robust equalizer can be readily designed.

With a white noise-like training sequence, $\boldsymbol{P}_{\text{Estimation uncertainty}}$ will be approximately diagonal. Furthermore, in this example, it turns out, that the taps in the channel are only weakly correlated due to the pulse shaping. Our assumption, when deriving (7.51), that the coefficients in the FIR model were completely uncorrelated is thus approximately valid. With this assumption the correlation matrix for the uncertainties due to the fading, $\boldsymbol{P}_{\text{Time variation}}$, is approximately diagonal. The total correlation matrix for the uncertainties, $\boldsymbol{P}_{\Delta b}$, is thus approximately diagonal. Using this in (7.44) and (7.45), we see that the term $\bar{E}[\Delta b(q^{-1})\Delta b^H(q)]$ in (7.34) just becomes a constant (not a polynomial) which is added to the normalized noise variance ρ. Thus, in this example, an approximatively optimal robust equalizer is achieved just by *increasing the noise variance in the design.* The noise variance is in fact increased by an amount equal to the energy of the uncertainties in the channel.

The BER for the nominal and the robust DFE is presented in Figure 7.5 for a collection of scenarios with different SNR and mobile speeds. The performance gained with the robust DFE as compared to the nominal DFE can be seen in Figure 7.6. The performance gain increases with increasing speed of the mobile, and is largest at large SNR. This is natural. In these cases the channel varies significantly, while the noise level is small so the the nominal DFE will have relatively high gain in its feedforward filter. This combined with an incorrect channel causes the higher BER for the nominal DFE, which is based on the assumption of a perfect channel estimate.

We can thus conclude that the robust DFE can be used to reduce the effects of moderate fading over a frame. This can then be an alternative to tracking of the channel and adapting the equalizer to the channel variations over the frame.

7.4 Robustness of Space-Time Equalizers

When we studied the space-time DFE and the space-time MLSE in Chapters 3 and 4 we could note that, contrary to what one might expect, the indirectly tuned equalizers performed reasonably well even when the signal to interference ratio was very low. One would think that the channel estimates would be so poor at low signal to interference ratios such that the indirectly tuned equalizers would have a very poor performance. This is however not always the case. In this section we explain why. We will show that the indirectly tuned equalizers have a certain degree of built-in *self-robustification*. First, we design equalizers that are robust to the expected error in the channel estimate by using the averaged MSE criterion introduced in Section 7.1. Second, we note that the *nominal* space-time DFE or MLSE, designed based on estimated spatial covariance matrices, is in some cases automatically robustified, in a way which is similar to a robust design which explicitly takes model errors into account. The explanation for this phenomenon is that when an estimated spatial covariance matrix is utilized in the nominal design, the structure and magnitude of this matrix will affect the design in ways similar to a robust design. The gains in directions where the model uncertainty is large will be reduced.

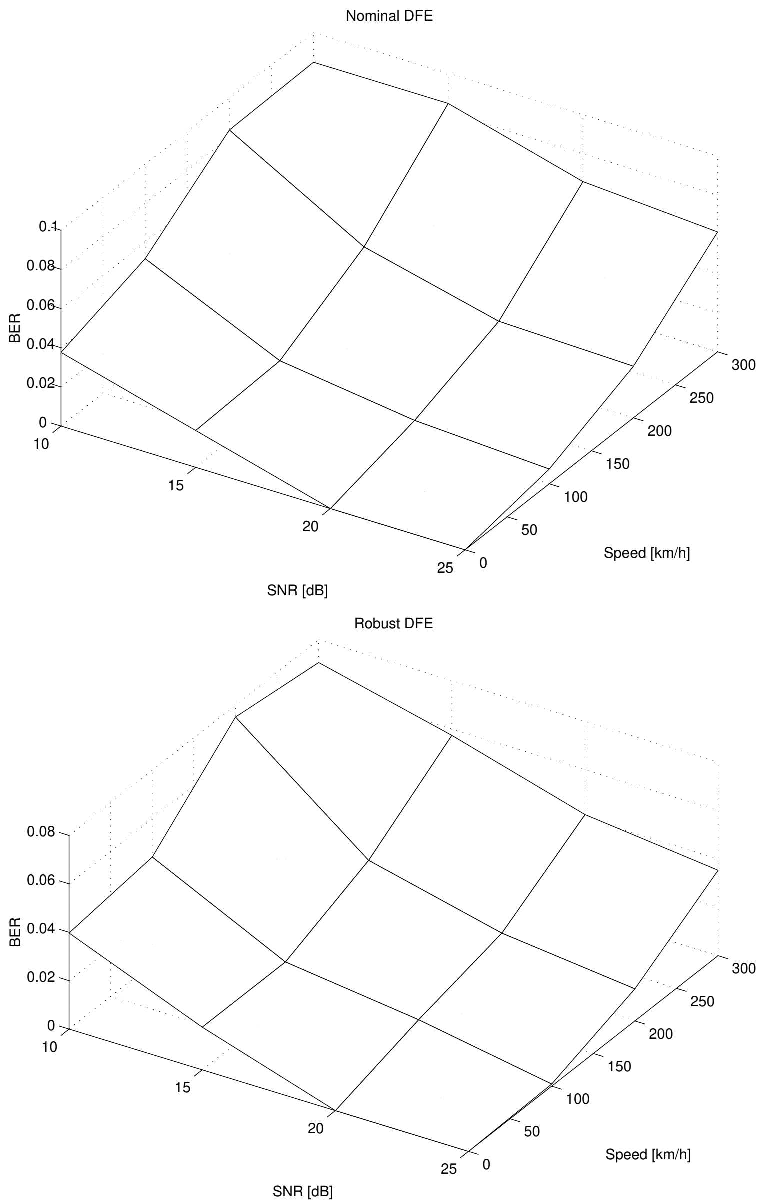


Figure 7.5: BER for the nominal and the robust DFE's.

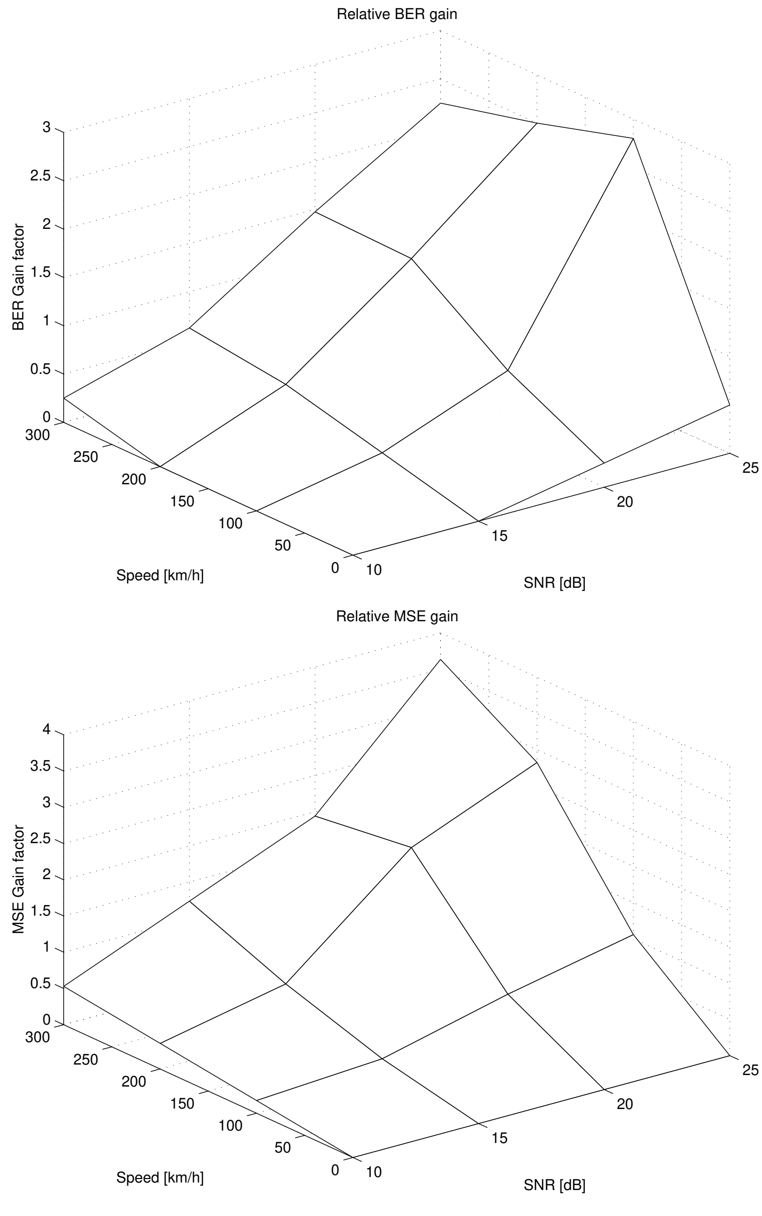


Figure 7.6: Performance gain, BER_{nom}/BER_{rob} and MSE_{nom}/MSE_{rob}, for the robust equalizer obtained by the averaged design, as compared to the nominal DFE based on the nominal estimated channel model.

7.4.1 Channel Model

To simplify the presentation we will assume that the co-channel interferers are temporally white and thus have only a single tap in their channels. The received signals collected from the M antenna elements, $\boldsymbol{y}(t) = [y_1(t)\ y_2(t)\ \ldots\ y_M(t)]^T$, can then, similar to (2.75), be modeled as

$$\boldsymbol{y}(t) = \boldsymbol{B}_0\boldsymbol{d}_0(t) + \sum_{i=1}^{K} \boldsymbol{b}_i d_i(t) + \boldsymbol{v}(t) \tag{7.55}$$

where $d_i(t)$, $i = 1, \ldots, K$ are the symbols transmitted by the K co-channel interferers and $\boldsymbol{b}_i$ are the single vector taps of the co-channel interferers, of dimensions $M \times 1$. The vector $\boldsymbol{d}_0(t) = [d_0(t)\ d_0(t-1)\ \ldots\ d_0(t-nb_0)]^T$ consists of lagged values of the symbol sequence for the desired signal with $\boldsymbol{B}_0$ being the corresponding channel matrix of size $M \times (nb_0 + 1)$. The vector $\boldsymbol{v}(t)$ represents spatially and temporally white thermal noise. We will assume that the transmitted symbols are temporally white with zero mean and unit variance, i.e.

$$E[d(t)d^*(t)] = 1. \tag{7.56}$$

The unit variance can always be arranged by a scaling. The symbol sequences for the interferers are assumed to be uncorrelated with the symbol sequence of the desired signal.

7.4.2 Channel Estimation Errors

The channel matrix for the desired signal, $\boldsymbol{B}_0$, is here assumed to be estimated by estimating the coefficients of the FIR channel model with the least squares method, as in Section 2.2.

