BLIND ESTIMATION WITHOUT PRIORS: PERFORMANCE, CONVERGENCE, AND EFFICIENT IMPLEMENTATION

A Dissertation

Presented to the Faculty of the Graduate School

of Cornell University

in Partial Fulfillment of the Requirements for the Degree of

Doctor of Philosophy

by

Philip Schniter

May 2000

BLIND ESTIMATION WITHOUT PRIORS: PERFORMANCE, CONVERGENCE, AND EFFICIENT IMPLEMENTATION

Philip Schniter, Ph.D.

Cornell University 2000

We address the problem of estimating a distorted signal in the presence of noise in the case that very little prior knowledge is available regarding the nature of the signal, the distortion, or the noise. More specifically, we assume that the distorting system is linear but otherwise unknown, that the signal is drawn from a sequence of independent and identically distributed random variables of unknown distribution, and that the noise is independent of the signal but also has unknown distribution. We refer to this problem as "blind estimation without priors," where the term "blind" captures the notion that signal estimates are obtained blindly with regard to knowledge of the distortion and interference.

Since its origins nearly half a century ago, there has evolved a large body of theoretical and practical knowledge regarding the blind estimation problem. Even so, very fundamental questions still remain. For example: (i) How good are blind estimates compared to their non-blind counterparts? (ii) When do we know that a blind estimation algorithm will return estimates of the desired signal versus a

component of the interference? Though both of these questions have long histories within the research community, existing results have been either approximate and/or limited to special cases that leave out many problems of practical interest.

This dissertation presents answers to the questions above for the well-known Shalvi-Weinstein (SW) and constant modulus (CM) approaches to blind linear estimation. All results are derived in a general setting: vector-valued infinite impulse response channels, constrained vector-valued auto-regressive moving-average estimators, and near-arbitrary forms of signal and interference. First, we derive concise expressions tightly upper bounding the mean-squared error (MSE) of SW and CM-minimizing estimates which are principally a function of the optimum MSE achievable in the same setting. Second, we derive similar bounds for the average squared parameter error in CM-based blind channel identification. Third, we present sufficient conditions for gradient-descent (GD) initializations that guarantee convergence to CM-minimizing estimates of the desired user. These conditions principally involve the signal-to-interference ratio of the initial estimates. Finally, we propose a novel approach to CM-GD implementation that greatly reduces the implementation complexity of the standard CM-GD algorithm while still retaining its mean behavior.

Biographical Sketch

Philip Schniter was born in Evanston IL in the summer of 1970. He received the B.S. and M.S. degrees in electrical and computer engineering from the University of Illinois at Urbana-Champaign in 1992 and 1993, respectively. From 1993 to 1996 he was employed by Tektronix Inc. in Beaverton OR as a Systems Engineer. There he worked on signal processing aspects of video and communications instrumentation design, including algorithms, software, and hardware architectures.

Since 1996 he has been working toward the Ph.D. degree in electrical engineering at Cornell University, Ithaca, NY, where he received the 1998 Schlumberger Fellowship and the 1998-99 Intel Foundation Fellowship. His research interests are in the areas of signal processing and data communication, with special emphasis on adaptive systems.

To my parents, Jorg and Martha Schniter, for their love and support,

and

in fond memory of David F. C. Simmonds.

Acknowledgements

I've had a phenomenal time during my four years in Ithaca, and I attribute my good fortune to the wonderful people I've come to know through my Cornell experience.

First and foremost, I would like to thank my supervisor Prof. C. Richard Johnson, Jr., for his generosity and support over the years. I will always remember Rick as "the advisor who made everything possible." His friendliness, wisdom, spontaneity, and sense of adventure make him unique in the academic community. Many thanks also go to my committee members, Profs. Lang Tong and Venu Veeravalli; they have provided me with invaluable guidance, encouragement, and inspiration throughout the last four years. These acknowledgements would be far from complete without heartfelt thanks to Profs. Doug Jones and Mike Orchard for their sage advice before and during my studies at Cornell, as well as Prof. Vince Poor and Dr. John Treichler for their genuine friendliness, support, and interest in my work.

Next, I want to thank the colleagues of mine who made Rhodes Hall a thoroughly entertaining, if not always productive, place to work. In order of appearance... Tom "Shotgun" Endres: seeing your John Deere hat during my first campus visit, I knew I belonged in the research group; Raul "Buddy Boy" Casas: a more friendly, helpful, spontaneous, and entertaining officemate cannot be imagined (and I still owe you a super-soaking); Knox Carey: your friendship and Linux wizardry were

huge helps during my first year at Cornell—long live the pedal steel; Rick "Vanilla" Brown: whether it be Linux, latex, math, or the job hunt, you've been generous with your help and plentiful with your (base) humor—its been a fun ride; Wonzoo "The Dragon" Chung: your appetite for knowledge, cross-cultural understanding, and large quantities of (jumping!) hot food has earned my deepest respect; Jaiganesh "Jai" Balakrishnan: your keen insights and kind ways are have been much appreciated; Andy Klein and Don Anair: ahhh, fond memories of rides in the van, walks on the beach, and all the good clean fun we've had; Azzedine Touzni: from our first meeting, your open-mindedness, generosity, and cultured European ways have always impressed me—it was great having you in Ithaca; Prof. Bill "6σ" "signum" Sethares: the late-night discussions, plastic-sax jams, and spaghetti dinners are fondly remembered—thanks for your sage advice and willingness to lend an ear anytime. Thanks also go to the visiting researchers Sangarapillai "Lambo" Lambotharan, Prof. Michael Green, and Jim Behm, for their assistance and companionship.

Indispensable to the completion of my degree were the many hours of ultimate frisbee at Helen-Newman, Belle-Sherman, and Cass Park; summer swimming at the 6-mile Creek reservoir and winter swimming at Teagle; hiking and cross-country skiing in the local hills and forests; playing/hearing live music at the local clubs and festivals; and raging at the countless (yet always memorable) parties. These activities wouldn't have been the same if it wasn't for my good friends Katherine Alaimo, Marisa Alcorta, Kristin Arend, Corien Bakermans, Ami Ben-Yaacov, Francis Chan, Peter Choi, Diane Decker, Peter Dees, Maile and Todd Deppe, Jesse Ernst, Thamora Fishel, Jen Fiskin, Anne Gallagher, Pete Hedlund, Ruth Hufbauer, Bill Huston, Laci, Loden, Tawainga Katsvairo, Frank Adelstein, Ketch and Critter,

Brady Moss, Matt Nemeth, Eric Odell, Kevin Pilz, Deepak Ramani, Mary Reeves, Art Rodgers, Art Schaub, Martin Schlaepfer, Sunjya & Lia Schweig, Kurt Smemo, Bob Turner, Mace Vaughan, Dave Walker, Rob Young, and Brooke Zanatell. Special mention is given to the following dear friends, for all that they have added to my life over the last four years: Tony Anderson, Ruby Beil, Ryan Budney, Heather Clark, and Beth Conrey. Finally, saving the best for last, I want to thank the sweet and wonderful Caroline Simmonds for all the sunshine she's brought to my last year and a half in Ithaca.

My studies were funded by the National Science Foundation (under grants MIP-9509011, MIP-9528363, and ECS-9811297), Applied Signal Technology Inc., the Intel Foundation, and the Schlumberger Foundation.

Table of Contents

List of Tables

List of Figures

List of Abbreviations

Abbreviation	Journal Name
ASSPM	IEEE Acoustics Speech and Signal Processing Magazine
ATT	AT&T Technical Journal
BSTJ	Bell System Technical Journal
GEO	Geoexploration
GP	Geophysical Prospecting
IJACSP	Internat. Journal of Adaptive Control & Signal Processing
OE	Optical Engineering
ETS	Educational Testing Service Research Bulletin
PROC	Proceedings of the IEEE
PSY	Psychometrika
SEP	Stanford Exploration Project
SP	Signal Processing
SPL	IEEE Signal Processing Letters
SPM	IEEE Signal Processing Magazine
TASSP	IEEE Trans. on Acoustics, Speech, and Signal Processing
TCOM	IEEE Trans. on Communications
TIT	IEEE Trans. on Information Theory
TSP	IEEE Trans. on Signal Processing

Abbreviation	Conference Name
ALL	Allerton Conf. on Communication, Control, and Computing
ASIL	Asilomar Conf. on Signals, Systems and Computers
CISS	Conf. on Information Science and Systems
GLOBE	IEEE Global Telecommunications Conf.
ICASSP	IEEE Internat. Conf. on Acoustics, Speech, and Signal Processing
ICC	IEEE Intern. Conf. on Communication
NNSP	IEEE Workshop on Neural Networks for Signal Processing
SPAWC	IEEE Internat. Workshop on Signal Processing Advances in Wireless Communications
SPIE	The Internat. Society for Optical Engineering
WCNC	IEEE Wireless Communication and Networking Conf.
WICASS	Internat. Workshop on Independent Component Analysis and Signal Separation

Abbreviation	Meaning
ARMA	Auto-Regressive Moving-Average
ASPE	Average-Squared Parameter Error
AWGN	Additive White Gaussian Noise
BEWP	Blind Estimation Without Priors
BIBO	Bounded-Input Bounded-Output
BMSE	Bayesian Mean-Squared Error
BPSK	Binary Phase Shift Keying
BSE	Baud-Spaced Equalizer
CDMA	Code Division Multiple Access
CM	Constant Modulus
CMA	Constant Modulus Algorithm
DSE	Dithered Signed Error
EMSE	Excess Mean-Squared Error
FCR	Full Column Rank
FIR	Finite Impulse Response
FSE	Fractionally-Spaced Equalizer
GD	Gradient Descent
HDTV	High Definition Television
HOS	Higher-Order Statistics
i.i.d.	Independent and Identically Distributed
IIR	Infinite Impulse Response
ISI	Inter-Symbol Interference
LMS	Least Mean Square
LTI	Linear Time-Invariant
MAI	Multi-Access Interference
MAP	Maximum A Posteriori
MIMO	Multiple Input Multiple Output
ML	Maximum Likelihood
MMSE	Minimum Mean-Squared Error
MSE	Mean-Squared Error
PAM	Pulse Amplitude Modulation
PBLE	Perfect Blind Linear Estimation
PSD	Positive Semi-Definite
ROC	Region of Convergence
QAM	Quadrature Amplitude Modulation
QPSK	Quadrature Phase Shift Keying
SE	Signed Error
SER	Symbol Error Rate
SINR	Signal to Interference-Plus-Noise Ratio
SISO	Single Input Single Output
SNR	Signal to Noise Ratio
SOS	Second-Order Statistics
SPIB	Signal Processing Information Base[1]
SW	Shalvi-Weinstein

[1] See `http://spib.rice.edu/spib/microwave.html`.

Abbreviation	Meaning
UD	Update Decimated
UMSE	Conditionally-Unbiased Mean-Squared Error
ZF	Zero Forcing

List of Symbols

Notation	Definition
$\mathrm{E}\{\cdot\}$	Expectation
$(\cdot)^t$	Transposition
$(\cdot)^*$	Conjugation
$(\cdot)^H$	Hermitian transpose (i.e., conjugate transpose)
$(\cdot)^\dagger$	Moore-Penrose pseudo-inverse
$\mathrm{tr}(\cdot)$	Trace operator for square matrices
$\mathrm{row}(\cdot)$	Row span of a matrix
$\mathrm{col}(\cdot)$	Column span of a matrix
$\mathrm{diag}(\cdot)$	Extraction of diagonal matrix elements
$\lambda_{\min}(\cdot)$	Minimum eigenvalue
$\lambda_{\max}(\cdot)$	Maximum eigenvalue
$\sigma(\cdot)$	Singular value
ℓ_p	The space of sequences $\{x_n\}$ such that $\sum_n \lvert x_n\rvert^p < \infty$
$\lVert\mathbf{x}\rVert_p$	ℓ_p norm: $\sqrt[p]{\sum_n \lvert x_n\rvert^p}$
$\lVert\mathbf{x}\rVert_{\mathbf{A}}$	Norm defined by $\sqrt{\mathbf{x}^H\mathbf{A}\mathbf{x}}$ for positive definite Hermitian $\mathbf{A}$
$\mathbf{I}$	Identity matrix
$\mathbf{e}_i$	Column vector with 1 at the i^{th} entry ($i \geq 0$) and zeros elsewhere
$\mathbb{R}^p$	The field of p-dimensional real-valued vectors
$\mathbb{C}^p$	The field of p-dimensional complex-valued vectors
$\mathrm{Re}(\cdot)$	Extraction of real-valued component
$\mathrm{Im}(\cdot)$	Extraction of imaginary-valued component
$\mathrm{sgn}(\cdot)$	Real-valued sign operator: $\mathrm{sgn}(x) = 1$ for $x \geq 0$, else $\mathrm{sgn}(x) = -1$
$\mathrm{csgn}(\cdot)$	Complex-valued sign operator: $\mathrm{csgn}(x) := \mathrm{sgn}(\mathrm{Re}\,x) + j\,\mathrm{sgn}(\mathrm{Im}\,x)$
$\nabla_{\mathbf{f}}$	Gradient with respect to $\mathbf{f}$: $\nabla_{\mathbf{f}} := \frac{\partial}{\partial \mathbf{f}_{\mathrm{r}}} + j\frac{\partial}{\partial \mathbf{f}_{\mathrm{i}}}$ for $\mathbf{f}_{\mathrm{r}} = \mathrm{Re}\,\mathbf{f}$, $\mathbf{f}_{\mathrm{i}} = \mathrm{Im}\,\mathbf{f}$
$\mathrm{bndr}(\cdot)$	Boundary of a set
$\mathrm{intr}(\cdot)$	Interior of a set

Chapter 1

Introduction

Estimation of a distorted signal in noise is a classic problem that finds important application in areas including, but not limited to,

- data communication [Gitlin Book 92], [Lee Book 94], [Proakis Book 95],
- radar signal processing [Haykin Book 92],
- sensor array processing [Compton Book 88], [VanVeen ASSPM 88],
- geophysical exploration [Mendel Book 83], [Robinson Book 86],
- speech processing [Deller Book 93],
- image analysis [Jain Book 89],
- biomedicine [Akay Book 96], and
- control systems [Anderson Book 89], [Doyle Book 91].

A mathematical model describing this problem is

$$\boxed{\boldsymbol{r} \;=\; H(\boldsymbol{s}) + \boldsymbol{w}.} \tag{1.1}$$

Here the observed vector $\boldsymbol{r}$ is modeled as a signal vector $\boldsymbol{s}$ distorted by function $H(\cdot)$ and corrupted by additive noise $\boldsymbol{w}$.

Different estimation problems can be characterized by the assumptions placed on $H(\cdot)$, $\boldsymbol{s}$, and $\boldsymbol{w}$. Though central limit theorem arguments are commonly used to justify a Gaussian noise model for $\boldsymbol{w}$, the assumptions on $H(\cdot)$ and $\boldsymbol{s}$ can differ significantly from one application to the next. For example, is $H(\cdot)$ completely known? If not, is it because $H(\cdot)$ is inherently random? Assuming random $H(\cdot)$, do we know its distribution? If not, can variabilities in the distribution or structure of $H(\cdot)$ be described with a small set of parameters? Similar questions can be asked about the signal $\boldsymbol{s}$ or about the noise $\boldsymbol{w}$ when the Gaussian assumption is not adequate. In general, the introduction of accurate prior assumptions about $H(\cdot)$, $\boldsymbol{s}$, and $\boldsymbol{w}$, allows the design of estimators with increased performance (though perhaps more complicated implementation). If prior assumptions are inaccurate, however, estimator performance can suffer significantly.

There exist many applications of estimation theory where prior knowledge is lacking and accurate assumptions are hard to come by. In this dissertation we focus on a relatively extreme lack of knowledge. Specifically, we assume that

1. $H(\cdot)$ is linear but otherwise *unknown*,
2. the signal vector $\boldsymbol{s}$ is composed of statistically independent and identically distributed random variables with *unknown* distribution, and
3. the noise vector $\boldsymbol{w}$ is composed of identically distributed random variables, statistically independent of $\boldsymbol{s}$, with *unknown* distribution.

This set of (non-)assumptions is thought to accurately describe many estimation scenarios encountered in, e.g., data communication over dispersive physical media [Johnson PROC 98]; beamforming with sensor arrays [Paulraj Chap 98], [vanderVeen PROC 98]; seismic deconvolution [Donoho Chap 81]; image de-blurring

Table 1.1: Examples of the blind estimation problem.

application	unknown distortion $H(\cdot)$	independent signal $\boldsymbol{s}$
data communication	channel dispersion	information sequence
beamforming	array geometry	array target signal
seismic deconvolution	seismic wavelet	reflectivity series
image de-blurring	lens blurring characteristic	edges in natural scenes
speech separation	mixing matrix	voice excitations

[Kundur SPM 96a], [Kundur SPM 96b]; and separation of speech mixtures [Torkkola WICASS 99].

One of key distinguishing features of our problem setup is the unknown linear structure of $H(\cdot)$. In the five previously mentioned applications, this feature often corresponds to a lack of knowledge about the dispersion pattern of the communication channel, geometry of the array, shape of the "seismic wavelet," blurring characteristics of the lens, or speaker mixing matrix, respectively. (See Table 1.1.)

Another key feature of our setup is the independent and identically distributed (i.i.d.) nature of $\boldsymbol{s}$. Considering the same five applications, this assumption corresponds to the i.i.d. nature of information transmission sequences, array target signals, seismic "reflectivity series," edges in natural scenes (see, e.g., [Bell Chap 96]), and voice excitation signals, respectively. (See Table 1.1.)

The term *blind estimation* has been used to describe problems of this type since the estimation of the signal is performed blindly[1] with respect to the channel and noise characteristics. For similar reasons, *blind identification* denotes the identi-

[1] The term "blind" seems to have originated in a 1975 paper by Stockham et al. which concerned the restoration of old phonograph records [Stockham PROC 75].

fication of the unknown system $H(\cdot)$ in the context of unknown signal and noise. Though the term "blind estimation" is sometimes used to describe situations where, for instance, $H(\cdot)$ is unknown but the signal distribution *is* known, we stress that our problem setup is different in that it *assumes no prior knowledge* about distribution of signal $\boldsymbol{s}$ and structure of distortion $H(\cdot)$ (with the exception, respectively, of independence and linearity). Hence, we refer to our problem as "blind estimation without priors" (BEWP).

To readers unfamiliar with BEWP, it may seem surprising that estimation of $\boldsymbol{s}$ and $H(\cdot)$ is even possible! Section 2.1 provides intuition as to when and why the independence and linearity assumptions are strong enough to allow accurate estimation of these quantities. More specifically, Section 2.1 discusses inherent ambiguities in the estimation of $\boldsymbol{s}$ and $H(\cdot)$, requirements for perfect blind estimation of $\boldsymbol{s}$, and blind estimation schemes which generate perfect estimates of $\boldsymbol{s}$ under these requirements. The approach taken by Section 2.1 is heavily influenced by Donoho's classic chapter [Donoho Chap 81].

As we shall see in Section 2.1, the situations allowing perfect blind estimation are ideal in the sense that they require an invertible distortion function $H(\cdot)$ and the absence of noise $\boldsymbol{w}$. Motivated by the non-ideality of practical estimation problems, the remainder of the dissertation focuses on the general case: non-invertible $H(\cdot)$, arbitrary signal $\boldsymbol{s}$, and arbitrary interference $\boldsymbol{w}$. A detailed description of the general model under which all results are derived is given in Section 2.2. As a means of measuring blind estimation performance in non-ideal cases, the mean-squared error criterion is introduced in Section 2.3.

One of the admissible criteria for BEWP has become popularly known as the *Shalvi-Weinstein* (SW) criterion following[2] a 1990 paper by Shalvi and Weinstein [Shalvi TIT 90]. Background on the SW approach is provided by Section 2.4. Though yielding perfect blind estimation under ideal conditions, the question remains: *How good are SW estimates in general?* This question is answered in Chapter 3, where we derive tight and general upper bounding expressions for the mean-squared error (MSE) of SW estimates.

The remainder of the dissertation focuses on the properties of linear estimation via the *constant modulus* (CM) criterion[3]—the most widely implemented and studied approach to blind estimation. The CM approach was conceived of independently by Godard in 1980 [Godard TCOM 80] and Treichler & Agee in 1983 [Treichler TASSP 83] as a means of recovering linearly distorted complex-valued communication signals with rapidly varying phase. Numerous applications of the CM criterion have emerged since its inception in the early 1980's and with them a large body of academic research (see, e.g., the citations in [Johnson PROC 98]). The popularity of the CM-minimizing estimator can be attributed to the existence of a computationally efficient algorithm for its implementation and the reportedly excellent performance of the resulting estimates. Though a more complete introduction on the CM criterion will be given in Section 2.5, we give a short preview below that will help outline the contents of Chapters 4–7.

Say that the coefficients of vector $\boldsymbol{f}$ can be adjusted to generate a linear estimate $y = \boldsymbol{f}^t\boldsymbol{r}$. (Recall from (1.1) that $\boldsymbol{r}$ is the vector of observed data.) The CM-minimizing estimators are then defined by the set of estimators $\boldsymbol{f}$ that locally

[2]Though named after Shalvi and Weinstein, the SW estimator was analyzed in detail by Donoho in 1981 [Donoho Chap 81] and was originally proposed in the psychometric literature of the early 1950s (see [Kaiser PSY 58]).

[3]Also known as Godard's criterion or the "minimum dispersion" criterion.

minimize the so-called CM cost:

$$J_{\rm c} \;=\; \mathrm{E}\big\{\big(|y|^2-1\big)^2\big\} \;=\; \mathrm{E}\big\{\big(|\boldsymbol{f}^t\boldsymbol{r}|^2-1\big)^2\big\}, \tag{1.2}$$

where $\mathrm{E}\{\cdot\}$ denotes expectation. Notice from (1.2) that the phrase "CM-minimizing" is synonymous with "dispersion-minimizing" given a target modulus of 1. Though $J_{\rm c}$ has a simple form, it turns out that there are no closed form expressions for its local minimizers given general $\boldsymbol{s}$, $\boldsymbol{w}$, and $H(\cdot)$. The lack of closed-form expressions for the CM-minimizing estimators has made performance characterization in the general case historically difficult and has left the blind estimation community wondering: *How good are CM-minimizing estimates in general, and what factors affect their performance?* We answer this question in Chapter 4 via tight and general bounding expressions for the mean-squared error (MSE) of CM-minimizing estimates. The bounding expressions have a simple and meaningful form which yields significant intuition about the fundamental properties of CM estimates.

Let us now consider a simple example in which the elements of $\boldsymbol{s}=(s_0, s_1, s_2, \dots)^t$ are identically distributed random variables taking on values $\{-1, 1\}$ and where the noise is absent. Since $H(\cdot)$ is assumed linear, it can be assigned a matrix representation $\boldsymbol{H}$, allowing the estimate to be written as $y = \boldsymbol{f}^t\boldsymbol{H}\boldsymbol{s}$. Notice now that the CM cost $J_{\rm c}$ attains its minimum value of zero both with $\boldsymbol{f}$ such that $\boldsymbol{f}^t\boldsymbol{H} = (1, 0, 0, 0, \dots)$ as well as with $\boldsymbol{f}$ such that $\boldsymbol{f}^t\boldsymbol{H} = (0, 1, 0, 0, \dots)$ since, in both cases, perfect estimates of particular elements in $\boldsymbol{s}$ are attained. In the former case we have $y = s_0$, i.e., perfect estimation of the first signal element, while in the latter case we have $y = s_1$, i.e., perfect estimation of the second signal element. While in some applications the difference between s_0 and s_1 might signify a mere one-sample delay in the desired signal estimate—a trivial ambiguity, in other applications it might signify the difference between estimating the desired signal versus an interferer—a nontrivial ambiguity. This example shows that the CM cost functional J_c is inherently multimodal, i.e., has more than one minimizer.

Due to the general lack of closed form expressions for CM-minimizing estimators, gradient descent (GD) methods are typically employed to locate these estimators. In the case of multiple minimizers (with some undesired), proper initialization of the GD algorithm will be of critical importance since it completely determines the minimizer to which the GD algorithm will converge. In other words, a "good" initialization will result in descent estimates of the desired signal, while a "bad" initialization might result in estimates of an interferer, hence poor estimates of the desired signal. With this problem in mind, Chapter 5 considers the question: *How can we guarantee that CM-minimizing GD algorithms will converge to a "useful" setting?* The answers obtained are in the form of CM-GD initialization conditions sufficient to ensure convergence to the desired source.

Though so far we have only considered blind signal estimation, Chapter 6 is concerned instead with blind system identification. It seeks answers to the question: *How can the CM criterion be used to identify the distortion* $\boldsymbol{H}$*, and how good is the resulting identification?* Chapter 6 presents bounds on the average squared parameter error (ASPE) of blind channel estimates for a particular CM-minimizing identification scheme.

As discussed previously, gradient descent methods are commonly used to determine the CM-minimizing estimators since closed-form solutions are unavailable under general conditions. The constant modulus algorithm (CMA) is the most commonly implemented CM gradient descent method and its particularly simple implementation makes it convenient for use in a wide range of applications. In high data-rate communication applications, for example, the low computational complexity associated with CMA is critical to its feasibility as a practical approach to blind adaptive equalization. In fact, implementers claim that the adaptive equalizer may claim as much as 80% of the total receiver circuitry [Treichler SPM 96, p. 73], motivating, if possible, even further reduction in the computational complexity of CMA. In Chapter 7, we present *a novel CM-GD scheme that eliminates the estimator update multiplications required by standard CM-GD while retaining its transient and steady-state mean behaviors.* For readers familiar with stochastic gradient descent algorithms of the LMS type [Haykin Book 96], our scheme may be considered as a variant of the "signed-error" approach [Sethares TSP 92] whose novelty stems from the incorporation of a carefully chosen dither signal [Gray TIT 93].

Chapter 8, the final chapter, summarizes the main results of the dissertation and gives suggestions for future work.

Fig. 1.1 summarizes the organization of the dissertation.

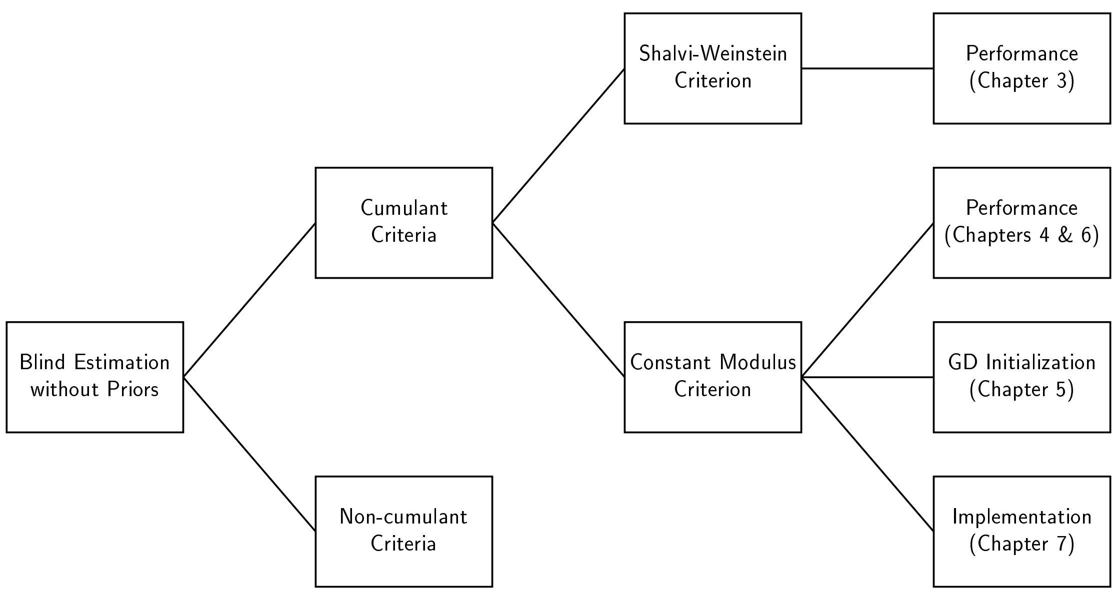

Figure 1.1: Dissertation map. Topics discussed in Chapter 2 unless otherwise labelled.

As discussed in the beginning of this introduction, the applications of blind estimation/identification are many and widespread. It should be stressed that the principle results of this dissertation are completely general and thus apply to any application for which the "independent linear model" holds. At times, however, we will find it instructive to present examples and simulation studies that target a particular application. Because data communication applications seem to have received the largest share of attention from the blind estimation community over the last twenty years and because they represent the author's field of expertise, the majority of examples in this dissertation will focus on the data communication application. In this application, $\boldsymbol{H}$ symbolizes the communication "channel" and so we refer to it using this terminology in the sequel. Similarly, we will often refer to $\boldsymbol{s}$ as the "source" or "symbol sequence" and to $\boldsymbol{f}$ as the "receiver" or "equalizer."

A word on notation. In general, we use bold lower-case letters to designate vectors and bold upper-case letters to designate matrices. Note that in some cases italicized and non-italicized versions of the same letter, such as $\boldsymbol{f}$ and $\mathbf{f}$, will be used to denote distinctly different quantities. We hope that this will not confuse the reader. Consult the list of symbols on page xvii for further information on mathematical notation.

Chapter 2

Background

2.1 Introduction to Blind Estimation without Priors

In Section 2.1 we give a tutorial introduction to the "blind estimation without priors" (BEWP) problem. We start by defining the BEWP problem for a relatively simple model. After finding that classical estimation techniques are not well suited to the problem setup (due to insufficient prior knowledge), we are forced to re-examine the few (but key!) assumptions made in BEWP. From an examination of fundamental properties of linear combinations of independent random variables, linear approaches to the estimation problem are found to be practical. Furthermore, these same properties suggest that perfect blind linear estimation is possible (under certain conditions) through maximization of properly defined estimation criteria. Examples of such criteria are then given, including the kurtosis criterion which (as we will see in Section 2.5) is central to the remainder of the dissertation.

2.1.1 A Simple Model

Consider the stationary single-input single-output (SISO) system model shown in Fig. 2.1. The notation $\{z_n\} \sim z$ will be used to denote the situation that $\{z_n\}$ is a sequence of random variables distributed identically to some random variable z. For our purposes, n is a discrete time index. Regarding the system in Fig. 2.1, we assume

- signal: i.i.d.[1] $\{s_n\} \sim s$ with finite nonzero variance,
- noise: i.i.d. $\{w_n\} \sim w$ independent of $\{s_n\}$ with finite variance, and
- channel: SISO, linear time-invariant (LTI), with impulse response $\{h_n\} \in \ell_2$.

Since the channel is time invariant and the signal and noise are both stationary, the observation $\{r_n\}$ will be stationary with some r such that $\{r_n\} \sim r$. Furthermore, i.i.d. signal and noise sequences imply ergodic $\{r_n\}$. Finally, all quantities are real-valued.

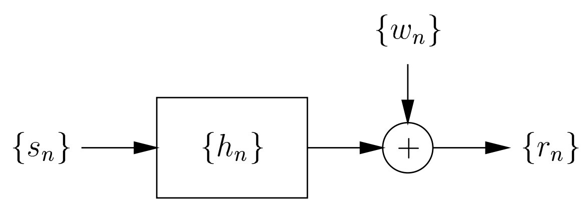

Figure 2.1: Stationary SISO observation model.

At times it will be convenient to consider a collection of observations such that the most recent time index is n. Such collections will be represented by the vector $\boldsymbol{r}_n = (r_n, r_{n-1}, r_{n-2}, \ldots)^t$. This collection may be finite or infinite and its size may be fixed or grow with n depending on how the observations are collected.

[1] independent and identically distributed

The observed vector can be related to an appropriately-defined source vector $\boldsymbol{s}_n = (s_n, s_{n-1}, s_{n-2}, \dots)^t$, noise vector $\boldsymbol{w}_n = (w_n, w_{n-1}, w_{n-2}, \dots)^t$, and (possibly infinite dimensional) Toeplitz matrix $\boldsymbol{H}$, as follows:

$$\boldsymbol{r}_n = \boldsymbol{H}\boldsymbol{s}_n + \boldsymbol{w}_n. \tag{2.1}$$

2.1.2 The Problem with Classical Techniques

In designing an estimator, a logical starting point is to consider classical methods such as maximum likelihood, maximum a posteriori, and Bayesian mean-squared error. Reference materials for classic estimation theory include [Kay Book 93], [Poor Book 94], [Porat Book 94], and [VanTrees Book 68]. The goal of this section is to show that the classical methods are not compatible with the BEWP problem.

We start with the maximum likelihood (ML) criterion. The ML estimates of $\boldsymbol{H}$ and $\boldsymbol{s}_n$ are defined as the global maximizers of the so-called likelihood function $p(\boldsymbol{r}_n|\boldsymbol{H}, \boldsymbol{s}_n)$. The likelihood function is specified by the conditional density of the observation $\boldsymbol{r}_n$ as a function of hypothesized channel $\boldsymbol{H}$ and signal $\boldsymbol{s}_n$. Intuitively, the ML estimates $\hat{\boldsymbol{H}}|_{\mathrm{ML}}$ and $\hat{\boldsymbol{s}}_n|_{\mathrm{ML}}$ are those that make the actual observation $\boldsymbol{r}_n$ the most likely out of all possible observations. Inherent to the likelihood function $p(\boldsymbol{r}_n|\boldsymbol{H}, \boldsymbol{s}_n)$ is a statistical model relating $\boldsymbol{H}$ and $\boldsymbol{s}_n$ to $\boldsymbol{r}_n$, which, according to the additive noise model (2.1), will be governed by the distribution of noise $\boldsymbol{w}_n$. Without any description of the distribution of $\boldsymbol{w}_n$, though, it is unclear how to proceed with the ML approach.

As an aside, we note that a ML approach has been applied to *noiseless invertible* blind source separation, a special case of the BEWP problem with $\boldsymbol{w}_n = \boldsymbol{0}$ and square invertible $\boldsymbol{H}$, by Cardoso [Cardoso PROC 98]. In his ML formulation, the likelihood function takes the form of $p(\boldsymbol{r}_n|\boldsymbol{H}, \boldsymbol{s}_n, q)$ where q is the (unknown) marginal distribution of the i.i.d. elements in $\boldsymbol{s}_n$. There is no known extension of this approach to the noisy non-invertible case, however; consider the following comments recently made by Cardoso.

> *"Taking noise effects into account...is futile at low SNR because the blind source separation problem becomes too difficult"* —[Cardoso PROC 98]

> *"The most challenging open problem in blind source separation probably is the extension to convolutive mixtures."* —[Cardoso PROC 98]

Next we consider the maximum a posteriori (MAP) criterion. The MAP estimates of $\boldsymbol{H}$ and $\boldsymbol{s}_n$ are defined as the global maximizers of the posterior density

$$p(\boldsymbol{H}, \boldsymbol{s}_n | \boldsymbol{r}_n) \;=\; \frac{p(\boldsymbol{r}_n | \boldsymbol{H}, \boldsymbol{s}_n) p(\boldsymbol{H}, \boldsymbol{s}_n)}{p(\boldsymbol{r}_n)}.$$

The often convenient right-hand side of the previous equation is a result of Bayes' Theorem [Papoulis Book 91]. Here again we are impeded by our lack of knowledge regarding the distributions of $\boldsymbol{w}_n$ and $\boldsymbol{s}_n$. It should be mentioned the MAP criterion has also been applied to the noiseless invertible blind source separation problem [Knuth WICASS 99].

Finally, we consider the Bayesian mean-squared error (BMSE) criterion. Say that we are interested in estimates $\{\hat{s}_n\}$ minimizing the mean-squared error (MSE) relative to a ν-delayed version of the signal:

$$J_{\mathrm{m},\nu} \;=\; \mathrm{E}\{|\hat{s}_n - s_{n-\nu}|^2\} \tag{2.2}$$

where $\mathrm{E}\{\cdot\}$ denotes expectation. It is well known that the minimum MSE (MMSE) estimate is given by the conditional mean $\hat{s}_n|_{\mathrm{MSE}} = \mathrm{E}\{s_{n-\nu}|\boldsymbol{r}_n\}$ which requires knowledge of the conditional density $p(s_{n-\nu}|\boldsymbol{r}_n)$. Here again we are stuck.

If we restrict our search to linear estimators (noting that $\mathrm{E}\{s_{n-\nu}|\boldsymbol{r}_n\}$ is in general nonlinear), then it can be shown that

$$\hat{s}_n|_{\mathrm{MSE}} \;=\; \boldsymbol{f}^t \boldsymbol{r}_n \quad \text{for} \quad \boldsymbol{f} \;=\; \left(\mathrm{E}\{\boldsymbol{r}_n \boldsymbol{r}_n^t\}\right)^{-1} \mathrm{E}\{\boldsymbol{r}_n s_{n-\nu}\}.$$

Though ergodicity implies that it is possible to identify $\mathrm{E}\{\boldsymbol{r}_n \boldsymbol{r}_n^t\}$ given enough observed data, it is not clear how to obtain $\mathrm{E}\{\boldsymbol{r}_n s_{n-\nu}\}$ from the observed sequence when $\{s_n\}$ and $\boldsymbol{H}$ are unknown.

To conclude, the lack of knowledge about signal and noise distributions in the BEWP problem prevents the application of classical approaches to estimation, namely the ML, MAP, and Bayesian-MSE methods.

2.1.3 Ambiguities Inherent to BEWP

With the apparent failure of classical estimation methods, one would be justified in questioning whether the BEWP problem actually has a solution. In this section partially answer this question by pointing out ambiguities in the BEWP problem formulation which cannot be resolved. In later sections we will examine the possibility of accurate blind estimation modulo these ambiguities.

Given that the estimator has knowledge of only the observation $\{r_n\}$ in the system of Fig. 2.1, we ask the question: *Can model quantities be altered in a way that does not effectively alter the observation?* If the answer is yes, then there exists inherent ambiguity in the BEWP problem setup.

First note that simultaneously scaling the signal by α and the channel by α^{-1}, where α is a fixed non-zero gain, will not affect the observation $\{r_n\}$:

$$\begin{aligned} r_n &= \{s_n\} * \{h_n\} + \{w_n\} \\ &= \sum_m s_m h_{n-m} + w_n \\ &= \sum_m (\alpha s_m)(\alpha^{-1} h_{n-m}) + w_n \\ &= \{\alpha s_n\} * \{\alpha^{-1} h_n\} + \{w_n\}. \end{aligned}$$

Next, note that advancing the signal by ν time-steps while delaying the channel by the same amount also yields an identical observation:

$$\begin{aligned} r_n &= \{s_n\} * \{h_n\} + \{w_n\} \\ &= \sum_m s_m h_{n-m} + w_n \\ &= \sum_m s_{m+\nu} h_{n-(m+\nu)} + w_n \\ &= \sum_m s_{m+\nu} h_{(n-\nu)+m} + w_n \\ &= \{s_{n+\nu}\} * \{h_{n-\nu}\} + \{w_n\}. \end{aligned}$$

Thus an *ambiguity in absolute gain and delay is inherent to BEWP*. Since these ambiguities are considered tolerable in many applications, we accept them as necessary consequences of the BEWP formulation and forge onward.

Finally, the effect of a Gaussian source is considered. Recalling that Gaussianity is preserved under linear combinations, a source process $\{s_n\} \sim s$ where s is Gaussian would give stationary Gaussian channel output $\{x_n\} = \{s_n\} * \{h_n\}$. Now, a stationary Gaussian process $\{x_n\}$ is completely characterized by its mean and covariance, hence its power spectrum[2]. The power spectrum of $\{x_n\}$ is, in turn, completely determined by the power spectrum of the $\{s_n\}$ and the magnitude of

[2]Power spectrum is defined as the discrete-time Fourier transform of the autocorrelation sequence.

the frequency response of $\{h_n\}$. The important point here is that, when $s \sim \{s_n\}$ is Gaussian, the phase component of the frequency response of $\{h_n\}$ does *not* enter into the statistical description of $\{x_n\}$. Thus, there is no way to tell whether $\{s_n\}$ was processed by an arbitrary allpass filter before processing by the linear system $\{h_n\}$. As a consequence, any statistically-derived estimate of $\{s_n\} \sim s$ with Gaussian s will be subject to an ambiguity in phase response[3]. Most applications consider such this form of ambiguity as severe and intolerable. For this reason we say that *the BEWP problem is ill-posed when the marginal distribution of the source* $\{s_n\}$ *is Gaussian.*

In summary, the BEWP problem setup does not admit the estimation of the absolute gain/delay of $\{s_n\}$, not does it allow the reliable estimation of $\{s_n\} \sim s$ with Gaussian distribution. However, the accurate estimation of possibly scaled or shifted non-Gaussian processes $\{s_n\}$ is of significant practical interest in, e.g., the applications mentioned in Chapter 1. The remainder of the dissertation focuses on blind estimation of non-Gaussian $\{s_n\}$ subject to inherent ambiguity in absolute gain and phase.

2.1.4 On Linear Combinations of I.I.D. Random Variables

In this section we examine some properties of linear combinations of i.i.d. random variables. Such properties are of great interest to the study of BEWP because "i.i.d.-ness" and linearity are the *only* modeling assumptions. Through this examination, we aim to build intuition regarding admissible estimation strategies for BEWP. Since our aim is instructional, some formalities have been omitted, and so readers

[3]When the linear system $\{h_n\}$ is known to be minimum phase (but otherwise unknown), it is possible to perfectly recover Gaussian $\{s_n\}$ in the absence of noise by passing $\{x_n\}$ through a whitening filter. In BEWP, however, we cannot assume that $\{h_n\}$ is minimum phase.

are encouraged to consult [Donoho Chap 81] and [Kagan Book 73] for further detail.

Through the definitions and lemmas below, we establish a so-called partial ordering between random variables which will be denoted by "$\stackrel{\bullet}{\geq}$". It will be shown that "$x \stackrel{\bullet}{\geq} y$" has the interpretation "x is farther from Gaussian than y is," though we do not assume this property from the outset.

Definition 2.1. *Two random variables y and s are said to be* equivalent*, denoted by $y \doteq s$, if there exist constants μ and $\alpha \neq 0$ such that $\alpha s + \mu$ has the same probability distribution as y.*

Definition 2.2. *The relation $y \stackrel{\bullet}{\leq} s$ means that $y \doteq \sum_n q_n s_n$ for i.i.d. $\{s_n\} \sim s$ and some $\{q_n\} \in \ell_2$. The relation $y \stackrel{\bullet}{<} s$ is short for "$\stackrel{\bullet}{\leq}$ but not $\doteq$".*

Definition 2.3. *A sequence $\{q_n\}$ is said to be* trivial *if there exists one and only one index n such that $|q_n| > 0$.*

Lemma 2.1 (KLR). *Consider i.i.d. $\{z_n\} \sim z$. Then z is Gaussian iff z has finite variance and the relation $z \doteq \sum_n q_n z_n$ holds for some nontrivial set $\{q_n\} \in \ell_2$.*

Proof. See Theorem 5.6.1 of [Kagan Book 73]. □

To paraphrase the above KLR Lemma, the *only* distribution preserved under nontrivial linear combinations of i.i.d. random variables is the Gaussian distribution.

Lemma 2.2. *The relation $\stackrel{\bullet}{\leq}$ is a "partial ordering" because (i) if $z \stackrel{\bullet}{\leq} y$ and $y \stackrel{\bullet}{\leq} s$ then $z \stackrel{\bullet}{\leq} s$ and (ii) if $s \stackrel{\bullet}{\leq} y$ and $y \stackrel{\bullet}{\leq} s$ then $s \doteq y$.*

Proof. Statement (i) follows from Definition 2.2: if $z \doteq \sum_n a_n y_n$ and $y \doteq \sum_n b_n s_n$, then $z \doteq \sum_{n,m} a_n b_m s_{n,m}$. For (ii), suppose that $y = \sum_n a_n s_n$ and $s = \sum_n b_n y_n$, so that $s = \sum_n a_n b_n s_{n,m}$. Then by KLR, s is Gaussian if either $\{a_n\}$ or $\{b_n\}$ is nontrivial. When s is Gaussian, KLR implies that y is also Gaussian, hence $s \doteq y$.

On the other hand, if both $\{a_n\}$ and $\{b_n\}$ are trivial, then $s \doteq y$ follows immediately. The claim that $\stackrel{\bullet}{\leq}$ is a partial ordering follows by definition after (i) and (ii). (See, e.g., [Naylor Book 82, p. 556].) ☐

Theorem 2.1. *Consider i.i.d.* $\{s_n\} \sim s$ *and Gaussian* z*. Then* $z \stackrel{\bullet}{\leq} \sum_n q_n s_n \stackrel{\bullet}{\leq} s$ *with strict ordering unless either*

1. s *is Gaussian, in which case* $z \doteq \sum_n q_n s_n \doteq s$*, or*

2. s *is non-Gaussian but* $\{q_n\}$ *is trivial, in which case* $z \stackrel{\bullet}{<} \sum_n q_n s_n \doteq s$*.*

Proof. First we tackle the right side. Definition 2.2 yields $\sum_n q_n s_n \stackrel{\bullet}{\leq} s$ directly. For trivial $\{q_n\}$ it is obvious that $\sum_n q_n s_n \doteq s$, and KLR implies $\sum_n q_n s_n \doteq s$ when s is Gaussian. Now the left side. Definition 2.2 and KLR imply that $y \stackrel{\bullet}{\nless} z$ for Gaussian z and any y, hence $z \stackrel{\bullet}{\leq} \sum_n q_n s_n$. If s is non-Gaussian and $\{q_n\}$ is trivial, then $z \stackrel{\bullet}{\neq} s$, so we must have $z \stackrel{\bullet}{<} s_n$. If s is Gaussian, then $z \doteq \sum_n q_n s_n$ follows from KLR. ☐

Theorem 2.1 can be interpreted as follows: *nontrivial linear combinations of i.i.d. random variables are "closer to Gaussian" than are the original random variables.* In the next section, we investigate the implications of these properties of $\stackrel{\bullet}{\leq}$ for the BEWP problem.

2.1.5 Implications for Blind Linear Estimation

Fig. 2.2 depicts linear estimation of a signal $\{s_n\}$ processed by a SISO linear system and corrupted by additive noise. Theorem 2.1 has powerful implications for linear estimation approaches to BEWP because the resulting estimates y_n are linear combinations of the desired i.i.d. symbols s_n. Recalling the gain and delay ambiguities inherent to BEWP (discussed in Section 2.1.3), our goal is to obtain linear estimates of the form

$$\{y_n\} \;=\; \{\alpha s_{n-\nu}\} \quad \text{for some} \;\; \alpha \neq 0,\; \nu.$$

(The bracketed notation in the previous expression indicates the estimation of a sequence of random variables.) The attainment of such estimates will be referred to as *perfect blind linear estimation* (PBLE).

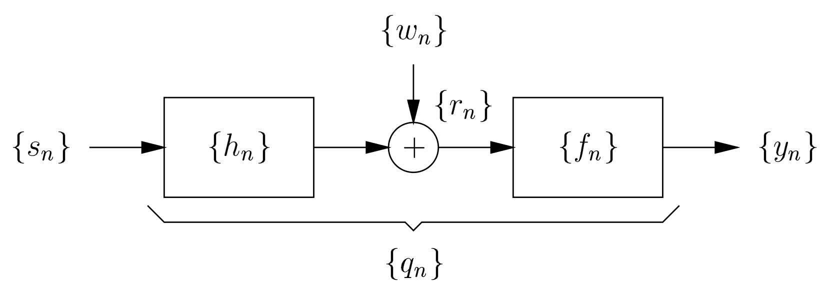

Figure 2.2: Blind linear estimation.

In this section we assume that $\boldsymbol{r}_n$, some collection of observations up to time n, has the same (though possibly infinite) length for all n. It will be convenient to collect the estimator coefficients $\{f_n\}$ into vector $\boldsymbol{f}$ having the same length as $\boldsymbol{r}_n$ and constructed so that $y_n = \boldsymbol{f}^t\boldsymbol{r}_n$. In the same way that we use $\{r_n\} \sim r$ to denote the case that r_n is distributed identically to r for all n, we use $\{\boldsymbol{r}_n\} \sim \boldsymbol{r}$ to denote the case that $\boldsymbol{r}_n$ is (jointly) distributed identically to $\boldsymbol{r}$ for all n.

We have already encountered one situation under which PBLE is not possible: the case of Gaussian $s \sim \{s_n\}$. Here we point out two more situations. According to Fig. 2.2, the estimates can be written

$$y_n \;=\; \sum_i \underbrace{\Big(\sum_m f_m h_{i-m}\Big)}_{q_i} s_{n-i} + \sum_i f_i w_{n-i}$$

Since $s_{n-\nu}$ is independent of both $\{s_{n-i}\}\big|_{i\neq\nu}$ and $\{w_n\}$ for any ν, PBLE occurs if and only if $\{q_i\}$ is trivial and noise is absent (i.e., $w_n = 0 \ \forall n$). But is it always possible to adjust $\{f_n\}$ so that $\{q_n\}$ is trivial? The answer is no; $\{h_n\}$ must be invertible in the sense of Definition 2.4.

Definition 2.4. *A system with impulse response $\{h_n\}$ is said to be* invertible *if there exists another system with impulse response $\{f_n\} \in \ell_2$ such that the cascaded system response $\{q_n\}$, where $q_n = \sum_m f_m h_{n-m}$, is trivial.*

To summarize, the PBLE conditions for the system in Fig. 2.2 are

1. non-Gaussian i.i.d. signal $\{s_n\}$,
2. invertible channel $\{h_n\}$, and
3. the absence of noise $\{w_n\}$.

For the remainder of this section *we assume satisfaction of the PBLE conditions* in order to study the properties of perfect blind estimates and propose estimation schemes capable of generating such estimates. We admit that such assumptions of ideality are artificial in the sense that they give no concrete information about BEWP under general (non-ideal) conditions. In fact, performance characterization in *non-ideal* settings provides one of the major themes for this dissertation, and is the subject of Chapters 3–6. For now, however, realize that perfect performance under ideal conditions is a reasonable requirement for serious consideration of any blind estimation scheme and forms a natural point from which to construct such schemes. In light of these comments, we focus the remainder of Section 2.1.5 on the search for admissible BEWP estimation criteria—a necessary starting point from which more general analyses will proceed.

Applying Theorem 2.1 to the ideal linear estimation scenario gives the following important result, written in the notation of Fig. 2.2.

Corollary 2.1. *Assuming Gaussian z and satisfaction of the PBLE conditions, the following holds.*

$$z \overset{\bullet}{<} y \doteq \sum_n q_n s_n \overset{\bullet}{\leq} s.$$

Furthermore, the above ordering is strict unless perfect blind estimation is achieved, in which case the right side becomes "$\doteq$".

Corollary 2.1 suggests the construction of blind estimation strategies which adjust the coefficients of the linear estimator so that the estimates $\{y_n\}$ are "as far from Gaussian as possible." To state this idea more precisely, say that $G(y)$ is some criterion of goodness that is a continuous function of the marginal distribution of the estimates $y \sim \{y_n\}$. (The ergodicity of $\{y_n\}$ ensures that in practice such information can be well-estimated from data records of adequate length.) Assume also that $G(y)$ is invariant to the scale of y. (Note that the gain ambiguity inherent to BEWP implies that this latter assumption can be made at no extra cost.) Then Theorem 2.2 states a necessary and sufficient condition for the construction of admissible criteria $G(\cdot)$.

Definition 2.5. *We say that $G(\cdot)$ "agrees with $\overset{\bullet}{<}$" if $x \overset{\bullet}{<} y$ implies $G(x) < G(y)$.*

Theorem 2.2 (Donoho). *Under satisfaction of PBLE conditions, locally maximizing estimators $\boldsymbol{f}_\star = \arg\max_{\boldsymbol{f}} G(\boldsymbol{f}^t \boldsymbol{x})$ generate perfect blind estimates $y = \boldsymbol{f}_\star^t \boldsymbol{x}$ for any non-Gaussian $\boldsymbol{x}$ iff $G(\cdot)$ agrees with $\overset{\bullet}{<}$.*

Proof. Informally speaking, this result follows from Corollary 2.1 and Definition 2.5. See [Donoho Chap 81] for more rigorous arguments. $\square$

2.1.6 Examples of Admissible Criteria for Linear BEWP

Theorem 2.2 presents a necessary and sufficient condition for a criterion $G(y)$ to be admissible, i.e., generate perfect blind linear estimates under ideal BEWP conditions. In this section we give examples of admissible criteria.

First consider the m^{th}-order cumulant of y, defined below using $j := \sqrt{-1}$ and using $p_y(\cdot)$ to denote the probability density function of y.

$$\mathcal{C}_m(y) \;:=\; \left[\left(-j\frac{d}{dt}\right)^m \log \int e^{jt\tilde{y}} p_y(\tilde{y})d\tilde{y}\right]\bigg|_{t=0} \quad \text{for } m = 1, 2, 3, \ldots$$

Cumulants have the following convenient property. For a linear combination of i.i.d. $\{s_n\} \sim s$,

$$\mathcal{C}_m\left(\sum_n q_n s_n\right) \;=\; \mathcal{C}_m(s) \sum_n q_n^m. \tag{2.3}$$

(Consult [Cadzow SPM 96] for other properties of cumulants.) The m^{th}-order normalized cumulant is defined by the ratio

$$\bar{\mathcal{C}}_m(y) \;:=\; \frac{\mathcal{C}_m(y)}{\mathcal{C}_2^{m/2}(y)}. \tag{2.4}$$

Normalization makes $\bar{\mathcal{C}}_m(y)$ insensitive to the scaling of y. We now show that $|\bar{\mathcal{C}}_m(y)|$ agrees with $\dot{<}$. Substituting $y_n = \sum_n q_n s_n$ into (2.4) and using the cumulant property (2.3),

$$\bar{\mathcal{C}}_m(y) \;=\; \frac{\mathcal{C}_m(s)\sum_n q_n^m}{\left(\mathcal{C}_2(s)\sum_n q_n^2\right)^{\frac{m}{2}}} \;=\; \bar{\mathcal{C}}_m(s)\frac{\sum_n q_n^m}{\left(\sum_n q_n^2\right)^{\frac{m}{2}}}. \tag{2.5}$$

For $m > 2$, the rightmost fraction in (2.5) is ≤ 1 with equality iff $\{q_n\}$ is trivial. Thus, for $m > 2$ we know $x \,\dot{<}\, y \;\Rightarrow\; |\bar{\mathcal{C}}_m(x)| < |\bar{\mathcal{C}}_m(y)|$ and that $x \doteq y \;\Leftrightarrow\; |\bar{\mathcal{C}}_m(x)| = |\bar{\mathcal{C}}_m(y)|$. We conclude that $|\bar{\mathcal{C}}_m(y)|$ agrees with $\dot{<}$, making $|\bar{\mathcal{C}}_m(y)|$ an admissible criterion for blind linear estimation without priors.

For criteria of the form $|\bar{\mathcal{C}}_m(y)|$, choosing a small value for m is encouraged by the fact that higher-order cumulants diverge/vanish before lower-order cumulants do. In other words, lower-order cumulants are suited to a wider class of problems than are higher-order cumulants. The fourth-order cumulant $\mathcal{C}_4(y)$, often referred to as *kurtosis* and denoted by $\mathcal{K}(y)$, has a particularly long history within the blind estimation community. (A detailed historical account will be given in Section 2.4.) The popularity of the kurtosis criterion may be related to a particular advantage of the choice $m = 4$: it is the smallest $m > 2$ which yields non-zero $\mathcal{C}_m(y)$ for symmetric densities $p_y(\cdot)$. The kurtosis criterion is, in fact, central to the focus of this dissertation since the estimation schemes analyzed in Chapters 3–7 are based, either directly or indirectly, on $\bar{\mathcal{C}}_4(y)$. (This point will be illuminated in Section 2.5.)

It is also possible to construct admissible criteria that use the entire distribution of y as opposed to the partial information given by cumulant ratios. For example, candidate criteria can be derived from Kullback-Leibler divergence or differential entropy under suitable normalization. We conclude this subsection with a brief outline of the admissibility of such methods. Differential entropy [Cover Book 91] is defined as

$$H(y) \;:=\; -\int p_y(\tilde{y}) \log p_y(\tilde{y})\, d\tilde{y}.$$

Using variational calculus, it can be shown that for i.i.d. $\{s_n\} \sim s$,

$$-H\Big(\sum_n q_n s_n\Big) \;\leq\; -H(s)$$

for $\sum_n q_n^2 = 1$, with strict inequality for non-Gaussian s and nontrivial $\{q_n\}$. The condition $\sum_n q_n^2 = 1$ can be enforced by considering only normalized estimates y/σ_y. Thus $-H(y/\sigma_y)$ agrees with $\stackrel{\bullet}{<}$, making $-H(y/\sigma_y)$ an admissible criterion for blind linear estimation without priors [Donoho Chap 81].

2.1.7 Summary and Unanswered Questions

Sections 2.1.4–2.1.6 motivated estimation approaches to the BEWP problem that took advantage of fundamental properties of linear combinations of i.i.d. random variables. It was shown that in the ideal case (i.e., non-Gaussian i.i.d. source, invertible channel, and no noise), the maximization of smooth scale-independent functionals of the marginal distribution of linear estimates y is sufficient to specify perfect blind linear estimators, i.e., estimators generating signal estimates that, modulo unavoidable ambiguity in absolute gain and delay, are otherwise perfect.

The underlined words in the previous paragraph point out the key limitations imposed in our (tutorially motivated) development. Challenging these limitations raises a number of questions:

- What can be said about the *non-ideal* cases, i.e., those which violate the PBLE conditions? Although we expect *imperfect* blind linear estimates in non-ideal cases, can we demonstrate that such estimates are still "good" in some meaningful sense? Chapters 3–6 aim to answer this question for the kurtosis-based and dispersion-based blind estimation criteria described in Sections 2.4–2.5 assuming the general linear model of Section 2.2 and using the (unbiased) mean-squared error criterion of Section 2.3 as the measure of "goodness."

- Though criteria built on the *marginal* distribution of linear estimates were shown to be adequate for PBLE in the ideal case, is something to be gained by consideration of the *joint* distribution of a subset of previous estimates $\{y_n, y_{n-1}, y_{n-2}, \dots\}$ in the non-ideal case? Two heuristic examples of this approach are CRIMNO [Chen OE 92] and vector CMA [Yang SPL 98], [Touzni SPL 00]. Though interesting, the consideration of criteria built on joint distributions is outside the scope of this dissertation.

- The restriction to *linear* estimators was a key step in making use of fundamental properties on linear combinations of i.i.d. random variables. Allowing nonlinear estimators would force us to consider completely different solutions to the BEWP problem. Furthermore, we expect that general results on nonlinear estimators would be much harder to obtain than those for linear estimators. Yet we know that non-linear blind estimation techniques have incredible potential; as evidence, consider the popularity of decision feedback approaches to blind symbol estimation for data communication [Casas Chap 00]. Though of great importance, the blind non-linear estimation problem is also outside the scope of this dissertation.

2.2 A General Linear Model

In this section we describe the system model illustrated in Fig. 2.3, which we assume *for the remainder of the dissertation.*

We will now describe the linear time-invariant multi-channel model of Fig. 2.3 in some detail. Say that the desired symbol sequence $\{s_n^{(0)}\}$ and K sources of interference $\{s_n^{(1)}\}, \ldots, \{s_n^{(K)}\}$ each pass through separate linear "channels" before being observed at the receiver. The interference processes may correspond, e.g., to interference signals or additive noise processes. In addition, say that the receiver uses a sequence of P-dimensional vector observations $\{\mathbf{r}_n\}$ to estimate (a possibly delayed version of) the desired source sequence, where the case $P > 1$ corresponds to a receiver that employs multiple sensors and/or samples at an integer multiple of the

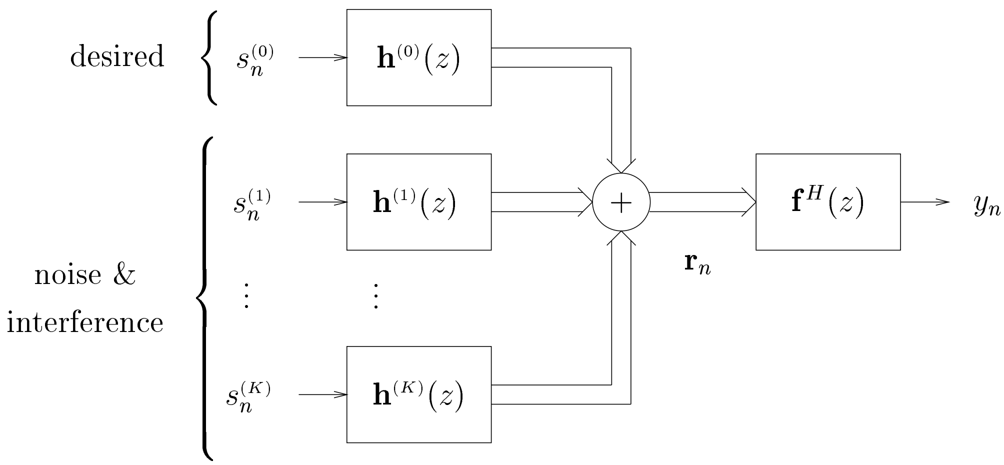

Figure 2.3: MIMO linear system model with K sources of interference.

symbol rate. The observations $\mathbf{r}_n$ can be written

$$\mathbf{r}_n = \sum_{k=0}^{K} \sum_{i=0}^{\infty} \mathbf{h}_i^{(k)} s_{n-i}^{(k)} \tag{2.6}$$

where $\{\mathbf{h}_n^{(k)}\}$ denote the impulse response coefficients of the linear time-invariant (LTI) channel $\mathbf{h}^{(k)}(z)$. We assume that $\mathbf{h}^{(k)}(z)$ is causal and bounded-input bounded-output (BIBO) stable. Fig. 2.3 can be referred to as a multiple-input multiple-output (MIMO) linear model.

From the vector-valued observation sequence $\{\mathbf{r}_n\}$, the receiver generates a sequence of linear estimates $\{y_n\}$ of $\{s_{n-\nu}^{(k)}\}$, where ν is a fixed integer. Using $\{\mathbf{f}_n\}$ to denote the impulse response of the linear estimator $\mathbf{f}(z)$, the estimates are formed as

$$y_n = \sum_{i=-\infty}^{\infty} \mathbf{f}_i^H \mathbf{r}_{n-i}. \tag{2.7}$$

We will assume that the linear system $\mathbf{f}(z)$ is BIBO stable with *constrained* ARMA structure, i.e., the p^{th} element of $\mathbf{f}(z)$ takes the form

$$[\mathbf{f}(z)]_p = \frac{\sum_{i=0}^{L_b^{[p]}} b_i^{[p]} z^{-n_i^{[p]}}}{1 + \sum_{i=1}^{L_a^{[p]}} a_i^{[p]} z^{-m_i^{[p]}}} \tag{2.8}$$

where the $L_b^{[p]}+1$ "active" numerator coefficients $\{b_i^{[p]}\}$ and the $L_a^{[p]}$ active denominator coefficients $\{a_i^{[p]}\}$ are constrained to the polynomial indices $\{n_i^{[p]}\}$ and $\{m_i^{[p]}\}$, respectively.

It will be convenient to collect the impulse response coefficients $\{\mathbf{f}_n\}$ into a (possibly infinite dimensional) vector

$$\boldsymbol{f} \ := \ (\ldots, \mathbf{f}_{-2}^t, \mathbf{f}_{-1}^t, \mathbf{f}_0^t, \mathbf{f}_1^t, \mathbf{f}_2^t, \ldots)^t \tag{2.9}$$

and the corresponding observations $\{\mathbf{r}_n\}$ into a vector

$$\boldsymbol{r}(n) \ := \ (\ldots, \mathbf{r}_{n+2}^t, \mathbf{r}_{n+1}^t, \mathbf{r}_n^t, \mathbf{r}_{n-1}^t, \mathbf{r}_{n-2}^t, \ldots)^t \tag{2.10}$$

so that

$$y_n \ = \ \boldsymbol{f}^H \boldsymbol{r}(n).$$

Note that, due to the constraints on $\mathbf{f}(z)$ made explicit in (2.8), not all $\boldsymbol{f}$ may be attainable. So, we denote the set of $\boldsymbol{f}$ that are attainable by $\mathcal{F}_{\mathrm{a}}$. As an example, when $\mathbf{f}(z)$ is causal FIR,

$$\begin{aligned} \boldsymbol{f} \ &= \ (\mathbf{f}_0^t, \mathbf{f}_1^t, \ldots, \mathbf{f}_{N_f-1}^t)^t \\ \boldsymbol{r}(n) \ &= \ (\mathbf{r}_n^t, \mathbf{r}_{n-1}^t, \ldots, \mathbf{r}_{n-N_f+1}^t)^t, \end{aligned}$$

and thus $\mathcal{F}_{\mathrm{a}}$ equals $\mathbb{C}^{N_f}$.

In the sequel, we focus heavily on the global channel-plus-estimator $q^{(k)}(z) := \mathbf{f}^H(z)\mathbf{h}^{(k)}(z)$. The impulse response coefficients of $q^{(k)}(z)$ can be written

$$q_n^{(k)} \ = \ \sum_{i=-\infty}^{\infty} \mathbf{f}_i^H \mathbf{h}_{n-i}^{(k)}, \tag{2.11}$$

allowing the estimates to be written as

$$y_n \ = \ \sum_{k=0}^{K} \sum_{i=-\infty}^{\infty} q_i^{(k)} s_{n-i}^{(k)}.$$

Adopting the following vector notation helps to streamline the remainder of the work.

$$\begin{aligned}
\boldsymbol{q}^{(k)} &:= (\ldots, q_{-1}^{(k)}, q_0^{(k)}, q_1^{(k)}, \ldots)^t, \\
\boldsymbol{q} &:= (\cdots,\ q_{-1}^{(0)}, q_{-1}^{(1)}, \ldots, q_{-1}^{(K)},\ q_0^{(0)}, q_0^{(1)}, \ldots, q_0^{(K)},\ q_1^{(0)}, q_1^{(1)}, \ldots, q_1^{(K)},\ \cdots)^t, \\
\boldsymbol{s}^{(k)}(n) &:= (\ldots, s_{n+1}^{(k)}, s_n^{(k)}, s_{n-1}^{(k)}, \ldots)^t, \\
\boldsymbol{s}(n) &:= (\cdots,\ s_{n+1}^{(0)}, s_{n+1}^{(1)}, \ldots, s_{n+1}^{(K)},\ s_n^{(0)}, s_n^{(1)}, \ldots, s_n^{(K)},\ s_{n-1}^{(0)}, s_{n-1}^{(1)}, \ldots, s_{n-1}^{(K)},\ \cdots)^t.
\end{aligned}$$

For instance, the estimates can be rewritten concisely as

$$y_n = \sum_{k=0}^{K} \boldsymbol{q}^{(k)t} \boldsymbol{s}^{(k)}(n) = \boldsymbol{q}^t \boldsymbol{s}(n). \tag{2.12}$$

The length of $\boldsymbol{q}$ (and of $\boldsymbol{s}(n)$) will be denoted by N_q.

The source-specific unit vector $\mathbf{e}_\nu^{(k)}$ will also prove convenient. $\mathbf{e}_\nu^{(k)}$ is a column vector with a single nonzero element of value 1 located such that

$$\boldsymbol{q}^t \mathbf{e}_\nu^{(k)} = q_\nu^{(k)}.$$

At times we will also use the standard basis element $\mathbf{e}_\nu$, which has its nonzero element located at index ν.

We now point out two important properties of $\boldsymbol{q}$. First, recognize that a particular channel and set of estimator constraints will restrict the set of attainable global responses, which we will denote by $\mathcal{Q}_\mathrm{a}$. For example, when the estimator is finite impulse response (FIR) but otherwise unconstrained (i.e., $\mathcal{F}_\mathrm{a} = \mathbb{C}^{N_f}$), (2.11) implies that $\boldsymbol{q} \in \mathcal{Q}_\mathrm{a} = \mathrm{row}(\boldsymbol{\mathcal{H}})$, where

$$\boldsymbol{\mathcal{H}} := \begin{pmatrix}
\mathbf{h}_0^{(0)} \cdots \mathbf{h}_0^{(K)} & \mathbf{h}_1^{(0)} \cdots \mathbf{h}_1^{(K)} & \mathbf{h}_2^{(0)} \cdots \mathbf{h}_2^{(K)} & \cdots \\
\mathbf{0} \ \cdots \ \mathbf{0} & \mathbf{h}_0^{(0)} \cdots \mathbf{h}_0^{(K)} & \mathbf{h}_1^{(0)} \cdots \mathbf{h}_1^{(K)} & \cdots \\
\vdots \quad\quad \vdots & \vdots \quad\quad \vdots & \vdots \quad\quad \vdots & \\
\mathbf{0} \ \cdots \ \mathbf{0} & \mathbf{0} \ \cdots \ \mathbf{0} & \mathbf{h}_0^{(0)} \cdots \mathbf{h}_0^{(K)} & \cdots
\end{pmatrix}. \tag{2.13}$$

Restricting the estimator to be sparse or autoregressive, for example, would generate different attainable sets $\mathcal{Q}_{\mathrm{a}}$. Second, BIBO stable $\mathbf{f}(z)$ and $\mathbf{h}^{(k)}(z)$ imply BIBO stable $q^{(k)}(z)$, so that $\|\boldsymbol{q}^{(k)}\|_p$ exists for all $p \geq 1$, and thus $\|\boldsymbol{q}\|_p$ does as well.

