The Core of FSE-CMA Behavior Theory[1]

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

Abstract

This chapter presents the basics of the current theory regarding the behavior of blind fractionally-spaced and/or spatial-diversity equalizers (FSE) adapted via the constant modulus algorithm (CMA). The constant modulus algorithm, which was developed in the late 1970s and disclosed in the early 1980s, performs a stochastic gradient descent of a cost function that penalizes the dispersion of the equalizer output from a constant value. The constant modulus (CM) cost function leads to a blind algorithm because evaluation of the CM cost at the receiver does not rely on access to a replica of the transmitted source, as in so-called "trained" scenarios. The capability for blind start-up makes certain communication systems feasible in circumstances that do not admit training. The analytically convenient feature of the fractionally-spaced realization of a linear equalizer is the potential for perfect equalization in the absence of channel noise given a finite impulse response equalizer of time span matching that of the finite impulse response channel. The conditions for perfect equalization coupled with some mild conditions on the source can be used to establish convergence to perfect performance with FSE parameter adaptation by CMA from any equalizer parameter initialization. The FSE-CMA behavior theory presented here merges the taxonomy of the behavior theory of trained adaptive equalization and recent robustness analysis of FSE-CMA with violation of the conditions leading to perfect equalization and global asymptotic optimality of FSE-CMA.

[1]To appear as a chapter in the book *Unsupervised Adaptive Filtering*, Simon Haykin, ed., Wiley: New York, 1999.

1.1 Introduction

The revolution in data communications technology can be dated from the invention of automatic and adaptive channel equalization in the late 1960s... Many engineers contributed to this revolution, but the early inventions of Robert W. Lucky, particularly data-driven equalizer adaptation, were the largest factor in realizing higher-speed data communication in commercial equipment.

[Gitlin, Hayes, and Weinstein, *Data Communication Principles*, 1992, p. viii]

There are several applications in digital data communications when start-up and retraining of an adaptive equalizer has to be accomplished without the aid of a training sequence. Hence, the system has to be trained "blind"... We are interested in those circumstances where the eye is closed, and the conventional decision-directed operation will fail... It is recognized that, in exchange for not requiring data decisions, blind equalization algorithms may require one or two more orders of magnitudes of time to converge. There are two basic algorithms for blind equalization: the constant modulus algorithm (CMA)...

[ibid, p. 585]

Motivation

The desire to move data at high rates across transmission media with limited bandwidth has prompted the development of sophisticated communications systems, e.g., voiceband modems and microwave radio relay systems. Success in those applications has led to great interest in other communication scenarios in which economic or regulatory considerations limit the available transmission bandwidth. An important example of such an application is the wireless and cable distribution of digital television.

Central to the successful employment of most high data rate transmission systems is the use of *adaptive equalization* to counteract the disruptive effects of the signal's propagation from the transmitter to the receiver. The equalizer's importance, coupled with the fact that it tends to consume most of the receiver's computational resources and implementation cost, has made it the focus of much analytical and practical attention. Initially, high data rate communication systems utilized session-oriented point-to-point links that accommodated cooperative equalizer training. By training, we mean the transmission of a symbol sequence known in advance by the receiver and usually preceded by a clearly identifiable synchronization burst. The more recent emergence of digital multi-point and broadcast systems has produced communication scenarios where training is infeasible or prohibited. In this chapter we are interested in "blind" adaptive equalizers, that is, those that do not need training to achieve convergence from an unacceptable equalizer setting to an acceptable one. In a style intended for the engineer with a first-year-graduate level acquaintance with digital communication systems, this chapter presents the core of the behavior theory of the most popular of blind equalization algorithms, CMA, in the so-called "fractionally-spaced" configuration that dominates current practice. The theory chosen here was selected for its utility in illuminating pragmatic design guidelines.

History

The concept of an adaptive digital linear equalizer was introduced and realized in the 1960s. (See [Qureshi PROC 85] for an excellent survey of pre-1985 advances in trained adaptive equalization and numerous references regarding the highlights cited here.) The received signal's sampling interval matched the baud interval, that is, the time between transmission of consecutive source symbols. The baud-spaced equalizer tapped-delay-line length was selected to provide an accurate delayed inverse of a mixed-phase but finite-duration impulse response (FIR) channel. The common theoretical assumption of infinite equalizer length can be attributed to the recognition that an infinitely long tapped-delay line would be required for perfect[2] equalization of a FIR channel even in the absence of channel noise. Algorithms with and without training were introduced in the 1960s (e.g., [Lucky BSTJ 66]). The format with training quickly dominated telephony practice, at least for session start-up, and decision-directed LMS assumed the role of the fundamental blind method for subsequent tracking.

The 1970s witnessed the emergence of fractionally-spaced equalizer implementations, that is, those that used sampling rates faster than the source symbol rate. Improved band-edge equalization capabilities and reduced sensitivity to timing synchronization errors were cited as motivation [Ungerboeck TCOM 76]. The practical necessity of "tap leakage" for long fractionally-spaced equalizers was the most significant adaptive equalizer algorithm modification [Gitlin Book 92]. Performance analyses for both fractionally-spaced and baud-spaced equalizers commonly included assumptions of effectively infinite equalizer length, which permitted perfect equalization and easy translation between time and frequency domain interpretations.

During the 1980s, linear equalization methods capable of blind start-up moved from concept into practice. Blind equalization is desirable in multi-point and broadcast systems and necessary in non-invasive test and intercept scenarios. Even in point-to-point communication systems, blind equalization has been adopted for various reasons, including capacity gain and procedural convenience. For performance reasons, fractional spacing of the equalizer became preferred where technologically feasible. However, performance analysis of blind equalizers remained focused almost exclusively on baud-spaced realizations [Haykin Book 94].

During the 1990s, blind equalization has been incorporated into several emerging communication technologies [Treichler SPM 96], [Treichler PROC 98], e.g., digital cable TV. Also in the 1990s, realization of the ideal capabilities of fractionally-spaced data-adaptive equalizers, especially blind finite-length varieties

[2]Perfect equalization denotes situations in which the equalizer output sequence equals the transmitted symbol sequence up to a (fixed) unknown amplitude and delay.

[Tong TIT 95], have energized the study of finite-length fractionally-spaced blind equalizers. See, for example, [Johnson PROC 98], appearing in the special issue [Liu PROC 98]. The advantages that result from utilizing time-diversity (i.e., fractional sampling) also occur in spatial diversity systems (e.g., those employing multiple sensors or cross polarity) and code diversity systems (e.g., short-code DS-CDMA) [Paulraj SPM 97].

Our Goal: Behavior Theory Basics and Design Guidelines

The pedagogy employed here is fundamentally similar to that used in a variety of widely cited textbooks (e.g., [Proakis Book 95], [Gitlin Book 92], [Lee Book 94]) for trained, baud-spaced equalization theory, based on minimization of the mean squared error (MSE) in recovery of the training sequence. This approach is bolstered by recent results—to be described in this chapter—on the similarity of the locations of MSE minima and minima of the (blind) constant modulus (CM) cost function. Such similarities prompt the adoption of a taxonomy associated with design rules for trained stochastic gradient descent procedures, such as the LMS algorithm, to a stochastic gradient descent approach for minimizing the CM cost via the constant modulus algorithm (CMA). This results in design guidelines—to be developed and dissected in this chapter—regarding adaptive algorithm step-size selection, equalizer length, and equalizer parameter (re)initialization.

Content Map

Against this backdrop we present a map of the contents of this chapter, carrying us from a fractionally-spaced equalizer problem formulation to an understanding of the design guidelines for blind CMA-FSE, that is, CMA-based adaptation of a fractionally-spaced equalizer.

- Section 1.2 introduces the fractionally-spaced equalizer problem formulation. A multichannel model is adopted and the capability for perfect symbol recovery is established with a fractionally-spaced equalizer in the absence of channel noise. With the addition of channel noise, the Wiener solution (with a necessarily nonzero minimum mean-squared, delayed-source recovery error) is formulated for an infinite-duration impulse response (IIR) linear equalizer and for a finite-duration impulse response (FIR) linear equalizer. The resulting minimum mean-squared error is dissected in terms of its factors (i.e., noise power, equalizer length, channel convolution matrix singular values, and target system delay). Given a description of transient and asymptotic performance of the underlying average system behavior, a brief distillation of step-size and equalizer length design guidelines for LMS-FSE is also provided as the background against which CMA-FSE design guidelines will be composed.
- Section 1.3 begins with definition of the CM (or CMA 2-2) criterion and combines the perfect equalization requirements with some generic assumptions on the source statistics to result in a set of requirements for CMA-FSE's global asymptotic optimality. A series of 2-tap FSE CM cost functions and

CMA trajectories is used to illustrate the basic robustness properties with violations of each of these conditions. Approximate perturbation analyses of the effects of channel noise and equalizer length and a geometric analysis of the achieved MSE of a CM-minimizing equalizer in the presence of channel noise are exploited for insight. Differences in CMA-FSE relative to LMS-FSE are highlighted with examination of convergence rate and excess MSE.

- Section 1.4 focuses on three design choices in CMA-FSE implementation: adaptive equalizer parameter update step-size, equalizer length, and equalizer parameter initialization. Guidelines are developed through an example-driven tutorial approach. Single-spike initialization for CMA-BSE and double-spike initialization for CMA-FSE are discussed in terms of magnitude and location. In step-size selection, the tradeoffs in LMS design (i.e., (i) between convergence rate and excess MSE and (ii) between tracking error and gradient approximation error) are noted to drive CMA step-size selection as well. A similar tradeoff between improved modeling accuracy and increased excess mean squared error is discussed for increases in equalizer length.
- Section 1.5 presents three case studies, each of which yields a blind equalizer capable of dealing with a particular problem class (specifically, voice channel modem, cable-borne HDTV, and microwave radio) represented by signals and channel models in a publicly accessible database.

Notation

The following tables present the abbreviations and mathematical notation used throughout this chapter.

Table 1.1: Acronyms and Abbreviations

Acronym	Definition
BER	Bit error rate
BPSK	Binary phase-shift keying
BS	Baud-spaced
BSE	Baud-spaced equalizer
CM	Constant modulus
CMA	Constant modulus algorithm
DD	Decision-directed
EMSE	Excess mean-squared error
FIR	Finite-duration impulse response
FS	Fractionally-spaced
FSE	Fractionally-spaced equalizer
i.i.d.	Independent and identically distributed
IIR	Infinite-duration impulse response
ISI	Inter-symbol interference
LMS	Least mean-square
MMSE	Minimum mean-squared error
MSE	Mean-squared error
ODE	Ordinary differential equation
PAM	Pulse amplitude modulation
PBE	Perfect blind equalizability
pdf	Probability density function
PSK	Phase-shift keying
QAM	Quadrature amplitude modulation
QPSK	Quadrature phase-shift keying
SER	Symbol error rate
SNR	Signal to noise ratio
SPIB	Signal Processing Information Base
SVD	Singular value decomposition
ZF	Zero-forcing

Table 1.2: Mathematical Notation

Symbol	Definition
$(\cdot)^T$	Transposition
$(\cdot)^*$	Conjugation
$(\cdot)^H$	Hermitian transpose (i.e., conjugate transpose)
$(\cdot)^\dagger$	Moore-Penrose pseudo-inverse
$\mathrm{tr}(\cdot)$	Trace operator
$\lambda_{\min}(\cdot)$	Minimum eigenvalue
$\lambda_{\max}(\cdot)$	Maximum eigenvalue
$\\|\mathbf{x}\\|_p$	ℓ_p norm: $\sqrt[p]{\sum_n \|x_n\|^p}$
$\\|\mathbf{x}\\|_\mathbf{A}$	Norm defined by $\sqrt{\mathbf{x}^H \mathbf{A}\mathbf{x}}$ for positive definite Hermitian $\mathbf{A}$
$\\|\mathbf{x}(e^{j\omega})\\|$	L_2 norm: $\sqrt{\sum_n \int_{-\pi}^{\pi} \|x_n(e^{j\omega})\|^2 d\omega}$
$\mathbf{I}$	Identity matrix
$\mathbf{e}_i$	Column vector with 1 at the i^{th} entry $(i \geq 0)$ and zeros elsewhere
$\mathbb{R}$	The field of real-valued scalars
$\mathbb{C}$	The field of complex-valued scalars
$\mathrm{Re}\{\cdot\}$	Extraction of real-valued component
$\mathrm{Im}\{\cdot\}$	Extraction of imaginary-valued component
$\mathcal{Z}\{\cdot\}$	z-transform, i.e., $\mathcal{Z}\{x_n\} = \sum_n x_n z^{-n}$ for allowable $z \in \mathbb{C}$
$\mathrm{E}\{\cdot\}$	Expectation
$\nabla_\mathbf{f}$	Gradient with respect to $\mathbf{f}$: $\nabla_\mathbf{f} = \frac{\partial}{\partial \mathbf{f}_\mathrm{r}} + j\frac{\partial}{\partial \mathbf{f}_\mathrm{i}}$ where $\mathbf{f}_\mathrm{r} = \mathrm{Re}\{\mathbf{f}\}$ and $\mathbf{f}_\mathrm{i} = \mathrm{Im}\{\mathbf{f}\}$
$\mathcal{H}_\mathbf{f}$	Hessian with respect to $\mathbf{f}$: $\mathcal{H}_\mathbf{f} = \nabla_\mathbf{f} \nabla_{\mathbf{f}^T}$, for real-valued $\mathbf{f}$
	Bold lower-case Roman typeface designates vectors
	Bold upper-case Roman typeface designates matrices

Table 1.3: System Model Quantities

Symbol	Definition
T	Symbol period
n	Index for quantities sampled at baud intervals: $t = nT$
k	Index for fractionally sampled quantities: $t = kT/P$
δ	System delay (non-negative, integer-valued)
q_n	System impulse response coefficient
s_n	Source symbol
y_n	System/equalizer output
ν_n	Filtered noise contribution to system output
$q(z)$	System transfer function $\mathcal{Z}\{q_n\}$
$s(z)$	z-transformed source sequence $\mathcal{Z}\{s_n\}$
$y(z)$	z-transformed output sequence $\mathcal{Z}\{y_n\}$
$\nu(z)$	z-transformed noise sequence $\mathcal{Z}\{\nu_n\}$
$\mathbf{q}$	Vector of BS system response coefficients $\{q_n\}$
$\mathbf{s}(n)$	Vector of past source symbols $\{s_n\}$ at time n
$\mathbf{H}$	(Multi)channel convolution matrix
σ_s^2	Variance of source sequence: $\mathrm{E}\{\|s_n\|^2\}$
κ_s	Normalized kurtosis of source process: $\mathrm{E}\{\|s_n\|^4\}/\sigma_s^4$
κ_g	Normalized kurtosis of a Gaussian process

Table 1.4: Multirate Model Quantities

Symbol	Definition
P	Fractional sampling factor
h_k	Channel impulse response coefficient
f_k	Equalizer impulse response coefficient
w_k	Additive channel noise sample
r_k	Channel output (i.e., receiver input) sample
x_k	Noiseless channel output sample
$\mathbf{h}$	Vector of FS channel coefficients $\{h_k\}$
$\mathbf{f}$	Vector of FS equalizer coefficients $\{f_k\}$
$\mathbf{w}(n)$	Vector of FS channel noise samples $\{w_k\}$ at time n
$\mathbf{r}(n)$	Vector of FS channel outputs $\{r_k\}$ at time n
$\mathbf{x}(n)$	Vector of noiseless FS channel outputs $\{x_k\}$ at time n
σ_w^2	Variance of additive noise process: $\mathrm{E}\{\|w_k\|^2\}$
σ_r^2	Variance of FS received signal: $\mathrm{E}\{\|r_k\|^2\}$
κ_w	Normalized kurtosis of additive noise process: $\mathrm{E}\{\|w_k\|^4\}/\sigma_w^4$