As in (2.21), the complete, least squares estimated, channel matrix can thus be expressed as

$$\hat{\boldsymbol{B}}_0 = \hat{\boldsymbol{R}}^H_{\boldsymbol{d}_0\boldsymbol{y}}\hat{\boldsymbol{R}}^{-1}_{\boldsymbol{d}_0\boldsymbol{d}_0} = \hat{\boldsymbol{R}}_{\boldsymbol{y}\boldsymbol{d}_0}\hat{\boldsymbol{R}}^{-1}_{\boldsymbol{d}_0\boldsymbol{d}_0} \tag{7.57}$$

where

$$\hat{\boldsymbol{R}}_{\boldsymbol{y}\boldsymbol{d}_0} = \frac{1}{N_{eff}} \sum_{t=t_{min}}^{t_{max}} \boldsymbol{y}(t)\boldsymbol{d}_0^H(t) \tag{7.58}$$

and

$$\hat{\boldsymbol{R}}_{\boldsymbol{d}_0\boldsymbol{d}_0} = \frac{1}{N_{eff}} \sum_{t=t_{min}}^{t_{max}} \boldsymbol{d}_0(t)\boldsymbol{d}_0^H(t) \tag{7.59}$$

where t_{min} and t_{max} are chosen such that all involved entities are well defined and $N_{eff} = t_{max} - t_{min} + 1$ is the effective length of the training sequence. By separating out the error terms induced by the co-channel interferers and the noise, using (7.55), the channel estimate can be written as

$$\hat{\boldsymbol{B}}_0 = \boldsymbol{B}_0 + \sum_{i=1}^{K} \boldsymbol{b}_i \hat{\boldsymbol{R}}_{d_i\boldsymbol{d}_0} \hat{\boldsymbol{R}}_{\boldsymbol{d}_0\boldsymbol{d}_0}^{-1} + \hat{\boldsymbol{R}}_{\boldsymbol{v}\boldsymbol{d}_0} \hat{\boldsymbol{R}}_{\boldsymbol{d}_0\boldsymbol{d}_0}^{-1} \tag{7.60}$$

where

$$\hat{\boldsymbol{R}}_{d_i\boldsymbol{d}_0} = \frac{1}{N_{eff}} \sum_{t=t_{min}}^{t_{max}} d_i(t)\boldsymbol{d}_0{}^H(t) \tag{7.61}$$

and

$$\hat{\boldsymbol{R}}_{\boldsymbol{v}\boldsymbol{d}_0} = \frac{1}{N_{eff}} \sum_{t=t_{min}}^{t_{max}} \boldsymbol{v}(t)\boldsymbol{d}_0^H(t). \tag{7.62}$$

The error in the channel matrix estimate is thus given by

$$\sum_{i=1}^{K} \boldsymbol{b}_i \hat{\boldsymbol{R}}_{d_i\boldsymbol{d}_0} \hat{\boldsymbol{R}}_{\boldsymbol{d}_0\boldsymbol{d}_0}^{-1} + \hat{\boldsymbol{R}}_{\boldsymbol{v}\boldsymbol{d}_0} \hat{\boldsymbol{R}}_{\boldsymbol{d}_0\boldsymbol{d}_0}^{-1}. \tag{7.63}$$

We will denote the negative of this term with $\Delta\boldsymbol{B}_0$, i.e.

$$\boldsymbol{\Delta B}_0 = -\sum_{i=1}^{K} \boldsymbol{b}_i \hat{\boldsymbol{R}}_{d_i\boldsymbol{d}_0} \hat{\boldsymbol{R}}_{\boldsymbol{d}_0\boldsymbol{d}_0}^{-1} + \hat{\boldsymbol{R}}_{\boldsymbol{v}\boldsymbol{d}_0} \hat{\boldsymbol{R}}_{\boldsymbol{d}_0\boldsymbol{d}_0}^{-1} \tag{7.64}$$

such that we can write a probabilistic model of the channel as

$$\boldsymbol{B}_0 = \hat{\boldsymbol{B}}_0 + \Delta\boldsymbol{B}_0 \tag{7.65}$$

or equivalently

$$\boldsymbol{b}_0(q^{-1}) = \hat{\boldsymbol{b}}_0(q^{-1}) + \Delta\boldsymbol{b}_0(q^{-1}) \tag{7.66}$$

where the taps of $\Delta \boldsymbol{b}_0(q^{-1})$ are the columns in $\Delta \boldsymbol{B}_0$. In other words, we regard the nominal (unknown) true channel as a member of a set, a stochastic variable, with average value $\hat{\boldsymbol{B}}_0$. The nominal channel model is thus our present estimate of the channel, $\hat{\boldsymbol{B}}_0$ or $\hat{\boldsymbol{b}}_0(q^{-1})$, and the uncertainty in the channel is given by the error term $\Delta \boldsymbol{B}_0$ or $\Delta \boldsymbol{b}(q^{-1})$.

We will use (7.65), or (7.66), as an uncertain channel model when designing a robust equalizer in the following section. We assume that we do not know the symbols transmitted by the interferers and thus cannot identify their channels. Consequently, the error term $\Delta \boldsymbol{B}_0$ or $\Delta \boldsymbol{b}_0(q^{-1})$ is unknown.

7.4.3 Spatial Robustness of the Space-Time FIR-DFE

In Appendix 7.A.3 we show that the covariance of the channel estimation errors, $E[\Delta \boldsymbol{b}_0(q^{-1})\Delta \boldsymbol{b}_0^H(q)]$, can be approximated with a constant matrix as

$$E[\Delta \boldsymbol{b}_0(q^{-1})\Delta \boldsymbol{b}_0^H(q)] \approx \frac{1}{nb+1}\Delta \boldsymbol{B}_0 \boldsymbol{B}_0^H \tag{7.67}$$

In Appendix 7.A.2 we show that if the approximation (7.67) holds, then the robust DFE, for the scenario studied here, can be realized with FIR feedforward and feedback filters. Furthermore this robust DFE is equal to a nominal DFE design based on the channel estimate $\hat{\boldsymbol{b}}_0(q^{-1})$, if we in the design equations replace the spatial covariance matrix for the noise plus interference, $\boldsymbol{R}_{\boldsymbol{nn}}$, with a constant matrix $\boldsymbol{P}_0$ given by

$$\boldsymbol{P}_0 = \boldsymbol{R}_{\boldsymbol{nn}} + \frac{1}{nb+1}\Delta \boldsymbol{B}_0 \Delta \boldsymbol{B}_0^H. \tag{7.68}$$

The spatial color of the noise plus interference should thus be replaced by the sum of the spatial color of the noise and interference *and* the spatial color of the uncertainties in the channel estimate.

Consider now the estimate of the spatial color of the residuals

$$\hat{\boldsymbol{R}}_{\boldsymbol{rr}} = \frac{1}{N_{eff}} \sum_{t=t_{min}}^{t_{max}} \boldsymbol{r}(t)\boldsymbol{r}^H(t) \tag{7.69}$$

where $\boldsymbol{r}(t)$ are the residuals from the estimation with the least squares method

$$\boldsymbol{r}(t) = \boldsymbol{y}(t) - \hat{\boldsymbol{B}}_0 \boldsymbol{d}_0(t). \tag{7.70}$$

Using (7.55) and (7.65) the residual $\boldsymbol{r}(t)$ can be expressed as

$$\boldsymbol{r}(t) = \Delta \boldsymbol{B}_0 \boldsymbol{d}_0(t) + \boldsymbol{n}(t) \tag{7.71}$$

where $\boldsymbol{n}(t)$ represent noise plus interference. The sample matrix estimate in (7.69) can thus be expressed as

$$\begin{aligned}\hat{\boldsymbol{R}}_{\boldsymbol{rr}} &= \frac{1}{N_{eff}} \sum_{t=t_{min}}^{t_{max}} \boldsymbol{r}(t)\boldsymbol{r}^H(t) \\ &= \frac{1}{N_{eff}} \sum_{t=t_{min}}^{t_{max}} [\Delta \boldsymbol{B}_0 \boldsymbol{d}_0(t) + \boldsymbol{n}(t)][\Delta \boldsymbol{B}_0 \boldsymbol{d}_0(t) + \boldsymbol{n}(t)]^H. \end{aligned}\tag{7.72}$$

Since $\boldsymbol{d}_0(t)$ and $\boldsymbol{n}(t)$ are assumed to be uncorrelated we can do the following approximation, given that N is not too small

$$\begin{aligned}\hat{\boldsymbol{R}}_{\boldsymbol{rr}} \approx\ & \frac{1}{N_{eff}} \sum_{t=t_{min}}^{t_{max}} \Delta \boldsymbol{B}_0 \boldsymbol{d}_0(t) \boldsymbol{d}_0^H(t) \Delta \boldsymbol{B}_0^H \\ & + \frac{1}{N_{eff}} \sum_{t=t_{min}}^{t_{max}} \boldsymbol{n}(t)\boldsymbol{n}^H(t). \end{aligned}\tag{7.73}$$

Furthermore, since $d_0(t)$ is assumed to be approximately white with unit variance

$$\hat{\boldsymbol{R}}_{\boldsymbol{rr}} \approx \frac{1}{nb+1} \Delta \boldsymbol{B}_0 \Delta \boldsymbol{B}_0^H + \boldsymbol{R}_{\boldsymbol{nn}} \tag{7.74}$$

where $\Delta \boldsymbol{B}_0$ is given by (7.64) and $\boldsymbol{R}_{\boldsymbol{nn}}$ is the spatial noise plus interference covariance defined in (7.116). This is of course a crude approximation if the training sequence is short. Asymptotically, with increasing length of the training sequence, it will however hold exactly.

Comparing (7.74) with (7.68) we thus conclude that

$$\hat{\boldsymbol{R}}_{\boldsymbol{rr}} \approx \boldsymbol{P}_0. \tag{7.75}$$

Thus, assume we have an FIR channel with temporally white interferers. A nominal space-time DFE with FIR filters in the feedforward and the feedback links, is designed to minimize the MSE over the training sequence. The LS-estimated channels (7.57) and the estimated spatial color (7.69) of the residuals are used in the tuning, as for the IS-DFE in Section 3.3.3. This equalizer is then *automatically robustified* against the channel estimation errors caused by strong co-channel interferers. If the co-channel interferers are temporally colored, it is likely that the IS-DFE is still *spatially* self-robustifying. However, a temporal color on the interference can potentially be utilized to *improve* the performance further. It is important to note, that if the number of uncorrelated strong interferers is too large compared to the number of antennas, then spatial-only interference suppression of the interference will not suffice.

If we study the estimated channels to the individual antennas in the presence of strong co-channel interferers, we can be led to believe that they are completely useless as they can be strongly distorted. Yet they turn out to be adequate for the IS-DFE in Chapter 3. The above analysis explains why this is the case. The individual channels may be distorted, but if we view the channels as a group, we will find that the channels are distorted in the spatial subspace spanned by the co-channel interferers channel taps. If the number of antennas is large enough, there will be a remaining spatial subspace where the channels will be distorted by the thermal noise only. The DFE-filtering can be performed using signals from this subspace, while suppressing signals from the subspace corrupted by the co-channel interferers.

7.4.4 Spatial Robustness of the MMF-MLSE and MMF-DFE

As for the IS-DFE, one can also wonder if short training sequences are adequate for obtaining a reasonable performance with the indirectly tuned MLSE in Section 4.4.2 and the indirectly tuned MMF-DFE in Section 3.3.5. The channels to the individual antennas will also here be severely distorted if strong co-channels interferers are present. The tuning of the MMF filters will however have many similarities with the tuning of the filters in the MISO-DFE discussed in the previous section. This can be understood if one considers the fact that the MMF filters in the MMF-DFE and the multi-channel MLSE can, as the filters in the DFE, be tuned with an MMSE criterion as is discussed in Sections 3.3.5 and 4.4.3. Basically the only difference is then that when tuning an MMF-filter, we use a "decision feedback" filter that is double sided (and complex conjugated symmetric).

From this it is easy to realize that the a spatially robust MLSE/MMF-DFE can be constructed approximately by tuning the MMF and "feedback" filters using the least squares estimated channel coefficients and the spatial color of the residuals. This is why the performance of the indirectly tuned multichannel MLSE using the spatial covariance of the residuals in the example in Section 4.5 in Chapter 4 does not degrade so seriously when the strength of the co-channel interferers increases, even though the relative errors in the channel estimates do. The same applies to the simulations with the MMF-DFE presented in Section 3.4 in Chapter 3. Again, of course, if we have too many uncorrelated strong interferers compared to the number of antennas, spatial-only suppression of the interferers will not suffice.