Throughout the dissertation, we make the following assumptions on the $K+1$ source processes:

S1) For all k, $\{s_n^{(k)}\}$ is zero-mean i.i.d.

S2) The processes $\{s_n^{(0)}\}, \ldots, \{s_n^{(K)}\}$ are jointly statistically independent.

S3) For all k, $\mathrm{E}\{|s_n^{(k)}|^2\} = \sigma_s^2 \neq 0$.

S4) When discussing the SW criterion, $\mathcal{K}(s_n^{(0)}) \neq 0$, and when discussing the CM criterion, $\mathcal{K}(s_n^{(0)}) < 0$.

S5) If, for any k, $q^{(k)}(z)$ or $\{s_n^{(k)}\}$ is not real-valued, then $\mathrm{E}\{s_n^{(k)2}\} = 0$ for all k.

At this point we make a few observations about S1)–S5).

- Though S1) specifies that each source process must be identically distributed, it allows the sources to be distributed differently from one another.

- Though S1) requires that all sources of interference be white, the model (2.6) is capable of representing coloration in the *observed* interference through proper construction of the channels $\mathbf{h}^{(k)}(z)$ for $k \geq 1$.

- S3) can be asserted w.l.o.g. since interference power may be absorbed in the channels $\mathbf{h}^{(k)}(z)$.

- For the SW criterion (in Chapter 3), S4) requires that the desired source must be non-Gaussian, since $\mathcal{K}(s_n) = 0$ when $\{s_n\}$ is a Gaussian process satisfying S1) and S5). For the CM criterion (in Chapters 4–7), we impose the more stringent requirement of sub-Gaussian $\{s_n^{(0)}\}$. There is no restriction on the distribution of the interferers $\{s_n^{(k)}\}\big|_{k\neq 0}$, however.

- S5) requires all sources to be "circularly-symmetric" in the complex plane when any of the global responses or sources are complex-valued. (E.g., QAM sources are circularly symmetric while PAM sources are not.)

Kurtosis $\mathcal{K}(\cdot)$, introduced in Section 2.1.6 as another name for the fourth-order (auto-) cumulant $\mathcal{C}_4$, has a simple expression for zero-mean random processes. Specifically, we write the kurtosis of zero-mean $\{s_n^{(k)}\}$ as

$$\mathcal{K}_s^{(k)} \;:=\; \mathcal{K}(s_n^{(k)}) \;=\; \mathrm{E}\{|s_n^{(k)}|^4\} - 2\,\mathrm{E}^2\{|s_n^{(k)}|^2\} - \left|\mathrm{E}\{(s_n^{(k)})^2\}\right|^2. \tag{2.14}$$

The following kurtosis-based quantities will be used in Chapter 3. The definitions speak for themselves.

$$\mathcal{K}_s^{\min} \;:=\; \min_{0\le k\le K} \mathcal{K}_s^{(k)} \tag{2.15}$$

$$\mathcal{K}_s^{\max} \;:=\; \max_{0\le k\le K} \mathcal{K}_s^{(k)} \tag{2.16}$$

$$\rho_{\min} \;:=\; \frac{\mathcal{K}_s^{\min}}{\mathcal{K}_s^{(0)}} \tag{2.17}$$

$$\rho_{\max} \;:=\; \frac{\mathcal{K}_s^{\max}}{\mathcal{K}_s^{(0)}}. \tag{2.18}$$

We define the *normalized* kurtosis of zero-mean $\{s_n^{(k)}\}$ as

$$\kappa_s^{(k)} \;:=\; \frac{\mathrm{E}\big\{|s_n^{(k)}|^4\big\}}{\mathrm{E}^2\big\{|s_n^{(k)}|^2\big\}}. \tag{2.19}$$

Under the following definition of κ_g,

$$\kappa_g \;:=\; \begin{cases} 3, & s_n^{(k)} \in \mathbb{R},\ \forall k, n \\ 2, & \text{otherwise,} \end{cases} \tag{2.20}$$

and S3)-S5), the normalized and standard kurtoses are related through

$$\mathcal{K}(s_n^{(k)}) \;=\; (\kappa_s^{(k)} - \kappa_g)\sigma_s^4.$$

(See Appendix 4.A.1.) Note that, under S1) and S5), κ_g represents the normalized kurtosis of a Gaussian source. The following normalized-kurtosis-based quantities will be used in Chapters 4–6:

$$\kappa_s^{\min} \;:=\; \min_{0\le k\le K} \kappa_s^{(k)} \tag{2.21}$$

$$\kappa_s^{\max} \;:=\; \max_{0\le k\le K} \kappa_s^{(k)}. \tag{2.22}$$

Note that $\rho_{\min}$ and $\rho_{\max}$ from (2.17)–(2.18) can be written as

$$\rho_{\min} \;=\; \frac{\kappa_g - \kappa_s^{\min}}{\kappa_g - \kappa_s^{(0)}} \tag{2.23}$$

$$\rho_{\max} \;=\; \frac{\kappa_g - \kappa_s^{\max}}{\kappa_g - \kappa_s^{(0)}}. \tag{2.24}$$

2.3 Mean-Squared Error Criteria

The mean-squared error (MSE) criterion, defined below in (2.25), constitutes a well-known and useful measure of estimate performance. As a means of quantifying the performance of blind estimates, we would like to compare their MSE to the minimum achievable MSE given identical sources, channels, and estimator constraints. The inherent gain ambiguity associated with BEWP estimates (discussed in Section 2.1.3) prevents straightforward application of the MSE criterion, however. To circumvent the ambiguity problem, we employ the so-called conditionally unbiased MSE criterion, discussed below in Section 2.3.2. Unbiased MSE is directly related to signal-to-interference-plus-noise ratio (SINR), as shown in Section 2.3.3.

2.3.1 The Mean-Squared Error Criterion

The well-known MSE criterion is defined below in terms of estimate y_n and estimand $s^{(0)}_{n-\nu}$.

$$J_{\mathrm{m},\nu}(y_n) \;:=\; \mathrm{E}\big\{|y_n - s^{(0)}_{n-\nu}|^2\big\}. \tag{2.25}$$

Using S1)–S3), we can rewrite the previous equation in terms of global response $\boldsymbol{q}$:

$$J_{\mathrm{m},\nu}(\boldsymbol{q}) \;=\; \|\boldsymbol{q} - \mathbf{e}^{(0)}_{\nu}\|_2^2\, \sigma_s^2. \tag{2.26}$$

Denoting MMSE quantities by the subscript "m," Appendix 2.A shows that in the unconstrained (non-causal) IIR case, S1)–S3) imply that the MMSE channel-plus-estimator is

$$q^{(\ell)}_{\mathrm{m},\nu}(z) \;=\; z^{-\nu}\mathbf{h}^{(0)H}(\tfrac{1}{z^*})\left(\sum_k \mathbf{h}^{(k)}(z)\mathbf{h}^{(k)H}(\tfrac{1}{z^*})\right)^{\dagger}\mathbf{h}^{(\ell)}(z) \quad \text{for} \;\; \ell = 0,\dots,K, \tag{2.27}$$

while in the FIR case, S1)–S3) imply

$$\boldsymbol{q}_{\mathrm{m},\nu} \;=\; \boldsymbol{\mathcal{H}}^t(\boldsymbol{\mathcal{H}}^*\boldsymbol{\mathcal{H}}^t)^{\dagger}\boldsymbol{\mathcal{H}}^*\mathbf{e}^{(0)}_{\nu}. \tag{2.28}$$

Note from (2.28) that $\boldsymbol{q}_{\mathrm{m},\nu}$ is the projection of $\mathbf{e}^{(0)}_{\nu}$ onto the row space of $\boldsymbol{\mathcal{H}}^*$.

2.3.2 Unbiased Mean-Squared Error

We have earlier argued that, since both symbol power and channel gain are unknown in the BEWP scenario, blind estimates are bound to suffer gain ambiguity. To ensure that our estimator performance evaluation is meaningful in the face of such ambiguity, we base our evaluation on normalized versions of the blind estimates, where the normalization factor is chosen to be the receiver gain $q^{(0)}_{\nu}$. Given that the estimate y_n can be decomposed into signal and interference terms as

$$y_n \;=\; q^{(0)}_{\nu}s^{(0)}_{n-\nu} + \bar{\boldsymbol{q}}^t\bar{\boldsymbol{s}}(n), \tag{2.29}$$

where

$$\begin{aligned}\bar{\boldsymbol{q}} &:= \text{“}\boldsymbol{q} \text{ with the } q_\nu^{(0)} \text{ term removed”}\\ \bar{\boldsymbol{s}}(n) &:= \text{“}\boldsymbol{s}(n) \text{ with the } s_{n-\nu}^{(0)} \text{ term removed”},\end{aligned}$$

the normalized estimate $y_n/q_\nu^{(0)}$ can be referred to as "conditionally unbiased" since $\mathrm{E}\{y_n/q_\nu^{(0)}|s_{n-\nu}^{(0)}\} = s_{n-\nu}^{(0)}$.

The conditionally-unbiased MSE (UMSE) associated with y_n, an estimate of $s_{n-\nu}^{(0)}$, is then defined

$$J_{\mathrm{u},\nu}(y_n) := \mathrm{E}\big\{|y_n/q_\nu^{(0)} - s_{n-\nu}^{(0)}|^2\big\}. \tag{2.30}$$

Substituting (2.29) into (2.30), we find that

$$J_{\mathrm{u},\nu}(\boldsymbol{q}) = \frac{\mathrm{E}\big\{|\bar{\boldsymbol{q}}^t\bar{\boldsymbol{s}}(n)|^2\big\}}{|q_\nu^{(0)}|^2} = \frac{\|\bar{\boldsymbol{q}}\|_2^2}{|q_\nu^{(0)}|^2}\,\sigma_s^2, \tag{2.31}$$

where the second equality invokes assumptions S1)–S3).

2.3.3 Signal to Interference-Plus-Noise Ratio

Signal to interference-plus-noise ratio (SINR) is defined below.

$$\mathrm{SINR}_\nu := \frac{\mathrm{E}\big\{|q_\nu^{(0)}s_{n-\nu}^{(0)}|^2\big\}}{\mathrm{E}\big\{|\bar{\boldsymbol{q}}^t\bar{\boldsymbol{s}}(n)|^2\big\}} = \frac{|q_\nu^{(0)}|^2}{\|\bar{\boldsymbol{q}}\|_2^2}, \tag{2.32}$$

Note from (2.31) and (2.32) that SINR and UMSE have the simple relation

$$\mathrm{SINR}_\nu = \frac{\sigma_s^2}{J_{\mathrm{u},\nu}}.$$

2.4 The Shalvi-Weinstein Criterion

The so-called Shalvi-Weinstein (SW) criterion [Shalvi TIT 90] is defined as

$$\max\big|\mathcal{K}(y_n)\big| \quad \text{such that} \quad \sigma_y = 1, \tag{2.33}$$

where $\mathcal{K}(y)$ denotes kurtosis, previously defined in (2.14).

Though the criterion (2.33) has been attributed (in name) to Shalvi and Weinstein, it has a history that long predates the publication of [Shalvi TIT 90]. In fact, use of kurtosis as a blind estimation criterion can be traced back to Saunders [Saunders ETS 53] (see also [Kaiser PSY 58]) in the context of factor analysis, a technique used in the analysis of data stemming from psychology experiments [Nunnally Book 78]. Moreover, kurtosis-based blind estimation schemes were being implemented on electronic computers[4] as early as 1954! Two of these early techniques were popularly referred to as "quartimax" and "varimax." During the late 1970's, Wiggins [Wiggins GEO 77] used varimax for geophysical exploration (as discussed in Chapter 1). To better fit his application context, he renamed the method "minimum entropy deconvolution." Various other researchers, such as Claerbout [Claerbout SEP 78], Godfrey [Godfrey SEP 78], Gray [Gray Thesis 79], and Ooe and Ulrych [Ooe GP 79] studied and extended the minimum-entropy methods, but it was not until Donoho's work in 1981 [Donoho Chap 81] that constrained kurtosis maximization was rigorously analyzed and formally linked[5] to Shannon entropy (thereby justifying Wiggins' "minimum entropy" terminology). Section 2.1 of this thesis presented a tutorial summary of [Donoho Chap 81].

It was established in Section 2.1.6 that maximization of the normalized cumulant $|\bar{\mathcal{C}}_4(y_n)|$ leads to perfect blind estimation under ideal conditions. Since, by definition,

$$|\bar{\mathcal{C}}_4(y)| \;=\; \left|\frac{\mathcal{C}_4(y)}{\big(\mathcal{C}_2(y)\big)^2}\right| \;=\; \frac{|\,\mathcal{K}(y)|}{\sigma_y^4} \;=\; |\,\mathcal{K}(y)| \;\; \text{when} \;\; \sigma_y = 1, \tag{2.34}$$

the SW criterion will also yield perfect blind estimates under ideal conditions. Per-

[4] Neuhaus and Wrigley realized that this criterion "involved calculations too extensive for a desk calculator or punch-card mechanical computer. Consequently, they programmed the quartimax criterion for the Illiac... the University of Illinois electronic computer." [Kaiser PSY 58].

[5] It is interesting to note, however, that the first suggestion of a link between kurtosis and entropy came in 1954 [Ferguson PSY 54], just a few years after Shannon's revolutionary work [Shannon BSTJ 48]!

formance analysis of the SW criterion under the general *non*-ideal model of Section 2.2 will be given in Chapter 3. A brief review of previous work on SW criterion analysis appears in Section 3.1.

2.5 The Constant Modulus Criterion

The constant modulus (CM) criterion specifies minimization of the CM cost J_c, defined below in terms of the estimates $\{y_n\}$ and a design parameter γ.

$$J_\mathrm{c}(y_n) \;:=\; \mathrm{E}\big\{\big(|y_n|^2 - \gamma\big)^2\big\}. \tag{2.35}$$

Note that the CM criterion penalizes the dispersion of estimates $\{y_n\}$ from the fixed value γ.

Independently conceived by Godard [Godard TCOM 80] and Treichler and Agee [Treichler TASSP 83] in the early 1980s, minimization of the CM cost has become perhaps the most studied and implemented means of blind equalization for data communication over dispersive channels (see, e.g., [Johnson PROC 98] and the references within) and has also been used successfully as a means of blind beamforming (see, e.g., [Shynk TSP 96]). Consider, as evidence, the following quotes from lead researchers in the field.

> *"The most widely tested and used-in-practice blind equalizer."*
>
> —[Proakis SPIE 91]

> *"The most widely used blind equalization technique."*
>
> —[Liu PROC 98]

> *"The workhorse for blind equalization of QAM signals."*
>
> —[Treichler PROC 98]

The popularity of the CM criterion is usually attributed to

1. the excellent MSE performance of CM-minimizing estimates, and

2. the existence of a simple adaptive algorithm ("CMA" [Godard TCOM 80, Treichler TASSP 83]) for estimation and tracking of the CM-minimizing estimator $\mathbf{f}_c(z)$.

The close relationship between MMSE and CM-minimizing estimates was first conjectured in the seminal works by Godard and Treichler/Agee, and provides the theme for the recently-published comprehensive survey [Johnson PROC 98]. In Chapter 4, we quantify the MSE performance of CM-minimizing estimates and make this conjectured relationship precise. A brief review of previous work on this topic will be given in Section 4.1.

The SW and CM criteria, both a function of the second and fourth order cumulants of the estimate, are closely related. Expanding (2.35) and substituting (2.14),

$$\begin{aligned} \mathrm{E}\Big\{\big(|y|^2-\gamma\big)^2\Big\} &= \mathrm{E}\big\{|y|^4\big\} - 2\gamma\sigma_y^2 + \gamma^2 \\ &= \mathcal{K}(y) + 3\sigma_y^4 - 2\gamma\sigma_y^2 + \gamma^2 \\ &= \left(\frac{\mathcal{K}(y)}{\sigma_y^4} + 3\right)\sigma_y^4 - 2\gamma\sigma_y^2 + \gamma^2 \\ &= \Bigg(\underbrace{\operatorname{sgn}\big(\mathcal{K}(y)\big)\cdot\frac{|\mathcal{K}(y)|}{\sigma_y^4} + 3}_{\substack{\text{gain}\\ \text{independent}}}\Bigg)\underbrace{\sigma_y^4 - 2\gamma\sigma_y^2 + \gamma^2}_{\substack{\text{strictly}\\ \text{gain dependent}}} \end{aligned} \tag{2.36}$$

for the case of a real-valued sources. (The circularly-symmetric complex-valued source case is identical with the exception that the constant "3" in (2.36) is replaced by a "2.") (2.36) shows that the CM cost decouples into two components: one component which is strictly independent of σ_y, and another component which is strictly dependent on σ_y. The gain dependent component is of little interest because we have already established that absolute gain estimation is impossible. Then, since $\sigma_y^4 \geq 0$, minimization of the CM cost is equivalent to maximization of the gain independent component, and thus maximization of $|\,\mathcal{K}(y)\,|$ subject to $\sigma_y = 1$ as long as $\mathrm{sgn}(\mathcal{K}(y)) < 0$. This latter requirement is satisfied in typical data communication applications, but will fail in, e.g., speech applications. The close relationship between the SW and CM criteria was first noticed in [Li TSP 95] and later established under more general conditions in [Regalia SP 99].

Appendix

2.A Derivation of MMSE Estimators

In this section we derive the MMSE (i.e., Weiner) estimators for the linear model (2.12) under assumptions S1)-S3).

The IIR derivation will be carried out in the z-domain, where we use $s^{(k)}(z)$, $y(z)$, and $\mathbf{r}(z)$ to denote the z-transforms of $\{s_n^{(k)}\}$, $\{y_n\}$, and $\{\mathbf{r}_n\}$, respectively. From S1)–S3) we adopt the definition

$$\mathrm{E}\big(s^{(k)}(z)s^{(\ell)*}(\tfrac{1}{z^*})\big) \;:=\; \sigma_s^2\delta_{k-\ell}, \tag{2.37}$$

where $\delta_{k-\ell}$ denotes the Kronecker delta. Starting with the orthogonality principle of MMSE estimation

$$0 \;=\; \mathrm{E}\Big(\mathbf{r}^*(\tfrac{1}{z^*})\big(y_{\mathrm{m}}(z) - z^{-\nu}s^{(0)}(z)\big)\Big),$$

using $y_{\mathrm{m}}(z)$ to denote the sequence of MMSE estimates, we can apply (2.37) and z-domain equivalents of (2.6) and (2.12) to obtain

$$\begin{aligned}
0 &= \mathrm{E}\left(\sum_k \mathbf{h}^{(k)*}(\tfrac{1}{z^*})s^{(k)*}(\tfrac{1}{z^*})\left(\sum_\ell \mathbf{f}^H_{\mathrm{m},\nu}(z)\mathbf{h}^{(\ell)}(z)s^{(\ell)}(z) - z^{-\nu}s^{(0)}(z)\right)\right)\\
&= \sum_k\sum_\ell \mathbf{h}^{(k)*}(\tfrac{1}{z^*})\mathbf{f}^H_{\mathrm{m},\nu}(z)\mathbf{h}^{(\ell)}(z)\,\mathrm{E}\left(s^{(k)*}(\tfrac{1}{z^*})s^{(\ell)}(z)\right)\\
&\quad - z^{-\nu}\sum_k \mathbf{h}^{(k)*}(\tfrac{1}{z^*})\,\mathrm{E}\left(s^{(k)*}(\tfrac{1}{z^*})s^{(0)}(z)\right)\\
&= \sum_k \mathbf{h}^{(k)*}(\tfrac{1}{z^*})\mathbf{f}^H_{\mathrm{m},\nu}(z)\mathbf{h}^{(k)}(z)\sigma_s^2 - z^{-\nu}\mathbf{h}^{(0)*}(\tfrac{1}{z^*})\sigma_s^2\\
&= \left(\sum_k \mathbf{h}^{(k)*}(\tfrac{1}{z^*})\mathbf{h}^{(k)t}(z)\right)\mathbf{f}^*_{\mathrm{m},\nu}(z) - z^{-\nu}\mathbf{h}^{(0)*}(\tfrac{1}{z^*}).
\end{aligned}$$

Thus the (conjugate) MMSE estimator is

$$\mathbf{f}^*_{\mathrm{m},\nu}(z) \;=\; \left(\sum_k \mathbf{h}^{(k)*}(\tfrac{1}{z^*})\mathbf{h}^{(k)t}(z)\right)^{\dagger}\mathbf{h}^{(0)*}(\tfrac{1}{z^*})z^{-\nu}$$

which may be plugged into the z-domain equivalent of (2.11) to yield

$$\boldsymbol{q}^{(\ell)}_{\mathrm{m},\nu}(z) \;=\; z^{-\nu}\mathbf{h}^{(0)H}(\tfrac{1}{z^*})\Big(\sum_k \mathbf{h}^{(k)}(z)\mathbf{h}^{(k)H}(\tfrac{1}{z^*})\Big)^{\dagger}\mathbf{h}^{(\ell)}(z).$$

The FIR derivation is analogous, though performed in the time domain. Using $y_{\mathrm{m}}(n)$ to denote the MMSE estimates, the orthogonality principle can be stated as

$$\mathbf{0} \;=\; \mathrm{E}\Big(\boldsymbol{r}^*(n)\big(y_{\mathrm{m}}(n) - \mathbf{e}^{(0)t}_{\nu}\boldsymbol{s}(n)\big)\Big), \tag{2.38}$$

then using (2.7), (2.11), (2.38), source assumptions S1)–S3), and the fact that $\boldsymbol{r}(n) = \boldsymbol{\mathcal{H}}\boldsymbol{s}(n)$, we have

$$\begin{aligned}
\mathbf{0} \;&=\; \mathrm{E}\Big(\boldsymbol{\mathcal{H}}^*\boldsymbol{s}^*(n)\big(\boldsymbol{f}^H_{\mathrm{m},\nu}\boldsymbol{\mathcal{H}}\boldsymbol{s}(n) - \mathbf{e}^{(0)t}_{\nu}\boldsymbol{s}(n)\big)\Big) \\
&=\; \boldsymbol{\mathcal{H}}^*\,\mathrm{E}\big(\boldsymbol{s}^*(n)\boldsymbol{s}^t(n)\big)\boldsymbol{\mathcal{H}}^t\boldsymbol{f}^*_{\mathrm{m},\nu} - \boldsymbol{\mathcal{H}}^*\,\mathrm{E}\big(\boldsymbol{s}^*(n)\boldsymbol{s}^t(n)\big)\mathbf{e}^{(0)}_{\nu} \\
&=\; \boldsymbol{\mathcal{H}}^*\boldsymbol{\mathcal{H}}^t\boldsymbol{f}^*_{\mathrm{m},\nu}\sigma_s^2 - \boldsymbol{\mathcal{H}}^*\mathbf{e}^{(0)}_{\nu}\sigma_s^2 \\
&=\; \boldsymbol{\mathcal{H}}^*\boldsymbol{\mathcal{H}}^t\boldsymbol{f}^*_{\mathrm{m},\nu} - \boldsymbol{\mathcal{H}}^*\mathbf{e}^{(0)}_{\nu}.
\end{aligned}$$

Thus the (conjugate) MMSE estimator is

$$\boldsymbol{f}^*_{\mathrm{m},\nu} \;=\; \big(\boldsymbol{\mathcal{H}}^*\boldsymbol{\mathcal{H}}^t\big)^{\dagger}\boldsymbol{\mathcal{H}}^*\mathbf{e}^{(0)}_{\nu}$$

which yields

$$\boldsymbol{q}_{\mathrm{m},\nu} \;=\; \boldsymbol{\mathcal{H}}^t\boldsymbol{f}^*_{\mathrm{m},\nu} \;=\; \boldsymbol{\mathcal{H}}^t\big(\boldsymbol{\mathcal{H}}^*\boldsymbol{\mathcal{H}}^t\big)^{\dagger}\boldsymbol{\mathcal{H}}^*\mathbf{e}^{(0)}_{\nu}.$$

Chapter 3

Bounds for the MSE performance of SW Estimators[1]

3.1 Introduction

It was proven independently in [Donoho Chap 81] and [Shalvi TIT 90] that unconstrained linear estimators locally maximizing the SW criterion yield perfect blind estimates of a single non-Gaussian i.i.d. source transmitted through a noiseless invertible linear channel. In practical situations, however, we expect constrained estimators, noise and/or interference of a potentially non-Gaussian nature, and possibly non-invertible channels. Are Shalvi-Weinstein (SW) estimators useful in these cases? How do SW estimators compare to optimal (linear) estimators, say, in a mean square sense?

For a finite impulse response (FIR), but otherwise unconstrained, estimator and a noiseless FIR channel, Regalia and Mboup studied various properties of SW minimizers [Regalia TSP 99]. Though they provided evidence that the SW and MMSE

[1]The main results of this chapter also appear in the manuscript [Schniter TSP tbd2].

estimators are closely related in most cases, their approach did not lead to upper bounds on the performance of the SW estimator.

Recently, Feng and Chi studied the properties of unconstrained infinite-dimensional SW estimators of a non-Gaussian source in the presence of Gaussian noise [Feng TSP 99], [Feng TSP 00]. Using a frequency-domain approach, they observed relationships between the Weiner and SW estimators that bear similarity[2] to the time-domain relationships derived previously by Regalia and Mboup. The complexity of the analytical relationships derived by Feng and Chi prevents their translation into meaningful statements about the MSE performance of SW estimators, however.

In this chapter we study the performance of constrained ARMA SW estimators under the assumptions of the model in Section 2.2: desired source with arbitrary non-Gaussian distribution, interference with arbitrary distribution, and vector IIR (or FIR) channels. The main contributions of this chapter are (i) a simple test for the existence of a SW estimator for the desired source (defined more rigorously in Section 3.2.1), and (ii) bounding expressions for the MSE of SW estimators that are a function of the minimum MSE attainable under the same conditions. These bounds, derived under the multi-source linear model of Section 2.2, provide a formal link between the SW and Wiener estimators in a very general context.

The organization of the chapter is as follows. Section 3.2 derives bounds for the MSE performance of SW estimators, Section 3.3 presents the results of numerical simulations demonstrating the efficacy of our bounding techniques, and Section 3.4 concludes the chapter.

[2]Keep in mind that Regalia and Mboup studied constrained estimators in noiseless settings while Feng and Chi studied unconstrained estimators in noisy settings.

3.2 SW Performance under General Additive Interference

In this section we derive tight bounds for the UMSE of SW symbol estimators that

- have a closed-form expression,
- support arbitrary additive interference,
- support complex-valued channels and estimators, and
- support IIR (as well as FIR) channels and estimators.

Section 3.2.1 outlines our approach, Section 3.2.2 presents the main results, and Section 3.2.3 comments on these results. Proof details appear in Appendix 3.A.

3.2.1 The SW-UMSE Bounding Strategy

Since $y_n = \boldsymbol{q}^t \boldsymbol{s}(n)$ for $\boldsymbol{q} \in \mathcal{Q}_{\mathrm{a}}$, source assumptions S1)-S5) imply that [Porat Book 94]

$$\mathcal{K}(y_n) = \sum_k \|\boldsymbol{q}^{(k)}\|_4^4 \, \mathcal{K}_s^{(k)} \tag{3.1}$$

$$\sigma_y^2 = \|\boldsymbol{q}^{(k)}\|_2^2 \sigma_s^2. \tag{3.2}$$

This allows us to rewrite the SW criterion (2.33) as

$$\max_{\boldsymbol{q} \in \mathcal{Q}_{\mathrm{a}} \cap \mathcal{Q}_{\mathrm{s}}} \left| \sum_k \|\boldsymbol{q}^{(k)}\|_4^4 \, \mathcal{K}_s^{(k)} \right|$$

where $\mathcal{Q}_{\mathrm{s}}$ denotes the set of unit-norm global responses: $\mathcal{Q}_{\mathrm{s}} := \{\boldsymbol{q} \text{ s.t. } \|\boldsymbol{q}\|_2 = 1\}$.

Though the SW criterion admits multiple solutions, we are only interested in those that correspond to the estimation of the 0^{th} user's symbols at delay ν. We define the set of global responses *associated*[3] with the {user, delay} pair $\{0, \nu\}$ as follows:

$$\mathcal{Q}_\nu^{(0)} \;:=\; \Big\{ \boldsymbol{q} \text{ s.t. } |q_\nu^{(0)}| > \max_{(k,\delta)\neq(0,\nu)} |q_\delta^{(k)}| \Big\}.$$

The set[4] of SW global responses associated with the $\{0, \nu\}$ pair is then defined by the following local maxima:

$$\{\boldsymbol{q}_{\mathrm{sw},\nu}\} \;:=\; \Big\{ \arg \max_{\boldsymbol{q}\in\mathcal{Q}_\mathrm{a}\cap\mathcal{Q}_\mathrm{s}} \Big| \sum_k \|\boldsymbol{q}^{(k)}\|_4^4\, \mathcal{K}_s^{(k)} \Big| \Big\} \cap \mathcal{Q}_\nu^{(0)}.$$

It is not possible to write general closed-form expressions for $\{\boldsymbol{q}_{\mathrm{sw},\nu}\}$, making it difficult to characterize their performance. In fact, Appendix 3.A.1 shows that $\{\boldsymbol{q}_{\mathrm{sw},\nu}\}$ may be empty, though for the discussion below we assume that this is not the case.

Consider a reference global response $\boldsymbol{q}_{\mathrm{r},\nu} \in \mathcal{Q}_\mathrm{a} \cap \mathcal{Q}_\mathrm{s} \cap \mathcal{Q}_\nu^{(0)}$. In other words, $\boldsymbol{q}_{\mathrm{r},\nu}$ is an attainable unit-norm response associated with user/delay $\{0, \nu\}$. When $\boldsymbol{q}_{\mathrm{r},\nu}$ is in the vicinity of a $\boldsymbol{q}_{\mathrm{sw},\nu}$ (the meaning of which will be made more precise later), we know that

$$\Big| \sum_k \|\boldsymbol{q}_{\mathrm{sw},\nu}^{(k)}\|_4^4\, \mathcal{K}_s^{(k)} \Big| \;\geq\; \Big| \sum_k \|\boldsymbol{q}_{\mathrm{r},\nu}^{(k)}\|_4^4\, \mathcal{K}_s^{(k)} \Big| \;=\; \big|\mathcal{K}(y_\mathrm{r})\big|.$$

Thus this $\boldsymbol{q}_{\mathrm{sw},\nu}$ lies in the following set of global responses:

$$\mathcal{Q}_\mathrm{sw}(\boldsymbol{q}_{\mathrm{r},\nu}) \;:=\; \Big\{ \boldsymbol{q} \;\text{ s.t. }\; \Big| \sum_k \|\boldsymbol{q}^{(k)}\|_4^4\, \mathcal{K}_s^{(k)} \Big| \geq \big|\mathcal{K}(y_\mathrm{r})\big| \Big\} \cap \mathcal{Q}_\nu^{(0)} \cap \mathcal{Q}_\mathrm{s}. \tag{3.3}$$

from which an SW-UMSE upper bound may be computed:

$$J_{\mathrm{u},\nu}(\boldsymbol{q}_{\mathrm{sw},\nu}) \;\leq\; \max_{\boldsymbol{q}\in\mathcal{Q}_\mathrm{sw}(\boldsymbol{q}_{\mathrm{r},\nu})} J_{\mathrm{u},\nu}(\boldsymbol{q}). \tag{3.4}$$

[3] Note that under S1)–S3), a particular user/delay combination is "associated" with an estimate if and only if that user/delay contributes more energy to the estimate than any other user/delay.

[4] We refer to the SW responses as a set to avoid establishing existence or uniqueness of local maxima within $\mathcal{Q}_\nu^{(0)}$ at this time.

Note that (3.4) avoids explicit consideration of the admissibility constraints of $\mathcal{Q}_{\rm a}$; they are implicitly incorporated via reference $\boldsymbol{q}_{{\rm r},\nu} \in \mathcal{Q}_{\rm a}$. Also note that the tightness of the upper bound (3.4) will depend on the size and shape of $\mathcal{Q}_{\rm sw}(\boldsymbol{q}_{{\rm r},\nu})$, motivating careful choice of $\boldsymbol{q}_{{\rm r},\nu}$. In the sequel we choose the scaled MMSE reference $\boldsymbol{q}_{{\rm r},\nu} = \boldsymbol{q}_{{\rm m},\nu}/\|\boldsymbol{q}_{{\rm m},\nu}\|_2$ (when $\boldsymbol{q}_{{\rm m},\nu} \in \mathcal{Q}_\nu^{(0)}$) since it is an established benchmark with a closed-form expression.

Two simplifications will ease the evaluation of bound (3.4). The first is the removal of absolute value signs in the definition (3.3). Recognize that for $\boldsymbol{q}$ sufficiently close to $\boldsymbol{e}_\nu^{(0)}$, $\mathrm{sgn}\big(\sum_k \|\boldsymbol{q}^{(k)}\|_4^4\,\mathcal{K}_s^{(k)}\big) = \mathrm{sgn}\big(\mathcal{K}_s^{(k)}\big)$, in which case

$$\left|\sum_k \|\boldsymbol{q}^{(k)}\|_4^4\,\mathcal{K}_s^{(k)}\right| \;=\; \mathrm{sgn}\big(\mathcal{K}_s^{(k)}\big)\sum_k \|\boldsymbol{q}^{(k)}\|_4^4\,\mathcal{K}_s^{(k)}. \tag{3.5}$$

Our bounds will impose conditions that ensure this behavior.

Next, since both the SW and UMSE criteria are invariant to phase rotation of $\boldsymbol{q}$ (i.e., scalar multiplication of $\boldsymbol{q}$ by $e^{j\phi}$ for $\phi \in \mathbb{R}$), we can restrict the our attention to the set of "de-rotated" global responses $\{\boldsymbol{q} \text{ s.t. } q_\nu^{(0)} \in \mathbb{R}^+\}$. For de-rotated responses $\boldsymbol{q} \in \mathcal{Q}_{\rm s} \cap \mathcal{Q}_\nu^{(0)}$, we know $q_\nu^{(0)} = \sqrt{1-\|\bar{\boldsymbol{q}}\|_2^2}$, which implies that such $\boldsymbol{q}$ are completely described by their interference response $\bar{\boldsymbol{q}}$ (as described in Section 2.3.2). Moreover, these interference responses lie within $\bar{\mathcal{Q}}_\nu^{(0)}$, the projection of $\mathcal{Q}_\nu^{(0)} \cap \mathcal{Q}_{\rm s}$ onto $\{\bar{\boldsymbol{q}}\}$:

$$\bar{\mathcal{Q}}_\nu^{(0)} \;:=\; \Big\{\bar{\boldsymbol{q}} \text{ s.t. } \sqrt{1-\|\bar{\boldsymbol{q}}\|_2^2} > \max_{(k,\delta)\neq(0,\nu)} |q_\delta^{(k)}|\Big\}.$$

(See Fig. 3.1 for the construction of $\bar{\mathcal{Q}}_\nu^{(0)}$, whose boundary is illustrated by the thick shaded curves.) Using this parameterization, (2.31) and (3.2) imply

$$\begin{aligned}
\left.\frac{J_{{\rm u},\nu}(\boldsymbol{q}_{{\rm sw},\nu})}{\sigma_s^2}\right|_{\boldsymbol{q}\in\mathcal{Q}_{\rm s}\cap\mathcal{Q}_\nu^{(0)}} \;&=\; \frac{\|\bar{\boldsymbol{q}}\|_2^2}{1-\|\bar{\boldsymbol{q}}\|_2^2} \\
\left.\sum_k \|\boldsymbol{q}^{(k)}\|_4^4\,\mathcal{K}_s^{(k)}\right|_{\boldsymbol{q}\in\mathcal{Q}_{\rm s}\cap\mathcal{Q}_\nu^{(0)}} \;&=\; \big(1-\|\bar{\boldsymbol{q}}\|_2^2\big)^2\,\mathcal{K}_s^{(0)} + \sum_k \|\bar{\boldsymbol{q}}^{(k)}\|_4^4\,\mathcal{K}_s^{(k)}.
\end{aligned} \tag{3.6}$$

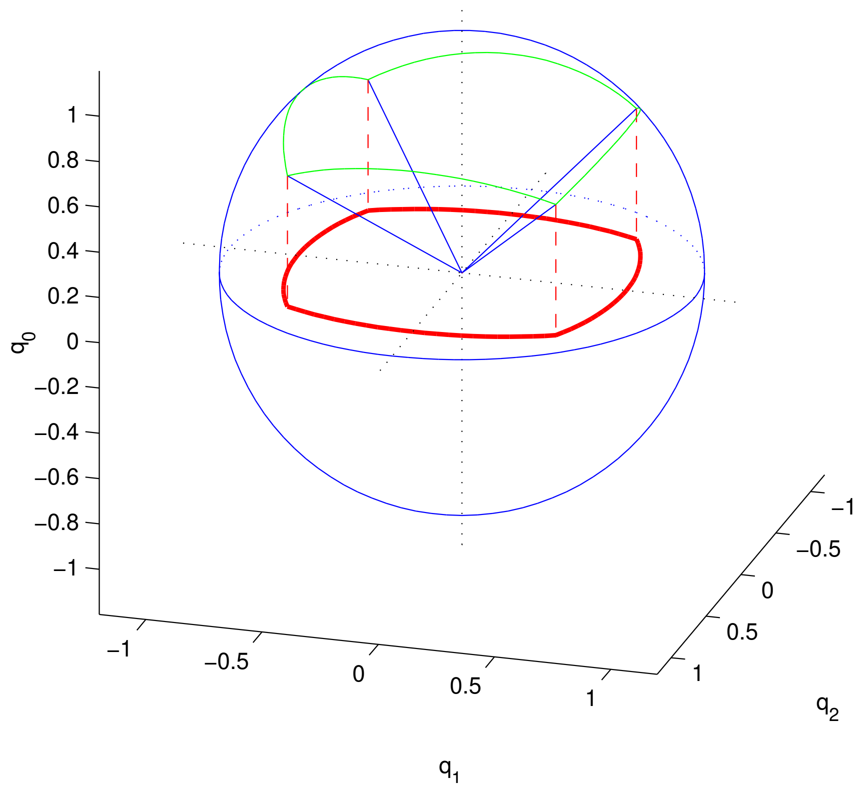

Figure 3.1: $\bar{\mathcal{Q}}_{\nu}^{(0)}$, created by projecting $\mathcal{Q}_{\nu}^{(0)} \cap \mathcal{Q}_{\mathrm{s}}$ onto the interference space, is illustrated here for the three-dimensional case. The boundary of $\bar{\mathcal{Q}}_{\nu}^{(0)}$ is demarcated by the thick shaded curves.

With the two simplifications above, (3.4) becomes

$$J_{\mathrm{u},\nu}(\boldsymbol{q}_{\mathrm{sw},\nu}) \;\leq\; \max_{\bar{\boldsymbol{q}} \in \bar{\mathcal{Q}}_{\mathrm{sw}}(\boldsymbol{q}_{\mathrm{r},\nu})} J_{\mathrm{u},\nu}(\bar{\boldsymbol{q}}).$$

where $\bar{\mathcal{Q}}_{\mathrm{sw}}$ is the following $\{\bar{\boldsymbol{q}}\}$-space projection of $\mathcal{Q}_{\mathrm{sw}}$:

$$\bar{\mathcal{Q}}_{\mathrm{sw}}(\boldsymbol{q}_{\mathrm{r},\nu}) \;:=\; \begin{cases} \Big\{\bar{\boldsymbol{q}} \in \bar{\mathcal{Q}}_{\nu}^{(0)} \text{ s.t. } \big(1-\|\bar{\boldsymbol{q}}\|_2^2\big)^2 \mathcal{K}_s^{(0)} + \sum_k \|\bar{\boldsymbol{q}}^{(k)}\|_4^4 \mathcal{K}_s^{(k)} \geq \mathcal{K}(y_{\mathrm{r}})\Big\}, \\ \hfill \text{for} \quad \mathcal{K}_s^{(0)} > 0, \\ \Big\{\bar{\boldsymbol{q}} \in \bar{\mathcal{Q}}_{\nu}^{(0)} \text{ s.t. } \big(1-\|\bar{\boldsymbol{q}}\|_2^2\big)^2 \mathcal{K}_s^{(0)} + \sum_k \|\bar{\boldsymbol{q}}^{(k)}\|_4^4 \mathcal{K}_s^{(k)} \leq \mathcal{K}(y_{\mathrm{r}})\Big\}, \\ \hfill \text{for} \quad \mathcal{K}_s^{(0)} < 0. \end{cases} \tag{3.7}$$

Finally, since $J_{\mathrm{u},\nu}(\bar{\boldsymbol{q}})$ is strictly increasing in $\|\bar{\boldsymbol{q}}\|_2$ (over its valid range), we claim

$$J_{\mathrm{u},\nu}(\boldsymbol{q}_{\mathrm{sw},\nu}) \;\leq\; \frac{b_*^2}{1-b_*^2} \quad \text{where} \quad b_* \;:=\; \max_{\bar{\boldsymbol{q}} \in \bar{\mathcal{Q}}_{\mathrm{sw}}(\boldsymbol{q}_{\mathrm{r},\nu})} \|\bar{\boldsymbol{q}}\|_2. \tag{3.8}$$

The constrained maximization of b_* can be restated as the following minimization.

$$b_* \;=\; \min b \text{ s.t. } \Big\{\bar{\boldsymbol{q}} \in \bar{\mathcal{Q}}_{\mathrm{sw}}(\boldsymbol{q}_{\mathrm{r},\nu}) \Rightarrow \|\bar{\boldsymbol{q}}\|_2 \leq b\Big\} \tag{3.9}$$

Fig. 3.2 presents a summary of the bounding procedure in the interference response space $\{\bar{\boldsymbol{q}}\}$. The set of attainable interference responses is denoted by $\bar{\mathcal{Q}}_{\mathrm{a}}$, which can be interpreted as a projection of $\mathcal{Q}_{\mathrm{a}} \cap \mathcal{Q}_{\mathrm{s}} \cap \mathcal{Q}_{\nu}^{(0)}$ onto $\{\bar{\boldsymbol{q}}\}$. Notice that the reference response $\bar{\boldsymbol{q}}_{\mathrm{r},\nu}$ and the SW response $\bar{\boldsymbol{q}}_{\mathrm{sw},\nu}$ both lie in $\bar{\mathcal{Q}}_{\mathrm{a}}$. Though the exact location of $\bar{\boldsymbol{q}}_{\mathrm{sw},\nu}$ is unknown, we know that it is contained by $\bar{\mathcal{Q}}_{\mathrm{sw}}(\boldsymbol{q}_{\mathrm{r},\nu})$, depicted in Fig. 3.2 by the shaded region. Thus, an upper bound on the UMSE of the SW estimator can be calculated using b_*, the maximum interference radius over $\bar{\mathcal{Q}}_{\mathrm{sw}}(\boldsymbol{q}_{\mathrm{r},\nu})$. As a cautionary note, there exist situations where the shape of $\bar{\mathcal{Q}}_{\mathrm{sw}}(\boldsymbol{q}_{\mathrm{r},\nu})$ prevents containment by a $\bar{\boldsymbol{q}}$-space ball. In the next section we present conditions (derived in Appendix 3.A.1) which avoid these problematic situations.

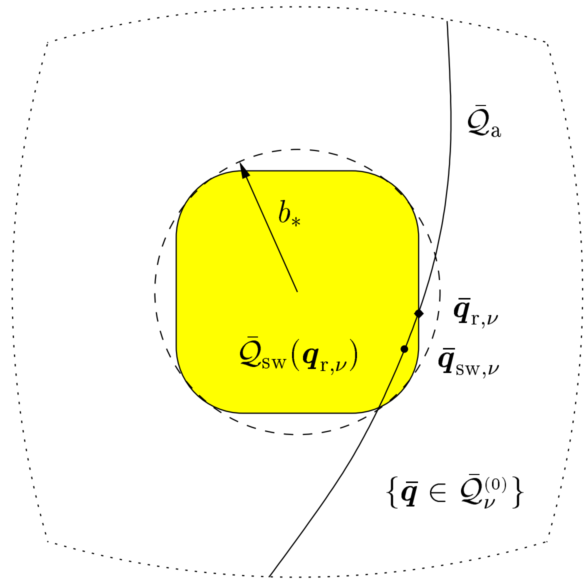

Figure 3.2: Illustration of SW-UMSE bounding technique in the interference response space $\{\bar{\boldsymbol{q}}\}$.

3.2.2 The SW-UMSE Bounds

In this section we present SW-UMSE bounds based on the method described in Section 3.2.1. Proofs appear in Appendix 3.A.

Theorem 3.1. *When $\mathcal{K}(y_{\mathrm{m}})$, the kurtosis of estimates generated by the Wiener estimator associated with the desired user at delay ν, obeys*

$$\begin{cases} \mathcal{K}_s^{(0)} \;\geq\; \mathcal{K}(y_{\mathrm{m}}) \;>\; \big(\mathcal{K}_s^{(0)} + \mathcal{K}_s^{\max}\big)/4, & \textit{for } \mathcal{K}_s^{(0)} > 0, \\ \mathcal{K}_s^{(0)} \;\leq\; \mathcal{K}(y_{\mathrm{m}}) \;<\; \big(\mathcal{K}_s^{(0)} + \mathcal{K}_s^{\min}\big)/4, & \textit{for } \mathcal{K}_s^{(0)} < 0, \end{cases} \tag{3.10}$$

the UMSE of SW estimators associated with the same user/delay can be upper *bounded by* $J_{\mathrm{u},\nu}\big|_{\mathrm{sw},\nu}^{\max,\mathcal{K}(y_{\mathrm{m}})}$, *where*

$$J_{\mathrm{u},\nu}\big|_{\mathrm{sw},\nu}^{\max,\mathcal{K}(y_{\mathrm{m}})} := \begin{cases} \dfrac{1-\sqrt{(\rho_{\max}+1)\frac{\mathcal{K}(y_{\mathrm{m}})}{\mathcal{K}_s^{(0)}}-\rho_{\max}}}{\rho_{\max}+\sqrt{(\rho_{\max}+1)\frac{\mathcal{K}(y_{\mathrm{m}})}{\mathcal{K}_s^{(0)}}-\rho_{\max}}}\,\sigma_s^2, & \text{for } \mathcal{K}_s^{(0)} > 0, \\ \dfrac{1-\sqrt{(\rho_{\min}+1)\frac{\mathcal{K}(y_{\mathrm{m}})}{\mathcal{K}_s^{(0)}}-\rho_{\min}}}{\rho_{\min}+\sqrt{(\rho_{\min}+1)\frac{\mathcal{K}(y_{\mathrm{m}})}{\mathcal{K}_s^{(0)}}-\rho_{\min}}}\,\sigma_s^2, & \text{for } \mathcal{K}_s^{(0)} < 0. \end{cases} \tag{3.11}$$

Furthermore, (3.10) guarantees the existence of a SW estimator associated with this user/delay when $\boldsymbol{q}$ *is FIR.*

While Theorem 3.1 presents a closed-form SW-UMSE bounding expression in terms of the kurtosis of the MMSE estimates, it is also possible to derive lower and upper bounds in terms of the UMSE of the MMSE estimator.

Theorem 3.2. *If* $J_{\mathrm{u},\nu}(\boldsymbol{q}_{\mathrm{m},\nu}) < J_{\mathrm{o}}\sigma_s^2$, *where*

$$J_{\mathrm{o}} := \begin{cases} 2\sqrt{(1+\rho_{\max})^{-1}}-1 & \mathcal{K}_s^{(0)} > 0,\ \ \mathcal{K}_s^{\min} \geq 0 \\ \dfrac{1-\sqrt{1-(3-\rho_{\max})(1+\rho_{\min})/4}}{\rho_{\min}+\sqrt{1-(3-\rho_{\max})(1+\rho_{\min})/4}}, & \mathcal{K}_s^{(0)} > 0,\ \ \mathcal{K}_s^{\min} < 0,\ \ \mathcal{K}_s^{\min} \neq -\mathcal{K}_s^{(0)} \\ \frac{3-\rho_{\max}}{5+\rho_{\max}} & \mathcal{K}_s^{(0)} > 0,\ \ \mathcal{K}_s^{\min} < 0,\ \ \mathcal{K}_s^{\min} = -\mathcal{K}_s^{(0)} \\ 2\sqrt{(1+\rho_{\min})^{-1}}-1 & \mathcal{K}_s^{(0)} < 0,\ \ \mathcal{K}_s^{\max} \leq 0 \\ \dfrac{1-\sqrt{1-(3-\rho_{\min})(1+\rho_{\max})/4}}{\rho_{\max}+\sqrt{1-(3-\rho_{\min})(1+\rho_{\max})/4}}, & \mathcal{K}_s^{(0)} < 0,\ \ \mathcal{K}_s^{\max} > 0,\ \ \mathcal{K}_s^{\max} \neq -\mathcal{K}_s^{(0)} \\ \frac{3-\rho_{\min}}{5+\rho_{\min}} & \mathcal{K}_s^{(0)} < 0,\ \ \mathcal{K}_s^{\max} > 0,\ \ \mathcal{K}_s^{\max} = -\mathcal{K}_s^{(0)} \end{cases} \tag{3.12}$$

the UMSE of SW estimators associated with the same user/delay can be bounded as follows:

$$J_{\mathrm{u},\nu}(\boldsymbol{q}_{\mathrm{m},\nu}) \ \leq\ J_{\mathrm{u},\nu}(\boldsymbol{q}_{\mathrm{sw},\nu}) \ \leq\ J_{\mathrm{u},\nu}\big|_{\mathrm{sw},\nu}^{\max,\mathcal{K}(y_{\mathrm{m}})} \ \leq\ J_{\mathrm{u},\nu}\big|_{\mathrm{sw},\nu}^{\max,J_{\mathrm{u},\nu}(\boldsymbol{q}_{\mathrm{m},\nu})},$$

where

$$J_{\mathrm{u},\nu}\Big|_{\mathrm{sw},\nu}^{\max,J_{\mathrm{u},\nu}(\boldsymbol{q}_{\mathrm{m},\nu})} := \begin{cases} \dfrac{1-\sqrt{(1+\rho_{\max})\left(1+\frac{J_{\mathrm{u},\nu}(\boldsymbol{q}_{\mathrm{m},\nu})}{\sigma_s^2}\right)^{-2}-\rho_{\max}}}{\rho_{\max}+\sqrt{(1+\rho_{\max})\left(1+\frac{J_{\mathrm{u},\nu}(\boldsymbol{q}_{\mathrm{m},\nu})}{\sigma_s^2}\right)^{-2}-\rho_{\max}}}\,\sigma_s^2 & \mathcal{K}_s^{(0)}>0,\ \ \mathcal{K}_s^{\min}\geq 0 \\ \dfrac{1-\sqrt{(1+\rho_{\max})\left(1+\frac{J_{\mathrm{u},\nu}(\boldsymbol{q}_{\mathrm{m},\nu})}{\sigma_s^2}\right)^{-2}\left(1+\rho_{\min}\frac{J^2_{\mathrm{u},\nu}(\boldsymbol{q}_{\mathrm{m},\nu})}{\sigma_s^4}\right)-\rho_{\max}}}{\rho_{\max}+\sqrt{(1+\rho_{\max})\left(1+\frac{J_{\mathrm{u},\nu}(\boldsymbol{q}_{\mathrm{m},\nu})}{\sigma_s^2}\right)^{-2}\left(1+\rho_{\min}\frac{J^2_{\mathrm{u},\nu}(\boldsymbol{q}_{\mathrm{m},\nu})}{\sigma_s^4}\right)-\rho_{\max}}}\,\sigma_s^2 & \mathcal{K}_s^{(0)}>0,\ \ \mathcal{K}_s^{\min}<0 \\ \dfrac{1-\sqrt{(1+\rho_{\min})\left(1+\frac{J_{\mathrm{u},\nu}(\boldsymbol{q}_{\mathrm{m},\nu})}{\sigma_s^2}\right)^{-2}-\rho_{\min}}}{\rho_{\min}+\sqrt{(1+\rho_{\min})\left(1+\frac{J_{\mathrm{u},\nu}(\boldsymbol{q}_{\mathrm{m},\nu})}{\sigma_s^2}\right)^{-2}-\rho_{\min}}}\,\sigma_s^2 & \mathcal{K}_s^{(0)}<0,\ \ \mathcal{K}_s^{\max}\leq 0 \\ \dfrac{1-\sqrt{(1+\rho_{\min})\left(1+\frac{J_{\mathrm{u},\nu}(\boldsymbol{q}_{\mathrm{m},\nu})}{\sigma_s^2}\right)^{-2}\left(1+\rho_{\max}\frac{J^2_{\mathrm{u},\nu}(\boldsymbol{q}_{\mathrm{m},\nu})}{\sigma_s^4}\right)-\rho_{\min}}}{\rho_{\min}+\sqrt{(1+\rho_{\min})\left(1+\frac{J_{\mathrm{u},\nu}(\boldsymbol{q}_{\mathrm{m},\nu})}{\sigma_s^2}\right)^{-2}\left(1+\rho_{\max}\frac{J^2_{\mathrm{u},\nu}(\boldsymbol{q}_{\mathrm{m},\nu})}{\sigma_s^4}\right)-\rho_{\min}}}\,\sigma_s^2 & \mathcal{K}_s^{(0)}<0,\ \ \mathcal{K}_s^{\max}>0. \end{cases} \tag{3.13}$$

Furthermore, (3.12) guarantees the existence of a SW estimator associated with this user/delay when $\boldsymbol{q}$ *is FIR.*

Equation (3.13) leads to an elegant approximation of the *extra* UMSE of SW estimators:

$$\mathcal{E}_{\mathrm{u},\nu}(\boldsymbol{q}_{\mathrm{sw},\nu}) \ := \ J_{\mathrm{u},\nu}(\boldsymbol{q}_{\mathrm{sw},\nu}) - J_{\mathrm{u},\nu}(\boldsymbol{q}_{\mathrm{m},\nu}).$$

Theorem 3.3. *If* $J_{\mathrm{u},\nu}(\boldsymbol{q}_{\mathrm{m},\nu}) < J_{\mathrm{o}}\sigma_s^2$*, then the extra UMSE of SW estimators can be* *bounded as* $\mathcal{E}_{\mathrm{u},\nu}(\boldsymbol{q}_{\mathrm{sw},\nu}) \leq \mathcal{E}_{\mathrm{u},\nu}\Big|_{\mathrm{c},\nu}^{\max,J_{\mathrm{u},\nu}(\boldsymbol{q}_{\mathrm{m},\nu})}$, *where*

$$
\begin{aligned}
&\mathcal{E}_{\mathrm{u},\nu}\Big|_{\mathrm{c},\nu}^{\max,J_{\mathrm{u},\nu}(\boldsymbol{q}_{\mathrm{m},\nu})} \\
&\quad := \quad J_{\mathrm{u},\nu}\Big|_{\mathrm{sw},\nu}^{\max,J_{\mathrm{u},\nu}(\boldsymbol{q}_{\mathrm{m},\nu})} - J_{\mathrm{u},\nu}(\boldsymbol{q}_{\mathrm{m},\nu}) \\
&\quad = \quad \begin{cases}
\frac{1}{2\sigma_s^2}\rho_{\max} J^2_{\mathrm{u},\nu}(\boldsymbol{q}_{\mathrm{m},\nu}) + \mathcal{O}\big(J^3_{\mathrm{u},\nu}(\boldsymbol{q}_{\mathrm{m},\nu})\big) & \mathcal{K}_s^{(0)} > 0, \ \ \mathcal{K}_s^{\max} \geq 0 \\
\frac{1}{2\sigma_s^2}(\rho_{\max} - \rho_{\min}) J^2_{\mathrm{u},\nu}(\boldsymbol{q}_{\mathrm{m},\nu}) + \mathcal{O}\big(J^3_{\mathrm{u},\nu}(\boldsymbol{q}_{\mathrm{m},\nu})\big) & \mathcal{K}_s^{(0)} > 0, \ \ \mathcal{K}_s^{\max} < 0 \\
\frac{1}{2\sigma_s^2}\rho_{\min} J^2_{\mathrm{u},\nu}(\boldsymbol{q}_{\mathrm{m},\nu}) + \mathcal{O}\big(J^3_{\mathrm{u},\nu}(\boldsymbol{q}_{\mathrm{m},\nu})\big) & \mathcal{K}_s^{(0)} < 0, \ \ \mathcal{K}_s^{\max} \leq 0 \\
\frac{1}{2\sigma_s^2}(\rho_{\min} - \rho_{\max}) J^2_{\mathrm{u},\nu}(\boldsymbol{q}_{\mathrm{m},\nu}) + \mathcal{O}\big(J^3_{\mathrm{u},\nu}(\boldsymbol{q}_{\mathrm{m},\nu})\big) & \mathcal{K}_s^{(0)} < 0, \ \ \mathcal{K}_s^{\max} > 0
\end{cases}
\end{aligned}
\tag{3.14}
$$

Equation (3.14) implies that the extra UMSE of SW estimators is upper bounded by approximately the *square* of the minimum UMSE. Fig. 3.3 plots the upper bound on SW-UMSE and extra SW-UMSE from (3.13) as a function of $J_{\mathrm{u},\nu}(\boldsymbol{q}_{\mathrm{m},\nu})/\sigma_s^2$ for various values of $\rho_{\min}$ and $\rho_{\max}$. The second-order approximation based on (3.14) appears very good for all but the largest values of UMSE.

3.2.3 Comments on the SW-UMSE Bounds

Implicit Incorporation of $\mathcal{Q}_{\mathrm{a}}$

First, recall that the SW-UMSE bounding procedure incorporated $\mathcal{Q}_{\mathrm{a}}$, the set of attainable global responses, *only* in the requirement that $\boldsymbol{q}_{\mathrm{r},\nu} \in \mathcal{Q}_{\mathrm{a}} \cap \mathcal{Q}_{\mathrm{s}} \cap \mathcal{Q}_\nu^{(0)}$. Thus Theorems 3.1–3.3, written under the reference choice $\boldsymbol{q}_{\mathrm{r},\nu} = \boldsymbol{q}_{\mathrm{m},\nu}/\|\boldsymbol{q}_{\mathrm{m},\nu}\|_2 \in \mathcal{Q}_{\mathrm{a}} \cap \mathcal{Q}_{\mathrm{s}} \cap \mathcal{Q}_\nu^{(0)}$, implicitly incorporate the channel and/or estimator constraints that define $\mathcal{Q}_{\mathrm{a}}$. For example, if $\boldsymbol{q}_{\mathrm{m},\nu}$ is the MMSE response constrained to the set of causal IIR estimators, then SW-UMSE bounds based on this $\boldsymbol{q}_{\mathrm{m},\nu}$ will implicitly incorporate the causality constraint. The implicit incorporation of the attainable set $\mathcal{Q}_{\mathrm{a}}$ makes these bounding theorems quite general and easy to use.

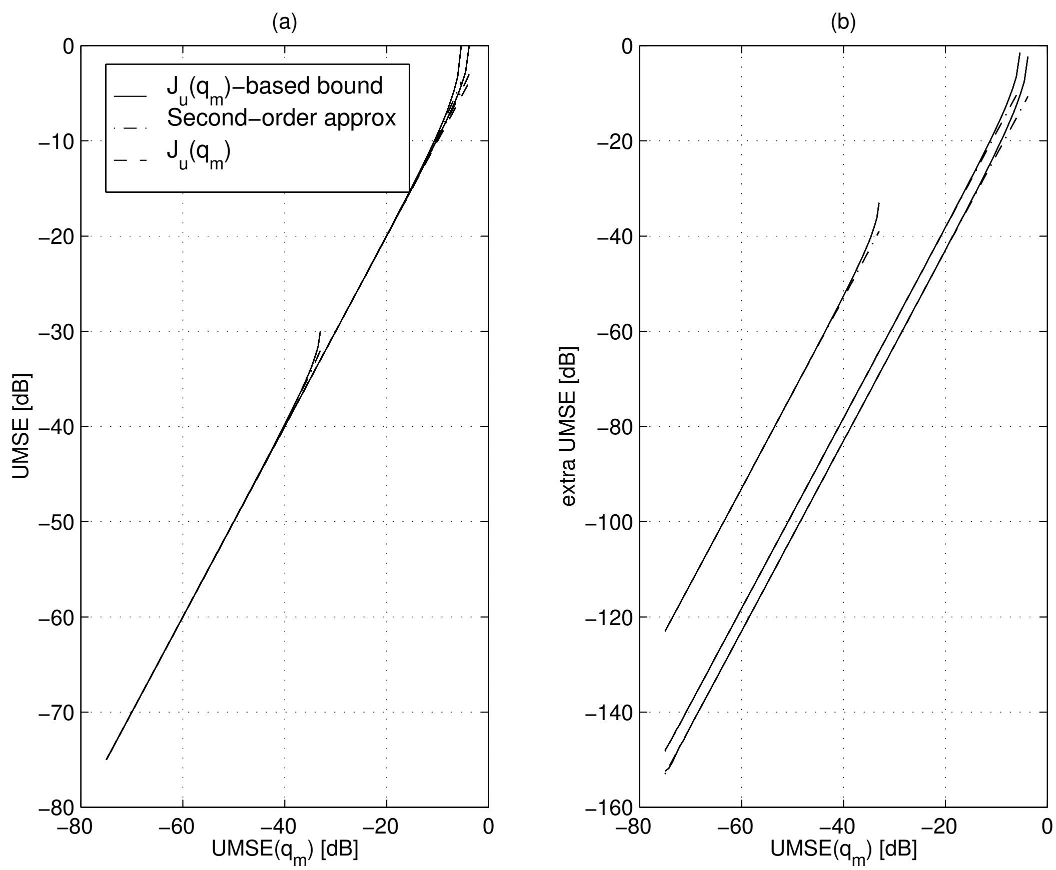

Figure 3.3: Upper bound on (a) SW-UMSE and (b) extra SW-UMSE versus $J_{\mathrm{u},\nu}(\boldsymbol{q}_{\mathrm{m},\nu})$ (when $\sigma_s^2 = 1$) from (3.13) with second-order approximation from (3.14). From left to right, $\{\rho_{\min}, \rho_{\max}\} = \{1000, 0\}$, $\{1, -2\}$, and $\{1, 0\}$.

Effect of $\rho_{\min}$ and $\rho_{\max}$

When $\rho_{\min} = 1$ (as a result of $\mathcal{K}_s^{(0)} < 0$, $\mathcal{K}_s^{\max} \le 0$, and $\mathcal{K}_s^{\min} = \mathcal{K}_s^{(0)}$) or when $\rho_{\max} = 1$ (as a result of $\mathcal{K}_s^{(0)} > 0$, $\mathcal{K}_s^{\min} \ge 0$, and $\mathcal{K}_s^{\max} = \mathcal{K}_s^{(0)}$) the expressions in Theorems 3.1–3.3 simplify:

$$\begin{aligned} J_{\mathrm{u},\nu}(\boldsymbol{q}_{\mathrm{sw},\nu}) &\le \frac{1 - \sqrt{2\frac{\mathcal{K}(y_{\mathrm{m}})}{\mathcal{K}_s^{(0)}} - 1}}{1 + \sqrt{2\frac{\mathcal{K}(y_{\mathrm{m}})}{\mathcal{K}_s^{(0)}} - 1}}\,\sigma_s^2 && \text{when} \quad \frac{1}{2} < \frac{\mathcal{K}(y_{\mathrm{m}})}{\mathcal{K}_s^{(0)}} \le 1, \\ &\le \frac{1 - \sqrt{2\left(1 + \frac{J_{\mathrm{u},\nu}(\boldsymbol{q}_{\mathrm{m},\nu})}{\sigma_s^2}\right)^{-2} - 1}}{1 + \sqrt{2\left(1 + \frac{J_{\mathrm{u},\nu}(\boldsymbol{q}_{\mathrm{m},\nu})}{\sigma_s^2}\right)^{-2} - 1}}\,\sigma_s^2 && \text{when} \quad \frac{J_{\mathrm{u},\nu}(\boldsymbol{q}_{\mathrm{m},\nu})}{\sigma_s^2} < \sqrt{2} - 1, \\ &= J_{\mathrm{u},\nu}(\boldsymbol{q}_{\mathrm{m},\nu}) + \frac{1}{2\sigma_s^2} J_{\mathrm{u},\nu}^2(\boldsymbol{q}_{\mathrm{m},\nu}) + \mathcal{O}\big(J_{\mathrm{u},\nu}^3(\boldsymbol{q}_{\mathrm{m},\nu})\big). \end{aligned}$$

Note that in these cases, the SW-UMSE upper bound is independent of the specific distribution of the desired and interfering sources, respectively.

In data communication applications, the case $\rho_{\min} = 1$ is typical as it results from, e.g.,

a) sub-Gaussian desired source in the presence of Gaussian noise, or

b) constant-modulus desired source in the presence of non-super-Gaussian interference.

The case $\rho_{\min} > 1$, on the other hand, might arise from a non-CM (and possibly shaped) desired source constellations in the presence of interference that is "more sub-Gaussian." In fact, source assumption S4) allows for arbitrarily large $\rho_{\min}$, which could result from a nearly-Gaussian desired source in the presence of non-Gaussian interference. Though Theorems 3.1–3.3 remain valid for arbitrarily high $\rho_{\min}$, the requirements placed on Wiener performance (via J_{o}) become more stringent (recall Fig. 3.3).

Generalization of Perfect SW-Estimation Property

Finally, we note that the $J_{\mathrm{u},\nu}(\boldsymbol{q}_{\mathrm{m},\nu})$-based SW-UMSE bound in Theorem 3.2 implies that the perfect SW-estimation property, proven under more restrictive conditions in [Shalvi TIT 90], extends to the general multi-source linear model of Fig. 2.3:

Corollary 3.1. *SW estimators are perfect (up to scaling) when Wiener estimators are perfect.*

Proof. From Theorem 3.2, $J_{\mathrm{u},\nu}(\boldsymbol{q}_{\mathrm{m},\nu}) = 0 \;\Rightarrow\; J_{\mathrm{u},\nu}(\boldsymbol{q}_{\mathrm{sw},\nu}) = 0$. Hence, the estimators are perfect up to a (fixed) scale factor. $\square$

3.3 Numerical Examples

Here we present the results of experiments which compare the UMSE bounds of Theorem 3.1 and Theorem 3.2 with the UMSE characterizing SW estimators found by gradient descent[5] under various source/interference environments. In all experiments, ten non-Gaussian sources are mixed using a matrix $\boldsymbol{\mathcal{H}}$ whose entries are generated randomly from a real-valued zero-mean Gaussian distribution. The estimator $\boldsymbol{f}$ observes the mixture in the presence of AWGN (at SNR of 40dB) and generates estimates of a particular source using $N_f = 8$ adjustable parameters. Note that the number of sensors is less than the number of sources and that noise is present, implying that $\boldsymbol{\mathcal{H}}$ is not full column rank and perfect estimation is not possible.

Figs. 3.4(a)–3.7(a) plot the UMSE upper bounds $J_{\mathrm{u},\nu}\big|_{\mathrm{sw},\nu}^{\max,\mathcal{K}(y_{\mathrm{m}})}$ and $J_{\mathrm{u},\nu}\big|_{\mathrm{sw},\nu}^{\max,J_{\mathrm{u},\nu}(\boldsymbol{q}_{\mathrm{m},\nu})}$ for comparison with $J_{\mathrm{u},\nu}(\boldsymbol{q}_{\mathrm{sw},\nu})$. As a means of "zooming in" on the

[5]Gradient descent results were obtained by the MATLAB routine "`fmincon`," which was initialized randomly in a small ball around the MMSE estimator.

small differences in UMSE, Figs. 3.4(b)–3.7(b) plot the extra-UMSE upper bounds $\mathcal{E}_{\mathrm{u},\nu}\big|_{\mathrm{c},\nu}^{\max,\mathcal{K}(\boldsymbol{q}_{\mathrm{m},\nu})}$ and $\mathcal{E}_{\mathrm{u},\nu}\big|_{\mathrm{c},\nu}^{\max,J_{\mathrm{u},\nu}(\boldsymbol{q}_{\mathrm{m},\nu})}$. In all plots, the $J_{\mathrm{u},\nu}(\boldsymbol{q}_{\mathrm{m},\nu})$-based bounds are denoted by solid lines, the $\mathcal{K}(\boldsymbol{q}_{\mathrm{m},\nu})$-based bounds are denoted by $\bullet$'s, and the gradient-descent values are denoted by $\times$'s.

In Fig. 3.4 ten BPSK sources (i.e., $\mathcal{K}_s^{(k)} = -2$) mix with Gaussian noise. Note, from Fig. 3.4(a), the tightness of the bounds for all but the largest values of $J_{\mathrm{u},\nu}(\boldsymbol{q}_{\mathrm{m},\nu})$.