Table 1.5: Multichannel Model Quantities

Symbol	Definition
P	Number of subchannels
L_h	Order of subchannel polynomials
L_f	Order of subequalizer polynomials
L_g	Order of subchannel polynomials' GCD
$h_n^{(p)}$	Impulse response coefficient n of p^{th} subchannel
$f_n^{(p)}$	Impulse response coefficient n of p^{th} subequalizer
$w_n^{(p)}$	Noise sample added to p^{th} subchannel at time n
$r_n^{(p)}$	Output of subchannel p at time n
$x_n^{(p)}$	Noiseless output of subchannel p at time n
$\mathbf{h}^{(p)}$	Vector of BS subchannel response coefficients $\{h_n^{(p)}\}$
$\mathbf{f}^{(p)}$	Vector of BS subequalizer response coefficients $\{f_n^{(p)}\}$
$h^{(p)}(z)$	Transfer function of p^{th} subchannel $\mathcal{Z}\{h_n^{(p)}\}$
$f^{(p)}(z)$	Transfer function of p^{th} subequalizer $\mathcal{Z}\{f_n^{(p)}\}$
$\mathbf{h}_n$	Vector-valued channel impulse response coefficient
$\mathbf{f}_n$	Vector-valued equalizer impulse response coefficient
$\mathbf{w}_n$	Vector-valued additive channel noise sample
$\mathbf{r}_n$	Vector-valued channel output (i.e., receiver input) sample
$\mathbf{x}_n$	Vector-valued noiseless channel output sample
$\mathbf{h}(z)$	Vector-valued channel transfer function $\mathcal{Z}\{\mathbf{h}_n\}$
$\mathbf{f}(z)$	Vector-valued equalizer transfer function $\mathcal{Z}\{\mathbf{f}_n\}$
$\mathbf{w}(z)$	Vector-valued z-transform of noise $\mathcal{Z}\{\mathbf{w}_n\}$
$\mathbf{r}(z)$	Vector-valued z-transform of channel output $\mathcal{Z}\{\mathbf{r}_n\}$
$\mathbf{x}(z)$	Vector-valued z-transform of noiseless channel output $\mathcal{Z}\{\mathbf{x}_n\}$

Table 1.6: Equalizer Design Quantities

Symbol	Definition
$J_{\mathrm{m}}^{(\delta)}(\cdot)$	MSE cost function for system delay δ
$J_{\mathrm{cm}}(\cdot)$	CM cost function
$\mathcal{E}_{\mathrm{m}}^{(\delta)}$	MMSE associated with system delay δ
$\mathcal{E}_{\chi}$	Excess MSE
$\mathbf{f}_{\mathrm{z}}^{(\delta)}$	Zero-forcing equalizer associated with system delay δ
$\mathbf{f}_{\mathrm{m}}^{(\delta)}$	Wiener equalizer associated with system delay δ
$\mathbf{f}_{\mathrm{c}}^{(\delta)}$	CM equalizer associated with system delay δ
$\mathbf{q}_{\mathrm{z}}^{(\delta)}$	System response achieved by ZF equalizer $\mathbf{f}_{\mathrm{z}}^{(\delta)}$
$\mathbf{q}_{\mathrm{m}}^{(\delta)}$	System response achieved by Wiener equalizer $\mathbf{f}_{\mathrm{m}}^{(\delta)}$
$\mathbf{q}_{\mathrm{c}}^{(\delta)}$	System response achieved by CM equalizer $\mathbf{f}_{\mathrm{c}}^{(\delta)}$
$\mathbf{R}_{r,r}$	Received signal autocorrelation matrix: $\mathrm{E}\{\mathbf{r}(n)\mathbf{r}^H(n)\}$
$\mathbf{R}_{x,x}$	Noiseless received signal autocorrelation matrix: $\mathrm{E}\{\mathbf{x}(n)\mathbf{x}^H(n)\}$
$\mathbf{d}_{r,s}^{(\delta)}$	Cross-correlation between the received and desired signal: $\mathrm{E}\{\mathbf{r}(n)s_{n-\delta}\}$
τ_{cma}	Time constant of CMA local convergence
μ	Step-size used in LMS and CMA
γ	CMA dispersion constant

1.2 MMSE Equalization and LMS

This section formulates the communications channel model and the fractionally-spaced equalization problem. In addition, it highlights basic results for zero-forcing and minimum mean-square-error (MMSE) linear equalizers and their adaptive implementation via the least mean squares (LMS) stochastic gradient descent algorithm. Section 1.3 will leverage these concepts to draw a parallel with CM receiver theory.

Although MMSE equalization is, in general, not optimal in the sense of minimizing symbol error rate (SER), it is perhaps the most widely used method in modem (among various other communication system) designs for inter-symbol interference (ISI) limited channels. Theoretically, the combination of coding and linear MMSE equalization offers a practical way to achieve channel capacity (even when SER is not minimized!) [Cioffi TCOM 95]. One advantage of the mean-squared-error (MSE) cost function is that it is quadratic and therefore unimodal (i.e., it is not complicated by the possibility of a MSE-minimizing algorithm (e.g., LMS) converging to a false local minimum). With all the merits of MMSE equalization, we will be motivated to compare blind equalizers, such as those minimizing the CM criterion, to MMSE equalizers.

Channel Models

Consider the equalization of linear, time-invariant, FIR channels transmitting an information symbol sequence $\{s_n\}$ as shown in Fig. 1.1(a). In a single-sensor scenario, the continuous-time received baseband signal $r(t)$ has the following form:

$$r(t) \quad = \quad \sum_{i=-\infty}^{\infty} s_i h(t-iT) + w(t), \tag{1.1}$$

where T is the symbol period, $h(t)$ is the continuous-time channel impulse response, and $w(t)$ represents additive channel noise. For simplicity, $w(t)$ is typically assumed to be a white Gaussian noise process. The model of the channel impulse response includes the (possibly unknown) pulse-shaping filter at the transmitter, the impulse response of the linear approximation to the propagation channel, and the receiver front-end filter (i.e., any filter prior to the equalizer).

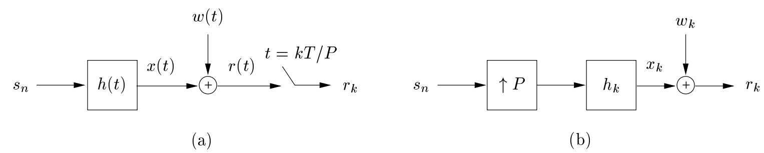

Figure 1.1: Two equivalent single-sensor models: (a) the continuous-time channel model, and (b) the discrete-time multirate channel model.

The discrete-time *multirate* channel model shown in Fig. 1.1(b) is obtained by uniformly sampling $r(t)$ at an integer[3] fraction of the symbol period, T/P. The fractionally-spaced (FS) channel output is then given by

$$r_k \;\stackrel{\Delta}{=}\; r\big(k\tfrac{T}{P}\big) \;=\; \sum_i s_i \underbrace{h\big(k\tfrac{T}{P} - iT\big)}_{h_{k-iP}} + \underbrace{w\big(k\tfrac{T}{P}\big)}_{w_k} \tag{1.2}$$

$$= \underbrace{\sum_i s_i h_{k-iP}}_{x_k} + w_k, \tag{1.3}$$

where the x_k are the (FS) noiseless channel outputs, the h_k are FS samples of the channel impulse response, and the w_k are FS samples of the channel noise process. (Throughout the chapter, the index "n" is reserved for baud-spaced quantities while the index "k" is applied to fractionally-spaced quantities.) For finite-duration channels, it is convenient to collect the fractionally-sampled channel response coefficients into the vector

$$\mathbf{h} = (h_0, h_1, h_2, \ldots, h_{(L_h+1)P-1})^T, \tag{1.4}$$

where L_h denotes the length of the channel impulse response in symbol intervals.

A particularly useful equivalent to the multirate model is the symbol-rate *multichannel* model shown in Fig. 1.2, where the p^{th} subchannel ($p = \{1, \ldots, P\}$) is obtained by sub-sampling $\mathbf{h}$ by the factor P. The respective multichannel quantities are, for the p^{th} subchannel,

$$h_n^{(p)} \stackrel{\Delta}{=} h_{(n+1)P-p}, \quad x_n^{(p)} \stackrel{\Delta}{=} x_{(n+1)P-p}, \quad r_n^{(p)} \stackrel{\Delta}{=} r_{(n+1)P-p}, \quad w_n^{(p)} \stackrel{\Delta}{=} w_{(n+1)P-p}. \tag{1.5}$$

Denoting the vector-valued channel response samples (at baud index n) and their z-transform by

$$\mathbf{h}_n \stackrel{\Delta}{=} \begin{pmatrix} h_n^{(1)} \\ \vdots \\ h_n^{(P)} \end{pmatrix} \;\stackrel{\mathcal{Z}}{\longrightarrow}\; \mathbf{h}(z), \tag{1.6}$$

we arrive at the following system of equations (in both time- and z-domains):

$$\mathbf{x}_n = \sum_{i=0}^{L_h} \mathbf{h}_i s_{n-i}, \qquad \mathbf{r}_n = \mathbf{x}_n + \mathbf{w}_n, \tag{1.7}$$

$$\mathbf{x}(z) = \mathbf{h}(z)s(z), \qquad \mathbf{r}(z) = \mathbf{x}(z) + \mathbf{w}(z), \tag{1.8}$$

Above, $\mathbf{x}_n$ denotes the vector-valued multichannel output without noise while $\mathbf{r}_n$ denotes the (noisy) received vector signal. Note that the multichannel vector quantities are indexed at the baud rate. We shall find this multichannel structure convenient in the sequel.

[3] In general, fractionally-spaced equalizers may operate at non-integer multiples of the baud rate. For simplicity, however, we restrict our attention to integer multiples.

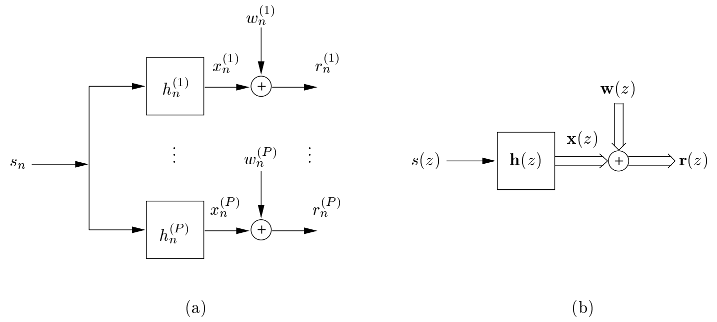

Figure 1.2: (a) Multichannel model and (b) its z-domain equivalent.

Though our derivation of the multichannel model originates from the single-sensor application of Fig. 1.1, the multichannel formulation applies directly to situations in which multiple sensors are used, with or without oversampling (see, e.g., [Moulines TSP 95]). In other words, the disparity achieved by oversampling in time or in space results in the same mathematical model. This is evident in Fig. 1.2, where $h_n^{(p)}$ would characterize the impulse response coefficients of the (BS) physical channel from the transmitter to the p^{th} sensor.

Linear Equalizers

For a fixed system delay δ, a fractionally-spaced equalizer (FSE) $\mathbf{f}$ is a linear estimator of the input $s_{n-\delta}$ given the multirate observation r_k or, equivalently, the multichannel observation $\mathbf{r}_n$ (both shown in Fig. 1.3). In the multirate setup, the T/P-spaced equalizer impulse response is specified by the coefficients $\{f_k\}$. In the multichannel setup, the baud-spaced vector-valued equalizer impulse response samples and their z-transform are given by

$$\mathbf{f}_n \stackrel{\Delta}{=} \begin{pmatrix} f_n^{(1)} \\ \vdots \\ f_n^{(P)} \end{pmatrix} \;\stackrel{\mathcal{Z}}{\longrightarrow}\; \mathbf{f}(z). \tag{1.9}$$

where $f_n^{(p)}$ are the coefficients of the p^{th} subequalizer. The multirate and multichannel equalizer coefficients are connected through the relationship $f_{nP+p-1} = f_n^{(p)}$.

The equalizer output y_n that estimates $s_{n-\delta}$ is given by the convolution

$$y_n \;=\; \sum_i \mathbf{f}_i^T \mathbf{r}_{n-i}, \tag{1.10}$$

and can be expressed in the z-domain as

$$y(z) \;=\; \mathbf{f}^T(z)\mathbf{r}(z) \;=\; \underbrace{\mathbf{f}^T(z)\mathbf{h}(z)}_{q(z)}\,s(z) + \underbrace{\mathbf{f}^T(z)\mathbf{w}(z)}_{\nu(z)} \tag{1.11}$$

$$\;=\; q(z)s(z) + \nu(z). \tag{1.12}$$

The corresponding multichannel system is depicted in Fig. 1.3. The transfer function $q(z)$ is often called the *combined channel-equalizer response* or the *system response*. (We will use the latter terminology for the remainder of the chapter.) Note that, as a polynomial in z, the system response is a baud-rate quantity. It is important to realize that, once restrictions are placed on the channel and/or equalizer, not all system responses may be attainable.

Next we consider the case where the channel and equalizer impulse responses are restricted to be finite in duration. In such a case, the estimate of $s_{n-\delta}$ is obtained from the past L_f+1 multichannel observations $\mathbf{r}_n$, where L_f+1 denotes the length of the multichannel equalizer. Specifically, we have

$$y_n \;=\; \sum_{i=0}^{L_f} \mathbf{f}_i^T \mathbf{r}_{n-i} \;=\; \mathbf{f}^T\mathbf{r}(n), \tag{1.13}$$

where

$$\mathbf{f} \triangleq \begin{pmatrix} \mathbf{f}_0 \\ \vdots \\ \mathbf{f}_{L_f} \end{pmatrix} = \begin{pmatrix} f_0 \\ \vdots \\ f_{P(L_f+1)-1} \end{pmatrix} \quad \text{and} \quad \mathbf{r}(n) \triangleq \begin{pmatrix} \mathbf{r}_n \\ \vdots \\ \mathbf{r}_{n-L_f} \end{pmatrix} = \begin{pmatrix} r_{(n+1)P-1} \\ \vdots \\ r_{(n-L_f)P} \end{pmatrix}. \tag{1.14}$$

The vector-valued received signal in (1.7) can be written as

$$\underbrace{\begin{pmatrix} \mathbf{r}_n \\ \vdots \\ \mathbf{r}_{n-L_f} \end{pmatrix}}_{\mathbf{r}(n)} = \underbrace{\begin{pmatrix} \mathbf{h}_0 & \cdots & \mathbf{h}_{L_h} & & \\ & \ddots & & \ddots & \\ & & \mathbf{h}_0 & \cdots & \mathbf{h}_{L_h} \end{pmatrix}}_{\mathbf{H}} \underbrace{\begin{pmatrix} s_n \\ \vdots \\ s_{n-L_f-L_h} \end{pmatrix}}_{\mathbf{s}(n)} + \underbrace{\begin{pmatrix} \mathbf{w}_n \\ \vdots \\ \mathbf{w}_{n-L_f} \end{pmatrix}}_{\mathbf{w}(n)} \tag{1.15}$$

$$\mathbf{r}(n) \;=\; \mathbf{H}\mathbf{s}(n) + \mathbf{w}(n). \tag{1.16}$$

As evident in (1.14), our construction ensures that the fractionally-sampled coefficients of the vector quantities $\mathbf{r}(n)$, $\mathbf{w}(n)$, $\mathbf{x}(n) \triangleq \mathbf{H}\mathbf{s}(n)$, and $\mathbf{f}$ are well-ordered with respect to the multirate time-index[4]. Substituting the received vector expression (1.16) into (1.13), we obtain the following system output, occurring at the baud rate:

$$y_n \;=\; \mathbf{f}^T\mathbf{H}\mathbf{s}(n) + \mathbf{f}^T\mathbf{w}(n) \tag{1.17}$$

$$\;=\; \mathbf{q}^T\mathbf{s}(n) + \nu_n. \tag{1.18}$$

The vector $\mathbf{q}$ represents the system impulse response (whose coefficients are sampled at the baud rate) and the quantity ν_n denotes the filtered channel-noise contribution to the system output.

[4]Note that the ordering of $\mathbf{h}_n$ in (1.6) implies $\mathbf{h} = (h_0, h_1, \ldots, h_{(L_h+1)P-1})^T \neq (\mathbf{h}_0^T, \mathbf{h}_1^T, \cdots, \mathbf{h}_{L_h}^T)^T$.

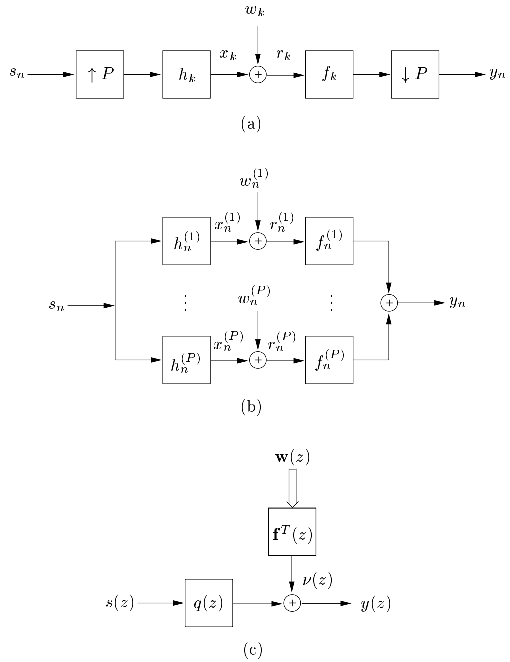

Figure 1.3: Equivalent system models: (a) multirate, (b) multichannel, and (c) their z-domain representation.