7.A Appendix

7.A.1 Robust MMSE Filtering

In this appendix, to simplify the notation, polynomials are represented with capital non-boldface letters leaving out the explicit dependence of q^{-1} or z^{-1}, i.e. for example

$$b(q^{-1}) \quad \text{is exchanged for} \quad B. \tag{7.76}$$

We can model a general communication channel as

$$y(t) = \mathcal{G}(q^{-1})d(t) + \mathcal{H}(q^{-1})v(t) \tag{7.77}$$

where $y(t)$ is the received signal, $\mathcal{G}(q^{-1})$ and $\mathcal{H}(q^{-1})$ are general IIR filters, $d(t)$ are the transmitted symbols and $v(t)$ represents the innovations for the noise.

Let us parametrize the communication channel as

$$\begin{aligned} \mathcal{G} = q^{-k}(\mathcal{G}_0 + \Delta\mathcal{G}) &= q^{-k}\left(\frac{B_0}{A_0} + \frac{\Delta B}{A_1}\right) \\ &= q^{-k}\frac{(B_0 A_1 + A_0 \Delta B)}{A_0 A_1} \triangleq q^{-k}\frac{B}{A} \end{aligned} \tag{7.78}$$

$$\mathcal{H} = \mathcal{H}_0 + \Delta\mathcal{H} = \frac{M_0}{N_0} + \frac{\Delta M}{N_1} = \frac{M_0 N_1 + N_0 \Delta M}{N_0 N_1} \triangleq \frac{M}{N} \tag{7.79}$$

where the polynomials $A(q^{-1})$ and $N(q^{-1})$ are assumed to be stable and where ΔB and ΔM are polynomials with random variables as coefficients. The problem formulation of Section 7.1 is here generalized, by allowing the transmitted sequence $d(t)$ to be colored. It is modeled by

$$d(t) = \frac{C}{D}e(t) \tag{7.80}$$

where D is stable and C has zeros in $|z| \leq 1$. The signal $e(t)$ is white, with zero mean and $E[e(t)]^2 = \lambda_e$. The aim is to obtain an estimator

$$\hat{d}(t-\ell|t) = \frac{S}{R}y(t) - \frac{Q}{P}\tilde{d}(t-\ell-1) \tag{7.81}$$

with $R(q^{-1})$ and $P(q^{-1})$ stable, such that the average mean square error

$$J = \bar{E}E|d(t-\ell) - \tilde{d}(t-\ell|t)|^2 = \bar{E}E|z(t-\ell)|^2 \tag{7.82}$$

is minimized. Above, ℓ denotes the smoothing lag and

$$y(t) = q^{-k}\frac{B}{A}\,\frac{C}{D}e(t) + \frac{M}{N}v(t) \tag{7.83}$$

is the measurement signal while $\tilde{d}(t-\ell-1)$ represent previous decision, obtained at time $t-\ell-1$ by feeding $\hat{d}(t-\ell-1)$ through a decision device.

If all previous decisions are correct, we can proceed along the lines of [90]. However, if an incorrect decision is made, an error burst might occur. Decision errors will be modeled by a fictitious zero mean white noise, $\kappa(t)$, added to the correct sequence $d(t)$. By means of $\kappa(t)$ we can thus write the decisions, $\tilde{d}(t)$, as

$$\tilde{d}(t) = d(t) + \kappa(t). \tag{7.84}$$

For further use, assume that $e(t)$, $v(t)$ and $\kappa(t)$ are all mutually uncorrelated and introduce the noise variance ratios

$$\rho \triangleq \frac{E|v(t)|^2}{E|e(t)|^2} \quad ; \quad \eta \triangleq \frac{E|\kappa(t)|^2}{E|e(t)|^2}.$$

The estimation error, to be optimized, can be written as

$$\begin{aligned}
z(t-\ell) &= d(t-\ell) - \hat{d}(t-\ell|t) \\
&= q^{-\ell}\frac{C}{D}e(t) - q^{-k}\frac{S}{R}\,\frac{B}{A}\,\frac{C}{D}e(t) - \frac{S}{R}\,\frac{M}{N}\,v(t) \\
&\quad + q^{-\ell-1}\frac{Q}{P}\,\frac{C}{D}e(t) + q^{-\ell-1}\frac{Q}{P}\kappa(t) \\
&= \frac{(q^{-\ell}RAP - q^{-k}PSB + q^{-\ell-1}QRA)}{RAPD}Ce(t) - \frac{S}{R}\,\frac{M}{N}v(t) \\
&\quad + q^{-\ell-1}\frac{Q}{P}\kappa(t).
\end{aligned} \tag{7.85}$$

In order to minimize the averaged estimation error $\bar{E}E(z(t-\ell))^2$ with respect to S, R, Q and P define the following polynomials:

$$H \triangleq \bar{E}(B)NC \triangleq B_0 A_1 N_0 N_1 C \tag{7.86}$$

$$\tilde{B}\tilde{B}^H \triangleq \bar{E}(BB^H) \triangleq B_0 B_0^H A_1 A_1^H + A_0 A_0^H \bar{E}(\Delta B \Delta B^H) \tag{7.87}$$

$$\tilde{M}\tilde{M}^H \triangleq \bar{E}(MM^H) \triangleq M_0 M_0^H N_1 N_1^H + N_0 N_0^H \bar{E}(\Delta M \Delta M^H) \tag{7.88}$$

$$\tilde{\gamma}\tilde{\gamma}^H \triangleq \bar{E}(AA^H DD^H MM^H) \triangleq AA^H DD^H \bar{E}(MM^H) \tag{7.89}$$

$$\tilde{\tau}\tilde{\tau}^H \triangleq \bar{E}(NN^H CC^H BB^H) \triangleq NN^H CC^H \bar{E}(BB^H). \tag{7.90}$$

In order to derive the design equations for the optimal equalizer, a variational approach, presented in [3] for nominal design, and [91] for robust design, will be used. Thus, the estimation error (7.85) should, on average over the model class, be orthogonal to any linear variation of the estimate $\hat{d}(t-\ell|t)$ which can be obtained from the signals available at time t. We denote such a variation by

$$n(t) = \underbrace{\mathcal{G}_1 \boldsymbol{y}(t)}_{n_1(t)} + \underbrace{\mathcal{G}_2 \tilde{d}(t-\ell-1)}_{n_2(t)}$$

where $\mathcal{G}_1$ and $\mathcal{G}_2$ are arbitrary but stable and causal transfer functions.

The first linear equation

Orthogonality of $z(t-\ell)$ will now be assured with respect to $n_1(t)$ and $n_2(t)$ separately. Thus, for $n_1(t)$ we obtain, by using (7.83), (7.85) and the assumption that κ is uncorrelated to other signals,

$$\bar{E}E(z(t-\ell)n_1^H(t)) =$$

$$= \bar{E}E\left(\frac{(q^{-\ell}RAP - q^{-k}PSB + q^{-\ell-1}QAR)C}{RAPD}e(t) - \frac{S}{R}\ \frac{M}{N}v(t)\right)$$
$$\times \quad \left[\mathcal{G}_1\left(q^{-k}\frac{B}{A}\ \frac{C}{D}e(t) + \frac{M}{N}v(t)\right)\right]^H$$

$$= \frac{E|e(t)|^2}{2\pi i}\bar{E}\oint_{|z|=1}\left(\frac{(z^{-\ell}RAPN - z^{-k}PSBN + z^{-\ell-1}QARN)}{RAPDN}\right.$$
$$\times \quad \left. z^k\frac{CC^HB^HN^H}{D^HA^HN^H} - \rho\ \frac{APD}{APD}\ \frac{SM}{RN}\ \frac{M^HA^HD^H}{N^HA^HD^H}\right)\mathcal{G}_1^H\frac{dz}{z}$$

$$= \frac{\lambda_e}{2\pi i}\oint\left(\frac{z^{k-\ell}RAPNCC^H\bar{E}(B^H)N^H - PSCC^HNN^H\bar{E}(BB^H)}{RAPDND^HA^HN^H}\right)\mathcal{G}_1^H\frac{dz}{z}$$
$$+ \quad \frac{\lambda_e}{2\pi i}\oint\left(\frac{z^{k-\ell-1}QRANCC^H\bar{E}(B^H)N^H - \rho PSAA^HDD^H\bar{E}(MM^H)}{RAPDND^HA^HN^H}\right)\mathcal{G}_1^H\frac{dz}{z}$$

$$= \frac{\lambda_e}{2\pi i}\oint\left(\frac{(z^{k-\ell}PRANCH^H + z^{k-\ell-1}QRANCH^H)}{RAPDND^HA^HN^H}\right)\mathcal{G}_1^H\frac{dz}{z}$$
$$- \quad \frac{\lambda_e}{2\pi i}\oint\left(\frac{PS(\tilde{\tau}\tilde{\tau}^H + \rho\tilde{\gamma}\tilde{\gamma}^H)}{RAPDND^HA^HN^H}\right)\mathcal{G}_1^H\frac{dz}{z}$$

where (7.86), (7.89) and (7.90) were used in the last equality.

Orthogonality of the error $z(t-\ell)$ with respect to $n_1(t)$ is fulfilled if all poles inside the unit circle of the integrand are canceled by zeros. Since A, N and D are assumed stable, while P and R in the equalizer must be stable, there must therefore exist a polynomial $L_2^H(z)$, such that

$$z^{k-\ell}PRANCH^H + z^{k-\ell-1}QRANCH^H - PS(\tilde{\tau}\tilde{\tau}^H + \rho\tilde{\gamma}\tilde{\gamma})$$
$$= zL_2^HRAPDN \quad (7.91)$$

or equivalently

$$RAN(z^{k-\ell}PCH^H + z^{k-\ell-1}QCH^H - zPDL_2^H) \\ = PS(\tilde{\tau}\tilde{\tau} + \rho\tilde{\gamma}\tilde{\gamma}). \quad (7.92)$$

The second linear equation and the spectral factorization

For the variational term $n_2(t)$, we obtain, by using (7.80), (7.84), (7.85) and the assumption that the noise $v(t)$ is uncorrelated with $e(t)$ and $\kappa(t)$,

$$\begin{aligned}
&\bar{E}E(z(t-\ell)n_2^H(t)) = \\
&= \bar{E}E\left(\frac{(q^{-\ell}RAP - q^{-k}PSB + q^{-\ell-1}QAR)C}{RAPD}e(t) + q^{-\ell-1}\frac{Q}{P}\kappa(t)\right) \\
&\quad \times \left[\mathcal{G}_2^H\left(q^{-\ell-1}\frac{C}{D}e(t) + q^{-\ell-1}\kappa(t)\right)\right]^H \\
&= \frac{E|e(t)|^2}{2\pi i}\bar{E}\oint_{|z|=1}\left(\frac{(zRAP - z^{-k+\ell+1}PSB + QAR)CC^H}{RAPDD^H}\right. \\
&\quad \left. +\eta\frac{Q}{P}\frac{RAD}{RAD}\frac{D^H}{D^H}\right)\mathcal{G}_2^H\frac{dz}{z} \\
&= \frac{\lambda_e}{2\pi i}\oint\frac{(zRAP - z^{-k+\ell+1}PS\bar{E}(B) + QAR)CC^H}{RAPDD^H}\mathcal{G}_2^H\frac{dz}{z} \\
&\quad +\frac{\lambda_e}{2\pi i}\oint\frac{\eta QRADD^H}{RAPDD^H}\mathcal{G}_2^H\frac{dz}{z}. \qquad (7.93)
\end{aligned}$$

Orthogonality of the error $z(t-\ell)$ with respect to $n_2(t)$ is here fulfilled if all poles of the integrand inside the unit circle are canceled by zeros. There must therefore exist a polynomial $L_1^H(z)$, such that

$$zRAPCC^H - z^{\ell-k+1}PS\bar{E}(B)CC^H + QARCC^H + \eta QARDD^H \\ = zL_1^H ARPD \quad (7.94)$$

or equivalently

$$P(zRACC^H - z^{\ell-k+1}S\bar{E}(B)CC^H - zL_1^H RAD) = \\ -(CC^H + \eta DD^H)QAR. \quad (7.95)$$