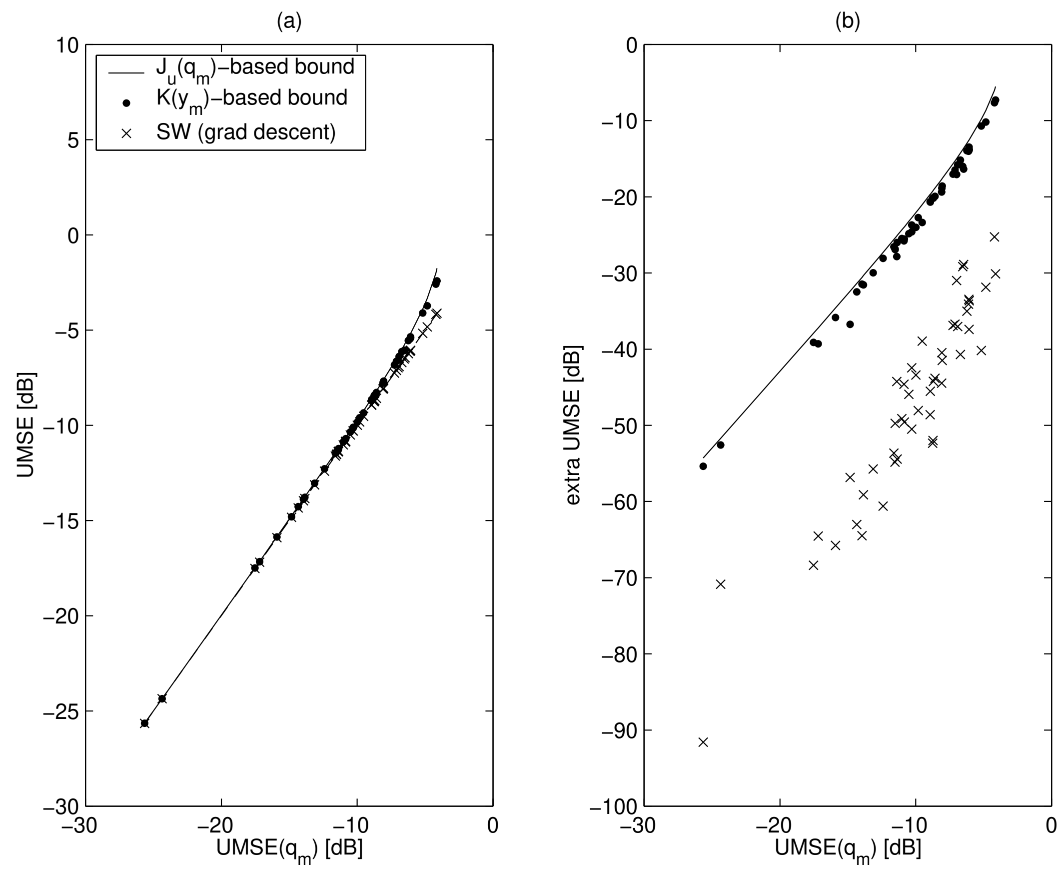

Figure 3.4: Bounds on SW-UMSE for $N_f = 8$, 10 BPSK sources, AWGN at -40dB, and random $\boldsymbol{\mathcal{H}}$.

Fig. 3.5 considers ten super-Gaussian sources, with $\mathcal{K}_s^{(k)} = 2$, in the presence of Gaussian noise. From (3.13) we do not expect SW performance to differ from the BPSK case (where $\mathcal{K}_s^{(k)} = -2$), and this notion is confirmed by comparison of Fig. 3.4 and Fig. 3.5.

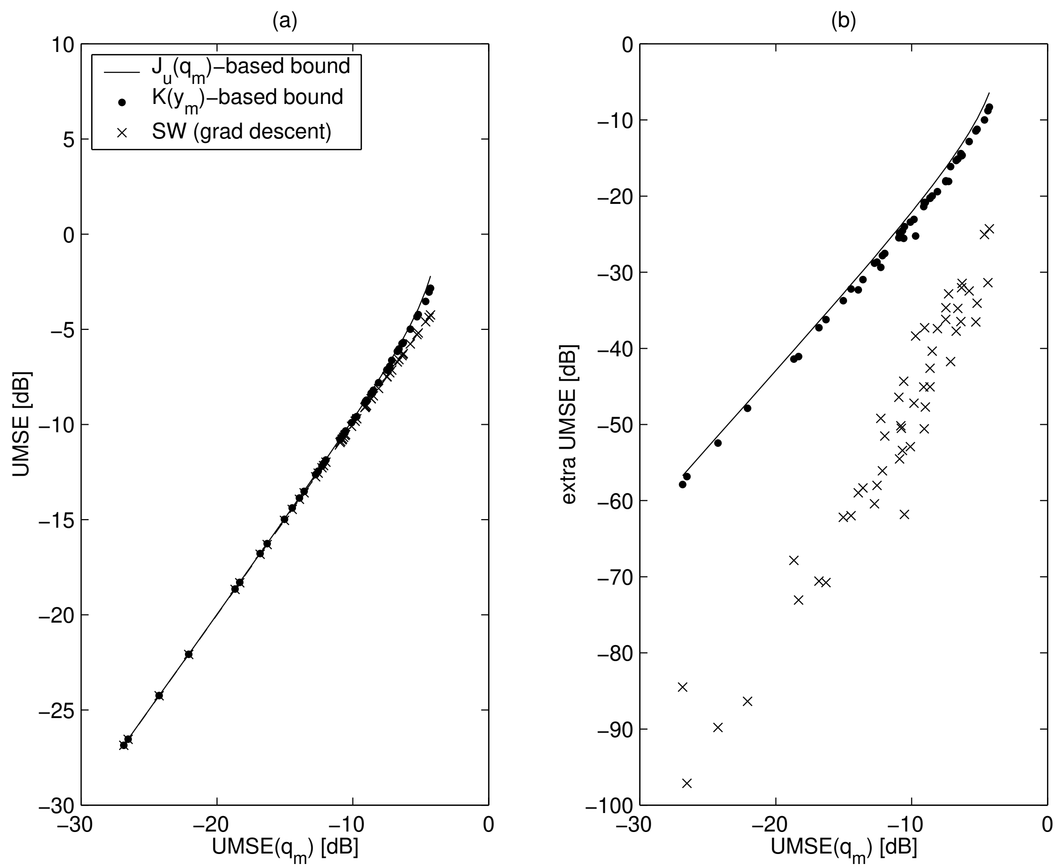

Figure 3.5: Bounds on SW-UMSE for $N_f = 8$, 10 sources with $\mathcal{K}_s^{(k)} = 2$, AWGN at -40dB, and random $\boldsymbol{\mathcal{H}}$.

Fig. 3.6 examines the estimation of a near-Gaussian signal ($\mathcal{K}_s^{(0)} = 0.1$) in the presence of BPSK and AWGN interference. Comparing this experiment to the previous two, notice that here $J_{\mathrm{u},\nu}\big|_{\mathrm{sw},\nu}^{\max,\mathcal{K}(y_{\mathrm{m}})}$ is appreciably tighter than $J_{\mathrm{u},\nu}\big|_{\mathrm{sw},\nu}^{\max,J_{\mathrm{u},\nu}(\boldsymbol{q}_{\mathrm{m},\nu})}$ for larger values of $J_{\mathrm{u},\nu}(\boldsymbol{q}_{\mathrm{m},\nu})$.

Finally, Fig. 3.7 examines the performance of a super-Gaussian signal ($\mathcal{K}_s^{(0)} = 1$) in the presence of impulsive-type noise ($\mathcal{K}_s^{(k)} = 100$). When $\rho_{\max} \gg 1$, (3.10) and (3.12) imply that we can only guarantee the existence of SW estimators in situations

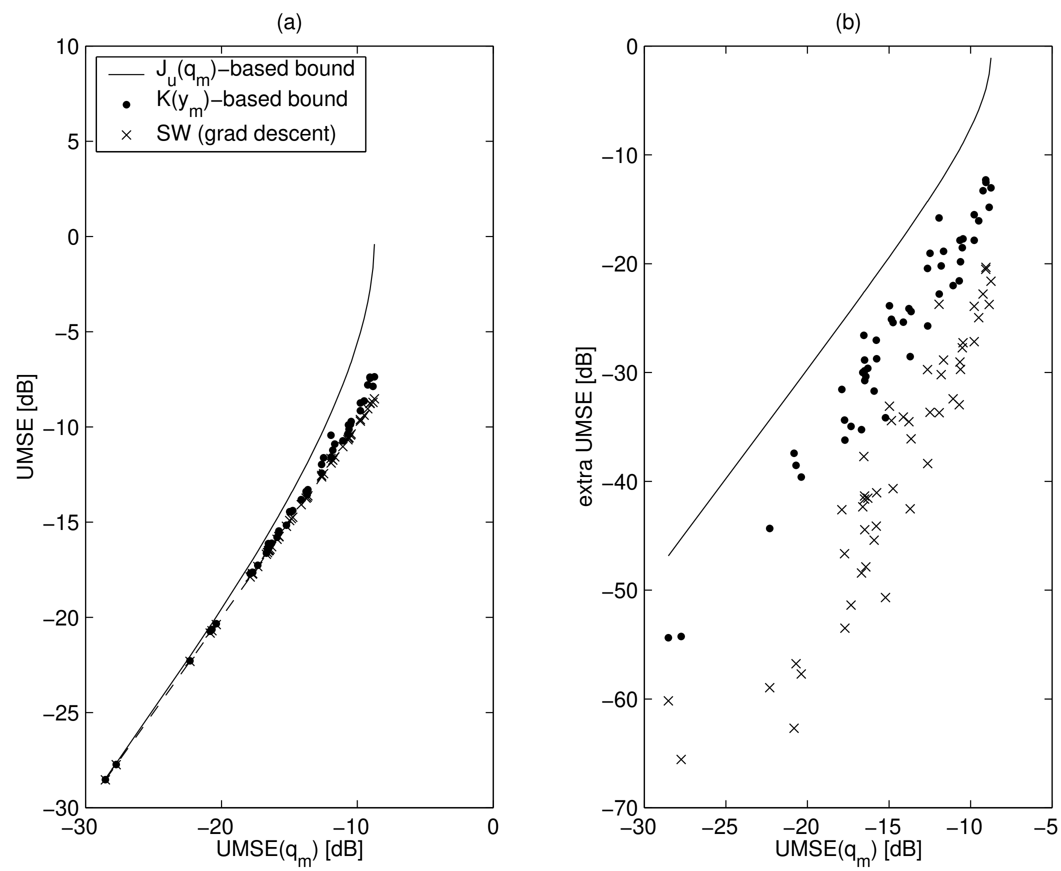

Figure 3.6: Bounds on SW-UMSE for $N_f = 8$, 5 BPSK sources, 5 sources with $\mathcal{K}_s^{(k)} = 0.1$ (one of which is desired), AWGN at -40dB, and random $\boldsymbol{\mathcal{H}}$.

where MMSE estimates are relatively good. As the interference environment in this experiment corresponds to $\rho_{\max} = 100$, UMSE bounds exist only when $J_{u,\nu}(\boldsymbol{q}_{m,\nu}) < -23$ dB.

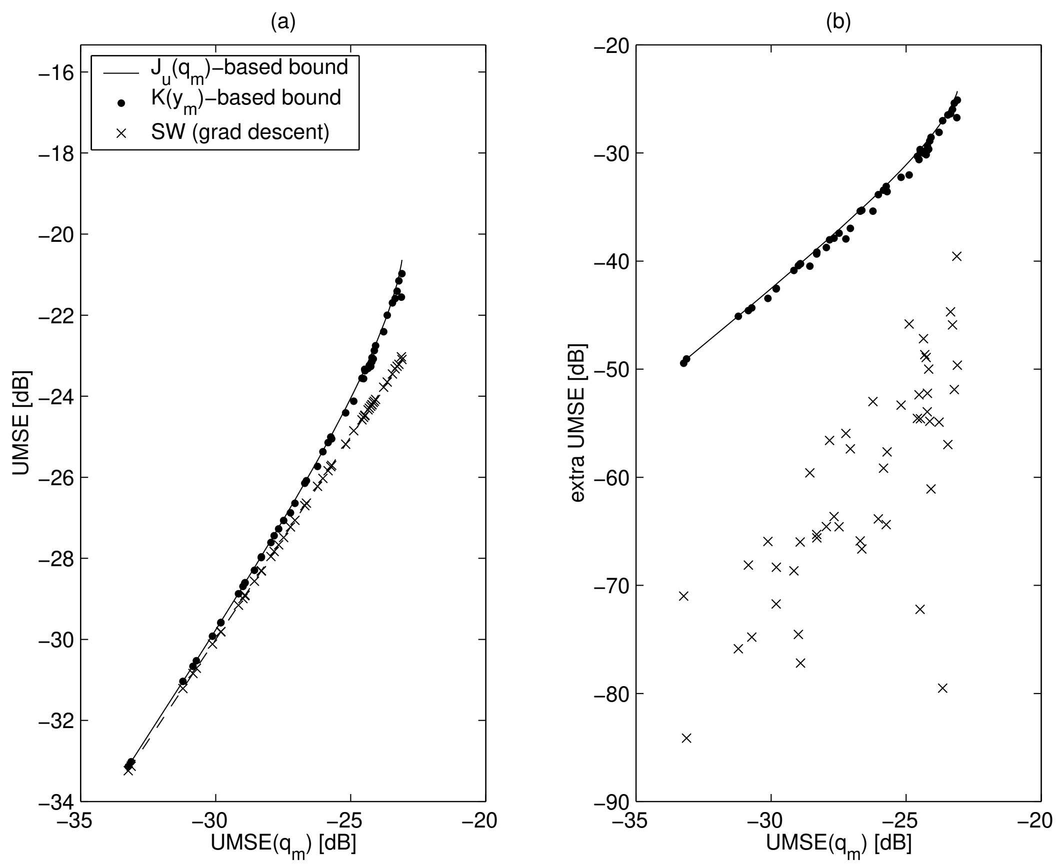

Figure 3.7: Bounds on SW-UMSE for $N_f = 8$, 5 sources with $\mathcal{K}_s^{(k)} = 100$, 5 sources with $\mathcal{K}_s^{(k)} = 1$ (one of which is desired), AWGN at -40dB, and random $\boldsymbol{\mathcal{H}}$.

3.4 Conclusions

In this chapter we have derived conditions under which SW estimators exist and derived bounds for the UMSE of SW estimators. The existence conditions are simple tests which guarantee a SW estimator for the desired user at a particular delay, and these existence arguments have been proven for vector-valued FIR channels and constrained vector-valued FIR estimators. The UMSE bounds hold for vector-valued FIR/IIR channels, constrained FIR/IIR estimators, and nearly arbitrary source and interference distributions. The first bound is a function of the kurtosis of the MMSE estimates, while the second bound is a function of the minimum UMSE of MMSE estimators. Analysis of the second bound shows that the extra UMSE of SW estimators is upper bounded by approximately the square of the minimum UMSE. Thus, SW estimators are very close (in a MSE sense) to optimum linear estimators when the minimum MSE is small. Numerical simulations suggest that the bounds are reasonably tight.

Appendix

3.A Derivation Details for SW-UMSE Bounds

This appendix contains the proofs of the theorems and lemmas found in Section 3.2.

3.A.1 Proof of Theorem 3.1

In this section, we are interested in deriving an expression for the interference radius b_* defined in (3.9) and establishing conditions under which this radius is well defined. Rather than working with (3.9) directly, we find it easier to use the equivalent definition

$$\begin{aligned} b_* &= \min b \text{ s.t. } \Big\{ \|\bar{\boldsymbol{q}}\|_2 > b \Rightarrow \bar{\boldsymbol{q}} \notin \bar{\mathcal{Q}}_{\mathrm{sw}}(\boldsymbol{q}_{\mathrm{r},\nu}) \Big\} \qquad (3.15) \\ &= \min b \text{ s.t. } \begin{cases} \Big\{ \|\bar{\boldsymbol{q}}\|_2 > b \;\Rightarrow\; \left(1 - \|\bar{\boldsymbol{q}}\|_2^2\right)^2 \mathcal{K}_s^{(0)} + \sum_k \|\bar{\boldsymbol{q}}^{(k)}\|_4^4 \,\mathcal{K}_s^{(k)} < \mathcal{K}(y_{\mathrm{r}}) \Big\}, \\ \qquad\qquad \text{when} \quad \mathcal{K}_s^{(0)} > 0, \\ \Big\{ \|\bar{\boldsymbol{q}}\|_2 > b \;\Rightarrow\; \left(1 - \|\bar{\boldsymbol{q}}\|_2^2\right)^2 \mathcal{K}_s^{(0)} + \sum_k \|\bar{\boldsymbol{q}}^{(k)}\|_4^4 \,\mathcal{K}_s^{(k)} > \mathcal{K}(y_{\mathrm{r}}) \Big\}, \\ \qquad\qquad \text{when} \quad \mathcal{K}_s^{(0)} < 0. \end{cases} \end{aligned}$$

We handle the super-Gaussian case (i.e., $\mathcal{K}_s^{(0)} > 0$) first. The following statements are equivalent.

$$\begin{aligned} \mathcal{K}(y_{\mathrm{r}}) &> \left(1 - \|\bar{\boldsymbol{q}}\|_2^2\right)^2 \mathcal{K}_s^{(0)} + \sum_k \|\bar{\boldsymbol{q}}^{(k)}\|_4^4 \,\mathcal{K}_s^{(k)} \\ 0 &> \underbrace{\left(1 - \frac{\mathcal{K}(y_{\mathrm{r}})}{\mathcal{K}_s^{(0)}}\right)}_{C_{\mathrm{r}}} - 2\|\bar{\boldsymbol{q}}\|_2^2 + \|\bar{\boldsymbol{q}}\|_2^4 + \sum_k \|\bar{\boldsymbol{q}}^{(k)}\|_4^4 \frac{\mathcal{K}_s^{(k)}}{\mathcal{K}_s^{(0)}} \qquad (3.16) \end{aligned}$$

Now using the fact that

$$\sum_k \|\bar{\boldsymbol{q}}^{(k)}\|_4^4 \,\mathcal{K}_s^{(k)} \;\leq\; \sum_k \|\bar{\boldsymbol{q}}^{(k)}\|_4^4 \,\mathcal{K}_s^{\max} \;\leq\; \sum_k \|\bar{\boldsymbol{q}}^{(k)}\|_2^4 \,\mathcal{K}_s^{\max} \;=\; \|\bar{\boldsymbol{q}}\|_2^4 \,\mathcal{K}_s^{\max}, \quad (3.17)$$

and the definition of $\rho_{\max}$, the following is a sufficient condition for (3.16):

$$0 \;>\; (1+\rho_{\max})\|\bar{\boldsymbol{q}}\|_2^4 - 2\|\bar{\boldsymbol{q}}\|_2^2 + C_{\mathrm{r}}. \tag{3.18}$$

Since $1+\rho_{\max} > 0$, the set of $\{\|\bar{\boldsymbol{q}}\|_2^2\}$ satisfying (3.18) is equivalent to the set of points $\{x\}$ that lie between the roots $\{x_1, x_2\}$ of the quadratic

$$P_2(x) \;=\; (1+\rho_{\max})x^2 - 2x + C_{\mathrm{r}}.$$

Because $\bar{\boldsymbol{q}}$ is an interference response, we need not consider all values of $\|\bar{\boldsymbol{q}}\|_2$. As explained below, we only need to concern ourselves about $0 \le \|\bar{\boldsymbol{q}}\|_2 < \sqrt{2}^{-1}$. This implies that a valid upper bound on b_*^2 is given by the smaller root of $P_2(x)$ when (i) this smaller root is non-negative real and (ii) the larger root of $P_2(x)$ is ≥ 0.5.

When both roots of $P_2(x)$ lie in the interval $[0, 0.5)$, there exist two valid regions in the interference space with absolute kurtosis larger than at the reference, i.e., $\bar{\mathcal{Q}}_{\mathrm{sw}}(\boldsymbol{q}_{\mathrm{r},\nu})$ becomes disjoint. The "inner" part of $\bar{\mathcal{Q}}_{\mathrm{sw}}(\boldsymbol{q}_{\mathrm{r},\nu})$ allows UMSE bounding since it can be contained by $\{\bar{\boldsymbol{q}} : \|\bar{\boldsymbol{q}}\|_2 \le b_1\}$ for a positive interference radius b_1, but the "outer" part of $\bar{\mathcal{Q}}_{\mathrm{sw}}(\boldsymbol{q}_{\mathrm{r},\nu})$ does *not* permit UMSE bounding in this manner. As an example of this behavior, Fig. 3.8(a) plots the quadratic $P_2(x)$ and Fig. 3.8(b) the region $\bar{\mathcal{Q}}_{\mathrm{sw}}(\boldsymbol{q}_{\mathrm{r},\nu})$ in two-dimensional interference space. As in Fig. 3.2, the attainable set $\bar{\mathcal{Q}}_{\mathrm{a}}$ is denoted by the curved line, the boundary of valid interference region $\bar{\mathcal{Q}}_{\nu}^{(0)}$ by the dotted lines, the SW response by the dot, and the reference response by the diamond. $\bar{\mathcal{Q}}_{\nu}^{(0)}$ is the projection of $\mathcal{Q}_{\nu}^{(0)} \cap \mathcal{Q}_{\mathrm{s}}$ onto $\{\bar{\boldsymbol{q}}\}$.

Disjointness of $\bar{\mathcal{Q}}_{\mathrm{sw}}(\boldsymbol{q}_{\mathrm{r},\nu})$ arises from an interfering source $k \ne 0$ such that $|\,\mathcal{K}_s^{(k)}\,| > |\,\mathcal{K}_s^{(0)}\,|$. In these scenarios, the point of highest kurtosis in the "outer" regions of $\bar{\mathcal{Q}}_{\mathrm{sw}}(\boldsymbol{q}_{\mathrm{r},\nu})$ occurs at points near the boundary of $\bar{\mathcal{Q}}_{\nu}^{(0)}$ of the form $\bar{\boldsymbol{q}} = (\dots, 0, 0, e^{j\theta}\sqrt{2}^{-1}, 0, 0, \dots)^t$. Thus, when $x_2 \ge 0.5$, we can be assured that all valid interference responses (i.e., $\bar{\boldsymbol{q}} \in \bar{\mathcal{Q}}_{\nu}^{(0)}$) with kurtosis greater than the reference can

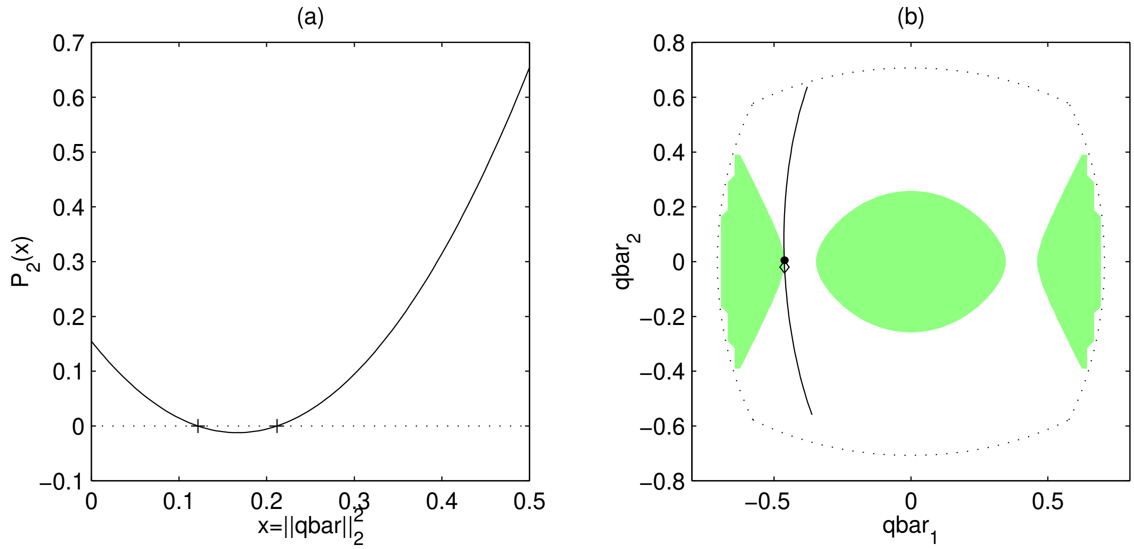

Figure 3.8: Example of $P_2(x)$ and disjoint $\bar{\mathcal{Q}}_{\rm sw}(\boldsymbol{q}_{{\rm r},\nu})$ when larger root of $P_2(x)$ is < 0.5.

be bounded by some radius b_1. In contrast to Fig. 3.8, Fig. 3.9 demonstrates a situation where both root requirements are satisfied.

The roots of $P_2(x)$ are given by

$$\{x_1, x_2\} \;=\; \frac{1 \pm \sqrt{1-(1+\rho_{\max})C_{\rm r}}}{1+\rho_{\max}}, \quad \text{assuming } x_1 \leq x_2,$$

which are both non-negative real when $0 \leq C_{\rm r} \leq (1+\rho_{\max})^{-1}$. It can be shown that $x_2 \geq 0.5$ when $C_{\rm r} \leq (3-\rho_{\max})/4$. Since $\rho_{\max} \geq 1$ implies $(3-\rho_{\max})/4 \leq (1+\rho_{\max})^{-1}$, both root requirements are satisfied when $0 \leq C_{\rm r} \leq (3-\rho_{\max})/4$, or equivalently when

$$\mathcal{K}_s^{(0)} \;\geq\; \mathcal{K}(y_{\rm r}) \;\geq\; \big(\mathcal{K}_s^{(0)} + \mathcal{K}_s^{\max}\big)/4. \tag{3.19}$$

Notice that when (3.19) is satisfied, the facts $\mathcal{K}_s^{(0)} > 0$ and $(\mathcal{K}_s^{(0)} + \mathcal{K}_s^{\max}) > 0$ imply $\mathrm{sgn}(\mathcal{K}_s^{(0)}) = \mathrm{sgn}(\mathcal{K}(y_{\rm r}))$, thus confirming the validity of (3.5).

Under satisfaction of (3.19), equation (3.8) implies the following bound for the

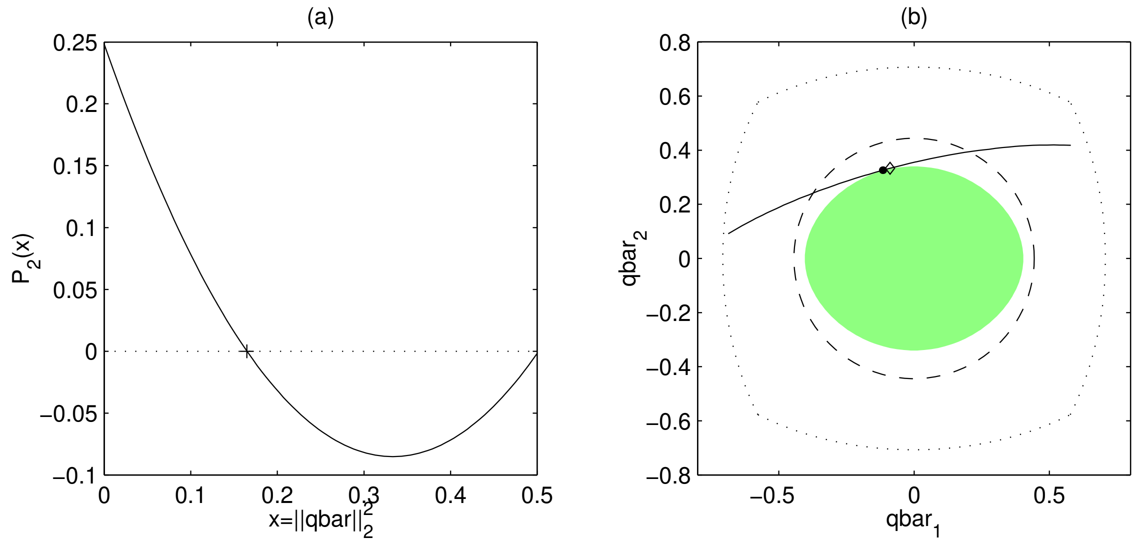

Figure 3.9: Example of $P_2(x)$, $\bar{\mathcal{Q}}_{\rm sw}(\boldsymbol{q}_{{\rm r},\nu})$, and bounding radius when larger root of $P_2(x)$ is > 0.5.

super-Gaussian case:

$$\begin{aligned} J_{{\rm u},\nu}(\boldsymbol{q}_{{\rm sw},\nu}) &\leq \frac{b_*^2}{1-b_*^2} \\ &\leq \frac{x_1^2}{1-x_1^2} \\ &= \frac{1-\sqrt{(1+\rho_{\max})\frac{\mathcal{K}(y_{\rm r})}{\mathcal{K}_s^{(0)}}-\rho_{\max}}}{\rho_{\max}+\sqrt{(1+\rho_{\max})\frac{\mathcal{K}(y_{\rm r})}{\mathcal{K}_s^{(0)}}-\rho_{\max}}} \quad \text{when } \mathcal{K}_s^{(0)} > 0. \qquad (3.20) \end{aligned}$$

The difference between b_* and x_1 accounts for the space between $\bar{\mathcal{Q}}_{\rm sw}(\boldsymbol{q}_{{\rm r},\nu})$ and the bounding radius in Fig. 3.9.

As for the existence of an attainable SW global response associated with the desired user at delay ν, i.e., $\boldsymbol{q}_{{\rm sw},\nu} \in \mathcal{Q}_{\rm a} \cap \mathcal{Q}_{\rm s} \cap \mathcal{Q}_\nu^{(0)}$, we work in the $\bar{\boldsymbol{q}}$ space and establish the existence of a kurtosis local maximum $\bar{\boldsymbol{q}}_{{\rm sw},\nu}$ in the set $\bar{\mathcal{Q}}_{\rm a} \cap \bar{\mathcal{Q}}_\nu^{(0)}$. For simplicity, we assume that the space $\bar{\boldsymbol{q}}$ is finite dimensional. We will exploit the Weierstrass theorem [Luenberger Book 69, p. 40], which says that a continuous cost functional must have a local maximum on a compact set if there exist points in the interior of the set which give cost strictly higher than anywhere on the boundary. The approach is illustrated in Fig. 3.10.

By definition, all points in $\bar{\mathcal{Q}}_{\mathrm{sw}}(\boldsymbol{q}_{\mathrm{r},\nu})$ give kurtosis $\geq \mathcal{K}(y_{\mathrm{r}})$, the kurtosis everywhere on the boundary of $\bar{\mathcal{Q}}_{\mathrm{sw}}(\boldsymbol{q}_{\mathrm{r},\nu})$. To make this inequality strict, we expand $\bar{\mathcal{Q}}_{\mathrm{sw}}(\boldsymbol{q}_{\mathrm{r},\nu})$ to form the new set $\bar{\mathcal{Q}}'_{\mathrm{sw}}(\boldsymbol{q}_{\mathrm{r},\nu})$ defined in terms of boundary kurtosis $\mathcal{K}(y_{\mathrm{r}})-\epsilon$ (for arbitrarily small $\epsilon > 0$). Thus, all points on the boundary of $\bar{\mathcal{Q}}'_{\mathrm{sw}}(\boldsymbol{q}_{\mathrm{r},\nu})$ will give kurtosis strictly less than $\mathcal{K}(y_{\mathrm{r}})$. But how do we know that such a set $\bar{\mathcal{Q}}'_{\mathrm{sw}}(\boldsymbol{q}_{\mathrm{r},\nu})$ exists? We simply need to reformulate (3.16) with ϵ-smaller $\mathcal{K}(y_{\mathrm{r}})$, resulting in ϵ-larger C_r and a modified quadratic $P_2(x)$ in sufficient condition (3.18). As long as the new roots (call them x'_1 and x'_2) satisfy $x'_1 \in [0, 0.5)$ and $x'_2 > 0.5$, the set $\bar{\mathcal{Q}}'_{\mathrm{sw}}(\boldsymbol{q}_{\mathrm{r},\nu})$ is well defined. This property can be guaranteed, for arbitrarily small ϵ, by replacing (3.19) with the stricter condition

$$\mathcal{K}_s^{(0)} \;\geq\; \mathcal{K}(y_{\mathrm{r}}) \;>\; \big(\mathcal{K}_s^{(0)} + \mathcal{K}_s^{\max}\big)/4. \tag{3.21}$$

To summarize, (3.21) guarantees the existence of a closed and bounded set $\bar{\mathcal{Q}}'_{\mathrm{sw}}(\boldsymbol{q}_{\mathrm{r},\nu}) \subset \bar{\mathcal{Q}}_{\nu}^{(0)}$ containing an interior point $\bar{\boldsymbol{q}}_{\mathrm{r},\nu}$ with kurtosis strictly greater than all points on the set boundary.

Due to attainability requirements, our local maximum search must be constrained to the relative interior of the $\bar{\mathcal{Q}}_{\mathrm{a}}$ manifold (which has been embedded in a possibly higher-dimensional $\bar{\boldsymbol{q}}$-space; see Fig. 3.10 for an illustration of a one-dimensional $\bar{\mathcal{Q}}_{\mathrm{a}}$ embedded in $\mathbb{R}^2$). Can we apply the Weierstrass theorem on this manifold? First, we know the $\bar{\mathcal{Q}}_{\mathrm{a}}$ manifold intersects $\bar{\mathcal{Q}}'_{\mathrm{sw}}(\boldsymbol{q}_{\mathrm{r},\nu})$, namely at the point $\bar{\boldsymbol{q}}_{\mathrm{r},\nu}$. Second, we know that the relative boundary of the $\bar{\mathcal{Q}}_{\mathrm{a}}$ manifold occurs outside $\bar{\mathcal{Q}}'_{\mathrm{sw}}(\boldsymbol{q}_{\mathrm{r},\nu})$, namely at infinity. These two observations imply that the boundary of $\bar{\mathcal{Q}}_{\mathrm{a}} \cap \bar{\mathcal{Q}}'_{\mathrm{sw}}(\boldsymbol{q}_{\mathrm{r},\nu})$ relative to $\bar{\mathcal{Q}}_{\mathrm{a}}$ must be a subset of the boundary of $\bar{\mathcal{Q}}'_{\mathrm{sw}}(\boldsymbol{q}_{\mathrm{r},\nu})$. Hence, the interior of $\bar{\mathcal{Q}}_{\mathrm{a}} \cap \bar{\mathcal{Q}}'_{\mathrm{sw}}(\boldsymbol{q}_{\mathrm{r},\nu})$ relative to $\bar{\mathcal{Q}}_{\mathrm{a}}$ contains points which give kurtosis strictly higher than those on the boundary of $\bar{\mathcal{Q}}_{\rm a} \cap \bar{\mathcal{Q}}'_{\rm sw}(\boldsymbol{q}_{{\rm r},\nu})$ relative to $\bar{\mathcal{Q}}_{\rm a}$. (See Fig. 3.10 for an illustration of these relationships.) Finally, the domain (i.e., $\bar{\mathcal{Q}}_{\rm a} \cap \bar{\mathcal{Q}}'_{\rm sw}(\boldsymbol{q}_{{\rm r},\nu})$ relative to $\bar{\mathcal{Q}}_{\rm a}$) is closed and bounded, hence compact. Thus the Weierstrass theorem ensures the existence of a local kurtosis maximum in the interior of $\bar{\mathcal{Q}}_{\rm a} \cap \bar{\mathcal{Q}}'_{\rm sw}(\boldsymbol{q}_{{\rm r},\nu})$ relative to $\bar{\mathcal{Q}}_{\rm a}$ under (3.21). Recalling the one-to-one correspondence between points in $\bar{\mathcal{Q}}_{\rm a} \cap \bar{\mathcal{Q}}'_{\rm sw}(\boldsymbol{q}_{{\rm r},\nu})$ and points in $\mathcal{Q}_{\rm a} \cap \mathcal{Q}_{\rm s} \cap \mathcal{Q}'_{\rm sw}(\boldsymbol{q}_{{\rm r},\nu})$, there exists an attainable SW response for the desired user/delay: $\boldsymbol{q}_{{\rm sw},\nu} \in \mathcal{Q}_{\rm a} \cap \mathcal{Q}_{\rm s} \cap \mathcal{Q}_{\nu}^{(0)}$.

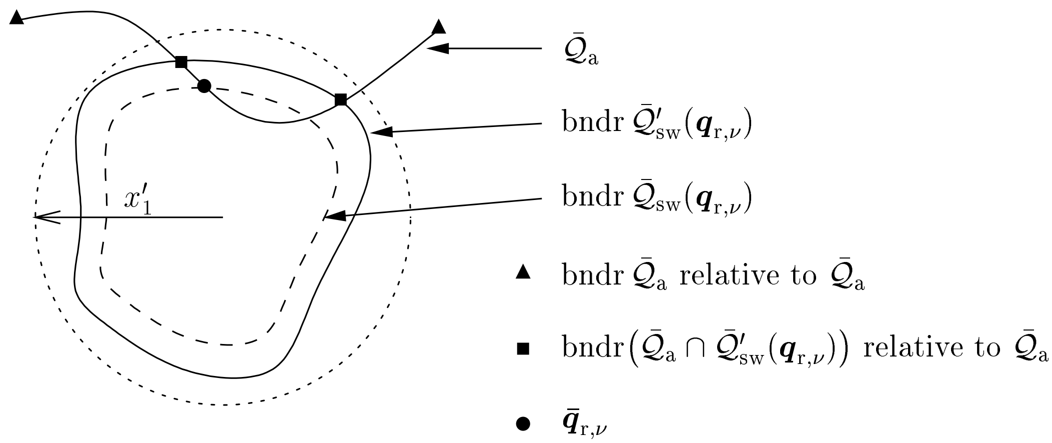

Figure 3.10: Illustration of local minima existence arguments.

Using a similar development for the sub-Gaussian case, we find that when

$$\mathcal{K}_s^{(0)} \;\leq\; \mathcal{K}(y_{\rm r}) \;<\; \big(\mathcal{K}_s^{(0)} + \mathcal{K}_s^{\min}\big)/4, \tag{3.22}$$

the UMSE of SW estimators can be bounded as follows:

$$J_{{\rm u},\nu}(\boldsymbol{q}_{{\rm sw},\nu}) \;\leq\; \frac{1 - \sqrt{(1+\rho_{\min})\frac{\mathcal{K}(y_{\rm r})}{\mathcal{K}_s^{(0)}} - \rho_{\min}}}{\rho_{\min} + \sqrt{(1+\rho_{\min})\frac{\mathcal{K}(y_{\rm r})}{\mathcal{K}_s^{(0)}} - \rho_{\min}}} \quad \text{when } \mathcal{K}_s^{(0)} < 0. \tag{3.23}$$

Furthermore, (3.22) implies that a SW estimator exists within $\mathcal{Q}_{\rm a} \cap \mathcal{Q}_{\rm s} \cap \mathcal{Q}_{\nu}^{(0)}$.

Choosing the scaled Weiner reference $\boldsymbol{q}_{{\rm r},\nu} = \boldsymbol{q}_{{\rm m},\nu}/\|\boldsymbol{q}_{{\rm m},\nu}\|_2$ in equations (3.20)–(3.23) gives Theorem 3.1. Note that we only consider the case $\boldsymbol{q}_{{\rm m},\nu} \in \mathcal{Q}_{\nu}^{(0)}$.

3.A.2 Proof of Theorem 3.2

From Appendix 3.A.1, we know that the expressions in Theorem 3.1 hold for any reference estimates associated with the desired source at delay ν. We consider the case of a super-Gaussian source (i.e., $\mathcal{K}_s^{(0)} > 0$) first. Noting that $J_{\mathrm{u},\nu}\big|_{\mathrm{sw},\nu}^{\max,\mathcal{K}(y_\mathrm{r})}$ in (3.11) is a strictly decreasing function of $\mathcal{K}(y_\mathrm{r})/\,\mathcal{K}_s^{(0)}$ (over its valid range), an upper bound for $J_{\mathrm{u},\nu}\big|_{\mathrm{sw},\nu}^{\max,\mathcal{K}(y_\mathrm{r})}$ follows from a lower bound of $\mathcal{K}(y_\mathrm{r})/\,\mathcal{K}_s^{(0)}$. From (3.6),

$$\begin{aligned}
\frac{\mathcal{K}(y_\mathrm{r})}{\mathcal{K}_s^{(0)}} &= \left(1 - \|\bar{\boldsymbol{q}}_{\mathrm{r},\nu}\|_2^2\right)^2 + \sum_k \|\bar{\boldsymbol{q}}_{\mathrm{r},\nu}^{(k)}\|_4^4 \frac{\mathcal{K}_s^{(k)}}{\mathcal{K}_s^{(0)}} \\
&\geq \left(1 - \|\bar{\boldsymbol{q}}_{\mathrm{r},\nu}\|_2^2\right)^2 + \frac{\mathcal{K}_s^{\min}}{\mathcal{K}_s^{(0)}} \|\bar{\boldsymbol{q}}_{\mathrm{r},\nu}\|_4^4 \\
&\geq \begin{cases} 1 - 2\|\bar{\boldsymbol{q}}_{\mathrm{r},\nu}\|_2^2 + \|\bar{\boldsymbol{q}}_{\mathrm{r},\nu}\|_2^4, & \mathcal{K}_s^{\min} \geq 0, \\ 1 - 2\|\bar{\boldsymbol{q}}_{\mathrm{r},\nu}\|_2^2 + (1+\rho_{\min})\|\bar{\boldsymbol{q}}_{\mathrm{r},\nu}\|_2^4, & \mathcal{K}_s^{\min} < 0. \end{cases}
\end{aligned}$$

When $\mathcal{K}_s^{(0)} > 0$ and $\mathcal{K}_s^{\min} \geq 0$, we see that

$$\begin{aligned}
1 - 2\|\bar{\boldsymbol{q}}_{\mathrm{r},\nu}\|_2^2 + \|\bar{\boldsymbol{q}}_{\mathrm{r},\nu}\|_2^4 &= \left(1 - \|\bar{\boldsymbol{q}}_{\mathrm{r},\nu}\|_2^2\right)^2 \\
&= \left(1 + \frac{\|\bar{\boldsymbol{q}}_{\mathrm{r},\nu}\|_2^2}{1 - \|\bar{\boldsymbol{q}}_{\mathrm{r},\nu}\|_2^2}\right)^{-2} \\
&= \left(1 + \frac{J_{\mathrm{u},\nu}(\boldsymbol{q}_{\mathrm{r},\nu})}{\sigma_s^2}\right)^{-2}
\end{aligned}$$

which implies the bound

$$J_{\mathrm{u},\nu}\big|_{\mathrm{sw},\nu}^{\max,\mathcal{K}(y_\mathrm{r})} \leq \frac{1 - \sqrt{(1+\rho_{\max})\left(1 + \frac{J_{\mathrm{u},\nu}(\boldsymbol{q}_{\mathrm{r},\nu})}{\sigma_s^2}\right)^{-2} - \rho_{\max}}}{\rho_{\max} + \sqrt{(1+\rho_{\max})\left(1 + \frac{J_{\mathrm{u},\nu}(\boldsymbol{q}_{\mathrm{r},\nu})}{\sigma_s^2}\right)^{-2} - \rho_{\max}}} \tag{3.24}$$

as long as

$$1 \geq \frac{\mathcal{K}(y_\mathrm{r})}{\mathcal{K}_s^{(0)}} > \frac{1+\rho_{\max}}{4}. \tag{3.25}$$

Having just shown that $\mathcal{K}(y_\mathrm{r})/\,\mathcal{K}_s^{(0)} \geq (1 + J_{\mathrm{u},\nu}(\boldsymbol{q}_{\mathrm{r},\nu})/\sigma_s^2)^{-2}$, a sufficient condition for the right inequality of (3.25) is $(1 + J_{\mathrm{u},\nu}(\boldsymbol{q}_{\mathrm{r},\nu})/\sigma_s^2)^{-2} > (1+\rho_{\max})/4$ which can be restated as

$$\frac{J_{\mathrm{u},\nu}(\boldsymbol{q}_{\mathrm{r},\nu})}{\sigma_s^2} \;<\; -1+2\sqrt{(1+\rho_{\max})^{-1}}. \tag{3.26}$$

For the left inequality in (3.25), we use (3.6) and (3.17) to bound

$$\begin{aligned}
\frac{\mathcal{K}(y_{\mathrm{r}})}{\mathcal{K}_s^{(0)}} &\leq 1-2\|\bar{\boldsymbol{q}}_{\mathrm{r},\nu}\|_2^2+(1+\rho_{\max})\|\bar{\boldsymbol{q}}_{\mathrm{r},\nu}\|_2^4 \\
&= \big(1-\|\bar{\boldsymbol{q}}_{\mathrm{r},\nu}\|_2^2\big)^2\left(1+\rho_{\max}\frac{\|\boldsymbol{q}_{\mathrm{r},\nu}\|_2^4}{\big(1-\|\bar{\boldsymbol{q}}_{\mathrm{r},\nu}\|_2^2\big)^2}\right) \\
&= \left(1+\frac{J_{\mathrm{u},\nu}(\boldsymbol{q}_{\mathrm{r},\nu})}{\sigma_s^2}\right)^{-2}\left(1+\rho_{\max}\frac{J^2_{\mathrm{u},\nu}(\boldsymbol{q}_{\mathrm{r},\nu})}{\sigma_s^4}\right).
\end{aligned}$$

Thus $1 \geq (1+J_{\mathrm{u},\nu}(\boldsymbol{q}_{\mathrm{r},\nu})/\sigma_s^2)^{-2}(1+\rho_{\max}J^2_{\mathrm{u},\nu}(\boldsymbol{q}_{\mathrm{r},\nu})/\sigma_s^4)$ is sufficient for the left side of (3.25), which can be restated simply as $J_{\mathrm{u},\nu}(\boldsymbol{q}_{\mathrm{r},\nu})/\sigma_s^2 \leq 2(\rho_{\max}-1)^{-1}$. But, using the fact that $\rho_{\max} \geq 1$, it can be shown that $-1+2\sqrt{(1+\rho_{\max})^{-1}} \leq 2(\rho_{\max}-1)^{-1}$, and thus (3.26) remains as a sufficient condition for bound (3.24).

For the case when $\mathcal{K}_s^{(0)} > 0$ and $\mathcal{K}_s^{\min} < 0$, we have just shown that

$$\begin{aligned}
\frac{\mathcal{K}(y_{\mathrm{r}})}{\mathcal{K}_s^{(0)}} &\geq 1-2\|\bar{\boldsymbol{q}}_{\mathrm{r},\nu}\|_2^2+(1+\rho_{\min})\|\bar{\boldsymbol{q}}_{\mathrm{r},\nu}\|_2^4 \\
&= \left(1+\frac{J_{\mathrm{u},\nu}(\boldsymbol{q}_{\mathrm{r},\nu})}{\sigma_s^2}\right)^{-2}\left(1+\rho_{\min}\frac{J^2_{\mathrm{u},\nu}(\boldsymbol{q}_{\mathrm{r},\nu})}{\sigma_s^4}\right),
\end{aligned}$$

which implies the bound

$$J_{\mathrm{u},\nu}\Big|_{\mathrm{sw},\nu}^{\max,\mathcal{K}(y_{\mathrm{r}})} \leq \frac{1-\sqrt{(1+\rho_{\max})\left(1+\frac{J_{\mathrm{u},\nu}(\boldsymbol{q}_{\mathrm{r},\nu})}{\sigma_s^2}\right)^{-2}\left(1+\rho_{\min}\frac{J^2_{\mathrm{u},\nu}(\boldsymbol{q}_{\mathrm{r},\nu})}{\sigma_s^4}\right)-\rho_{\max}}}{\rho_{\max}+\sqrt{(1+\rho_{\max})\left(1+\frac{J_{\mathrm{u},\nu}(\boldsymbol{q}_{\mathrm{r},\nu})}{\sigma_s^2}\right)^{-2}\left(1+\rho_{\min}\frac{J^2_{\mathrm{u},\nu}(\boldsymbol{q}_{\mathrm{r},\nu})}{\sigma_s^4}\right)-\rho_{\max}}} \tag{3.27}$$

as long as (3.25) holds. A sufficient condition for the right inequality in (3.25) would then be $1-2\|\bar{\boldsymbol{q}}_{\mathrm{r},\nu}\|_2^2+(1+\rho_{\min})\|\bar{\boldsymbol{q}}_{\mathrm{r},\nu}\|_2^4 > (1+\rho_{\max})/4$, or equivalently

$$(1+\rho_{\min})\|\bar{\boldsymbol{q}}_{\mathrm{r},\nu}\|_2^4-2\|\bar{\boldsymbol{q}}_{\mathrm{r},\nu}\|_2^2+(3-\rho_{\max})/4 \;>\; 0.$$

It can be shown that the quadratic inequality above is satisfied by

$$\|\bar{\boldsymbol{q}}_{\mathrm{r},\nu}\|_2^2 \;<\; \begin{cases} \frac{1-\sqrt{1-(3-\rho_{\max})(1+\rho_{\min})/4}}{1+\rho_{\min}}, & \rho_{\min} \neq -1, \\ (3-\rho_{\max})/8 & \rho_{\min} = -1, \end{cases}$$

and since $J_{\mathrm{u},\nu}(\boldsymbol{q}) = \|\bar{\boldsymbol{q}}\|_2^2/(1-\|\bar{\boldsymbol{q}}\|_2^2)$ is strictly increasing in $\|\bar{\boldsymbol{q}}\|_2^2$, the following must be sufficient for the right inequality of (3.25).

$$\frac{J_{\mathrm{u},\nu}(\boldsymbol{q}_{\mathrm{r},\nu})}{\sigma_s^2} \;<\; \begin{cases} \frac{1-\sqrt{1-(3-\rho_{\max})(1+\rho_{\min})/4}}{\rho_{\min}+\sqrt{1-(3-\rho_{\max})(1+\rho_{\min})/4}}, & \rho_{\min} \neq -1, \\ \frac{3-\rho_{\max}}{5+\rho_{\max}} & \rho_{\min} = -1. \end{cases} \tag{3.28}$$

For the left inequality in (3.25) we have the same sufficient condition as when $\mathcal{K}_s^{(0)} > 0$ and $\mathcal{K}_s^{\min} \geq 0$, namely, $J_{\mathrm{u},\nu}(\boldsymbol{q}_{\mathrm{r},\nu})/\sigma_s^2 \leq 2(\rho_{\max}-1)^{-1}$. Again, this latter condition is less stringent than (3.28), implying that (3.28) is sufficient for bound (3.27).

The case of a sub-Gaussian source can be treated in a similar manner, and the results are the same as (3.24), (3.26), (3.27), and (3.28), modulo a swapping of $\rho_{\min}$ and $\rho_{\max}$. The final results are collected in Theorem 3.2 under the choice $\boldsymbol{q}_{\mathrm{r},\nu} = \boldsymbol{q}_{\mathrm{m},\nu}/\|\boldsymbol{q}_{\mathrm{m},\nu}\|_2$. The UMSE conditions above guarantee that $\boldsymbol{q}_{\mathrm{m},\nu} \in \mathcal{Q}_\nu^{(0)}$.

3.A.3 Proof of Theorem 3.3

Here, we reformulate the upper bound (3.13). To simplify the presentation of the proof, the shorthand notation $J := J_{\mathrm{u},\nu}(\boldsymbol{q}_{\mathrm{m},\nu})/\sigma_s^2$ will be used.

Starting with the case that $\mathcal{K}_s^{(0)} > 0$ and $\mathcal{K}_s^{\min} \geq 0$, (3.13) says

$$\frac{J_{\mathrm{u},\nu}\big|_{\mathrm{sw},\nu}^{\max, J_{\mathrm{u},\nu}(\boldsymbol{q}_{\mathrm{m},\nu})}}{\sigma_s^2} = \frac{1-\sqrt{(1+\rho_{\min})(1+J)^{-2}-\rho_{\min}}}{\rho_{\min}+\sqrt{(1+\rho_{\min})(1+J)^{-2}-\rho_{\min}}},$$

from which routine manipulations yield

$$\frac{J_{\mathrm{u},\nu}\big|_{\mathrm{sw},\nu}^{\max, J_{\mathrm{u},\nu}(\boldsymbol{q}_{\mathrm{m},\nu})}}{\sigma_s^2} = \frac{1-\sqrt{1-\big((\rho_{\min}-1)+\rho_{\min}(2J+J^2)\big)(2J+J^2)}}{(\rho_{\min}-1)+\rho_{\min}(2J+J^2)}.$$

For $x \in \mathbb{R}$ such that $|x| < 1$, the binomial series [Rudin Book 76] may be used to claim

$$\sqrt{1-x} \;=\; 1 - \frac{x}{2} - \frac{x^2}{8} - \mathcal{O}(x^3).$$

Applying the previous expression with $x = \big((\rho_{\min} - 1) + \rho_{\min}(2J + J^2)\big)(2J + J^2)$, we find that

$$\begin{aligned}
\frac{J_{\mathrm{u},\nu}\big|_{\mathrm{sw},\nu}^{\max,J_{\mathrm{u},\nu}(\boldsymbol{q}_{\mathrm{m},\nu})}}{\sigma_s^2} &= \frac{1}{2}\,\frac{\big((\rho_{\min}-1)+\rho_{\min}(2J+J^2)\big)(2J+J^2)}{(\rho_{\min}-1)+\rho_{\min}(2J+J^2)} \\
&\quad + \frac{1}{8}\,\frac{\big((\rho_{\min}-1)+\rho_{\min}(2J+J^2)\big)^2(2J+J^2)^2}{(\rho_{\min}-1)+\rho_{\min}(2J+J^2)} + \mathcal{O}(J^3) \\
&= J + \frac{\rho_{\min}}{2}J^2 + \mathcal{O}(J^3).
\end{aligned}$$

Finally, subtraction of J gives the first case in (3.14).

For the case of $\mathcal{K}_s^{(0)} > 0$ and $\mathcal{K}_s^{\min} < 0$, (3.13) says

$$\frac{J_{\mathrm{u},\nu}\big|_{\mathrm{sw},\nu}^{\max,J_{\mathrm{u},\nu}(\boldsymbol{q}_{\mathrm{m},\nu})}}{\sigma_s^2} \;=\; \frac{1-\sqrt{(1+\rho_{\min})(1+J)^{-2}(1+\rho_{\max}J^2)-\rho_{\min}}}{\rho_{\min}+\sqrt{(1+\rho_{\min})(1+J)^{-2}(1+\rho_{\max}J^2)-\rho_{\min}}},$$

from which routine manipulations yield

$$\begin{aligned}
&\frac{J_{\mathrm{u},\nu}\big|_{\mathrm{sw},\nu}^{\max,J_{\mathrm{u},\nu}(\boldsymbol{q}_{\mathrm{m},\nu})}}{\sigma_s^2} \;= \\
&\quad \frac{\rho_{\max}J^2 + 1 - \sqrt{1+(2-2\rho_{\max})J+(1+\rho_{\max}+\rho_{\min}\rho_{\max}-5\rho_{\min})J^2+\mathcal{O}(J^3)}}{(\rho_{\min}-1)+2\rho_{\min}J+(\rho_{\min}-\rho_{\max})J^2}.
\end{aligned}$$

As before, we use the binomial series expansion for $\sqrt{1-x}$, but now with $x = (2\rho_{\max}-2)J + (5\rho_{\min}-1-\rho_{\max}-\rho_{\min}\rho_{\max})J^2 + \mathcal{O}(J^3)$. After some algebra, we find

$$\frac{J_{\mathrm{u},\nu}\big|_{\mathrm{sw},\nu}^{\max,J_{\mathrm{u},\nu}(\boldsymbol{q}_{\mathrm{m},\nu})}}{\sigma_s^2} \;=\; J + \frac{1}{2}\,\frac{(\rho_{\min}-\rho_{\max})(\rho_{\min}-1)J^2+\mathcal{O}(J^3)}{(\rho_{\min}-1)+2\rho_{\min}J+(\rho_{\min}-\rho_{\max})J^2}$$

Finally we apply the series approximation

$$\frac{1}{1-y} \;=\; 1 + y + \mathcal{O}(y^2)$$

with $y = -(2\rho_{\mathrm{min}} J + (\rho_{\mathrm{min}} - \rho_{\mathrm{max}})J^2)/(\rho_{\mathrm{min}} - 1)$ for $\rho_{\mathrm{min}} \neq 1$. Straightforward algebra yields

$$\frac{J_{\mathrm{u},\nu}\Big|_{\mathrm{sw},\nu}^{\mathrm{max},J_{\mathrm{u},\nu}(\boldsymbol{q}_{\mathrm{m},\nu})}}{\sigma_s^2} = J + \frac{1}{2}(\rho_{\mathrm{min}} - \rho_{\mathrm{max}})J^2 + \mathcal{O}(J^3).$$

Taking the limit $\rho_{\mathrm{min}} \to 1$, it is evident that no problems arise at the point $\rho_{\mathrm{min}} = 1$. Subtraction of J from the last statement gives the second case in (3.14).

Chapter 4

Bounds for the MSE performance of CM Estimators[1]

4.1 Introduction

Under "ideal" conditions, CM-minimizing estimators are perfect modulo unavoidable gain and delay ambiguities. In other words, CM-minimizing $\{y_n\} = \{\alpha s_{n-\nu}^{(0)}\}$ for some α and ν. More specifically, for a single source satisfying S1), S4) and S5), the perfect CM-estimation property has been proven for

- unconstrained doubly-infinite estimators with BIBO channels: [Foschini ATT 85], and
- causal FIR estimators with full-column rank (FCR) $\boldsymbol{\mathcal{H}}$: [Tong CISS 92], [Li TSP 96a].

(As we show in Section 4.2.3, the perfect CM-estimation property can also be extended to the multi-source linear model described in Section 2.2.) But how good

[1]The main results of this chapter also appear in the manuscript [Schniter TIT 00].

are CM-minimizing estimators under non-ideal conditions?

The last decade has seen a plethora of papers giving evidence for the "robustness" of CM performance in situations where the CM-minimizing (and MMSE) estimators are not perfect. Most of these studies, however, focus on *particular* features of the system model that prevent perfect estimation, such as

1. the presence of additive white Gaussian noise (AWGN) corrupting the observation (e.g., [Fijalkow TSP 97], [Zeng TIT 98], [Liu SP 99]),

2. channels that do not provide adequate diversity (e.g., [Fijalkow TSP 97], [Li TSP 96b]), or

3. estimators with an insufficient number of adjustable parameters (e.g., [Endres TSP 99], [Regalia TSP 99]).

A notable exception is the work of Zeng et al. [Zeng TSP 99], in which an algorithm is given to bound the MSE of CM-minimizing estimators for the case of a single source transmitted through a finite-duration impulse response (FIR) linear channel in the presence of AWGN. The channel model assumed by [Zeng TSP 99] is general enough to incorporate most combinations of the three conditions above, though not as general as the multi-source model of Fig. 2.3. The bounding algorithm in [Zeng TSP 99] is rather involved, however, preventing a direct link between the MSE performance of CM and Wiener estimators. (See Table 4.1.)

The main contribution of this chapter is a (closed-form) bound on the MSE performance of CM-minimizing estimators that is a simple function of the MSE performance of Weiner estimators. This bound, derived under the multi-source linear model in Section 2.2, provides the most formal link (established to date) between the CM and Wiener estimators, and as such, the most general testament to the MSE-robustness of the CM criterion.

The organization of the chapter is as follows. Section 4.2 derives the bound for the MSE performance of the CM criterion, Section 4.3 presents the results of numerical simulations demonstrating the efficacy of our bounding technique, and Section 4.4 concludes the chapter.

4.2 CM Performance under General Additive Interference

An algorithm for bounding the MSE performance of CM minimizers has been derived by Zeng et al. for the case of a real-valued i.i.d. source, a FIR channel, AWGN, and a finite-length estimator. The development for FCR and non-FCR $\boldsymbol{\mathcal{H}}$ appear in [Zeng TIT 98] and [Zeng TSP 99], respectively. Using the notation established in Section 2.2, the algorithm of [Zeng TSP 99] is summarized in Table 4.1. Though the relatively complicated Zeng algorithm generates reasonably tight CM-UMSE upper bounds (as we shall see in Section 4.3), we have found that it is possible to derive tight bounds for the UMSE of CM-minimizing symbol estimators that

- have a closed-form expression,
- support arbitrary additive interference,
- support complex-valued channels and estimators, and
- support IIR (as well as FIR) channels and estimators.

We will now derive such bounds. Section 4.2.1 outlines our approach, Section 4.2.2 presents the main results, and Section 4.2.3 comments on these results. Proof details appear in Appendix 4.A.

Table 4.1: Zeng et al.s' CM-UMSE bounding algorithm [Zeng TSP 99].

<u>Assumptions:</u>

The desired ($k=0$) channel is FIR with coefficients $\{\mathbf{h}_i^{(0)}\} \in \mathbb{R}^P$.
AWGN of variance σ_w^2 is present at each of P sensors, so that

$$\mathbf{r}_n \;=\; \sum_{i=0}^{N_h-1}\left(\mathbf{h}_i^{(0)} s_{n-i}^{(0)} + \tfrac{\sigma_w}{\sigma_s}\sum_{k=1}^{P} \mathbf{e}_k s_{n-i}^{(k)}\right).$$

The sources are real-valued and satisfy (S1)-(S5).
The dispersion constant is $\gamma = \mathrm{E}\{|s_n^{(0)}|^4\}/\sigma_s^2$.
The estimator $\boldsymbol{f} = (\mathbf{f}_0^t, \ldots, \mathbf{f}_{N_f-1}^t)^t$ has N_f coefficients of size $P \times 1$.

<u>Definitions:</u>

$$\mathbf{H} := \begin{pmatrix} \mathbf{h}_0^{(0)} & \mathbf{h}_1^{(0)} & \cdots & \mathbf{h}_{N_h-1}^{(0)} & & & \\ & \ddots & & & \ddots & & \\ & & & \mathbf{h}_0^{(0)} & \mathbf{h}_1^{(0)} & \cdots & \mathbf{h}_{N_h-1}^{(0)} \end{pmatrix} \in \mathbb{R}^{PN_f \times (N_f+N_h-1)}.$$

$$\mathbf{R} := \mathbf{H}\mathbf{H}^t + (\tfrac{\sigma_w}{\sigma_s})^2\mathbf{I}.$$

$$\boldsymbol{\Phi} := \mathbf{I} + (\tfrac{\sigma_w}{\sigma_s})^2\,(\mathbf{H}^t\mathbf{H})^{\dagger} = \begin{pmatrix} \mathbf{C}_{11} & \mathbf{b}_1 & \mathbf{C}_{12} \\ \mathbf{b}_1^t & a & \mathbf{b}_2^t \\ \mathbf{C}_{12} & \mathbf{b}_2 & \mathbf{C}_{22} \end{pmatrix},$$

set $a := [\boldsymbol{\Phi}]_{\nu,\nu}$, $\mathbf{b} := \begin{pmatrix}\mathbf{b}_1 \\ \mathbf{b}_2\end{pmatrix}$ and $\mathbf{C} := \begin{pmatrix}\mathbf{C}_{11} & \mathbf{C}_{12} \\ \mathbf{C}_{12} & \mathbf{C}_{22}\end{pmatrix}$.

<u>Calculations:</u>

$$\boldsymbol{q}_{\mathrm{m},\nu}^{(0)} = \mathbf{H}^t\mathbf{R}^{-1}\mathbf{H}\mathbf{e}_\nu.$$

$\boldsymbol{q}_{\mathrm{mI}}^{(0)} = \boldsymbol{q}_{\mathrm{m},\nu}^{(0)}(0:\nu{-}1,\ \nu{+}1:N_q{-}1)/\boldsymbol{q}_{\mathrm{m},\nu}^{(0)}(\nu)$, using MATLAB notation.

$$\alpha_{\mathrm{r},\nu} = \sqrt{\frac{\gamma\|\boldsymbol{q}_{\mathrm{m},\nu}^{(0)}\|_{\boldsymbol{\Phi}}^2}{3\|\boldsymbol{q}_{\mathrm{m},\nu}^{(0)}\|_{\boldsymbol{\Phi}}^4 - (3-\gamma)\|\boldsymbol{q}_{\mathrm{m},\nu}^{(0)}\|_4^4}}.$$

$$\boldsymbol{q}_{\mathrm{r},\nu}^{(0)} = \alpha_{\mathrm{r},\nu}\boldsymbol{q}_{\mathrm{m},\nu}^{(0)}.$$

$$J_{\mathrm{c},\nu}(\boldsymbol{q}_{\mathrm{r},\nu}^{(0)}) = 3\|\boldsymbol{q}_{\mathrm{r},\nu}^{(0)}\|_{\boldsymbol{\Phi}}^4 - 2\gamma\|\boldsymbol{q}_{\mathrm{r},\nu}^{(0)}\|_{\boldsymbol{\Phi}}^2 - (3-\gamma)\|\boldsymbol{q}_{\mathrm{r},\nu}^{(0)}\|_4^4 + \gamma^2.$$

$$\boldsymbol{q}_{\mathrm{oI}}^{(0)} = -\mathbf{C}^{-1}\mathbf{b}.$$

$$\theta_{\mathrm{o}} = (a - \mathbf{b}^t\mathbf{C}^{-1}\mathbf{b})^{-1}.$$

$$\delta_{\mathrm{o}} = \|\boldsymbol{q}_{\mathrm{mI}}^{(0)} - \boldsymbol{q}_{\mathrm{oI}}^{(0)}\|_{\mathbf{C}}.$$

<u>UMSE Bound:</u>

For the quartic polynomial $D(\delta) = c_1^2(\delta) - 4c_2(\delta)c_0$, where

$$c_0 = \gamma^2 - J_{\mathrm{c},\nu}(\boldsymbol{q}_{\mathrm{r},\nu}^{(0)}),$$
$$c_1(\delta) = -2\gamma(\delta^2 + \theta_{\mathrm{o}}^{-1}),$$
$$c_2(\delta) = 3(\delta^2 + \theta_{\mathrm{o}}^{-1})^2 - (3-\gamma)\big(1 + (\delta + \|\boldsymbol{q}_{\mathrm{oI}}^{(0)}\|_4)^4\big),$$

find $\{\delta_1 < \cdots < \delta_m\}$ = real-valued roots of $D(\delta)$, and
set $\delta_\star = \min\{\delta_i \mid \delta_i > \delta_{\mathrm{o}}\}$.
If $\delta_\star \neq \emptyset$, $D(\delta_{\mathrm{o}}) \geq 0$, and $c_2(\delta) > 0$ for all $\delta \in [\delta_{\mathrm{o}}, \delta_\star]$, then

$$\mathrm{UMSE}(\boldsymbol{q}_{\mathrm{c},\nu}^{(0)}) \leq \delta_\star^2 + \theta_o^{-1} - 1,$$

else unable to compute bound.

4.2.1 The CM-UMSE Bounding Strategy

Say that $\boldsymbol{q}_{\mathrm{r},\nu}$ is an attainable global reference response for the desired user ($k = 0$) at some fixed delay ν. Formally, $\boldsymbol{q}_{\mathrm{r},\nu} \in \mathcal{Q}_{\mathrm{a}} \cap \mathcal{Q}_\nu^{(0)}$, where

$$\mathcal{Q}_\nu^{(0)} := \left\{ \boldsymbol{q} \text{ s.t. } |q_\nu^{(0)}| > \max_{(k,\delta)\neq(0,\nu)} |q_\delta^{(k)}| \right\}.$$

$\mathcal{Q}_\nu^{(0)}$ defines the set of global responses *associated*[2] with user 0 at delay ν. The set[3] of (attainable) locally CM-minimizing global responses for the desired user at delay ν will be denoted by $\{\boldsymbol{q}_{\mathrm{c},\nu}\}$ and defined as:

$$\{\boldsymbol{q}_{\mathrm{c},\nu}\} := \left\{ \arg\min_{\boldsymbol{q}\in\mathcal{Q}_{\mathrm{a}}} J_{\mathrm{c}}(\boldsymbol{q}) \right\} \cap \mathcal{Q}_\nu^{(0)}.$$

In general, it is not possible to determine closed-form expressions for $\{\boldsymbol{q}_{\mathrm{c},\nu}\}$, making it difficult to evaluate the UMSE of CM-minimizing estimators.

When $\boldsymbol{q}_{\mathrm{r},\nu}$ is in the vicinity of a $\boldsymbol{q}_{\mathrm{c},\nu}$ (the meaning of which will be made more precise later) then, by definition, this $\boldsymbol{q}_{\mathrm{c},\nu}$ must have CM cost less than or equal to the cost at $\boldsymbol{q}_{\mathrm{r},\nu}$. In this case, $\exists \boldsymbol{q}_{\mathrm{c},\nu} \in \mathcal{Q}_{\mathrm{c}}(\boldsymbol{q}_{\mathrm{r},\nu})$, where

$$\mathcal{Q}_{\mathrm{c}}(\boldsymbol{q}_{\mathrm{r},\nu}) := \left\{ \boldsymbol{q} \;\text{ s.t. }\; J_{\mathrm{c}}(\boldsymbol{q}) \le J_{\mathrm{c}}(\boldsymbol{q}_{\mathrm{r},\nu}) \right\} \cap \mathcal{Q}_\nu^{(0)}. \tag{4.1}$$

This approach implies the following CM-UMSE upper bound:

$$J_{\mathrm{u},\nu}(\boldsymbol{q}_{\mathrm{c},\nu}) \le \max_{\boldsymbol{q}\in\mathcal{Q}_{\mathrm{c}}(\boldsymbol{q}_{\mathrm{r},\nu})} J_{\mathrm{u},\nu}(\boldsymbol{q}). \tag{4.2}$$

Note that the maximization on the right of (4.2) does not explicitly involve the admissibility constraint $\mathcal{Q}_{\mathrm{a}}$; the constraint is implicitly incorporated through $\boldsymbol{q}_{\mathrm{r},\nu}$.