In measuring the performance of linear equalizers, we will be considering the mean-squared error (MSE) criterion. Given a fixed system delay δ, for mutually uncorrelated symbol and noise processes with variances $\mathrm{E}\{|s_n|^2\} = \sigma_s^2$ and $\mathrm{E}\{|w_k|^2\} = \sigma_w^2$, the MSE achieved by linear equalizer $\mathbf{f}$ is defined by

$$
\begin{aligned}
J_{\mathrm{m}}^{(\delta)}(\mathbf{f}) &\triangleq \mathrm{E}\{|\mathbf{f}^T\mathbf{r}(n) - s_{n-\delta}|^2\} &(1.19)\\
&= \mathrm{E}\{|\mathbf{q}^T\mathbf{s}(n) - s_{n-\delta} + \mathbf{f}^T\mathbf{w}(n)|^2\} &(1.20)\\
&= \underbrace{\sigma_s^2\|\mathbf{q}-\mathbf{e}_\delta\|_2^2}_{\text{ISI \& bias}} + \underbrace{\sigma_w^2\|\mathbf{f}\|_2^2}_{\text{noise enhancement}}. &(1.21)
\end{aligned}
$$

where $\mathbf{e}_\delta$ denotes a vector with a one in the δ^{th} position ($\delta \geq 0$) and zeros elsewhere.

From (1.21) we note that the MSE of a linear equalizer comes from two sources: (i) inter-symbol interference (ISI) plus bias and (ii) noise enhancement. ISI measures the effect of residual interference from other transmitted symbols, while bias refers to an incorrect amplitude estimation of the desired symbol. Both are minimized by making the system response close to the unit impulse, which leads to the so-called *zero-forcing equalizer* (see, e.g., [Lee Book 94]). For certain channels, however, reducing ISI and bias leads to an increase in equalizer norm and thus an enhancement in noise power. The *MMSE equalizer* achieves the optimal trade-off between ISI reduction and noise enhancement (for a particular noise level) in the sense of minimizing $J_m^{(\delta)}(\mathbf{f})$. In Section 1.3, we shall discuss how CM receivers, not designed to minimize the MSE criterion, achieve a similar compromise. In the remainder of the chapter, the terms "receiver" and "equalizer" will be used interchangeably.

Zero-Forcing Receivers

An equalizer capable of *perfect symbol recovery* in the absence of noise, that is, $y_n = s_{n-\delta}$ for some fixed delay δ, is called a zero-forcing (ZF) equalizer [Lucky BSTJ 66] and is denoted by $\mathbf{f}_z^{(\delta)}$. From (1.12), we see that perfect symbol recovery requires $\nu(z) = 0$ and $q(z) = z^{-\delta}$, implying an absence of channel noise and a particular relationship between the subchannels and subequalizers (discussed in detail below).

For nontrivial channels with finite-duration impulse response, baud-spaced ZF equalizers of finite-duration impulse response do not exist for reasons that will become evident in the discussion below. In contrast, FIR fractionally-spaced ZF equalizers do exist under particular conditions. A sufficient condition for perfect symbol recovery, referred to as *strong perfect equalization* (SPE), is that *any* system response $\mathbf{q}$ can be achieved through proper choice of equalizer $\mathbf{f}$. Applicable to both FIR and IIR channels, SPE guarantees perfect symbol recovery for any delay $0 \leq \delta \leq L_f + L_h$ (in the absence of noise). The SPE requirement may seem overly strong since it may be satisfactory to attain perfect equalization at only one particular delay. However, the class of channels allowing perfect symbol recovery for a restricted range of delays consists of primarily trivial channels. In other words, without SPE, it is usually impossible to achieve perfect symbol

recovery for a fixed delay.

A necessary and sufficient condition for SPE is that the channel matrix $\mathbf{H}$ is full column rank, which has implications for the subchannel and subequalizer polynomials $h^{(p)}(z) = \mathcal{Z}\{h_n^{(p)}\}$ and $f^{(p)}(z) = \mathcal{Z}\{f_n^{(p)}\}$, respectively. A fundamental requirement for SPE is that the subchannel polynomials must not all share a common zero, that is, $\{h^{(p)}(z)\}$ must be coprime[5]. This is often described by the condition: $\forall z,\ \mathbf{h}(z) \neq \mathbf{0}$. It can be shown [Tong TIT 95] that when the $\{h^{(p)}(z)\}$ are coprime, there exists a minimum equalizer length for which the channel matrix $\mathbf{H}$ has full column rank, thus ensuring SPE. Specifically, $L_f \geq L_h - 1$ is a sufficient equalizer length condition when the subchannels are coprime.

The subchannel polynomials are coprime if and only if there exists a set $\{f^{(p)}(z)\}$ that satisfies the Bezout equation [Kailath Book 80, Fuhrmann Book 96]:

$$
1 = \sum_{p=1}^{P} f^{(p)}(z)\, h^{(p)}(z) = \mathbf{f}^T(z)\mathbf{h}(z). \qquad (1.22)
$$

In other words, equalizer polynomials which satisfy the Bezout equation specify ZF equalizers.

We summarize our statements about perfect equalization in the following set of equivalences, valid in the absence of channel noise:

- Satisfaction of strong perfect equalization (SPE) conditions,
- Channel matrix $\mathbf{H}$ of full column-rank,
- Existence of zero-forcing equalizer for all system delays δ, where $0 \leq \delta \leq L_f + L_h$.
- Bezout equation satisfied.

To gain further insight into the SPE condition, it is useful to examine what happens when the subchannels are not coprime. For example, consider the case where $g(z) = 1 + g_1 z^{-1}$ can be factored out of every subchannel polynomial $\{h^{(1)}(z), \dots, h^{(P)}(z)\}$, leaving $\{\bar{h}^{(1)}(z), \dots, \bar{h}^{(P)}(z)\}$. It becomes clear that, for any set $\{f^{(p)}(z)\}$,

$$
\sum_{p=1}^{P} f^{(p)}(z)\, h^{(p)}(z) = (1 + g_1 z^{-1}) \sum_{p=1}^{P} f^{(p)}(z)\, \bar{h}^{(p)}(z) \neq 1. \qquad (1.23)
$$

Thus, the presence of the common subchannel factor $g(z)$ prevents the Bezout equation from being satisfied, making the ideal system response $q(z) = 1$ unattainable.

When the subchannels are not coprime, it may still be possible to approximate the perfect system response with a finite-length equalizer. In this case the equalizer is designed so that the remaining subchannel-

[5]Note that while coprimeness ensures the absence of any zero common to *all* subchannels in the set $\{h^{(p)}(z)\}$, it allows the existence of zeros common to strict subsets of $\{h^{(p)}(z)\}$.

subequalizer combinations approximate a delayed inverse of $g(z)$, that is,

$$\sum_{p=1}^{P} f^{(p)}(z)\,\bar{h}^{(p)}(z) \quad \approx \quad g^{-1}(z)\,z^{-\delta} \tag{1.24}$$

In general, the approximation improves as the equalizer length is increased, though performance will depend on the choice of system delay $\delta : 0 \le \delta \le L_f + L_{\bar{h}}$. The implication here is that long enough equalizers can well approximate zero-forcing equalizers even in the presence of common subchannel roots (as long as the common roots do not lie on the z-plane's unit circle).

We can also examine the effect of common zero(s) in the time domain via a decomposition of the channel matrix $\mathbf{H}$. If an order L_g polynomial $g(z)$ can be factored out of every subchannel, then it can be factored out of each row of the vector polynomial $\mathbf{h}(z)$ leaving $\bar{\mathbf{h}}(z)$ (of order $L_h - L_g$). We exploit this in the decomposition $\mathbf{H} = \bar{\mathbf{H}}\mathbf{G}$, where

$$\bar{\mathbf{H}} = \begin{pmatrix} \bar{\mathbf{h}}_0 & \cdots & \bar{\mathbf{h}}_{L_h-L_g} & & \\ & \ddots & & \ddots & \\ & & \bar{\mathbf{h}}_0 & \cdots & \bar{\mathbf{h}}_{L_h-L_g} \end{pmatrix} \quad \text{and} \quad \mathbf{G} = \begin{pmatrix} g_0 & \cdots & g_{L_g} & & \\ & \ddots & & \ddots & \\ & & g_0 & \cdots & g_{L_g} \end{pmatrix}. \tag{1.25}$$

The matrix $\bar{\mathbf{H}}$ is full column rank with dimension $P(L_f+1) \times (L_f+L_h-L_g+1)$, while $\mathbf{G}$ is full row rank with dimension $(L_f+L_h-L_g+1) \times (L_f+L_h+1)$. Since the rank of $\bar{\mathbf{H}}$ cannot exceed $(L_f+L_h-L_g+1)$ and $\mathbf{H}$ has (L_f+L_h+1) columns, the choice of $L_g > 0$ prevents $\mathbf{H}$ from achieving full column rank.

Finally, it is worth mentioning that the presence of a common subchannel root is associated with P roots of the FS channel polynomial (i.e., $h_0 + h_1 z^{-1} + \cdots + h_{P(L_h+1)} z^{-P(L_h+1)}$) lying equally spaced on a circle in the complex plane [Tong TIT 95]. In the case of $P = 2$, this implies that common subchannel roots are equivalent to FS channel roots reflected across the origin.

Wiener Receivers

For a fixed system delay δ, the Wiener receiver $\mathbf{f}_{\mathrm{m}}^{(\delta)}$ estimates the source symbol $s_{n-\delta}$ by minimizing the MSE cost

$$J_{\mathrm{m}}^{(\delta)}(\mathbf{f}) \quad = \quad \mathrm{E}\{|\mathbf{f}^T\mathbf{r}(n) - s_{n-\delta}|^2\}, \tag{1.26}$$

$$\mathbf{f}_{\mathrm{m}}^{(\delta)} \quad \triangleq \quad \arg\min_{\mathbf{f}} J_{\mathrm{m}}^{(\delta)}(\mathbf{f}). \tag{1.27}$$

For notational simplicity, the remainder of Section 1.2 assumes that the input s_n is a zero-mean, unit-variance ($\sigma_s^2 = \mathrm{E}\{|s_n|^2\} = 1$), uncorrelated random process. Furthermore, we assume that the channel noise $\{w_k\}$ is an uncorrelated process with variance σ_w^2 that is uncorrelated with the source.

The theory of MMSE estimation is well established and widely accessible (see, e.g., [Haykin Book 96]). Since we will find it convenient to refer to the geometrical aspects of MMSE estimation, especially later in

our presentation of the MSE properties of the CM receiver, we introduce some of the basic concepts at this point. Minimizing MSE can be translated into the problem of finding the minimum Euclidean distance of a vector to a plane spanned by observations. Consider the Hilbert space defined by the joint probability distributions of all random variables [Caines Book 88]. By the orthogonality principle [Haykin Book 96], the Wiener receiver's output, say $\bar{y}_n = \mathbf{r}^T(n)\mathbf{f}_{\mathrm{m}}^{(\delta)}$, is obtained by projecting $s_{n-\delta}$ onto the subspace $\mathcal{Y}$ spanned by the observation contained in $\mathbf{r}(n)$, as shown in Fig. 1.4.

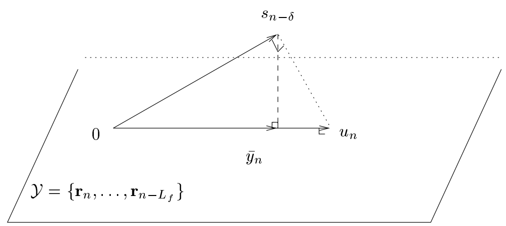

Figure 1.4: The geometrical interpretation of the Wiener estimator.

Figure 1.4 illustrates how, as an estimate of $s_{n-\delta}$, the Wiener output $\bar{y}_n$ is conditionally biased, that is,

$$\mathrm{E}\{\bar{y}_n | s_{n-\delta}\} \neq s_{n-\delta}. \tag{1.28}$$

The conditionally unbiased estimate of $s_{n-\delta}$, denoted here by u_n, is given by scaling $\bar{y}_n$ such that its projection onto the direction of $s_{n-\delta}$ is $s_{n-\delta}$ itself. It is important to note that SER performance is measured via u_n (rather than $\bar{y}_n$), and thus the comparison of the Wiener and CM receivers should also be made through the conditionally unbiased estimates. This idea will be revisited in Section 1.3.

For simplicity, we focus the remainder of the section on real-valued source, noise, channel, and equalizer quantities. Note, however, that the z-transforms of these real-valued quantities will be complex-valued.

The IIR Wiener Equalizer. For the IIR equalizer, we assume for simplicity that $\delta = 0$ and drop the superscript notation on $\mathbf{f}_{\mathrm{m}}^{(0)}$. This simplification is justified by the fact that the MSE performance of the IIR equalizer is independent of delay choice δ. The (non-causal) IIR Wiener receiver $\mathbf{f}_{\mathrm{m}}$ can be derived most easily in the z-domain. By the orthogonality principle,

$$\mathrm{E}\Big\{\mathbf{r}(z)\big(\mathbf{r}^T(\tfrac{1}{z^*})\mathbf{f}_{\mathrm{m}}(\tfrac{1}{z^*}) - s(\tfrac{1}{z^*})\big)^*\Big\} \quad = \quad \mathbf{0}. \tag{1.29}$$

Then, solving for $\mathbf{f}_{\mathrm{m}}(z)$,

$$\mathrm{E}\{\mathbf{r}(z)\mathbf{r}^H(\tfrac{1}{z^*})\}\mathbf{f}_{\mathrm{m}}^*(\tfrac{1}{z^*}) \quad = \quad \mathbf{h}(z). \tag{1.30}$$

Finally, using $\mathbf{r}(z) = \mathbf{h}(z)s(z) + \mathbf{w}(z)$ and the Matrix Inversion Lemma[6] [Kailath Book 80],

$$\mathbf{f}_{\mathrm{m}}(z) \;=\; \frac{1}{\mathbf{h}^T(z)\mathbf{h}^*(\frac{1}{z^*}) + \sigma_w^2}\mathbf{h}^*(\tfrac{1}{z^*}), \tag{1.31}$$

where $\sigma_w^2 = \mathrm{E}\{|w_k|^2\}$ is the noise power. By setting $z = e^{j\omega}$, we obtain the frequency response of the Wiener equalizer:

$$\mathbf{f}_{\mathrm{m}}(e^{j\omega}) \;=\; \frac{1}{\|\mathbf{h}(e^{j\omega})\|^2 + \sigma_w^2}\mathbf{h}^*(e^{j\omega}). \tag{1.32}$$

The denominator term $\|\mathbf{h}(e^{j\omega})\|^2 = \sum_{p=1}^{P} |h^{(p)}(e^{j\omega})|^2$ is sometimes called the *folded channel spectrum* [Lee Book 94, p. 228]. Recall that $h^{(p)}(e^{j\omega})$ is the frequency response of the p^{th} subchannel.

System Response. The combined channel-equalizer response $q_{\mathrm{m}}(z)$ resulting from IIR Wiener equalization is

$$q_{\mathrm{m}}(z) \;=\; \mathbf{f}_{\mathrm{m}}^T(z)\mathbf{h}(z) \;=\; \frac{\mathbf{h}^H(\frac{1}{z^*})\mathbf{h}(z)}{\mathbf{h}^H(\frac{1}{z^*})\mathbf{h}(z) + \sigma_w^2}, \tag{1.33}$$

or in the frequency domain

$$q_{\mathrm{m}}(e^{j\omega}) \;=\; \frac{\|\mathbf{h}(e^{j\omega})\|^2}{\|\mathbf{h}(e^{j\omega})\|^2 + \sigma_w^2}. \tag{1.34}$$

As $\sigma_w^2 \to 0$, $q_{\mathrm{m}}(z) \to 1$ as long as the subchannels have no common roots on the unit circle (i.e., $\forall\omega$, $\exists\, p,\ell$ s.t. $h^{(p)}(e^{j\omega}) \neq h^{(\ell)}(e^{j\omega})$). In this case, the folded spectrum has no nulls and perfect symbol recovery is achieved.