It is evident from (7.95) that AR must be a factor of P, and from (7.92) that RAN must be a factor of PS. Let us therefore choose

$$P = AR \quad (7.96)$$

$$S = S_0 N \quad (7.97)$$

where the polynomial S_0 is as yet undetermined. The use of (7.96) and (7.97) in (7.92) and (7.95) then gives

$$z^{-\ell+k}ARCH^H + z^{k-\ell-1}QCH^H - zARDL_2^H \\ = S_0(\tilde{\tau}\tilde{\tau}^H + \rho\tilde{\gamma}\tilde{\gamma}^H) \quad (7.98)$$

$$zARCC^H - z^{\ell-k+1}S_0N\bar{E}(B)CC^H - zL_1^H RAD \\ = -Q(CC^H + \eta DD^H). \quad (7.99)$$

Define

$$CC^H + \eta DD^H \triangleq VV^H. \quad (7.100)$$

By multiplying (7.98) by VV^H, multiplying (7.99) by $z^{-1-\ell+k}CH^H$, and using (7.86), we obtain

$$z^{-\ell+k}ARCH^H VV^H + z^{k-\ell-1}QCH^H VV^H - zARDL_2^H VV^H \\ = S_0 VV^H(\tilde{\tau}\tilde{\tau}^H + \rho\tilde{\gamma}\tilde{\gamma}^H) \quad (7.101)$$

and

$$z^{-\ell+k}ARCC^H CH^H + z^{k-\ell-1}QCH^H VV^H - z^{-\ell+k}L_1^H RADCH^H \\ = S_0 HC^H CH^H. \quad (7.102)$$

Subtracting equation (7.102) from (7.101) yields

$$z^{-\ell+k}ARCH^H\eta DD^H - zARDL_2^H VV^H + z^{-\ell+k}L_1^H RADCH^H$$
$$= S_0(VV^H\tilde{\tau}\tilde{\tau}^H + \rho VV^H\tilde{\gamma}\tilde{\gamma}^H - CC^H HH^H). \quad (7.103)$$

Define the polynomial spectral factorization

$$\gamma\beta\beta^H \triangleq VV^H\tilde{\tau}\tilde{\tau}^H + \rho VV^H\tilde{\gamma}\tilde{\gamma}^H - CC^H HH^H \quad (7.104)$$

where β is stable and monic, while γ is a scalar. The filter denominator R is a factor of the left-hand side of (7.103). Since R must be stable, equation (7.103) indicates that we should choose

$$R = \beta. \quad (7.105)$$

Since also A is a factor of the left-hand side of (7.103), we must require it to be a factor of S_0 on the right-hand side. Thus,

$$S_0 = S_1 A \quad (7.106)$$

for some S_1. Substituting S_1A for S_0 in (7.98) reveals that A must be a factor of Q. Thus, for some Q_1,

$$Q = Q_1 A. \quad (7.107)$$

Summary of the design equations

Equations (7.99) and (7.103) constitute our coupled polynomial equations. Multiply (7.99) by z^{-1} and change sign of (7.103). By also substituting q for z, and by using (7.86), (7.100), (7.104), (7.105), (7.106) and (7.107), these equations can be written as

$$A\beta CC^H - q^{\ell-k}S_1AHC^H - L_1^H\beta AD = -q^{-1}Q_1AVV^H \quad (7.108)$$

$$-q^{-\ell+k}\eta CH^H DD^H + qL_2^H DVV^H - q^{-\ell+k}L_1^H DCH^H$$
$$= -S_1\gamma\beta^H. \quad (7.109)$$

Note that $VV^H = CC^H + \eta DD^H$. Using (7.87)-(7.90), the spectral factorization defined by (7.104) can be expressed as

$$
\begin{aligned}
\gamma\beta\beta^H = {} & (CC^H + \eta DD^H)NN^H CC^H (B_0 B_0^H A_1 A_1^H + A_0 A_0^H \bar{E}(\Delta B \Delta B^H)) \\
& + \rho(CC^H + \eta DD^H) AA^H DD^H (M_0 M_0^H N_1 N_1^H \\
& + N_0 N_0^H \bar{E}(\Delta M \Delta M)) - CC^H CC^H NN^H B_0 A_1 B_0^H A_1^H \qquad (7.110)
\end{aligned}
$$

or

$$
\begin{aligned}
\gamma\beta\beta^H = {} & CC^H CC^H NN^H A_0 A_0^H \bar{E}(\Delta B \Delta B^H) + \eta DD^H NN^H CC^H \tilde{B}\tilde{B}^H \\
& + \rho CC^H DD^H AA^H \tilde{M}\tilde{M}^H + \rho\eta DD^H DD^H AA^H \tilde{M}\tilde{M}^H. \qquad (7.111)
\end{aligned}
$$

This expression reduces to (7.20) for white symbol sequences ($C = D = 1$). By using the definition (7.100) and canceling the common factor A in (7.108), the linear equations (7.108), (7.109) can be expressed as

$$
\begin{aligned}
& \beta CC^H + q^{-1}(CC^H + \eta DD^H)Q_1 \\
= {} & q^{\ell-k} HC^H S_1 + \beta D L_1^H \qquad (7.112) \\
& -q^{-\ell+k}\eta CH^H DD^H + qD(CC^H + \eta DD^H)L_2^H \\
= {} & -\gamma\beta^H S_1 + q^{-\ell+k} DCH^H L_1^H. \qquad (7.113)
\end{aligned}
$$

Summarizing (7.96), (7.97), (7.105), (7.106) and (7.107), it becomes evident that the filters of the optimal decision feedback equalizer (7.81) have the following structure

$$
\frac{S}{R} = \frac{N_0 N_1 A_0 A_1 S_1}{\beta} \quad ; \quad \frac{Q}{P} = \frac{Q_1 A}{AR} = \frac{Q_1}{\beta} \qquad (7.114)
$$

where S_1, L_1^H, Q_1 and L_2^H are obtained as the unique solution to (7.112), (7.113). In the same way as for the nominal solution, discussed in [90], these two equations can be reformulated as a linear system of equations in S_1 and L_1^H. Once S_1 and L_1^H are obtained, Q_1 can be calculated. See also Appendix 3.A.1 of Chapter 3.

7.A.2 Derivation of a Spatially Robust DFE

As usual, we will adopt the assumption that the DFE feeds back only correct decisions. The averaged MSE-optimal DFE can then be designed as a linear filter. Since only the second order properties of the processed signals will affect the DFE design, we can replace the sum of all interferers and the noise in (7.55) with a, for our purpose, equivalent interference term, $\boldsymbol{n}(t)$,

$$\boldsymbol{n}(t) = \sum_{i=1}^{K} \boldsymbol{b}_i d_i(t) + \boldsymbol{v}(t) \tag{7.115}$$

with the spatial covariance matrix

$$\boldsymbol{R_{nn}} = E\left[\left(\sum_{i=1}^{K} \boldsymbol{b}_i d_i(t) + \boldsymbol{v}(t)\right)\left(\sum_{i=1}^{K} \boldsymbol{b}_i d_i(t) + \boldsymbol{v}(t)\right)^H\right]. \tag{7.116}$$

We will now attempt to construct a DFE, with FIR structures of the feed-forward and feedback links, that is robust to the channel estimation errors expressed in (7.63). It is not given that the MSE-optimal DFE can be realized with FIR-filters but we will here show that for our case, with temporally white interferers, it is approximately true.

Let us express the DFE as:

$$\hat{d}_0(t-\ell|t) = \boldsymbol{s}(q^{-1})\boldsymbol{y}(t) - Q(q^{-1})d_0(t-\ell-1). \tag{7.117}$$

where $\boldsymbol{s}(q^{-1})$ is a $1 \times M$ vector FIR filter and $Q(q^{-1})$ is a polynomial. Note that we use $d_0(t-\ell-1)$ in the feedback filter since we assume correct past decisions.

To obtain a robust DFE we minimize the criterion

$$J = \bar{E}E|d_0(t-\ell) - \hat{d}_0(t-\ell|t)|^2 = \bar{E}E|z(t)|^2 \tag{7.118}$$

where

$$\begin{aligned} z(t) &= d_0(t-\ell) - \hat{d}_0(t-\ell|t) = \\ &(q^{-\ell} - \boldsymbol{s}(q^{-1})(\boldsymbol{b}_0(q^{-1}) + \Delta\boldsymbol{b}_0(q^{-1})) \\ &+q^{-\ell-1}Q(q^{-1}))d_0(t) - \boldsymbol{s}(q^{-1})\boldsymbol{n}(t). \end{aligned} \tag{7.119}$$

The expectation E is with respect to the noise and symbol realizations, $\boldsymbol{n}(t)$ and $d(t)$, while the expectation $\bar{E}$ is with respect to the channel uncertainty $\Delta\boldsymbol{b}_0(q^{-1})$, which is assumed to be zero mean, and have a known covariance matrix $\bar{E}[\Delta\boldsymbol{b}_0(q^{-1})\Delta\boldsymbol{b}_0^H(q)]$.

If we assume that the training sequence for the desired signal is approximately white and that the noise level is small compared to the interference level then it can be shown, as in Appendix 7.A.3, that the expectation $\bar{E}[\Delta\boldsymbol{b}_0(q^{-1})\Delta\boldsymbol{b}_0^H(q)]$ is approximately a *constant* matrix and can be estimated as

$$\bar{E}[\Delta\boldsymbol{b}_0(q^{-1})\Delta\boldsymbol{b}_0^H(q)] \approx \frac{1}{nb+1}\Delta\boldsymbol{B}_0\Delta\boldsymbol{B}_0^H. \tag{7.120}$$

We will here make the assumtion that indeed

$$\bar{E}[\Delta\boldsymbol{b}_0(q^{-1})\Delta\boldsymbol{b}_0^H(q)] = \frac{1}{nb+1}\Delta\boldsymbol{B}_0\Delta\boldsymbol{B}_0^H. \tag{7.121}$$

We will below show that with this assumption, the robust DFE, for the scenario studied here, can be realized with FIR feedforward and feedback filters.

In order to minimize the criterion in (7.118), $z(t)$ has to be orthogonal to any filtered version of the input signals $\boldsymbol{y}(t)$ and to $d_0(t-\ell-1)$. The signal $z(t)$ thus has to be orthogonal to

$$e_1(t) \;\triangleq\; \boldsymbol{g}_1(q^{-1})\boldsymbol{y}(t) \tag{7.122}$$

and to

$$e_2(t) \;\triangleq\; g_2(q^{-1})d_0(t-\ell-1) \tag{7.123}$$

where $\boldsymbol{g}_1(q^{-1})$, of dimension $1 \times M$, and $g_2(q^{-1})$ are arbitrary stable and causal *IIR* filters.