The tightness of the upper bound (4.2) will depend on the size and shape of $\mathcal{Q}_{\mathrm{c}}(\boldsymbol{q}_{\mathrm{r},\nu})$, motivating careful selection of the reference $\boldsymbol{q}_{\mathrm{r},\nu}$. Notice that the size of

[2] Note that under S1)–S3), a particular {user, delay} combination is "associated" with an estimate if and only if that {user, delay} contributes more energy to the estimate than any other {user, delay}.

[3] We refer to the CM-minimizing responses as a set to avoid establishing existence or uniqueness of local minima within $\mathcal{Q}_{\mathrm{a}} \cap \mathcal{Q}_\nu^{(0)}$ at this time.

$\mathcal{Q}_c(\boldsymbol{q}_{r,\nu})$ can usually be reduced via replacement of $\boldsymbol{q}_{r,\nu}$ with $\beta_r \boldsymbol{q}_{r,\nu}$, where $\beta_r := \arg\min_\beta J_c(\beta \boldsymbol{q}_{r,\nu})$. This implies that the direction (rather than the size) of $\boldsymbol{q}_{r,\nu}$ is important; the tightness of the CM-UMSE bound (4.2) will depend on collinearity of $\boldsymbol{q}_{r,\nu}$ and $\boldsymbol{q}_{c,\nu}$. Fig. 4.1 presents an illustration of this idea.

Zeng has shown that in the case of an i.i.d. source, a FIR channel and AWGN noise, $\boldsymbol{q}_{c,\nu}$ is nearly collinear to the MMSE response $\boldsymbol{q}_{m,\nu}$ [Zeng TSP 99]. These findings, together with the abundant interpretations of the MMSE estimator and the existence of closed-form expressions for $\boldsymbol{q}_{m,\nu}$ (e.g., (2.27) and (2.28)), suggest the reference choice $\boldsymbol{q}_{r,\nu} = \boldsymbol{q}_{m,\nu}$.

Determining a CM-UMSE upper bound from (4.2) can be accomplished as follows. Since both $J_c(\boldsymbol{q})$ and $J_{u,\nu}(\boldsymbol{q})$ are invariant to phase rotation of $\boldsymbol{q}$ (i.e., scalar multiplication of $\boldsymbol{q}$ by $e^{j\phi}$ for $\phi \in \mathbb{R}$), we can restrict our attention to the set of "de-rotated" responses $\{\boldsymbol{q} \text{ s.t. } q_\nu^{(0)} \in \mathbb{R}^+\}$. Such $\boldsymbol{q}$ allow parameterization in terms of gain $a = \|\boldsymbol{q}\|_2$ and interference response $\bar{\boldsymbol{q}}$ (defined in Section 2.3.2), where $\|\bar{\boldsymbol{q}}\|_2 \le a$. In terms of the pair $(a, \bar{\boldsymbol{q}})$, the upper bound in (4.2) may then be rewritten

$$\max_{\boldsymbol{q} \in \mathcal{Q}_c(\beta_r \boldsymbol{q}_{r,\nu})} J_{u,\nu}(\boldsymbol{q}) \;=\; \max_a \left(\max_{\bar{\boldsymbol{q}}:\, (a,\bar{\boldsymbol{q}}) \in \mathcal{Q}_c(\beta_r \boldsymbol{q}_{r,\nu})} J_{u,\nu}(a, \bar{\boldsymbol{q}}) \right).$$

Under particular conditions on the gain a and the reference $\boldsymbol{q}_{r,\nu}$ (made explicit in Section 4.2.2), there exists a minimum interference gain

$$b_*(a) \;:=\; \min b(a) \text{ s.t. } \Big\{ (a, \bar{\boldsymbol{q}}) \in \mathcal{Q}_c(\beta_r \boldsymbol{q}_{r,\nu}) \;\;\Rightarrow\;\; \|\bar{\boldsymbol{q}}\|_2 \le b(a) \Big\}, \tag{4.3}$$

which can be used in the containment:

$$\Big\{ (a, \bar{\boldsymbol{q}}) \in \mathcal{Q}_c(\beta_r \boldsymbol{q}_{r,\nu}) \Big\} \;\subset\; \Big\{ (a, \bar{\boldsymbol{q}}) \;\text{ s.t. }\; \|\bar{\boldsymbol{q}}\|_2 \le b_*(a) \Big\},$$

implying

$$\max_{\bar{\boldsymbol{q}}:\, (a,\bar{\boldsymbol{q}}) \in \mathcal{Q}_c(\beta_r \boldsymbol{q}_{r,\nu})} J_{u,\nu}(a, \bar{\boldsymbol{q}}) \;\;\le\;\; \max_{\bar{\boldsymbol{q}}:\, \|\bar{\boldsymbol{q}}\|_2 \le b_*(a)} J_{u,\nu}(a, \bar{\boldsymbol{q}}).$$

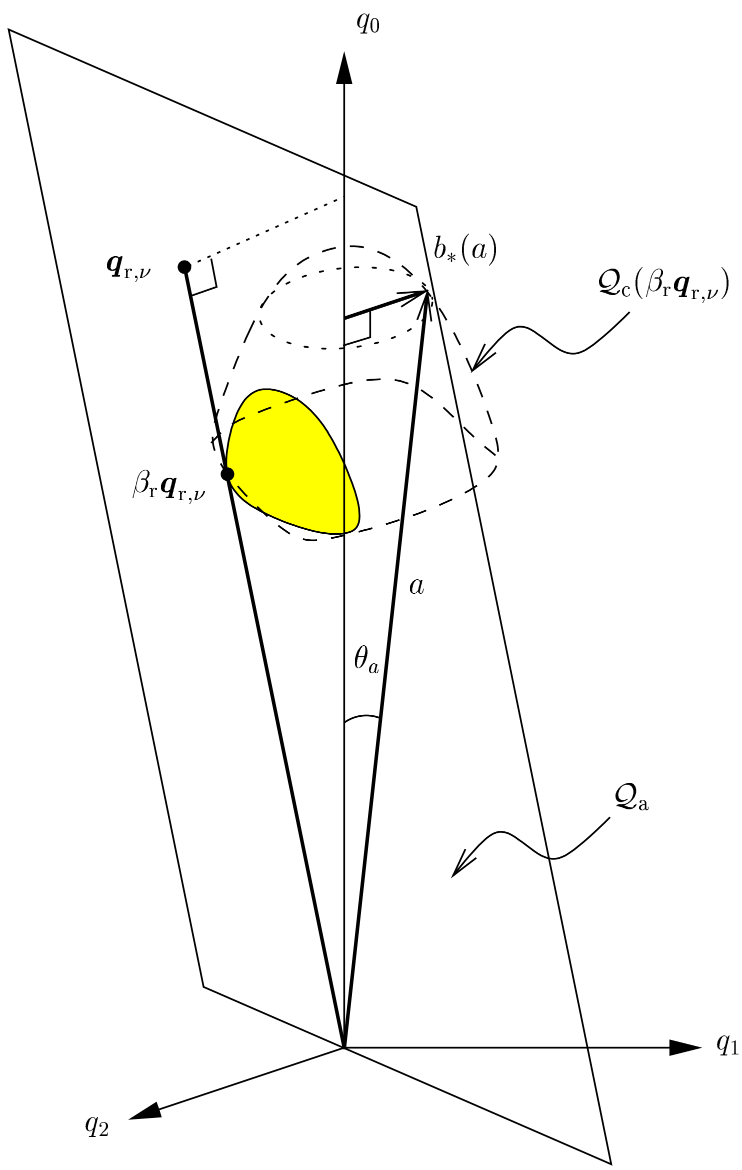

Figure 4.1: Illustration of CM-UMSE upper-bounding technique using reference $\boldsymbol{q}_{r,\nu}$.

Applying (2.31) to the previous statement yields

$$\begin{aligned}\max_{\bar{\boldsymbol{q}}:\,\|\bar{\boldsymbol{q}}\|_2\le b_*(a)} J_{\mathrm{u},\nu}(a,\bar{\boldsymbol{q}}) &= \max_{\bar{\boldsymbol{q}}:\,\|\bar{\boldsymbol{q}}\|_2\le b_*(a)} \left(\frac{\|\bar{\boldsymbol{q}}\|_2^2}{a^2-\|\bar{\boldsymbol{q}}\|_2^2}\right)\sigma_s^2 \\ &= \left(\frac{b_*^2(a)}{a^2-b_*^2(a)}\right)\sigma_s^2,\end{aligned}$$

and putting these arguments together, we arrive at the CM-UMSE bound

$$J_{\mathrm{u},\nu}(\boldsymbol{q}_{\mathrm{c},\nu}) \;\le\; \max_a \left(\frac{b_*^2(a)}{a^2-b_*^2(a)}\right)\sigma_s^2. \tag{4.4}$$

The roles of various quantities can be summarized using Fig. 4.1. Starting with the (attainable) global reference response $\boldsymbol{q}_{\mathrm{r},\nu}$, the scalar β_{r} minimizes the CM cost that characterizes all scaled versions of $\boldsymbol{q}_{\mathrm{r},\nu}$. Since the CM minimum $\boldsymbol{q}_{\mathrm{c},\nu}$ is known to lie within the set $\mathcal{Q}_{\mathrm{c}}(\beta_{\mathrm{r}}\boldsymbol{q}_{\mathrm{r},\nu})$, delineated in Fig. 4.1 by long-dashed lines, the maximum UMSE within $\mathcal{Q}_{\mathrm{c}}(\beta_{\mathrm{r}}\boldsymbol{q}_{\mathrm{r},\nu})$ forms a valid upper bound for CM-UMSE.[4] Determining the maximum UMSE within $\mathcal{Q}_{\mathrm{c}}(\beta_{\mathrm{r}}\boldsymbol{q}_{\mathrm{r},\nu})$ is accomplished by first deriving $b_*(a)$, the smallest upper bound on interference gain for all $\boldsymbol{q}\in\mathcal{Q}_{\mathrm{c}}(\beta_{\mathrm{r}}\boldsymbol{q}_{\mathrm{r},\nu})$ that have a total gain of a, and then finding the particular combination of $\{a, b_*(a)\}$ that maximizes UMSE. The angle θ_a shown in Fig. 4.1 gives a simple trigonometric interpretation of the UMSE bound (4.4): $J_{\mathrm{u},\nu}(\boldsymbol{q}_{\mathrm{c},\nu}) \le \max_a \tan^2(\theta_a)$. Also apparent from Fig. 4.1 is the notion that the valid range for a will depend on the choice of $\boldsymbol{q}_{\mathrm{r},\nu}$.

4.2.2 Derivation of the CM-UMSE Bounds

In this section we derive CM-UMSE bounds based on the method described in Section 4.2.1. The main steps in the derivation are presented as lemmas, with proofs appearing in Appendix 4.A.

[4]Though a tighter CM-UMSE bound would follow from use of the fact that $\exists\, \boldsymbol{q}_{\mathrm{c},\nu}\in\mathcal{Q}_{\mathrm{c}}(\beta_{\mathrm{r}}\boldsymbol{q}_{\mathrm{r},\nu})\cap\mathcal{Q}_{\mathrm{a}}$ (denoted by the shaded area in Fig. 4.1), the set $\mathcal{Q}_{\mathrm{c}}(\beta_{\mathrm{r}}\boldsymbol{q}_{\mathrm{r},\nu})\cap\mathcal{Q}_{\mathrm{a}}$ is too difficult to describe analytically.

The first step is to express the CM cost (2.35) in terms of the global response $\boldsymbol{q}$ (defined in Section 2.2).

Lemma 4.1. *The CM cost may be written in terms of global response* $\boldsymbol{q}$ *as*

$$\frac{J_{\mathrm{c}}(\boldsymbol{q})}{\sigma_s^4} \;=\; \sum_k (\kappa_s^{(k)} - \kappa_g)\|\boldsymbol{q}^{(k)}\|_4^4 + \kappa_g\|\boldsymbol{q}\|_2^4 - 2(\gamma/\sigma_s^2)\|\boldsymbol{q}\|_2^2 + (\gamma/\sigma_s^2)^2. \tag{4.5}$$

Similar expressions for the CM cost have been generated for the case of a desired user in AWGN (see, e.g., [Johnson PROC 98]).

The CM cost expression (4.5) can now be used to compute the CM cost at scaled versions of a reference $\boldsymbol{q}_{\mathrm{r},\nu}$.

Lemma 4.2. *For any* $\boldsymbol{q}_{\mathrm{r},\nu}$*,*

$$\beta_{\mathrm{r}} = \arg\min_{\beta} J_{\mathrm{c}}(\beta\boldsymbol{q}_{\mathrm{r},\nu}) \;=\; \frac{1}{\|\boldsymbol{q}_{\mathrm{r},\nu}\|_2}\sqrt{\left(\frac{\gamma}{\sigma_s^2}\right)\frac{1}{\kappa_{y_{\mathrm{r}}}}},$$

and

$$J_{\mathrm{c}}(\beta_{\mathrm{r}}\boldsymbol{q}_{\mathrm{r},\nu}) \;=\; \gamma^2\left(1 - \kappa_{y_{\mathrm{r}}}^{-1}\right), \tag{4.6}$$

where $\kappa_{y_{\mathrm{r}}}$ *is the normalized kurtosis of the estimates generated by the reference* $\boldsymbol{q}_{\mathrm{r},\nu}$*.*

The expression for $J_{\mathrm{c}}(\beta_{\mathrm{r}}\boldsymbol{q}_{\mathrm{r},\nu})$ in (4.6) leads directly to an expression for $\mathcal{Q}_{\mathrm{c}}(\beta_{\mathrm{r}}\boldsymbol{q}_{\mathrm{r},\nu})$, from which the minimum interference gain $b_*(a)$ of (4.3) can be derived.

Lemma 4.3. *The non-negative gain* $b_*(a)$ *satisfying definition (4.3) can be upper bounded as*

$$b_*(a) \;\leq\; a\sqrt{\frac{1-\sqrt{1-(\rho_{\min}+1)\frac{C(a,\boldsymbol{q}_{\mathrm{r},\nu})}{a^4}}}{\rho_{\min}+1}} \quad \textit{when} \quad 0 \leq \frac{C(a,\boldsymbol{q}_{\mathrm{r},\nu})}{a^4} \leq \frac{3-\rho_{\min}}{4}, \tag{4.7}$$

where $C(a,\boldsymbol{q}_{\mathrm{r},\nu})$ *is defined in (4.22).*

Equations (4.4) and (4.7) lead to an upper bound for the UMSE of CM-minimizing estimators.

Theorem 4.1. *When there exists a Wiener estimator associated with the desired user at delay ν generating estimates with kurtosis $\kappa_{y_{\mathrm{m}}}$ obeying*

$$\frac{1+\rho_{\min}}{4} \;<\; \frac{\kappa_g - \kappa_{y_{\mathrm{m}}}}{\kappa_g - \kappa_s^{(0)}} \;\leq\; 1, \tag{4.8}$$

the UMSE of CM-minimizing estimators associated with the same user/delay can be upper bounded by $J_{\mathrm{u},\nu}\big|_{\mathrm{c},\nu}^{\max,\kappa_{y_{\mathrm{m}}}}$, where

$$J_{\mathrm{u},\nu}\big|_{\mathrm{c},\nu}^{\max,\kappa_{y_{\mathrm{m}}}} \;:=\; \frac{1-\sqrt{(\rho_{\min}+1)\,\frac{\kappa_g-\kappa_{y_{\mathrm{m}}}}{\kappa_g-\kappa_s^{(0)}}-\rho_{\min}}}{\rho_{\min}+\sqrt{(\rho_{\min}+1)\,\frac{\kappa_g-\kappa_{y_{\mathrm{m}}}}{\kappa_g-\kappa_s^{(0)}}-\rho_{\min}}}\,\sigma_s^2. \tag{4.9}$$

Furthermore, (4.8) guarantees the existence of a CM-minimizing estimator associated with this user/delay when $\boldsymbol{q}$ is FIR.

While Theorem 4.1 presents a closed-form CM-UMSE bounding expression in terms of the kurtosis of the MMSE estimates, it is also possible to derive lower and upper bounds in terms of the UMSE of MMSE estimators.

Theorem 4.2. *If Wiener UMSE $J_{\mathrm{u},\nu}(\boldsymbol{q}_{\mathrm{m},\nu}) < J_{\mathrm{o}}\sigma_s^2$, where*

$$J_{\mathrm{o}} \;:=\; \begin{cases} 2\sqrt{(1+\rho_{\min})^{-1}}-1 & \kappa_s^{\max} \leq \kappa_g \\ \frac{1-\sqrt{1-(3-\rho_{\min})(1+\rho_{\max})/4}}{\rho_{\max}+\sqrt{1-(3-\rho_{\min})(1+\rho_{\max})/4}}, & \kappa_s^{\max} > \kappa_g,\;\; \rho_{\max} \neq -1, \\ \frac{3-\rho_{\min}}{5+\rho_{\min}} & \kappa_s^{\max} > \kappa_g,\;\; \rho_{\max} = -1. \end{cases} \tag{4.10}$$

the UMSE of CM-minimizing estimators associated with the same user/delay can be upper bounded as follows:

$$J_{\mathrm{u},\nu}(\boldsymbol{q}_{\mathrm{m},\nu}) \;\leq\; J_{\mathrm{u},\nu}(\boldsymbol{q}_{\mathrm{c},\nu}) \;\leq\; J_{\mathrm{u},\nu}\big|_{\mathrm{c},\nu}^{\max,\kappa_{y_{\mathrm{m}}}} \;\leq\; J_{\mathrm{u},\nu}\big|_{\mathrm{c},\nu}^{\max,J_{\mathrm{u},\nu}(\boldsymbol{q}_{\mathrm{m},\nu})},$$

where

$$J_{\mathrm{u},\nu}\Big|_{\mathrm{c},\nu}^{\max,J_{\mathrm{u},\nu}(\boldsymbol{q}_{\mathrm{m},\nu})} := \begin{cases} \dfrac{1-\sqrt{(1+\rho_{\min})\left(1+\frac{J_{\mathrm{u},\nu}(\boldsymbol{q}_{\mathrm{m},\nu})}{\sigma_s^2}\right)^{-2}-\rho_{\min}}}{\rho_{\min}+\sqrt{(1+\rho_{\min})\left(1+\frac{J_{\mathrm{u},\nu}(\boldsymbol{q}_{\mathrm{m},\nu})}{\sigma_s^2}\right)^{-2}-\rho_{\min}}}\,\sigma_s^2 & \kappa_s^{\max} \leq \kappa_g \\ \dfrac{1-\sqrt{(1+\rho_{\min})\left(1+\frac{J_{\mathrm{u},\nu}(\boldsymbol{q}_{\mathrm{m},\nu})}{\sigma_s^2}\right)^{-2}\left(1+\rho_{\max}\frac{J^2_{\mathrm{u},\nu}(\boldsymbol{q}_{\mathrm{m},\nu})}{\sigma_s^4}\right)-\rho_{\min}}}{\rho_{\min}+\sqrt{(1+\rho_{\min})\left(1+\frac{J_{\mathrm{u},\nu}(\boldsymbol{q}_{\mathrm{m},\nu})}{\sigma_s^2}\right)^{-2}\left(1+\rho_{\max}\frac{J^2_{\mathrm{u},\nu}(\boldsymbol{q}_{\mathrm{m},\nu})}{\sigma_s^4}\right)-\rho_{\min}}}\,\sigma_s^2 & \kappa_s^{\max} > \kappa_g. \end{cases} \tag{4.11}$$

Furthermore, (4.10) guarantees the existence of a CM-minimizing estimator associated with this user/delay when $\boldsymbol{q}$ *is FIR.*

Note that the two cases of J_{o} in (4.10) and of $J_{\mathrm{u},\nu}\big|_{\mathrm{c},\nu}^{\max,J_{\mathrm{u},\nu}(\boldsymbol{q}_{\mathrm{m},\nu})}$ in (4.11) coincide as $\kappa_s^{\max} \to \kappa_g$.

Equation (4.11) leads to an elegant approximation of the *extra* UMSE of CM-minimizing estimators:

$$\mathcal{E}_{\mathrm{u},\nu}(\boldsymbol{q}_{\mathrm{c},\nu}) := J_{\mathrm{u},\nu}(\boldsymbol{q}_{\mathrm{c},\nu}) - J_{\mathrm{u},\nu}(\boldsymbol{q}_{\mathrm{m},\nu}).$$

Theorem 4.3. *If* $J_{\mathrm{u},\nu}(\boldsymbol{q}_{\mathrm{m},\nu}) < J_{\mathrm{o}}\sigma_s^2$*, then the extra UMSE of CM-minimizing estimators can be bounded as* $\mathcal{E}_{\mathrm{u},\nu}(\boldsymbol{q}_{\mathrm{c},\nu}) \leq \mathcal{E}_{\mathrm{u},\nu}\big|_{\mathrm{c},\nu}^{\max,J_{\mathrm{u},\nu}(\boldsymbol{q}_{\mathrm{m},\nu})}$*, where*

$$\begin{aligned} \mathcal{E}_{\mathrm{u},\nu}\Big|_{\mathrm{c},\nu}^{\max,J_{\mathrm{u},\nu}(\boldsymbol{q}_{\mathrm{m},\nu})} &:= J_{\mathrm{u},\nu}\Big|_{\mathrm{c},\nu}^{\max,J_{\mathrm{u},\nu}(\boldsymbol{q}_{\mathrm{m},\nu})} - J_{\mathrm{u},\nu}(\boldsymbol{q}_{\mathrm{m},\nu}) \\ &= \begin{cases} \frac{1}{2\sigma_s^2}\rho_{\min} J^2_{\mathrm{u},\nu}(\boldsymbol{q}_{\mathrm{m},\nu}) + \mathcal{O}\big(J^3_{\mathrm{u},\nu}(\boldsymbol{q}_{\mathrm{m},\nu})\big) & \kappa_s^{\max} \leq \kappa_g \\ \frac{1}{2\sigma_s^2}(\rho_{\min} - \rho_{\max}) J^2_{\mathrm{u},\nu}(\boldsymbol{q}_{\mathrm{m},\nu}) + \mathcal{O}\big(J^3_{\mathrm{u},\nu}(\boldsymbol{q}_{\mathrm{m},\nu})\big) & \kappa_s^{\max} > \kappa_g \end{cases} \end{aligned} \tag{4.12}$$

Equation (4.12) implies that the extra UMSE of CM-minimizing estimators is upper bounded by approximately the *square* of the minimum UMSE. Fig. 4.2 plots the upper bound on CM-UMSE and extra CM-UMSE from (4.11) as a function of

$J_{\mathrm{u},\nu}(\boldsymbol{q}_{\mathrm{m},\nu})/\sigma_s^2$ for various values of $\rho_{\min}$ and $\rho_{\max}$. The second-order approximation based on (4.12) appears very good for all but the largest values of UMSE.

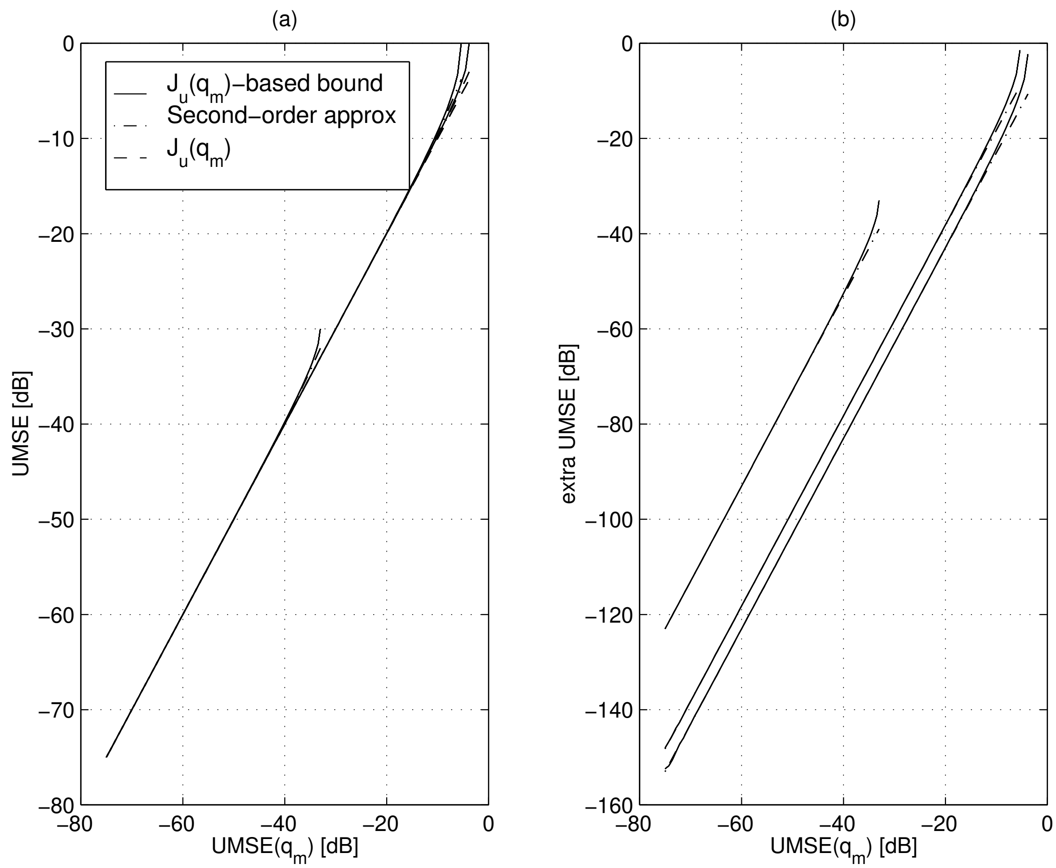

Figure 4.2: Upper bound on (a) CM-UMSE and (b) extra CM-UMSE versus $J_{\mathrm{u},\nu}(\boldsymbol{q}_{\mathrm{m},\nu})$ (when $\sigma_s^2 = 1$) from (4.11) with second-order approximation from (4.12). From left to right, $\{\rho_{\min}, \rho_{\max}\} = \{1000, 0\}$, $\{1, -2\}$, and $\{1, 0\}$.

4.2.3 Comments on the CM-UMSE Bounds

Implicit Incorporation of $\mathcal{Q}_{\mathrm{a}}$

First, recall that the CM-UMSE bounding procedure incorporated $\mathcal{Q}_{\mathrm{a}}$, the set of attainable global responses, *only* in the requirement that $\boldsymbol{q}_{\mathrm{r},\nu} \in \mathcal{Q}_{\mathrm{a}}$. Thus Theorems 4.1–4.3, written under the reference choice $\boldsymbol{q}_{\mathrm{r},\nu} = \boldsymbol{q}_{\mathrm{m},\nu} \in \mathcal{Q}_{\mathrm{a}} \cap \mathcal{Q}_{\nu}^{(0)}$, implicitly incorporate the channel and/or estimator constraints that define $\mathcal{Q}_{\mathrm{a}}$. For example, if $\boldsymbol{q}_{\mathrm{m},\nu}$ is the MMSE response constrained to a set of finitely-parameterized ARMA estimators, then CM-UMSE bounds based on this $\boldsymbol{q}_{\mathrm{m},\nu}$ will implicitly incorporate the causality constraint. The implicit incorporation of the attainable set $\mathcal{Q}_{\mathrm{a}}$ makes these bounding theorems quite general and easy to use.

Effect of $\rho_{\min}$

When $\kappa_s^{\max} \le \kappa_g$ and $\rho_{\min} = \frac{\kappa_g - \kappa_s^{\min}}{\kappa_g - \kappa_s^{(0)}} = 1$, the expressions in Theorems 4.1–4.3 simplify:

$$
\begin{aligned}
J_{\mathrm{u},\nu}(\boldsymbol{q}_{\mathrm{c},\nu}) \;&\le\; \frac{1 - \sqrt{2\frac{\kappa_g - \kappa_{y_{\mathrm{m}}}}{\kappa_g - \kappa_s^{(0)}} - 1}}{1 + \sqrt{2\frac{\kappa_g - \kappa_{y_{\mathrm{m}}}}{\kappa_g - \kappa_s^{(0)}} - 1}}\,\sigma_s^2 \quad \text{when} \quad \frac{1}{2} < \frac{\kappa_g - \kappa_{y_{\mathrm{m}}}}{\kappa_g - \kappa_s^{(0)}} \le 1, \\
&\le\; \frac{1 - \sqrt{2\left(1 + \frac{J_{\mathrm{u},\nu}(\boldsymbol{q}_{\mathrm{m},\nu})}{\sigma_s^2}\right)^{-2} - 1}}{1 + \sqrt{2\left(1 + \frac{J_{\mathrm{u},\nu}(\boldsymbol{q}_{\mathrm{m},\nu})}{\sigma_s^2}\right)^{-2} - 1}}\,\sigma_s^2 \quad \text{when} \quad \frac{J_{\mathrm{u},\nu}(\boldsymbol{q}_{\mathrm{m},\nu})}{\sigma_s^2} < \sqrt{2} - 1, \\
&=\; J_{\mathrm{u},\nu}(\boldsymbol{q}_{\mathrm{m},\nu}) + \frac{1}{2\sigma_s^2} J_{\mathrm{u},\nu}^2(\boldsymbol{q}_{\mathrm{m},\nu}) + \mathcal{O}\big(J_{\mathrm{u},\nu}^3(\boldsymbol{q}_{\mathrm{m},\nu})\big).
\end{aligned}
$$

Typical scenarios leading to $\rho_{\min} = 1$ include

a) sub-Gaussian desired source in the presence of AWGN, or

b) constant-modulus desired source in the presence of non-super-Gaussian interference.

Note that in the two cases above, the CM-UMSE upper bound is independent of the specific distribution of the desired and interfering sources, respectively.

The case $\rho_{\min} > 1$, on the other hand, might arise from the use of dense (and/or shaped) source constellations in the presence of interfering sources that are "more sub-Gaussian." In fact, source assumption S4) allows for arbitrarily large $\rho_{\min}$, which could result from a nearly-Gaussian desired source in the presence of non-Gaussian interference. Though Theorems 4.1–4.3 remain valid for arbitrarily high $\rho_{\min}$, the requirements placed on $\boldsymbol{q}_{\mathrm{m},\nu}$ via J_{o} become more stringent (recall Fig. 4.2).

Generalization of Perfect CM-Estimation Property

Finally, we note that the $J_{\mathrm{u},\nu}(\boldsymbol{q}_{\mathrm{m},\nu})$-based CM-UMSE bound in Theorem 4.2 implies that the perfect CM-estimation property, proven under more restrictive conditions in [Foschini ATT 85]-[Li TSP 96a], extends to the general multi-source linear model of Fig. 2.3:

Corollary 4.1. *CM-minimizing estimators are perfect (up to a scaling) when Wiener estimators are perfect.*

Proof. From Theorem 4.2, $J_{\mathrm{u},\nu}(\boldsymbol{q}_{\mathrm{m},\nu}) = 0 \;\Rightarrow\; J_{\mathrm{u},\nu}(\boldsymbol{q}_{\mathrm{c},\nu}) = 0$. Hence, the estimators are perfect up to a (fixed) scale factor. □

4.3 Numerical Examples

In Sections 4.3.1–4.3.3, we compare the UMSE bounds in (4.9) and (4.11) to the UMSE bound of the Zeng et al. method of Table 4.1, to the UMSE of the CM-minimizing estimators found by gradient descent,[5] and to the minimum UMSE (i.e., that obtained by the MMSE solution). The results suggest that, over a wide range of conditions, (i) the CM-UMSE bounds are close to the CM-UMSE found by gradient descent, and (ii) the CM-UMSE performance is close to the optimal

[5] Gradient descent results were obtained via the MATLAB routine "`fminu`," which was initialized randomly in a small ball around the MMSE estimator.

UMSE performance. In other words, the CM-UMSE bounds are tight, and the CM-minimizing estimator is robust in a MSE sense.

4.3.1 Performance versus Estimator Length for Fixed Channel

In practical equalization applications, CM-minimizing estimators will not be perfect due to violation of the FCR $\boldsymbol{\mathcal{H}}$ requirement discussed in Section 2.5. For instance, even in the absence of noise and interferers, insufficient estimator length can lead to a matrix $\boldsymbol{\mathcal{H}}$ that is wider than tall, thus preventing FCR. For FIR channels with adequate "diversity," it is well known that there exists a finite estimator length sufficient for the achievement of FCR $\boldsymbol{\mathcal{H}}$. When diversity is not adequate, however, as with a baud-spaced scalar channel (i.e., $P = 1$) or with multiple channels sharing common zeros,[6] there exists no finite sufficient length. Consequently, the performance of the CM criterion under so-called "channel undermodelling" and "lack of disparity" has been a topic of recent interest (see, e.g., [Fijalkow TSP 97], [Li TSP 96b], [Endres TSP 99], [Regalia TSP 99]).

Using the $T/2$-spaced microwave channel impulse response model #5 from the Signal Processing Information Base (SPIB) database, CM-minimizing estimator performance was calculated versus estimator length. Fig. 4.3(a) plots the UMSE of CM-minimizing estimators as predicted by various bounds and by gradient descent. Note that all methods yield CM-UMSE bounds nearly indistinguishable from the minimum UMSE. Fig. 4.3(b) plots the same information in the form of extra CM-UMSE (i.e., CM-UMSE minus minimum UMSE), and once again we see that the bounds are tight and give nearly identical performance. For the higher equalizer

[6] See, e.g., [Johnson PROC 98] or [Johnson Chap 99] for more information on length and diversity requirements.

lengths, it is apparent that numerical inaccuracies prevented the CM gradient descent procedure from finding the true minimum (resulting in $\times$'s above the upper bound line).

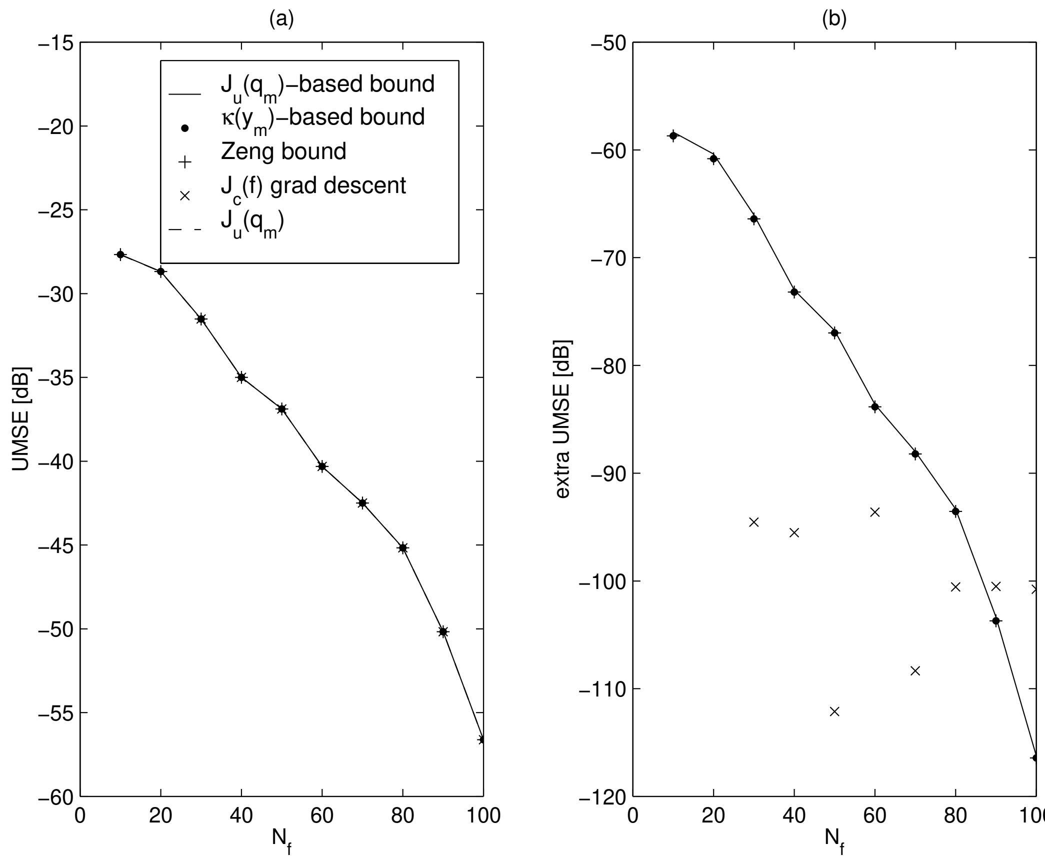

Figure 4.3: Bounds on CM-UMSE versus estimator length N_f for SPIB microwave channel #5 and 8-PAM.

4.3.2 Performance versus AWGN for Fixed Channel

Using the same microwave channel model, we conducted a different experiment in which AWGN was introduced at various power levels (for fixed equalizer length $N_f = 20$). Fig. 4.4(a) shows that the UMSE predicted by the CM bounds is very close to that predicted by gradient descent for all but the highest levels of AWGN, and as before, the CM-UMSE performance is quite close to the minimum UMSE performance. Fig. 4.4(b) reveals slight differences in bound performance: Zeng et al.'s algorithmic bound appears slightly tighter than our closed-form bounds at lower SNR.

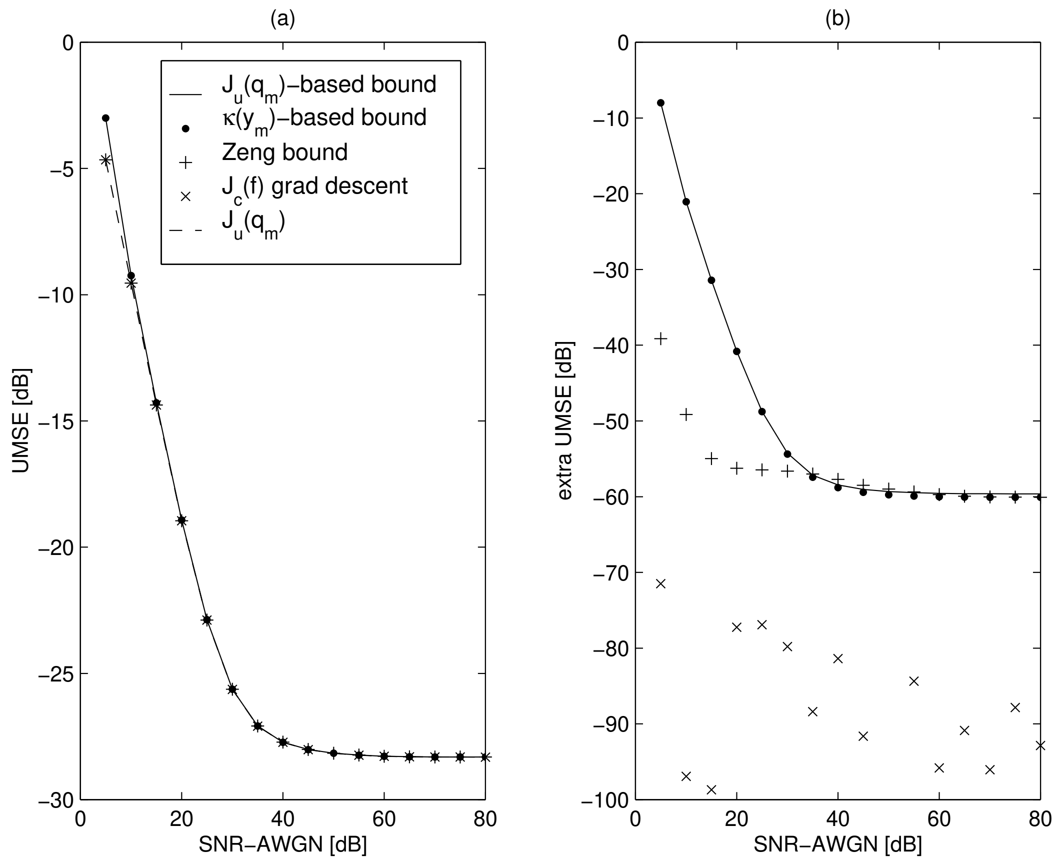

Figure 4.4: Bounds on CM-UMSE versus SNR of AWGN for SPIB microwave channel 5, $N_f = 20$, and 8-PAM.

4.3.3 Performance with Random Channels

While the convolutive nature of the channel in equalization applications gives $\boldsymbol{\mathcal{H}}$ a block-Toeplitz structure, other applications (e.g., beamforming) may lead to $\boldsymbol{\mathcal{H}}$

with a more general, non-Toeplitz, structure. When the number of sources is greater than the estimator length (which, in our model, is always the case when noise is present), the channel matrix $\boldsymbol{\mathcal{H}}$ will be non-FCR and different estimation techniques will yield different levels of performance.

Here we present the results of experiments where $\boldsymbol{\mathcal{H}}$ was generated with zero-mean Gaussian entries. Fig. 4.5 corresponds to a desired source having constant modulus (i.e., $\kappa_s^{(0)} = 1$) in the presence of AWGN and constant modulus interference, Fig. 4.6 corresponds to a nearly-Gaussian desired source in the same interference environment, and Fig. 4.7 corresponds to a desired source with constant modulus in the presence of AWGN and super-Gaussian interference. As with our previous experiments, Fig.s 4.5–4.7 demonstrate that (i) the closed-form CM-UMSE bounds are tight and (ii) that the CM-minimizing estimators generate nearly-MMSE estimates under arbitrary forms of additive interference.

4.4 Conclusions

In this chapter we have derived, for the general multi-source linear model of Fig. 2.3, two closed-form bounding expressions for the UMSE of CM-minimizing estimators. The first bound is based on the kurtosis of the MMSE estimates, while the second is based on the UMSE of the MMSE estimators. Analysis of the second bound shows that the *extra* UMSE of CM-minimizing estimators is upper bounded by approximately the *square* of the minimum UMSE. Thus, the CM-minimizing estimator generates nearly-MMSE estimates when the minimum MSE is small. Numerical simulations suggest that the bounds are tight (w.r.t. the performance of CM-minimizing estimators designed by gradient descent).

This work confirms the longstanding conjecture (see, e.g., [Godard TCOM 80]

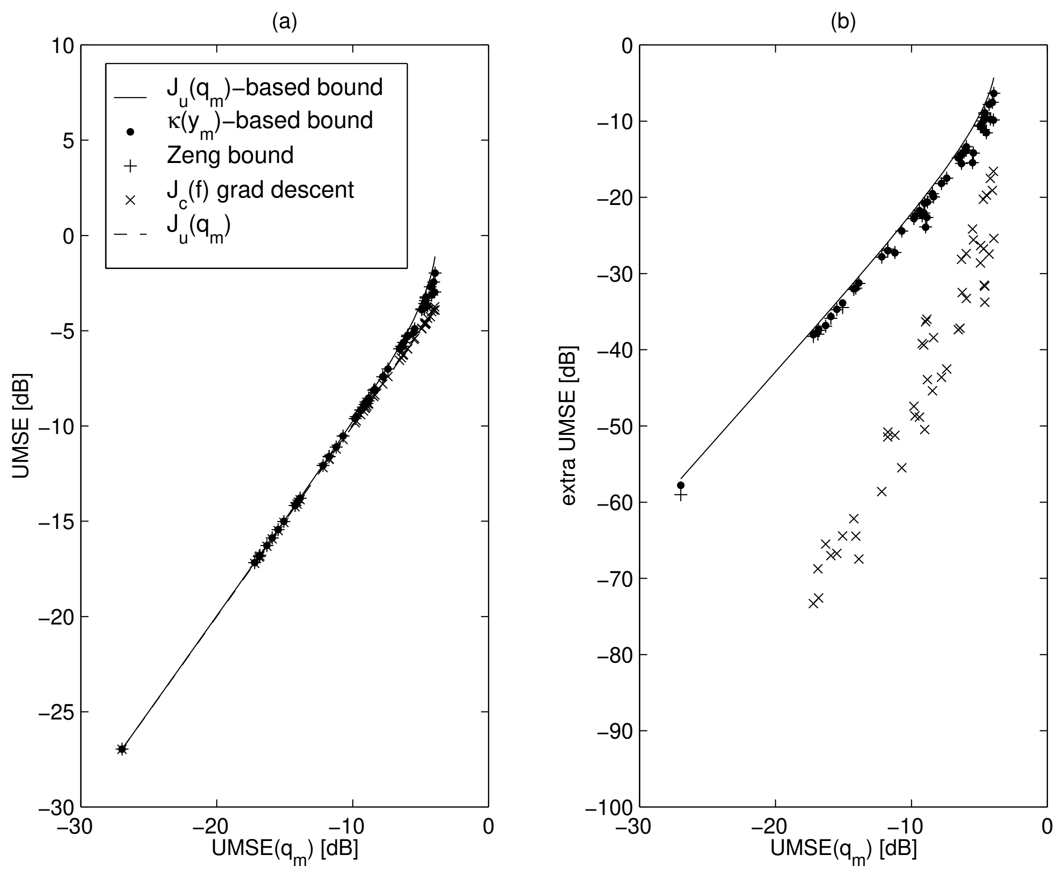

Figure 4.5: Bounds on CM-UMSE for $N_f = 8$, 10 BPSK sources, AWGN at -40dB, and random $\boldsymbol{\mathcal{H}}$.

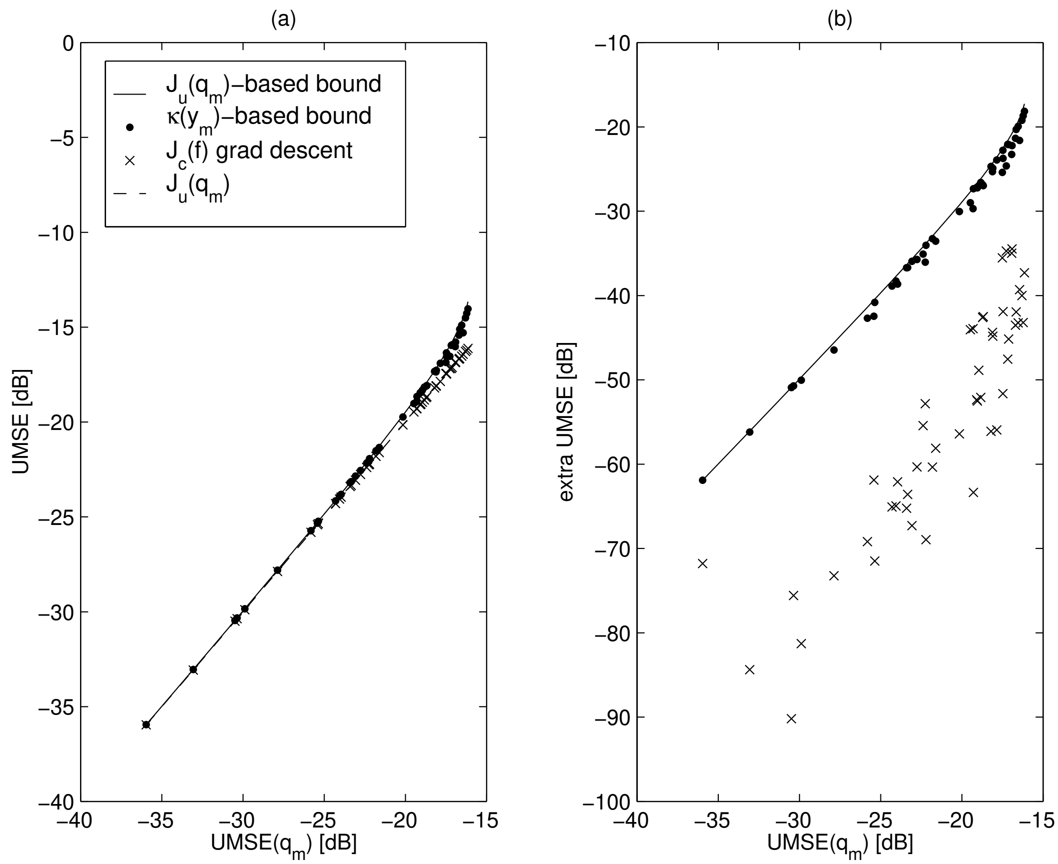

Figure 4.6: Bounds on CM-UMSE for $N_f = 8$, 5 BPSK sources, 5 sources with $\kappa_s^{(k)} = 2.9$ (one of which is desired), AWGN at -40dB, and random $\boldsymbol{\mathcal{H}}$.

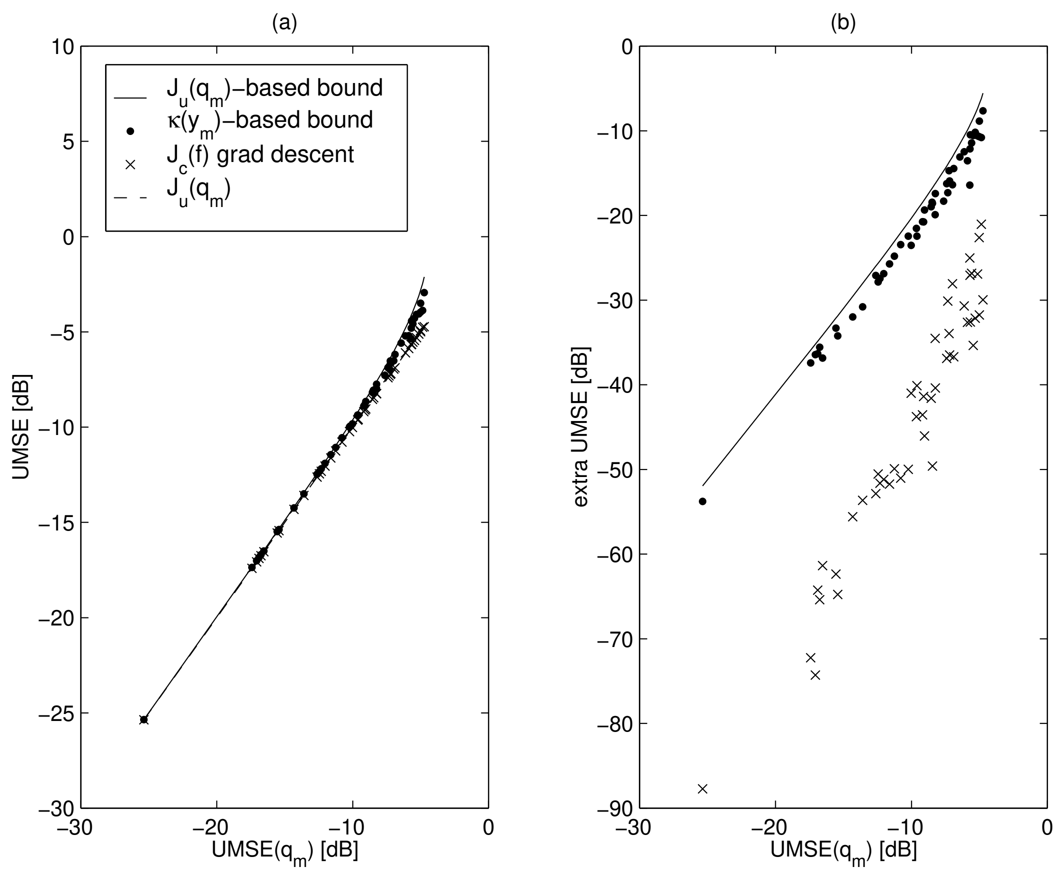

Figure 4.7: Bounds on CM-UMSE for $N_f = 8$, 5 BPSK sources (one of which is desired), 5 sources with $\kappa_s^{(k)} = 4$, AWGN at -40dB, and random $\boldsymbol{\mathcal{H}}$.

and [Treichler TASSP 83]) that the MSE performance of the CM-minimizing estimator is robust to *general* linear channels and general (multi-source) additive interference. As such, our results supersede previous work demonstrating the MSE-robustness of CM-minimizing estimators in special cases (e.g., when only AWGN is present, when the channel does not provide adequate diversity, or when the estimator has an insufficient number of adjustable parameters).

Appendix

4.A Derivation Details for CM-UMSE Bounds

This appendix contains the proofs of the theorems and lemmas found in Section 4.2.

4.A.1 Proof of Lemma 4.1

In this section we derive an expression for the CM cost J_c in terms of the global response $\boldsymbol{q}$. From (2.14) and (2.35),

$$\begin{aligned} J_c(y_n) &= \mathrm{E}\big\{|y_n|^4\big\} - 2\gamma\,\mathrm{E}\big\{|y_n|^2\big\} + \gamma^2 \\ &= \mathcal{K}(y_n) + 2\,\mathrm{E}^2\big\{|y_n|^2\big\} + \big|\mathrm{E}\big\{y_n^2\big\}\big|^2 - 2\gamma\,\mathrm{E}\big\{|y_n|^2\big\} + \gamma^2. \end{aligned} \tag{4.13}$$

Source assumptions S1)–S2) imply [Porat Book 94]

$$\mathcal{K}(y_n) = \sum_k \|\boldsymbol{q}^{(k)}\|_4^4\,\mathcal{K}(s_n^{(k)}). \tag{4.14}$$

From S3), S5), and the definitions of $\kappa_s^{(k)}$ and κ_g in (2.19) and (2.20),

$$\begin{aligned} \mathcal{K}(s_n^{(k)}) &= \begin{cases} \mathrm{E}\{|s_n^{(k)}|^4\} - 3\sigma_s^4, & \text{real-valued } \{s_n^{(k)}\} \\ \mathrm{E}\{|s_n^{(k)}|^4\} - 2\sigma_s^4, & \mathrm{E}\{s_n^{(k)2}\} = 0 \end{cases} \\ &= \mathrm{E}\{|s_n^{(k)}|^4\} - \kappa_g\sigma_s^4 \\ &= (\kappa_s^{(k)} - \kappa_g)\sigma_s^4. \end{aligned} \tag{4.15}$$

Similarly, S1)–S3) and S5) imply

$$\mathrm{E}\big\{|y_n|^2\big\} = \sum_k \|\boldsymbol{q}^{(k)}\|_2^2\sigma_s^2 = \|\boldsymbol{q}\|_2^2\sigma_s^2 \tag{4.16}$$

$$\mathrm{E}\big\{y_n^2\big\} = \begin{cases} \|\boldsymbol{q}\|_2^2\sigma_s^2, & \text{real-valued } \{s_n^{(k)}\}, \\ 0, & \mathrm{E}\{s_n^{(k)2}\} = 0\ \forall k. \end{cases} \tag{4.17}$$

Plugging (4.14)–(4.17) into (4.13), we arrive at (4.5).

4.A.2 Proof of Lemma 4.2

In this section, we are interested in computing $\beta_{\mathrm{r}} = \arg\min_\beta J_{\mathrm{c}}(\beta \boldsymbol{q}_{\mathrm{r},\nu})$. For any $\boldsymbol{q}_{\mathrm{r},\nu}$, (4.5) implies

$$\frac{J_{\mathrm{c}}(\beta \boldsymbol{q}_{\mathrm{r},\nu})}{\sigma_s^4} = \beta^4 \sum_k (\kappa_s^{(k)} - \kappa_g) \|\boldsymbol{q}_{\mathrm{r},\nu}^{(k)}\|_4^4 + \beta^4 \kappa_g \|\boldsymbol{q}_{\mathrm{r},\nu}\|_2^4 - 2\beta^2 (\gamma/\sigma_s^2) \|\boldsymbol{q}_{\mathrm{r},\nu}\|_2^2 + (\gamma/\sigma_s^2)^2. \tag{4.18}$$

Taking the partial derivative of (4.18) w.r.t. β,

$$\frac{\partial}{\partial \beta}\left\{\frac{J_{\mathrm{c}}(\beta \boldsymbol{q}_{\mathrm{r},\nu})}{\sigma_s^4}\right\} = 4\beta\left(\beta^2\left(\sum_k (\kappa_s^{(k)} - \kappa_g)\|\boldsymbol{q}_{\mathrm{r},\nu}^{(k)}\|_4^4 + \kappa_g \|\boldsymbol{q}_{\mathrm{r},\nu}\|_2^4\right) - (\gamma/\sigma_s^2)\|\boldsymbol{q}_{\mathrm{r},\nu}\|_2^2\right).$$

If we use (4.14)–(4.17) and definitions (2.19) and (2.20) to write the previous expression in terms of the normalized kurtosis of the reference estimates

$$\kappa_{y_{\mathrm{r}}} := \left.\frac{\mathrm{E}\{|y_n|^4\}}{\sigma_y^4}\right|_{\boldsymbol{q}_{\mathrm{r},\nu}} = \sum_k (\kappa_s^{(k)} - \kappa_g)\frac{\|\boldsymbol{q}_{\mathrm{r},\nu}^{(k)}\|_4^4}{\|\boldsymbol{q}_{\mathrm{r},\nu}\|_2^4} + \kappa_g, \tag{4.19}$$

we obtain

$$\frac{\partial}{\partial \beta}\left\{\frac{J_{\mathrm{c}}(\beta \boldsymbol{q}_{\mathrm{r},\nu})}{\sigma_s^4}\right\} = 4\beta\big(\beta^2 \kappa_{y_{\mathrm{r}}} \|\boldsymbol{q}_{\mathrm{r},\nu}\|_2^4 - (\gamma/\sigma_s^2)\|\boldsymbol{q}_{\mathrm{r},\nu}\|_2^2\big).$$

Setting the partial derivative equal to zero,

$$\beta_{\mathrm{r}} = \frac{1}{\|\boldsymbol{q}_{\mathrm{r},\nu}\|_2}\sqrt{\left(\frac{\gamma}{\sigma_s^2}\right)\kappa_{y_{\mathrm{r}}}^{-1}}.$$

Finally, plugging the expression for β_{r} into (4.18), we arrive at the expression for $J_{\mathrm{c}}(\beta_{\mathrm{r}} \boldsymbol{q}_{\mathrm{r},\nu})$ given in (4.6).

4.A.3 Proof of Lemma 4.3

In this section, we are interested in deriving an expression for the interference radius $b_*(a)$ defined in (4.3) and establishing conditions under which this radius is well defined. Rather than working with (4.3) directly, we find it easier to use the equivalent definition

$$b_*(a) \;=\; \min b(a) \text{ s.t. } \Big\{ \|\bar{\boldsymbol{q}}\|_2 > b(a) \;\;\Rightarrow\;\; J_{\mathrm{c}}(a, \bar{\boldsymbol{q}}) > J_{\mathrm{c}}(\beta_{\mathrm{r}} \boldsymbol{q}_{\mathrm{r},\nu}) \Big\}. \tag{4.20}$$

First we rewrite the CM cost expression (4.5) in terms of gain $a = \|\boldsymbol{q}\|_2$ and interference response $\bar{\boldsymbol{q}}$ (defined in Section 2.3.2). Using the fact that $|q_\nu^{(0)}|^2 = a^2 - \|\bar{\boldsymbol{q}}\|_2^2$,

$$\begin{aligned}
\sum_k (\kappa_s^{(k)} - \kappa_g) \|\boldsymbol{q}^{(k)}\|_4^4 &= (\kappa_s^{(0)} - \kappa_g) |q_\nu^{(0)}|^4 + \sum_k (\kappa_s^{(k)} - \kappa_g) \|\bar{\boldsymbol{q}}^{(k)}\|_4^4 \\
&= (\kappa_s^{(0)} - \kappa_g) \big(a^4 - 2a^2 \|\bar{\boldsymbol{q}}\|_2^2 + \|\bar{\boldsymbol{q}}\|_2^4 \big) + \sum_k (\kappa_s^{(k)} - \kappa_g) \|\bar{\boldsymbol{q}}^{(k)}\|_4^4.
\end{aligned}$$

Plugging the previous expression into (4.5), we find that

$$\begin{aligned}
\frac{J_{\mathrm{c}}(a, \bar{\boldsymbol{q}})}{\sigma_s^4} &= \sum_k (\kappa_s^{(k)} - \kappa_g) \|\bar{\boldsymbol{q}}^{(k)}\|_4^4 + \kappa_s^{(0)} a^4 - 2(\kappa_s^{(0)} - \kappa_g) a^2 \|\bar{\boldsymbol{q}}\|_2^2 + (\kappa_s^{(0)} - \kappa_g) \|\bar{\boldsymbol{q}}\|_2^4 \\
&\quad - 2(\gamma/\sigma_s^2) a^2 + (\gamma/\sigma_s^2)^2.
\end{aligned} \tag{4.21}$$

From (4.6) and (4.21), the following statements are equivalent:

$$\begin{aligned}
J_{\mathrm{c}}(\beta_{\mathrm{r}} \boldsymbol{q}_{\mathrm{r},\nu}) &< J_{\mathrm{c}}(a, \bar{\boldsymbol{q}}) \\
0 &< \sum_k (\kappa_s^{(k)} - \kappa_g) \|\bar{\boldsymbol{q}}^{(k)}\|_4^4 + (\kappa_s^{(0)} - \kappa_g) \Big(-2a^2 \|\bar{\boldsymbol{q}}\|_2^2 + \|\bar{\boldsymbol{q}}\|_2^4 \Big) \\
&\quad + \kappa_s^{(0)} a^4 - 2 \left(\frac{\gamma}{\sigma_s^2} \right) a^2 + \left(\frac{\gamma}{\sigma_s^2} \right)^2 \kappa_{y_{\mathrm{r}}}^{-1} \\
0 &> \frac{1}{\kappa_s^{(0)} - \kappa_g} \sum_k (\kappa_s^{(k)} - \kappa_g) \|\bar{\boldsymbol{q}}^{(k)}\|_4^4 - 2a^2 \|\bar{\boldsymbol{q}}\|_2^2 + \|\bar{\boldsymbol{q}}\|_2^4 \\
&\quad + \underbrace{\frac{1}{\kappa_s^{(0)} - \kappa_g} \left(\kappa_s^{(0)} a^4 - 2 \left(\frac{\gamma}{\sigma_s^2} \right) a^2 + \left(\frac{\gamma}{\sigma_s^2} \right)^2 \kappa_{y_{\mathrm{r}}}^{-1} \right)}_{C(a, \boldsymbol{q}_{\mathrm{r},\nu})}.
\end{aligned} \tag{4.22}$$

The reversal of inequality in (4.22) occurs because $\kappa_s^{(0)} - \kappa_g < 0$ (as implied by S4)). Since (2.21) defined $\kappa_s^{\min} = \min_{0 \le k \le K} \kappa_s^{(k)}$, we know that $\kappa_s^{\min} - \kappa_g < 0$. Combining this with the fact that $0 \leq \frac{\|\bar{\boldsymbol{q}}^{(k)}\|_4^4}{\|\bar{\boldsymbol{q}}^{(k)}\|_2^4} \leq 1$, we have

$$\begin{aligned}\sum_k (\kappa_s^{(k)} - \kappa_g)\|\bar{\boldsymbol{q}}^{(k)}\|_4^4 &\geq (\kappa_s^{\min} - \kappa_g)\sum_k \|\bar{\boldsymbol{q}}^{(k)}\|_4^4 \\ &\geq (\kappa_s^{\min} - \kappa_g)\sum_k \|\bar{\boldsymbol{q}}^{(k)}\|_2^4 \\ &= (\kappa_s^{\min} - \kappa_g)\|\bar{\boldsymbol{q}}\|_2^4.\end{aligned}$$

Thus, the following is a sufficient condition for (4.22):

$$0 \;>\; \Bigg(1 + \underbrace{\frac{\kappa_s^{\min} - \kappa_g}{\kappa_s^{(0)} - \kappa_g}}_{\rho_{\min}}\Bigg)\|\bar{\boldsymbol{q}}\|_2^4 - 2a^2\|\bar{\boldsymbol{q}}\|_2^2 + C(a, \boldsymbol{q}_{\mathrm{r},\nu}). \tag{4.23}$$

Because $1 + \rho_{\min}$ is positive, the set of $\{\|\bar{\boldsymbol{q}}\|_2^2\}$ that satisfy (4.23) is equivalent to the set of points $\{x\}$ that lie between the roots $\{x_1, x_2\}$ of the quadratic

$$P_a(x) \;=\; (1 + \rho_{\min})\, x^2 - 2a^2 x + C(a, \boldsymbol{q}_{\mathrm{r},\nu}).$$

Because $\bar{\boldsymbol{q}}$ is an interference response, not all values of $\|\bar{\boldsymbol{q}}\|_2$ are valid. As explained below, we only need to concern ourselves about $0 \leq \|\bar{\boldsymbol{q}}\|_2 < a\sqrt{2}^{-1}$. This implies that a valid upper bound on $b_*^2(a)$ from (4.3) is given by the smaller root of $P_a(x)$ when (i) this smaller root is non-negative real and (ii) the larger root of $P_a(x)$ is $\geq a^2/2$.

When both roots of $P_a(x)$ lie in the interval $[0, a^2/2)$, there exist two valid regions in the gain-a interference space with CM cost smaller than at the reference, i.e., the set $\{\bar{\boldsymbol{q}} : (a, \bar{\boldsymbol{q}}) \in \mathcal{Q}_{\mathrm{c}}(\beta_{\mathrm{r}}\boldsymbol{q}_{\mathrm{r},\nu})\}$ becomes disjoint. The "inner" part of this disjoint set allows UMSE bounding since it can be contained by $\{\bar{\boldsymbol{q}} : \|\bar{\boldsymbol{q}}\|_2 \leq b_1(a)\}$ for a positive interference radius $b_1(a)$, but the "outer" part of the set does *not* permit practical bounding. Such disjointness of $\mathcal{Q}_{\mathrm{c}}(\beta_{\mathrm{r}}\boldsymbol{q}_{\mathrm{r},\nu})$ arises from a source $k \neq 0$ such that $\kappa_s^{(k)} < \kappa_s^{(0)}$. In these scenarios, the point of lowest CM cost in the "outer" regions of $\{\bar{\boldsymbol{q}} : (a, \bar{\boldsymbol{q}}) \in \mathcal{Q}_{\mathrm{c}}(\beta_{\mathrm{r}}\boldsymbol{q}_{\mathrm{r},\nu})\}$ occurs at points on the boundary of $\mathcal{Q}_\nu^{(0)}$ of the form $\bar{\boldsymbol{q}} = (\ldots, 0, 0, ae^{j\theta}\sqrt{2}^{-1}, 0, 0, \ldots)^t$ and hence with $\|\bar{\boldsymbol{q}}\|_2^2 = a^2/2$. Thus, when $x_2 \geq a^2/2$, we can be assured that all valid interference responses (i.e., $\{\bar{\boldsymbol{q}} : (a, \bar{\boldsymbol{q}}) \in \mathcal{Q}_\nu^{(0)}\}$) with CM cost less than the reference can be bounded by some radius b_1.