MMSE of IIR Wiener Receiver. Using (1.11), the estimation error of the IIR Wiener receiver is given by

$$e(z) \;=\; s(z) - y(z) \;=\; \big(1 - \mathbf{f}_{\mathrm{m}}^T(z)\mathbf{h}(z)\big)s(z) - \mathbf{f}_{\mathrm{m}}^T(z)\mathbf{w}(z). \tag{1.35}$$

Then, with the assumption that $\sigma_s^2 = 1$, the power spectrum of the error sequence of the Wiener filter has the form

$$S_e(\omega) \;=\; |1 - \mathbf{f}_{\mathrm{m}}^T(e^{j\omega})\mathbf{h}(e^{j\omega})|^2 + \sigma_w^2\|\mathbf{f}_{\mathrm{m}}(e^{j\omega})\|^2 \tag{1.36}$$

$$\;=\; \frac{\sigma_w^2}{\|\mathbf{h}(e^{j\omega})\|^2 + \sigma_w^2}. \tag{1.37}$$

The MSE $\mathcal{E}_{\mathrm{m}}$ of the Wiener filter is then given by

$$\mathcal{E}_{\mathrm{m}} \;=\; \frac{1}{2\pi}\int_{-\pi}^{\pi} \frac{\sigma_w^2}{\|\mathbf{h}(e^{j\omega})\|^2 + \sigma_w^2}\, d\omega. \tag{1.38}$$

The FIR Wiener Equalizer. When the vector equalizer polynomial $\mathbf{f}_{\mathrm{m}}^{(\delta)}(z)$ is of finite order, it is convenient to derive the Wiener equalizer in the time-domain. From Fig. 1.4, the Wiener receiver satisfies the

[6] The matrix inversion lemma is commonly written as $\mathbf{A}^{-1} = (\mathbf{B}^{-1} + \mathbf{C}\mathbf{D}^{-1}\mathbf{C}^H)^{-1} = \mathbf{B} - \mathbf{B}\mathbf{C}(\mathbf{D} + \mathbf{C}^H\mathbf{B}\mathbf{C})^{-1}\mathbf{C}^H\mathbf{B}$ where $\mathbf{A}$ and $\mathbf{B}$ are positive definite $M \times M$ matrices, $\mathbf{D}$ is a positive definite $N \times N$ matrix, and $\mathbf{C}$ is an $M \times N$ matrix. In deriving (1.31) we use the inversion lemma to find $\mathbf{A}^{-1}$, where $\mathbf{A} = \mathrm{E}\{\mathbf{r}(z)\mathbf{r}^H(\frac{1}{z^*})\}$, by choosing $\mathbf{B} = 1/\sigma_w^2$, $\mathbf{C} = \mathbf{h}(z)$, and $\mathbf{D} = 1/\sigma_s^2 = 1$.

orthogonality principle: $\big(s_{n-\delta} - y_n\big) \perp \mathbf{r}(n)$, or more specifically,

$$\mathrm{E}\big\{\mathbf{r}(n)(s_{n-\delta} - y_n)\big\} = \mathbf{0}. \tag{1.39}$$

Equation (1.39) leads, via (1.13), to the Wiener-Hopf equation whose solution is the Wiener receiver [Haykin Book 96]:

$$\mathbf{R}_{r,r}\mathbf{f}_{\mathrm{m}}^{(\delta)} = \mathbf{d}_{r,s}^{(\delta)} \quad\Rightarrow\quad \mathbf{f}_{\mathrm{m}}^{(\delta)} = \mathbf{R}_{r,r}^{\dagger}\mathbf{d}_{r,s}^{(\delta)}. \tag{1.40}$$

Here $\mathbf{R}_{r,r} \triangleq \mathrm{E}\{\mathbf{r}(n)\mathbf{r}^T(n)\}$, $\mathbf{d}_{r,s}^{(\delta)} \triangleq \mathrm{E}\{\mathbf{r}(n)\, s_{n-\delta}\}$, and $(\cdot)^{\dagger}$ denotes the Moore-Penrose pseudo-inverse [Strang Book 88]. In the case of mutually uncorrelated input and noise processes and $\sigma_s^2 = 1$, $\mathbf{R}_{r,r} = \mathbf{H}\mathbf{H}^T + \sigma_w^2\mathbf{I}$ and $\mathbf{d}_{r,s}^{(\delta)} = \mathbf{H}\mathbf{e}_\delta$. This yields the following expression for the FIR Wiener equalizer:

$$\mathbf{f}_{\mathrm{m}}^{(\delta)} \;=\; (\mathbf{H}\mathbf{H}^T + \sigma_w^2\mathbf{I})^{\dagger}\mathbf{H}\mathbf{e}_\delta. \tag{1.41}$$

With the help of the singular value decomposition (SVD) [Strang Book 88], we can obtain an interpretation of the FIR Wiener receiver that is analogous to (1.33). Let $\mathbf{H}$ have the following SVD:

$$\mathbf{H} = \mathbf{U}\mathbf{\Sigma}\mathbf{V}^T, \quad \text{with} \quad \mathbf{\Sigma} = \mathrm{diag}\{\varsigma_0, \ldots, \varsigma_{L_f+L_h}\}, \tag{1.42}$$

where $\mathbf{\Sigma}$ has the same dimensions as $\mathbf{H}$ (which may not be square) and has the singular values $\{\varsigma_i\}$ on its first diagonal. We then have

$$\mathbf{f}_{\mathrm{m}}^{(\delta)} \;=\; \mathbf{U}(\mathbf{\Sigma}\mathbf{\Sigma}^T + \sigma_w^2\mathbf{I})^{\dagger}\mathbf{\Sigma}\mathbf{V}^T\mathbf{e}_\delta. \tag{1.43}$$

The above formula resembles the frequency-domain solution given in (1.31)-(1.32). In fact, (1.43) can be rewritten as

$$\mathbf{f}_{\mathrm{m}}^{(\delta)} \;=\; \mathbf{U}\,\mathrm{diag}\left\{\frac{\varsigma_0}{\varsigma_0^2 + \sigma_w^2}, \cdots, \frac{\varsigma_{L_f+L_h}}{\varsigma_{L_f+L_h}^2 + \sigma_w^2}\right\}\mathbf{V}^T\mathbf{e}_\delta. \tag{1.44}$$

Note that the terms involving the singular values $\{\varsigma_i\}$ in (1.44) have a form reminiscent of the frequency response in (1.32).

System Response. The Wiener system response is given by

$$\mathbf{q}_{\mathrm{m}}^{(\delta)} \;=\; \mathbf{H}^T\mathbf{f}_{\mathrm{m}}^{(\delta)} \;=\; \mathbf{H}^T(\mathbf{H}\mathbf{H}^T + \sigma_w^2\mathbf{I})^{\dagger}\mathbf{H}\mathbf{e}_\delta. \tag{1.45}$$

To obtain a form similar to that of the IIR case, we use the SVD expressions (1.42) and (1.43) to obtain

$$\mathbf{q}_{\mathrm{m}}^{(\delta)} \;=\; \mathbf{V}\mathbf{\Sigma}^T(\mathbf{\Sigma}\mathbf{\Sigma}^T + \sigma_w^2\mathbf{I})^{\dagger}\mathbf{\Sigma}\mathbf{V}^T\mathbf{e}_\delta, \tag{1.46}$$

where the singular values of the channel matrix play the role of magnitude spectrum of the channel in the IIR case. When $\mathbf{H}$ has column dimension less than or equal to its row dimension[7] column rank $\mathbf{H}$ we have

$$\mathbf{q}_{\mathrm{m}}^{(\delta)} \;=\; \mathbf{V}\,\mathrm{diag}\left\{\frac{\varsigma_0^2}{\varsigma_0^2 + \sigma_w^2}, \cdots, \frac{\varsigma_{L_f+L_h}^2}{\varsigma_{L_f+L_h}^2 + \sigma_w^2}\right\}\mathbf{V}^T\mathbf{e}_\delta. \tag{1.47}$$

[7] If $\mathbf{H}$ has more columns than rows, that is, the equalizer is "undermodelled" with respect to the channel, then $\mathbf{q}_{\mathrm{m}}^{(\delta)} = \mathbf{V}\,\mathrm{diag}\left\{\frac{\varsigma_0^2}{\varsigma_0^2+\sigma_w^2}, \cdots, \frac{\varsigma_{L_f+L_h}^2}{\varsigma_{L_f+L_h}^2+\sigma_w^2}, 0, \ldots, 0\right\}\mathbf{V}^T\mathbf{e}_\delta$.

Note the similarity between (1.33) and (1.47). Again, as $\sigma_w^2 \to 0$, $\mathbf{q}_{\rm m}^{(\delta)} \to \mathbf{e}_\delta$, and perfect symbol recovery is achieved. In this case, the Wiener and ZF equalizers are identical ($\mathbf{f}_{\rm m}^{(\delta)} = \mathbf{f}_{\rm z}^{(\delta)}$). We note in advance that, for full column rank $\mathbf{H}$ (i.e., $\varsigma_i > 0$) and $\sigma_w^2 = 0$, CM receivers also achieve perfect symbol recovery up to a fixed phase ambiguity. On the other hand, by increasing $\sigma_w^2/\sigma_s^2 \to \infty$, $\mathbf{q}_{\rm m}^{(\delta)}$ approaches the origin. Interestingly, this property is *not* shared by the CM receivers, as we shall describe in Section 1.3.

MMSE of FIR Wiener Receiver. For a given system delay δ and equalizer length $P(L_f+1)$, the MMSE $\mathcal{E}_{\rm m}^{(\delta)}$ is defined as $J_{\rm m}^{(\delta)}(\mathbf{f}_{\rm m}^{(\delta)})$. A simplified expression may be obtained by substituting the Wiener expression (1.41) into

$$\mathcal{E}_{\rm m}^{(\delta)} = \|\mathbf{H}^T\mathbf{f}_{\rm m}^{(\delta)} - \mathbf{e}_\delta\|_2^2 + \sigma_w^2\|\mathbf{f}_{\rm m}^{(\delta)}\|_2^2, \tag{1.48}$$

or can be obtained by first applying the orthogonality principle (1.39) to the MSE definition (1.26) and then substituting the expression (1.41) as follows:

$$\begin{aligned}\mathcal{E}_{\rm m}^{(\delta)} &= \mathrm{E}\{|\mathbf{r}^T(n)\mathbf{f}_{\rm m}^{(\delta)} - s_{n-\delta}|^2\} \\ &= \mathrm{E}\{-s_{n-\delta}(\mathbf{r}^T(n)\mathbf{f}_{\rm m}^{(\delta)} - s_{n-\delta})\} \\ &= 1 - \mathbf{e}_\delta^T\mathbf{H}^T(\mathbf{H}\mathbf{H}^T + \sigma_w^2\mathbf{I})^\dagger\mathbf{H}\mathbf{e}_\delta. \end{aligned} \tag{1.49}$$

Effects of various parameters on the MMSE can be analyzed using the SVD. Substituting (1.42) into (1.49), we get

$$\mathcal{E}_{\rm m}^{(\delta)} = 1 - \sum_{i=0}^{L_f+L_h} \frac{\varsigma_i^2}{\varsigma_i^2+\sigma_w^2}|v_{\delta,i}|^2 = \sum_{i=0}^{L_f+L_h} \frac{\sigma_w^2}{\varsigma_i^2+\sigma_w^2}|v_{\delta,i}|^2, \tag{1.50}$$

where $v_{\delta,i}$ is the $(\delta,i)^{th}$ entry of $\mathbf{V}$, that is, the δ^{th} entry of the i^{th} right singular vector of $\mathbf{H}$.

From (1.50) we see that $\mathcal{E}_{\rm m}^{(\delta)}$ depends on four factors: (i) the noise power, (ii) the subequalizer order L_f, (iii) the singular values and singular vectors of the channel matrix $\mathbf{H}$, and (iv) the system delay δ.

- *Effects of Noise:* As $\sigma_w^2 \to 0$, $\mathcal{E}_{\rm m}^{(\delta)}$ decreases, but not necessarily to zero. The only case in which $\mathcal{E}_{\rm m}^{(\delta)}$ approaches zero (for all δ) is when $\varsigma_i > 0$, that is, $\mathbf{H}$ has full column rank. For non-zero σ_w^2, (1.48) indicates that the MMSE equalizer achieves a compromise between noise gain (i.e., $\|\mathbf{f}_{\rm m}^{(\delta)}\|_2$) and ISI (i.e., $\|\mathbf{q}_{\rm m}^{(\delta)} - \mathbf{e}_\delta\|_2$).

- *Subequalizer Order L_f:* $\mathcal{E}_{\rm m}^{(\delta)}$ is a non-increasing function of L_f. Using the Toeplitz distribution theorem [Gray TIT 72], it can be shown that the FIR MMSE (1.50) approaches the IIR MMSE (1.38) as $L_f \to \infty$, which is intuitively satisfying. In practice, the selection of L_f leads to a tradeoff between desired performance and implementation complexity.

- *Effects of Channel:* For FIR equalizers, (1.50) suggests a relatively complex relationship between MMSE and the singular values and right singular vectors of the channel matrix. In the case of IIR equalizers, there exists a much simpler relationship between channel properties and MMSE performance. Specifically, (1.38) indicates that common subchannel roots near the unit circle cause an increase in MMSE. Though the relationship between MMSE and subchannel roots is less obvious in the case of an FIR equalizer, it has been shown that increasing the proximity of subchannel roots decreases the product of the singular values (see [Fijalkow SPWSSAP 96] and [Casas DSP 97]). Furthermore, the effects of near-common subchannel roots are more severe when noise power is large. Unfortunately, however, a direct link between subchannel root locations and FIR MMSE has yet to be found.

- *System Delay δ:* For FIR Wiener receivers, selection of δ may affect MMSE significantly. This can be seen in (1.49), where the δ^{th} diagonal element of the matrix quantity $\mathbf{H}^T(\mathbf{H}\mathbf{H}^T + \sigma_w^2\mathbf{I})^\dagger\mathbf{H}$ is extracted by the $\mathbf{e}_\delta$ pair. Figure 1.22 in Section 1.4 shows an example of $\mathcal{E}_{\rm m}^{(\delta)}$ for various equalizer lengths. Typically, a low-MSE "trough" exists for system delays in the vicinity of the channel's center of gravity, and system delays outside of this trough exhibit markedly higher MMSE. This can be contrasted to the performance of the IIR Wiener receiver which is invariant to delay choice. We note in advance that system delay is not a direct design parameter with CMA-adapted FIR equalizers (as discussed in Section 1.4).

The LMS Algorithm

The LMS algorithm [Widrow Book 85] is one of the most widely used stochastic gradient descent (SGD) algorithms for adaptively minimizing MSE. In terms of the instantaneous squared error

$$\hat{J}_{\rm m}^{(\delta)}(n) \triangleq \frac{1}{2}|y_n - s_{n-\delta}|^2, \tag{1.51}$$

the real-valued LMS parameter-vector update equation is

$$\begin{aligned}\mathbf{f}(n+1) &= \mathbf{f}(n) - \mu\nabla_{\mathbf{f}}\hat{J}_{\rm m}^{(\delta)}(n) && (1.52)\\ &= \mathbf{f}(n) - \mu\mathbf{r}(n)(y_n - s_{n-\delta}). && (1.53)\end{aligned}$$

where $\nabla_{\mathbf{f}}$ denotes the gradient with respect to the equalizer coefficient vector and μ is a (small) positive step-size. In practice, a training sequence sent by the transmitter and known a priori by receiver is used to supply the $s_{n-\delta}$ term in (1.53).

Standard analysis of LMS considers the transient and steady-state properties of the algorithm separately. Detailed expositions on LMS can be found in, e.g., [Haykin Book 96], [Widrow Book 85], [Gitlin Book 92], and [Macchi Book 95]. We shall review the basic behavior of the LMS algorithm (e.g., excess MSE and convergence rate) in the equalization context for later comparison with the CM-minimizing algorithm CMA. Many similarities can be found between LMS and CMA because both attempt to minimize their respective costs ($J_{\rm m}^{(\delta)}$ and $J_{\rm cm}$) using a stochastic gradient descent technique.

Transient behavior. The principal item of interest in the transient behavior of LMS is convergence rate. Because the MSE cost $J_{\rm m}^{(\delta)}$ is quadratic, the Hessian is constant throughout the parameter space and thus convergence rate analysis is straightforward. (The Hessian matrix, defined as $\nabla_{\mathbf{f}} \nabla_{\mathbf{f}^T} J_{\rm m}^{(\delta)}$, determines the curvature of the cost surface.) Below we derive bounds on the convergence rate of an FIR equalizer.