Orthogonality with respect to $e_1(t) = \boldsymbol{g}_1(q^{-1})\boldsymbol{y}(t)$ gives:

$$\begin{aligned}
&\bar{E}E[z(t)e_1^H(t)] = \\
&\quad \bar{E}E[(q^{-\ell} - \boldsymbol{s}(q^{-1})(\boldsymbol{b}_0(q^{-1}) + \Delta\boldsymbol{b}_0(q^{-1})) \\
&\quad +q^{-\ell-1}Q(q^{-1}))d_0(t) - \boldsymbol{s}(q^{-1})\boldsymbol{n}(t)] \\
&\quad [\boldsymbol{g}_1(q^{-1})((\boldsymbol{b}_0(q^{-1}) + \Delta\boldsymbol{b}_0(q^{-1}))d_0(t) + \boldsymbol{n}(t))]^H \\
&\quad = 0.
\end{aligned} \tag{7.124}$$

Evaluating the expectation with respect to the signals $d_0(t)$ and $\boldsymbol{n}(t)$ gives

$$\begin{aligned}\frac{1}{2\pi i}\bar{E}\oint_{|z|=1}&[(z^{-\ell}-\boldsymbol{s}(z^{-1})(\boldsymbol{b}_0(z^{-1})+\Delta\boldsymbol{b}_0(z^{-1}))\\&+z^{-\ell-1}Q(z^{-1}))(\boldsymbol{b}_0^H(z)+\Delta\boldsymbol{b}_0^H(z))-\\&\boldsymbol{s}(z^{-1})\boldsymbol{R_{nn}}]\boldsymbol{g}_1^H(z)\frac{dz}{z}=0.\end{aligned} \tag{7.125}$$

Moving the expectation, $\bar{E}$, inside the integration and observing that $\bar{E}\Delta\boldsymbol{b}(q^{-1})=\boldsymbol{b}_0$ gives

$$\begin{aligned}\frac{1}{2\pi i}\oint_{|z|=1}&[(z^{-\ell}\boldsymbol{b}_0^H(z)+z^{-\ell-1}Q(z^{-1})\boldsymbol{b}_0^H(z)\\&-\boldsymbol{s}(z^{-1})(\boldsymbol{b}_0(z^{-1})\boldsymbol{b}_0^H(z)+\bar{E}[\Delta\boldsymbol{b}_0(z)\Delta\boldsymbol{b}_0^H(z^{-1})]\\&+\boldsymbol{R_{nn}})]\boldsymbol{g}_1^H(z)\frac{dz}{z}=0.\end{aligned} \tag{7.126}$$

In order for this expression to be zero for *all* admissible $\boldsymbol{g}_1(z^{-1})$, we need to require that the integrand has no poles inside the unit circle. Since the factor $\boldsymbol{g}_1^H(z)$ has no poles inside the unit circle we are left with the requirement

$$\begin{aligned}&z^{-\ell}\boldsymbol{b}_0^H(z)+z^{-\ell-1}Q(z^{-1})\boldsymbol{b}_0^H(z)\\&\quad-\boldsymbol{s}(z^{-1})(\boldsymbol{b}_0(z^{-1})\boldsymbol{b}_0^H(z)+\bar{E}(\Delta\boldsymbol{b}_0(z^{-1})\Delta\boldsymbol{b}_0^H(z))+\boldsymbol{R_{nn}})\\&\quad=z\boldsymbol{l}_2^H(z)\end{aligned} \tag{7.127}$$

where $\boldsymbol{l}_2(z^{-1})$ is a causal and stable column vector polynomial to be determined.

Orthogonality with respect to $e_2(t)=g_2(q^{-1})d_0(t-\ell-1)$ gives:

$$\begin{aligned}&\bar{E}E[z(t)e_2^H(t)]=\\&\quad=\bar{E}E[(q^{-\ell}-\boldsymbol{s}(q^{-1})(\boldsymbol{b}_0(q^{-1})+\Delta\boldsymbol{b}_0(q^{-1}))\\&\quad+q^{-\ell-1}Q(q^{-1}))d_0(t)-\boldsymbol{s}(q^{-1})\boldsymbol{n}(t)]\\&\quad[g_2(q^{-1})q^{-\ell-1}d_0(t)]^H=0.\end{aligned} \tag{7.128}$$

Evaluating the expectation with respect to the signals $d(t)$ and $\boldsymbol{n}(t)$ gives:

$$\begin{aligned}\frac{1}{2\pi i}\bar{E}\oint_{|z|=1}&(z-z^{\ell+1}\boldsymbol{s}(z^{-1})(\boldsymbol{b}_0(z^{-1})+\Delta\boldsymbol{b}_0(z^{-1}))\\&+Q(z^{-1}))g_2^H(z)\frac{dz}{z}=0.\end{aligned} \tag{7.129}$$

Similar to equation (7.125), in order for this equation to be zero for all admissible $g_2(z^{-1})$, we must have

$$z - z^{\ell+1}\boldsymbol{s}(z^{-1})\boldsymbol{b}_0(z^{-1}) + Q(z^{-1}) = z l_1^H(z) \tag{7.130}$$

where $l_1(z^{-1})$ is a causal and stable polynomial in z^{-1} to be determined.

Multiplying (7.130) with z^{-1} and rearranging gives

$$1 + z^{-1}Q(z^{-1}) = z^{\ell}\boldsymbol{s}(z^{-1})\boldsymbol{b}_0(z^{-1}) + l_1^H(z). \tag{7.131}$$

Subtracting equation (7.131) multiplied with $z^{-\ell}\boldsymbol{b}_0^H(z)$ from the right, from equation (7.127) gives

$$\begin{aligned} z\boldsymbol{l}_2^H(z) = &-\boldsymbol{s}(z^{-1})\left(\bar{E}\left[\Delta\boldsymbol{b}_0(z^{-1})\Delta\boldsymbol{b}_0^H(z)\right] + \boldsymbol{R_{nn}}\right) \\ &+ z^{-\ell}l_1^H(z)\boldsymbol{b}_0^H(z). \end{aligned} \tag{7.132}$$

Let us now assume that the double sided polynomial $\bar{E}\left[\Delta\boldsymbol{b}_0(z^{-1})\Delta\boldsymbol{b}_0(z)\right]$ has the double sided order δnb_0, i.e. it contains powers of z from $-\delta nb_0$ to δnb_0. Looking at the negative powers of z in (7.132) we can see that

$$\deg\ \boldsymbol{s}(z^{-1}) = \ell - \delta nb_0. \tag{7.133}$$

However, the optimal equalizer must have $\deg\ \boldsymbol{s}(z^{-1}) \geq \ell$, otherwise the estimate $\hat{d}(t-\ell)$ does not make use of all samples of $\boldsymbol{y}(t)$ that are affected by $d(t-\ell)$, namely $\boldsymbol{y}(t-\ell+nb_0)$, $\boldsymbol{y}(t-\ell+nb_0-1)$,..., $\boldsymbol{y}(t-\ell)$.

We can thus conclude that in order for the coupled polynomial FIR-DFE design equations (7.131) and (7.132) to have a solution, we must have $\delta nb_0 = 0$ and the double sided matrix polynomial, $\bar{E}\left[\Delta\boldsymbol{b}_0(z^{-1})\Delta\boldsymbol{b}_0(z)\right]$, has to be a constant matrix. This means that the uncertainties in the vector channel taps need to be uncorrelated.

However, with the assumption in (7.121) we have

$$\bar{E}[\Delta\boldsymbol{b}(z^{-1})\Delta\boldsymbol{b}^H(z)] + \boldsymbol{R_{nn}} = \frac{1}{nb+1}\Delta\boldsymbol{B}_0\Delta\boldsymbol{B}_0^H + \boldsymbol{R_{nn}} \tag{7.134}$$

and we can thus replace $\bar{E}[\Delta\boldsymbol{b}(z^{-1})\Delta\boldsymbol{b}^H(z)] + \boldsymbol{R_{nn}}$ with the constant matrix

$$\boldsymbol{P}_0 = \frac{1}{nb+1}\Delta\boldsymbol{B}_0\Delta\boldsymbol{B}_0^H + \boldsymbol{R_{nn}}. \tag{7.135}$$

Looking at the positive powers of z in (7.130) we can deduce that

$$\deg\, l_1(z^{-1}) = \ell. \tag{7.136}$$

Furthermore, since we have concluded that $\delta nb_0 = 0$ and thus $\deg\, \boldsymbol{s}(z^{-1}) = \ell$, we can conclude from (7.130), by looking at the negative powers of z, that

$$\deg\, Q(z^{-1}) = nb_0 - 1. \tag{7.137}$$

Finally, using $\deg\, l_1(z^{-1}) = \ell$ in (7.132) gives

$$\deg\, \boldsymbol{l}_2(z^{-1}) = nb_0 - 1. \tag{7.138}$$

We thus have the following orders of the polynomials

$$\deg\, \boldsymbol{s}(z^{-1}) = \deg\, l_1(z^{-1}) = \ell \tag{7.139}$$

$$\deg\, Q(z^{-1}) = \deg\, \boldsymbol{l}_2(z^{-1}) = nb_0 - 1. \tag{7.140}$$

Again, similar to the treatment of the DFE's in Appendix 3.A.1 and 3.A.2, the coupled polynomial equations (7.131) and (7.132) can be solved by converting them to a system of linear equations in the coefficients.

Define

$$\boldsymbol{\mathcal{B}} \triangleq \begin{pmatrix} \boldsymbol{b}_0^T & 0 & \ldots & 0 \\ \boldsymbol{b}_1^T & \boldsymbol{b}_0^T & \ldots & 0 \\ \ldots & \ldots & \ldots & \ldots \\ \boldsymbol{b}_\ell^T & \boldsymbol{b}_{\ell-1}^T & \ldots & \boldsymbol{b}_0^T \end{pmatrix} \tag{7.141}$$

$$\boldsymbol{\mathcal{P}}_0 \triangleq \begin{pmatrix} \boldsymbol{P}_0 & 0 & \cdots & \cdots & 0 \\ 0 & \boldsymbol{P}_0 & 0 & \cdots & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & \cdots & \cdots & 0 & \boldsymbol{P}_0 \end{pmatrix} \tag{7.142}$$

$$\boldsymbol{\mathcal{S}}^T \triangleq \begin{pmatrix} \boldsymbol{s}_0 & \boldsymbol{s}_1 & \ldots & \boldsymbol{s}_\ell & l_{1,\ell}^* & l_{1,\ell-1}^* & \ldots & l_{1,0}^* \end{pmatrix} \tag{7.143}$$

$$\boldsymbol{\mathcal{C}}^T \triangleq \begin{pmatrix} 0 & 0 & \ldots & 1 & 0 & 0 & \ldots & 0 \end{pmatrix} \tag{7.144}$$

where the "1" in $\boldsymbol{\mathcal{C}}$ appears in element nr $\ell+1$. It can now be be shown from (7.131) and (7.132), along the same lines as in Appendix 3.A.1 and 3.A.2, that

$$\boldsymbol{\mathcal{S}} = \begin{pmatrix} \boldsymbol{\mathcal{B}} & \boldsymbol{I} \\ \boldsymbol{\mathcal{P}}_0 & -\boldsymbol{\mathcal{B}}^H \end{pmatrix}^{-1} \boldsymbol{\mathcal{C}} \tag{7.145}$$

where $\boldsymbol{I}$ is an identity matrix. Let us at this point exchange the complex variable z for the shift operator q in the notation. Whenever the matrix in (7.145) is invertible, we can extract $\boldsymbol{s}(q^{-1})$ from $\boldsymbol{\mathcal{S}}$ (and $l_1(q^{-1})$) and subsequently $Q(q^{-1})$ can be obtained from equation (7.130) as

$$Q(q^{-1}) = q^{\ell+1}\boldsymbol{s}(q^{-1})\boldsymbol{b}_0(q^{-1}) + q(l_1^H(q) - 1). \tag{7.146}$$

The robust DFE for the channel (7.55) with FIR channels for the desired signal and co-channel interferers without temporal color can thus, under the assumption (7.121), be realized with FIR filters in the feedforward and the feedback links. Furthermore, compared to the nominal design, the spatial covariance matrix for the noise plus interference, $\boldsymbol{R_{nn}}$, should be replaced with a constant matrix $\boldsymbol{P}_0$ given by

$$\boldsymbol{P}_0 = \boldsymbol{R_{nn}} + \frac{1}{nb+1}\Delta\boldsymbol{B}_0\Delta\boldsymbol{B}_0^H. \tag{7.147}$$

The estimate of the spatial color of the noise should thus be replaced by the sum of the spatial color of the noise and interference and the spatial color of the uncertainties in the estimated channel.