Solving for the roots of $P_a(x)$ yields (with the convention $x_1 \leq x_2$)

$$\begin{aligned}\{x_1, x_2\} &= \frac{a^2 \pm \sqrt{a^4 - (\rho_{\min}+1)C(a, \boldsymbol{q}_{\mathrm{r},\nu})}}{\rho_{\min}+1} \\ &= a^2 \left(\frac{1 \pm \sqrt{1 - (\rho_{\min}+1)\frac{C(a,\boldsymbol{q}_{\mathrm{r},\nu})}{a^4}}}{\rho_{\min}+1} \right),\end{aligned}$$

and both roots are non-negative real when $0 \leq C(a, \boldsymbol{q}_{\mathrm{r},\nu})/a^4 \leq (\rho_{\min}+1)^{-1}$. It can be shown that $x_2 > a^2/2$ occurs when $C(a, \boldsymbol{q}_{\mathrm{r},\nu})/a^4 \leq (3-\rho_{\min})/4$. Since $\rho_{\min} \geq 1$ implies $(3-\rho_{\min})/4 \leq (\rho_{\min}+1)^{-1}$, both root requirements are satisfied when

$$0 \;\leq\; \frac{C(a, \boldsymbol{q}_{\mathrm{r},\nu})}{a^4} \;\leq\; \frac{3-\rho_{\min}}{4}. \tag{4.24}$$

4.A.4 Proof of Theorem 4.1

In this section we use the expression for $b_*(a)$ from (4.7) and a suitably chosen reference response $\boldsymbol{q}_{\mathrm{r},\nu} \in \mathcal{Q}_{\mathrm{a}} \cap \mathcal{Q}_\nu^{(0)}$ to upper bound $J_{\mathrm{u},\nu}(\boldsymbol{q}_{\mathrm{c},\nu})$. Plugging (4.7) in (4.4),

$$\begin{aligned}\frac{J_{\mathrm{u},\nu}(\boldsymbol{q}_{\mathrm{c},\nu})}{\sigma_s^2} \;&\leq\; \max_a \frac{1 - \sqrt{1 - (\rho_{\min}+1)\frac{C(a,\boldsymbol{q}_{\mathrm{r},\nu})}{a^4}}}{\rho_{\min} + \sqrt{1 - (\rho_{\min}+1)\frac{C(a,\boldsymbol{q}_{\mathrm{r},\nu})}{a^4}}} \\ &\quad \text{when} \quad 0 \leq \frac{C(a, \boldsymbol{q}_{\mathrm{r},\nu})}{a^4} \leq \frac{3-\rho_{\min}}{4}.\end{aligned} \tag{4.25}$$

Note that the fraction on the right of (4.25) is non-negative and strictly increasing in $C(a, \boldsymbol{q}_{\mathrm{r},\nu})/a^4$ over the valid range of $C(a, \boldsymbol{q}_{\mathrm{r},\nu})/a^4$. Hence, finding a that maximizes this expression can be accomplished by finding a that maximizes $C(a, \boldsymbol{q}_{\mathrm{r},\nu})/a^4$. To find these maxima, we first rewrite $C(a, \boldsymbol{q}_{\mathrm{r},\nu})/a^4$ from (4.22):

$$\frac{C(a, \boldsymbol{q}_{\mathrm{r},\nu})}{a^4} \;=\; C_1 \left(\frac{1}{2}(a^2)^{-2} - \kappa_{y_\mathrm{r}} \left(\frac{\gamma}{\sigma_s^2} \right)^{-1} (a^2)^{-1} + C_2 \right),$$

where C_1 and C_2 are independent of a. Computing the partial derivative with respect to the quantity a^2,

$$\frac{\partial}{\partial(a^2)} \left\{ \frac{C(a, \boldsymbol{q}_{\mathrm{r},\nu})}{a^4} \right\} \;=\; C_1\,(a^2)^{-3} \left(\kappa_{y_\mathrm{r}} \left(\frac{\gamma}{\sigma_s^2} \right)^{-1} a^2 - 1 \right).$$

Setting the partial derivative to zero yields the unique finite maximum

$$a_{\max}^2 \;=\; \left(\frac{\gamma}{\sigma_s^2} \right) \kappa_{y_\mathrm{r}}.$$

Plugging $a_{\max}^2$ into (4.22) gives the simple result

$$\frac{C(a_{\max}, \boldsymbol{q}_{\mathrm{r},\nu})}{a_{\max}^4} \;=\; \frac{\kappa_s^{(0)} - \kappa_{y_\mathrm{r}}}{\kappa_s^{(0)} - \kappa_g} \;=\; 1 - \frac{\kappa_g - \kappa_{y_\mathrm{r}}}{\kappa_g - \kappa_s^{(0)}},$$

and the $C(a, \boldsymbol{q}_{\mathrm{r},\nu})/a^4$ requirement (4.24) translates into

$$\frac{1 + \rho_{\min}}{4} \;\le\; \frac{\kappa_g - \kappa_{y_\mathrm{r}}}{\kappa_g - \kappa_s^{(0)}} \;\le\; 1. \tag{4.26}$$

Finally, plugging $C(a_{\max}, \boldsymbol{q}_{\mathrm{r},\nu})/a_{\max}^4$ into (4.25) gives

$$\frac{J_{\mathrm{u},\nu}(\boldsymbol{q}_{\mathrm{c},\nu})}{\sigma_s^2} \;\le\; \frac{1 - \sqrt{(1+\rho_{\min})\,\frac{\kappa_g - \kappa_{y_\mathrm{r}}}{\kappa_g - \kappa_s^{(0)}} - \rho_{\min}}}{\rho_{\min} + \sqrt{(1+\rho_{\min})\,\frac{\kappa_g - \kappa_{y_\mathrm{r}}}{\kappa_g - \kappa_s^{(0)}} - \rho_{\min}}}. \tag{4.27}$$

We now establish the existence of an attainable CM-minimizing global response associated with the desired user at delay ν, i.e., $\boldsymbol{q}_{\mathrm{c},\nu} \in \mathcal{Q}_\mathrm{a} \cap \mathcal{Q}_\nu^{(0)}$. For simplicity, we assume that the space $\mathbf{q}$ is finite dimensional. We will exploit the Weierstrass theorem [Luenberger Book 69, p. 40], which says that a continuous cost functional has a local minimum in a compact set if there exist points in the interior of the set which give cost lower than anywhere on the boundary.

By definition, all points in $\mathcal{Q}_{\rm c}(\beta_{\rm r}\boldsymbol{q}_{{\rm r},\nu})$ have CM cost less than or equal to $J_{\rm c}(y_{\rm r})$, the CM cost everywhere on the boundary of $\mathcal{Q}_{\rm c}(\beta_{\rm r}\boldsymbol{q}_{{\rm r},\nu})$. To make this inequality strict, we expand $\mathcal{Q}_{\rm c}(\beta_{\rm r}\boldsymbol{q}_{{\rm r},\nu})$ to form the new set $\mathcal{Q}'_{\rm c}(\beta_{\rm r}\boldsymbol{q}_{{\rm r},\nu})$, defined in terms of boundary cost $J_{\rm c}(y_{\rm r}) + \epsilon$ (for arbitrarily small $\epsilon > 0$). Thus, all points on the boundary of $\mathcal{Q}'_{\rm c}(\beta_{\rm r}\boldsymbol{q}_{{\rm r},\nu})$ will have CM cost strictly greater than $J_{\rm c}(y_{\rm r})$. But how do we know that such a set $\mathcal{Q}'_{\rm c}(\beta_{\rm r}\boldsymbol{q}_{{\rm r},\nu})$ exists? We simply need to reformulate (4.22) with ϵ-larger $J_{\rm c}(y_{\rm r})$, resulting in ϵ-larger $C(a, \boldsymbol{q}_{{\rm r},\nu})$ and a modified quadratic $P_a(x)$ in sufficient condition (4.23). As long as the new roots (call them x'_1 and x'_2) satisfy $x'_1 \in [0, a^2/2)$ and $x'_2 > a^2/2$, the set $\{\bar{\boldsymbol{q}} : (a, \bar{\boldsymbol{q}}) \in \mathcal{Q}'_{\rm c}(\beta_{\rm r}\boldsymbol{q}_{{\rm r},\nu})\}$ is well defined, and as long as this holds for the worst-case a (i.e., $a_{\max}$), $\mathcal{Q}'_{\rm c}(\beta_{\rm r}\boldsymbol{q}_{{\rm r},\nu})$ will itself be well defined. This behavior can be guaranteed, for arbitrarily small ϵ, by replacing (4.26) with the stricter condition

$$\frac{1+\rho_{\min}}{4} \;<\; \frac{\kappa_g - \kappa_{y_{\rm r}}}{\kappa_g - \kappa_s^{(0)}} \;\leq\; 1. \tag{4.28}$$

To summarize, (4.28) guarantees the existence of a closed and bounded set $\mathcal{Q}'_{\rm c}(\beta_{\rm r}\boldsymbol{q}_{{\rm r},\nu})$ containing an interior point $\beta_{\rm r}\boldsymbol{q}_{{\rm r},\nu}$ with CM cost strictly smaller than all points on the set boundary.

Due to attainibility requirements, our local minimum search must be constrained to the relative interior of the $\mathcal{Q}_{\rm a}$ manifold (which has been embedded in a possibly higher-dimensional $\boldsymbol{q}$-space). Can we apply the Weierstrass theorem on this manifold? First, we know the $\mathcal{Q}_{\rm a}$ manifold intersects $\mathcal{Q}'_{\rm c}(\beta_{\rm r}\mathbf{q}_{{\rm r},\nu})$, namely at the point $\beta_{\rm r}\boldsymbol{q}_{{\rm r},\nu}$. Second, we know that the relative boundary of the $\mathcal{Q}_{\rm a}$ manifold occurs outside $\mathcal{Q}'_{\rm c}(\beta_{\rm r}\mathbf{q}_{{\rm r},\nu})$, namely at infinity. These two observations imply that the boundary of $\mathcal{Q}_{\rm a} \cap \mathcal{Q}'_{\rm c}(\beta_{\rm r}\boldsymbol{q}_{{\rm r},\nu})$ relative to $\mathcal{Q}_{\rm a}$ must be a subset of the boundary of $\mathcal{Q}'_{\rm c}(\beta_{\rm r}\boldsymbol{q}_{{\rm r},\nu})$. Hence, the interior of $\mathcal{Q}_{\rm a} \cap \mathcal{Q}'_{\rm c}(\beta_{\rm r}\boldsymbol{q}_{{\rm r},\nu})$ relative to $\mathcal{Q}_{\rm a}$ contains points which give kurtosis strictly higher than those on the boundary of $\mathcal{Q}_{\rm a} \cap \mathcal{Q}'_{\rm c}(\beta_{\rm r}\boldsymbol{q}_{{\rm r},\nu})$ relative to $\mathcal{Q}_{\rm a}$.

Finally, the domain (i.e., $\mathcal{Q}_\mathrm{a} \cap \mathcal{Q}'_\mathrm{c}(\beta_\mathrm{r}\boldsymbol{q}_{\mathrm{r},\nu})$ relative to $\mathcal{Q}_\mathrm{a}$) is closed and bounded, hence compact. Thus the Weierstrass theorem ensures the existence of a local CM minimum in the interior of $\mathcal{Q}_\mathrm{a} \cap \mathcal{Q}'_\mathrm{c}(\beta_\mathrm{r}\boldsymbol{q}_{\mathrm{r},\nu})$ relative to $\mathcal{Q}_\mathrm{a}$ under (4.28). Recalling that $\mathcal{Q}'_\mathrm{c}(\beta_\mathrm{r}\boldsymbol{q}_{\mathrm{r},\nu}) \subset \mathcal{Q}^{(0)}_\nu$, we see that there exists an attainable locally CM-minimizing response associated with the desired user at delay ν.

Theorem 4.1 follows directly from (4.27) with reference choice $\boldsymbol{q}_{\mathrm{r},\nu} = \boldsymbol{q}_{\mathrm{m},\nu} \in \mathcal{Q}_\mathrm{a} \cap \mathcal{Q}^{(0)}_\nu$. Note that we restrict ourselves to $\boldsymbol{q}_{\mathrm{m},\nu} \in \mathcal{Q}^{(0)}_\nu$, which may not always be the case.

4.A.5 Proof of Theorem 4.2

In this section we find an upper bound for $J_{\mathrm{u},\nu}(\boldsymbol{q}_{\mathrm{c},\nu})$ that involves the UMSE of reference estimators, $J_{\mathrm{u},\nu}(\boldsymbol{q}_{\mathrm{r},\nu})$, rather than the kurtosis of reference estimates, κ_{y_r}. The choosing the reference to be the MMSE estimator can be considered a special case. The conditions we establish below will guarantee that $\boldsymbol{q}_{\mathrm{m},\nu} \in \mathcal{Q}^{(0)}_\nu$.

We will take advantage of the fact that $J_{\mathrm{u},\nu}\big|^{\max,\kappa_{y_\mathrm{r}}}_{\mathrm{c},\nu}$ in (4.27) is a strictly decreasing function of $\frac{\kappa_g - \kappa_{y_\mathrm{r}}}{\kappa_g - \kappa_s^{(0)}}$ over its valid range. From (4.19),

$$\frac{\kappa_g - \kappa_y}{\kappa_g - \kappa_s^{(0)}} = \frac{\sum_k (\kappa_g - \kappa_s^{(k)}) \|\boldsymbol{q}^{(k)}\|_4^4}{(\kappa_g - \kappa_s^{(0)}) \|\boldsymbol{q}\|_2^4} = \frac{(\kappa_g - \kappa_s^{(0)}) |q_\nu^{(0)}|^4 + \sum_k (\kappa_g - \kappa_s^{(k)}) \|\bar{\boldsymbol{q}}^{(k)}\|_4^4}{(\kappa_g - \kappa_s^{(0)}) \|\boldsymbol{q}\|_2^4}.$$

Examining the previous equation, $0 \le \frac{\|\bar{\boldsymbol{q}}\|_4^4}{\|\bar{\boldsymbol{q}}\|_2^4} \le 1$ implies that

$$\sum_k (\kappa_g - \kappa_s^{(k)}) \|\bar{\boldsymbol{q}}^{(k)}\|_4^4 \ \ge\ (\kappa_g - \kappa_s^{\max}) \|\bar{\boldsymbol{q}}\|_4^4 \ \ge\ \begin{cases} 0, & \kappa_s^{\max} \le \kappa_g \\ (\kappa_g - \kappa_s^{\max}) \|\bar{\boldsymbol{q}}\|_2^4, & \kappa_s^{\max} > \kappa_g \end{cases} \tag{4.29}$$

and

$$\sum_k (\kappa_g - \kappa_s^{(k)}) \|\bar{\boldsymbol{q}}^{(k)}\|_4^4 \ \le\ (\kappa_g - \kappa_s^{\min}) \|\bar{\boldsymbol{q}}\|_4^4 \ \le\ (\kappa_g - \kappa_s^{\min}) \|\bar{\boldsymbol{q}}\|_2^4. \tag{4.30}$$

Note that in (4.30) and the super-Gaussian case of (4.29), equality is reached by global responses of the form $\bar{\boldsymbol{q}} = \alpha \mathbf{e}_i^{(k)}$, where k corresponds to the source with minimum and maximum kurtosis, respectively.

Considering first the sub-Gaussian interference case ($\kappa_s^{\max} \leq \kappa_g$), we claim

$$\frac{\kappa_g - \kappa_y}{\kappa_g - \kappa_s^{(0)}} \;\geq\; \frac{|q_\nu^{(0)}|^4}{\|\boldsymbol{q}\|_2^4} \;=\; \left(1 + \frac{J_{\mathrm{u},\nu}(\boldsymbol{q})}{\sigma_s^2}\right)^{-2} \tag{4.31}$$

since the definition of $J_{\mathrm{u},\nu}(\boldsymbol{q})$ in (2.31) implies

$$\|\boldsymbol{q}\|_2^4 \;=\; \left(|q_\nu^{(0)}|^2 + \|\bar{\boldsymbol{q}}\|_2^2\right)^2 \;=\; |q_\nu^{(0)}|^4 \left(1 + \frac{\|\bar{\boldsymbol{q}}\|_2^2}{|q_\nu^{(0)}|^2}\right)^2 \;=\; |q_\nu^{(0)}|^4 \left(1 + \frac{J_{\mathrm{u},\nu}(\boldsymbol{q})}{\sigma_s^2}\right)^2.$$

Applying (4.31) to (4.27), we obtain

$$\frac{J_{\mathrm{u},\nu}\Big|_{\mathrm{c},\nu}^{\max,\kappa_{y\mathrm{r}}}}{\sigma_s^2} \;\leq\; \frac{1 - \sqrt{(1+\rho_{\min})\left(1 + \frac{J_{\mathrm{u},\nu}(\boldsymbol{q}_{\mathrm{r},\nu})}{\sigma_s^2}\right)^{-2} - \rho_{\min}}}{\rho_{\min} + \sqrt{(1+\rho_{\min})\left(1 + \frac{J_{\mathrm{u},\nu}(\boldsymbol{q}_{\mathrm{r},\nu})}{\sigma_s^2}\right)^{-2} - \rho_{\min}}}$$

when (4.28) is satisfied. Inequality (4.31) implies that

$$\frac{1+\rho_{\min}}{4} \;<\; \left(1 + \frac{J_{\mathrm{u},\nu}(\boldsymbol{q}_{\mathrm{r},\nu})}{\sigma_s^2}\right)^{-2} \quad\Leftrightarrow\quad \frac{J_{\mathrm{u},\nu}(\boldsymbol{q}_{\mathrm{r},\nu})}{\sigma_s^2} \;<\; -1 + \frac{2}{\sqrt{\rho_{\min}+1}}$$

is sufficient for the left inequality of (4.28). Turning our attention to the right inequality of (4.28), we can use (4.30) to say

$$\frac{\kappa_g - \kappa_y}{\kappa_g - \kappa_s^{(0)}} \;\leq\; \frac{|q_\nu^{(0)}|^4}{\|\boldsymbol{q}\|_2^4} + \rho_{\min}\frac{\|\bar{\boldsymbol{q}}\|_2^4}{\|\boldsymbol{q}\|_2^4} \;=\; \left(1 + \frac{J_{\mathrm{u},\nu}(\boldsymbol{q})}{\sigma_s^2}\right)^{-2} \left(1 + \rho_{\min}\frac{J_{\mathrm{u},\nu}^2(\boldsymbol{q})}{\sigma_s^4}\right) \tag{4.32}$$

since (2.31) implies

$$\frac{\|\bar{\boldsymbol{q}}\|_2^4}{\|\boldsymbol{q}\|_2^4} \;=\; \left(\frac{\|\bar{\boldsymbol{q}}\|_2^2 + |q_\nu^{(0)}|^2}{\|\bar{\boldsymbol{q}}\|_2^2}\right)^{-2} \;=\; \left(1 + \frac{\sigma_s^2}{J_{\mathrm{u},\nu}(\boldsymbol{q})}\right)^{-2} \;=\; \frac{J_{\mathrm{u},\nu}^2(\boldsymbol{q})}{\sigma_s^4}\left(1 + \frac{J_{\mathrm{u},\nu}(\boldsymbol{q})}{\sigma_s^2}\right)^{-2}.$$

Then, inequality (4.32) implies that a sufficient condition for the right side of equation (4.28) is

$$\left(1 + \frac{J_{\mathrm{u},\nu}(\boldsymbol{q}_{\mathrm{r},\nu})}{\sigma_s^2}\right)^{-2} \left(1 + \rho_{\min}\frac{J_{\mathrm{u},\nu}^2(\boldsymbol{q}_{\mathrm{r},\nu})}{\sigma_s^4}\right) \;\leq\; 1 \quad\Leftrightarrow\quad \frac{J_{\mathrm{u},\nu}(\boldsymbol{q}_{\mathrm{r},\nu})}{\sigma_s^2} \;\leq\; \frac{2}{\rho_{\min}-1}.$$

Using $\rho_{\min} \geq 1$, it can be shown that $-1 + 2/\sqrt{\rho_{\min}+1} \leq 2/(\rho_{\min}-1)$. Thus, satisfaction of our sufficient condition for the left inequality in (4.28) suffices for both inequalities in (4.28).

Treatment of the super-Gaussian interference case ($\kappa_s^{\max} > \kappa_g$) is analogous. With the methods used to obtain (4.32), equation (4.29) implies

$$\begin{aligned}\frac{\kappa_g - \kappa_y}{\kappa_g - \kappa_s^{(0)}} &\geq \frac{|q_\nu^{(0)}|^4}{\|\boldsymbol{q}\|_2^4} + \underbrace{\frac{\kappa_g - \kappa_s^{\max}}{\kappa_g - \kappa_s^{(0)}}}_{\rho_{\max}} \frac{\|\bar{\boldsymbol{q}}\|_2^4}{\|\boldsymbol{q}\|_2^4} \\ &= \left(1 + \frac{J_{\mathrm{u},\nu}(\boldsymbol{q})}{\sigma_s^2}\right)^{-2} \left(1 + \rho_{\max}\frac{J_{\mathrm{u},\nu}^2(\boldsymbol{q})}{\sigma_s^4}\right). \end{aligned} \tag{4.33}$$

Applying (4.33) to (4.27), we obtain

$$\frac{J_{\mathrm{u},\nu}\big|_{\mathrm{c},\nu}^{\max,\kappa_{y_\mathrm{r}}}}{\sigma_s^2} \leq \frac{1 - \sqrt{(1+\rho_{\min})\left(1 + \frac{J_{\mathrm{u},\nu}(\boldsymbol{q}_{\mathrm{r},\nu})}{\sigma_s^2}\right)^{-2}\left(1 + \rho_{\max}\frac{J_{\mathrm{u},\nu}^2(\boldsymbol{q}_{\mathrm{r},\nu})}{\sigma_s^4}\right) - \rho_{\min}}}{\rho_{\min} + \sqrt{(1+\rho_{\min})\left(1 + \frac{J_{\mathrm{u},\nu}(\boldsymbol{q}_{\mathrm{r},\nu})}{\sigma_s^2}\right)^{-2}\left(1 + \rho_{\max}\frac{J_{\mathrm{u},\nu}^2(\boldsymbol{q}_{\mathrm{r},\nu})}{\sigma_s^4}\right) - \rho_{\min}}}$$

as long as (4.28) is satisfied. Substituting $|q_\nu^{(0)}|^2 = \|\boldsymbol{q}\|_2^2 - \|\bar{\boldsymbol{q}}\|_2^2$ in (4.33), we find that

$$\frac{\kappa_g - \kappa_y}{\kappa_g - \kappa_s^{(0)}} \geq 1 - 2\frac{\|\bar{\boldsymbol{q}}\|_2^2}{\|\boldsymbol{q}\|_2^2} + (1+\rho_{\max})\frac{\|\bar{\boldsymbol{q}}\|_2^4}{\|\boldsymbol{q}\|_2^4},$$

hence a sufficient condition for the left inequality of (4.28) becomes $(1+\rho_{\min})/4 < 1 - 2\|\bar{\boldsymbol{q}}_{\mathrm{r},\nu}\|_2^2/\|\boldsymbol{q}_{\mathrm{r},\nu}\|_2^2 + (1+\rho_{\max})(\|\bar{\boldsymbol{q}}_{\mathrm{r},\nu}\|_2^2/\|\boldsymbol{q}_{\mathrm{r},\nu}\|_2^2)^2$, or equivalently

$$(1+\rho_{\max})\left(\frac{\|\bar{\boldsymbol{q}}_{\mathrm{r},\nu}\|_2^2}{\|\boldsymbol{q}_{\mathrm{r},\nu}\|_2^2}\right)^2 - 2\frac{\|\bar{\boldsymbol{q}}_{\mathrm{r},\nu}\|_2^2}{\|\boldsymbol{q}_{\mathrm{r},\nu}\|_2^2} + (3-\rho_{\min})/4 < 0.$$

It can be shown that the quadratic inequality above is satisfied by

$$\frac{\|\bar{\boldsymbol{q}}_{\mathrm{r},\nu}\|_2^2}{\|\boldsymbol{q}_{\mathrm{r},\nu}\|_2^2} < \begin{cases} \frac{1-\sqrt{1-(3-\rho_{\min})(1+\rho_{\max})/4}}{1+\rho_{\max}}, & \rho_{\max} \neq -1, \\ (3-\rho_{\min})/8, & \rho_{\max} = -1, \end{cases}$$

and since $J_{\mathrm{u},\nu}(\boldsymbol{q}) = \big(\|\bar{\boldsymbol{q}}\|_2^2/\|\boldsymbol{q}\|_2^2\big)/\big(1 - \|\bar{\boldsymbol{q}}\|_2^2/\|\boldsymbol{q}\|_2^2\big)$ is strictly increasing in $\|\bar{\boldsymbol{q}}\|_2^2\|\boldsymbol{q}\|_2^2$, the following must be sufficient for the left inequality of (4.28).

$$\frac{J_{\mathrm{u},\nu}(\boldsymbol{q}_{\mathrm{r},\nu})}{\sigma_s^2} \;<\; \begin{cases} \frac{1-\sqrt{1-(3-\rho_{\min})(1+\rho_{\max})/4}}{\rho_{\max}+\sqrt{1-(3-\rho_{\min})(1+\rho_{\max})/4}}, & \rho_{\max} \neq -1, \\ \frac{3-\rho_{\min}}{5+\rho_{\min}}, & \rho_{\max} = -1. \end{cases} \tag{4.34}$$

As for the right inequality of (4.28), it can be shown that the quantities in (4.34) are smaller than $2/(\rho_{\min} - 1)$. Thus, satisfaction of (4.34) suffices for both inequalities in (4.28).

4.A.6 Proof of Theorem 4.3

Here, we reformulate the upper bound (4.11). To simplify the presentation of the proof, the shorthand notation $J := J_{\mathrm{u},\nu}(\boldsymbol{q}_{\mathrm{m},\nu})/\sigma_s^2$ will be used.

Starting with the non-super-Gaussian case (i.e., $\kappa_s^{\max} \le \kappa_g$), (4.11) says

$$\frac{J_{\mathrm{u},\nu}\big|_{\mathrm{c},\nu}^{\max,J_{\mathrm{u},\nu}(\boldsymbol{q}_{\mathrm{m},\nu})}}{\sigma_s^2} \;=\; \frac{1-\sqrt{(1+\rho_{\min})(1+J)^{-2}-\rho_{\min}}}{\rho_{\min}+\sqrt{(1+\rho_{\min})(1+J)^{-2}-\rho_{\min}}},$$

from which routine manipulations yield

$$\frac{J_{\mathrm{u},\nu}\big|_{\mathrm{c},\nu}^{\max,J_{\mathrm{u},\nu}(\boldsymbol{q}_{\mathrm{m},\nu})}}{\sigma_s^2} \;=\; \frac{1-\sqrt{1-\big((\rho_{\min}-1)+\rho_{\min}(2J+J^2)\big)(2J+J^2)}}{(\rho_{\min}-1)+\rho_{\min}(2J+J^2)}.$$

For $x \in \mathbb{R}$ such that $|x| < 1$, the binomial series [Rudin Book 76] may be used to claim

$$\sqrt{1-x} \;=\; 1 - \frac{x}{2} - \frac{x^2}{8} - \mathcal{O}(x^3).$$

Applying the previous expression with $x = \big((\rho_{\min}-1)+\rho_{\min}(2J+J^2)\big)(2J+J^2)$, we find that

$$\begin{aligned}\frac{J_{\mathrm{u},\nu}\big|_{\mathrm{c},\nu}^{\max,J_{\mathrm{u},\nu}(\boldsymbol{q}_{\mathrm{m},\nu})}}{\sigma_s^2} &= \frac{1}{2}\frac{\big((\rho_{\min}-1)+\rho_{\min}(2J+J^2)\big)(2J+J^2)}{(\rho_{\min}-1)+\rho_{\min}(2J+J^2)} \\ &\quad +\frac{1}{8}\frac{\big((\rho_{\min}-1)+\rho_{\min}(2J+J^2)\big)^2(2J+J^2)^2}{(\rho_{\min}-1)+\rho_{\min}(2J+J^2)}+\mathcal{O}(J^3) \\ &= J+\frac{\rho_{\min}}{2}J^2+\mathcal{O}(J^3).\end{aligned}$$

Finally, subtraction of J gives the first case in (4.12).

For the super-Gaussian case (i.e., $\kappa_s^{\max} > \kappa_g$), (4.11) says

$$\frac{J_{\mathrm{u},\nu}\big|_{\mathrm{c},\nu}^{\max,J_{\mathrm{u},\nu}(\boldsymbol{q}_{\mathrm{m},\nu})}}{\sigma_s^2} = \frac{1-\sqrt{(1+\rho_{\min})(1+J)^{-2}(1+\rho_{\max}J^2)-\rho_{\min}}}{\rho_{\min}+\sqrt{(1+\rho_{\min})(1+J)^{-2}(1+\rho_{\max}J^2)-\rho_{\min}}},$$

from which routine manipulations yield

$$\frac{J_{\mathrm{u},\nu}\big|_{\mathrm{c},\nu}^{\max,J_{\mathrm{u},\nu}(\boldsymbol{q}_{\mathrm{m},\nu})}}{\sigma_s^2} = \frac{\rho_{\max}J^2+1-\sqrt{1+(2-2\rho_{\max})J+(1+\rho_{\max}+\rho_{\min}\rho_{\max}-5\rho_{\min})J^2+\mathcal{O}(J^3)}}{(\rho_{\min}-1)+2\rho_{\min}J+(\rho_{\min}-\rho_{\max})J^2}.$$

As before, we use the binomial series expansion for $\sqrt{1-x}$, but now with $x = (2\rho_{\max}-2)J+(5\rho_{\min}-1-\rho_{\max}-\rho_{\min}\rho_{\max})J^2+\mathcal{O}(J^3)$. After some algebra, we find

$$\frac{J_{\mathrm{u},\nu}\big|_{\mathrm{c},\nu}^{\max,J_{\mathrm{u},\nu}(\boldsymbol{q}_{\mathrm{m},\nu})}}{\sigma_s^2} = J+\frac{1}{2}\frac{(\rho_{\min}-\rho_{\max})(\rho_{\min}-1)J^2+\mathcal{O}(J^3)}{(\rho_{\min}-1)+2\rho_{\min}J+(\rho_{\min}-\rho_{\max})J^2}$$

Finally we apply the series approximation

$$\frac{1}{1-y} = 1+y+\mathcal{O}(y^2)$$

with $y = -(2\rho_{\min}J+(\rho_{\min}-\rho_{\max})J^2)/(\rho_{\min}-1)$ for $\rho_{\min}\neq 1$. Straightforward algebra yields

$$\frac{J_{\mathrm{u},\nu}\big|_{\mathrm{c},\nu}^{\max,J_{\mathrm{u},\nu}(\boldsymbol{q}_{\mathrm{m},\nu})}}{\sigma_s^2} = J+\frac{1}{2}(\rho_{\min}-\rho_{\max})J^2+\mathcal{O}(J^3).$$

Taking the limit $\rho_{\min}\to 1$, it is evident that no problems arise at the point $\rho_{\min}=1$. Subtraction of J from the last statement gives the second case in (4.12).

Chapter 5

Sufficient Conditions for the Local Convergence of CM Algorithms[1]

5.1 Introduction

Perhaps the greatest challenge facing successful application of the CM criterion in arbitrary interference environments results from the difficulty in determining CM-minimizing estimates of the desired source (as opposed to mistakenly estimating an interferer). The potential for "interference capture" is a direct consequence of the fact that the CM criterion exhibits multiple local minima in the estimator parameter space, each corresponding to a CM estimator of a particular source at a particular delay. Such multi-modality might be suspected from (2.35); the CM criterion is based on a particular property of the estimates $\{y_n\}$, and one can imagine a case in which this property is satisfied to a similar extent by, e.g., $\{y_n\} \approx \{s_n^{(0)}\}$ and $\{y_n\} \approx \{s_n^{(1)}\}$ when $\{s_n^{(0)}\}$ and $\{s_n^{(1)}\}$ have the same statistics.

Various "multiuser" modifications of the CM criterion have been proposed to

[1]The main results of this chapter also appear in the manuscript [Schniter TSP 00].

jointly estimate all sub-Gaussian sources present in a multi-source environment. Some of these techniques add a non-negative term to the CM criterion which penalizes correlation between any pair of L parallel estimator outputs, forcing the L estimators to generate estimates of distinct sources [Batra GLOBE 95], [Papadias SPL 96], [Touzni ICASSP 98]. Other techniques use the CM criterion in a successive interference cancellation scheme, whereby estimates of the n strongest sub-Gaussian sources are used to remove their respective contributions to the received signal before estimation of the $n+1^{th}$ source [Treichler TASSP 85b]. Both of these approaches, however, require knowledge of the number of (non-Gaussian) sources, result in significant increase in computational complexity when the number of sources is large, and generate estimates with questionable MSE performance. Instead, we focus on the well-known standard CM (or "Godard" [Godard TCOM 80]) criterion and consider desired-source convergence as an outcome of proper initialization.

Closed-form expressions for CM estimators do not generally exist, and thus gradient descent (GD) methods provide the typical means of solving for these estimators. Because exact gradient descent requires statistical knowledge of the received process that is not usually available in practical situations, *stochastic* GD algorithms such as CMA are used to estimate and track the (possibly time-varying) CM estimator. It is widely accepted, however, that small step-size stochastic GD algorithms exhibit mean transient and steady-state behaviors very close to those of exact GD under typical operating conditions [Benveniste Book 90], [Ljung Book 99]. Hence, we circumvent the details of stochastic adaptation by restricting our attention to (exact) GD minimization of the CM cost. An important property of GD minimization is that the location of algorithm initialization completely determines the stationary point to which the GD trajectory will eventually converge. The description of the CM-GD regions-of-convergence (ROC) in terms of estimator parameters appears to be a very difficult problem, however, and attempts at finding closed-form expressions for the ROC boundaries have thus far been unsuccessful [Chung Thesis 99], [Gu TSP 99].

In this chapter, we derive three sufficient conditions under which CM-GD minimization will generate an estimator for the desired source. The conditions are expressed in terms of statistical properties of the initial estimates, specifically, CM cost, kurtosis, and signal to interference-plus-noise ratio (SINR). Earlier attempts at describing the interference capture or "local convergence" properties of CMA have been made by Treichler and Larimore in [Treichler TASSP 85a] and Li and Ding in [Li TSP 95]. Treichler and Larimore constructed a simplifying approximation to the mean behavior of CMA for the case of a constant envelope signal in tonal interference and inferred the roles of initial signal-to-interference ratio and initial estimator parameterization on desired convergence. Li and Ding derived a sufficient kurtosis condition for the local convergence of the Shalvi-Weinstein (SW) algorithm [Shalvi TIT 90] and suggested that the condition applies to small-stepsize CMA as well. Our analysis and simulations suggest that the local convergence behavior of CMA differs from that of the SW algorithm, contrasting certain claims of [Li TSP 95].

The organization of the chapter is as follows. Section 5.2 derives initialization conditions sufficient for CM-GD convergence to desired source estimates, Section 5.3 discusses the implications of these conditions on choice of CM-GD initialization scheme, Section 5.4 presents numerical simulations verifying our analyses, and Section 5.5 concludes the chapter.

5.2 Sufficient Conditions for Local Convergence of CM-GD

5.2.1 The Main Idea

The set of global responses associated with the desired source ($k = 0$) at estimation delay ν will be denoted $\mathcal{Q}_\nu^{(0)}$ and defined as follows.

$$\mathcal{Q}_\nu^{(0)} \;:=\; \left\{ \boldsymbol{q} \;\text{ s.t. }\; |q_\nu^{(0)}| > \max_{(k,\delta)\neq(0,\nu)} |q_\delta^{(k)}| \right\}. \tag{5.1}$$

Note that under S1)–S3), the previous definition associates an estimator with a particular {source, delay} combination if and only if that {source, delay} contributes more energy to the estimate than any other {source, delay}. Choosing, as a reference set, the global responses on the boundary of $\mathcal{Q}_\nu^{(0)}$ with minimum CM cost,

$$\{\boldsymbol{q}_{\mathrm{r}}\} \;:=\; \arg \min_{\boldsymbol{q}\in \mathrm{bndr}(\mathcal{Q}_\nu^{(0)})} J_{\mathrm{c}}(\boldsymbol{q}), \tag{5.2}$$

we will denote the set of all global responses in $\mathcal{Q}_\nu^{(0)}$ with CM cost no higher than $J_{\mathrm{c}}(\boldsymbol{q}_{\mathrm{r}})$ by

$$\mathcal{Q}_{\mathrm{c}}(\boldsymbol{q}_{\mathrm{r}}) \;:=\; \big\{ \boldsymbol{q} \;\text{ s.t. }\; J_{\mathrm{c}}(\boldsymbol{q}) \leq J_{\mathrm{c}}(\boldsymbol{q}_{\mathrm{r}}) \big\} \cap \mathcal{Q}_\nu^{(0)}.$$

The main idea is this. Since all points in a CM gradient descent (CM-GD) trajectory have CM cost less than or equal to the cost at initialization, a CM-GD trajectory initialized within $\mathcal{Q}_{\mathrm{c}}(\boldsymbol{q}_{\mathrm{r}})$ must be entirely contained in $\mathcal{Q}_{\mathrm{c}}(\boldsymbol{q}_{\mathrm{r}})$ and thus in $\mathcal{Q}_\nu^{(0)}$. In other words, when a particular response $\boldsymbol{q}$ yields sufficiently small CM cost, CM-GD initialized from $\boldsymbol{q}$ will preserve the {source, delay} combination associated with $\boldsymbol{q}$. Note that initializing within $\mathcal{Q}_{\mathrm{c}}(\boldsymbol{q}_{\mathrm{r}})$ is sufficient, but not necessary, for eventual CM-GD convergence to a stationary point in $\mathcal{Q}_\nu^{(0)}$.

Since the size and shape of $\mathcal{Q}_\mathrm{c}(\boldsymbol{q}_\mathrm{r})$ are not easily characterizable, we find it more useful to derive sufficient CM-GD initialization conditions in terms of well-known statistical quantities such as kurtosis and SINR. It was shown in Section 2.5 that CM cost and kurtosis are closely related, and we shall see that translation between these two quantities is relatively straightforward. Translation of the initial CM-cost condition into an initial SINR condition is more difficult but can be accomplished through definition of $\mathrm{SINR}_{\min,\nu}$, the SINR above which all $\boldsymbol{q}$ have scaled versions in $\mathcal{Q}_\mathrm{c}(\boldsymbol{q}_\mathrm{r})$:

$$\mathrm{SINR}_{\min,\nu} \;:=\; \min x \text{ s.t. } \left\{ \forall \boldsymbol{q} : \mathrm{SINR}_\nu(\boldsymbol{q}) \geq x,\; \exists\, a_* \text{ s.t. } \frac{a_* \boldsymbol{q}}{\|\boldsymbol{q}\|_2} \in \mathcal{Q}_\mathrm{c}(\boldsymbol{q}_\mathrm{r}) \right\}. \tag{5.3}$$

If initializations in the set $\{\boldsymbol{q} : \mathrm{SINR}_\nu(\boldsymbol{q}) \geq \mathrm{SINR}_{\min,\nu}\}$ are scaled so that they lie within $\mathcal{Q}_\mathrm{c}(\boldsymbol{q}_\mathrm{r})$, the resulting CM-GD trajectories will remain within $\mathcal{Q}_\mathrm{c}(\boldsymbol{q}_\mathrm{r})$ and hence within $\mathcal{Q}_\nu^{(0)}$. In other words, when a particular response $\boldsymbol{q}$ yields sufficiently high SINR, CM-GD initialized from a properly scaled version of $\boldsymbol{q}$ will preserve the source/delay combination associated with $\boldsymbol{q}$. This sufficient SINR property is formalized below.

Since $\mathcal{Q}_\mathrm{c}(\boldsymbol{q}_\mathrm{r})$ and $\mathrm{SINR}_\nu(\boldsymbol{q})$ are all invariant to phase rotation (i.e., scalar multiplication by $e^{j\phi}$ for $\phi \in \mathbb{R}$) of $\boldsymbol{q}_\mathrm{r}$ and $\boldsymbol{q}$, respectively, we can (w.l.o.g.) restrict our attention to the "de-rotated" set of global responses $\{\boldsymbol{q} \text{ s.t. } q_\nu^{(0)} \in \mathbb{R}^+\}$. Such $\boldsymbol{q}$ allow parameterization in terms of gain $a = \|\boldsymbol{q}\|_2$ and interference response $\bar{\boldsymbol{q}}$ (defined in Section 2.3.3) where $\|\bar{\boldsymbol{q}}\|_2 \leq a$. In terms of the pair $(a, \bar{\boldsymbol{q}})$, the SINR (2.32) can be written

$$\mathrm{SINR}_\nu(a, \bar{\boldsymbol{q}}) = \frac{a^2 - \|\bar{\boldsymbol{q}}\|_2^2}{\|\bar{\boldsymbol{q}}\|_2^2},$$

so that (5.3) becomes

$$\begin{aligned}\mathrm{SINR}_{\min,\nu} &:= \min x \\ &\text{s.t. } \left\{\forall(a,\bar{\boldsymbol{q}}) : \frac{a^2 - \|\bar{\boldsymbol{q}}\|_2^2}{\|\bar{\boldsymbol{q}}\|_2^2} \geq x,\ \exists\, a_* \text{ s.t. } (a_*, \tfrac{a_*}{a}\bar{\boldsymbol{q}}) \in \mathcal{Q}_\mathrm{c}(\boldsymbol{q}_\mathrm{r})\right\}. \end{aligned} \tag{5.4}$$

Under particular conditions on a and $\boldsymbol{q}_\mathrm{r}$ (made explicit in Section 5.2.2), there exists a maximum interference gain b, specified as a function of system gain a, below which all $\bar{\boldsymbol{q}}$ are contained in $\mathcal{Q}_\mathrm{c}(\boldsymbol{q}_\mathrm{r})$:

$$b_{\max}(a) \ := \ \max b(a) \text{ s.t. } \left\{\forall\, \bar{\boldsymbol{q}} : \|\bar{\boldsymbol{q}}\|_2 \leq b(a),\ (a,\bar{\boldsymbol{q}}) \in \mathcal{Q}_\mathrm{c}(\boldsymbol{q}_\mathrm{r})\right\}. \tag{5.5}$$

For an illustration of a, $b_{\max}(a)$, and $\mathcal{Q}_\mathrm{c}(\boldsymbol{q}_\mathrm{r})$, see Fig. 5.1. Now, consider the quantity

$$\mathrm{SINR}_\nu(a, b_{\max}) \ := \ \frac{a^2 - b_{\max}^2(a)}{b_{\max}^2(a)}.$$

Note that $\mathrm{SINR}(a, b_{\max}) = \cot^2(\theta)$ for the θ in Fig. 5.1. Since $\mathrm{SINR}_\nu(a, b_{\max})$ is a decreasing function of $b_{\max}(a)$ (over its valid domain), definition (5.5) implies that

$$\frac{a^2 - \|\bar{\boldsymbol{q}}\|_2^2}{\|\bar{\boldsymbol{q}}\|_2^2} \geq \mathrm{SINR}_\nu(a, b_{\max}) \quad \Rightarrow \quad (a,\bar{\boldsymbol{q}}) \in \mathcal{Q}_\mathrm{c}(\boldsymbol{q}_\mathrm{r}).$$

Using the previous expression to minimize SINR in accordance with (5.4) yields the key quantities defined in (5.3):

$$\mathrm{SINR}_{\min,\nu} \ = \ \min_a \mathrm{SINR}_\nu(a, b_{\max}) \tag{5.6}$$

$$a_* \ = \ \arg\min_a \mathrm{SINR}_\nu(a, b_{\max}). \tag{5.7}$$

To summarize, when $\mathrm{SINR}_\nu(\boldsymbol{q}) \geq \mathrm{SINR}_{\min,\nu}$ and $\|\boldsymbol{q}\|_2 = a_*$, CM-GD initialized from $\boldsymbol{q}$ will preserve the {source, delay} combination associated with $\boldsymbol{q}$.

5.2.2 Derivation of Sufficient Conditions

In this section we formalize the previously-described initialization conditions for CM-GD local convergence. The main steps in the derivation are presented as theorems and lemmas, with proofs appearing in Appendix 5.A.

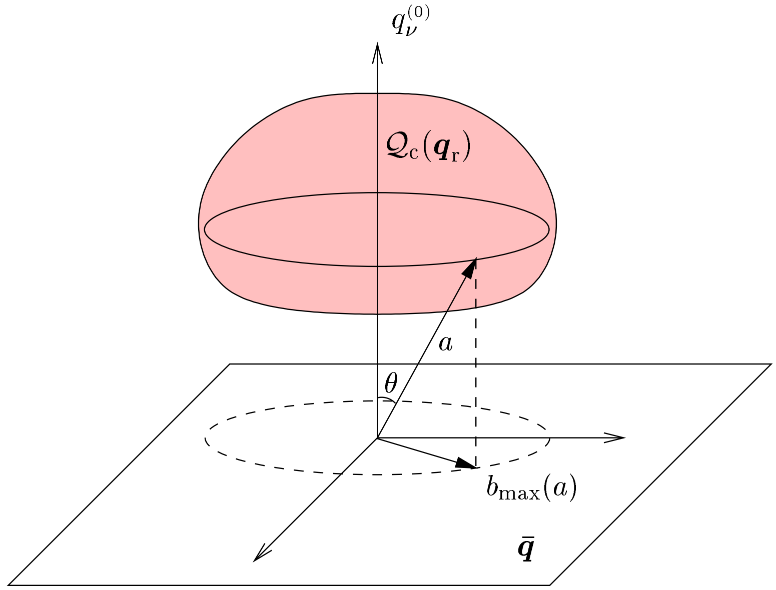

Figure 5.1: Illustration of maximum interference gain $b_{\max}(a)$ below which all global responses with gain a are contained in the CM cost region $\mathcal{Q}_c(\boldsymbol{q}_r)$.

Before presentation of the results, we introduce the following notation that is specific to this chapter.

$$\kappa_s^{\text{imin}} := \begin{cases} \min_{0 \le k \le K} \kappa_s^{(k)}, & \dim(\boldsymbol{q}^{(0)}) > 1 \\ \min_{1 \le k \le K} \kappa_s^{(k)}, & \dim(\boldsymbol{q}^{(0)}) = 1 \end{cases} \tag{5.8}$$

$$\rho_{\text{imin}} := \frac{\kappa_g - \kappa_s^{\text{imin}}}{\kappa_g - \kappa_s^{(0)}}, \tag{5.9}$$

$$\sigma_y^2\big|_{\text{crit}} := \gamma \left(\frac{4}{\kappa_s^{(0)} + \kappa_s^{\text{imin}} + 2\kappa_g} \right). \tag{5.10}$$

Above, "dim" denotes the dimension of a vector. Note that the second case in (5.8) applies only when the desired source contributes zero intersymbol interference (ISI). Thus, κ_s^{imin} equals κ_s^{min} in most cases.

Lemma 5.1. *The minimum CM cost on the boundary of $\mathcal{Q}_\nu^{(0)}$ is*

$$J_c(\boldsymbol{q}_r) = \min_{\boldsymbol{q} \in \text{bndr}(\mathcal{Q}_\nu^{(0)})} J_c(\boldsymbol{q}) = \gamma^2 \left(1 - \frac{4}{\kappa_s^{(0)} + \kappa_s^{\text{imin}} + 2\kappa_g} \right). \tag{5.11}$$

Theorem 5.1. *If* $\{y_n\}$ *are initial estimates of the desired source at delay* ν *(i.e.,* $y_n = \boldsymbol{q}_{\text{init}}^t \boldsymbol{s}(n)$ *for* $\boldsymbol{q}_{\text{init}} \in \mathcal{Q}_\nu^{(0)} \cap \mathcal{Q}_{\text{a}}$*) with CM cost*

$$J_{\text{c}}(y_n) \;<\; \gamma^2 \left(1 - \frac{4}{\kappa_s^{(0)} + \kappa_s^{\text{imin}} + 2\kappa_g}\right), \tag{5.12}$$

then estimators resulting from subsequent CM-minimizing gradient descent will also yield estimates of the desired source at delay ν*.*

Theorem 5.2. *If* $\{y_n\}$ *are initial estimates of the desired source at delay* ν *(i.e.,* $y_n = \boldsymbol{q}_{\text{init}}^t \boldsymbol{s}(n)$ *for* $\boldsymbol{q}_{\text{init}} \in \mathcal{Q}_\nu^{(0)} \cap \mathcal{Q}_{\text{a}}$*) with variance* $\sigma_y^2 = \sigma_y^2\big|_{crit}$ *and normalized kurtosis*

$$\kappa_y \;<\; \kappa_y^{\text{crit}} \;:=\; \frac{1}{4}\big(\kappa_s^{(0)} + \kappa_s^{\text{imin}} + 2\kappa_g\big), \tag{5.13}$$

then estimators resulting from subsequent CM-minimizing gradient descent will also yield estimates of the desired source at delay ν*.*

Theorem 5.3. *If* $\kappa_s^{(0)} \;\le\; (\kappa_s^{\text{imin}} + 2\kappa_g)/3$*, and if* $\{y_n\}$ *are initial estimates with variance* $\sigma_y^2 = \sigma_y^2\big|_{crit}$ *and* $\text{SINR}_\nu(y_n) > \text{SINR}_{\text{min},\nu}$*, where*

$$\text{SINR}_{\text{min},\nu} \;=\; \begin{cases} \frac{\sqrt{1+\rho_{\text{imin}}}}{2-\sqrt{1+\rho_{\text{imin}}}}, & \kappa_s^{\max} \le \kappa_g, \\[2ex] \left\{ \begin{array}{ll} \frac{\rho_{\max}+\sqrt{1-(1+\rho_{\max})(3-\rho_{\text{imin}})/4}}{1-\sqrt{1-(1+\rho_{\max})(3-\rho_{\text{imin}})/4}}, & \rho_{\max} \ne -1 \\[1ex] \frac{5+\rho_{\text{imin}}}{3-\rho_{\text{imin}}}, & \rho_{\max} = -1 \end{array} \right\}, & \kappa_s^{\max} > \kappa_g, \end{cases} \tag{5.14}$$

then estimators resulting from subsequent CM-minimizing gradient descent will also yield estimates of the desired source at delay ν*.*

We now make a few comments on the theorems. First, notice the presence of gain condition $\sigma_y^2 = \sigma_y^2\big|_{\text{crit}}$ in Theorems 5.2 and 5.3. One might wonder how necessary this gain condition is, and our answer is twofold. As we have derived *sufficient* conditions for convergence, none are strictly necessary. But it *is* possible to construct scenarios where satisfaction of all *except* the gain condition results in misconvergence. The existence of such scenarios implies that initial kurtosis cannot be the sole indicator of CM-GD convergence, as suggested in [Li TSP 95]. Fortunately, it appears that such scenarios are quite rare unless σ_y^2 is far from $\sigma_y^2\big|_{\text{crit}}$ or unless the SINR and/or kurtosis conditions are themselves near violation. Thus, in practice, successful CM-GD convergence should be quite robust to small violations of $\sigma_y^2 = \sigma_y^2\big|_{\text{crit}}$.

Finally, it should be pointed out that the relatively complicated expressions in Theorem 5.3 simplify under the operating conditions commonly encountered in, e.g., data communication. When the sources of interference are non-super-Gaussian (i.e., $\kappa_s^{\max} \leq \kappa_g$) and none have kurtosis less than the desired source (i.e., $\kappa_s^{(0)} \leq \kappa_s^{(k)}$), we find that $\rho_{\min} = 1$, and thus $\text{SINR}_{\min,\nu} = 1 + \sqrt{2}$ or 3.8 dB.

5.3 Implications for CM Initialization Schemes

In the previous section we have shown that there exist statistical properties of initial estimates which guarantee that subsequent CM gradient descent will produce an estimator of the same source at the same delay. In this section we suggest how one might satisfy these initialization conditions.

We consider CM initialization procedures capable of being described by the following two-step procedure: 1) design of one or more initialization hypotheses, 2) choice among hypotheses. Note that most popular CM initialization procedures, such as the single-spike scheme discussed below, fall within this general framework.

In evaluating a CM-GD initialization scheme, we must then consider the difficulty in both the design and evaluation of initialization hypotheses. The theorems in the previous section suggest that when a particular source or delay is desired, initialization hypotheses should be designed to either (i) maximize SINR, or (ii) minimize CM cost or kurtosis *when the initial estimates are known to correspond to a desired source/delay combination.*

5.3.1 The "Single-Spike" Initialization

The so-called single-spike initialization, first proposed in [Godard TCOM 80], is quite popular in single-user environments. Single-spike initializations for single-sensor baud-spaced equalizers (i.e., $P = 1$) are characterized by impulse responses with a single nonzero coefficient, i.e., $\mathbf{f}(z) = z^{-\delta}$. There exists a straightforward extension to multi-rate/multi-channel (i.e., $P > 1$) estimators: $\mathbf{f}(z) = \mathbf{1}z^{-\delta}$ for $\mathbf{1} := (\sqrt{1/P}, \ldots, \sqrt{1/P})^t \in \mathbb{R}^P$. For $P = 2$, this has been called the "double-spike" initialization [Johnson Chap 99]. The spike position is often an important design parameter, as we explain below.

Since the spike method yields an initial global response equaling (a delayed version of) the channel response, the kurtosis of the initial estimates can be expressed directly in terms of the channel coefficients $\{\mathbf{h}_i^{(k)}\}$:

$$\kappa_y \;=\; \sum_k (\kappa_s^{(k)} - \kappa_g) \frac{\|\boldsymbol{q}^{(k)}\|_4^4}{\|\boldsymbol{q}\|_2^4} + \kappa_g \;=\; \frac{\sum_{k,i} |\mathbf{1}^t \mathbf{h}_i^{(k)}|^4 (\kappa_s^{(k)} - \kappa_g)}{\left(\sum_{k,i} |\mathbf{1}^t \mathbf{h}_i^{(k)}|^2\right)^2} + \kappa_g.$$

If we assume a single sub-Gaussian user in the presence of additive white Gaussian noise (AWGN) of variance σ_w^2 at each sensor, the previous expression simplifies to

$$\kappa_y \;=\; (\kappa_s^{(0)} - \kappa_g) \frac{\sum_i |\mathbf{1}^t \mathbf{h}_i^{(0)}|^4}{\left(\sum_i |\mathbf{1}^t \mathbf{h}_i^{(0)}|^2 + \sigma_w^2/\sigma_s^2\right)^2} + \kappa_g. \tag{5.15}$$

Table 5.1 shows initial kurtoses κ_y from (5.15) for Signal Processing Information Base (SPIB) microwave channel models in AWGN (resulting in 20dB SNR at channel output), along with the critical kurtosis κ_y^{crit} from (5.13). From Table 5.1 we see that the single-spike initialization procedure generates estimates with kurtosis less

Table 5.1: Single-spike kurtoses for SPIB microwave channel models and 20 dB SNR.

Channel #	1	2	3	4	5	6	7	8	9	κ_y^{crit}
BPSK	1.17	1.17	1.41	1.98	1.94	1.76	1.16	1.69	1.70	2
8-PAM	1.87	1.86	2.02	2.37	2.34	2.23	1.86	2.19	2.19	2.38
QPSK	1.09	1.15	1.26	1.49	1.47	1.38	1.24	1.35	1.35	1.5
64-QAM	1.43	1.48	1.53	1.68	1.67	1.62	1.53	1.60	1.60	1.69

than the critical value for all SPIB channels. The implication is that *CM gradient descent from a single-spike initialization with magnitude chosen in accordance with Theorem 5.2 typically preserves the estimation delay of the initial estimates.* Similar conjectures have been made in [Li TSP 95] and [Johnson Chap 99].

Since MMSE performance is known to vary (significantly) with estimation delay, the connection between Wiener and CM performance established in Chapter 4 implies that the MSE performance of CM-minimizing estimates should also vary with estimation delay. Thus, from our observations on the local convergence of single-spike initializations, we conclude that the asymptotic MSE performance CM-GD estimates can be directly linked to the choice of initial spike delay.

5.3.2 Initialization Using Partial Information

Though the single-spike scheme has desirable properties in (noisy) single-user applications, one would not expect it to yield reliable estimates of the desired source when in the presence of significant sub-Gaussian interference since single-spike initialized CM-GD might lock onto a sub-Gaussian interferer instead of the desired source. With partial knowledge of the desired user's channel, however, it may be possible to construct rough guesses of the desired estimator that are good enough for use as CM-GD initializations. Then, if the initialization satisfies the sufficient conditions in the previous section, we know that CM-GD can be used to design an an estimator with nearly optimal MSE performance (as discussed in Section 2.5). The "partial knowledge" may come in various forms, for example, short training records in semi-blind equalization applications, rough direction-of-arrival knowledge in array applications, spreading sequences in code-division multiple access (CDMA) applications, or desired polarization angle in cross-pole interference cancellation.

We have seen that various criteria could be used to design and evaluate initialization hypotheses. Since reliable evaluation of higher-order statistics typically require a larger sample size than second-order statistics, the design and/or evaluation of SINR-based initializations might be advantageous when sample size is an issue. For this reason, SINR-based methods will be considered for the remainder of this section. Still, good results have been reported for kurtosis-based CM-GD initialization schemes for CDMA applications when sample size is not an issue [Schniter ALL 98].

The SINR-maximizing linear estimator is given by Wiener estimator $\mathbf{f}_{\mathrm{m},\nu}(z)$. It was shown in Appendix 2.A that

$$\mathbf{f}_{\mathrm{m},\nu}(z) \;=\; z^{-\nu}\sigma_s^2\Big(\mathrm{E}\left\{\mathbf{r}(z)\mathbf{r}^H(\tfrac{1}{z^*})\right\}\Big)^{\dagger}\mathbf{h}^{(0)}(\tfrac{1}{z^*}) \tag{5.16}$$

where $(\cdot)^{\dagger}$ denotes pseudo-inverse. As evident from (5.16), design of $\mathbf{f}_{\mathrm{m},\nu}(z)$ requires knowledge of the desired channel $\mathbf{h}^{(0)}(z)$ in addition to the auto-correlation of the received signal. Though various methods exist for the design of blind SINR-maximizing (i.e., MSE-minimizing) estimators based on partial knowledge of $\mathbf{h}^{(0)}(z)$, the Wiener expression (5.16) suggests the following CM initialization when given only a channel estimate $\hat{\mathbf{h}}^{(0)}(z)$ and knowledge of $\mathrm{E}\left\{\mathbf{r}(z)\mathbf{r}^H(1/z^*)\right\}$.

$$\mathbf{f}_{\text{init}}(z) \;=\; z^{-\nu}\Big(\mathrm{E}\left\{\mathbf{r}(z)\mathbf{r}^H(\tfrac{1}{z^*})\right\}\Big)^{\dagger}\hat{\mathbf{h}}^{(0)}(\tfrac{1}{z^*}). \tag{5.17}$$

Note that (5.17) may require additional scaling to satisfy the σ_y^2-requirements of Theorems 5.2 and 5.3.

5.4 Numerical Examples

In Fig. 5.2, CM-GD minimization trajectories conducted in estimator space are plotted in global (i.e. channel-plus-estimator) space ($\boldsymbol{q} \in \mathbb{R}^2$) to demonstrate the key results of this chapter. CM-GD can be described by the update equation $\boldsymbol{f}(n+1) = \boldsymbol{f}(n) - \mu\nabla_{\boldsymbol{f}} J_{\mathrm{c}}$, where $\boldsymbol{f} = (\ldots, \mathbf{f}_{-1}^t, \mathbf{f}_0^t, \mathbf{f}_1^t, \ldots)^t$ is a vector containing the estimator parameter coefficients, μ is a vanishingly small positive stepsize, and $\nabla_{\boldsymbol{f}}$ denotes the gradient with respect to $\boldsymbol{f}$. When the estimator is FIR we can write $\boldsymbol{q} = (\boldsymbol{f}^H\boldsymbol{\mathcal{H}})^t$, implying the global-response CM-GD update equation $\boldsymbol{q}(n+1) = \boldsymbol{q}(n) - \mu\boldsymbol{\mathcal{H}}^t\big(\nabla_{\boldsymbol{f}} J_{\mathrm{c}}\big)^*$. In all experiments, we use a two-parameter estimator and a FIR channel that corresponds to the following arbitrarily-chosen channel matrix $\boldsymbol{\mathcal{H}}$ (having condition number 3):

$$\boldsymbol{\mathcal{H}} = \begin{pmatrix} 0.2940 & -0.0596 \\ 0.1987 & 0.9801 \end{pmatrix}.$$

Fig. 5.2(a)-(c) depict the $\mathcal{Q}_\nu^{(0)}$ boundaries as dash-dotted lines, the $\mathrm{SINR}_{\min,\nu}$ boundaries as dashed lines, and the κ_y^{crit} boundaries as dotted lines. Note that in Fig. 5.2(a), the dash-dotted and dotted lines are coincident, while in Fig. 5.2(c), the dash-dotted, dashed, and dotted lines are coincident. The three sub-plots in Fig. 5.2 differ only in the kurtosis of the desired source: in Fig. 5.2(a) the sources have $\kappa_s^{(0)} = 1$ and $\kappa_s^{(1)} = 1$, in Fig. 5.2(b) $\kappa_s^{(0)} = 2$ and $\kappa_s^{(1)} = 1$, while in Fig. 5.2(c) $\kappa_s^{(0)} = 2$ and $\kappa_s^{(1)} = 4$.

The following three behaviors can be observed in every sub-plot of Fig. 5.2. First, all trajectories entering into $\mathcal{Q}_{\mathrm{c}}(\boldsymbol{q}_{\mathrm{r}})$ (denoted by the shaded region between the dash-dotted lines) converge to an estimator for the desired source, confirming Theorem 5.1. Next, all trajectories initialized with small enough kurtosis (indicated by the region between the dotted lines) and proper gain (indicated by the fat shaded arc) converge to an estimator for the desired source, thus confirming Theorem 5.2. Finally, all trajectories initialized with high enough SINR (indicated by the region between the dashed lines) and proper gain (again indicated by the fat shaded arc) converge to estimators for the desired source, confirming Theorem 5.3.

Fig. 5.2 suggests that the sufficient-SINR condition of Theorem 5.3 is more restrictive than the sufficient-kurtosis condition of Theorem 5.2, which in turn is more restrictive than the J_{c}-based condition of Theorem 5.1: the sufficient-SINR region (between the dashed lines) is contained by the sufficient-kurtosis region (between the dotted lines), which is contained by the sufficient-J_{c} region (between the dash-dotted lines). The relative ordering of these three conditions is, in fact, formally implied by the proofs in Appendix 5.A.

We stress again that initial kurtosis or SINR is not sufficient for desired local convergence; initial estimator gain plays an important role. This is demonstrated by Fig. 5.2(a)-(b) wherein some trajectories initialized within the $\mathrm{SINR}_\nu(\boldsymbol{q}) > \mathrm{SINR}_{\min,\nu}$ region (between the dashed lines), but with insufficient initial gain, converge to the undesired equilibria $\boldsymbol{q} = (0, \pm 1)^t$. Though recognized in [Treichler TASSP 85a], this fact was overlooked in [Li TSP 95], resulting in some overly strong claims about the convergence behavior of CMA.

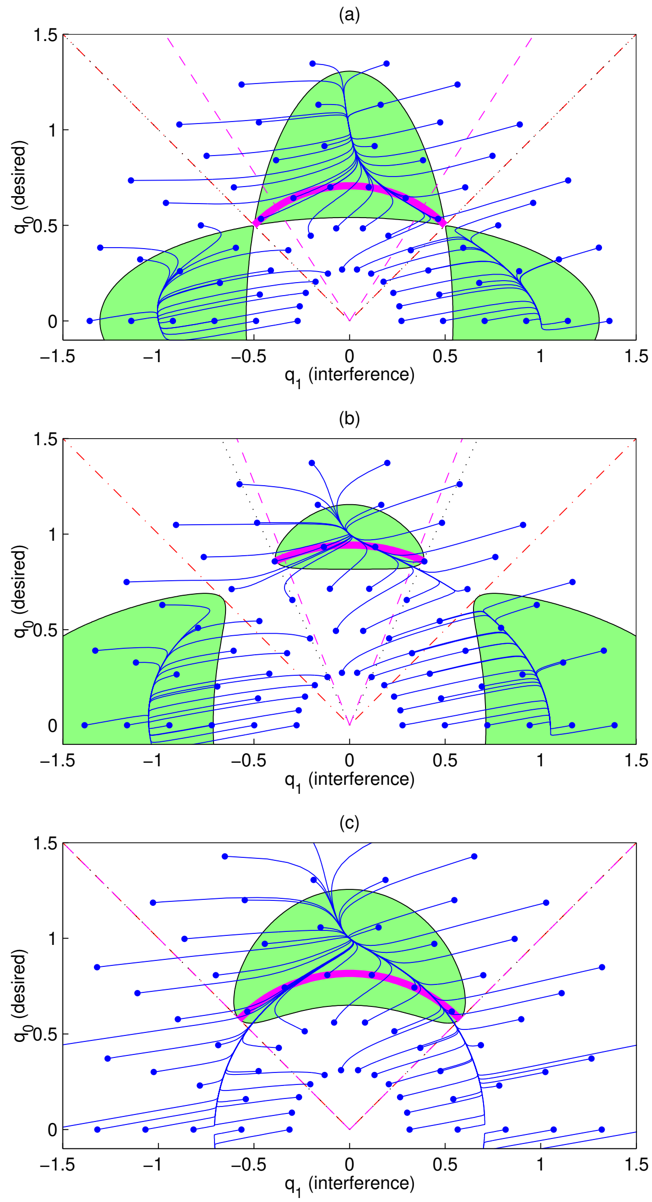

Figure 5.2: CM-GD trajectories in channel-plus-estimator space ($\boldsymbol{q} \in \mathbb{R}^2$) for (a) $\kappa_s^{(0)} = 1$ and $\kappa_s^{(1)} = 1$, (b) $\kappa_s^{(0)} = 2$ and $\kappa_s^{(1)} = 1$, and (c) $\kappa_s^{(0)} = 2$ and $\kappa_s^{(1)} = 4$. Also shown are $\mathcal{Q}_\nu^{(0)}$ boundaries (dash-dotted), $\mathrm{SINR}_{\min,\nu}$ boundaries (dashed), κ_y^{crit} boundaries (dotted), and $J_{\mathrm{c}}(\boldsymbol{q}) < J_{\mathrm{c}}(\boldsymbol{q}_{\mathrm{r}})$ regions (shaded). Fat shaded arcs denote global responses giving $\sigma_y^2 = \sigma_y^2\big|_{\mathrm{crit}}$. Notice that dotted and dash-dotted lines are coincident in (a), while dotted, dash-dotted, and dashed lines are coincident in (c).

In Fig. 5.3 we examine probability of CM-GD convergence to desired {source, delay} versus SINR for higher-dimensional estimators. CM gradient descents randomly initialized in a ball around $\mathbf{f}_{\mathrm{m},\nu}$ (and subsequently normalized according to Theorem 5.3) were conducted using random channel matrices $\{\boldsymbol{\mathcal{H}}\} \in \mathbb{R}^{10\times 11}$ with zero-mean Gaussian elements. Every data point in Fig. 5.3 represents an average of 500 CM-GD simulations. Fig. 5.3(a) demonstrates $\kappa_s^{(0)} = 1$ and ten interfering sources with $\kappa_s^{(k)} = 1$; Fig. 5.3(b) demonstrates $\kappa_s^{(0)} = 2$, five interfering sources with $\kappa_s^{(k)} = 1$, and five interfering sources with $\kappa_s^{(k)} = 2$; while Fig. 5.3(c) demonstrates $\kappa_s^{(0)} = 2$, five interfering sources with $\kappa_s^{(k)} = 2$, and five interfering sources with $\kappa_s^{(k)} = 4$.

Fig. 5.3 also confirms the claim of Theorem 5.3: all properly-scaled CM-GD initializations with SINR_ν greater than $\mathrm{SINR}_{\min,\nu}$ converge to the desired source. Recalling that the SINR condition is sufficient, but not necessary, for desired convergence, it is interesting to note that both Fig. 5.2 and Fig. 5.3 suggest that the sufficiency of our SINR condition becomes "looser" as the kurtosis of the desired source rises above the minimum interference kurtosis (i.e., as ρ_{imin} increases).

5.5 Conclusions

In this chapter we have derived, under the general linear model of Fig. 2.3, three sufficient conditions for the convergence of CM-minimizing gradient descent to an estimator for a particular source at a particular delay. The sufficient conditions are expressed in terms of statistical properties of initial estimates, i.e., estimates generated by an estimator parameterization from which the gradient descent procedure is initialized. More specifically, we have proven that when initial estimates result in sufficiently low CM cost, or in sufficiently low kurtosis *and* a particular variance,

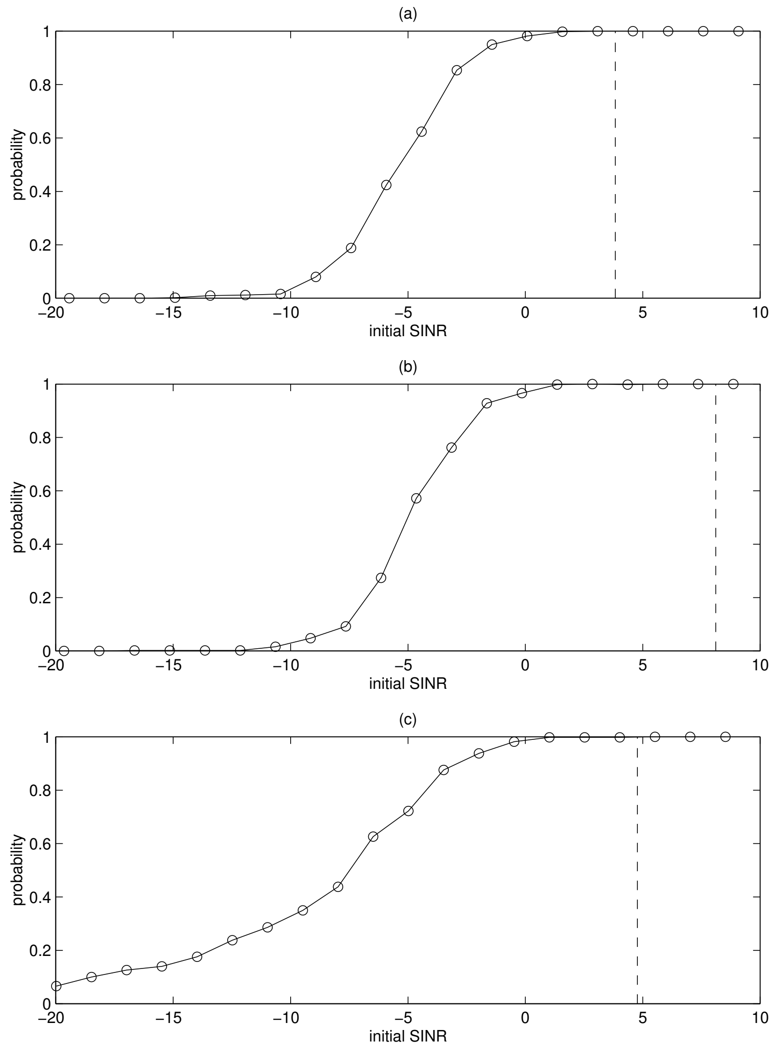

Figure 5.3: Estimated probability of convergence to desired source/delay for random channels and random initializations scaled according to Theorem 5.3 and plotted as a function of initialization SINR. In (a) $\kappa_s^{(0)} = 1$ with interfering $\kappa_s^{(k)} \in \{1, 3\}$, in (b) $\kappa_s^{(0)} = 2$ with interfering $\kappa_s^{(k)} \in \{1, 2\}$, and in (c) $\kappa_s^{(0)} = 2$ with interfering $\kappa_s^{(k)} \in \{2, 4\}$. $\mathrm{SINR}_{\min,\nu}$ from (5.14) shown by dashed lines.