The instantaneous error can be partitioned into three components: the noise and ISI contributions to Wiener equalization and the error resulting from any deviations from Wiener equalization[8].

$$\begin{aligned} e_n &= y_n - s_{n-\delta} && (1.54)\\ &= \mathbf{r}^T(n)\mathbf{f}(n) - \mathbf{x}^T(n)\mathbf{f}_{\rm z}^{(\delta)} && (1.55)\\ &= \underbrace{\mathbf{r}^T(n)\big(\mathbf{f}(n)-\mathbf{f}_{\rm m}^{(\delta)}\big)}_{\text{from parameter errors}} + \underbrace{\mathbf{x}^T(n)\big(\mathbf{f}_{\rm m}^{(\delta)}-\mathbf{f}_{\rm z}^{(\delta)}\big)}_{\text{from ISI \& bias}} + \underbrace{\mathbf{w}^T(n)\mathbf{f}_{\rm m}^{(\delta)}}_{\text{from noise}}. && (1.56)\end{aligned}$$

In deriving (1.56), we used the definition $\mathbf{x}(n) = \mathbf{H}\mathbf{s}(n)$ and the ZF equalizer property $\mathbf{H}^T\mathbf{f}_{\rm z}^{(\delta)} = \mathbf{e}_\delta$ that appeared earlier in this section.

We define the equalizer *parameter error* vector as $\tilde{\mathbf{f}}(n) \triangleq \mathbf{f}(n) - \mathbf{f}_{\rm m}^{(\delta)}$. Substituting (1.56) into the LMS update equation (1.53) and subtracting $\mathbf{f}_{\rm m}^{(\delta)}$ from both sides, the parameter error can be seen to evolve as

$$\tilde{\mathbf{f}}(n+1) = \big(\mathbf{I} - \mu\mathbf{r}(n)\mathbf{r}^T(n)\big)\tilde{\mathbf{f}}(n) - \mu\mathbf{r}(n)\mathbf{x}^T(n)\big(\mathbf{f}_{\rm m}^{(\delta)} - \mathbf{f}_{\rm z}^{(\delta)}\big) - \mu\mathbf{r}(n)\mathbf{w}^T(n)\mathbf{f}_{\rm m}^{(\delta)}. \qquad (1.57)$$

With a reasonably small step-size, the parameters change very slowly with respect to the signal and noise vectors in (1.56). To exploit this dual time-scale nature of the LMS update, we define the average parameter error vector $\mathbf{g}(n) \triangleq \mathrm{E}\{\tilde{\mathbf{f}}(n)\}$. Then, assuming that the signal and noise are mutually uncorrelated and denoting the autocorrelation matrix of the (real-valued) noiseless received signal by $\mathbf{R}_{x,x} \triangleq \mathrm{E}\{\mathbf{x}(n)\mathbf{x}^T(n)\}$, (1.57) implies that the average parameter error evolves as

$$\mathbf{g}(n+1) = \big(\mathbf{I} - \mu\mathbf{R}_{r,r}\big)\mathbf{g}(n) - \mu\mathbf{R}_{x,x}\big(\mathbf{f}_{\rm m}^{(\delta)} - \mathbf{f}_{\rm z}^{(\delta)}\big) - \mu\sigma_w^2\mathbf{f}_{\rm m}^{(\delta)}. \qquad (1.58)$$

We have utilized the assumption that $\mathbf{f}(n)$ is changing slowly with respect to the data $\mathbf{r}(n)$, so that the average of their products can be well approximated by the product of their averages.

Our expression for average parameter error can be simplified with the observation that

$$\mathbf{R}_{x,x}\big(\mathbf{f}_{\rm m}^{(\delta)} - \mathbf{f}_{\rm z}^{(\delta)}\big) + \sigma_w^2\mathbf{f}_{\rm m}^{(\delta)} = \mathbf{R}_{r,r}\mathbf{f}_{\rm m}^{(\delta)} - \mathbf{R}_{x,x}\mathbf{f}_{\rm z}^{(\delta)} = \mathbf{d}_{r,s}^{(\delta)} - \mathbf{H}\mathbf{H}^T\mathbf{f}_{\rm z}^{(\delta)} = \mathbf{d}_{r,s}^{(\delta)} - \mathbf{H}\mathbf{e}_\delta = \mathbf{0},$$

giving the LMS average parameter update equation

$$\mathbf{g}(n+1) = \big(\mathbf{I} - \mu\mathbf{R}_{r,r}\big)\mathbf{g}(n). \qquad (1.59)$$

[8] In (1.56), $\mathbf{f}(n)$ is allowed to be of arbitrary length. While a Wiener equalizer always exists to match the length of $\mathbf{f}(n)$, the ZF equalizer $\mathbf{f}_{\rm z}^{(\delta)}$ must, in general, satisfy the length condition $L_f \geq L_h - 1$. Thus, in (1.56), $\mathbf{f}_{\rm z}^{(\delta)}$ must be zero-padded when $L_f > L_h - 1$, while the $\mathbf{f}_{\rm m}^{(\delta)}$ in the "from ISI & bias" term must be zero padded when $L_f < L_h - 1$.

Notice that (1.59) specifies a linear homogeneous difference equation with state transition matrix $(\mathbf{I} - \mu\mathbf{R}_{r,r})$. For stability of the average-error-system (1.59), the eigenvalues of $(\mathbf{I} - \mu\mathbf{R}_{r,r})$ must lie between -1 and 1. Letting $\lambda_{\max}$ be the maximum eigenvalue of $\mathbf{R}_{r,r}$, this implies that the stability of the average system is guaranteed when

$$0 < \mu < \frac{2}{\lambda_{\max}}. \qquad (1.60)$$

A classical implication of (1.59) is that, for μ obeying (1.60), the average LMS parameter error decays to zero as the equalizer adapts. In other words, LMS converges in mean to the Wiener solution $\mathbf{f}_{\rm m}^{(\delta)}$.

Mean-*square* convergence requires, in addition to (1.60), that [Haykin Book 96]

$$\sum_i \frac{\mu\lambda_i}{2 - \mu\lambda_i} < 1,$$

which, for small step-sizes, can be approximated by

$$\mu < \frac{2}{\sum_i \lambda_i} = \frac{2}{P(L_f+1)\sigma_r^2}. \qquad (1.61)$$

Equation (1.61) indicates a step-size requirement proportional to equalizer length and received signal power.

Asymptotic Performance. When any of the perfect symbol recovery conditions is violated, no fixed equalizer can zero the error $e_n = y_n - s_{n-\delta}$ used by the LMS algorithm. Thus, with a non-vanishing step-size μ, an LMS-derived parameter vector $\mathbf{f}(n)$ will never settle to the MMSE solution but instead "jitter" around it. Therefore, the actual steady-state MSE achieved by LMS is greater than the MMSE. We call the difference between MSE and MMSE the *excess mean square error* (EMSE).

EMSE can be approximated by a function of step-size and equalizer length. Using (1.56) and the assumptions that $\mathbf{x}(n)$, $\mathbf{w}(n)$, and $\tilde{\mathbf{f}}(n)$ are uncorrelated, we can obtain an expression for the steady-state MSE in terms of EMSE and MMSE [Haykin Book 96]:

$$\underbrace{\mathrm{E}\{e_n^2\}}_{\text{MSE}} = \underbrace{\mathrm{tr}\big(\mathbf{R}_{r,r}\,\mathrm{E}\{\tilde{\mathbf{f}}(n)\tilde{\mathbf{f}}^T(n)\}\big)}_{\text{EMSE, } \mathcal{E}_\chi} + \underbrace{\|\mathbf{f}_{\rm m}^{(\delta)} - \mathbf{f}_{\rm z}^{(\delta)}\|_{\mathbf{R}_{x,x}}^2 + \sigma_w^2\|\mathbf{f}_{\rm m}^{(\delta)}\|_2^2}_{\text{MMSE, } \mathcal{E}_m}. \qquad (1.62)$$

The EMSE term $\mathcal{E}_\chi$ is, in general, difficult to analyze. However, for small μ, a close approximation to $\mathcal{E}_\chi$ can be found using various independence assumptions on $\mathbf{r}(n)$, $\mathbf{s}(n)$, and $\tilde{\mathbf{f}}(n)$. In that case,

$$\mathcal{E}_\chi \approx \frac{\mu}{2}P(L_f+1)\,\sigma_r^2\,\mathcal{E}_{\rm m}, \qquad (1.63)$$

where $\sigma_r^2 \triangleq \mathrm{E}\{|r_k|^2\}$ is the received signal power [Haykin Book 96]. Note that EMSE is proportional to the product of step-size, equalizer length, received signal power, and MMSE. Recall that $P(L_f+1)$ represents the total number of adapted equalizer coefficients.

LMS Design Implications. The form of the EMSE expression (1.63) has a number of design implications for LMS-based adaptive equalization.

For one, there is the step-size tradeoff: a large step-size gives fast convergence to an equalizer with large EMSE while a small step-size gives slow convergence to an equalizer with small EMSE. This suggests making μ as small as possible within allowed limits on convergence time. When the channel impulse response is time-varying, however, the situation becomes more complicated: step-sizes that are too small may not be able to adequately track the channel variations, resulting in what is known as *tracking lag* [Widrow PROC 76]. Tracking lag prevents the equalizer from remaining in close vicinity to the time-varying $\mathbf{f}_{\mathrm{m}}^{(\delta)}$ which ultimately increases MSE. Increasing the step-size may reduce tracking lag but may increase $\mathcal{E}_\chi$ via (1.63). Thus, even in time-varying situations, step-sizes that are too large ultimately increase steady-state MSE via (1.62). Hence, the best step-size is a compromise between these situations.

There is a similar tradeoff with equalizer length $P(L_f+1)$. Equalizers that are too short result in a high MMSE, which prohibits a low steady-state MSE via (1.62). On the other hand, very long equalizers may result in a large EMSE via (1.63), also preventing good steady-state MSE performance. Thus, the optimal equalizer length lies somewhere in between.

The analysis of Section 1.3 will demonstrate that CMA shares many of the same behavior features of LMS. Hence, many of the CMA design guidelines in Section 1.4 can be related back to the LMS design guidelines discussed in this section.

1.3 The CM Criterion and CMA

This section focuses on the properties of the CM criterion and the behavior of the Constant Modulus Algorithm (CMA). What will be a recurring theme can be summarized by *Godard's conjecture*, an observation made by Godard in his seminal paper on blind adaptive (baud-spaced) equalization. In [Godard TCOM 80], Godard observed that the MSE performance of CMA is close to that of the MMSE-optimal (Wiener) equalizer:

> *"It should also be noted that the equalizer coefficients minimizing the dispersion functions closely approximate those which minimize the mean-squared error."* —Godard (1980).

Treichler and Agee made a similar claim in their independent development of the CM criterion [Treichler TASSP 83]. Throughout this section, we shall provide evidence supporting Godard's conjecture.

This section is organized as follows. We first introduce the CM criterion and present what are known as the Perfect Blind Equalization (PBE) conditions. Simple illustrated examples are provided to understand the effects of violating these conditions. Next, properties of the local and global minima of the CM cost function (referred to as *CM receivers*) are discussed and compared to that of the well-known Wiener and zero-forcing receivers discussed in Section 1.2. This is followed by summaries of important analytical work concerning the robustness of CM receivers to the presence of noise and channel undermodelling. Both perturbation-based approximations and bounds from a geometrical approach are used to predict CM performance in a practical setting. The section concludes by summarizing the transient and asymptotic behavior of CMA, a stochastic gradient descent (SGD) algorithm popularly employed to adapt blind equalizers, which minimizes the CM criterion. Various comparisons are drawn between CMA and the well-known LMS algorithm (discussed in Section 1.2).

Properties of the CM Criterion

Functional Form. Blind equalization can be considered as the estimation of (unknown) source symbols $\{s_n\}$ from the receiver input sequence $\{\mathbf{r}_n\}$, or equivalently $\{\mathbf{r}(n)\}$. In our formulation of the problem, the linear filter $\mathbf{f}$ generates the symbol estimate $y_n = \mathbf{f}^T\mathbf{r}(n)$, and we desire that $y_n \approx s_{n-\delta}$ for some fixed integer delay δ.

In motivating the CM criterion, we first consider the family of Bussgang techniques used in blind equalization [Bellini (Haykin) 94]. Some intuition behind these Bussgang techniques comes from a consideration of trained and decision-directed LMS-based equalizer adaptation, described by the update equation

$$\mathbf{f}(n+1) \quad = \quad \mathbf{f}(n) - \mu\,\mathbf{r}^*(n)\,e_n. \tag{1.64}$$

In the case of a trained update, $e_n = y_n - s_{n-\delta}$, and in the case of a decision-directed (DD) update, $e_n = y_n - \hat{s}_{n-\delta}$, where $\hat{s}_{n-\delta}$ is the symbol estimate generated by the decision device. Since high symbol-error rates prevent the success of DD equalization in cold start-up applications, we are motivated to find a more reliable error signal e_n. Consider instead applying a memoryless nonlinearity $g(\cdot)$ to the equalizer output and forming the error

$$e_n = y_n - g(y_n). \tag{1.65}$$

An update based on (1.65) would satisfy our criterion for *blind* equalization since e_n can be constructed solely from the received signal. Such algorithms are referred to as "Bussgang" [Bussgang MIT 52] and achieve equilibrium when the equalizer output satisfies $\mathrm{E}\{\mathbf{r}(n)y_n\} = \mathrm{E}\{\mathbf{r}(n)g(y_n)\}$ (via (1.64) and (1.65)).

In the reception of complex source alphabets such as QAM, the receiver must not only remove ISI, but also synchronize its demodulator to the carrier frequency and phase employed by the transmitter. In a blind scenario (i.e., in the absence of a pilot signal), carrier synchronization *before* equalization (i.e., in a non-decision-directed mode) is especially difficult with large QAM constellations. Similarly, blind equalization before carrier synchronization is often a prohibitively difficult task: to the typical blind symbol estimation algorithm, carrier offset appears as a (potentially rapid) time-variation in the channel response. Rapid time-variations are known to hinder the convergence of a SGD algorithms such as (1.64), which rely on some degree of (cyclo)stationarity for proper operation. Thus, practical considerations motivate the decoupling of blind equalization from carrier recovery [Treichler PROC 98].

The search for a carrier-phase independent algorithm led Godard to consider blind equalization techniques based only on the signal modulus $|y_n|$. He proposed the minimization of the following "dispersion" cost [Godard TCOM 80]:

$$\begin{aligned} J_p &\triangleq \frac{1}{2p}\mathrm{E}\big\{\big(|y_n|^p - \gamma\big)^2\big\} \\ &= \frac{1}{2p}\mathrm{E}\big\{\big(|\mathbf{f}^T\mathbf{r}(n)|^p - \gamma\big)^2\big\}, \end{aligned} \tag{1.66}$$

where the real constant γ is chosen as a function of the source alphabet and of the integer p. Specifically, Godard showed that for sub-Gaussian sources (described in **(A4)** below), J_p is minimized by system responses generating zero ISI, and he showed that choosing $\gamma = \mathrm{E}\{|s_n|^{2p}\}/\,\mathrm{E}\{|s_n|^p\}$ ensures that local minima of J_p exist at the perfectly-equalizing system responses [Godard TCOM 80]. In the case of $p = 2$, we refer to (1.66) as the CM cost or the *CM criterion*, and denote it by J_{cm}:

$$J_{\mathrm{cm}} \triangleq \frac{1}{4}\mathrm{E}\big\{\big(|y_n|^2 - \tfrac{\mathrm{E}\{|s_n|^4\}}{\sigma_s^2}\big)^2\big\} = \frac{1}{4}\mathrm{E}\big\{\big(|\mathbf{f}^T\mathbf{r}(n)|^2 - \tfrac{\mathrm{E}\{|s_n|^4\}}{\sigma_s^2}\big)^2\big\}, \tag{1.67}$$

Equation (1.67) shows that the CM criterion penalizes dispersion of the squared output modulus $|y_n|^2$ away from the constant $\frac{\mathrm{E}\{|s_n|^4\}}{\sigma_s^2}$. As a phase-independent criterion for blind source recovery, minimizing

dispersion seems intuitively satisfying for CM sources[9]. Remarkably, the CM criterion works almost as well with non-CM sources. We note that the CM idea was independently proposed for constant-modulus source sequences (i.e., $\exists\,\rho \in \mathbb{R}$ s.t. $\forall n,\ |s_n| = \rho$) by Treichler and Agee in [Treichler TASSP 83].