7.A.3 Expectation of Channel Error Spectrum

If the taps in $\Delta\boldsymbol{b}_0(q^{-1})$ are uncorrelated, then $\bar{E}[\Delta\boldsymbol{b}_0(q^{-1})\Delta\boldsymbol{b}_0^H(q)]$ will indeed be a constant matrix. Thus consider the contribution, $\Delta\boldsymbol{B}_i$, from interferer i to the matrix representation (7.63) of the channel estimation error, $\Delta\boldsymbol{B}_0$:

$$\Delta\boldsymbol{B}_i \;\triangleq\; \boldsymbol{b}_i\hat{\boldsymbol{R}}_{d_i\boldsymbol{d}_0}\hat{\boldsymbol{R}}^{-1}_{\boldsymbol{d}_0\boldsymbol{d}_0} \tag{7.148}$$

where $\boldsymbol{b}_i$ is the single vector tap of the channel for interferer i, with $\hat{\boldsymbol{R}}_{d_i\boldsymbol{d}_0}$ and $\hat{\boldsymbol{R}}_{\boldsymbol{d}_0\boldsymbol{d}_0}$ defined in (7.61) and (7.59), respectively.

We thus have to consider how the columns, $\Delta\boldsymbol{b}_{ik}$, in

$$\Delta\boldsymbol{B}_i = [\Delta\boldsymbol{b}_{i0} \;\; \dots \;\; \Delta\boldsymbol{b}_{i,nb_k}] \tag{7.149}$$

in (7.148), corresponding to different time lags, are correlated. This correlation is given by

$$\begin{aligned}
&\bar{E}[\Delta\boldsymbol{b}_{ik}\Delta\boldsymbol{b}_{il}^H] = \\
&\quad = \bar{E}[\boldsymbol{b}_i\hat{\boldsymbol{R}}_{d_i\boldsymbol{d}_0}(\hat{\boldsymbol{R}}^{-1}_{\boldsymbol{d}_0\boldsymbol{d}_0})_k(\boldsymbol{b}_i\hat{\boldsymbol{R}}_{d_i\boldsymbol{d}_0}(\hat{\boldsymbol{R}}^{-1}_{\boldsymbol{d}_0\boldsymbol{d}_0})_l)^H] \\
&\quad = \bar{E}[\boldsymbol{b}_i\hat{\boldsymbol{R}}_{d_i\boldsymbol{d}_0}(\hat{\boldsymbol{R}}^{-1}_{\boldsymbol{d}_0\boldsymbol{d}_0})_k(\hat{\boldsymbol{R}}^{-1}_{\boldsymbol{d}_0\boldsymbol{d}_0})_l^H\hat{\boldsymbol{R}}^H_{d_i\boldsymbol{d}_0}\boldsymbol{b}_i^H]
\end{aligned} \tag{7.150}$$

where $(\hat{\boldsymbol{R}}^{-1}_{\boldsymbol{d}_0\boldsymbol{d}_0})_k$ and $(\hat{\boldsymbol{R}}^{-1}_{\boldsymbol{d}_0\boldsymbol{d}_0})_l$ are the k:th and the l:th columns, respectively, of the matrix $\hat{\boldsymbol{R}}^{-1}_{\boldsymbol{d}_0\boldsymbol{d}_0}$. We can simplify this equation if we assume that the training sequence for the desired signal is chosen such that it is approximately white. There are good reasons for such a choice. It will for example make the channel estimation more well conditioned and it will also facilitate the synchronization of the system. For example, the training sequences in GSM are chosen such that they are approximately uncorrelated over the number of lags necessary to make $\hat{\boldsymbol{R}}_{\boldsymbol{d}_0\boldsymbol{d}_0}$ close to an identity matrix [89]. Thus, assuming that $\hat{\boldsymbol{R}}_{\boldsymbol{d}_0\boldsymbol{d}_0} \approx \boldsymbol{I}$ we obtain

$$\begin{aligned}
&\bar{E}[\Delta\boldsymbol{b}_{ik}\Delta\boldsymbol{b}_{il}^H] \approx \\
&\approx \frac{1}{(N-nb_0)^2}\bar{E}[\boldsymbol{b}_i\sum_{t_1,t_2=nb_0+1}^{N} d_i(t_1)d_0^H(t_1-k)d_0(t_2-l)d_i^H(t_2)\boldsymbol{b}_i^H].
\end{aligned} \tag{7.151}$$

With the training sequence being approximately white, the terms in the sum in (7.151) will only contribute significantly if $t_1 - k = t_2 - l$. We can thus replace t_2 with $t_2 = t_1 - k + l$, giving

$$\bar{E}[\Delta \boldsymbol{b}_{ik} \Delta \boldsymbol{b}_{il}^H] \approx$$
$$\approx \frac{1}{(N - nb_0)^2} \bar{E}[\boldsymbol{b}_i \sum_{t_1 = nb_0 + 1}^{N} d_i(t_1) d_0^H(t_1 - k) d_0(t_1 - k) d_i^H(t_1 - k + l) \boldsymbol{b}_i^H]. \tag{7.152}$$

We will also assume that the symbol sequence of the interferers are white and thus only terms with $k = l$ will contribute significantly. Therefore we approximately have

$$\bar{E}[\Delta \boldsymbol{b}_{ik} \Delta \boldsymbol{b}_{il}^H] \approx 0 \quad \text{when } k \neq l \tag{7.153}$$

and thus the errors in the different channel taps of the channel estimate, originating from a specific interferer, are approximately uncorrelated. Consequently the errors in the different channel taps of the channel estimate from the combined effect of all interferers (and the temporally white thermal noise) are approximately uncorrelated.

The cross-terms between different taps in the expectation $\bar{E}[\Delta \boldsymbol{b}_0(q^{-1}) \Delta \boldsymbol{b}_0^H(q)]$ will thus approximately cancel and we obtain the approximation

$$E[\Delta \boldsymbol{b}_0(q^{-1}) \Delta \boldsymbol{b}_0^H(q)] \approx \frac{1}{nb + 1} \Delta \boldsymbol{B}_0 \boldsymbol{B}_0^H \tag{7.154}$$

Bibliography

[1] L. Acar and R. T. Compton, “”The performance of LMS adaptive array with frequency hopped signals,” *IEEE Trans. Aerospace Electron. Syst.*, vol. AES-21, pp. 360–370, May 1985.

[2] A. Ahlén and M. Sternad, “Optimal deconvolution based on polynomial methods,” *IEEE Transactions on Acoustics, Speech and Signal Processing*, vol. 37, pp. 217–26, 1989.

[3] A. Ahlén and M. Sternad, “Wiener filter design using polynomial equations,” *IEEE Transactions on Signal Processing*, vol. 39, pp. 2387–2399, 1991.

[4] S. Anderson, M Millnert, M. Viberg, and B. Wahlberg, “An adaptive array for mobile communication systems,” *IEEE Transactions on Vehicular Technology*, vol. 40, no. 1, pp. 230–236, February 1991.

[5] S. Andersson, U. Forssén, J. Karlsson, T. Witzschel, P. Fischer, and A. Krug, “Ericsson/Mannesmann GSM field-trials with adaptive antennas,” in *Proceedings of the IEEE Vehicular Technology Conference*, vol. 3, Phoenix, USA, May 1997, pp. 1587–1591.

[6] D. Asztely. “On antenna arrays in mobile communication systems, fast fading and GSM base station receiver algorithms,”, Master’s thesis, Royal Institute of Technology, Department of Signals, Sensors and Systems, Stockholm, Sweden, February 1995.

[7] D. Asztely and B. Ottersten, “MLSE and spatio-temporal interference rejection combining with antenna arrays,” in *Proceedings of EUSIPCO’98*, Rhodos, Greece, September 8-11 1998.

[8] P. Balaban and J. Salz, “Optimum diversity combining and equalization in digital data transmission with applications to cellular mobile

radio – part i: Theoretical considerations," *IEEE Transactions on Communications*, vol. 40, no. 5, pp. 885–94, May 1992.

[9] P. Balaban and J. Salz, "Optimum diversity combining and equalization in digital data transmission with applications to cellular mobile radio – part ii: Numerical results," *IEEE Transactions on Communications*, vol. 40, no. 5, pp. 895–907, May 1992.

[10] B.D.Veen and K.M. Buckley, "Beamforming: A versitile approach to spatial filtering," *IEEE ASSP Magazine*, pp. 4–24, April 1988.

[11] S. Bigi, "On the use of system identification for design purposes and parameter estimation,", Licentiat thesis, Uppsala University, Uppsala, Sweden, 1995.

[12] B. Bjerke. "A comparison of low complexity receivers for TDMA mobile wireless communications,", Master's thesis, Northeastern University, Boston, MA, USA, December 1997.

[13] G.E. Bottomley and K. Jamal, "Adaptive arrays and MLSE equalization," in *Proceedings of 45th Vehicular Technology Conference*, vol. 1, Chicago, Illinois, U.S.A., 1995, pp. 50–54.

[14] Richard L. Burden and J. Douglas Faires, *Numerical Analysis.* Boston, MA, USA.: PWS-KENT Publishing Company, 1989.

[15] M. Cedervall and R. Moses, "Decoupled maximum likelihood angle estimation for coherent signals," in *29th Asilomar Conference on Signals, Systems and Computers*, Pacific Grove, CA, U.S.A., Oct 30 - Nov 1 1995.

[16] R. C. Chung and P. R. Bélanger, "Minimum–sensitivity filter for linear time–invariant stochastic systems with uncertain parameters," *IEEE Transactions on Automatic Control*, vol. 21, no. 1, pp. 98–100, February 1976.

[17] M. V. Clark, L. J. Greenstein, W. K. Kennedy, and M. Shafi, "Matched filter performance bounds for diversity combining in digital mobile radio," in *Globecom*, 1991.

[18] M. V. Clark, L. J. Greenstein, W. K. Kennedy, and M. S. Shafi, "Optimum linear diversity receivers for mobile communications," *IEEE Transactions on Vehicular Technology*, vol. 43, no. 1, pp. 47–56, February 1994.

[19] R.T. JR. Compton, *Adaptive Antennas.* New Jersey: Prentice-Hall, Inc, 1988.

[20] Zhi Ding, "Multipath channel identification based on partial system information," *IEEE Transactions on Signal Processing*, pp. 235–240, January 1997.

[21] J. F. Diouris, M. Diop, and J. Saillard, "Digital generation of a reference signal for adaptive arrays applications," in *42nd IEEE Vehicular Technology Conference*, vol. 1, Denver, Colorado, May 10-13 1992, pp. 17–20.

[22] J-F. Diouris, B Feuvrie, and J. Saillard, "Adaptive multisensor receiver for mobile communications," *Ann. Télécommun*, vol. 48, no. 1-2, pp. 35–46, 1993.

[23] A. Duel–Hallen, "Equalizers for multiple input/multiple output channels and PAM systems with cyclostationary input sequences," *IEEE Journal on Selected Areas in Communications*, vol. 10, no. 3, pp. 630–639, April 1992.

[24] ETSI/PT 12, *GSM Technical Specification 05.04: Modulation*, Version 3.1.2.

[25] D.D. Falconer, M. Abdulraham, N.W.K. Lo, B.R. Petersen, and A.U.H. Sheik, "Advances in equalization and diversity for portable wireless systems," *Digital Signal Processing*, vol. 3, pp. 148–62, 1993.

[26] G.D. Forney, "Maximum-likelihood sequence estimation of digital sequences in the presence of intersymbol interference," *IEEE Transactions on Information Theory*, vol. 18, no. 3, pp. 363–378, May 1972.