CM-GD will preserve the source/delay combination associated with the initial estimator. In addition, we have proven that when the SINR of the initial estimates (with respect to a particular source/delay combination) is above a prescribed threshold and the estimates have a particular variance, CM-GD will converge to an estimator of the same source/delay. These results suggest ways in which *a priori* channel knowledge may be used to predict and control the convergence behavior of CMA and are of particular importance in multi-user applications.

Appendix

5.A Derivation Details for Local Convergence Conditions

This appendix contains the proofs of the theorems and lemmas found in Section 5.2.2.

5.A.1 Proof of Lemma 5.1

We are interested in computing the minimum CM cost on the boundary of the set $\mathcal{Q}_\nu^{(0)}$. The approach we take is to minimize J_{c} over a set containing $\mathrm{bndr}(\mathcal{Q}_\nu^{(0)})$ which, as shown below, still yields a minimum within $\mathrm{bndr}(\mathcal{Q}_\nu^{(0)})$. Specifically, we consider the set

$$\left\{\boldsymbol{q} : q_i^{(\ell)} = q_\nu^{(0)} \text{ for } (\ell,i) \neq (0,\nu)\right\} \;\supset\; \mathrm{bndr}(\mathcal{Q}_\nu^{(0)}).$$

If (ℓ, i) represents the {source, delay} pair of minimum interference kurtosis [recall the definition of κ_s^{imin} in (5.8)], we henceforth use $\breve{\boldsymbol{q}}^{(k)}$ to denote $\boldsymbol{q}^{(k)}$ with the terms $q_\nu^{(0)}$ and $q_i^{(\ell)}$ removed. Then we have

$$\min_{(\ell,i)\neq(0,\nu)} \sum_k (\kappa_s^{(k)} - \kappa_g)\|\boldsymbol{q}^{(k)}\|_4^4\Big|_{q_\nu^{(0)}=q_i^{(\ell)}} = (\kappa_s^{(0)} + \kappa_s^{\mathrm{imin}} - 2\kappa_g)|q_\nu^{(0)}|^4 + \sum_k (\kappa_s^{(k)} - \kappa_g)\|\breve{\boldsymbol{q}}^{(k)}\|_4^4,$$

and

$$\min_{(\ell,i)\neq(0,\nu)} \|\boldsymbol{q}\|_2^2\Big|_{q_\nu^{(0)}=q_i^{(\ell)}} = 2|q_\nu^{(0)}|^2 + \|\breve{\boldsymbol{q}}\|_2^2.$$

Plugging the two previous equations into (4.5), we find that

$$\min_{(\ell,i)\neq(0,\nu)} \min_{q_\nu^{(0)}=q_i^{(\ell)}} J_{\mathrm{c}}(\boldsymbol{q}) \quad\Leftrightarrow\quad \min_{q_\nu^{(0)},\breve{\boldsymbol{q}}} J_{\mathrm{c}}(q_\nu^{(0)}, \breve{\boldsymbol{q}})$$

where

$$\begin{aligned}\frac{J_c\left(q_\nu^{(0)}, \breve{\boldsymbol{q}}\right)}{\sigma_s^4} &:= (\kappa_s^{(0)} + \kappa_s^{\text{imin}} - 2\kappa_g)|q_\nu^{(0)}|^4 + \textstyle\sum_k (\kappa_s^{(k)} - \kappa_g)\|\breve{\boldsymbol{q}}^{(k)}\|_4^4 \\ &\quad + \kappa_g \left(2|q_\nu^{(0)}|^2 + \|\breve{\boldsymbol{q}}\|_2^2\right)^2 - 2(\gamma/\sigma_s^2)\left(2|q_\nu^{(0)}|^2 + \|\breve{\boldsymbol{q}}\|_2^2\right) + (\gamma/\sigma_s^2)^2 \\ &= (\kappa_s^{(0)} + \kappa_s^{\text{imin}} + 2\kappa_g)|q_\nu^{(0)}|^4 - 4\left((\gamma/\sigma_s^2) - \kappa_g\|\breve{\boldsymbol{q}}\|_2^2\right)|q_\nu^{(0)}|^2 + J_c(\breve{\boldsymbol{q}})/\sigma_s^4.\end{aligned} \tag{5.18}$$

Zeroing the partial derivative of $J_c\left(q_\nu^{(0)}, \breve{\boldsymbol{q}}\right)$ w.r.t. $|q_\nu^{(0)}|^2$ yields

$$\arg\min_{|q_\nu^{(0)}|^2} J_c\left(q_\nu^{(0)}, \breve{\boldsymbol{q}}\right) = 2\frac{(\gamma/\sigma_s^2) - \kappa_g\|\breve{\boldsymbol{q}}\|_2^2}{\kappa_s^{(0)} + \kappa_s^{\text{imin}} + 2\kappa_g}, \tag{5.19}$$

thus

$$\begin{aligned}\frac{J_c\left(q_\nu^{(0)}\big|_{\min}, \breve{\boldsymbol{q}}\right)}{\sigma_s^4} &= \frac{J_c(\breve{\boldsymbol{q}})}{\sigma_s^4} - 4\frac{\kappa_g^2\|\breve{\boldsymbol{q}}\|_2^4 - 2\kappa_g(\gamma/\sigma_s^2)\|\breve{\boldsymbol{q}}\|_2^2 + (\gamma/\sigma_s^2)^2}{\kappa_s^{(0)} + \kappa_s^{\text{imin}} + 2\kappa_g} \\ &= \sum_k (\kappa_s^{(k)} - \kappa_g)\|\breve{\boldsymbol{q}}^{(k)}\|_4^4 + \kappa_g d\|\breve{\boldsymbol{q}}\|_2^4 - 2(\gamma/\sigma_s^2)d\|\breve{\boldsymbol{q}}\|_2^2 \\ &\quad + (\gamma/\sigma_s^2)^2\left(1 - 4(\kappa_s^{(0)} + \kappa_s^{\text{imin}} + 2\kappa_g)^{-1}\right)\end{aligned} \tag{5.20}$$

using the abbreviation

$$d := \frac{\kappa_s^{(0)} + \kappa_s^{\text{imin}} - 2\kappa_g}{\kappa_s^{(0)} + \kappa_s^{\text{imin}} + 2\kappa_g}. \tag{5.21}$$

The gradient and Hessian analysis below reveals that, when $\kappa_s^{\text{imin}} \leq 2\kappa_g - \kappa_s^{(0)}$, the (unique) global minimum of $J_c\left(q_\nu^{(0)}\big|_{\min}, \breve{\boldsymbol{q}}\right)$ occurs at $\breve{\boldsymbol{q}} = \mathbf{0}$, implying [via (5.20)] that

$$\begin{aligned}\min_{\boldsymbol{q}\in\text{bndr}(\mathcal{Q}_\nu^{(0)})} \frac{J_c(\boldsymbol{q})}{\sigma_s^4} &= \min_{\breve{\boldsymbol{q}}} \frac{J_c\left(q_\nu^{(0)}\big|_{\min}, \breve{\boldsymbol{q}}\right)}{\sigma_s^4} \\ &= \left(\frac{\gamma}{\sigma_s^2}\right)^2\left(1 - \frac{4}{\kappa_s^{(0)} + \kappa_s^{\text{imin}} + 2\kappa_g}\right).\end{aligned} \tag{5.22}$$

Now for the gradient and Hessian analysis. For simplicity, we assume real-valued quantities, though a similar technique may be used to establish the same result in the complex-valued case. Using standard techniques (see, e.g., [Johnson IJACSP 95]), it can be shown that the component of the gradient $\nabla_{\check{\boldsymbol{q}}} J_c\big(q_\nu^{(0)}\big|_{\min}, \check{\boldsymbol{q}}\big)$ corresponding to the direction of $q_i^{(k)}$ is

$$\Big[\nabla_{\check{\boldsymbol{q}}} J_c\big(q_\nu^{(0)}\big|_{\min}, \check{\boldsymbol{q}}\big)\Big]_i^{(k)} \;=\; \Big((\kappa_s^{(k)} - \kappa_g)|q_i^{(k)}|^2 + d\left(\kappa_g\|\check{\boldsymbol{q}}\|_2^2 - (\gamma/\sigma_s^2)\right)\Big)\, q_i^{(k)}$$

and the component of the Hessian $\mathcal{H}_{\check{\boldsymbol{q}}} J_c\big(q_\nu^{(0)}\big|_{\min}, \check{\boldsymbol{q}}\big)$ corresponding to the directions of $q_i^{(k)}$ and $q_j^{(\ell)}$ is

$$\Big[\mathcal{H}_{\check{\boldsymbol{q}}} J_c\big(q_\nu^{(0)}\big|_{\min}, \check{\boldsymbol{q}}\big)\Big]_{i,j}^{(k,\ell)} =$$
$$\begin{cases} \left(3(\kappa_s^{(k)} - \kappa_g) + 2\kappa_g d\right)|q_i^{(k)}|^2 + d\left(\kappa_g\|\check{\boldsymbol{q}}\|_2^2 - (\gamma/\sigma_s^2)\right), & k=\ell \text{ and } i=j \\ 2\kappa_g d \cdot q_i^{(k)} q_j^{(\ell)}, & \text{else.} \end{cases}$$

All $\check{\boldsymbol{q}}$ setting the gradient to zero are stationary points, but only those stationary points with positive semi-definite (PSD) Hessians qualify as local minima.

From the previous expression, the stationary points are described by $\check{\boldsymbol{q}}$ with the elements

$$q_i^{(k)} \;=\; \begin{cases} 0, & \text{or} \\ \pm\sqrt{\frac{d\left((\gamma/\sigma_s^2) - \kappa_g\|\check{\boldsymbol{q}}\|_2^2\right)}{\kappa_s^{(k)} - \kappa_g}}, & \text{when } \kappa_s^{(k)} < \kappa_g \text{ and } \|\check{\boldsymbol{q}}\|_2^2 \le (\gamma/\sigma_s^2)/\kappa_g. \end{cases} \tag{5.23}$$

In stating the two conditions in (5.23), we have assumed $\kappa_s^{\text{imin}} \le 2\kappa_g - \kappa_s^{(0)}$ which implies $d \le 0$. Noting that $q_i^{(k)}$ is independent of i, we can write $\|\check{\boldsymbol{q}}\|_2^2 = \sum_k M_k |q_i^{(k)}|^2$ using the integer M_k to denote the number of nonzero $q_i^{(k)}$ in the candidate response. Solving for $\|\check{\boldsymbol{q}}\|_2^2$ and plugging it into (5.23) gives

$$q_i^{(k)} \;=\; \alpha_i^{(k)} \sqrt{d\left(\frac{\gamma}{\sigma_s^2}\right)\left(1 + \kappa_g d \sum_\ell \frac{M_\ell}{\kappa_s^{(\ell)} - \kappa_g}\right)^{-1} \frac{1}{\kappa_s^{(k)} - \kappa_g}}$$
$$\text{for} \quad \alpha_i^{(k)} \in \{-1, 0, 1\},$$

when $\kappa_s^{(k)} < \kappa_g$ and $\sum_\ell \frac{M_\ell}{\kappa_s^{(\ell)} - \kappa_g} < \frac{1}{\kappa_g d}$ (and where $\sum_i |\alpha_i^{(k)}| = M_k$).

The Hessian will be PSD if and only if all of its principle minors are PSD. The principle minors of $\mathcal{H}_{\breve{\boldsymbol{q}}} J_{\mathrm{c}}\left(q_\nu^{(0)}\big|_{\min}, \breve{\boldsymbol{q}}\right)$ take the form

$$d \cdot \frac{\gamma}{\sigma_s^2}\left(1+\kappa_g d \sum_\ell \frac{M_\ell}{\kappa_s^{(\ell)}-\kappa_g}\right)^{-1} \times \cdots$$
$$\begin{bmatrix} \left(3+\frac{2\kappa_g d}{\kappa_s^{(k)}-\kappa_g}\right)\alpha_i^{(k)2}-1 & \frac{2\kappa_g |d| \alpha_i^{(k)}\alpha_j^{(\ell)}}{\sqrt{(\kappa_s^{(k)}-\kappa_g)(\kappa_s^{(\ell)}-\kappa_g)}} \\ \frac{2\kappa_g |d| \alpha_i^{(k)}\alpha_j^{(\ell)}}{\sqrt{(\kappa_s^{(k)}-\kappa_g)(\kappa_s^{(\ell)}-\kappa_g)}} & \left(3+\frac{2\kappa_g d}{\kappa_s^{(\ell)}-\kappa_g}\right)\alpha_j^{(\ell)2}-1 \end{bmatrix}. \tag{5.24}$$

We now consider the three relevant combinations of $\{\alpha_i^{(k)2}, \alpha_j^{(\ell)2}\}$.

i) When $\{\alpha_i^{(k)2}, \alpha_j^{(\ell)2}\} = \{0, 0\}$, (5.24) implies that the principle minors take the form

$$-d \cdot \frac{\gamma}{\sigma_s^2}\left(1+\kappa_g d \sum_\ell \frac{M_\ell}{\kappa_s^{(\ell)}-\kappa_g}\right)^{-1} \mathbf{I}.$$

Then the requirements $d \le 0$ (since $\kappa_s^{\mathrm{imin}} \le 2\kappa_g - \kappa_s^{(0)}$) and $\sum_\ell \frac{M_\ell}{\kappa_s^{(\ell)}-\kappa_g} < \frac{1}{\kappa_g d}$ imply that principle minors of this form must be PSD.

ii) Now when $\{\alpha_i^{(k)2}, \alpha_j^{(\ell)2}\} = \{1, 0\}$, we have principle minors of the form

$$d \cdot \frac{\gamma}{\sigma_s^2}\left(1+\kappa_g d \sum_\ell \frac{M_\ell}{\kappa_s^{(\ell)}-\kappa_g}\right)^{-1} \begin{bmatrix} 2+\frac{2\kappa_g d}{\kappa_s^{(k)}-\kappa_g} & 0 \\ 0 & -1 \end{bmatrix}.$$

Using the definition of d from (5.21) it can be shown that the top left term in the bracketed quantity will be positive when

$$\kappa_s^{(k)} < \frac{4\kappa_g^2}{\kappa_s^{(k)}+\kappa_s^{\mathrm{imin}}+2\kappa_g}.$$

Then, since $\kappa_s^{\mathrm{imin}} \le 2\kappa_g - \kappa_s^{(0)}$ implies that $\kappa_s^{(k)} + \kappa_s^{\mathrm{imin}} + 2\kappa_g \le 4\kappa_g$, a sufficient condition for the previous inequality becomes $\kappa_s^{(k)} < \kappa_g$. Since this is required any time that $\alpha_i^{(k)} \ne 0$ (recalling (5.23)), the diagonal matrix will have elements of different sign. Hence, principle minors of this form can never be PSD.

iii) Finally, when $\{\alpha_i^{(k)2}, \alpha_j^{(\ell)2}\} = \{1, 1\}$, the principle minors become

$$d \cdot \frac{\gamma}{\sigma_s^2} \left(1 + \kappa_g d \sum_\ell \frac{M_\ell}{\kappa_s^{(\ell)} - \kappa_g}\right)^{-1} \times \cdots$$
$$\begin{bmatrix} 2+\frac{2\kappa_g d}{\kappa_s^{(k)}-\kappa_g} & \frac{2\kappa_g |d|}{\sqrt{(\kappa_s^{(k)}-\kappa_g)(\kappa_s^{(\ell)}-\kappa_g)}} \\ \frac{2\kappa_g |d|}{\sqrt{(\kappa_s^{(k)}-\kappa_g)(\kappa_s^{(\ell)}-\kappa_g)}} & 2+\frac{2\kappa_g d}{\kappa_s^{(\ell)}-\kappa_g} \end{bmatrix}.$$

We have just seen that the diagonal terms of the bracketed expression will be positive when $\kappa_s^{\text{imin}} \leq 2\kappa_g - \kappa_s^{(0)}$, and it is evident that the off-diagonal terms are always positive. But since the term multiplying the bracketed expression will be negative, the principle minor is purely negative, hence cannot be PSD.

Summing up the three points above, the Hessian will be PSD only when $\alpha_i^{(k)} = 0$ for all k and i. In other words, the unique local minimum of $J_{\text{c}}\big(q_\nu^{(0)}\big|_{\text{min}}, \breve{\boldsymbol{q}}\big)$ occurs at $\breve{\boldsymbol{q}} = \mathbf{0}$.

5.A.2 Proof of Theorem 5.1

If $\boldsymbol{q}_{\text{init}} \in \mathcal{Q}_\nu^{(0)}$ satisfies $J_{\text{c}}(\boldsymbol{q}_{\text{init}}) \leq J_{\text{c}}(\boldsymbol{q}_{\text{r}})$, then by definition, $\boldsymbol{q}_{\text{init}} \in \mathcal{Q}_{\text{c}}(\boldsymbol{q}_{\text{r}})$. (Note that, for $\boldsymbol{q}_{\text{init}}$ to be meaningful, we also require that it be an attainable global response, i.e., $\boldsymbol{q}_{\text{init}} \in \mathcal{Q}_{\text{a}}$.) Combining $\mathcal{Q}_{\text{c}}(\boldsymbol{q}_{\text{r}}) \subset \mathcal{Q}_\nu^{(0)}$ with the fact that a CM-GD trajectory initialized within $\mathcal{Q}_{\text{c}}(\boldsymbol{q}_{\text{r}})$ remains entirely within $\mathcal{Q}_{\text{c}}(\boldsymbol{q}_{\text{r}})$, we conclude that a CM-GD trajectory initialized at $\boldsymbol{q}_{\text{init}}$ remains entirely within $\mathcal{Q}_\nu^{(0)}$. Using the $J_{\text{c}}(\boldsymbol{q}_{\text{r}})$ expression (5.11) appearing in Lemma 4.2, we arrive at (5.12).

5.A.3 Proof of Theorem 5.2

Continuing the arguments used in the proof of Lemma 4.1, the CM cost expression (4.5) can be restated as follows.

$$\frac{J_{\mathrm{c}}(\boldsymbol{q})}{\sigma_s^4} \;=\; \kappa_y \|\boldsymbol{q}\|_2^4 - 2(\gamma/\sigma_s^2)\|\boldsymbol{q}\|_2^2 + (\gamma/\sigma_s^2)^2.$$

From Theorem 5.1, a CM cost satisfying (5.12) suffices to guarantee the desired CM-GD property. Normalizing (5.12) by σ_s^4 and plugging in the previous expression, we obtain the equivalent sufficient conditions

$$\begin{aligned}
0 &> \frac{J_{\mathrm{c}}(\boldsymbol{q})}{\sigma_s^4} - (\gamma/\sigma_s^2)^2 \left(1 - \frac{4}{\kappa_s^{(0)} + \kappa_s^{\mathrm{imin}} + 2\kappa_g}\right)\\
0 &> \kappa_y \|\boldsymbol{q}\|_2^4 - 2(\gamma/\sigma_s^2)\|\boldsymbol{q}\|_2^2 + (\gamma/\sigma_s^2)^2 \left(\frac{4}{\kappa_s^{(0)} + \kappa_s^{\mathrm{imin}} + 2\kappa_g}\right)\\
\kappa_y &< 2(\gamma/\sigma_s^2)\|\boldsymbol{q}\|_2^{-2} - (\gamma/\sigma_s^2)^2 \left(\frac{4}{\kappa_s^{(0)} + \kappa_s^{\mathrm{imin}} + 2\kappa_g}\right) \|\boldsymbol{q}\|_2^{-4}.
\end{aligned}$$

It is now apparent that the critical value of κ_y depends on the gain $\|\boldsymbol{q}\|_2$. Maximizing the critical kurtosis w.r.t. $\|\boldsymbol{q}\|_2$ can be accomplished by finding a which zeros the partial derivative of

$$2(\gamma/\sigma_s^2)(a^2)^{-1} - (\gamma/\sigma_s^2)^2 \left(\frac{4}{\kappa_s^{(0)} + \kappa_s^{\mathrm{imin}} + 2\kappa_g}\right) (a^2)^{-2}$$

w.r.t. a^2. Straightforward calculus reveals that the maximizing value of $\|\boldsymbol{q}\|_2^2$ is

$$\|\boldsymbol{q}\|_2^2\big|_{\max} \;=\; \left(\frac{\gamma}{\sigma_s^2}\right) \left(\frac{4}{\kappa_s^{(0)} + \kappa_s^{\mathrm{imin}} + 2\kappa_g}\right)$$

which implies that the maximum critical kurtosis is

$$\kappa_y \;=\; \left(\kappa_s^{(0)} + \kappa_s^{\mathrm{imin}} + 2\kappa_g\right)/4.$$

Since S1)-S3) imply that $\sigma_y^2 = \|\boldsymbol{q}\|_2^2 \sigma_s^2$, the expression for $\|\boldsymbol{q}\|_2^2\big|_{\max}$ above is easily rewritten in terms of estimate variance σ_y^2.

5.A.4 Proof of Theorem 5.3

Section 5.2.1 established that estimators yielding gain a_* and producing estimates of SINR_ν greater than $\mathrm{SINR}_{\min,\nu}$ are contained within the set $\mathcal{Q}_c(\boldsymbol{q}_r)$, and thus further CM-GD adaptation of these estimates will guarantee estimation of the desired source. In this section we will derive explicit formulas for the quantities $\mathrm{SINR}_{\min,\nu}$ and a_*. This will be accomplished through (5.6) and (5.7) after first solving for $b_{\max}(a)$ defined in (5.5).

To find $b_{\max}(a)$, (5.5) may be translated

$$b_{\max}(a) \;=\; \max b(a) \text{ s.t. } \Big\{ \|\bar{\boldsymbol{q}}\|_2 \le b(a) \;\;\Rightarrow\;\; J_c(a, \bar{\boldsymbol{q}}) < J_c(\boldsymbol{q}_r) \Big\}. \tag{5.25}$$

To proceed further, the CM cost expression (4.5) must be rewritten in terms of gain $a = \|\boldsymbol{q}\|_2$ and interference response $\bar{\boldsymbol{q}}$ (defined in Section 2.3.3). Using the fact that $|q_\nu^{(0)}|^2 = a^2 - \|\bar{\boldsymbol{q}}\|_2^2$,

$$\begin{aligned}
\sum_k (\kappa_s^{(k)} - \kappa_g)\|\boldsymbol{q}^{(k)}\|_4^4 &= (\kappa_s^{(0)} - \kappa_g)|q_\nu^{(0)}|^4 + \sum_k (\kappa_s^{(k)} - \kappa_g)\|\bar{\boldsymbol{q}}^{(k)}\|_4^4 \\
&= (\kappa_s^{(0)} - \kappa_g)\big(a^4 - 2a^2\|\bar{\boldsymbol{q}}\|_2^2 + \|\bar{\boldsymbol{q}}\|_2^4\big) + \sum_k (\kappa_s^{(k)} - \kappa_g)\|\bar{\boldsymbol{q}}^{(k)}\|_4^4.
\end{aligned}$$

Plugging the previous expression into (4.5), we find that

$$\begin{aligned}
\frac{J_c(a, \bar{\boldsymbol{q}})}{\sigma_s^4} &= \sum_k (\kappa_s^{(k)} - \kappa_g)\|\bar{\boldsymbol{q}}^{(k)}\|_4^4 + \kappa_s^{(0)} a^4 - 2(\kappa_s^{(0)} - \kappa_g)a^2\|\bar{\boldsymbol{q}}\|_2^2 + (\kappa_s^{(0)} - \kappa_g)\|\bar{\boldsymbol{q}}\|_2^4 \\
&\quad - 2(\gamma/\sigma_s^2)a^2 + (\gamma/\sigma_s^2)^2.
\end{aligned} \tag{5.26}$$

From (5.11) and (5.26), the following statements are equivalent:

$$
\begin{aligned}
J_{\mathrm{c}}(\boldsymbol{q}_{\mathrm{r}}) &> J_{\mathrm{c}}(a, \bar{\boldsymbol{q}}) \\
0 &> \sum_k (\kappa_s^{(k)} - \kappa_g)\|\bar{\boldsymbol{q}}^{(k)}\|_4^4 + (\kappa_s^{(0)} - \kappa_g)\Big(-2a^2\|\bar{\boldsymbol{q}}\|_2^2 + \|\bar{\boldsymbol{q}}\|_2^4\Big) \\
&\quad + \kappa_s^{(0)} a^4 - 2\left(\frac{\gamma}{\sigma_s^2}\right) a^2 + \left(\frac{\gamma}{\sigma_s^2}\right)^2 \left(\frac{4}{\kappa_s^{(0)} + \kappa_s^{\mathrm{imin}} + 2\kappa_g}\right) \\
0 &< \frac{1}{\kappa_s^{(0)} - \kappa_g} \sum_k (\kappa_s^{(k)} - \kappa_g)\|\bar{\boldsymbol{q}}^{(k)}\|_4^4 - 2a^2\|\bar{\boldsymbol{q}}\|_2^2 + \|\bar{\boldsymbol{q}}\|_2^4 \\
&\quad + \underbrace{\frac{1}{\kappa_s^{(0)} - \kappa_g}\left(\kappa_s^{(0)} a^4 - 2\left(\frac{\gamma}{\sigma_s^2}\right) a^2 + \left(\frac{\gamma}{\sigma_s^2}\right)^2 \left(\frac{4}{\kappa_s^{(0)} + \kappa_s^{\mathrm{imin}} + 2\kappa_g}\right)\right)}_{C(a,\boldsymbol{q}_{\mathrm{r}})}
\end{aligned}
\tag{5.27}
$$

The reversal of inequality in (5.27) occurs because $\kappa_s^{(0)} - \kappa_g < 0$ (as implied by S4)). Using the definition of $\kappa_s^{\max}$ in (2.22), $0 \leq \frac{\|\bar{\boldsymbol{q}}^{(k)}\|_4^4}{\|\bar{\boldsymbol{q}}^{(k)}\|_2^4} \leq 1$ implies that

$$
\sum_k (\kappa_s^{(k)} - \kappa_g)\|\bar{\boldsymbol{q}}^{(k)}\|_4^4 \;\leq\; (\kappa_s^{\max} - \kappa_g)\|\bar{\boldsymbol{q}}\|_4^4 \;\leq\; \begin{cases} 0, & \kappa_s^{\max} \leq \kappa_g \\ (\kappa_s^{\max} - \kappa_g)\|\bar{\boldsymbol{q}}\|_2^4, & \kappa_s^{\max} > \kappa_g. \end{cases} \tag{5.28}
$$

Thus, with $\rho_{\max}$ defined in (2.24), the following becomes a sufficient condition for (5.27).

$$
0 \;>\; \begin{cases} \|\bar{\boldsymbol{q}}\|_2^4 - 2a^2\|\bar{\boldsymbol{q}}\|_2^2 + C(a, \boldsymbol{q}_{\mathrm{r}}), & \kappa_s^{\max} \leq \kappa_g \\ (1 + \rho_{\max})\|\bar{\boldsymbol{q}}\|_2^4 - 2a^2\|\bar{\boldsymbol{q}}\|_2^2 + C(a, \boldsymbol{q}_{\mathrm{r}}), & \kappa_s^{\max} > \kappa_g. \end{cases} \tag{5.29}
$$

Focusing first on the super-Gaussian case ($\kappa_s^{\max} > \kappa_g$), we see from (5.29) that valid $b_{\max}^2(a)$ satisfying (5.25) can be determined by solving for the roots of

$$
P_1(x) \;=\; (1 + \rho_{\max})\, x^2 - 2a^2 x + C(a, \boldsymbol{q}_{\mathrm{r}}).
$$

Specifically, we are interested in the smaller root when $1 + \rho_{\max} > 0$ and the larger root when $1 + \rho_{\max} < 0$. In either of these two cases, the appropriate root has the form

$$b^2_{\max}(a) \;=\; a^2 \left(\frac{1 - \sqrt{1 - (\rho_{\max} + 1)\frac{C(a, \boldsymbol{q}_{\mathrm{r}})}{a^4}}}{\rho_{\max} + 1} \right). \tag{5.30}$$

When $1 + \rho_{\max} = 0$ instead, $P_1(x)$ becomes linear and

$$b^2_{\max}(a) \;=\; \frac{C(a, \boldsymbol{q}_{\mathrm{r}})}{2a^2}. \tag{5.31}$$

As a valid interference power, we require that $b^2_{\max}(a) \in [0, a^2]$. Straightforward manipulations show that, for all valid super-Gaussian values of $\rho_{\max}$ (i.e., $\rho_{\max} < 0$),

$$b^2_{\max}(a) \in [0, a^2] \quad \Leftrightarrow \quad 0 \;\le\; \frac{C(a, \boldsymbol{q}_{\mathrm{r}})}{a^4} \;\le\; 1 - \rho_{\max}. \tag{5.32}$$

From (5.29) it can be seen that the same arguments may be applied to the non-super-Gaussian case ($\kappa_s^{\max} \le \kappa_g$) by setting $\rho_{\max}$ to zero. This yields

$$b^2_{\max}(a) \;=\; a^2 \left(1 - \sqrt{1 - \frac{C(a, \boldsymbol{q}_{\mathrm{r}})}{a^4}} \right) \tag{5.33}$$

with the requirement that $0 \le C(a, \boldsymbol{q}_{\mathrm{r}})/a^4 \le 1$.

The expressions for $b^2_{\max}(a)$ in (5.30), (5.31), and (5.33) can now be used to calculate $\mathrm{SINR}_{\min,\nu}$ and a_* given in (5.6) and (5.7). First we tackle the super-Gaussian case ($\kappa_s^{\max} > \kappa_g$). Assuming for the moment that $\rho_{\max} \neq -1$, we plug (5.30) into (5.6) to obtain

$$\mathrm{SINR}_{\min,\nu} \;=\; \min_a \frac{\rho_{\max} + \sqrt{1 - (\rho_{\max} + 1)\frac{C(a, \boldsymbol{q}_{\mathrm{r}})}{a^4}}}{1 - \sqrt{1 - (\rho_{\max} + 1)\frac{C(a, \boldsymbol{q}_{\mathrm{r}})}{a^4}}}. \tag{5.34}$$

Since the fraction on the right of (5.34) is non-negative and strictly decreasing in $C(a, \boldsymbol{q}_{\mathrm{r}})/a^4$ over the valid range $C(a, \boldsymbol{q}_{\mathrm{r}})/a^4 \in [0, 1 - \rho_{\max}]$ identified by (5.32), finding a that minimizes this expression [in accordance with (5.7)] can be accomplished by finding a that maximizes $C(a, \boldsymbol{q}_{\mathrm{r}})/a^4$. To find these maxima, we first write $C(a, \boldsymbol{q}_{\mathrm{r}})/a^4$ using (5.27):

$$\frac{C(a, \boldsymbol{q}_{\mathrm{r}})}{a^4} \;=\; C_0 + C_1 \cdot (a^2)^{-1} + C_2 \cdot (a^2)^{-2},$$

where C_0, C_1, and C_2 are independent of a. Computing the partial derivative with respect to the quantity a^2 and setting it equal to zero, we find that

$$a_*^2 \;=\; -2\frac{C_2}{C_1} \;=\; \left(\frac{\gamma}{\sigma_s^2}\right)\left(\frac{4}{\kappa_s^{(0)}+\kappa_s^{\text{imin}}+2\kappa_g}\right). \tag{5.35}$$

Plugging a_*^2 into (5.27) and using the definition of ρ_{imin} in (5.9) gives the simple result $C(a_*, \boldsymbol{q}_{\text{r}})/a_*^4 = (3-\rho_{\text{imin}})/4$. With this value of a_*, requirement (5.32) translates into

$$\kappa_s^{(0)} \;\leq\; \min\left\{\frac{\kappa_s^{\text{imin}}+2\kappa_g}{3},\; 4\kappa_s^{\max}-\kappa_s^{\text{imin}}-2\kappa_g\right\}. \tag{5.36}$$

In the super-Gaussian case, we know that $\kappa_s^{(0)} < 4\kappa_s^{\max}-\kappa_s^{\text{imin}}-2\kappa_g$, hence (5.36) simplifies to

$$\kappa_s^{(0)} \;\leq\; \frac{\kappa_s^{\text{imin}}+2\kappa_g}{3}.$$

Finally, plugging a_* into (5.34) gives

$$\text{SINR}_{\min,\nu} \;=\; \frac{\rho_{\max}+\sqrt{1-(1+\rho_{\max})(3-\rho_{\text{imin}})/4}}{1-\sqrt{1-(1+\rho_{\max})(3-\rho_{\text{imin}})/4}}. \tag{5.37}$$

Revisiting the super-Gaussian case with $\rho_{\max} = -1$, we plug (5.31) into (5.6) and get

$$\text{SINR}_{\min,\nu} \;=\; \min_a\; 2\left(\frac{C(a,\boldsymbol{q}_{\text{r}})}{a^4}\right)^{-1} - 1.$$

Again, the quantity to be minimized is strictly decreasing in $C(a, \boldsymbol{q}_{\text{r}})/a^4$ over $[0, 1-\rho_{\max}]$. As above, maximization of $C(a, \boldsymbol{q}_{\text{r}})/a^4$ yields the a_* of (5.35) and the same condition on $\kappa_s^{(0)}$. Applying these to the previous equation,

$$\text{SINR}_{\min,\nu} \;=\; \frac{5+\rho_{\text{imin}}}{3-\rho_{\text{imin}}}. \tag{5.38}$$

For the non-super-Gaussian case ($\kappa_s^{\max} \leq \kappa_g$), we plug (5.33) into (5.6) and obtain

$$\text{SINR}_{\min,\nu} \;=\; \min_a \frac{\sqrt{1-\frac{C(a,\boldsymbol{q}_{\text{r}})}{a^4}}}{1-\sqrt{1-\frac{C(a,\boldsymbol{q}_{\text{r}})}{a^4}}}. \tag{5.39}$$

Since (5.39) equals (5.34) with $\rho_{\max} = 0$ (i.e., when $\kappa_s^{\max} = \kappa_g$), the non-super-Gaussian will have the same a_* as (5.35) and the same translation of (5.32) given by (5.36). After substituting $\kappa_s^{\max} = \kappa_g$ into (5.36), the non-super-Gaussian property implies that condition (5.36) simplifies again to

$$\kappa_s^{(0)} \;\leq\; \frac{\kappa_s^{\mathrm{imin}} + 2\kappa_g}{3}.$$

Plugging a_* from (5.35) into (5.39), the non-super-Gaussian $\mathrm{SINR}_{\min,\nu}$ becomes

$$\mathrm{SINR}_{\min,\nu} \;=\; \frac{\sqrt{1+\rho_{\mathrm{imin}}}}{2-\sqrt{1+\rho_{\mathrm{imin}}}}. \tag{5.40}$$

Finally, S1)-S3) imply that $a = \|\boldsymbol{q}\|_2^2 = \sigma_y^2/\sigma_s^2$, linking the critical gain a_* in (5.35) to the critical estimate variance $\sigma_y^2\big|_{\mathrm{crit}}$ in (5.10), yielding Theorem 5.3.

Chapter 6

Performance Bounds for CM-Based Channel Identification[1]

6.1 Introduction

Consider Fig. 6.1 (or the left half of Fig. 2.3) where a desired source sequence $\{s_n^{(0)}\}$ combines linearly with K interferers through vector channels $\{\mathbf{h}^{(0)}(z), \dots, \mathbf{h}^{(K)}(z)\}$. We now consider the problem of estimating the impulse response coefficients of channel $\{\mathbf{h}^{(0)}(z)\}$, rather than the symbol sequence $\{s_n^{(0)}\}$, knowing only the statistics of the received signal $\{\mathbf{r}_n\}$. The literature refers to this problem as blind channel identification [Tong PROC 98].

In Chapter 6 we analyze the performance of the blind channel identification scheme of Fig. 6.2, whereby M-delayed versions of the CM-minimizing symbol estimates $\{y_n\} \approx \{s_{n-\nu}^{(0)}\}$ are cross-correlated with the vector received samples $\{\mathbf{r}_{n-Q}, \dots, \mathbf{r}_n\}$, yielding the $Q+1$ vector channel parameter estimates $\{\hat{\mathbf{h}}_{\nu+M-Q}^{(0)}, \dots, \hat{\mathbf{h}}_{\nu+M}^{(0)}\}$. The δ^{th} parameter estimate $\hat{\mathbf{h}}_{\nu+M-\delta}^{(0)}$ can be expressed as

[1]The main results of this chapter also appear in the manuscript [Schniter TSP tbd].

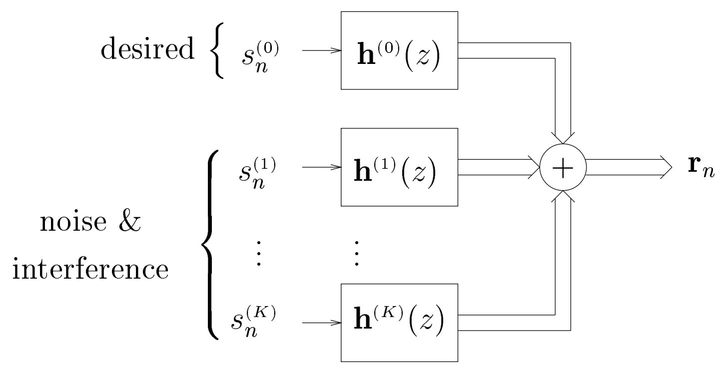

Figure 6.1: Linear system model with K sources of interference.

a biased version of the true parameter corrupted by an error term:

$$\begin{aligned}
\hat{\mathbf{h}}^{(0)}_{\nu+M-\delta} &= \mathrm{E}\left\{\mathbf{r}_{n-\delta}y^*_{n-M}\right\} \\
&= \mathrm{E}\left\{\sum_{k,j}\mathbf{h}^{(k)}_j s^{(k)}_{n-\delta-j} \sum_{\ell,i} q^{(\ell)*}_i s^{(\ell)*}_{n-M-i}\right\} \\
&= \sigma_s^2 \sum_{k,i}\mathbf{h}^{(k)}_{i+M-\delta}\, q^{(k)*}_i \qquad (6.1)\\
&= \underbrace{\sigma_s^2\, q^{(0)*}_\nu}_{\text{bias}}\Bigg(\mathbf{h}^{(0)}_{\nu+M-\delta} + \underbrace{\sum_{(k,i)\neq(0,\nu)} \mathbf{h}^{(k)}_{i+M-\delta}\, q^{(k)*}_i / q^{(0)*}_\nu}_{\text{error}}\Bigg).
\end{aligned}$$

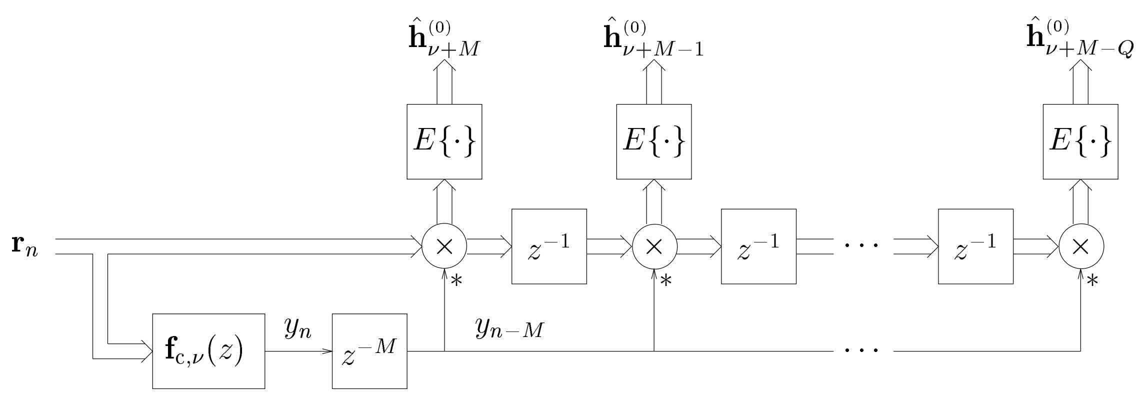

Figure 6.2: Blind channel identification using CM estimates $\{y_n\} \approx \{s^{(0)}_{n-\nu}\}$.

We note that the identification scheme in Fig. 6.2 bears similarity to the Gooch-

Harp method of channel identification [Gooch ICC 88] in Fig. 6.3, whereby the CM estimates $\{y_n\}$ are processed by a hard decision device $\mathcal{D}$ before cross-correlation. Due to the nonlinear operation $\mathcal{D}$, however, performance analysis of the Gooch-Harp scheme is difficult unless perfect decision-making (i.e., $d_n = s_{n-\nu}$) is assumed. In addition, forming reliable decisions requires carrier phase synchronization (an issue with passband data transmission [Proakis Book 95]) which is not required in the identification scheme of Fig. 6.2.

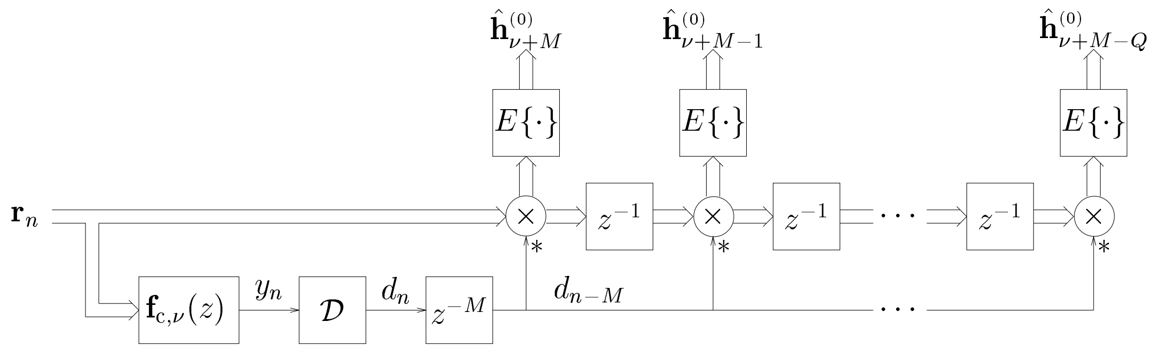

Figure 6.3: Gooch-Harp method of blind channel identification.

Many other methods of blind channel identification have been proposed (see, e.g., the citations in [Tong PROC 98]), most of which estimate channel coefficients from the observed data directly, i.e., without first forming blind symbol estimates. When $P > 1$ and the channel satisfies certain conditions, it is possible to accomplish blind identification using only the second-order statistics (SOS) of the observed process (see, e.g., the references in [Liu SP 96]). Most SOS-based techniques, however, are known to fail catastrophically when the channel order is overestimated (see, e.g., the discussion in [Abed-Meraim TSP 97]). An exception is the approach in [Abed-Meraim TSP 97] where, similar to Figs. 6.2 and 6.3, the channel coefficients are estimated using cross-correlations with blind symbol estimates based on linear prediction. The CM-based schemes in Figs. 6.2 and 6.3, however, do not rely on the satisfaction of any channel-identifiability conditions.

In this chapter, we derive upper bounds for the average squared parameter error (ASPE) of blind channel parameter estimates generated by the method of Fig. 6.2. The bounds are a function of the channel coefficients, the source kurtoses, and the symbol estimation delay. Next, we derive the expected ASPE that results when the correlations in Fig. 6.2 are estimated from N-length data blocks. Finally, we discuss the effect of stochastic-gradient symbol-estimator design (i.e., the use of CMA versus exact gradient descent), as well as the effect of residual carrier offset in $\mathbf{r}_n$, on ASPE. All results are derived in the multi-user vector-channel context of Section 2.2.

The organization of the chapter is as follows. Section 6.2 derives ASPE performance bounds and Section 6.3 covers implementational issues such as finite data effects, the use of CMA, and the presence of residual carrier offset. Section 6.4 presents the results of numerical simulations verifying our performance analyses, and Section 6.5 concludes the chapter. Proofs and technical details have been relegated to Appendix 6.A.

6.2 Blind Identification – Performance Bounds

We are interested in quantifying the sum-squared error of the $Q+1$ parameter estimates $\{\hat{\mathbf{h}}^{(0)}_{\nu+M-Q}, \dots, \hat{\mathbf{h}}^{(0)}_{\nu+M}\}$ relative to the true parameter subset $\{\mathbf{h}^{(0)}_{\nu+M-Q}, \dots, \mathbf{h}^{(0)}_{\nu+M}\}$. We tolerate arbitrary scaling of the total estimated channel response and define our average squared parameter error (ASPE) criterion as follows.

$$\mathcal{E}_{\hat{\boldsymbol{h}}} := \min_{\theta \in \mathbb{C}} \frac{1}{Q+1} \sum_{\delta=0}^{Q} \left\| \theta \, \hat{\mathbf{h}}^{(0)}_{\nu+M-\delta} - \mathbf{h}^{(0)}_{\nu+M-\delta} \right\|_2^2 \tag{6.2}$$

$$= \min_{\theta \in \mathbb{C}} \frac{1}{Q+1} \left\| \theta \underbrace{\begin{pmatrix} \hat{\mathbf{h}}^{(0)}_{\nu+M} \\ \vdots \\ \hat{\mathbf{h}}^{(0)}_{\nu+M-Q} \end{pmatrix}}_{\hat{\boldsymbol{h}}^{(0)}_{\nu+M}} - \underbrace{\begin{pmatrix} \mathbf{h}^{(0)}_{\nu+M} \\ \vdots \\ \mathbf{h}^{(0)}_{\nu+M-Q} \end{pmatrix}}_{\boldsymbol{h}^{(0)}_{\nu+M}} \right\|_2^2 . \tag{6.3}$$

Note that by choosing M and Q large enough, an arbitrarily large subset of the total channel response $\{\mathbf{h}^{(0)}_n\}$ may be estimated regardless of the symbol estimation delay ν.

Theorem 6.1. *For symbol estimation delay ν, the ASPE generated by the blind channel identification scheme in Fig. 6.2 can be upper bounded as*

$$\mathcal{E}_{\hat{\boldsymbol{h}}} \leq \frac{\|\boldsymbol{\mathcal{H}}_M\|^2 \, J_{\mathrm{u},\nu}\big|^{\max, J_{\mathrm{u},\nu}(\boldsymbol{q}_{\mathrm{m},\nu})}_{\mathrm{c},\nu}}{(Q+1)\sigma_s^2} \tag{6.4}$$

$$\leq \frac{\mathrm{E}\left\{\|\mathbf{r}_n\|_2^2\right\} \; J_{\mathrm{u},\nu}\big|^{\max, J_{\mathrm{u},\nu}(\boldsymbol{q}_{\mathrm{m},\nu})}_{\mathrm{c},\nu}}{\sigma_s^4} \tag{6.5}$$

when the Wiener symbol estimates satisfy the UMSE condition $J_{\mathrm{u},\nu}(\boldsymbol{q}_{\mathrm{m},\nu}) \leq J_{\mathrm{o}}\sigma_s^2$ in Theorem 4.2.

When the channels and estimator are FIR, the quantity $\boldsymbol{\mathcal{H}}_M$ is a block Toeplitz matrix closely related to $\boldsymbol{\mathcal{H}}$ defined in (2.13). Specifically, $\boldsymbol{\mathcal{H}}_M$ has $P(Q+1)$ rows and N_q columns, where N_q denotes the length of $\boldsymbol{q}$, with the following structure:

$$\boldsymbol{\mathcal{H}}_M := \tag{6.6}$$
$$\begin{pmatrix} \mathbf{h}^{(0)}_{M} \cdots \mathbf{h}^{(K)}_{M} & \mathbf{h}^{(0)}_{M+1} \cdots \mathbf{h}^{(K)}_{M+1} & \cdots & \mathbf{h}^{(0)}_{M+N_q-1} \cdots \mathbf{h}^{(K)}_{M+N_q-1} \\ \mathbf{h}^{(0)}_{M-1} \cdots \mathbf{h}^{(K)}_{M-1} & \mathbf{h}^{(0)}_{M} \cdots \mathbf{h}^{(K)}_{M} & \cdots & \mathbf{h}^{(0)}_{M+N_q-2} \cdots \mathbf{h}^{(K)}_{M+N_q-2} \\ \vdots \quad\quad \vdots & \vdots \quad\quad \vdots & & \vdots \quad\quad \vdots \\ \mathbf{h}^{(0)}_{M-Q} \cdots \mathbf{h}^{(K)}_{M-Q} & \mathbf{h}^{(0)}_{M-Q+1} \cdots \mathbf{h}^{(K)}_{M-Q+1} & \cdots & \mathbf{h}^{(0)}_{M-Q+N_q-1} \cdots \mathbf{h}^{(K)}_{M-Q+N_q-1} \end{pmatrix} .$$

(Note that with causal FIR channels, many of the elements of $\boldsymbol{\mathcal{H}}_M$ in (6.6) may be zero-valued.) When any of the channels or the estimator is IIR, $\boldsymbol{\mathcal{H}}_M$ can be defined as an operator (as is done in Appendix 6.A.1).

Theorem 6.1 gives an upper bound for the ASPE that is proportional to the norm of the channel operator[2] and the UMSE of ν-delayed Wiener symbol estimates (through the definition of $J_{\mathrm{u},\nu}\big|_{\mathrm{c},\nu}^{\max, J_{\mathrm{u},\nu}(\boldsymbol{q}_{\mathrm{m},\nu})}$ in (4.11)), as well as a looser bound that is proportional to the received power and the Wiener UMSE. Section 6.4 plots these upper bounds for comparison to the actual ASPE attained using the CM-minimizing estimator.

6.3 Blind Identification – Issues in Practical Implementation

6.3.1 ASPE with Finite-Data Correlation Approximation

In practice, the expectation operations in Fig. 6.2 will be replaced by some sort of block or exponential averages. In this section, we analyze the effect of block averaging on the parameter estimation error. The δ^{th} block parameter estimate is defined below for block size N.

$$\hat{\hat{\mathbf{h}}}^{(0)}_{\nu+M-\delta} \;:=\; \frac{1}{N}\sum_{n=0}^{N-1} \mathbf{r}_{n-\delta}\, y^*_{n-M}. \tag{6.7}$$

Lemma 6.1. *The expected ASPE using N-block estimates of the autocorrelations*

[2]When $\boldsymbol{\mathcal{H}}_M$ is a finite-dimensional matrix, $\|\boldsymbol{\mathcal{H}}_M\|$ equals the maximum singular value of $\boldsymbol{\mathcal{H}}_M$.

can be written

$$
\begin{aligned}
\mathrm{E}\{\mathcal{E}_{\hat{\hat{\boldsymbol{h}}}}\} \;&=\; \mathcal{E}_{\hat{\boldsymbol{h}}} + \frac{\sigma_s^4}{N(Q+1)} \frac{|\hat{\boldsymbol{h}}_{\nu+M}^{(0)H} \boldsymbol{h}_{\nu+M}^{(0)}|^2}{\big\|\hat{\boldsymbol{h}}_{\nu+M}^{(0)}\big\|_2^4} \\
&\sum_{\delta=0}^{Q} \left(\Big\| \sum_{k,i} \sqrt{\kappa_s^{(k)} - 1}\, \mathbf{h}_{i+M-\delta}^{(k)} q_i^{(k)*} \Big\|_2^2 + \sum_{p,k,j} \mathbf{h}_j^{(k)H} \mathbf{h}_{j+p}^{(k)} \sum_{(\ell,i)\neq(k,j+M-\delta)} q_i^{(\ell)} q_{i+p}^{(\ell)*} \right). \qquad (6.8)
\end{aligned}
$$

Simulations suggest that, for CM-minimizing estimators $\mathbf{f}_{c,\nu}(z)$ and typical values of N, the second term in (6.8) dominates the first. This implies that the performance of the proposed channel estimation scheme is in practice limited by the finite-data correlations rather than by the performance of the blind symbol estimates. The plots in Section 6.4 agree with this notion: improvement in symbol estimates gained through quantization of $\{y_n\}$ gives the Gooch-Harp scheme [Gooch ICC 88] only minor advantage in ASPE.

6.3.2 Stochastic Gradient Estimation of CM Equalizer

Practical implementations of the identification scheme in Fig. 6.2 will not have knowledge of the exact CM-minimizing symbol estimator $\mathbf{f}_{c,\nu}(z)$. Typically, $\mathbf{f}_{c,\nu}(z)$ will be replaced by an iteratively updated approximation to $\mathbf{f}_{c,\nu}(z)$ generated by the constant modulus algorithm (CMA), an algorithm which attempts stochastic gradient minimization of CM cost [Godard TCOM 80, Treichler TASSP 83]. For finite-length $\mathbf{f}(z)$, CMA updates the equalizer parameters $\{\mathbf{f}_0, \ldots, \mathbf{f}_{N_f-1}\}$ using the following rule (where n denotes the time step).

$$
\mathbf{f}_i(n+1) \;=\; \mathbf{f}_i(n) + \mu \mathbf{r}_{n-i} y_n^* \big(\gamma - |y_n|^2\big), \qquad 0 \le i \le N_f - 1. \qquad (6.9)
$$

Similar update rules can be derived for equalizer structures that employ feedback (resulting in an IIR equalizer) [Endres SPAWC 99]. In (6.9), μ is a small positive stepsize.

The operation of CMA can be considered as a two-stage process. Starting from an initialization $\mathbf{f}_{\mathrm{i}}(z)$, the CMA-updated estimator $\mathbf{f}(z)$ first converges to a neighborhood of the exact CM-minimizing estimator $\mathbf{f}_{\mathrm{c},\nu}(z)$ associated with some combination of source k and symbol estimation delay ν. The particular {source, delay}, as well as the time to convergence, depend on the initialization $\mathbf{f}_{\mathrm{i}}(z)$. Though various initialization procedures have been proposed (see, e.g., the references in [Johnson PROC 98]), none are known to work "perfectly" in all situations. In Chapter 5 of this thesis, however, we have shown that if the signal to interference-plus-noise (SINR) ratio of the estimates generated by $\mathbf{f}_{\mathrm{i}}(z)$ is above a prescribed threshold, then small stepsize CMA will converge to a neighborhood of the CM-minimum $\mathbf{f}_{\mathrm{c},\nu}(z)$ associated with the same {source, delay} as $\mathbf{f}_{\mathrm{i}}(z)$. Furthermore, for i.i.d. sub-Gaussian sources in the presence of AWGN, we have shown that this SINR threshold equals 3.8 dB.

Once the CMA-updated equalizer parameters have converged to a neighborhood of the local CM minimum $\mathbf{f}_{\mathrm{c},\nu}(z)$, averaging theory predicts that the CMA-updated equalizer trajectory converges almost surely and in mean to $\mathbf{f}_{\mathrm{c},\nu}(z)$ [Ljung Book 99]. In practical situations, however, the CMA-updated equalizer will "jitter" around this local minimum, where the amount of jitter is proportional to the stepsize μ and to the average size of the error term $\left(\gamma - |y_n|^2\right)$ in (6.9). It is possible to derive expressions for the *excess* MSE due to CMA, i.e., the difference between the expected MSE of CMA-generated symbol estimates and the MSE of CM-minimizing estimates. For example, CMA's excess MSE resulting from the use of source symbols drawn from a non-constant-modulus alphabet is characterized in [Fijalkow TSP 98]. The simulations in Section 6.4, however, seem to indicate that for typical block sizes N, the effects of finite-data correlation approximation overwhelm the effects of CMA-induced excess symbol estimation error. For this reason, we do not further investigate CMA-induced error.

Throughout our discussion of blind channel estimation we have been assuming that the CMA-derived symbol estimates are reasonably accurate, which would seem to require use of a data record long enough to support the convergence time of CMA. Even with very short data records, however, it may be possible to adapt CMA using a repeatedly concatenated version of the same data record (similar to "looped" LMS [Qureshi PROC 85]). Since source independence assumptions S1)–S2) become less valid as record length decreases, however, it is difficult to make solid claims about the convergence of such data-reusing CMA schemes. Though of practical importance, CMA data-reuse lies outside the scope of this thesis.

6.3.3 Effect of Residual Carrier Offset

In passband communication systems, it is typically the case that the receiver's down-conversion frequency is not exactly equal to the transmitter's up-conversion frequency, so that the quasi-baseband received signal $\{\mathbf{r}_n\}$ contains residual carrier frequency offset [Proakis Book 95]. This phenomenon changes the received signal model (2.6), shown in Fig. 6.1, to (6.10), shown in Fig. 6.4, for an offset frequency of ω radians per symbol-period.

$$\mathbf{r}_n \;=\; \sum_{k=0}^{K}\sum_{i=0}^{\infty}\mathbf{h}_i^{(k)} s_{n-i}^{(k)} e^{j\omega n}. \tag{6.10}$$

Throughout this subsection, we assume that the channel, equalizer, and source quantities are all complex-valued and use j to denote $\sqrt{-1}$.

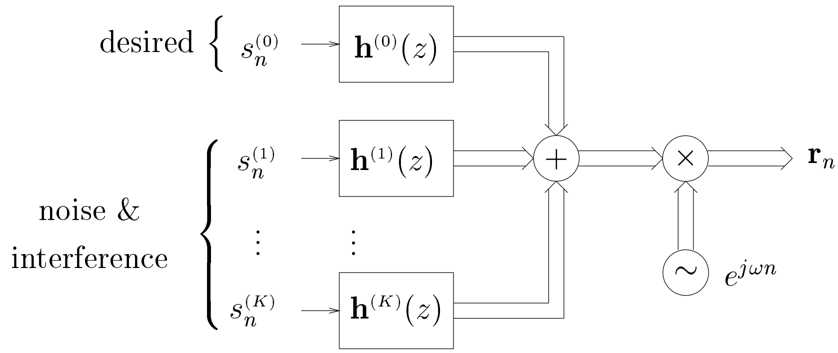

Figure 6.4: Linear system model with residual carrier offset

From (6.10), the symbol estimates become

$$\begin{aligned} y_n &= \sum_{m=-\infty}^{\infty} \mathbf{f}_m^H \mathbf{r}_{n-m} \\ &= \sum_{m,i,k} \mathbf{f}_m^H \mathbf{h}_i^{(k)} s_{n-m-i}^{(k)} e^{j\omega(n-m)} \\ &= e^{j\omega n} \sum_{k,i} \underbrace{\left(\sum_m \mathbf{f}_m^H e^{-j\omega m} \mathbf{h}_{i-m}^{(k)} \right)}_{\tilde{q}_i^{(k)}} s_{n-i}^{(k)}, \end{aligned} \tag{6.11}$$

which implies that the identification scheme in Fig. 6.2 returns the following channel parameter estimates.

$$\begin{aligned} \hat{\mathbf{h}}_{\nu+M-\delta}^{(0)} &= \mathrm{E}\left\{ \mathbf{r}_{n-\delta} y_{n-M}^* \right\} \\ &= \mathrm{E}\left\{ \sum_{k,p} \mathbf{h}_p^{(k)} s_{n-\delta-p}^{(k)} e^{j\omega(n-\delta)} \cdot e^{-j\omega n} \sum_{\ell,i} \tilde{q}_i^{(\ell)*} s_{n-M-i}^{(\ell)*} \right\} \\ &= \sigma_s^2 e^{-j\omega\delta} \sum_{k,i} \mathbf{h}_{i+M-\delta}^{(k)} \, \tilde{q}_i^{(k)*}. \end{aligned} \tag{6.12}$$

From (2.35) it is apparent that the CM criterion is insensitive to symbol estimate phase. Equation (6.11) then implies that $\tilde{\boldsymbol{q}}_{\mathrm{c},\nu}$ in the presence of carrier offset should equal $\boldsymbol{q}_{\mathrm{c},\nu}$ in the absence of offset (via $\mathbf{f}_{\mathrm{c},\nu}(z)$ in the presence of offset equaling $\mathbf{f}_{\mathrm{c},\nu}(z)$ in the absence of offset after rotation of the m^{th} impulse response coefficient by $e^{j\omega m}$).

A comparison of (6.1) and (6.12) reveals that when $\tilde{\boldsymbol{q}} = \boldsymbol{q}$ (as for CM-minimizing channel-equalizer responses), the only effect of carrier offset is a rotation of the δ^{th} channel parameter estimate $\hat{\mathbf{h}}^{(0)}_{\nu+M-\delta}$ by $e^{-j\omega\delta}$. Thus, if $(Q+1)\omega$ is small, the increase in ASPE caused by carrier frequency offset ω will be insignificant. (Recall that $Q+1$ denotes the time span, in symbols, of the set of channel coefficients that is being estimated.)

6.4 Numerical Examples

Figs. 6.5 and 6.6 each plot bounds (6.4) and (6.5) for the ASPE of the exact CM-minimizing estimator with exact cross-correlations compared to (i) the average ASPE achieved by the proposed CMA-based scheme using block length $N = 10^4$, (ii) the average ASPE achieved by the Gooch-Harp scheme [Gooch ICC 88] using block length N, (iii) the expected ASPE for the exact CM-minimizing estimator[3] using block length N (from (6.8)), and (iv) the ASPE for the exact CM-minimizing estimator with exact cross-correlations (from (6.2)).

Fig. 6.5 is based on a complex-valued $T/2$-spaced (i.e., $P = 2$) Signal Processing Information Base (SPIB) microwave channel response model #3, shortened to a 16 symbol duration, in various levels of additive white Gaussian noise (AWGN). The impulse response of channel #3 is shown in Fig. 6.7(a)-(b) and example channel estimation errors are plotted in Fig. 6.7(c)-(d). The complex-valued $T/2$-spaced equalizer $\mathbf{f}(z)$ had a time support of 10 symbols.

Fig. 6.6 is based on SPIB channel #2 and a number of different restrictions on symbol estimator length N_f. The impulse response of channel #2 is shown in

[3]The CM-minimizing estimator $\mathbf{f}_{c,\nu}(z)$ was determined numerically using Matlab's "fminunc" routine.

Fig. 6.8(a)-(b) and example channel estimation errors are plotted in Fig. 6.8(c)-(d). The SNR of AWGN was 40 dB.

The following were common to all experiments: the desired source was i.i.d. and drawn from a 64-QAM alphabet; ν was the MSE-minimizing symbol delay for the particular combination of channel, noise, and equalizer constraints; Q and M were adjusted so that all 32 $T/2$-spaced coefficients of the SPIB channel were estimated; and CMA was initialized at $\mathbf{f}_{c,\nu}(z)$ and adapted with stepsize $\mu = 10^{-3}$.

The fact that the upper-bound-(6.4) trace crosses the N-block traces in Figs. 6.5 and 6.6 should not cause alarm: (6.4) bounds the ASPE assuming *exact* cross-correlations, while the CMA, Gooch-Harp, and CM-N traces assume length-N block approximations of the cross-correlations.

6.5 Conclusions

We have analyzed the performance of a blind channel identification scheme based on the cross-correlation of CM-minimizing blind symbol estimates with the received signal. Leveraging recent results on the unbiased MSE of CM-minimizing symbol estimates, upper bounds on the average squared channel parameter estimation error (ASPE) were derived. Various implementational aspects were also considered, such as ASPE increase due to finite-data correlator approximations, stochastic gradient estimates of the CM-minimizing equalizer, and residual carrier frequency offset. Finally, experiments using SPIB microwave channel models were presented to verify the results of our analyses.

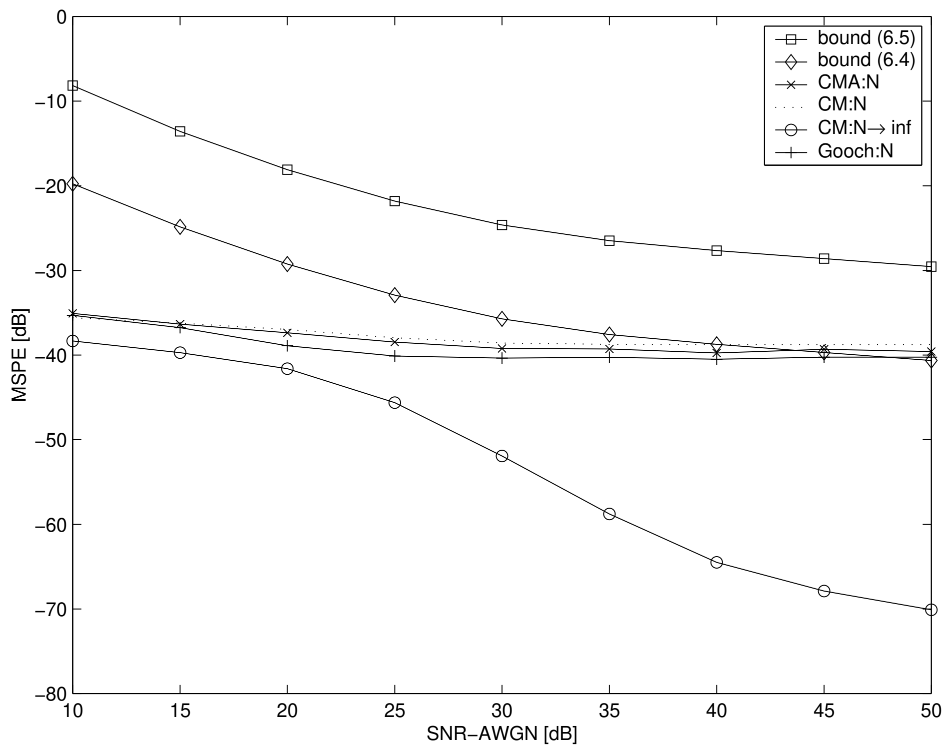

Figure 6.5: Mean-squared parameter error for 32-tap $T/2$-spaced SPIB microwave channel #3 versus SNR of AWGN.

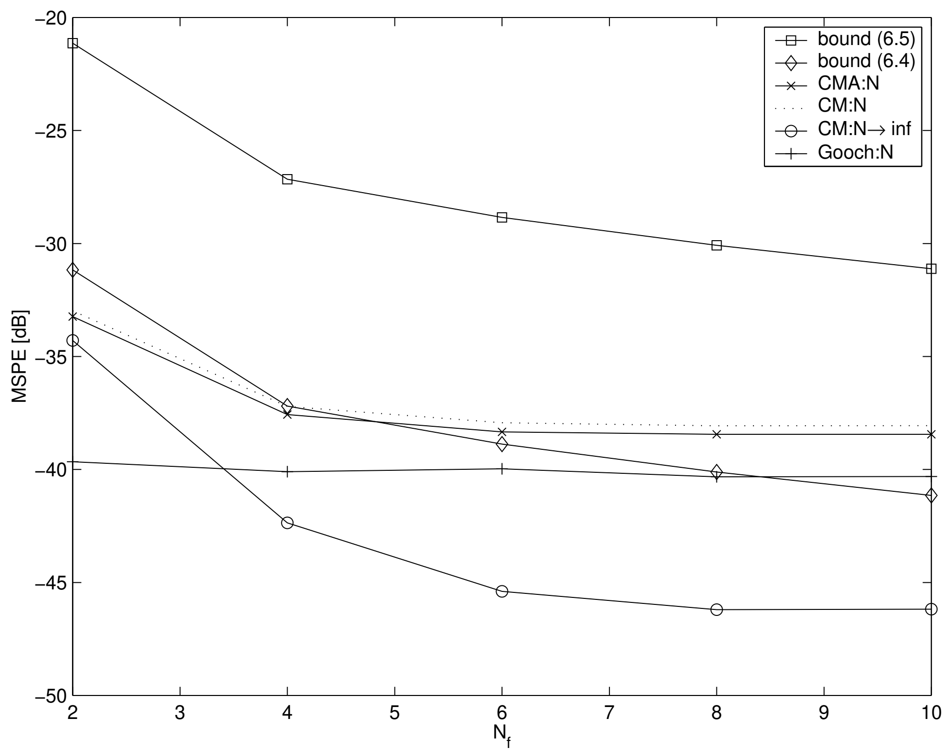

Figure 6.6: Mean-squared parameter error for 32-tap $T/2$-spaced SPIB microwave channel #3 versus symbol estimator length N_f.