It is important to note that Godard's blind equalization algorithm belongs to the Bussgang class. In fact, it can be shown to approximate the conditional-mean estimator of $|s_n|^2$ given $|y_n|^2$ [Bellini (Haykin) 94], a valid estimator in the absence of carrier phase information (or equivalently, when all rotations of the source constellation are assumed equally likely).

Having introduced the CM criterion as a phase-independent means of blind equalization (thus intended for complex-valued signals) it may seem strange that in the sequel we restrict our analysis to the case of real-valued quantities. We justify our position by claiming that nearly all of the intuition concerning the behavior of the CM criterion can be gained from the study of its real-valued incarnation with the benefit of a simplified presentation. We will attempt to note any situations in which complex-valued quantities lead to meaningful conceptual differences. With this in mind, we present a useful expansion of the CM cost in terms of the equalizer coefficients $\mathbf{f}$ for real-valued channels, real-valued i.i.d. noise, and real-valued i.i.d. sources. (See [Johnson PROC 98] for derivations of this and more general CM cost expressions.)

$$\begin{aligned} J_{\mathrm{cm}} = {} & \frac{1}{4}\sigma_s^4(\kappa_s - 3)\|\mathbf{H}^T\mathbf{f}\|_4^4 + \frac{3}{4}\sigma_s^4\|\mathbf{H}^T\mathbf{f}\|_2^4 + \frac{1}{4}\sigma_w^4(\kappa_w - 3)\|\mathbf{f}\|_4^4 + \frac{3}{4}\sigma_w^4\|\mathbf{f}\|_2^4 \\ & + \frac{3}{2}\sigma_s^2\sigma_w^2\|\mathbf{H}^T\mathbf{f}\|_2^2\,\|\mathbf{f}\|_2^2 - \frac{1}{2}\sigma_s^2\kappa_s(\sigma_s^2\|\mathbf{H}^T\mathbf{f}\|_2^2 + \sigma_w^2\|\mathbf{f}\|_2^2) + \frac{1}{4}\sigma_s^4\kappa_s^2. \end{aligned} \tag{1.68}$$

In the equation above, κ_s refers to the normalized kurtosis of the source process, defined below in (1.69). By analogy, κ_w denotes the kurtosis of the channel noise process $\{w_k\}$. The quantities σ_s^2 and σ_w^2 denote the source and noise variances, respectively. Note that, in the absence of noise, J_{cm} is a quartic function of the ℓ_4 and ℓ_2 norms of the channel-equalizer impulse response $\mathbf{H}^T\mathbf{f}$. The addition of noise brings a quartic dependence on the ℓ_4 and ℓ_2 norms of the equalizer impulse response $\mathbf{f}$.

As we shall see in the following sections, the local and global minimizers of J_{cm}, that is, the CM receivers $\mathbf{f}_c$, are of key importance. For example, they represent the asymptotic mean-convergence points of the constant modulus algorithm.

Perfect Blind Equalizability Conditions. The set of conditions under which all minimizers of J_{cm} accomplish perfect symbol recovery are known as the *Perfect Blind Equalization* (PBE) conditions [Foschini ATT 85, Fijalkow SPW 94, Li TSP 96a]. They are given below.

(A1) Full column-rank channel matrix $\mathbf{H}$,

(A2) No additive channel noise,

[9] Sources derived from M-PSK constellations are examples of CM sources, since every symbol in the M-PSK alphabet has the same magnitude (or "modulus"). M-QAM sources for $M > 4$ do not have the CM property since alphabet members differ in magnitude as well as phase. See Fig. 1.19 for an illustration.

Table 1.7: Summary of Channels Used for Two-Tap FSE Examples.

Name	$T/2$-spaced Impulse Response	Classification
$\mathbf{h}_{\mathrm{a}}$	$(-0.0901, 0.6853, 0.7170, -0.0901)^T$	well-behaved[11]
$\mathbf{h}_{\mathrm{b}}$	$(1.0, -0.5, 0.2, 0.3)^T$	well-behaved
$\mathbf{h}_{\mathrm{c}}$	$(-0.0086, 0.0101, 0.9999, -0.0086)^T$	nearly-common subchannel roots
$\mathbf{h}_{\mathrm{d}}$	$(1.0, -0.5, 0.2, 0.3, -0.2, -0.15)^T$	undermodelled

(A3) Symmetric i.i.d.[10] source, circularly symmetric ($\mathrm{E}\{s_n^2\} = 0$) in the complex-valued case.

(A4) Sub-Gaussian source: the normalized source kurtosis satisfies $\kappa_s < \kappa_g$.

The normalized source kurtosis κ_s is defined as

$$\kappa_s \triangleq \frac{\mathrm{E}\{|s_n|^4\}}{\sigma_s^4}, \tag{1.69}$$

where the kurtosis of a real-valued Gaussian random process is $\kappa_g = 3$ and the kurtosis of a proper complex-valued Gaussian random process is $\kappa_g = 2$. Note that Conditions **(A1)**-**(A2)** pertain to the channel-equalizer pair's ability to achieve perfect equalization, as discussed in Section 1.2. Conditions **(A3)** and **(A4)** pertain solely to blind equalization based on the CM criterion.

Illustrative Examples of CM Cost Surface Deformations. This section studies changes to the "shape" of the CM cost surface as a means of understanding the effect of various violations of the PBE conditions. Restricting our focus to the case of a real-valued two-tap $T/2$-spaced equalizer permits illustration of the cost surface as a function of equalizer parameters. We do not intend a rigorous analysis of CM robustness properties here—that will be the subject of later subsections.

Perfect Blind Equalizability. We first consider a 4-tap channel and 2-tap equalizer, both $T/2$-spaced (i.e., $P = 2$), satisfying the column rank condition **(A1)**. Associated with this model are the following channel convolution matrix and equalizer coefficient vector:

$$\mathbf{H} = \begin{pmatrix} h_1 & h_3 \\ h_0 & h_2 \end{pmatrix}, \quad \mathbf{f} = \begin{pmatrix} f_0 \\ f_1 \end{pmatrix}.$$

A square invertible $\mathbf{H}$ ensures unique zero-forcing solutions to $\pm\mathbf{e}_\delta = \mathbf{H}^T\mathbf{f}$ for $\delta \in \{0, 1\}$. As discussed in Section 1.2, the existence of this inverse requires that the even and odd subchannels, $h^{(1)}(z) = h_0 + h_2 z^{-1}$ and $h^{(2)}(z) = h_1 + h_3 z^{-1}$, respectively, do not share the same root location. Examples of 4-tap channels satisfying this non-common-root condition are given by $\mathbf{h}_{\mathrm{a}}$ and $\mathbf{h}_{\mathrm{b}}$ in Table 1.7.

[10]Examination of the derivation in [Johnson PROC 98] reveals that the CM cost expression yielding the global convergence properties requires only fourth-order statistical independence of the source process.

[11]"Well-behaved" indicates the absence of common or nearly-common subchannel roots.

For an i.i.d. source with kurtosis $\kappa_s = 1$ in the absence of channel noise (thus satisfying **(A2)**-**(A4)**), the CM cost is shown in Fig. 1.5 and 1.6 as a function of equalizer coefficients. In all upcoming contour plots, a "$*$" indicates a MMSE solution corresponding to an optimal system delay, while a "$\times$" indicates a MMSE solution corresponding to a sub-optimal system delay. In Fig. 1.6, however, zero MSE can be achieved at all system delays and thus all solutions are optimal. It can be seen that when the PBE conditions are satisfied, the MMSE solutions exactly coincide with the minima of J_{cm}. Each pair of CM minima symmetric with respect to the origin corresponds to the same system delay but to two different[12] choices of system polarity ($\pm$). Such sign ambiguity is inconsequential when, for example, the symbols have been differentially encoded. Thus, in Fig. 1.5, the four minima correspond to the four permutations of system delay ($\delta = 0, 1$) and polarity ($+, -$). Note the CM local maximum at the origin.

Sometimes it is more convenient to study J_{cm} as a function of system response $\mathbf{q}$. Note that when $\mathbf{H}$ is square and invertible, as in our current example, there exists a unique mapping between $\mathbf{f}$- and $\mathbf{q}$-spaces: $\mathbf{q} = \mathbf{H}^T\mathbf{f}$. One appeal of studying J_{cm} in $\mathbf{q}$-space follows from the normalization and alignment of the J_{cm} minima with the coordinate axes. For example, our current ZF system responses occur at $\mathbf{q} = (\pm 1, 0)$ and $(0, \pm 1)$ (see Fig. 1.7). Such symmetry will be exploited in the section. Under satisfaction of **(A1)** and **(A2)**, viewing J_{cm} from system space has the additional advantage that the CM cost is channel-independent (recall (1.68) with $\mathbf{H}^T\mathbf{f} = \mathbf{q}$).

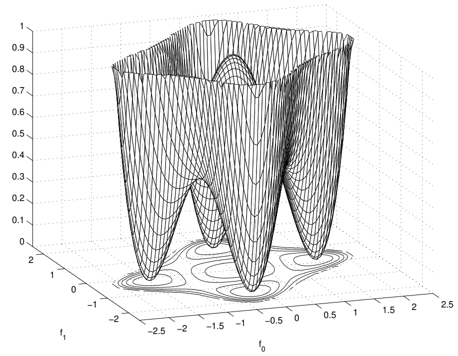

Figure 1.5: J_{cm} for BPSK, well-behaved channel $\mathbf{h}_{\mathrm{a}}$, and no noise, in equalizer ($\mathbf{f}$) space.

Effect of Channel Noise. Violation of **(A2)** occurs in any communication system where additive channel noise $\mathbf{w}(n)$ is present. We denote the *signal-to-noise ratio* (SNR) by $\mathrm{SNR} = 10\log_{10}(\sigma_s^2/\sigma_w^2)$ using the

[12]In the complex-valued case, each pair of minima would be replaced by a a continuum of minima spanning the full range $(0 - 2\pi)$ of allowable system phase.

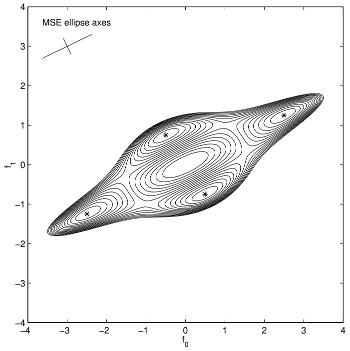

Figure 1.6: $J_{\rm cm}$ contours for BPSK, well-behaved channel $\mathbf{h}_{\rm b}$, and no noise, in equalizer ($\mathbf{f}$) space, with global MMSE minima marked by "$*$".

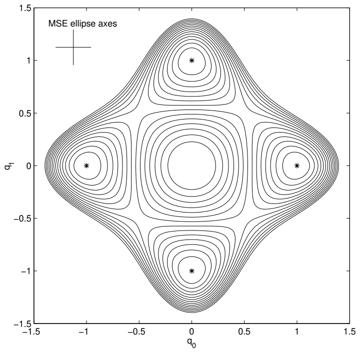

Figure 1.7: $J_{\rm cm}$ contours for a BPSK source in the absence of noise, in system ($\mathbf{q}$) space.

source and noise variances introduced in Section 1.2. The presence of noise causes the shape of the CM cost surface to change in such a way that the CM minima no longer correspond to the Wiener minima and such that different CM minima may result in different levels of performance. In other words, we now have strictly local (versus global) minima. This can be recognized in the contour depths in Fig. 1.8.

Most importantly, however, Fig. 1.8 suggests that, even at relatively high noise powers, the CM and MSE minima roughly correspond. Here we see evidence of Godard's conjecture. It is interesting to note that the suboptimal CM minima are further from their neighboring Wiener solutions than the optimal CM minima are from theirs. The robustness properties of the CM criterion to additive channel noise are the focus of sections presented in the sequel.

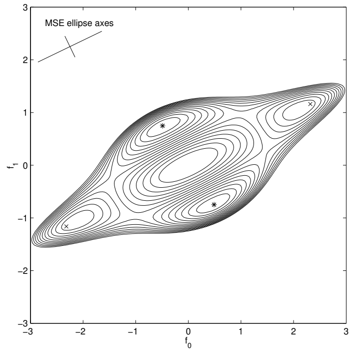

Figure 1.8: $J_{\rm cm}$ contours for channel $\mathbf{h}_{\rm b}$ and 20 dB SNR in equalizer ($\mathbf{f}$) space.

Effect of Nearly-Common Subchannel Roots. In Section 1.2 we discussed the effects of (exactly) common subchannel roots and demonstrated that, in general, they prevent perfect source recovery (hence violating **(A1)**). Though nearly-common subchannel roots do not share this problem, they may excite an undesirable phenomenon known as *noise gain*, a familiar concept from ZF and MMSE equalization theory. We will see that a similar phenomenon exists for equalization methods based on minimizing the CM criterion.

One way to understand the mechanics of noise gain is to consider the transformation from the ZF system responses $\{\mathbf{q}_{\rm z}^{(\delta)}\}$ to the respective ZF equalizers $\{\mathbf{f}_{\rm z}^{(\delta)}\}$. Nearly-common subchannel roots (i.e., $h_3/h_1 \approx h_2/h_0$) result in a transformation matrix $(\mathbf{H}^T)^\dagger$ with large eigenvalues. This has the effect of mapping certain ZF system responses to ZF equalizers with large norm, as illustrated[13] by Fig. 1.9. These

[13]Recall that with square invertible $\mathbf{H}$ and in the absence of noise the ZF and MMSE equalizers are identical. Thus the $*$'s in Fig. 1.9 denote the locations of the ZF equalizers.

large-norm equalizers result in significant noise gain, as indicated by the MSE formula (1.48). In order to better compromise between ISI cancellation and noise gain, we expect the MMSE optimal equalizers to be smaller in norm than their ZF counterparts. This is confirmed by the behavior of the Wiener expressions (1.32) and (1.44) as σ_w^2 is increased. Equation (1.44) also indicates that if $\mathbf{H}$ has small singular values (as would result from nearly-common subchannel roots), $\mathbf{f}_{\mathrm{m}}^{(\delta)}$ will be sensitive to even modest amounts of noise.

The effects of nearly-common subchannel roots on CM receivers are quite similar. Since nearly-common roots do not actually violate **(A1)**, the PBE conditions may still be satisfied, as shown by Fig. 1.9. They do excite a similar noise gain phenomenon, however, affecting CM receivers quite similarly to their Wiener counterparts (see Fig. 1.10). Note again that the better J_{cm} minima are in close proximity to the better Wiener solutions, offering evidence for Godard's conjecture. CM robustness to common subchannels has been formally addressed in [Fijalkow TSP 97].

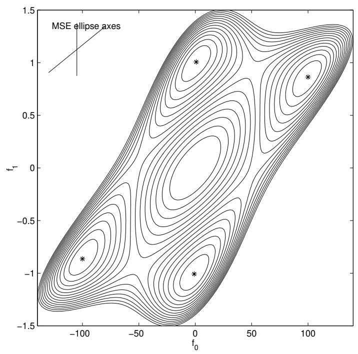

Figure 1.9: J_{cm} contours for nearly-common subchannel-roots channel $\mathbf{h}_{\mathrm{c}}$ and no noise in equalizer ($\mathbf{f}$) space. Note axis scaling.

<u>*Effect of Channel Undermodelling*</u>. When the equalizer length is short enough to violate **(A1)**, the equalizer is, in general, not capable of perfect symbol recovery. This can be demonstrated by leaving our equalizer length at two taps but extending the $\mathbf{h}_{\mathrm{b}}$ channel response to $(1.0, -0.5, 0.2, 0.3, -0.2, -0.15)^T = \mathbf{h}_{\mathrm{d}}$. Now we have

$$\mathbf{H} = \begin{pmatrix} h_1 & h_3 & h_5 \\ h_0 & h_2 & h_4 \end{pmatrix}, \quad \mathbf{f} = \begin{pmatrix} f_0 \\ f_1 \end{pmatrix}.$$

The dimensions of $\mathbf{H}$ indicate that three distinct system delays are now possible, and thus there exist six Wiener receivers (after allowing $\pm$ system polarities). Looking at Fig. 1.11, however, the number of CM receivers has not increased. It is important to note that the best CM minima are still near the best Wiener

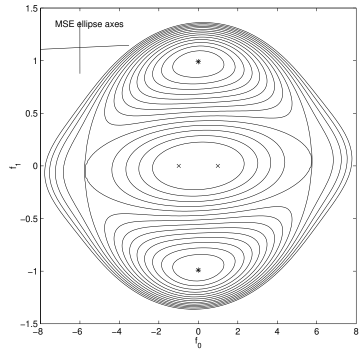

Figure 1.10: J_{cm} contours for nearly-common subchannel-roots channel $\mathbf{h}_{\mathrm{c}}$ and 20 dB SNR in equalizer ($\mathbf{f}$) space. Note axis scaling.

solutions, implying a certain degree of robustness to undermodelling and further evidence for Godard's conjecture. These robustness properties are discussed further in an upcoming subsection.