[27] G.J. Foschini and M.J. Gans, "On limits of wireless communications in a fading environment when using multiple antennas," *Wireless Personal Communications*, vol. 6, no. 3, pp. 311–?, March 1998.

[28] D. Gerlach. *Adaptive Transmitting Antenna Arrays at the Base Station in Mobile Radio Networks*, PhD thesis, Stanford University, Dept. of Electrical Engineering, Stanford, CA, USA, 1995.

[29] D. Gerlach, "Transmit antenna beamforming for the advanced mobile phone system," in *Proceedings of 29th Asilomar Conference on Signals, Systems & Computers*, Pacific Grove, California, U.S.A., October 30 - November 1 1995.

[30] D. Gerlach and A. Paulraj, "Adaptive transmitting antenna methods for multipath environments," in *Proceedings of Globecom'94*, 1994.

[31] G.D. Golden, G.J. Foschini, R.A. Valenzuela, and P.W. Wolniansky, "Detection algorithm and initial laboratory results using the v-blast space-time communication architecture," *Electronic Letters*, vol. 35, no. 1, pp. 14–15, January 1999.

[32] G. C. Goodwin, M. Gevers, and B. Ninnes, "Quantifying the error in estimated transfer functions with application to model order selection," *IEEE Transactions on Automatic Control*, vol. 37, no. 7, pp. 913–928, July 1992.

[33] G. C. Goodwin and M. E. Salgado, "A stochastic embedding approach for quantifying uncertainty in the estimation of restricted complexity models," *International Journal of Adaptive Control and Signal Processing*, vol. 3, no. 4, pp. 333–356, December 1989.

[34] B. Hagerman, "Downlink relative co-channel interference powers in cellular radio systems," in *Proceedings of VTC'95*, vol. 2, Rosemont, IL, USA, 1995, pp. 366–370.

[35] Monson H. Hayes, *Statistical Digital Signal Processing and Modeling*: John Wiley and Sons, inc., 1996.

[36] Didier Henrion, Martin Hromcik, Huibert Kwakernaak, and Sonja Pejchová, "Polynomial toolbox version 2.0,", http://www.math.utwente.nl/polbox, 1998, Toolbox for Matlab.

[37] K. Öhrn. *Design of Multivariable Cautious Discrete-time Wiener Filters*, PhD thesis, Uppsala University, Signals and Systems, Uppsala, Sweden, 1996.

[38] K. Öhrn, A. Ahlén, and M. Sternad, "A probabilistic approach to multivariable robust filtering and open-loop control," *IEEE Transactions on Automatic Control*, vol. 40, pp. 405–418, 1995.

[39] Y. C. Jenq, "A class of generalized decision feedback equalizers," *Journal of the Franklin Institute*, vol. 306, no. 6, pp. 513–22, December 1979.

[40] J. Jezěk and V. Kŭcera, "Efficient algorithms for matrix spectral factorization," *Automatica*, vol. 21, pp. 663–669, 1985.

[41] P. Jung, B. Steiner, and Y. Ma, "Maximum-likelihood detector for coherent receiver antena diversity," *Frequenz*, vol. 48, no. 3-4, pp. 94–99, 1994.

[42] J. Karlsson. "Adaptive antennas in GSM systems with non-synchronous base stations,", Master's thesis, Royal Institute of Technology, Stockholm, Sweden, March 1998.

[43] A.S. Kayrallah, R. Ramésh, G.E. Bottomley, and D. Koilpillai, "Improved channel estimation with side information," in *Proceedings of VTC'97*, vol. 2, Phoenix, AZ, USA, May 4-7 1997, pp. 1049–1051.

[44] A.V. Keerthi and J. J. Shynk, "Separation of cochannel signals in TDMA mobile radio," *IEEE Transactions on Signal Processing*, vol. 10, no. 46, pp. 2684–2697, October 1998.

[45] R. Krenz and K. Wesolowski, "Comparison of several space diversity techniques for MLSE receivers in mobile communications," in *Proceedings of PIMRC'94*, vol. 2, 1994.

[46] H. Krim and M. Viberg, "Two decades of array signal processing research: The parametric approach," *IEEE Signal Procesing Magazine*, pp. 67–94, July 1996.

[47] V. Kŭcera, "Factorization of rational spectral matrices: a survey of methods," *Preprints, IEE Control'91*, pp. 1074–1078, March 1991.

[48] P.A. Laurent, "Exact and approximative construction of digital phase modulations by superposition of amplitude modulated pulses (AMP)," *IEEE Transactions on Communications*, vol. COM-14, no. 2, pp. 150–160, February 1986.

[49] H.-N. Lee and G.J. Pottie, "Channel estimation based adaptive equalization/diversity combining for time-varying dispersive channels," in *Proceedings of VTC'97*, vol. 2, Phoenix, AZ, USA, May 4-7 1997, pp. 884–888.

[50] H.-N. Lee and G.J. Pottie, "Fast adaptive equalization/diversity combining for time-varying dispersive channels," *IEEE Transactions on Communications*, vol. 46, no. 9, pp. 1146–1162, September 1998.

[51] W.C. Lee, *Mobile Cellular Telecommunications*. New York: McGraw-Hill, 1995.

[52] J. Li, B. Halder, P. Stoica, and M. Viberg, "Computationally efficient angle estimation for signals with known waveforms," *IEEE Transactions on Signal Processing*, vol. 43, pp. 2154–2164, 1995.

[53] J.-W. Liang, J.-T. Chen, and A. Paulraj, "A two-stage hybrid approach for CCI/ISI reduction with space-time processing," *IEEE Communications Letters*, vol. 1, no. 6, pp. 163–165, November 1997.

[54] L. Lindbom. *A Wiener filtering approach to the design of tracking algorithms - With applications in mobile radio communications*, PhD thesis, Uppsala University, Signals Processing Group, Uppsala, Sweden, 1995.

[55] E. Lindskog, "Combatting co-channel interference in a TDMA system using interference estimates from adjacent frames," in *Proceedings of 29th Asilomar Conference on Signals, Systems & Computers*, Pacific Grove, California, U.S.A., October 30 - November 1 1995, pp. 367–371.

[56] E. Lindskog, "Making SMI-beamforming insensitive to the sampling timing for GSM signals," in *Proceedings of the Sixth International Symposium on Personal, Indoor and Mobile Radio Communications*, Toronto, Canada, 27-29 September 1995, pp. 664–668.

[57] E. Lindskog, "Array channel identification using directional of arrival parametrization," in *Proceedings of IEEE International Conference on Universal Personal Communications*, Cambridge, Massachusetts, U.S.A., September 29 - October 2 1996, pp. 999–1003.

[58] E. Lindskog, "Channel estimation exploiting pulse shaping information - a channel interpolation approach,", Technical Report UPTEC 97138R, Uppsala University, Signal Processing Group, PO Box 528, SE-751 20 Uppsala, Sweden, 1997, See http://www.signal.uu.se/Publications/preports.html.

[59] E. Lindskog, "Multi-channel maximum likelihood sequence estimation," in *Proceedings of the 47th IEEE Vehicular Technology Conference*, vol. 2, Phoenix, Arizona, USA, May 5-7 1997, pp. 715–719.

[60] E. Lindskog, A. Ahlén, and M. Sternad, "Combined spatial and temporal equalization using an adaptive antenna array and a decision feedback equalization scheme," in *Proceedings of Int. Conf. on Acoustics, Speech and Signal Processing*, Detroit, Michigan, U.S.A, May 8-12 1995.

[61] E. Lindskog, A. Ahlén, and M. Sternad, "Spatio-temporal equalization for multipath environments in mobile radio applications," in *Proceedings of the 45th IEEE Vehicular Technology Conference*, Rosemont, Illinois, USA, July 26-69 1995, pp. 399–403.

[62] E. Lindskog, M. Sternad, and A. Ahlén, "Designing decision feedback equalizers to be robust with respect to channel time variations," in *Proceedings of Nordic Radio Symposium seminar*, Uppsala, Sweden, November 10-11 1993.

[63] E. Lindskog, M. Sternad, and A. Ahlén, "Combined spatial and temporal equalization for mobile radio applications," in *Proceedings of Nordic Radio Symposium seminar*, Linköping, Sweden, October 26-27 1994, pp. 7–10.

[64] E. Lindskog and J. Strandell, "Multi-user channel estimation exploiting pulse shaping information," in *Proceedings of EUSIPCO'98*, Rhodos, Greece, September 8-11 1998.

[65] E. Lindskog and C. Tidestav, "Reduced rank equalization," in *Proceedings of PIMRC'98*, Boston, Massachusetts, USA, September 8-11 1998, pp. 1081–85.

[66] E. Lindskog and C. Tidestav, "Reduced rank channel estimation," in *Proceedings of VTC'99*, Houston, Texas, USA, May 17-21 1999.

[67] N.W.K. Lo, D.D. Falconer, and A.U.H. Sheik, "Adaptive equalization and diversity combining for mobile radio using interpolated channel estimates," *IEEE Transactions on Vehicular Technology*, vol. 40, no. 3, pp. 636–45, August 1991.

[68] 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, Jan. 1989.

[69] U. Madhow and M. Honig, "MMSE interference suppression for direct-sequence spread-spectrum CDMA," *IEEE Trans. Communications*, vol. 42, no. 12, pp. 3178–3188, December 1994.

[70] S. Miller and S. Schwartz, "Integrated spatial-temporal detectors for asynchronous Gaussian multiple access channels," *IEEE Transactions on Communications*, vol. 43, no. 2/3/4, pp. 396–411, February/March/April 1995.

[71] J.W. Modestino and V.M. Eyuboglu, "Integrated multielement receiver structures for spatially distributed interference channels," *IEEE Transactions on Information Theory*, vol. 32, no. 2, pp. 195–219, March 1986.

[72] M. Nagatsuka and R. Kohno, "A spatially and temporally optimal multi-user receiver using an array antenna for DS/CDMA," *IEICE Transactions on communications*, vol. E78-B, no. 11, pp. 1489–1497, November 1995.

[73] A. F. Naguib and A. Paulraj, "Performance of wireless CDMA with M-ary orthogonal modulation and cell site antenna arrays," *Journal on Selected Areas in Communication*, vol. 14, no. 9, pp. 1770–1783, December 1996.

[74] A.F. Naguib, B. Khalaj, A. Paulraj, and T. Kailath, "Adaptive channel equalization for TDMA digital cellular communications using antenna arrays," in *Procedings of International Conference on Acoustics, Speech and Signal Processing*, 1994.

[75] B. Chong Ng and M. Cedervall, "A structured channel estimator for maximum likelihood sequence detection in multipath fading channels," *IEEE Communication Letters*, vol. 1, no. 2, pp. 52–55, March 1997.

[76] Y. Ogawa, Y. Nagashima, and K. Itoh, "An adaptive antenna system for high-speed digital mobile communications," *IEEE Transactions on Communications*, vol. E75-B, no. 5, pp. 413–21, May 1992.

[77] Bjorn Ottersten, "Array processing for wireless communications," in *8th IEEE Signal Processing Workshop on Statistical Signal and Array Processing*, Corfu, June 24-26 1996, pp. 466–473.

[78] A. Paulraj and T Kailath, "Increasing capacity in wireless broadcast systems using distributed transmission and directional reception,", US Patent 5,345,599.

[79] A. Paulraj and E. Lindskog, "A taxonomy of space-time processing for wireless networks," in *Proceedings of COMMSPHERE'97*, Lausanne, Switzerland, 1997.

[80] A. Paulraj and E. Lindskog, "A taxonomy of space-time processing for wireless networks," *IEE Proceedings, Radar, Sonar and Navigation*, vol. 145, no. 1, pp. 25–31, February 1998.