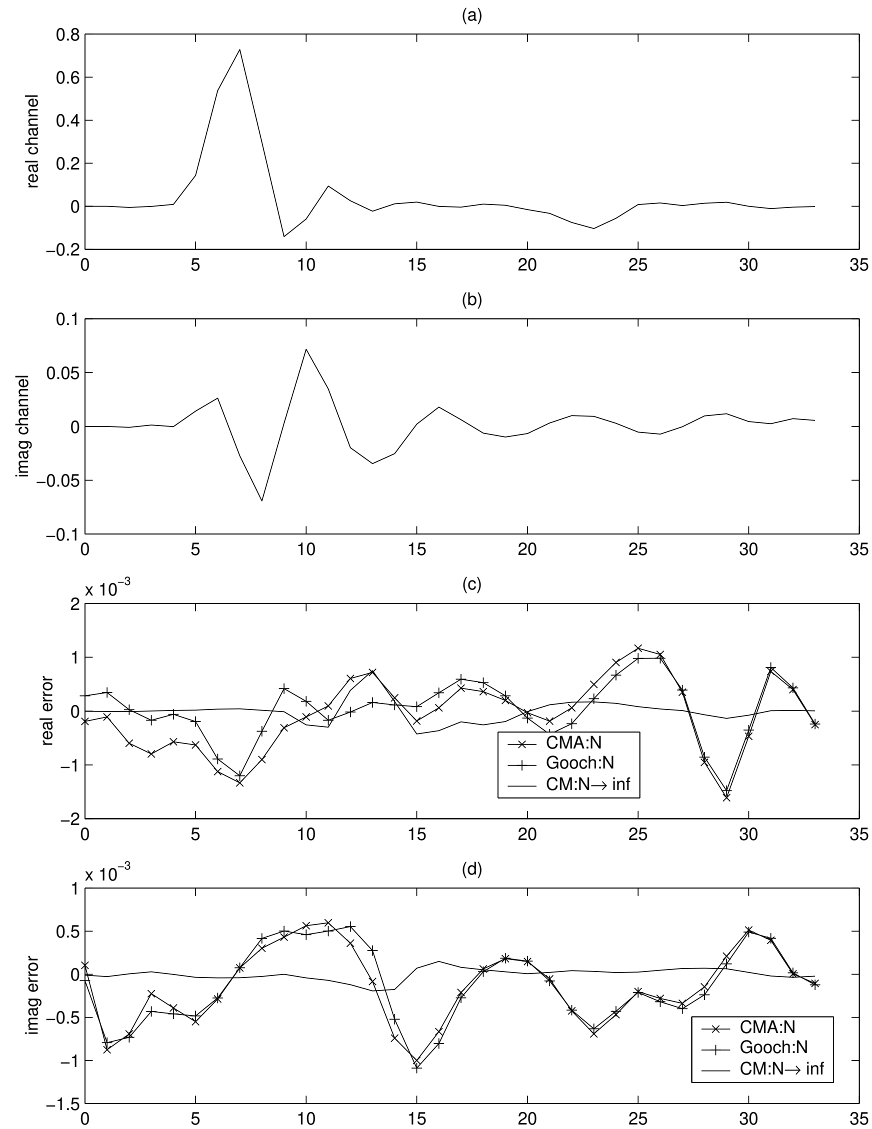

Figure 6.7: (a) Real and (b) imaginary components of 32-tap $T/2$-spaced SPIB microwave channel #3 impulse response. (c) Real and (d) imaginary components of estimation errors for SNR = 50 dB and $N_f = 10$.

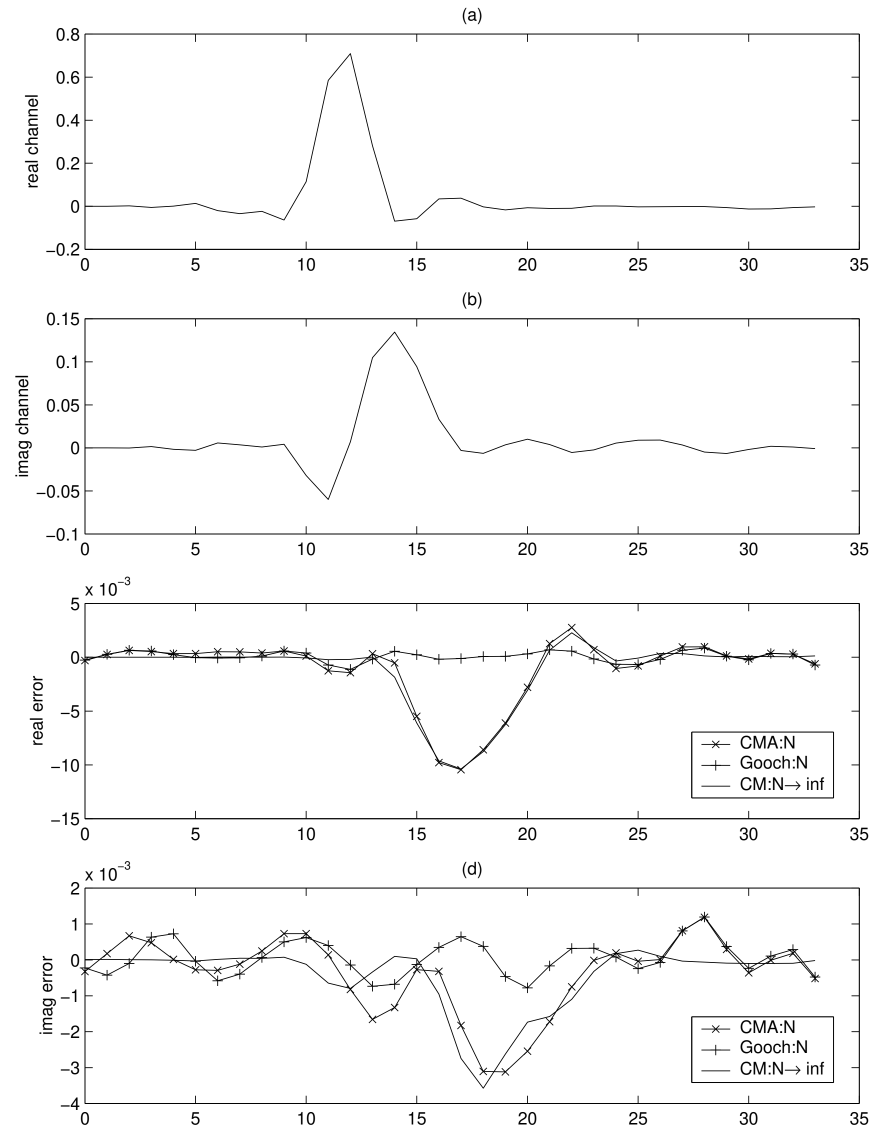

Figure 6.8: (a) Real and (b) imaginary components of 32-tap $T/2$-spaced SPIB microwave channel #2 impulse response. (c) Real and (d) imaginary components of estimation errors for SNR $= 40$ dB and $N_f = 10$.

Appendix

6.A Derivation Details for Channel Identification Bounds

This appendix contains proofs of the theorem and lemma found in Section 6.2 and Section 6.3, respectively.

6.A.1 Proof of Theorem 6.1

To ensure that our bound applies to both finite and infinite-dimensional channels and estimators, we formalize the definition of the operator $\boldsymbol{\mathcal{H}}$ (whose finite-dimensional definition appeared in (6.6)).

$$\boldsymbol{\mathcal{H}}_{M} : \ell_1(\mathbb{C}^P) \to \mathbb{C}^{P(Q+1)} \quad \text{s.t.} \quad \boldsymbol{\mathcal{H}}_{M}\,\boldsymbol{q} \;=\; \begin{pmatrix} \sum_{k,i} \mathbf{h}_{i+M}^{(k)} q_i^{(k)} \\ \sum_{k,i} \mathbf{h}_{i+M-1}^{(k)} q_i^{(k)} \\ \vdots \\ \sum_{k,i} \mathbf{h}_{i+M-Q}^{(k)} q_i^{(k)} \end{pmatrix}. \tag{6.13}$$

Recall that $\boldsymbol{q}$ and $\ell_1(\mathbb{C}^P)$ were defined in Section 2.2. When $\mathbf{h}^{(k)}(z)$ are FIR, $\boldsymbol{\mathcal{H}}_{M}$ reduces to a block Toeplitz matrix. The operator $\bar{\boldsymbol{\mathcal{H}}}_{M}$ will also prove useful:

$$\bar{\boldsymbol{\mathcal{H}}}_{M} : \ell_1(\mathbb{C}^P) \to \mathbb{C}^{P(Q+1)} \quad \text{s.t.} \quad \bar{\boldsymbol{\mathcal{H}}}_{M}\,\bar{\boldsymbol{q}} \;=\; \begin{pmatrix} \sum_{(k,i)\neq(0,\nu)} \mathbf{h}_{i+M}^{(k)} q_i^{(k)} \\ \sum_{(k,i)\neq(0,\nu)} \mathbf{h}_{i+M-1}^{(k)} q_i^{(k)} \\ \vdots \\ \sum_{(k,i)\neq(0,\nu)} \mathbf{h}_{i+M-Q}^{(k)} q_i^{(k)} \end{pmatrix}. \tag{6.14}$$

$\bar{\boldsymbol{\mathcal{H}}}_{M}$ is a version of $\boldsymbol{\mathcal{H}}_{M}$ with the components for the 0^{th} source at delay ν removed, and $\bar{\boldsymbol{q}}$ is a version of $\boldsymbol{q}$ with the $q_\nu^{(0)}$ element extracted.

Using (6.1) and the definitions of $\hat{\boldsymbol{h}}^{(0)}_{\nu+M}$ and $\boldsymbol{h}^{(0)}_{\nu+M}$ in (6.3), the operators $\boldsymbol{\mathcal{H}}_{M}$ and $\bar{\boldsymbol{\mathcal{H}}}_{M}$ allow us to write

$$\begin{aligned}
\hat{\boldsymbol{h}}^{(0)}_{\nu+M} &= \boldsymbol{\mathcal{H}}_{M}\, \boldsymbol{q}^* \sigma_s^2 \\
&= \bar{\boldsymbol{\mathcal{H}}}_{M}\, \bar{\boldsymbol{q}}^* \sigma_s^2 + \boldsymbol{h}^{(0)}_{\nu+M} q_\nu^{(0)*} \sigma_s^2.
\end{aligned} \tag{6.15}$$

Choosing $\theta = (q_\nu^{(0)*}\sigma_s^2)^{-1}$ in (6.3),

$$\mathcal{E}_{\hat{\boldsymbol{h}}} \;\le\; \frac{1}{Q+1} \left\| \frac{\hat{\boldsymbol{h}}^{(0)}_{\nu+M}}{q_\nu^{(0)*}\sigma_s^2} - \boldsymbol{h}^{(0)}_{\nu+M} \right\|_2^2 \;=\; \frac{1}{Q+1} \frac{\left\| \bar{\boldsymbol{\mathcal{H}}}_{M}\, \bar{\boldsymbol{q}}^* \right\|_2^2}{|q_\nu^{(0)}|^2}. \tag{6.16}$$

Definition of the induced norm

$$\|\boldsymbol{\mathcal{H}}_{M}\| \;:=\; \sup_{\boldsymbol{q}\neq \boldsymbol{0}} \frac{\|\boldsymbol{\mathcal{H}}_{M}\, \boldsymbol{q}^*\|_2}{\|\boldsymbol{q}\|_2} \tag{6.17}$$

(which, for finite-dimensional $\boldsymbol{\mathcal{H}}_{M}$, equals the largest singular value) allows further bounding of (6.16):

$$\mathcal{E}_{\hat{\boldsymbol{h}}} \;\le\; \frac{1}{Q+1} \frac{\|\bar{\boldsymbol{\mathcal{H}}}_{M}\|^2 \|\bar{\boldsymbol{q}}\|_2^2}{|q_\nu^{(0)}|^2}. \tag{6.18}$$

Recalling the definition of UMSE in (2.31), equation (6.18) becomes

$$\mathcal{E}_{\hat{\boldsymbol{h}}} \;\le\; J_{\mathrm{u},\nu}(\boldsymbol{q}_{\mathrm{c},\nu}) \frac{\|\bar{\boldsymbol{\mathcal{H}}}_{M}\|^2}{(Q+1)\sigma_s^2}. \tag{6.19}$$

When $\{\mathbf{h}^{(k)}(z)\}$, $\mathcal{Q}_{\mathrm{a}}$, and ν are such that $J_{\mathrm{u},\nu}(\boldsymbol{q}_{\mathrm{m},\nu}) \le J_{\mathrm{o}}\sigma_s^2$ for J_{o} in (4.10), Theorem 4.2 allows upper bounding of $J_{\mathrm{u},\nu}(\boldsymbol{q}_{\mathrm{c},\nu})$ and (6.19) becomes

$$\mathcal{E}_{\hat{\boldsymbol{h}}} \;\le\; J_{\mathrm{u},\nu}\Big|_{\mathrm{c},\nu}^{\max, J_{\mathrm{u},\nu}(\boldsymbol{q}_{\mathrm{m},\nu})} \frac{\|\bar{\boldsymbol{\mathcal{H}}}_{M}\|^2}{(Q+1)\sigma_s^2}. \tag{6.20}$$

Since

$$\begin{aligned}
\|\boldsymbol{\mathcal{H}}_{M}\|^2 &= \sup_{\|\boldsymbol{q}\|_2=1} \boldsymbol{q}^t \boldsymbol{\mathcal{H}}_{M}^H \boldsymbol{\mathcal{H}}_{M}\, \boldsymbol{q}^* \\
&\ge \sup_{\|\boldsymbol{q}\|_2=1,\, q_\nu^{(0)}=0} \boldsymbol{q}^t \boldsymbol{\mathcal{H}}_{M}^H \boldsymbol{\mathcal{H}}_{M}\, \boldsymbol{q}^* \\
&= \sup_{\|\bar{\boldsymbol{q}}\|_2=1} \bar{\boldsymbol{q}}^t \bar{\boldsymbol{\mathcal{H}}}_{M}^H \bar{\boldsymbol{\mathcal{H}}}_{M}\, \bar{\boldsymbol{q}}^* \\
&= \|\bar{\boldsymbol{\mathcal{H}}}_{M}\|^2,
\end{aligned}$$

(6.20) yields

$$\mathcal{E}_{\hat{\boldsymbol{h}}} \;\leq\; J_{\mathrm{u},\nu}\Big|_{\mathrm{c},\nu}^{\max, J_{\mathrm{u},\nu}(\boldsymbol{q}_{\mathrm{m},\nu})} \frac{\|\boldsymbol{\mathcal{H}}_M\|^2}{(Q+1)\sigma_s^2}. \tag{6.21}$$

Simplification of (6.21) is possible using the fact that

$$\begin{aligned}
\|\boldsymbol{\mathcal{H}}_M \boldsymbol{q}^*\|_2^2 &= \sum_{\delta=0}^{Q} \Big\| \sum_{k,i} \mathbf{h}_{i+M-\delta}^{(k)} q_i^{(k)*} \Big\|_2^2 \\
&\leq \sum_{\delta=0}^{Q} \Big(\sum_{k,i} \|\mathbf{h}_{i+M-\delta}^{(k)} q_i^{(k)*}\|_2^2 \Big) \\
&= \sum_{\delta=0}^{Q} \sum_{k,i} \|\mathbf{h}_{i+M-\delta}^{(k)}\|_2^2 \, |q_i^{(k)}|^2 \\
&\leq (Q+1) \sum_{k,i} \|\mathbf{h}_i^{(k)}\|_2^2 \, \|\boldsymbol{q}\|_2^2
\end{aligned}$$

which implies

$$\|\boldsymbol{\mathcal{H}}_M\|^2 \;\leq\; (Q+1) \sum_{k,i} \|\mathbf{h}_i^{(k)}\|_2^2. \tag{6.22}$$

Rewriting (6.22) using

$$\sum_{k,i} \|\mathbf{h}_i^{(k)}\|_2^2 \;=\; \frac{1}{\sigma_s^2} \mathrm{E}\Big\{ \Big\| \sum_{k,i} \mathbf{h}_i^{(k)} s_{n-i}^{(k)} \Big\|_2^2 \Big\} \;=\; \frac{1}{\sigma_s^2} \mathrm{E}\left\{ \|\mathbf{r}_n\|_2^2 \right\}$$

gives

$$\|\boldsymbol{\mathcal{H}}_M\|_2^2 \;\leq\; \frac{Q+1}{\sigma_s^4} \mathrm{E}\left\{ \|\mathbf{r}_n\|_2^2 \right\}, \tag{6.23}$$

which, combined with (6.21), leads to (6.5).

6.A.2 Proof of Lemma 6.1

The mean of the block parameter estimate is

$$\mathrm{E}\Big\{ \hat{\hat{\mathbf{h}}}_{\nu+M-\delta}^{(0)} \Big\} \;=\; \mathrm{E}\{\mathbf{r}_{n-\delta} y_{n-M}^*\} \;=\; \hat{\mathbf{h}}_{\nu+M-\delta}^{(0)}, \tag{6.24}$$

implying a variance of the form

$$\mathrm{var}\left(\hat{\hat{\mathbf{h}}}^{(0)}_{\nu+M-\delta}\right) = \mathrm{E}\left\{\left\|\hat{\hat{\mathbf{h}}}^{(0)}_{\nu+M-\delta} - \hat{\mathbf{h}}^{(0)}_{\nu+M-\delta}\right\|_2^2\right\}.$$

Rewriting (6.7) using expressions from Section 2.2

$$\begin{aligned}
\hat{\hat{\mathbf{h}}}^{(0)}_{\nu+M-\delta} &= \frac{1}{N}\sum_{n=0}^{N-1}\sum_{k,j}\mathbf{h}_j^{(k)} s^{(k)}_{n-\delta-j}\sum_{\ell,i} q_i^{(\ell)*} s^{(\ell)*}_{n-M-i} \\
&= \sum_{k,j,\ell,i}\mathbf{h}_j^{(k)} q_i^{(\ell)*}\frac{1}{N}\sum_{n=0}^{N-1} s^{(k)}_{n-\delta-j} s^{(\ell)*}_{n-M-i} \\
&= \sum_{k,i}\mathbf{h}^{(k)}_{i+M-\delta} q_i^{(k)*}\frac{1}{N}\sum_{n=0}^{N-1}|s^{(k)}_{n-M-i}|^2 \\
&\quad + \sum_{k,j}\sum_{(\ell,i)\neq(k,j+M-\delta)}\mathbf{h}_j^{(k)} q_i^{(\ell)*}\frac{1}{N}\sum_{n=0}^{N-1} s^{(k)}_{n-\delta-j} s^{(\ell)*}_{n-M-i}
\end{aligned} \tag{6.25}$$

and recalling (6.1), the variance can be rewritten as

$$\begin{aligned}
&\mathrm{var}\left(\hat{\hat{\mathbf{h}}}^{(0)}_{\nu+M-\delta}\right) \\
&= \mathrm{E}\left\{\left\|\sum_{k,i}\mathbf{h}^{(k)}_{i+M-\delta} q_i^{(k)*}\left(\frac{1}{N}\sum_{n=0}^{N-1}|s^{(k)}_{n-M-i}|^2 - \sigma_s^2\right)\right.\right. \\
&\qquad\left.\left. + \sum_{k,j}\sum_{(\ell,i)\neq(k,j+M-\delta)}\mathbf{h}_j^{(k)} q_i^{(\ell)*}\frac{1}{N}\sum_{n=0}^{N-1} s^{(k)}_{n-\delta-j} s^{(\ell)*}_{n-M-i}\right\|_2^2\right\}.
\end{aligned} \tag{6.26}$$

It can be shown that

$$\mathrm{E}\left\{\left(\frac{1}{N}\sum_{n=0}^{N-1}|s_n^{(k)}|^2 - \sigma_s^2\right)^2\right\} = \frac{(\kappa_s^{(k)} - 1)}{N}\sigma_s^4, \tag{6.27}$$

$$\mathrm{E}\left\{\left(\frac{1}{N}\sum_{n=0}^{N-1}|s^{(k)}_{n-a}|^2 - \sigma_s^2\right)\frac{1}{N}\sum_{m=0}^{N-1} s^{(k)*}_{m-j} s^{(\ell)}_{m-i}\right\} = \begin{cases}\text{nonzero} & (k,j) = (\ell,i) \\ 0 & \text{else,}\end{cases} \tag{6.28}$$

and

$$
\begin{aligned}
&\mathrm{E}\left\{\left(\frac{1}{N}\sum_{n=0}^{N-1} s_{n-b}^{(a)} s_{n-d}^{(c)*}\right)\left(\frac{1}{N}\sum_{m=0}^{N-1} s_{m-j}^{(k)*} s_{m-i}^{(\ell)}\right)\right\} \\
&= \begin{cases} \text{nonzero} & (a,b)=(c,d) \;\; \text{and} \;\; (k,j)=(\ell,i) \\ \dfrac{\sigma_s^4}{N} & \exists\, p \;\; \text{s.t.} \;\; (a,b)=(k,j{+}p) \;\; \neq \;\; (c,d)=(\ell,i{+}p) \\ 0 & \text{else.} \end{cases}
\end{aligned}
\tag{6.29}
$$

Plugging (6.27)-(6.29) into (6.26), we see that the terms labeled "nonzero" in (6.27)-(6.29) do not enter into the variance expression. Collected, the remaining terms yield (6.30).

$$
\begin{aligned}
\operatorname{var}\!\left(\hat{\hat{\mathbf{h}}}_{\nu+M-\delta}^{(0)}\right) &= \frac{\sigma_s^4}{N}\left(\left\|\sum_{k,i}\sqrt{\kappa_s^{(k)}-1}\,\mathbf{h}_{i+M-\delta}^{(k)} q_i^{(k)*}\right\|_2^2\right. \\
&\qquad \left. + \sum_p \sum_{k,j} \mathbf{h}_j^{(k)H}\mathbf{h}_{j+p}^{(k)} \sum_{(\ell,i)\neq(k,j+M-\delta)} q_i^{(\ell)} q_{i+p}^{(\ell)*}\right).
\end{aligned}
\tag{6.30}
$$

In terms of the N-block variance (6.30), the expected block-MSPE can be written

$$
\begin{aligned}
\mathrm{E}\{\mathcal{E}_{\hat{\boldsymbol{h}}}\} &= \mathrm{E}\left\{\frac{1}{Q+1}\sum_{\delta=0}^{Q}\left\|\theta_{\hat{\boldsymbol{h}}}\hat{\hat{\mathbf{h}}}_{\nu+M-\delta}^{(0)} - \mathbf{h}_{\nu+M-\delta}^{(0)}\right\|_2^2\right\} \\
&= \mathrm{E}\left\{\frac{1}{Q+1}\sum_{\delta=0}^{Q}\left(\left\|\theta_{\hat{\boldsymbol{h}}}\hat{\hat{\mathbf{h}}}_{\nu+M-\delta}^{(0)} - \theta_{\hat{\boldsymbol{h}}}\,\mathrm{E}\left\{\hat{\hat{\mathbf{h}}}_{\nu+M-\delta}^{(0)}\right\}\right\|_2^2\right.\right. \\
&\qquad \left.\left. + \left\|\theta_{\hat{\boldsymbol{h}}}\,\mathrm{E}\left\{\hat{\hat{\mathbf{h}}}_{\nu+M-\delta}^{(0)}\right\} - \mathbf{h}_{\nu+M-\delta}^{(0)}\right\|_2^2\right)\right\} \\
&= \frac{1}{Q+1}\sum_{\delta=0}^{Q}\left(|\theta_{\hat{\boldsymbol{h}}}|^2\,\mathrm{E}\left\{\left\|\hat{\hat{\mathbf{h}}}_{\nu+M-\delta}^{(0)} - \hat{\mathbf{h}}_{\nu+M-\delta}^{(0)}\right\|_2^2\right\}\right. \\
&\qquad \left. + \left\|\theta_{\hat{\boldsymbol{h}}}\hat{\mathbf{h}}_{\nu+M-\delta}^{(0)} - \mathbf{h}_{\nu+M-\delta}^{(0)}\right\|_2^2\right) \\
&= \mathcal{E}_{\hat{\boldsymbol{h}}} + \frac{|\theta_{\hat{\boldsymbol{h}}}|^2}{Q+1}\sum_{\delta=0}^{Q}\operatorname{var}\!\left(\hat{\hat{\mathbf{h}}}_{\nu+M-\delta}^{(0)}\right)
\end{aligned}
$$

where $\theta_{\hat{\boldsymbol{h}}}$ is the value of θ minimizing $\mathcal{E}_{\hat{\boldsymbol{h}}}$ in (6.2). It is straightforward to show that

$$
\theta_{\hat{\boldsymbol{h}}} = \frac{\hat{\boldsymbol{h}}_{\nu+M}^{(0)H}\boldsymbol{h}_{\nu+M}^{(0)}}{\left\|\hat{\boldsymbol{h}}_{\nu+M}^{(0)}\right\|_2^2}.
$$

Chapter 7

Dithered Signed-Error CMA[1]

7.1 Introduction

The constant modulus algorithm (CMA), conceived independently in [Godard TCOM 80] and [Treichler TASSP 83], is a stochastic gradient algorithm minimizing the CM criterion: $J_{\rm c} = {\rm E}\big\{(|y_n|^2 - \gamma)^2\big\}$. The positive constant γ is referred to as the dispersion constant and can be chosen in accordance with the source statistics when they are known. As an iterative update algorithm of FIR estimator $\boldsymbol{f}$, CMA takes the form

$$\boldsymbol{f}(n+1) \;=\; \boldsymbol{f}(n) + \mu \boldsymbol{r}^*(n)\,\underbrace{y_n\big(\gamma - |y_n|^2\big)}_{:=\,\psi(y_n)}, \tag{7.1}$$

where μ is a (small) positive step-size. The function $\psi(\cdot)$ identified in (7.1) is referred to as the CMA error function and will appear many times throughout this chapter.

Low-cost consumer applications (e.g., HDTV) motivate blind equalization tech-

[1]Based on "Dithered Signed-Error CMA: Robust, Computationally Efficient Blind Adaptive Equalization" by P. Schniter and C.R. Johnson, Jr., which appeared in *IEEE Transactions on Signal Processing*; vol. 47, no. 6, pp. 1592-1603; June 1999. ©1999 IEEE. (With permission.)

niques requiring minimum implementation cost. Although noted for its LMS-like complexity, CMA may be further simplified by transforming the bulk of its update multiplications into sign operations [Treichler TASSP 83]. A recent study suggests, however, that straightforward implementations of signed-error CMA (SE-CMA) do not inherit the desirable robustness properties of CMA [Brown ALL 97]. In this chapter we present a simple modification of SE-CMA, based on the judicious incorporation of controlled noise (sometimes referred to as "dither"), that results in an algorithm with robustness properties closely resembling the standard (unsigned) CMA. In fact, we show that the mean behavior of dithered signed-error CMA (DSE-CMA) is *identical* to CMA under realistically achievable conditions. The anticipated drawback to this dithering is a degradation in steady-state mean-square error (MSE) performance. Hence, we derive an expression for the excess MSE (EMSE) of DSE-CMA and discuss implications on step-size and equalizer-length selection. We note in advance that the EMSE expression for DSE-CMA bears close resemblance to an analogous expression derived for CMA in [Fijalkow TSP 98].

The chapter is partitioned as follows: Section 7.2 reviews CMA, discusses computationally-efficient versions of CMA, and introduces the new algorithm. The transient and steady-state properties of DSE-CMA are studied in Section 7.3 and result in the design guidelines of Section 7.4. Simulation results based on SPIB microwave channel models are presented in Section 7.5. Section 7.5 also includes a comparison study with a different robust and computationally-constrained implementation of CMA.

For simplicity, *this chapter assumes the case of real-valued quantities*[2] *and FIR estimators*. Recall from Section 4.1 that under these assumptions, the satisfaction of

[2]Extension to the complex-valued case is straightforward—see Section 7.6.

S1)–S4) and the presence of full column rank (FCR) $\boldsymbol{\mathcal{H}}$ (in the model of Section 2.2) are sufficient to guarantee that CM minimization yields perfect blind estimates, i.e., perfect estimates modulo unavoidable gain and delay ambiguities. In addition, *we assume that the source statistic ratio* $\mathrm{E}\{|s_n^{(0)}|^4\}/\,\mathrm{E}\{|s_n^{(0)}|^2\}$ *is known* since this allows removal of the gain ambiguity via the choice $\gamma = \gamma_\star$ for

$$\gamma_\star \;=\; \frac{\mathrm{E}\{|s_n^{(0)}|^4\}}{\mathrm{E}\{|s_n^{(0)}|^2\}}. \tag{7.2}$$

7.2 Computationally Efficient CMA

Straightforward implementations of LMS-like adaptive algorithms (such as CMA) require a multiplication between the error function and every regressor element (see update equation (7.1)). Many practical applications benefit from eliminating these N_f regressor multiplies. Signed-error (SE) algorithms present one method for doing so, whereby only the sign of the error function is retained [Macchi Book 95]. When a SE algorithm is combined with a power-of-two step-size, it is possible to construct a multiply-free fixed-point implementation of the equalizer update algorithm. The subsections below discuss two versions of SE-CMA. (For the remainder of the chapter, we restrict our focus to the case where all quantities are real-valued. Extensions to the complex-valued case are briefly discussed in Section 7.6.)

7.2.1 Signed-Error CMA

The real-valued SE-CMA algorithm [Treichler TASSP 83] is specified as

$$\boldsymbol{f}(n+1) \;=\; \boldsymbol{f}(n) + \mu \boldsymbol{r}(n) \underbrace{\mathrm{sgn}\big(y_n(\gamma - y_n^2)\big)}_{:=\,\xi(y_n)}, \tag{7.3}$$

where sgn$(\cdot)$ denotes the standard signum function. Equation (7.3) defines the SE-CMA error function $\xi(\cdot)$, depicted in Fig. 7.1.

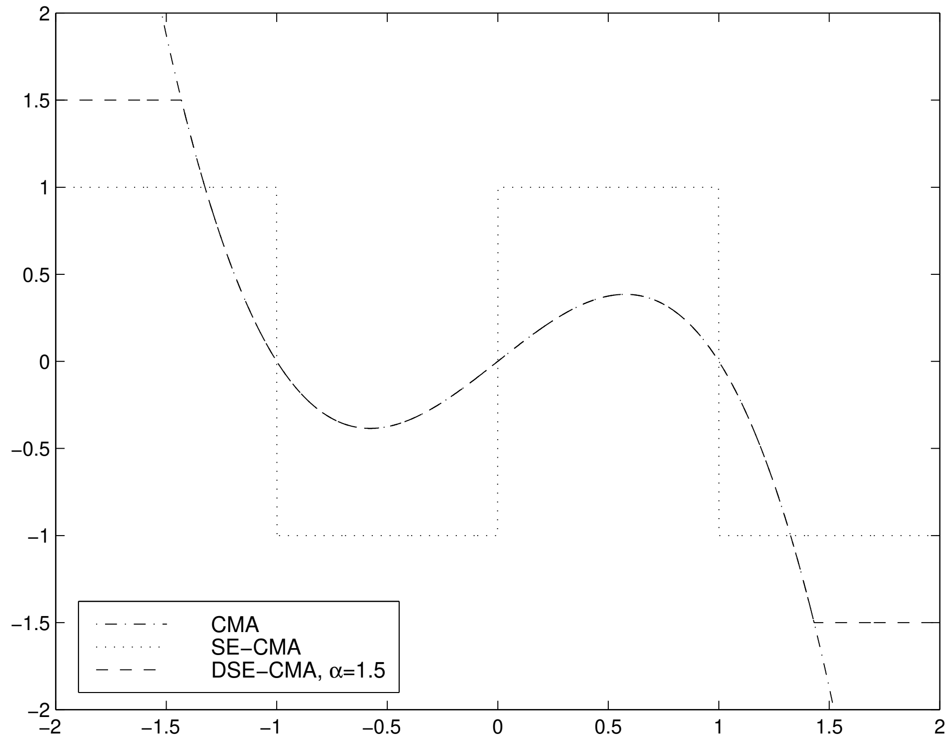

Figure 7.1: CMA, SE-CMA, and DSE-CMA error functions.

A recent investigation into SE-CMA has shown that, while satisfaction of FCR $\boldsymbol{\mathcal{H}}$ and the selection $\gamma = \gamma_\star$ ensure mean convergence to perfectly scaled blind estimates, the presence of non-FCR $\boldsymbol{\mathcal{H}}$ can severely hinder SE-CMA convergence behavior [Brown ALL 97]. Specifically, there may exist vast yet highly suboptimal regions in equalizer space in which the expected update in (7.3) is zero. Fig. 7.2 presents an example[3] of such behavior, in which the trajectory labelled "B" appears not to converge. (See Fig. 7.6 for examples of CMA trajectories under identical

[3]In Fig. 7.2, BPSK was transmitted over the following noiseless FIR vector channel: $\{\mathbf{h}_n\} = \left\{\binom{0.1}{0.3}, \binom{1}{-0.1}, \binom{0.5}{0.2}\right\}$.

conditions.) Thus, while computationally efficient, SE-CMA does not inherit the

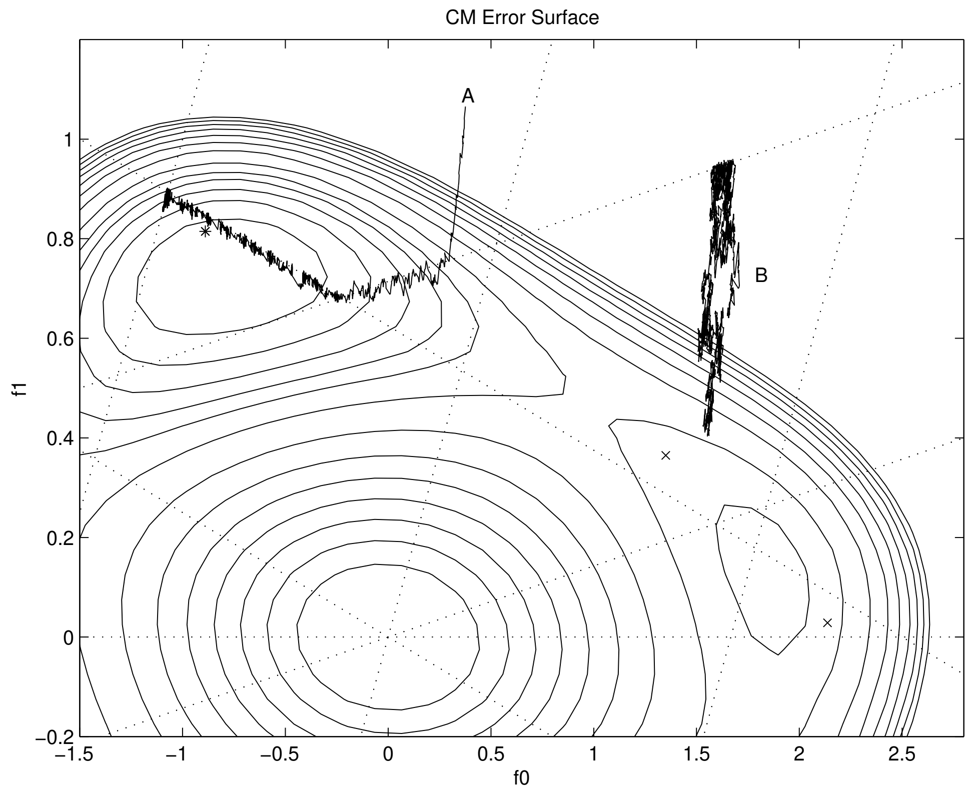

Figure 7.2: SE-CMA trajectories superimposed on J_c cost contours. Dotted lines delineate SE-CMA constant-gradient facets.

desirable robustness properties of CMA. This fact motivates the search for computationally efficient blind algorithms which *do* inherit these robustness properties. The following section describes one such algorithm.

7.2.2 Dithered Signed-Error CMA

"Gimme noise, noise, noise, noise ... "

—The Replacements, *Stink*, 1982.

Viewing the SE-CMA error function as a one bit quantizer, one might wonder whether a suitable dithering technique [Gray TIT 93] would help to remove the unwanted behavioral artifacts caused by the sign operator[4]. Dithering refers to the addition of a random signal before quantization in an attempt to preserve the information lost in the quantization process. From an additive noise perspective, dithering is an attempt to make the so-called quantization noise (see Fig. 7.3) white, zero-mean, and independent of the signal being quantized. One might expect that such quantization noise could be "averaged out" by a small step-size adaptive algorithm, yielding mean behavior identical to that of its unsigned counterpart. These ideas are made precise in Section 7.3.2.

The real-valued dithered signed-error constant modulus algorithm (DSE-CMA) is defined by the update

$$\boldsymbol{f}(n+1) \quad = \quad \boldsymbol{f}(n) + \mu\, \boldsymbol{r}(n) \underbrace{\alpha \operatorname{sgn}\big(y_n(\gamma - y_n^2) + \alpha d_n\big)}_{:= \varphi_\alpha(y_n, d_n)}, \tag{7.4}$$

where $\{d_n\}$ is an i.i.d. "dithering" process uniformly distributed on (-1,1], both γ and α are positive constants, and $\varphi_\alpha(y_n, d_n)$ is the DSE-CMA error function. The practical selection of the dispersion constant γ and the "dither amplitude" α are discussed in Section 7.4. It should become clear in the next section why α appears twice in (7.4).

In the sequel we shall see that the mean behavior of DSE-CMA closely matches that of standard (unsigned) CMA.

[4]The authors acknowledge a previous application of controlled noise to SE-LMS in the context of echo cancellation [Holte TCOM 81], [Bonnet ICASSP 84]. However, both the analyses and goals were substantially different than those in this chapter.

7.3 The Fundamental Properties of DSE-CMA

Sections 7.3.2–7.3.4 utilize an additive noise model of the dithered sign operation to characterize the transient and steady-state behaviors of DSE-CMA. Before proceeding, we present the details of this quantization noise model.

7.3.1 Quantization Noise Model of DSE-CMA

At first glance, the nonlinear sign operator in (7.4) appears to complicate the behavioral analysis of DSE-CMA. Fortunately, the theory of dithered quantizers allows us to subsume the sign operator by adopting a quantization-noise model of the DSE-CMA error function (see Fig. 7.3). Appendix 7.A collects the key results from classical quantization theory that allow us to formulate this model.

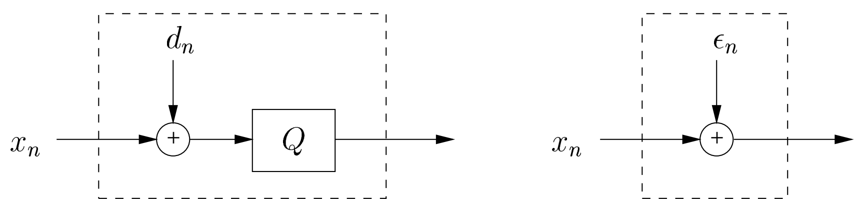

Figure 7.3: Quantization noise model (right) of the dithered quantizer (left).

DSE-CMA can be connected to the quantization literature with the observation that the operator $\alpha\,\mathrm{sgn}(\cdot)$ is identical to the two-level uniform quantizer $Q(\cdot)$, specified by

$$Q(x) \quad = \quad \begin{cases} \Delta/2 & x \geq 0, \\ -\Delta/2 & x < 0, \end{cases} \tag{7.5}$$

for quantizer spacing $\Delta = 2\alpha$. Furthermore, the specification that $\{d_n\}$ be uniformly distributed on $(-1, 1]$ ensures that $\{\alpha d_n\}$ satisfies the requirements for a valid dither process outlined in Appendix 7.A, so long as α is selected large enough to satisfy

$$\alpha \;\; \geq \;\; \big|\psi(y_n)\big| \tag{7.6}$$

for relevant values of the equalizer output y_n. Recall that $\psi(\cdot)$ denotes the CMA error function, defined in (7.1).

Employing the model of Fig. 7.3, we write the DSE-CMA error function in terms of the quantization noise ϵ_n,

$$\varphi_\alpha(y_n, d_n) \;\; = \;\; \psi(y_n) + \epsilon_n, \tag{7.7}$$

which leads to the following DSE-CMA update expression:

$$\boldsymbol{f}(n+1) \;\; = \;\; \boldsymbol{f}(n) + \mu \boldsymbol{r}(n)\big(\psi(y_n) + \epsilon_n\big). \tag{7.8}$$

When α and y_n satisfy (7.6), the properties of ϵ_n follow from equations (7.29), (7.30), and (7.32) in Appendix 7.A. Specifically, we have that ϵ_n is an uncorrelated random process whose first moment obeys

$$\mathrm{E}\big\{\epsilon_n \big| \psi(y_n)\big\} \;\; = \;\; \mathrm{E}\{\epsilon_n\} \;\; = \;\; 0, \tag{7.9}$$

and whose conditional second moment is given by

$$\mathrm{E}\big\{\epsilon_n^2 \big| \psi(y_n)\big\} \;\; = \;\; \alpha^2 - \psi^2(y_n). \tag{7.10}$$

In (7.9) and (7.10), the expectation is taken over the dither process, thus leaving a dependence on y_n.

7.3.2 DSE-CMA Transient Behavior

The average transient behavior of DSE-CMA is completely determined by the expected DSE-CMA error function, $\varphi_\alpha(y_n) := \mathrm{E}\big\{\varphi_\alpha(y_n, d_n)\big|y_n\big\}$. Equations (7.6)–(7.9) indicate that $\varphi_\alpha(\cdot)$ is a “hard-limited” version of the CMA error function,

$\psi(\cdot)$, i.e.,

$$\varphi_\alpha(y) \;=\; \begin{cases} \alpha & y : \psi(y) > \alpha, \\ \psi(y) & y : |\psi(y)| \leq \alpha, \\ -\alpha & y : \psi(y) < -\alpha. \end{cases} \tag{7.11}$$

Fig. 7.1 plots the various error functions $\varphi_\alpha(\cdot)$, $\psi(\cdot)$, and $\xi(\cdot)$ for comparison. In the theorems below, the implications of (7.11) are formalized in terms of DSE-CMA behavior over specific ranges of α.

Lemma 7.1. *Define*

$$\alpha_C \;:=\; 2(\gamma/3)^{3/2}. \tag{7.12}$$

The choice of dither amplitude $\alpha > \alpha_C$ ensures that $\varphi_\alpha(y) = \psi(y)$ for all equalizer outputs y satisfying the output amplitude constraint: $|y| \leq \psi^{-1}(\alpha)$.

Proof. By evaluating ψ at the locations where $\psi' = 0$, it can be seen that the "humps" of the cubic CMA error function (see Fig. 7.1) occur at heights $\pm 2(\gamma/3)^{3/2}$. Thus, $\psi^{-1}(\alpha)$ is unique and well-defined for $\alpha > 2(\gamma/3)^{3/2} = \alpha_{\mathrm{C}}$. Since (7.11) implies that such values of α prevent these humps from being clipped in forming the expected DSE-CMA error function, φ_α and ψ are identical over the interval $[-\psi^{-1}(\alpha), \psi^{-1}(\alpha)]$ when $\alpha > 2(\gamma/3)^{3/2}$. $\square$

For values $\alpha > \alpha_{\mathrm{C}}$, $\psi^{-1}(\alpha)$ is determined by the unique real-valued root of the cubic polynomial $-y^3 + \gamma y + \alpha$ and can be expressed as

$$\begin{aligned} \psi^{-1}(\alpha) \;=& \\ & \frac{1}{6}\left(12\sqrt{81\alpha^2 - 12\gamma^3} - 108\alpha\right)^{\frac{1}{3}} + 2\gamma\left(12\sqrt{81\alpha^2 - 12\gamma^3} - 108\alpha\right)^{-\frac{1}{3}}. \end{aligned} \tag{7.13}$$

From Equation (7.13), it can be shown that $\lim_{\alpha \to \alpha_{\mathrm{C}}^+} \psi^{-1}(\alpha) = 2\sqrt{\gamma/3}$.

Writing the system output as $y = \boldsymbol{r}^t \boldsymbol{f}$ for a (fixed) received vector $\boldsymbol{r}$ and arbitrary equalizer $\boldsymbol{f}$ allows the following equalizer-space interpretation of Lemma 7.1:

Theorem 7.1. *Denote the set of possible received vectors by $\mathcal{R}$, and define $\mathcal{F}_\alpha$ to be the convex hull formed by the set of hyperplanes $\mathcal{B}_\alpha := \{\boldsymbol{f} : |\boldsymbol{r}^t \boldsymbol{f}| = \psi^{-1}(\alpha)$ for $\boldsymbol{r} \in \mathcal{R}\}$. Then choice of dither amplitude $\alpha > \alpha_C$ ensures that the expected DSE-CMA update is identical to the CMA update for equalizers within $\mathcal{F}_\alpha$.*

Proof. Choose any two equalizers $\boldsymbol{f}_1$ and $\boldsymbol{f}_2$ that satisfy the output constraint $|\boldsymbol{r}^t \boldsymbol{f}| \leq \psi^{-1}(\alpha)$ for all $\boldsymbol{r} \in \mathcal{R}$. (Recall $\psi^{-1}(\alpha)$ is well defined for $\alpha > \alpha_{\mathrm{C}}$.) The triangle inequality implies that any convex combination of $\boldsymbol{f}_1$ and $\boldsymbol{f}_2$ also satisfies this output constraint. Lemma 7.1 ensures that, for $y = \boldsymbol{r}^t \boldsymbol{f}$ that satisfy the output amplitude constraint, $\varphi_\alpha(y) = \psi(y)$. Hence, the two updates are identical within $\mathcal{F}_\alpha$. $\square$

For an M-ary source, the set $\mathcal{S}$ of possible source vectors $\boldsymbol{s}$ is of size M^{N_q}. Then, in the absence of channel noise, we expect at most M^{N_q} equalizer input vectors $\boldsymbol{r} = \boldsymbol{\mathcal{H}}^t \boldsymbol{s}$. Hence, in this noiseless case, $\mathcal{F}_\alpha$ is the convex hull formed by the finite set of M^{N_q} hyperplanes $\mathcal{B}_\alpha = \{\boldsymbol{f} : |\boldsymbol{s}^t \boldsymbol{\mathcal{H}} \boldsymbol{f}| = \psi^{-1}(\alpha)$ for $\boldsymbol{s} \in \mathcal{S}\}$. In other words, $\mathcal{F}_\alpha$ is a polytope formed by the boundary set $\mathcal{B}_\alpha$. An illustrative example of $\mathcal{F}_\alpha$ and $\mathcal{B}_\alpha$ is provided by Fig. 7.5.

Next, we concern ourselves with neighborhoods of the zero-forcing (ZF) equalizers $\{\boldsymbol{f}_\delta : 0 \leq \delta < N_q\}$ which have the property $\boldsymbol{f}_\delta^t \boldsymbol{\mathcal{H}} = \mathbf{e}_\delta^t$. The ZF equalizers exist when $\boldsymbol{\mathcal{H}}$ is FCR.

Theorem 7.2. *Define*

$$\alpha_{ZF} \quad := \quad \max_{s \in S} |\psi(s)|. \tag{7.14}$$

Under FCR $\boldsymbol{\mathcal{H}}$, choice of dither amplitude $\alpha > \alpha_{ZF}$ ensures the existence of a neighborhood around every ZF solution $\boldsymbol{f}_\delta$ within which the expected DSE-CMA update is identical to the CMA update.

Proof. When $\boldsymbol{f} = \boldsymbol{f}_\delta$, we know $y_n = s_{n-\delta}$ for all s_n. In this case, (7.11) and the definition of α_{ZF} imply that $\psi(y_n) = \varphi_\alpha(y_n)$ for $\alpha \geq \alpha_{\text{ZF}}$. In other words, $\alpha \geq \alpha_{\text{ZF}}$ guarantees that the expected DSE-CMA update is identical to the CMA update *at* the zero-forcing solutions.

Now, consider an open ball $\mathcal{B}$ of radius ρ centered at $\boldsymbol{f}_\delta$. Equalizers within $\mathcal{B}$ can be parameterized as $\boldsymbol{f} = \boldsymbol{f}_\delta + \tilde{\boldsymbol{f}}$ for $\|\tilde{\boldsymbol{f}}\| < \rho$. Then there exists a finite constant K for which $|y_n - s_{n-\delta}| \leq \max_{\boldsymbol{s}\in\mathcal{S}} |\boldsymbol{s}^t \boldsymbol{\mathcal{H}} \tilde{\boldsymbol{f}}| < K\rho$. From the continuity of the polynomial function $\psi(\cdot)$, we claim the following: for any $\varepsilon := \alpha - \alpha_{\text{ZF}} > 0$ and any $\boldsymbol{r} \in \mathcal{R}$, there exists a $\rho > 0$ such that $\|\tilde{\boldsymbol{f}}\| < \rho$ implies $|\psi(\boldsymbol{r}^t\boldsymbol{f}) - \psi(\boldsymbol{r}^t\boldsymbol{f_\delta})| < \varepsilon$. Applying (7.11), we conclude that $\psi = \varphi_\alpha$ for any equalizer within the ball $\mathcal{B}$. $\square$

Note that the constant α_{ZF} may be less than α_{C}, in which case there would exist *isolated* "CMA-like" neighborhoods around the ZF solutions—i.e., neighborhoods not contained in any "CMA-like" convex hull.

Theorem 7.2 is of limited practical use since it requires FCR $\boldsymbol{\mathcal{H}}$. Fortunately, the concept is easily extended to the set of "open-eye" equalizers, $\mathcal{F}_{\text{OE}}$. Denoting the minimum distance between any pair of adjacent symbols in S by Δ_s, we define the set $\mathcal{F}_{\text{OE}}$ as[5]

$$\mathcal{F}_{\text{OE}} := \left\{ \boldsymbol{f} : \min_{\delta} \max_{\boldsymbol{r}\in\mathcal{R}} \boldsymbol{r}^t(\boldsymbol{f} - \boldsymbol{f}_\delta) < \Delta_s/2 \right\}.$$

[5]We acknowledge that the definition of $\mathcal{F}_{\text{OE}}$ is overly strict in that it bounds the outermost decision region from both sides. In addition, the definition of $\mathcal{F}_{\text{OE}}$ only makes sense in the context of bounded inputs $\boldsymbol{r}$. Although the AWGN channel model does not ensure bounded $\boldsymbol{r}$, all practical implementations do.

The corresponding set of open-eye equalizer outputs is defined by

$$Y_{\mathrm{OE}} := \big\{ y : \min_{s \in S} |y - s| < \Delta_s/2 \big\}.$$

For M-PAM, Y_{OE} becomes the open interval $(-s_{\max} - s_{\min},\ s_{\max} + s_{\min})$ minus the set of points halfway between adjacent elements of S. Here, $s_{\min}$ and $s_{\max}$ are used to denote the minimum and maximum positive-valued elements of S, respectively.

Theorem 7.3. *Define*

$$\alpha_{OE} \ := \ \max_{y \in Y_{OE}} |\psi(y)|. \tag{7.15}$$

Choice of dither amplitude $\alpha > \alpha_{OE}$ ensures the existence of a neighborhood around every open-eye equalizer, $\boldsymbol{f} \in \mathcal{F}_{OE}$, within which the expected DSE-CMA update is identical to the CMA update.

Proof. The proof is identical to that of Theorem 7.2 after replacing $s \in S$ by $y \in Y_{\mathrm{OE}}$. $\square$

In summary, α_{C} is the lower limit of α for which the convex set $\mathcal{F}_\alpha$ exists, while α_{ZF} and α_{OE} are the lower limits of α for which "CMA-like" local neighborhoods around the zero-forcing and open-eye equalizers exist, respectively. Table 7.1 quantifies the values of $\{\alpha_{\mathrm{C}}, \alpha_{\mathrm{ZF}}, \alpha_{\mathrm{OE}}\}$ for M-PAM alphabets, and Fig. 7.4 illustrates their relationship to the CMA error function. Note that the difference between α_{ZF} and α_{OE} narrows as the alphabet size increases. This can be attributed to the fact that the open-eye neighborhoods shrink as the constellation becomes more dense.

7.3.3 DSE-CMA Cost Surface

Studies of the multi-modal J_{c} cost surface give substantial insight into the transient behavior of CMA (see, e.g., [Johnson PROC 98]). Thus, we expect that an examination of J_{dse}, the cost stochastically minimized by DSE-CMA, should also prove

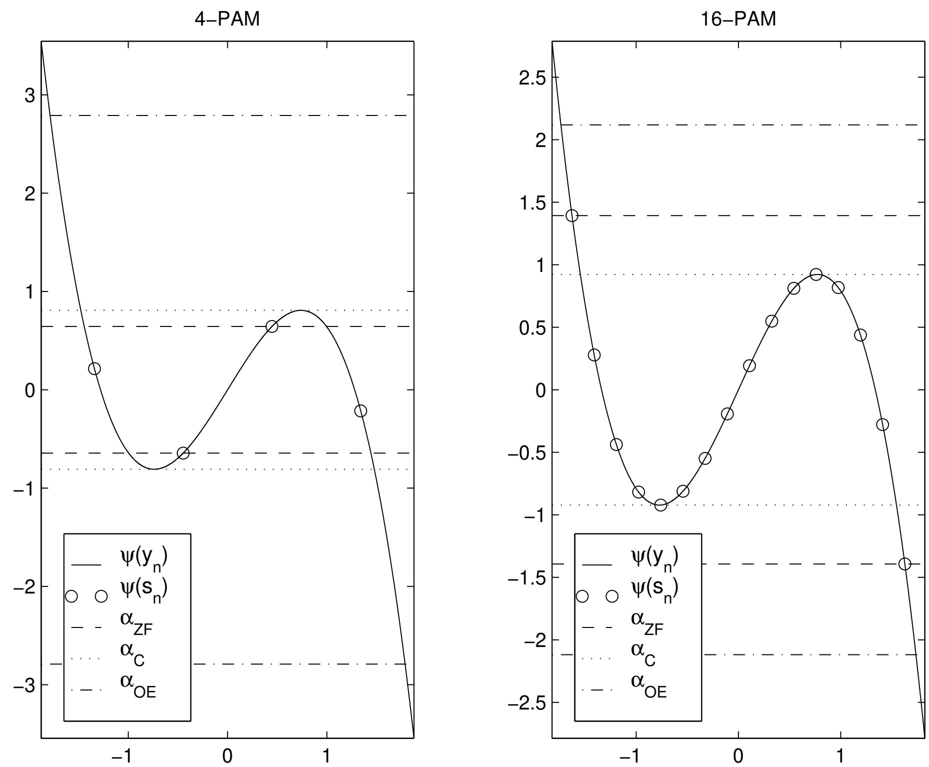

Figure 7.4: CMA error function and critical values of α for 4-PAM and 16-PAM sources.

Table 7.1: Critical values of α for M-PAM.

M	2	4	8	16	32
α_{C}	0.38	0.81	0.90	0.92	0.93
α_{ZF}	0	0.64	0.87	1.39	1.71
α_{OE}	6	2.79	2.24	2.12	2.09

worthwhile. First, however, we need to construct J_{dse}. Since we know that a gradient descent algorithm minimizing J has the general form $\boldsymbol{f}(n+1) = \boldsymbol{f}(n) - \mu \nabla_{\boldsymbol{f}} J$, we conclude from (7.4) that $\nabla_{\boldsymbol{f}} J_{\mathrm{dse}} = -\,\mathrm{E}\{\varphi_\alpha(y_n, d_n)\, \boldsymbol{r}(n)\}$. It is then possible to find $J_{\mathrm{dse}}(\boldsymbol{f})$ (to within a constant) by integrating $\nabla_{\boldsymbol{f}} J_{\mathrm{dse}}$ over N_f-dimensional equalizer space.

Fig. 7.5 shows an illustrative example of $J_{\mathrm{dse}}(\boldsymbol{f})$ contours superimposed on $J_{\mathrm{c}}(\boldsymbol{f})$ contours in equalizer space for $N_f = 2$. Note that the two sets of cost contours are identical within the convex polytope $\mathcal{F}_\alpha$ formed by the hyperplanes $\mathcal{B}_\alpha$. Outside $\mathcal{F}_\alpha$, the CMA cost contours rise much quicker than the DSE-CMA contours. This observation can be attributed to the fact that, for large $\|\boldsymbol{f}\|$, $J_{\mathrm{c}}(\boldsymbol{f})$ is proportional to $\|\boldsymbol{f}\|^4$ while the hard limiting on φ_α makes $J_{\mathrm{dse}}(\boldsymbol{f})$ proportional to $\|\boldsymbol{f}\|$. As a result, we expect that CMA exhibits much faster convergence for initializations far outside of $\mathcal{F}_\alpha$. Unlike standard SE algorithms [Macchi Book 95], though, DSE-CMA converges *as rapidly* as its unsigned version within $\mathcal{F}_\alpha$. Fortunately, there is no need to initialize the adaptive algorithm with large $\|\boldsymbol{f}\|$: the "power constraint property" of CMA [Zeng TIT 98] ensures that the CMA minima lie in a hyper-annulus that includes[6] $\|\boldsymbol{f}\| \approx 1$ (see, e.g., Fig. 7.8). Initialization of DSE-CMA is discussed in Section 7.4.

Fig. 7.6 shows two low-dimensional examples of a DSE-CMA trajectory overlaid on a CMA trajectory. Note that the DSE-CMA trajectories closely follow the CMA trajectories, but exhibit more parameter "jitter". The effect of this parameter variation on steady-state MSE performance is quantified in the next section.

Figures 7.5 and 7.6 both assume a BPSK source transmitted over noiseless FIR vector channel $\{\mathbf{h}_n\} = \left\{ \left(\begin{smallmatrix} 0.1 \\ 0.3 \end{smallmatrix}\right), \left(\begin{smallmatrix} 1 \\ -0.1 \end{smallmatrix}\right), \left(\begin{smallmatrix} 0.5 \\ 0.2 \end{smallmatrix}\right) \right\}$ and $\alpha = 1$.

[6] Assuming that the equalizer input is power-normalized, as occurs in practice.

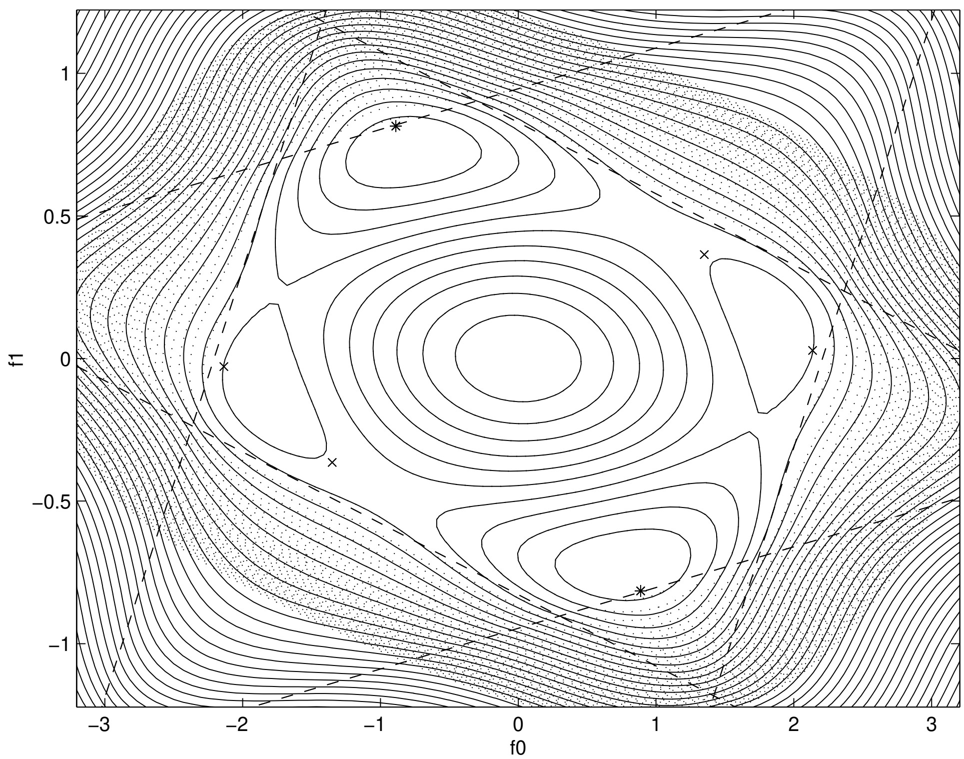

Figure 7.5: Superimposed DSE-CMA (solid) and CMA (dotted) cost contours in equalizer space. Dashed lines show the set of hyperplanes $\mathcal{B}_\alpha$ whose convex hull $\mathcal{F}_\alpha$ ensures expected DSE-CMA behavior identical to that of CMA.

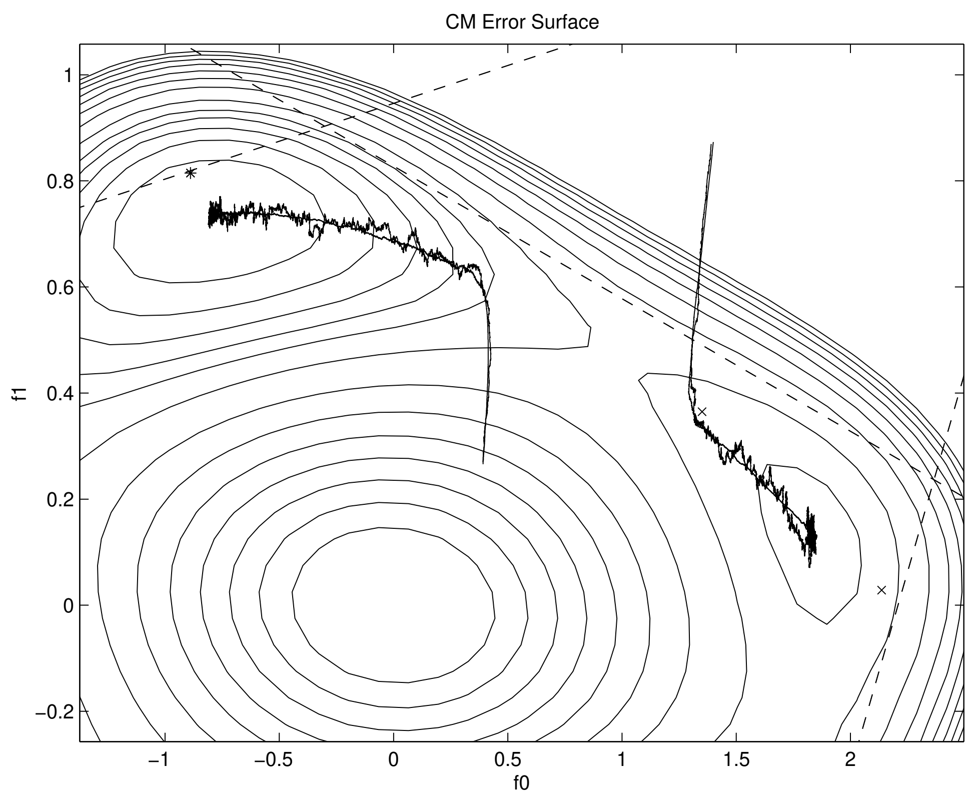

Figure 7.6: Trajectories of DSE-CMA (rough) overlaid on those of CMA (smooth) for $\mu = 5 \times 10^{-4}$. Solid lines are J_c contours, and dashed lines form the boundary set $\mathcal{B}_\alpha$.

7.3.4 DSE-CMA Steady-State Behavior

The principle disadvantage of DSE-CMA concerns its steady-state behavior: the addition of dither leads to an increase in excess mean-squared error (EMSE). EMSE is typically defined as the steady-state MSE above the level attained by the fixed locally minimum MSE solution. The subsections below quantify the EMSE of DSE-CMA assuming $\gamma = \gamma_\star$ and FCR $\boldsymbol{\mathcal{H}}$.

Small-Error Approximation of the CMA Update

By writing the equalizer output y_n in terms of the delayed source $s_{n-\delta}$ and defining the output error $e_n := y_n - s_{n-\delta}$, the CMA error function can be written as

$$\begin{aligned} \psi(y_n) &= \big(\gamma - |e_n + s_{n-\delta}|^2\big)(e_n + s_{n-\delta}), \\ &= -e_n^3 - 3s_{n-\delta}e_n^2 - (3s_{n-\delta}^2 - \gamma)e_n + \psi(s_{n-\delta}). \end{aligned}$$

For small output error (i.e., $|e_n| \ll 1$), the error function can be approximated by

$$\psi(y_n) \approx (\gamma - 3s_{n-\delta}^2)e_n + \psi(s_{n-\delta}). \tag{7.16}$$

In the absence of channel noise, we can write $e_n = \boldsymbol{r}^t(n)\tilde{\boldsymbol{f}}(n)$, using the parameter error vector $\tilde{\boldsymbol{f}}(n) := \boldsymbol{f}(n) - \boldsymbol{f}_\delta$ defined relative to the zero-forcing equalizer $\boldsymbol{f}_\delta$. For adequately small $\tilde{\boldsymbol{f}}(n)$, (7.16) implies that the CMA error function has the approximate form

$$\psi(y_n) \approx (\gamma - 3s_{n-\delta}^2)\boldsymbol{r}^t(n)\tilde{\boldsymbol{f}}(n) + \psi(s_{n-\delta}). \tag{7.17}$$

With FCR $\boldsymbol{\mathcal{H}}$ and a reasonably small step-size, we expect asymptotically small e_n. Thus, the small-error approximation (7.17) can be used to characterize the steady-state behavior of DSE-CMA.

The Excess MSE of DSE-CMA

We define EMSE at time index n as the expected squared error above that achieved by the (local) zero-forcing solution $\boldsymbol{f}_\delta$. Since $\boldsymbol{f}_\delta$ achieves zero error when $\boldsymbol{\mathcal{H}}$ is FCR,

$$J_{\mathrm{ex}}(n) \quad := \quad \mathrm{E}\big\{|\boldsymbol{r}^t(n)\tilde{\boldsymbol{f}}(n)|^2\big\}. \tag{7.18}$$

We are interested in quantifying the steady-state EMSE: $J_{\mathrm{ex}} := \lim_{n\to\infty} J_{\mathrm{ex}}(n)$. Our derivation of steady-state EMSE assumes the following:

(B1) The equalizer parameter error vector $\tilde{\boldsymbol{f}}(n)$ is statistically independent of the equalizer input $\boldsymbol{r}(n)$.

(B2) The dither amplitude α is chosen sufficiently greater than α_{ZF} so that $\alpha > |\psi(y_n)|$ for all y_n under consideration.

(B3) $\boldsymbol{\mathcal{H}}$ is FCR so that the zero-forcing solution attains zero error, i.e., $\mathrm{E}\big\{|s_{n-\delta} - \boldsymbol{r}^t(n)\boldsymbol{f}_\delta|^2\big\} = 0$.

(B4) The step-size is chosen small enough for the small-error approximation (7.16) to hold asymptotically.

The classical assumption (B1) implies that $\tilde{\boldsymbol{f}}(n)$ is independent of the source process $\{s_n\}$. Assumption (B2) is needed for the results of the quantization noise model in Section 7.3.1 to hold.

Using the facts that $\mathrm{tr}(A) = A$ for any scalar A, and that $\mathrm{tr}(\tilde{\boldsymbol{f}}^t\mathbf{A}\tilde{\boldsymbol{f}}) = \mathrm{tr}(\tilde{\boldsymbol{f}}\tilde{\boldsymbol{f}}^t\mathbf{A})$ and $\mathrm{E}\big\{\mathrm{tr}(\mathbf{A})\big\} = \mathrm{tr}\big(\mathrm{E}\{\mathbf{A}\}\big)$ for any matrix $\mathbf{A}$, the EMSE at time index n can be written

$$\begin{aligned} J_{\mathrm{ex}}(n) \quad &= \quad \mathrm{tr}\Big(\mathrm{E}\big\{\tilde{\boldsymbol{f}}(n)\tilde{\boldsymbol{f}}^t(n)\boldsymbol{r}(n)\boldsymbol{r}^t(n)\big\}\Big) \\ &= \quad \mathrm{tr}\Big(\mathrm{E}\big\{\tilde{\boldsymbol{f}}(n)\tilde{\boldsymbol{f}}^t(n)\big\}\,\mathrm{E}\big\{\boldsymbol{r}(n)\boldsymbol{r}^t(n)\big\}\Big), \end{aligned} \tag{7.19}$$

where the second step follows from (B1). Defining the expected equalizer outer product matrix $\boldsymbol{F}(n) := \mathrm{E}\{\tilde{\boldsymbol{f}}(n)\tilde{\boldsymbol{f}}^t(n)\}$ and the source-power-normalized regressor autocorrelation matrix $\boldsymbol{R} := \frac{1}{\sigma_s^2}\mathrm{E}\{\boldsymbol{r}(n)\boldsymbol{r}^t(n)\}$ we can write the EMSE as

$$J_{\mathrm{ex}}(n) \;=\; \sigma_s^2 \,\mathrm{tr}\big(\boldsymbol{R}\boldsymbol{F}(n)\big). \tag{7.20}$$

Note that, since $\{s_n\}$ is i.i.d. and $\boldsymbol{r}(n) = \boldsymbol{\mathcal{H}}\boldsymbol{s}(n)$, we have $\boldsymbol{R} = \boldsymbol{\mathcal{H}}\boldsymbol{\mathcal{H}}^t$.

Appendix 7.B uses the quantization noise model from Section 7.3.1 and the error function approximation from Section 7.3.4 to derive the following recursion for $\boldsymbol{F}(n)$, valid for equalizer lengths $N_f \gg 1$:

$$\boldsymbol{F}(n+1) \;=\; \boldsymbol{F}(n) - \mu(3-\kappa_s)\sigma_s^4\big(\boldsymbol{F}(n)\boldsymbol{R} + \boldsymbol{R}\boldsymbol{F}(n)\big) + \mu^2\alpha^2\sigma_s^2\boldsymbol{R}. \tag{7.21}$$

Using (7.20)–(7.21), Appendix 7.C derives the following approximation to the steady-state EMSE of DSE-CMA:

$$J_{\mathrm{ex}} \;\approx\; \frac{\mu\alpha^2 N_f\,\sigma_r^2}{2(3-\kappa_s)\sigma_s^2}, \tag{7.22}$$

where $\sigma_r^2 := \mathrm{E}\{|r_k|^2\}$. The approximation in (7.22) closely matches the outcomes of experiments conducted using microwave channel models obtained from the SPIB database. The simulation results are presented in Section 7.5.

Equation (7.22) can be compared to an analogous expression for the EMSE of CMA [Fijalkow TSP 98]:

$$J_{\mathrm{ex|cma}} \;\approx\; \frac{\mu N_f\,\sigma_r^2}{2(3-\kappa_s)}\left(\frac{\mathrm{E}\{s_n^6\}}{\sigma_s^6} - \kappa_s^2\right)\sigma_s^4. \tag{7.23}$$

It is apparent that the EMSE of CMA and DSE-CMA differ by the multiplicative factor

$$K_{\alpha,S} \;:=\; \frac{\alpha^2}{\mathrm{E}\{s_n^6\} - \kappa_s^2\sigma_s^6}, \tag{7.24}$$

via $J_{\text{ex}} = K_{\alpha,S} J_{\text{ex|cma}}$. Note the dependence on both the dither amplitude α and the source distribution. Table 7.2 presents values of $K_{\alpha,S}$ for various M-PAM sources and particular choices of α (to be discussed in Section 7.4.2).