<u>*Effect of Source Correlation*</u>. PBE condition **(A3)** specifies a white source sequence. To examine this condition, Fig. 1.12 shows the effect of a temporally correlated source on the CM cost surface using noiseless channel $\mathbf{h}_{\mathrm{b}}$ and the 4-PAM periodic sequence $\{s_n\} = \{\dots, 3, 1, -1, -3, \dots\}$. In comparison to Fig. 1.6, we can observe a "twisting" of the cost surface that leads to a significant separation between the CM and Wiener receivers. More dramatic effects are possible in higher-dimensional systems including a potential increase in the number of CM local minima [LeBlanc IJACSP 98].

In the complex-valued case, correlation between the real and imaginary components of the source yields $\mathrm{E}\{s_n^2\} \neq 0$, thus violating **(A3)**. [Axford TSP 98] and [Papadias ICASSP 97] discuss the presence of false CM minima under these conditions.

<u>*Effect of Source Kurtosis*</u>. Condition **(A4)** states that the source distribution must be sub-Gaussian (i.e., $\kappa_s < \kappa_g$) for perfect blind equalizability. This requirement is satisfied by all uniformly-distributed data constellations, such as M-PAM, M-PSK, and M-QAM. Even so, the value of the (sub-Gaussian) kurtosis still has an effect on the shape of the CM cost surface. Specifically, as the kurtosis (1.69) increases towards κ_g, the CM cost surface rises and all but the radial component of the gradient tends to vanish. Figure 1.13 shows the surface for a uniformly distributed 32-PAM source, for which $\kappa_s \approx 1.8$. (Compare to Fig. 1.5.) Note that the parameter-space positions of the CM minima are unaltered (as expected, since the PBE conditions are still satisfied). For more information on the effect of source kurtosis on the CM criterion, we refer interested

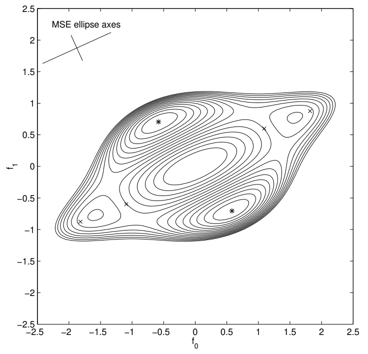

Figure 1.11: $J_{\rm cm}$ contours for undermodelled channel $\mathbf{h}_{\rm d}$ and no noise in equalizer ($\mathbf{f}$) space.

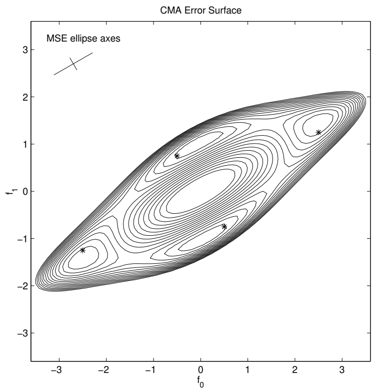

Figure 1.12: $J_{\rm cm}$ contours for channel $\mathbf{h}_{\rm b}$ and periodic source in equalizer ($\mathbf{f}$) space.

readers to [LeBlanc IJACSP 98, LeBlanc Thesis 95].

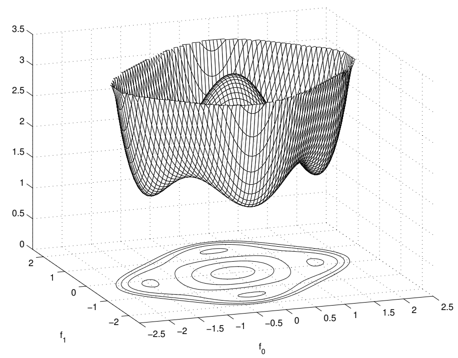

Figure 1.13: $J_{\rm cm}$ for non-CM source ($\kappa_s \approx 1.8$) and channel $\mathbf{h}_{\rm a}$ in equalizer ($\mathbf{f}$) space.

Properties of CM Stationary Points. Recall that, when the PBE conditions are satisfied, all minima of the CM cost function are global minima and coincide with the ZF and Wiener receivers. In such a case, the CM receivers achieve perfect source recovery.

In addition to minima, the CM cost surface exhibits saddle points and one maximum. These extrema, commonly called stationary points, are described in [Johnson IJACSP 95] for BPSK and a real-valued channel. With an i.i.d. source and in the absence of channel noise, the CM stationary points can be identified by having system responses $\mathbf{q}$ satisfying particular criteria. Specifically, a stationary point $\mathbf{q}_c$ satisfies the zero-gradient condition $\nabla_{\mathbf{f}} J_{\rm cm} = 0$, where the equalizer-space gradient $\nabla_{\mathbf{f}} J_{\rm cm}$ is given below as a function of $\mathbf{q}$ (under **(A2)** and **(A3)**):

$$\nabla_{\mathbf{f}} J_{\rm cm}(\mathbf{q}) \quad = \quad \sigma_s^4 \mathbf{H} \boldsymbol{\Delta}_{\mathbf{q}} \mathbf{q}. \tag{1.70}$$

The matrix $\boldsymbol{\Delta}_{\mathbf{q}}$ is defined (in both the real- and complex-valued cases) as follows:

$$\boldsymbol{\Delta}_{\mathbf{q}} \quad \triangleq \quad (\kappa_g \, \|\mathbf{q}\|_2^2 - \kappa_s)\mathbf{I} - (\kappa_g - \kappa_s)\mathrm{Diag}(\mathbf{q}\mathbf{q}^H), \tag{1.71}$$

where κ_g is the kurtosis of a Gaussian source. The matrix operation $\mathrm{Diag}(\cdot)$ retains the diagonal entries of its argument while replacing the other entries with zeros.

The stability of a CM stationary point (i.e., whether or not it is a local minima) is given by the positive-semidefiniteness of its equalizer-space Hessian, denoted by $\mathcal{H}_{\mathbf{f}} J_{\rm cm} \geq 0$. The Hessian is given below, for the case of real-valued quantities[14], as a function of $\mathbf{q}$ (again, under **(A2)** and **(A3)**):

$$\mathcal{H}_{\mathbf{f}} J_{\mathrm{cm}} = \sigma_s^4 \mathbf{H} \Psi_{\mathbf{q}} \mathbf{H}^T. \tag{1.72}$$

The matrix $\Psi_{\mathbf{q}}$ is defined as follows:

$$\Psi_{\mathbf{q}} \triangleq (\kappa_g \|\mathbf{q}\|_2^2 - \kappa_s)\mathbf{I} + 2\kappa_g \mathbf{q}\mathbf{q}^T - 3(\kappa_g - \kappa_s)\mathrm{diag}(\mathbf{q}\mathbf{q}^T). \tag{1.73}$$

In the case that $\mathbf{H}$ is full-column rank, the requirements for $\mathbf{q}$ to be a CM minimum reduce to $\mathbf{\Delta_q q} = 0$ and $\Psi_{\mathbf{q}} \geq 0$.

In summary, it can be shown (see, e.g., [Fijalkow SPW 94] for the real-valued case) that, under the PBE conditions,

- all CM minima correspond to ZF system responses, that is, $\mathbf{q}_c = \pm \mathbf{e}_\delta$ for $0 \geq \delta \geq L_f + L_h$;
- the only CM maximum occurs at the origin, that is, $\mathbf{q} = 0$; and
- the CM saddle points are described by system responses composed of coefficients that are either zero or a fixed magnitude: $\mathbf{q} = \tau_M \sum_{i=0}^{M-1} (\pm \mathbf{e}_i)$ where M, the number of nonzero elements in $\mathbf{q}$, obeys $2 \leq M \leq L_f + L_h + 1$ and where $\tau_M = \sqrt{\frac{\kappa_s}{\kappa_g(M-1)+\kappa_s}}$. (See Fig. 1.18.)

Notice that the number and type of stationary points of the CM cost are not affected by the source kurtosis κ_s (as long as the source is sub-Gaussian). However, the values of the CM costs at the stationary points, as well as the locations of the saddle points, remain a function of κ_s. (This can be verified by substituting the appropriate $\mathbf{q}$ into (1.68).) The relocation of saddle points is important because it affects the shape of the CM cost surface and, consequently, the performance of SGD minimization algorithms such as CMA. These issues are discussed further in an upcoming subsection.

Robustness Properties of the CM Criterion

This section summarizes important analytical results pertaining to the robustness of CM-minimizing equalizer performance to violations of the PBE conditions. Though we will focus on the robustness to noise and channel undermodelling, we note that evidence of CM robustness to nearly-common subchannel roots appears in [Fijalkow TSP 97], and evidence of robustness to correlated sources appears in [LeBlanc IJACSP 98].

The first two subsections below describe performance approximations obtained using perturbation techniques. In general, such techniques measure the effect of local deviations from a reference setting (e.g., a zero-forcing or MMSE receiver) and are generally valid for small deviations about this reference. Later in

[14] The complex-valued case requires calculating the partial derivatives of J_{cm} with respect to the real part and imaginary part of $\mathbf{f}$ separately. For example, the Hessian could be taken with respect to the real-valued vector $\bar{\mathbf{f}} = (\mathrm{Re}\,\mathbf{f}^T, \mathrm{Im}\,\mathbf{f}^T)^T$.

the section, we will study the effect of noise on the CM criterion using a technique that does not rely on perturbational approximations but, instead, uses a geometric approach to generate tight bounds on performance.

Perturbation in Noise Power. Noise is unconditionally present in practical environments, thus violating **(A2)**. In such cases, both the achieved system delay and the distance between subchannel roots may significantly impact the performance of a CM receiver [Fijalkow SPWSSAP 96]. A small amount of noise can be viewed as a perturbation of the (noise-free) parameters of a given equalizer design. For example, noise perturbation analysis has been considered around noise-free CM receivers in [Fijalkow TSP 97, Fijalkow GRETSI 95], ZF receivers in [Li TSP 96b], and Wiener receivers in [Zeng ICC 96],[Zeng ASIL 96]. This method is justified by the fact that the CM-cost is a smooth (polynomial) function of the equalizer parameters, and assumes that a CM receiver is indeed in the neighborhood of the ZF or MMSE receiver around which the perturbation analysis is performed. Such an assumption may not always be valid as shown in [Chung ICASSP 98] and [Gu CISS 98].

Following the approach in [Fijalkow TSP 97], the CM-cost (for a unit-variance source and Gaussian noise) can be written in terms of the noise-free cost plus a term proportional to the noise power σ_w^2:

$$J_{\mathrm{cm}} = \tilde{J}_{\mathrm{cm}} + \frac{\sigma_w^2}{4}\|\mathbf{f}\|_2^2 \big(2(\kappa_g \|\mathbf{H}^T\mathbf{f}\|_2^2 - \kappa_s) + \kappa_g \sigma_w^2 \|\mathbf{f}\|_2^2 \big), \tag{1.74}$$

where $\tilde{J}_{\mathrm{cm}}$ is the following noise-free CM cost function:

$$\tilde{J}_{\mathrm{cm}} \triangleq \frac{1}{4}\mathrm{E}\big\{ \big(|\mathbf{f}^T\mathbf{H}\mathbf{s}(n)|^2 - \tfrac{\mathrm{E}\{|s_n|^4\}}{\sigma_s^2} \big)^2 \big\}. \tag{1.75}$$

The equalizer minimizing J_{cm} must therefore strike a balance between the noise enhancement appearing as the second term of (1.74) and the CM penalty imposed by $\tilde{J}_{\mathrm{cm}}$. The behavior of CM receivers in the presence of noise can then be analyzed using a Taylor series expansion of the CM cost around the noise-free CM receiver. When **(A1)** is satisfied, [Fijalkow TSP 97] uses a first-order approximation of $\mathbf{q}$-space minima relocation as a function of noise power to yield the following MSE approximation:

$$J_{\mathrm{m}}^{(\delta)}(\mathbf{f}_c^{(\delta)}) \approx \sigma_w^2 \mathbf{e}_\delta^t (\mathbf{H}^t\mathbf{H})^{-1}\mathbf{e}_\delta + o(\sigma_w^2).$$

We note that the expression above is equivalent to a first-order noise-power expansion of MSE about the Wiener receiver. Under lack of disparity (i.e., the presence of common subchannel roots—violating **(A1)**) the noise-free CM receivers are difficult to express analytically. However, when the equalizer is long enough to mitigate the baud-spaced channel component lacking disparity (see (1.24)), sufficiently long CM receivers have been shown to yield MSE similar to that of the Wiener receivers when operating in the presence of channel noise [Fijalkow TSP 97].

[Li TSP 96b] studies the effect of noise on the baud-spaced CM receiver relative to the (IIR) ZF receiver. Similar to the techniques described above, they approximate the resulting MSE using a Taylor series expansion. In the presence of noise, the CM receivers are shown to yield lower MSE than the ZF receivers.

Since the ZF receivers are susceptible to noise gain, this result should not come as a surprise. Recall that the presence of nearly common subchannel roots can make the performance of ZF receivers very sensitive to noise.

Leveraging Godard's conjecture concerning the proximity of CM receivers to Wiener receivers, [Zeng ICC 96, Zeng ASIL 96] perform a noise perturbation analysis around the Wiener receiver. Under satisfaction of the remaining PBE conditions, this approach shows that the CM receiver is collinear with the Wiener receiver up to $O(\mathcal{E}_{\rm m}^3)$, where we have been using $\mathcal{E}_{\rm m}$ to denote the MSE of the Wiener receiver.

Perturbation in Equalizer Length. [Endres SPAWC 97, Endres ICASSP 97] study the robustness properties of the CM criterion (and of CMA) to the sub-optimal but realistic situation in which the number of FSE coefficients is less than that needed to perfectly cancel ISI, thus violating **(A1)**. The main conclusions of this work include evidence that "small" undermodelled channel coefficients produce only a mild deformation of the CM cost surface, and that Wiener receivers corresponding to system delays of better MSE performance have CM minima in closer proximity than those corresponding to system delays of worse performance. Endres's work also results in design guidelines for the choice of equalizer length, the subject of Section 1.4.

Endres presents an algebraic analysis describing the deformation of the CM cost surface from the case of PBE satisfaction as a result of (i) channel coefficients outside the time span of the FSE and (ii) equalizer coefficients lost in truncation, both violating **(A1)**. Using microwave channel models from the SPIB[15] database, it is shown that when the undermodelled channel coefficients are "small," the change in the CM cost is "small" and approximately equal to the scaled change in the MSE cost, suggesting that the Wiener minima and CM minima stay in a tight neighborhood.

As a means of studying the relationship between undermodelled CM cost and achieved system delay, [Endres SPAWC 97, Endres ICASSP 97] construct a second-order Taylor series approximation of the binary CM cost function about the Wiener solution. Since this approximation is quadratic in the FSE coefficients, a closed-form expression can be obtained for approximations of the CM minima corresponding to various system delays. A measure of the proximity of CM and Wiener minima is derived, suggesting that the Wiener receivers corresponding to system delays of good MSE performance have CM receivers in closer proximity than do the Wiener receivers corresponding to system delays of worse MSE performance.

Exact Analysis of the CM Criterion

Although both [Godard TCOM 80] and [Treichler TASSP 83] suggested that CMA offers approximately Wiener-like MSE performance, it was [Foschini ATT 85] who first established the equivalence between the CM and Wiener receivers under the perfect conditions outlined earlier. [Ding TCOM 91] first established the potential for CM local minima under violation of the SPE condition **(A1)**. Such loss of SPE prevents, in general, the equivalence between CM and Wiener receivers.

[15] The SPIB database resides at `http://spib.rice.edu/spib/select_comm.html`.