[81] A.J. Paulraj and C.B. Papadias, "Space-time processing for wireless communications," *Signal Processing Magazine*, November 1997.

[82] J.G. Proakis, *Digital Communications*: McGraw-Hill Book Company, Singapore, 1989.

[83] P.Vila, F. Pipon, D. Pirez, and L. Féty, "MMSE antenna diversity equalization of a jammed frequency-selective fading channel," in *Proceedings of International Conference of Acoustics, Speach and Signal Processing*, Detroit, Michigan, U.S.A., 1995, pp. 1872–1875.

[84] G. Raleigh, S. Diggavi, V. Jones, and A. Paulraj, "A blind adaptive transmit antenna algorithm for wireless communication," in *Proc. ICC*, 1995.

[85] J. Salz and J.H. Winters, "Effect of fading correlation on adaptive arrays in digital wreless communications," in *IEEE International Conference on Communications '93*, vol. 3, Geneva, Switzerland, 1993, pp. 1768–74.

[86] S. Sampei, *Applications of Digital Wireless Technologies to Global Wireless Communications*: Prentice Hall, 1997.

[87] T. Söderstrom and P. Stoica, *System Identification*: Prentice Hall, 1989.

[88] S. Simanpalli, "Adaptive array methods for mobile communications," in *Proceedings of Vehicular Technology Conference*, vol. 3, Stockholm, Sweden, June 8-10 1994, pp. 1503–06.

[89] R. Steele, *Mobile Radio Communications*: Pentech Press Limited, London, 1992.

[90] M. Sternad and A. Ahlén, "The structure and design of realizable decision feedback equalizers for IIR channels with colored noise," *IEEE Transactions on Information Theory*, vol. 36, pp. 848–858, 1990.

[91] M. Sternad and A. Ahlén, "Robust filtering and feedforward control based on probabilistic descriptions of model errors," *Automatica*, vol. 29, no. 3, pp. 661–79, May 1993.

[92] M. Sternad, A. Ahlén, and E. Lindskog, "Robust decision feedback equalizers," in *Proceedings of Int. Conf. on Acoustics, Speech and Signal Processing*, vol. 3, Minneapolis, Minnesota, U.S.A, April 1993, pp. 555–58.

[93] M. Sternad, K. Öhrn, and A. Ahlén, “Robust H_2 filtering for structured uncertainty: the performance of probabillistic and minimax schemes,” in *European Control Conference*, Rome, Italy, September 1995, pp. 87–92.

[94] P. Stoica and K.C. Sharman, “Maximum likelihood methods for direction-of-arrival estimation,” *IEEE Transactions on Acoustics, Speech and Signal Processing*, vol. 38, pp. 1132–43, July 1990.

[95] P. Stoica and K.C. Sharman, “Novel eigenanalysis method for direction estimation,” *IEE Procedings, Part F*, vol. 137, pp. 19–26, February 1990.

[96] P. Stoica and M. Viberg, “Maximum likelihood rank estimation in reduced-rank multivariate linear regressions,” *IEEE Transactions on Signal Processing*, vol. 44, no. 12, pp. 3069–3078, December 1996.

[97] M. Stojanovic, J. Catipovic, and J.G. Proakis, “Adaptive multichannel combining and equalization for underwater acoustic communications,” *Journal of the Acoustical Soceity of America*, vol. 94, no. 3, pp. 1621–1631, September 1993.

[98] M. Stojanovic, J. A. Catipovic, and J. G. Proakis, “Reduced-complexity simultaneous beamforming and equalization for underwater acoustic communications,” in *Proceedings of Oceans'93*, vol. III, IEEE, 1993.

[99] M. Stojanovic, J.G. Proakis, and J.A. Catipovic, “Spatial signal processing for time-dispersive digital communication channels,” in *SYSID*, vol. 3, Copenhagen Denmark, 1994, pp. 301–6.

[100] M. Stojanovic and Z. Zvonar, “Multichannel processing of broadband multiuser communication signals in shallow water acoustic channels,” *IEEE J. Oceanic Eng*, pp. 156–166, April 1996.

[101] J. Strandell and E. Lindskog, “Channel estimation by maximum likelihood projection onto a parametrized subspace,” in *Proceedings of EUSIPCO'98*, Rhodos, Greece, September 8-11 1998.

[102] B. Suard, A. Naguib, G. Xu, and A. Paulraj, “Performance analysis of CDMA mobile communication systems using antenna arrays,” in *Proc. ICASSP'93*, vol. VI, Minneapolis, MN, April 1993, pp. 153–156.

[103] S. C. Swales, M. Beach, D. Edwards, and J. P. McGeehan, "The performance enhancement of multibeam adaptive base-station antennas for cellular land mobile radio systems," *IEEE Transactions on Vehicular Technology*, vol. 39, no. 1, pp. 56–67, Feb. 1990.

[104] S.C. Swales, M.A. Beach, and D.J. Edwards, "Adaptive antenna arrays for future generation cellular communications networks," in *Sixth International Conference on Antenna and Propagation*, vol. 1, IEE, London, UK, 1989, pp. 321–325.

[105] S. Talwar, M. Viberg, and A. Paulraj, "Blind separation of synchronous co-channel digital signals using an antenna array. Part I. Algorithms," *IEEE Transactions on Signal Processing*, vol. 44, no. 5, pp. 1184–1197, May 1996.

[106] C. Tidestav, A. Ahlén, and M. Sternad, "Realizable MIMO decision feedback equalizers: Structure and design," *IEEE Transactions on Signal Processing*, Submitted.

[107] C. Tidestav, A. Ahlén, and M. Sternad, "Narrowband and broadband multiuser detection using a multivariable DFE," in *Proceedings of PIMRC'95*, Toronto, Canada, September 27-29 1995, pp. 732–736.

[108] C. Tidestav, A. Ahlén, and M. Sternad, "Realizable MIMO decision feedback equalizers: Structure and design,", Technical report, Uppsala University, Signals and Systems, PO Box 528, SE-751 20 Uppsala, Sweden, 1998.

[109] C. Tidestav and E. Lindskog, "Bootstrap equalization," in *Proceedings of ICUPC'98*, Florence, Italy, October 5-9 1998.

[110] C. Tidestav, M. Sternad, and A. Ahlén, "Reuse within a cell - interference rejection or multiuser detection," *IEEE Transactions on Communications*, To appear.

[111] Lang Tong, G. Xu, and T. Kailath, "Fast blind equalization via antenna arrays," in *Procedings of International Conference on Acoustics, Speech and Signal Processing*, vol. 4, Minneapolis, Minnesota, USA, April 27-30 1993, pp. 272–75.

[112] V. Torakh, A. Nabguib, N. Seshadri, and A. R. Calderbank, "Space-time coding for high data rate wireless communication: combined array processing and space-time coding,", Preprint.

[113] V. Torakh, A. Nagiub, N. Seshadri, and A. R. Calderbank, "Space-time coding for high data rate wireless communication: Practical considerations,", Submitted to *IEEE Trans. on Communications.*

[114] V. Torakh, N. Seshadri, and A. R. Calderbank, "Space-time codes for high data rate wireless communication: performance criterion and code construction,", Submitted to *IEEE Trans. on Information Theory.*

[115] T. Trump and B. Ottersten, "Estimation of nominal direction of arrival and angular spread using an array of sensors," *Signal Processing*, vol. 50, pp. 57–69, 1996.

[116] G. Ungerboeck, "Adaptive maximum-likelihood receiver for carrier-modulated data-transmission systems," *IEEE Transactions on Communications*, vol. 22, no. 5, pp. 624–636, May 1974.

[117] B. D. Van Veen and K. M. Buckley, "Beamforming: a versatile approach to spatial filtering," *IEEE Acoustics, Speech and Signal Processing Magazine*, pp. 4–24, April 1988.

[118] S. Verdú, "Minimum probability of error for asynchronous Gaussian multiple-access channels," *IEEE Trans. on Information Theory*, vol. IT-32, pp. 85–96, Jan. 1986.

[119] M. Viberg and B. Ottersten, "Sensor array processing based on subspace fitting," *IEEE Transactions on Signal Processing*, vol. 39, no. 5, pp. 1110–1121, May 1991.

[120] M. Viberg, B. Ottersten, and T. Kailath, "Detection and estimation in sensor arrays using weighted subspace fitting," *IEEE Transactions on Signal Processing*, vol. 39, no. 11, pp. 2436–2449, Nov. 1991.

[121] P. Vila, F. Pipon, D. Pirez, and L. Féty, "MLSE antenna diversity equalization of a jammed frequency-selective fading channel," in *Proceedings of EUSIPCO'94*, Edinburg, UK, 1994, pp. 1516–1519.

[122] A.J. Viterbi, "Error bounds for convolutional codes and an asymptotically optimum decoding algorithm," *IEEE Transactions on Information Theory*, vol. 13, pp. 260–269, April 1967.

[123] R.K. Wangsness, *Electromagnetic fields.* New York, NY, USA.: John Wiley & Sons, Inc., 1986.

[124] J. Ward and R. T. Compton, Jr., "High throughput slotted ALOHA packet radio networks with adaptive arrays," *IEEE Transactions on Communications*, vol. 41, no. 3, pp. 460–470, March 1993.

[125] M. Wax and I. Ziskind, "On unique localozation of multiple sources by passive sensor arrays," *IEEE Transactions on Acoustics, Speech, and Signal Processing*, vol. 37, pp. 996–1000, July 1989.

[126] Jr. William C Jakes, editor, *Microwave Mobile Communications*. New York, NY: Wiley, 1974.

[127] J. Winters, "On the capacity of radio communication systems with diversity in a Rayleigh fading environment," *IEEE Journal on Selected Areas in Communications*, vol. 5, no. 5, pp. 871–878, June 1987.

[128] J. H. Winters, "Optimum combining in digital mobile radio with cochannel interference," *IEEE Journal on Selected Areas in Communications*, vol. SAC-2, no. 4, pp. 528–39, July 1984.

[129] J. H. Winters, "Optimum combining for indoor radio systems with multiple users," *IEEE Transactions on Communications*, vol. COM-35, no. 11, pp. 1222–30, November 1987.

[130] Jack H. Winters, "Signal acquisition and tracking with adaptive arrays in wireless systems," in *Proc. 43rd Vehicular Technology Conf.*, vol. I, November 1993.

[131] P. W. Wolniansky, G. J. Foschini, G. D. Golden, and R. A. Valenzuela, "V-blast: An architecture for realizing very high data rates over the rich-scattering wireless channel," in *ISSSE-98*, Pisa, Italy, Sept 29 1998, See www.bell-labs.com/projects/blast/.

[132] G. Xu and H. Liu, "An effective transmission beamforming scheme for frequency-division-duplex digital wireless communication systems," in *Proc. ICASSP-95*, Detroit, MI, 1995, pp. 1729–1732.

[133] Y-S Yeh and D. O. Reudink, "Efficient spectrum utilization for mobile radio systems using space diversity," *IEEE Trans. Communications*, vol. COM-30, no. 3, pp. 447–455, March 1982.

[134] H. Yoshino, K. Fukawa, and H. Sukuki, "Interference cancelling equalizer (ice) for mobile radio communication," *IEEE Transactions on Vehicular Technology*, vol. 46, no. 4, pp. 849–861, November 1997.

[135] P. Zetterberg and B. Ottersten, "The spectrum efficiency of a base station antenna array system for spatially selective transmission," *IEEE Trans. on Vehicular Technology*, vol. 44, no. 3, pp. 651–660, August 1995.

[136] M. D. Zoltowski and J. Ramos, "Blind adaptive beamforming for CDMA Based PCS/Cellular," in *Proc. 29th Asilomar Conference on Signals, Systems and Computers*, Pacific Grove, CA, Oct. 31 - Nov. 2 1995.