7.4 DSE-CMA Design Guidelines

7.4.1 Selection of Dispersion Constant γ

We take the "Bussgang" approach used in [Godard TCOM 80], whereby γ is selected to ensure that the mean equalizer update is zero when perfect equalization been achieved. From (7.4), (7.11), and the system model in Section 2.2, we can write the mean update term of DSE-CMA at $\boldsymbol{f}_\delta$ (in the absence of noise) as $\mu \boldsymbol{\mathcal{H}}\, \mathrm{E}\big\{\boldsymbol{s}(n)\varphi_\alpha(s_{n-\delta})\big\}$. For an i.i.d. source, $\varphi_\alpha(s_{n-\delta})$ is independent of all but one element in $\boldsymbol{s}(n)$, namely $s_{n-\delta}$. Hence, we require that the value of γ in φ_α be chosen so that

$$\mathrm{E}\big\{s_{n-\delta}\, \varphi_\alpha(s_{n-\delta})\big\} \quad = \quad 0. \tag{7.25}$$

When $\alpha > \alpha_{\text{ZF}}$, Theorem 7.2 ensures the existence of a neighborhood around $\boldsymbol{f}_\delta$ within which $\varphi_\alpha(y_n) = \psi(y_n)$. For such α, (7.25) implies that γ should be chosen as for CMA: $\gamma = \mathrm{E}\{|s|^4\}/\sigma_s^2$ [Godard TCOM 80]. When $\alpha < \alpha_{\text{ZF}}$, closed form expressions for γ in the case of M-PAM DSE-CMA are difficult to derive. However, γ satisfying (7.25) for these cases can be determined numerically.

7.4.2 Selection of Dither Amplitude α

While Section 7.3.4 demonstrated that EMSE is proportional to α^2, Section 7.3.2 showed that larger values of α increase the region within which DSE-CMA behaves like CMA. *The selection of dither amplitude α is therefore a design tradeoff between CMA-like robustness and steady-state MSE performance.*

Theorems 7.1 and 7.2 imply that the choice $\alpha > \max\{\alpha_{\mathrm{C}}, \alpha_{\mathrm{ZF}}\}$ ensures that the zero-forcing equalizers are contained in the convex polytope $\mathcal{F}_\alpha$. Thus, under FCR $\boldsymbol{\mathcal{H}}$, $\alpha = \max\{\alpha_{\mathrm{C}}, \alpha_{\mathrm{ZF}}\}$ could be considered a useful design guideline, since the CMA minima are expected to be in close proximity to the zero-forcing solutions [Johnson PROC 98]. In fact, since $\mathcal{F}_\alpha$ is convex and contains the origin, we expect that a small-norm initialization (see Section 7.4.4) will lead to equalizer trajectories completely contained within $\mathcal{F}_\alpha$. Such a strategy is advantageous from the point of robustness.

In situations where the FCR $\boldsymbol{\mathcal{H}}$ condition is severely violated and CMA can do no better than "open the eye", selection of dither amplitude in the range

$$\max\{\alpha_{\mathrm{C}}, \alpha_{\mathrm{ZF}}\} < \alpha < \max\{\alpha_{\mathrm{C}}, \alpha_{\mathrm{OE}}\}$$

is recommended to retain CMA-like robustness.

Table 7.1 presents these critical values of α for various M-PAM constellations. Note that the value of α_{OE} for BPSK appears unusually large because near-closed-eye operating conditions for BPSK are quite severe.

7.4.3 Selection of Step-Size μ

As in "classical" LMS theory, the selection of step-size becomes a tradeoff between convergence rate and EMSE. If convergence rate is non-critical, α could be selected with robustness in mind and μ selected to meet steady-state MSE requirements.

Say that the goal was to attain the same steady-state MSE performance as CMA. Then when $\boldsymbol{\mathcal{H}}$ is FCR, μ should be chosen $K_{\alpha,S}^{-1}$ times that of CMA, where $K_{\alpha,S}$ was defined in (7.24). Table 7.2 presents values of $K_{\alpha,S}$ over the recommended range of

Table 7.2: Steady-state MSE relative performance factor.

M-PAM	2	4	8	16	32
$K_{\alpha,S}\big\|_{\alpha=\max\{\alpha_{\mathrm{C}},\alpha_{\mathrm{ZF}}\}}$	-	2.9	1.6	3.3	4.8
$K_{\alpha,S}\big\|_{\alpha=\max\{\alpha_{\mathrm{C}},\alpha_{\mathrm{OE}}\}}$	-	34	9.8	7.6	7.2

α and can be used to predict the typical range of CMA convergence speed relative to DSE-CMA (for equal steady-state performance).

When neither convergence rate nor steady-state MSE performance can be sacrificed, Table 7.2 suggests choosing α closer to $\max\{\alpha_{\mathrm{C}}, \alpha_{\mathrm{ZF}}\}$. In this case, CMA-like robustness is sacrificed instead. For such α, however, it becomes hard to predict the effects that non-FCR $\boldsymbol{\mathcal{H}}$ have on the transient and steady-state performance of DSE-CMA. Loosely speaking, as α is decreased below $\max\{\alpha_{\mathrm{C}}, \alpha_{\mathrm{ZF}}\}$, the performance of DSE-CMA becomes more like that of SE-CMA.

7.4.4 Initialization of DSE-CMA

The single-spike initialization [Godard TCOM 80] has become a popular initialization strategy for baud-spaced CMA, as has double-spike initialization [Johnson PROC 98], its $T/2$-spaced counterpart. The similarities between DSE-CMA and CMA suggest that these initialization strategies should work well for DSE-CMA as well.

In the interest of preserving CMA-like robustness, however, it is suggested the norm of the DSE-CMA initialization be kept small[7]. Under proper selection of α (i.e., $\alpha > \alpha_{\mathrm{C}}$), this strategy ensures that the parameter trajectories begin within

[7] This is consistent with various recommendations on the initialization of CMA: those given for single-user applications in [Chung ASIL 98], as well as those given for multi-user applications in Chapter 5 of this thesis.

Table 7.3: J_{ex} Deviation from predicted level for various SPIB channels and M-PAM.

M-PAM	2	4	8	16	32
SPIB #2	1.3%	-0.5%	-0.5%	-0.7%	-0.5%
SPIB #3	1.2%	-0.2%	-0.6%	-1.0%	-1.0%
SPIB #4	1.4%	-0.6%	-0.6%	-0.7%	-0.7%

the convex region $\mathcal{F}_\alpha$ (see Fig. 7.8). Extending this idea, Section 7.3.2 implies that large enough choices of α (e.g., $\alpha \approx \alpha_{\text{OE}}$) ensure that the *entire* mean trajectory will stay within $\mathcal{F}_\alpha$ (and for adequately small step-sizes, the actual trajectories should closely approximate the mean trajectory). To conclude, proper choices of initialization norm and dither amplitude α guarantee that the mean behavior of DSE-CMA never differs from that of CMA.

7.5 Simulation Results

7.5.1 Excess MSE for FCR $\boldsymbol{\mathcal{H}}$

Table 7.3 presents simulation results verifying the approximation of the excess MSE of DSE-CMA given in (7.22). The simulations were conducted using length-64 MMSE approximations of three (noiseless) SPIB microwave channels, length-62 $T/2$-spaced FSEs, and various i.i.d. M-PAM sources. The resulting $\boldsymbol{\mathcal{H}}$ was FCR. The step-sizes were chosen so that (B4) was satisfied, and the dither amplitude of $\alpha = 2$ satisfied (B2). Table 7.3 gives percentage deviations from the EMSE levels predicted by (7.22) which were obtained by averaging the results of 2.5×10^8 iterations. Overall, the simulation results closely match our approximation (7.22).

7.5.2 Average Transient Behavior

Throughout the chapter, we have emphasized the importance of performance evaluation in realistic (non-ideal) environments. It is only proper to present a comparison of DSE-CMA to CMA in this context as well. Fig. 7.7 shows ensemble-averaged MSE trajectories of the two algorithms operated under identical conditions and initialized at the same locations using various SPIB microwave channels. Noise levels (SNR = 40dB) and equalizer lengths ($N_f = 32$) were selected to represent typical applications while providing open-eye performance (for an 8-PAM source) at convergence. The following "double-spike" equalizer initialization was used in all simulations: taps 10 and 11 were set to 0.5 and all others were set to zero. Although (purposely) sub-optimal, this initialization represents a reasonable choice given the microwave channel profiles and the discussion in Section 7.4.4. As evident in Fig. 7.7, the DSE-CMA trajectories track the CMA trajectories closely until the effects of EMSE take over. Fig. 7.7 also suggests that the EMSE approximation in (7.22) remains a useful guideline even under practical noisy non-FCR channels.

Although parameter trajectory comparisons are impractical with length-32 equalizers, it is easy to visualize two-tap examples. Fig. 7.8 shows ensemble-averaged DSE-CMA trajectories overlaid on ensemble-averaged CMA trajectories for a noisy undermodelled channel and 4-PAM. The two trajectories in each pair correspond so closely that they are nearly indistinguishable from one another. The trajectories were initialized from various locations on the inner CMA power constraint boundary, and remain, for the most part, in $\mathcal{F}_\alpha$. Note that for trajectories that cross a single boundary plane in the set $\mathcal{B}_\alpha$, the expected DSE-CMA update differs from CMA for only *one* element in the set of possible received vectors $\mathcal{R}$. In other words, loss of CMA-like behavior outside $\mathcal{F}_\alpha$ occurs gradually.

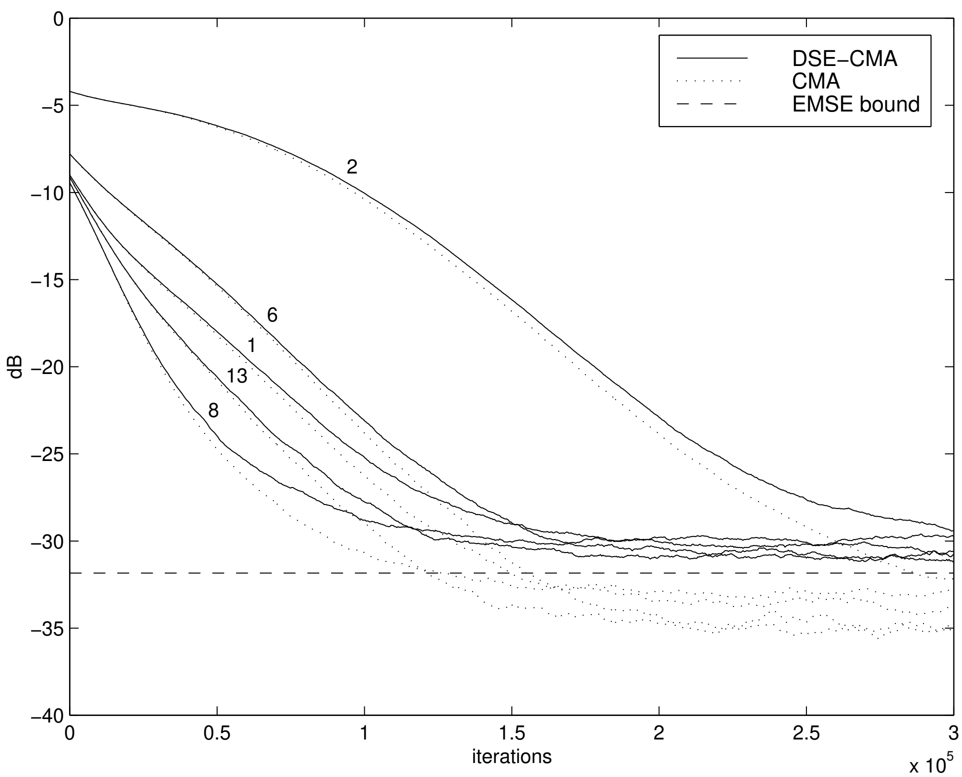

Figure 7.7: Averaged MSE trajectories for DSE-CMA and CMA initialized at the same locations using 8-PAM and (normalized) SPIB channels 1, 2, 6, 8, and 13. For all simulations: SNR = 40dB, $N_f = 32$, $\mu = 2 \times 10^{-5}$, and $\alpha = \alpha_{\text{OE}} = 2.25$.

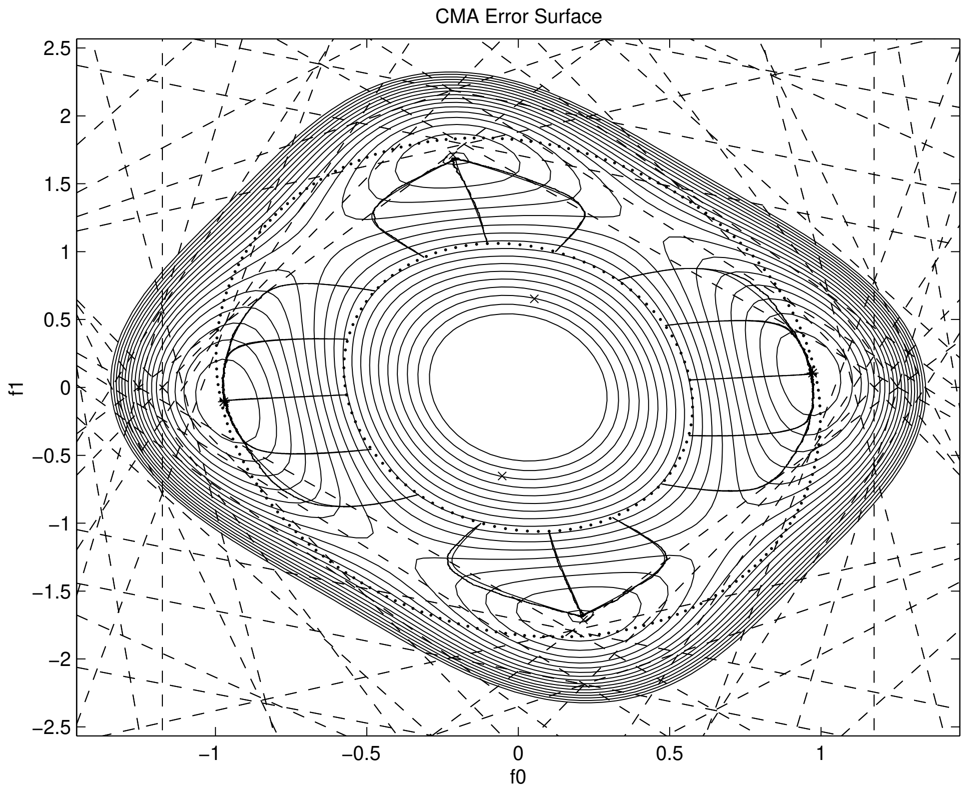

Figure 7.8: Averaged DSE-CMA and CMA tap trajectories initialized at the same locations and superimposed on CMA cost contours for channel $\{h_k\} = (0.1, 1, 0.5, -0.1, 0.2, 0.1)^t$, SNR = 30dB, 4-PAM, and $\alpha = 2$. Dotted lines indicate CMA power constraint boundaries and dashed lines indicate $\mathcal{B}_\alpha$.

7.5.3 Comparison with Update-Decimated CMA

One popular technique used to reduce the computation requirements of CMA involves updating the equalizer every D baud samples (rather than every sample as (7.1) suggests). This is possible in situations where the channel time variations are slow with respect to the equalizer adaptation speed. As an example, fixed-site microwave applications can often tolerate update decimations of $D = 16$ and higher [Treichler PROC 98]. The fundamental drawback to these decimated algorithms is that their convergence rates decrease in proportion to D.

Since DSE-CMA and update-decimated CMA (UD-CMA) both present strategies for computationally-efficient CMA-like blind adaptation, a comparison is in order. In Section 7.4.3 we discussed how DSE-CMA step-size may be selected to achieve steady-state MSE levels on par with CMA and argued that the penalty is DSE-CMA convergence rate $K_{\alpha,S}$ times slower than CMA. Although, for a given step-size, UD-CMA should achieve the same steady-state performance as CMA, we expect a convergence rate that is D times slower. Taken together, we anticipate advantages in using DSE-CMA in situations where the implementation budget demands a UD-CMA decimation factor $D > K_{\alpha,S}$. (Recall that typical values of $K_{\alpha,S}$ appear in Table 7.2.)

As verification of our claim, Fig. 7.9 presents ensemble-averaged MSE trajectories comparing DSE-CMA to UD-CMA for $\alpha = \alpha_{\text{OE}}$ and $D = 16$. The operating environment and design quantities used were the same as those of Fig. 7.7 with the exception that $\mu = 2 \times 10^{-4}$ for UD-CMA. This UD-CMA step-size was adjusted to equate steady-state performance, thus the advantage of DSE-CMA appears in the form of increased convergence rate. Checking Table 7.2, we find that, for dither amplitude α_{OE} and an 8-PAM source, DSE-CMA is expected to “beat” UD-CMA whenever D must be selected ≥ 10.

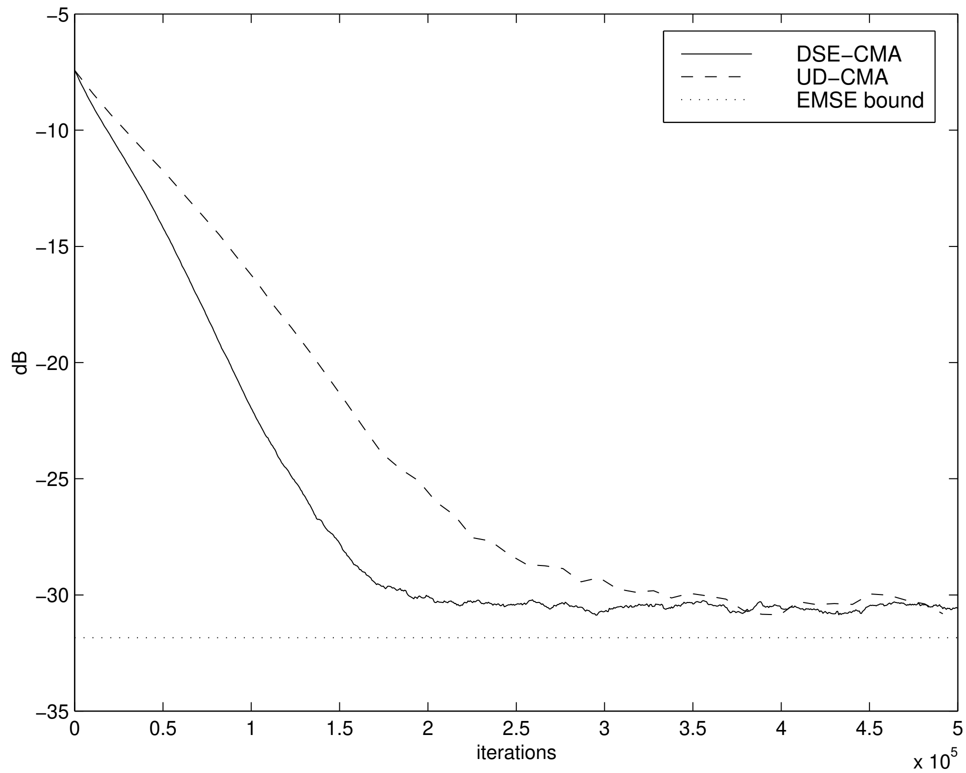

Figure 7.9: Averaged MSE trajectories for DSE-CMA and update-decimated CMA initialized at the same locations using 8-PAM and SPIB channel 8. Relevant parameters: SNR = 40dB, $N_f = 32$, $\alpha = \alpha_{\text{OE}} = 2.25$, and update-decimation factor $D = 16$.

7.6 Conclusions

This chapter has derived the fundamental properties of the dithered signed-error constant modulus algorithm. In summary, we have found that, under proper selection of algorithmic design quantities, the expected transient behavior of DSE-CMA is *identical* to that of CMA. Although the steady-state MSE of DSE-CMA is larger than that of CMA, its value is well characterized and can be accounted for in the design procedure.

With the exception of computational complexity, the new algorithm has been designed to mimic CMA, rather than "improve" on its performance. Our primary motivation for this is twofold. First, CMA is well-regarded by practitioners. It has established itself over the last 20 years as the most popular practical blind equalization algorithm, due in large part to its robustness properties [Johnson PROC 98]. It is precisely these robustness properties which we have attempted to preserve. Secondly, CMA has been extensively analyzed by theoreticians. The bulk of these analyses apply directly to DSE-CMA. As it is often the case that modifications of classic algorithms have disadvantages that outweigh the proposed advantages, the spirit of DSE-CMA is a computationally efficient algorithm that "leaves well enough alone."

Although we have restricted our focus to the real-valued case, a straightforward complex-valued extension of DSE-CMA is obtained by replacing the real-valued $\mathrm{sgn}(\cdot)$ in (7.4) with the complex-valued operator $\mathrm{csgn}(x) := \mathrm{sgn}(\mathrm{Re}\, x) + j\, \mathrm{sgn}(\mathrm{Im}\, x)$ and by replacing the real-valued dither process $\{d_n\}$ with the complex-valued $\{d_n^{(r)}\} + j\{d_n^{(i)}\}$. Here $j := \sqrt{-1}$, and the processes $\{d_n^{(r)}\}$ and $\{d_n^{(i)}\}$ are real-valued, independent, and distributed identically to $\{d_n\}$. It can be shown that, with minor modifications, the properties of real-valued DSE-CMA apply to its complex-valued counterpart [Schniter ASIL 98]. Hence, the design guidelines of Section 7.4 apply to both the real- and complex-valued cases.

Finally, we mention a potentially useful modification to DSE-CMA. In the case of SE-LMS, the extension of the sign operator to a multi-level quantizer has been shown to yield significant performance improvements at the expense of a modest increase in computational complexity [Duttweiler TASSP 82]. Perhaps multi-level quantization would yield similar advantages for DSE-CMA, most importantly a reduction in EMSE.

Appendix

7.A Properties of Non-Subtractively Dithered Quantizers

In this appendix we review the key results from the theory of dithered quantizers that allow us to formulate a quantization-noise model for the DSE-CMA error function. Figure 7.3 illustrates the model described below.

We define the quantization noise arising from the non-subtractively dithered quantization of information signal x_n as

$$\epsilon_n \quad = \quad Q(x_n + d_n) - x_n \tag{7.26}$$

for a dither process $\{d_n\}$ and for $Q(\cdot)$ defined in (7.5). When the quantizer spacing Δ is large enough to satisfy

$$\big|x_n + d_n\big| \quad \leq \quad \Delta \tag{7.27}$$

and the dither is the sum of L i.i.d. random variables uniformly distributed on $\left(-\frac{\Delta}{2}, \frac{\Delta}{2}\right]$ (and statistically independent of x_n), the quantization noise has the following properties [Gray TIT 93]:

$$\mathrm{E}\{\epsilon_n^L | x_n\} \quad = \quad \mathrm{E}\{\epsilon_n^L\}, \tag{7.28}$$

$$\mathrm{E}\{\epsilon_n \epsilon_m\} \quad = \quad \mathrm{E}\{\epsilon_n^2\}\delta_{n-m}. \tag{7.29}$$

In words, equations (7.28) and (7.29) state that the quantization noise ϵ_n is an uncorrelated random process whose L^{th} moment is uncorrelated with the information signal x_n. Note that, for all values of L, we have the important property that quantization noise ϵ_n is uncorrelated with the information signal x_n:

$$\mathrm{E}\{\epsilon_n | x_n\} \quad = \quad \mathrm{E}\{\epsilon_n\} \quad = \quad 0. \tag{7.30}$$

For $L = 1$, however, we have the property that the quantization noise power *is* correlated with the information signal:

$$\mathrm{E}\{\epsilon_n^2 | x_n\} \neq \mathrm{E}\{\epsilon_n^2\}. \tag{7.31}$$

Although dither processes characterized by higher values of L make the quantization noise "more independent" of the information signal x_n, it is not without penalty. For one, the average noise power $\mathrm{E}\{\epsilon_n^2\}$ increases [Gray TIT 93]. But more importantly, the class of information signals satisfying (7.27) for a fixed Δ shrinks. Take, for example, the case where $L = 2$, so that $\{d_n\}$ has a triangular distribution on $(-\Delta, \Delta]$. In this case, (7.27) is only guaranteed when $|x_n| = 0$. Worse yet, choices of $L \geq 3$ fail to meet (7.27) for any x_n. In other words, $\{d_n\}$ uniformly distributed on $\left(-\frac{\Delta}{2}, \frac{\Delta}{2}\right]$ is the only dither process that yields a useful quantization noise model for the two-level quantizer of (7.5).

We will now quantify $\mathrm{E}\{\epsilon_n^2 | x_n\}$ for uniformly distributed dither. Note that the quantization noise takes on the values: $\epsilon_n \in \left\{-\frac{\Delta}{2} - x_n, \frac{\Delta}{2} - x_n\right\}$ with conditional probabilities $\left\{\frac{1}{2} - \frac{x_n}{\Delta}, \frac{1}{2} + \frac{x_n}{\Delta}\right\}$, respectively. The conditional expectation then becomes

$$\begin{aligned} \mathrm{E}\{\epsilon_n^2 | x_n\} &= \left(\frac{1}{2} - \frac{x_n}{\Delta}\right)\left(\frac{\Delta}{2} + x_n\right)^2 + \left(\frac{1}{2} + \frac{x_n}{\Delta}\right)\left(\frac{\Delta}{2} - x_n\right)^2 \\ &= \frac{\Delta^2}{4} - x_n^2. \end{aligned} \tag{7.32}$$

7.B Derivation of $\boldsymbol{F}(n+1)$

This appendix derives a recursion for the DSE-CMA expected parameter-error-vector outer-product, $\boldsymbol{F}(n) := \mathrm{E}\{\tilde{\boldsymbol{f}}(n)\tilde{\boldsymbol{f}}^t(n)\}$. We assume that (B1)-(B4), stated in Section 7.3.4, hold. In the sequel, the notation $[a_{i,j}]$ will be used to denote a matrix whose $(i,j)^{th}$ entry is specified by $a_{i,j}$.

Under (B2), subtracting $\boldsymbol{f}_\delta$ from both sides of equation (7.8) yields $\tilde{\boldsymbol{f}}(n+1) = \tilde{\boldsymbol{f}}(n) + \mu \boldsymbol{r}(n)\big(\psi(y_n) + \epsilon_n\big)$. Thus, the expectation of the outer product of $\tilde{\boldsymbol{f}}(n+1)$ is

$$\begin{aligned}\boldsymbol{F}(n+1) &= \boldsymbol{F}(n) + \mu\,\mathrm{E}\big\{(\psi(y_n)+\epsilon_n)\tilde{\boldsymbol{f}}(n)\boldsymbol{r}^t(n)\big\} + \mu\,\mathrm{E}\big\{(\psi(y_n)+\epsilon_n)\boldsymbol{r}(n)\tilde{\boldsymbol{f}}^t(n)\big\} \\ &\quad + \mu^2\,\mathrm{E}\big\{\boldsymbol{r}(n)\boldsymbol{r}^t(n)\psi^2(y_n)\big\} + 2\mu^2\,\mathrm{E}\big\{\boldsymbol{r}(n)\boldsymbol{r}^t(n)\psi(y_n)\epsilon_n\big\} \\ &\quad + \mu^2\,\mathrm{E}\big\{\boldsymbol{r}(n)\boldsymbol{r}^t(n)\epsilon_n^2\big\}.\end{aligned}$$

The quantization noise properties (7.9) and (7.10) can be applied to simplify the previous expression.

$$\begin{aligned}\boldsymbol{F}(n+1) &= \boldsymbol{F}(n) + \mu\,\mathrm{E}\big\{\psi(y_n)\tilde{\boldsymbol{f}}(n)\boldsymbol{r}^t(n)\big\} + \mu\,\mathrm{E}\big\{\psi(y_n)\boldsymbol{r}(n)\tilde{\boldsymbol{f}}^t(n)\big\} \\ &\quad + \mu^2\alpha^2\,\mathrm{E}\big\{\boldsymbol{r}(n)\boldsymbol{r}^t(n)\big\}.\end{aligned}$$

Applying the small-error approximation $\psi(y_n) \approx (\gamma - 3s_{n-\delta}^2)\boldsymbol{r}^t(n)\tilde{\boldsymbol{f}}(n) + \psi(s_{n-\delta})$ from Section 7.3.4, the outer product recursion is well described, for small $\tilde{\boldsymbol{f}}(n)$, by

$$\begin{aligned}\boldsymbol{F}(n+1) &= \boldsymbol{F}(n) + \mu\,\mathrm{E}\big\{(\gamma - 3s_{n-\delta}^2)\tilde{\boldsymbol{f}}(n)\tilde{\boldsymbol{f}}^t(n)\boldsymbol{r}(n)\boldsymbol{r}^t(n)\big\} \\ &\quad + \mu\,\mathrm{E}\big\{(\gamma - 3s_{n-\delta}^2)\boldsymbol{r}(n)\boldsymbol{r}^t(n)\tilde{\boldsymbol{f}}(n)\tilde{\boldsymbol{f}}^t(n)\big\} + \mu\,\mathrm{E}\big\{\psi(s_{n-\delta})\tilde{\boldsymbol{f}}(n)\boldsymbol{r}^t(n)\big\} \\ &\quad + \mu\,\mathrm{E}\big\{\psi(s_{n-\delta})\boldsymbol{r}(n)\tilde{\boldsymbol{f}}^t(n)\big\} + \mu^2\alpha^2\,\mathrm{E}\big\{\boldsymbol{r}(n)\boldsymbol{r}^t(n)\big\}. \qquad (7.33)\end{aligned}$$

The individual terms in (7.33) are successively analyzed below.

The second and third terms in (7.33) are transposes of one another. For now we concentrate on the first of the pair, for which we can use (B1) and the fact that $\mathrm{E}\{\boldsymbol{r}(n)\boldsymbol{r}^t(n)\} = \sigma_s^2\boldsymbol{\mathcal{H}}\boldsymbol{\mathcal{H}}^t = \sigma_s^2\boldsymbol{R}$ to write

$$\mathrm{E}\big\{(\gamma - 3s_{n-\delta}^2)\tilde{\boldsymbol{f}}(n)\tilde{\boldsymbol{f}}^t(n)\boldsymbol{r}(n)\boldsymbol{r}^t(n)\big\} = \boldsymbol{F}(n)\Big(\sigma_s^2\gamma\boldsymbol{R} - 3\,\mathrm{E}\big\{s_{n-\delta}^2\boldsymbol{r}(n)\boldsymbol{r}^t(n)\big\}\Big).$$

Since $\mathrm{E}\big\{s_{n-\delta}^2\boldsymbol{r}(n)\boldsymbol{r}^t(n)\big\} = \boldsymbol{\mathcal{H}}\,\mathrm{E}\big\{s_{n-\delta}^2\boldsymbol{s}(n)\boldsymbol{s}^t(n)\big\}\boldsymbol{\mathcal{H}}^t$, we define the matrix $[a_{i,j}] =$

$\mathrm{E}\big\{s_{n-\delta}^2 \boldsymbol{s}(n)\boldsymbol{s}^t(n)\big\}$ with elements

$$a_{i,j} \;=\; \mathrm{E}\{s_{n-\delta}^2 s_{n-i} s_{n-j}\} \;=\; \begin{cases} 0 & i \neq j, \\ \sigma_s^4 & i = j \neq \delta, \\ \mathrm{E}\{s_{n-\delta}^4\} & i = j = \delta. \end{cases}$$

Then in matrix notation, $[a_{i,j}] = \sigma_s^4 \mathbf{I} + \big(\mathrm{E}\{s_{n-\delta}^4\} - \sigma_s^4\big)\mathbf{e}_\delta \mathbf{e}_\delta^t$ (where $\mathbf{e}_\delta$ is a vector with a one in the δ^{th} position and zeros elsewhere). Incorporating the definition of κ_s from (A3), we conclude that

$$\mathrm{E}\big\{s_{n-\delta}^2 \boldsymbol{r}(n)\boldsymbol{r}^t(n)\big\} \;=\; \sigma_s^4 \boldsymbol{\mathcal{H}}\boldsymbol{\mathcal{H}}^t + \sigma_s^4(\kappa_s - 1)\boldsymbol{\mathcal{H}}\mathbf{e}_\delta \mathbf{e}_\delta^t \boldsymbol{\mathcal{H}}^t.$$

For long equalizers (i.e., $N_f \gg 1$), the second term in the preceding equation is dominated by the first, so that we can approximate

$$\mathrm{E}\big\{s_{n-\delta}^2 \boldsymbol{r}(n)\boldsymbol{r}^t(n)\big\} \;\approx\; \sigma_s^4 \boldsymbol{\mathcal{H}}\boldsymbol{\mathcal{H}}^t \;=\; \sigma_s^4 \boldsymbol{R}.$$

Finally, since $\gamma = \sigma_s^2 \kappa_s$, these approximations yield

$$\begin{aligned} &\mu\,\mathrm{E}\big\{(\gamma - 3s_{n-\delta}^2)\tilde{\boldsymbol{f}}(n)\tilde{\boldsymbol{f}}^t(n)\boldsymbol{r}(n)\boldsymbol{r}^t(n)\big\} + \mu\,\mathrm{E}\big\{(\gamma - 3s_{n-\delta}^2)\boldsymbol{r}(n)\boldsymbol{r}^t(n)\tilde{\boldsymbol{f}}(n)\tilde{\boldsymbol{f}}^t(n)\big\} \\ &\qquad = -\mu(3 - \kappa_s)\sigma_s^4\big(\boldsymbol{F}(n)\boldsymbol{R} + \boldsymbol{R}\boldsymbol{F}(n)\big). \end{aligned}$$

As for the fourth and fifth terms of (7.33), notice that (B1) implies

$$\mathrm{E}\big\{\psi(s_{n-\delta})\tilde{\boldsymbol{f}}(n)\boldsymbol{r}^t(n)\big\} = \mathrm{E}\big\{\tilde{\boldsymbol{f}}(n)\big\}\,\mathrm{E}\big\{\psi(s_{n-\delta})\boldsymbol{s}^t(n)\big\}\boldsymbol{\mathcal{H}}^t.$$

As we know from Section 7.4.1, the dispersion constant is selected to force

$$\mathrm{E}\big\{\psi(s_{n-\delta})\boldsymbol{s}^t(n)\big\} = 0.$$

Thus, the fourth and fifth terms of (7.33) vanish.

Re-writing the final term of (7.33), the approximated outer product recursion (valid for small $\tilde{\boldsymbol{f}}(n)$ and $N_f \gg 1$) becomes

$$\boldsymbol{F}(n+1) \;=\; \boldsymbol{F}(n) - \mu(3 - \kappa_s)\sigma_s^4\big(\boldsymbol{F}(n)\boldsymbol{R} + \boldsymbol{R}\boldsymbol{F}(n)\big) + \mu^2\alpha^2\sigma_s^2\boldsymbol{R}.$$

7.C Derivation of J_{ex}

In this appendix, we use (7.21) to determine an expression for the steady-state EMSE achieved by DSE-CMA. A similarity transformation of the symmetric Toeplitz matrix $\boldsymbol{R}$ is employed to simplify the derivation: $\boldsymbol{R} = \boldsymbol{Q}\boldsymbol{\Lambda}\boldsymbol{Q}^t$, where the matrix $\boldsymbol{\Lambda}$ is diagonal and the matrix $\boldsymbol{Q}$ is orthogonal. Applying this transformation to $\boldsymbol{F}(n)$ yields $\boldsymbol{F}(n) = \boldsymbol{Q}\boldsymbol{X}(n)\boldsymbol{Q}^t$, where $\boldsymbol{X}(n)$ is, in general, *not* diagonal. Using the properties of the trace operator and the fact that $\boldsymbol{Q}^t\boldsymbol{Q} = \mathbf{I}$, we can express the EMSE from (7.20) in terms of the transformed variables:

$$J_{\text{ex}}(n) \;=\; \sigma_s^2 \,\text{tr}\big(\boldsymbol{\Lambda}\boldsymbol{X}(n)\big).$$

The diagonal nature of $\boldsymbol{\Lambda}$ implies $J_{\text{ex}}(n) = \sigma_s^2 \sum_i \lambda_i x_i(n)$, where λ_i and $x_i(n)$ represent the i^{th} diagonal elements of $\boldsymbol{\Lambda}$ and $\boldsymbol{X}(n)$, respectively.

The similarity transformation can be applied to (7.21) to obtain a recursion in terms of $\boldsymbol{X}(n)$.

$$\boldsymbol{X}(n+1) \;=\; \boldsymbol{X}(n) - \mu(3-\kappa_s)\sigma_s^4\big(\boldsymbol{X}(n)\boldsymbol{\Lambda} + \boldsymbol{\Lambda}\boldsymbol{X}(n)\big) + \mu^2\alpha^2\sigma_s^2\boldsymbol{\Lambda}.$$

For the characterization of J_{ex}, we are interested in only the steady-state values of the diagonal elements $x_i(n)$. In terms of the i^{th} element,

$$x_i(n+1) \;=\; x_i(n) - 2\mu(3-\kappa_s)\sigma_s^4 x_i(n)\lambda_i + \mu^2\alpha^2\sigma_s^2\lambda_i.$$

Because $|x_i(n+1) - x_i(n)| \to 0$ as $n \to \infty$, the limit of the previous equation becomes

$$2\mu(3-\kappa_s)\sigma_s^4 x_i\lambda_i \;=\; \mu^2\alpha^2\sigma_s^2\lambda_i,$$

where we have introduced the shorthand notation $x_i = \lim_{n\to\infty} x_i(n)$. We can now sum over i to obtain

$$J_{\text{ex}} \;=\; \frac{\mu\alpha^2}{2(3-\kappa_s)} \sum_i \lambda_i.$$

Using the fact that $\sum_i \lambda_i = \mathrm{tr}(\boldsymbol{R}) = \mathrm{E}\{\boldsymbol{r}^t(n)\boldsymbol{r}(n)\} = N_f\sigma_r^2/\sigma_s^2$, we finalize our approximation for J_{ex}, the asymptotic EMSE of DSE-CMA:

$$J_{\mathrm{ex}} = \frac{\mu\alpha^2 N_f \sigma_r^2}{2(3-\kappa_s)\sigma_s^2}.$$

Chapter 8

Concluding Remarks

This dissertation considers blind estimation without priors (BEWP): the estimation of an i.i.d. signal distorted by a multichannel linear system and corrupted by additive noise, wherein the distribution of the signal, the distribution of the noise, and the structure of the linear system are all unknown. As shown in Chapter 2, the independence of the signal and the linearity of the distortion lead to broad class of admissible estimation criteria for which perfect blind linear estimation (PBLE) is possible under ideal conditions. By PBLE, we mean perfect signal estimation modulo unknown (but fixed) delay and scaling—ambiguities that were shown to be inherent to the BEWP problem definition. It was shown that the admissible criteria are those rewarding something akin to "distance of estimate distribution from Gaussian" and include, as perhaps the simplest example, kurtosis maximization. It was also shown that kurtosis maximization is equivalent to dispersion minimization when the desired signal is sub-Gaussian, providing a link between the popular-in-practice constant modulus (CM) criterion and more formal elements of estimation theory.

8.1 Summary of Original Work

Chapters 3 and 4 investigated the performance of kurtosis-maximizing and dispersion-minimizing criteria, respectively, under a very general set of non-ideal conditions. (Recall from Section 2.2 that our model allowed vector-valued IIR channels, constrained vector-valued ARMA estimators, and near-arbitrary signal and interference distribution.) In this general setting, we derived (i) simple conditions for the existence of blind linear estimators and (ii) tight bounding expressions for the conditionally-unbiased mean squared estimation error (UMSE) of these estimators. The bounds are a function of (a) signal and interference kurtoses and (b) the UMSE of the optimal linear estimator under the same conditions. It is important to note that the bounds are *not* a direct function of the channel structure nor the interference spectrum; such features affect blind estimation performance indirectly through their effect on optimal performance. Perhaps the most important feature of these bounds is that they prove that there exist many situations in which *blind* linear performance is nearly identical to *optimal* linear performance. In other words, the absence of distributional or structural knowledge in the formulation of the linear estimation problem does not significantly hinder the resulting mean-squared error performance.

Notwithstanding the good performance of CM-minimizing (i.e., dispersion minimizing) estimates, there remains the question of how to *obtain* these estimates. When using gradient descent (GD) methods, as is typical in practice, there exists the possibility that the GD algorithm will converge to an estimator for a source of interference rather than for the desired signal. Should this happen, the resulting estimates will be useless. In response to this problem, Chapter 5 derived conditions on the GD initialization sufficient for desired user convergence. These conditions are a function of principally the signal to interference-plus-noise (SINR) ratio of the initial estimates. It should be noted that there exists a broad class of problems (including, e.g., the typical data communication application) for which the critical SINR is 3.8 dB. The implication of these initialization conditions is that estimation schemes capable of guaranteeing only modest desired-user estimates can be used to initialize CM-GD, thereby inheriting the near-optimal asymptotic performance of CM-minimizing estimates.

In Chapter 6 the focus shifted from blind estimation of symbol sequences to blind identification of channel impulse responses. There we analyzed the performance of a classical blind identification method in which blind symbol estimates are cross-correlated with delayed copies of the received signal. By considering linear symbol estimates that minimize the CM cost, we were able to leverage the results of Chapter 4 in the derivation of average squared parameter error (ASPE) bounds for blind channel estimates.

Chapter 7 studied the efficient implementation of CM-GD algorithms, motivated by the practical application of CM methods in low cost or otherwise computationally demanding scenarios. Specifically, we presented a novel CM-GD algorithm that eliminates the estimator update multiplications required by standard CM-GD (i.e., CMA) while retaining identical transient and steady-state mean behaviors. Our algorithm, referred to as the dithered signed-error CM algorithm (DSE-CMA), is a modification of the standard signed-error approach to stochastic gradient descent in which a judicious incorporation of dither results in mean behavior identical to that of unsigned CMA. Though the cost of dithering manifests as increased excess MSE, Chapter 7 characterizes the excess MSE performance of DSE-CMA so that implementers can choose algorithm parameters accordingly.

Table 8.1: Correspondence between dissertation chapters and journal submissions/publications.

Chapter	Journal Submission/Publication
3	"Existence and performance of Shalvi-Weinstein estimators," by P. Schniter and L. Tong, to be submitted to *IEEE Trans. on Signal Processing*, Apr. 2000.
4	"Bounds for the MSE performance of constant modulus estimators," by P. Schniter and C.R. Johnson, Jr., to appear in *IEEE Trans. on Information Theory*, 2000.
5	"Sufficient conditions for the local convergence of constant modulus algorithms," by P. Schniter and C.R. Johnson, Jr., to appear in *IEEE Trans. on Signal Processing*, 2000.
6	"Performance analysis of Godard-based blind channel identification," by P. Schniter, R. Casas, A. Touzni, and C.R. Johnson, Jr., submitted to *IEEE Trans. on Signal Processing*, Sep. 2000.
7	"Dithered signed-error CMA: Robust, computationally efficient, blind adaptive equalization," by P. Schniter and C.R. Johnson, Jr., *IEEE Trans. on Signal Processing*, vol. 47, no. 6, pp. 1592-1603, June 1999.

Table 8.1 lists the correspondence between the chapters of this dissertation and submissions/publications in IEEE journals.

8.2 Possible Future Work

Multiuser Extensions

With regard to the performance bounds and convergence conditions of Chapters 3–6, there exist natural extensions from the single-estimator model of Fig. 2.3 to a multi-estimator (or "joint" estimator) model. As a starting point, one might consider kurtosis maximizing or dispersion minimizing schemes with additional intra-user-correlation penalties similar to those discussed in [Papadias Chap 00]. Perhaps user-averaged or worst-user UMSE bounds could be calculated for such criteria. Generalizing further, one might wonder: Does penalizing each user's own adjacent-symbol correlations (i.e., even in the single-user case) yield increased robustness?

BEWP using General Criteria

The analyses in this dissertation target the kurtosis and dispersion criteria, both functions of only the second- and fourth-order moments of the linear estimates. Since it seems likely that incorporating additional information into an estimation criterion will lead to improved estimates, can we say anything about the performance of criteria that utilize the *entire* distribution of the estimate? (Examples of such criteria can be found in [Wu NNSP 99].) Going further, what is the optimal linear estimator for the BEWP problem? Though the intuition developed in Section 2.1 holds, it is not clear that the analytical techniques of Chapters 3 and 4 are applicable.

Finally, we might wonder: What is the optimal (perhaps non-linear) estimator for BEWP? Unfortunately, the intuition developed in Section 2.1 does not hold because the estimates are no longer linear combinations of i.i.d. random variables. Though the limiting performance of BEWP problem is of fundamental importance to the theory of blind estimation, we have unfortunately little to say about it at this time.

Bibliography

[Abed-Meraim TSP 97] K. Abed-Meraim, E. Moulines and P. Loubaton, "Prediction Error Method for Second-Order Blind Identification," *IEEE Trans. on Signal Processing*, vol. 45, no. 3, pp. 694-705, Mar. 1997.

[Akay Book 96] M. Akay, *Detection and Estimation Methods for Biomedical Signals*, New York, NY: Academic, 1996.

[Alberi SPAWC 99] M.L. Aliberi, R.A. Casas, I. Fijalkow, and C.R. Johnson, Jr., "Looping LMS versus fast least squares algorithms: Who gets there first?," in *Proc. IEEE Workshop on Signal Processing Advances in Wireless Communication* (Annapolis, MD), pp. 296-9, May 1999.

[Anderson Book 89] B.D.O. Anderson and J.B. Moore, *Optimal Control: Linear Quadratic Methods*, Englewood Cliffs, NJ: Prentice-Hall, 1989.

[Batra GLOBE 95] A. Batra and J.R. Barry, "Blind cancellation of co-channel interference," in *Proc. IEEE Global Telecommunications Conf.* (Singapore), pp. 157-62, 13-17 Nov. 1995.

[Bell Chap 96] A.J. Bell and T.J. Segnowski, "Edges are the 'independent components' of natural scenes," in *Advances in Neural Information Processing Systems,* ed. M. Mozer, et al., Cambridge, MA: MIT Press, 1996, pp. 145-151.

[Benveniste Book 90] A. Benveniste, M. M'etivier, and P. Priouret, *Adaptive Algorithms and Stochastic Approximations*, Paris, France: Springer-Verlag, 1990.

[Bonnet ICASSP 84] M. Bonnet and O. Macchi, "An echo canceller having reduced size word taps and using the sign algorithm with extra controlled noise," in *Proc. IEEE Internat. Conf. on Acoustics, Speech, and Signal Processing* (San Diego, CA), pp. 30.2.1-4, Mar. 1984.

[Brown ALL 97] D.R. Brown, P. Schniter, and C.R. Johnson, Jr., "Computationally efficient blind equalization," in *Proc. Allerton Conf. on Communication, Control, and Computing* (Monticello, IL), pp. 54-63, Sep. 1997.

[Cadzow SPM 96] J.A. Cadzow, "Blind deconvolution via cumulant extrema," *IEEE Signal Processing Magazine*, vol. 13, no. 3, pp. 24-42, May 1996.

[Cardoso PROC 98] J.F. Cardoso, “Blind signal separation: Statistical principles,” *Proceedings of the IEEE—Special Issue on Blind System Identification and Estimation*, vol. 86, no. 10, pp. Oct. 1998, 2009-2025.

[Casas Chap 00] R.A. Casas, T.J. Endres, A. Touzni, C.R. Johnson Jr., and J.R. Treichler, “Current approaches to blind decision-feedback equalization,” to appear in *Signal Processing Advances in Communications, vol. 1,* (eds. G.B. Giannakis, P. Stoica, Y. Hua, and L. Tong), Wiley, 2000.

[Chen OE 92] Y. Chen, C.L. Nikias, J.G. Proakis, “Blind equalization with criterion with memory nonlinearity,” *Optical Engineering*, vol. 31 no. 6, pp. 1200-1210, June 1992.

[Chung ASIL 98] W. Chung and C.R. Johnson, Jr., “Characterization of the regions of convergence of CMA adapted blind fractionally spaced equalizer,” in *Proc. Asilomar Conf. on Signals, Systems and Computers* (Pacific Grove, CA), pp. 493-7, Nov. 1998.

[Chung Thesis 99] W. Chung, “Geometrical Understanding of the Constant Modulus Algorithm: Adaptive Blind Equalization and Cross-Polarized Source Separation,” *M.S. Thesis*, Cornell University, Ithaca, NY, 1999.

[Claerbout SEP 78] J.F. Claerbout, “Minimum information deconvolution,” Stanford Exploration Project, Report 15, pp. 109-22, 1978.

[Compton Book 88] R.T. Compton, *Adaptive Antennas: Concepts and Performance*, Englewood Cliffs, NJ: Prentice-Hall, 1988.

[Cover Book 91] T.M. Cover and J.A. Thomas, *Elements of Information Theory*, New York, NY: Wiley, 1991.

[Deller Book 93] J. Deller, J.G. Proakis, and J.H.L. Hansen, *Discrete Time Processing of Speech Signals*, Englewood Cliffs, NJ: Prentice-Hall, 1993.

[Donoho Chap 81] D.L. Donoho, “On minimum entropy deconvolution,” in *Applied Time Series Analysis II,* ed. D. Findley, New York, NY: Academic, 1981, pp. 565-608.

[Doyle Book 91] J.C. Doyle, B. Francis, and A. Tannenbaum, *Feedback Control Theory*, New York, NY: Macmillan, 1991.

[Duttweiler TASSP 82] D.L. Duttweiler, “Adaptive filter performance with nonlinearities in the correlation multiplier,” *IEEE Trans. on Acoustics, Speech, and Signal Processing*, vol. 30, no. 4, pp. 578-86, Aug. 1982.

[Endres TSP 99] T.J. Endres, B.D.O. Anderson, C.R. Johnson, Jr., and M. Green, “Robustness to fractionally-spaced equalizer length using the constant modulus criterion,” *IEEE Trans. on Signal Processing*, vol. 47, no. 2, pp. 544-9, Feb. 1999.

[Endres SPAWC 99] T.J. Endres, C.H. Strolle, S.N. Hulyalkar, T.A. Schaffer, A. Shah, M. Gittings, C. Hollowell, A. Bhaskaran, J. Roletter, B. Paratore, "Carrier independent blind initialization of a DFE using CMA," in *Proc. IEEE Workshop on Signal Processing Advances in Wireless Communication* (Annapolis, MD), pp. 239-42, May 1999.

[Feng TSP 99] C.C. Feng and C.Y. Chi, "Performance of cumulant based inverse filters for blind deconvolution," *IEEE Trans. on Signal Processing*, vol. 47, no. 7, pp. 1922-35, July 1999.

[Feng TSP 00] C.C. Feng and C.Y. Chi, "Performance of Shalvi and Weinstein's deconvolution criteria for channels with/without zeros on the unit circle," *IEEE Trans. on Signal Processing*, vol. 48, no. 2, pp. 571-5, Feb. 2000.

[Ferguson PSY 54] G.A. Ferguson, "The concept of parsimony in factor analysis," *Psychometrika*, vol. 19, no. 4, pp. 281-90, Dec. 1954.

[Fijalkow TSP 97] I. Fijalkow, A. Touzni, and J.R. Treichler, "Fractionally spaced equalization using CMA: Robustness to channel noise and lack of disparity," *IEEE Trans. on Signal Processing*, vol. 45, no. 1, pp. 56-66, Jan. 1997.

[Fijalkow TSP 98] I. Fijalkow, C. Manlove, and C.R. Johnson, Jr., "Adaptive fractionally spaced blind CMA adaptation: Excess MSE," *IEEE Trans. on Signal Processing*, vol. 46, no. 1, pp. 227-31, Jan. 1998.

[Foschini ATT 85] G.J. Foschini, "Equalizing without altering or detecting data (digital radio systems)," *AT&T Technical Journal*, vol. 64, no. 8, pp. 1885-911, Oct. 1985.

[Gitlin Book 92] R.D. Gitlin, J.F. Hayes, and S.B. Weinstein, *Data Communications Principles*, New York, NY: Plenum Press, 1992.

[Godard TCOM 80] D.N. Godard, "Self-recovering equalization and carrier tracking in two-dimensional data communication systems," *IEEE Trans. on Communications*, vol. 28, no. 11, pp. 1867-75, Nov. 1980.

[Godfrey SEP 78] R.J. Godfrey, "An information-theoretic approach to deconvolution," Stanford Exploration Project, Report 14, pp. 157-182, 1978.

[Gooch ICC 88] R.P. Gooch and J.C. Harp, "Blind channel identification using the constant modulus adaptive algorithm," in *Proc. IEEE Intern. Conf. on Communication* (Philadelphia, PA), pp. 75-9, June 1988.

[Gray TIT 93] R.M. Gray and T.G. Stockham, Jr., "Dithered quantizers," *IEEE Trans. on Information Theory*, vol. 39, no. 3, pp. 805-12, May 1993.

[Gray Thesis 79] W.C. Gray, "Variable norm deconvolution," *Ph.D. dissertation*, Stanford University, Palo Alto, CA, 1979.

[Gu TSP 99] M. Gu and L. Tong, "Geometrical characterizations of constant modulus receivers," *IEEE Trans. on Signal Processing*, vol. 47, no. 10, pp. 2745-2756, Oct. 1999.

[Haykin Book 92] S. Haykin, Ed., and A. Steinhardt, Contributor, *Adaptive Radar Detection and Estimation*, New York, NY: Wiley, 1992.

[Haykin Book 94] S. Haykin, Ed., *Blind Deconvolution*, Englewood Cliffs, NJ: Prentice-Hall, 1994.

[Haykin Book 96] S. Haykin, *Adaptive Filter Theory*, 3rd ed., Englewood Cliffs, NJ: Prentice-Hall, 1996.

[Holte TCOM 81] N. Holte and S. Stueflotten, "A new digital echo canceller for two-wire subscriber lines," *IEEE Trans. on Communications*, vol. 29, no. 11, pp. 1573-80, Nov. 1981.

[Jain Book 89] A.K. Jain, *Fundamentals of Digital Image Processing*, Englewood Cliffs, NJ: Prentice Hall, Inc., 1989.

[Johnson IJACSP 95] C.R. Johnson, Jr. and B.D.O. Anderson, "Godard blind equalizer error surface characteristics: White, zero-mean, binary case," *Internat. Journal of Adaptive Control & Signal Processing*, vol. 9, pp. 301-324, July-Aug. 1995.

[Johnson PROC 98] C.R. Johnson, Jr., P. Schniter, T.J. Endres, J.D. Behm, D.R. Brown, and R.A. Casas, "Blind equalization using the constant modulus criterion: A review," *Proceedings of the IEEE—Special Issue on Blind System Identification and Estimation*, vol. 86, no. 10, pp. 1927-50, Oct. 98.

[Johnson Chap 99] C.R. Johnson, Jr., P. Schniter, I. Fijalkow, L. Tong, J.D. Behm, M.G. Larimore, D.R. Brown, R.A. Casas, T.J. Endres, S. Lambotharan, A. Touzni, H.H. Zeng, M. Green, and J.R. Treichler, "The core of FSE-CMA behavior theory," to appear in *Unsupervised Adaptive Filtering, Volume 2: Blind Deconvolution*, ed. Simon Haykin, New York, NY: Wiley, 2000.

[Kagan Book 73] A.M. Kagan, Y.U. Linnik, and C.R. Rao, *Characterization Problems in Mathematical Statistics*, New York, NY: Wiley, 1973.

[Kaiser PSY 58] H.F. Kaiser, "The varimax criterion for analytic rotation in factor analysis," *Psychometrika*, vol. 23, no. 3, pp. 187-200, Sep. 1958.

[Kay Book 93] S.M. Kay, *Fundamentals of Statistical Signal Processing: Estimation Theory*, Englewood Cliffs, NJ: Prentice-Hall, 1993.

[Knuth WICASS 99] K.H. Knuth, "A Bayesian approach to source separation," in *Proc. Internat. Workshop on Independent Component Analysis and Signal Separation* (Aussios, France), pp. 283-8, 1999.

[Kundur SPM 96a] D. Kundur and D. Hatzinakos, "Blind Image Deconvolution," *IEEE Signal Processing Magazine*, vol. 13, no. 3, pp. 43-64, May 1996.

[Kundur SPM 96b] D. Kundur and D. Hatzinakos, "Blind Image Deconvolution Revisited," *IEEE Signal Processing Magazine*, vol. 13, no. 3, pp. 61-63, Nov. 1996.

[Lee Book 94] E.A. Lee and D.G. Messerschmitt, *Digital Communication*, 2nd ed., Boston, MA: Kluwer Academic Publishers, 1994.

[Li TSP 95] Y. Li and Z. Ding, "Convergence analysis of finite length blind adaptive equalizers," *IEEE Trans. on Signal Processing*, vol. 43, no. 9, pp. 2120-9, Sep. 1995.

[Li TSP 96a] Y. Li and Z. Ding, "Global convergence of fractionally spaced Godard (CMA) adaptive equalizers," *IEEE Trans. on Signal Processing*, vol. 44, no.4, pp. 818-26, Apr. 1996.

[Li TSP 96b] Y. Li, K.J.R. Liu, and Z. Ding, "Length and cost dependent local minima of unconstrained blind channel equalizers," *IEEE Trans. on Signal Processing*, vol. 44, no. 11, pp. 2726-35, Nov. 1996.

[Liu SP 96] H. Liu, G. Xu, L. Tong, and T. Kailath, "Recent developments in blind channel equalization: From cyclostationarity to subspaces," *Signal Processing*, vol. 50, pp. 83-9, 1996.

[Liu PROC 98] R. Liu and L. Tong, "Scanning the issue," *Proceedings of the IEEE—Special Issue on Blind System Identification and Estimation*, vol. 86, no. 10, pp. 1903-6, 0ct. 1998.

[Liu SP 99] D. Liu and L. Tong, "An analysis of constant modulus algorithm for array signal processing," *Signal Processing*, vol. 73, pp. 81-104, 1999.

[Ljung Book 99] L. Ljung, *System Identification: Theory for the User, 2nd Ed.*, Englewood Cliffs, NJ: Prentice Hall, 1999.

[Luenberger Book 69] D.G. Luenberger, *Optimization by Vector Space Methods*, New York, NY: Wiley, 1968.

[Macchi Book 95] O. Macchi, *Adaptive Processing*, New York, NY: Wiley, 1995.

[Mendel Book 83] J. Mendel, *Optimal Seismic Deconvolution: An Estimation Based Approach*, New York, NY: Academic, 1983.

[Naylor Book 82] A.W. Naylor and G.R. Sell, *Linear Operator Theory in Engineering and Science*, New York, NY: Springer-Verlag, 1982.

[Nunnally Book 78] J.C. Nunnally, *Psychometric Theory*, New York, NY: McGraw-Hill, 1978.

[Ooe GP 79] M. Ooe and T.J. Ulrych, "Minimum entropy deconvolution with exponential transformation," *Geophysical Prospecting*, vol. 27, pp. 458-73, 1979.

[Oppenheim Book 89] A.V. Oppenheim and R.W. Schafer, *Discrete-Time Signal Processing*, Englewood Cliffs, NJ: Prentice-Hall, 1989.

[Papadias SPL 96] C.B. Papadias and A.J. Paulraj, "A constant modulus algorithm for multiuser signal separation in presence of delay spread using antenna arrays," *IEEE Signal Processing Letters*, vol. 4, no. 6, pp. 178-81, June 1997.

[Papadias Chap 00] C.B. Papadias, "Blind separation of independent sources based on multiuser kurtosis optimization criteria," to appear in *Unsupervised Adaptive Filtering, Volume 2: Blind Deconvolution*, ed. Simon Haykin, New York, NY: Wiley, 2000.

[Papoulis Book 91] A. Papoulis, *Probability, random variables, and stochastic processes*, New York, NY: McGraw-Hill, 1991.

[Paulraj Chap 98] A.J. Paulraj, C.B. Papadias, V.U. Reddy, A.J. van der Veen, "Blind space-time processing," in *Wireless Communications: Signal Processing Perspectives,* eds. H.V. Poor and G.W. Wornell, Upper Saddle River, NJ: Prentice Hall, 1998, pp. 179-210.

[Poor Book 94] H.V. Poor, *An Introduction to Signal Detection and Estimation*, New York, NY: Springer-Verlag, 1994.

[Porat Book 94] B. Porat, *Digital Processing of Random Signals*, Englewood Cliffs, NJ: Prentice-Hall, 1994.

[Proakis SPIE 91] J.G. Proakis and C.L. Nikias, "Blind equalization," *The Internat. Society for Optical Engineering*, vol. 1565, pp. 76-87, 1991.

[Proakis Book 95] J.G. Proakis, *Digital Communications*, 3rd ed., New York, NY: McGraw-Hill, 1995.

[Qureshi PROC 85] S.U.H. Qureshi, "Adaptive Equalization," *Proceedings of the IEEE*, vol. 73, no. 9, pp. 1349-87, Sep. 1985.

[Regalia SP 99] P. Regalia, "On the equivalence between the Godard and Shalvi-Weinstein schemes of blind equalization," *Signal Processing*, vol. 73, nos. 1-2, pp. 185-90, Feb. 1999.

[Regalia TSP 99] P. Regalia and M. Mboup, "Undermodeled equalization: A characterization of stationary points for a family of blind criteria," *IEEE Trans. on Signal Processing*, vol. 47, no. 3, pp. 760-70, Mar. 1999.

[Robinson Book 86] E.A. Robinson and T. Durrani, *Geophysical Signal Processing*, Engelwood Cliffs, NJ: Prentice Hall, Inc., 1986.

[Rudin Book 76] W. Rudin, *Principles of Mathematical Analysis, 3rd Ed.*, New York, NY: McGraw-Hill, 1976.

[Saunders ETS 53] D.R. Saunders, "An analytic method for rotation to orthogonal simple structure," *Educational Testing Service Research Bulletin*, vol. 53, no. 10, pp. ??, 1953.

[Schniter ALL 98] P. Schniter and C.R. Johnson, Jr., "Minimum-entropy blind acquisition/equalization for uplink DS-CDMA," in *Proc. Allerton Conf. on Communication, Control, and Computing* (Monticello, IL), pp. 401-10, Oct. 1998.

[Schniter ASIL 98] P. Schniter and C.R. Johnson, Jr., "Dithered signed-error CMA: The complex-valued case," in *Proc. Asilomar Conf. on Signals, Systems and Computers* (Pacific Grove, CA), pp. 1143-7, Nov. 1998.

[Schniter TSP 99] P. Schniter and C.R. Johnson, Jr., "Dithered signed-error CMA: Robust, computationally efficient, blind adaptive equalization," *IEEE Trans. on Signal Processing*, vol. 47, no. 6, pp. 1592-1603, June 1999.

[Schniter TIT 00] P. Schniter and C.R. Johnson, Jr., "Bounds for the MSE performance of constant modulus estimators," to appear in *IEEE Trans. on Information Theory*, 2000.

[Schniter TSP 00] P. Schniter and C.R. Johnson, Jr., "Sufficient conditions for the local convergence of constant modulus algorithms," to appear in *IEEE Trans. on Signal Processing*, 2000.

[Schniter TSP tbd] P. Schniter, R. Casas, A. Touzni, and C.R. Johnson, Jr., "Performance analysis of Godard-based blind channel identification," submitted to *IEEE Trans. on Signal Processing*, Sep. 1999.

[Schniter TSP tbd2] P. Schniter and L. Tong, "Existence and performance of Shalvi-Weinstein estimators," In preparation.

[Sethares TSP 92] W. A. Sethares, "Adaptive Algorithms with Nonlinear Data and Error Functions," *IEEE Trans. on Signal Processing*, vol. 40, no. 9, pp. 2199-206, Sept. 1992.

[Shalvi TIT 90] O. Shalvi and E. Weinstein, "New criteria for blind deconvolution of nonminimum phase systems (channels)," *IEEE Trans. on Information Theory*, vol. 36, no. 2, pp. 312-21, Mar. 1990.

[Shannon BSTJ 48] C.E. Shannon, "A mathematical theory of communication," *Bell System Technical Journal*, vol. 27, pp. 379-423, 623-56, 1948.

[Shynk TSP 96] J.J. Shynk and R.P. Gooch, "The constant modulus array for cochannel signal copy and direction finding," *IEEE Trans. on Signal Processing*, vol. 44, no. 3, pp. 652-60, Mar. 1996.

[Stockham PROC 75] T. Stockham, T. Cannon, and R. Ingebretsen, "Blind deconvolution through digital signal processing," *Proceedings of the IEEE*, vol. 63, pp. 678-92, Apr. 1975.

[Tong CISS 92] L. Tong, "A fractionally spaced adaptive blind equalizer," in *Proc. Conf. on Information Science and Systems* (Princeton, NJ), pp. 711-16, Mar. 1992.

[Tong PROC 98] L. Tong and S. Perreau, "Blind channel estimation: From subspace to maximum likelihood methods," *Proceedings of the IEEE special issue on Blind System Identification and Estimation*, vol. 86, no. 10, pp. 1951-68, Oct. 1998.

[Torkkola WICASS 99] K. Torkkola, "Blind separation for audio signals—Are we there yet?," in *Proc. Internat. Workshop on Independent Component Analysis and Signal Separation* (Aussois, France), pp. 239-44, Jan. 1999.

[Touzni SPL 00] A. Touzni, L. Tong, R.A. Casas, and C.R. Johnson, Jr., "Vector-CM stable equilibrium analysis," *IEEE Signal Processing Letters*, vol. 7, no. 2, pp. 31-3, Feb. 2000.

[Touzni ICASSP 98] A. Touzni, I. Fijalkow, M. Larimore, and J.R. Treichler, "A globally convergent approach for blind MIMO adaptive deconvolution," in *Proc. IEEE Internat. Conf. on Acoustics, Speech, and Signal Processing* (Seattle, WA), pp. 2385-8, May 1998.

[Treichler TASSP 83] J.R. Treichler and B.G. Agee, "A new approach to multipath correction of constant modulus signals," *IEEE Trans. on Acoustics, Speech, and Signal Processing*, vol. ASSP-31, no.2, pp. 459-72, Apr. 1983.

[Treichler TASSP 85b] J.R. Treichler and M.G. Larimore, "New processing techniques based on the constant modulus adaptive algorithm," *IEEE Trans. on Acoustics, Speech, and Signal Processing*, vol. ASSP-33, no.2, pp. 420-31, Apr. 1985.

[Treichler TASSP 85a] J.R. Treichler and M.G. Larimore, "The tone capture properties of CMA-based interference suppressors," *IEEE Trans. on Acoustics, Speech, and Signal Processing*, vol. ASSP-33, no.4, pp. 946-58, Aug. 1985.

[Treichler SPM 96] J.R. Treichler, I. Fijalkow, and C.R. Johnson, Jr., "Fractionally-spaced equalizers: How long should they really be?," *IEEE Signal Processing Magazine*, vol. 13, No. 3, pp. 65-81, May 1996.

[Treichler PROC 98] J.R. Treichler, M.G. Larimore, and J.C. Harp, "Practical blind demodulators for high-order QAM signals," *Proceedings of the IEEE—Special Issue on Blind System Identification and Estimation*, vol. 86, no. 10, pp. 1907-26, Oct. 1998.

[vanderVeen PROC 98] A.J. van der Veen, "Algebraic methods for deterministic blind beamforming," *Proceedings of the IEEE—Special Issue on Blind System Identification and Estimation*, vol. 86, no. 10, pp. 1987-2008, Oct. 1998.

[VanTrees Book 68] H.L. Van Trees, *Detection, Estimation, and Modulation Theory, vol. 1*, New York, NY: Wiley, 1968.

[VanVeen ASSPM 88] B.D. Van Veen and K.M. Buckley, "Beamforming: a versatile approach to spatial filtering," *IEEE Acoustics Speech and Signal Processing Magazine*, vol. 5, pp. 4-24, 1988.

[Wiggins GEO 77] R.A. Wiggins, "Minimum entropy deconvolution," *Geoexploration*, vol. 16, pp. 21-35, 1978.

[Wu NNSP 99] H.C. Wu and J.C. Principe, "A Gaussianity measure for blind source separation insensitive to the sign of kurtosis," in *Proc. IEEE Workshop on Neural Networks for Signal Processing* (Madison, WI), pp. 58-66, Aug. 1999.

[Yang SPL 98] V.Y. Yang and D.L. Jones, "A vector constant modulus algorithm for shaped constellation equalization," *IEEE Signal Processing Letters*, vol. 5, no. 4, pp. 89-91, Apr. 1998.

[Zeng TIT 98] H.H. Zeng, L. Tong and C.R. Johnson, Jr., "Relationships between the constant modulus and Wiener receivers," *IEEE Trans. on Information Theory*, vol. 44, no. 4, pp. 1523-38, July 1998.

[Zeng TSP 99] H.H. Zeng, L. Tong, and C.R. Johnson, Jr., "An analysis of constant modulus receivers," *IEEE Trans. on Signal Processing*, vol. 47, no. 11, pp. 2990-9, Nov. 1999.