The presence of noise, violating **(A2)**, is another factor preventing the theoretical equivalence between the CM and Weiner receivers. In fact, most CM analysis is made quite difficult by the presence of noise. For example, the task of characterizing the full set of CM-cost stationary points—let alone distinguishing which are local minima—becomes far more complicated than in the noiseless SPE case. Although the perturbation analyses previously described can be used to assess the MSE performance of CM receivers (and are reasonably accurate at high SNR), such approaches rely on *approximations* of the CM cost function and, as such, have two main drawbacks. First, they presume the existence of a CM receiver in a neighborhood of the reference (e.g., the zero-forcing or Wiener receiver), which may not be the case. Second, they make it difficult to characterize the accuracy of the analysis given a particular SNR.

We describe here an alternative approach [Zeng TIT 98] to the analysis of the CM criterion that does not involve an approximation of the CM cost function. This approach, commonly referred to as the "exact CM analysis," yields the following results:

1. a signal space property and power constraints that apply to all CM receivers,
2. an analytical expression that can be used to verify the existence of a CM local minimum in a particular neighborhood of a Wiener receiver,
3. an analytical description of the regions which may contain CM local minima in the neighborhoods of Wiener receivers, and
4. upper and lower bounds on the MSE of these CM receivers.

Unlike techniques based on equilibrium point analysis and CM cost surface local curvature, this approach to analyzing FIR CM receivers is geometrical and is based on the Weierstrass maximum theorem [Luenberger Book 90, p. 40]. As illustrated in Fig. 1.14, we attempt to find a compact region of equalizer parameter space, called $\mathcal{B}$, with boundary $\partial\mathcal{B}$ and interior reference $\mathbf{f}_{\rm r}$, such that the CM cost $J_{\rm cm}$ everywhere on the boundary is greater than the CM cost at the reference. As a consequence of the Weierstrass theorem, the continuity of the CM cost function implies that at least one CM local minimum exists in $\mathcal{B}$.

Critical in this analysis is the selection of the shape of $\mathcal{B}$, the location of $\mathcal{B}$, and the location of the reference $\mathbf{f}_{\rm r}$. In defining the region $\mathcal{B}$, the first goal is for it to be as small as possible. This should lead to more accurate descriptions of the local minimum and its MSE performance. The second goal is to relate the region $\mathcal{B}$ to the location of the neighboring Wiener receiver. In addition to the description of $\mathcal{B}$ in the parameter space, we give a description in the Hilbert space of the observations that provides important

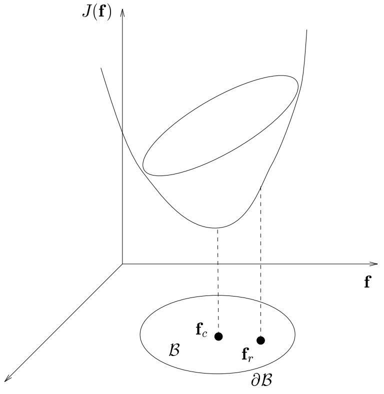

Figure 1.14: Illustration of the main idea underlying the "exact analysis."

physical interpretations.

We emphasize that these results apply only to CM local minima sufficiently near the corresponding Wiener solutions. Therefore, if CM local minima exist within the regions defined by our approach, then one of these minima must be the globally MSE-optimal CM receiver.

The Signal Space Property of CM Receivers. As pointed out in [Ericson TIT 71], a receiver designed by any "reasonable" criterion of goodness includes a matched filter as its front-end. Recalling that system model (1.16) formulates the time-n received vector as $\mathbf{r}(n) = \mathbf{H}\mathbf{s}(n) + \mathbf{w}(n)$, the presence of a matched filter is equivalent to a receiver having the *signal space property*: the receiver is a linear combination of the columns of the channel matrix $\mathbf{H}$. The signal space property is important for several reasons. First, receivers with the signal space property will filter out all noise orthogonal to the signal space. Second, under the SPE condition *and* the signal space property, the analysis of the receiver parameter $\mathbf{f}$ is equivalent to the analysis of system response parameter $\mathbf{q}$. The equivalence between the equalizer and system spaces considerably simplifies the analysis that follows.

Noting that the output power can be written $\mathrm{E}\{|y_n|^2\} = \mathbf{f}^T\,\mathrm{E}\{\mathbf{r}(n)\mathbf{r}^T(n)\}\mathbf{f} = \mathbf{f}^T\mathbf{R}_{r,r}\mathbf{f} = \|\mathbf{f}\|^2_{\mathbf{R}_{r,r}}$, we state what is known as the *CM power constraint*:

Theorem 1.1 *All local minima of the CM cost function are in the signal subspace $\mathcal{Y}_{\mathbf{H}}$. (That is, for any CM local minimum $\mathbf{f}_c$, there exists a vector $\mathbf{v}$ such that $\mathbf{f}_c = \mathbf{H}\mathbf{v}$.) Furthermore, when $\sigma_w^2 > 0$, the output*

power of any CM receiver $\mathbf{f}_c$ satisfies

$$\frac{\kappa_s}{3} < \|\mathbf{f}_c\|^2_{\mathbf{R}_{r,r}} < 1. \tag{1.76}$$

Proof: See [Zeng TIT 98].

For BPSK transmission in the absence of channel noise, the CM power constraint (1.76) first appeared in [Johnson IJACSP 95]. Geometrically, it implies that *all* CM receivers lie in an elliptical "shell" in parameter space, as illustrated by Fig. 1.15 for the two-parameter case.

It is interesting to compare the output power of a CM receiver with that of the ZF and Wiener receivers. For the ZF receiver corresponding to system delay δ, we have (assuming $\sigma_s^2 = 1$ and $\sigma_w^2 > 0$)

$$\|\mathbf{f}_z^{(\delta)}\|^2_{\mathbf{R}_{r,r}} = \mathbf{e}_\delta^T\mathbf{H}^\dagger(\mathbf{H}\mathbf{H}^T + \sigma_w^2\mathbf{I})(\mathbf{H}^T)^\dagger\mathbf{e}_\delta = 1 + \sigma_w^2\|\mathbf{f}_z^{(\delta)}\|_2^2 > 1.$$

Contrast this to the Wiener receiver, for which it is known that

$$\|\mathbf{f}_{\mathrm{m}}^{(\delta)}\|^2_{\mathbf{R}_{r,r}} < 1.$$

Therefore, the output powers of Wiener receivers are always less than 1, whereas the output powers of ZF receivers are greater than 1. As SNR decreases to zero, the output power of Wiener receivers approaches zero and the output power of ZF receivers approaches infinity, but the output power of CM receivers stays between $\frac{\kappa_s}{3}$ and 1. This condition, particularly the lower bound, is useful in determining if a CM local minimum exists near the Wiener receiver.

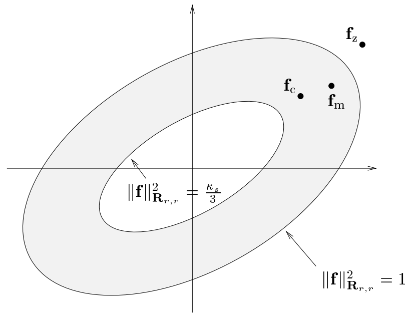

Figure 1.15: Region of CM local minima.

In the sequel, we assume that the SPE condition **(A1)** is satisfied (i.e., $\mathbf{H}$ is full column-rank). Thanks to the signal space property, portions of the following analysis are performed in the system parameter (i.e., $\mathbf{q}$) space.

Location of CM Receivers. Our estimates of the CM receiver's location and achieved MSE are obtained by first specifying a neighborhood $\mathcal{B}$ near the Wiener receiver and then comparing the CM cost on the boundary $\partial\mathcal{B}$ with that of a reference $\mathbf{q}_\mathrm{r}$ contained in $\mathcal{B}$.

For the remainder of the section we assume that the source is BPSK, implying that $\kappa_s = \sigma_s^2 = 1$.

The Neighborhood. The neighborhood has different but equivalent definitions in the equalizer parameter space and the equalizer output space. To keep the notation simple, we consider a receiver $\mathbf{f}$ that estimates the first symbol s_0 of a transmitted source vector $\mathbf{s}$ given the noisy observation $\mathbf{r} = \mathbf{H}\mathbf{s} + \mathbf{w}$. In this context, the receiver gain θ and the conditionally unbiased MSE (UMSE) are defined as follows:

$$\theta \triangleq \mathbf{f}^T\mathbf{H}\mathbf{e}_0, \qquad \mathrm{UMSE} \triangleq \mathrm{E}\{|\tfrac{1}{\theta}\mathbf{f}^T\mathbf{r} - s_0|^2\}. \tag{1.77}$$

Note that $\frac{1}{\theta}\mathbf{f}^T\mathbf{r}$ is a conditionally unbiased estimate of s_0 in the sense that

$$\mathrm{E}\left\{\tfrac{1}{\theta}\mathbf{f}^T\mathbf{r}\,\middle|\, s_0\right\} = s_0. \tag{1.78}$$

Shown in Fig. 1.16 is the geometry associated with the linear estimation of s_0 from $\mathbf{r}$. The output of any linear estimator must lie in the plane $\mathcal{Y}$ spanned by the components of $\mathbf{r}$. The output y_m of the Wiener receiver is obtained by projecting s_0 onto $\mathcal{Y}$. If we scale y_m to u_m such that the projection of u_m onto the direction of s_0 is s_0 itself, we obtain the so-called (conditionally) unbiased minimum mean-square error (U-MMSE) estimate of s_0. Indeed, u_m is conditionally unbiased since $E\{u_\mathrm{m}|s_0\} = s_0$. Furthermore, it is easy to see from Fig. 1.16 that u_m has the shortest distance to s_0 (and hence the minimum MSE) among all conditionally unbiased estimates. Note that the output of a conditionally unbiased estimator must lie on the line $\overline{AB}$ in the figure.

Shown in the shaded area of Fig. 1.16 is a neighborhood of estimates whose receiver gain (obtained by projecting the estimate in the direction of s_0) is bounded in the interval $(\theta_\mathrm{L}, \theta_\mathrm{U})$, and whose corresponding conditionally unbiased estimates of s_0 have mean-square error at most ρ_U^2 greater than $\mathrm{MSE}(u_\mathrm{m})$. In other words, the estimates in the shaded region have extra UMSE upper bounded by ρ_U^2.

To translate this neighborhood to the parameter space, let the system response $\mathbf{q} = \mathbf{H}^T\mathbf{f}$, where $\mathbf{q} = (q_0, q_1, \ldots, q_{L_f+L_h})^T$, have the following parameterization:

$$\begin{aligned} \theta &\triangleq q_0 = \mathbf{e}_0^T\mathbf{q} && (1.79)\\ \mathbf{q}_\mathrm{I} &\triangleq \frac{1}{q_0}(q_0, q_1, \ldots, q_{L_f+L_h})^T. && (1.80) \end{aligned}$$

The receiver output y can then be expressed as

$$y = \underbrace{\theta}_{\text{gain}} \cdot s_0 + \underbrace{\sum_{i\neq 0} q_i s_i}_{\text{interference}} + \underbrace{\mathbf{f}^T\mathbf{w}}_{\text{noise}}, \tag{1.81}$$

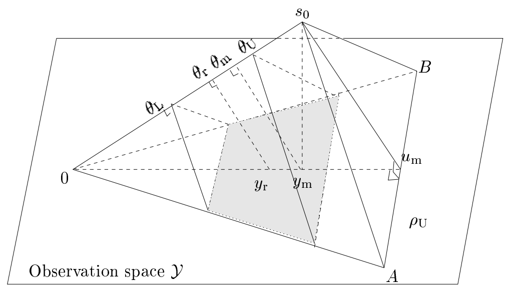

Figure 1.16: The region $\mathcal{B}$ in the Hilbert space of the observations.

where θ is the receiver gain, or bias. Scaling y by $1/\theta$, we have the (conditionally) unbiased estimate of s_0:

$$u \triangleq \frac{y}{\theta} = s_0 + \underbrace{\frac{1}{\theta}\Big(\sum_{i\neq 0} q_i s_i + \mathbf{f}^T\mathbf{w}\Big)}_{\text{equivalent noise}}. \tag{1.82}$$

Therefore, the receiver gain and UMSE of $\mathbf{q}$ is given by θ and $\mathrm{MSE}(u)$ respectively. Hence, the shaded neighborhood in Fig. 1.16 can be described as

$$\{\mathbf{q} : \theta_\mathrm{L} < \theta < \theta_\mathrm{U},\ \mathrm{MSE}(u) - \mathrm{MSE}(u_\mathrm{m}) < \rho_\mathrm{U}^2\}. \tag{1.83}$$

In this definition, θ_L (θ_U) specifies the lower (upper) bound on CM receiver gain, while ρ_U^2 specifies the upper bound on extra UMSE.

Although the neighborhood defined above is related to specific characteristics of a receiver (i.e., UMSE and bias), its relationship to the receiver coefficients ($\mathbf{q}$) is not given explicitly. To locate the CM receiver using this neighborhood, it is necessary to translate the neighborhood above to one that is specified in terms of the system parameter space. Given the Wiener receiver $\mathbf{q}_\mathrm{m} = \theta_\mathrm{m}\binom{1}{\mathbf{q}_{\mathrm{m}\mathrm{I}}}$, it can be shown [Zeng TIT 98] that an equivalent neighborhood, illustrated in the parameter space by Fig. 1.17, is given by

$$\begin{aligned} \mathcal{B}(\mathbf{q}_\mathrm{m}, \rho_\mathrm{U}, \theta_\mathrm{L}, \theta_\mathrm{U}) &\triangleq \{\mathbf{q} : \theta_\mathrm{L} < \theta < \theta_\mathrm{U},\ \|\mathbf{q}_\mathrm{I} - \mathbf{q}_{\mathrm{m}\mathrm{I}}\|_\mathbf{C} < \rho_\mathrm{U}\} && (1.84)\\ &= \{\mathbf{q} : \theta_\mathrm{L} < \theta < \theta_\mathrm{U},\ \mathrm{MSE}(u) - \mathrm{MSE}(u_\mathrm{m}) < \rho_\mathrm{U}^2\}, && (1.85) \end{aligned}$$

where matrix $\mathbf{C}$ is a matrix formed by removing the first row and column from $\big(\mathbf{I} + \sigma_w^2\mathbf{H}^\dagger(\mathbf{H}^\dagger)^T\big)$.

The Reference $\mathbf{q}_\mathrm{r}$. In relating the CM receiver to its Wiener counterpart, we choose, in the direction of the Wiener receiver $\mathbf{q}_\mathrm{m} = \theta_\mathrm{m}\binom{1}{\mathbf{q}_{\mathrm{m}\mathrm{I}}}$, the reference point

$$\mathbf{q}_r \triangleq \theta_\mathrm{r}\begin{pmatrix} 1 \\ \mathbf{q}_{m\mathrm{I}} \end{pmatrix} \tag{1.86}$$

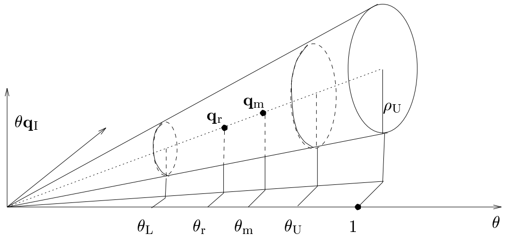

Figure 1.17: A cone-type region: $\mathcal{B}(\mathbf{q}_{\mathrm{m}}, \rho_{\mathrm{U}}, \theta_{\mathrm{L}}, \theta_{\mathrm{U}})$.

with the minimum CM cost. In other words, θ_{r} is chosen to minimize the CM cost of $\mathbf{q} = \theta\binom{1}{\mathbf{q}_{\mathrm{mI}}}$ with respect to θ, which can be shown to obey

$$\theta_{\mathrm{r}}^2 \;=\; \frac{\theta_{\mathrm{m}}}{3 - 2\theta_{\mathrm{m}}^2 - 2\theta_{\mathrm{m}}^2\|\mathbf{q}_{\mathrm{mI}}\|_4^4}. \tag{1.87}$$

In analyzing CM local minima in the neighborhood of Wiener receivers, the role of $\mathbf{q}_{\mathrm{r}}$ turns out to be more than only technical: it can be shown that $\mathbf{q}_{\mathrm{r}}$ is a very good approximation of the CM receiver.

The theorem below provides us with a test for the existence of a CM receiver in $\mathcal{B}(\mathbf{q}_{\mathrm{m}}, \rho_{\mathrm{U}}, \theta_{\mathrm{L}}, \theta_{\mathrm{U}})$.

Theorem 1.2 *Given the Wiener receiver* $\mathbf{q}_{\mathrm{m}} = \theta_{\mathrm{m}}\binom{1}{\mathbf{q}_{\mathrm{mI}}}$ *and the neighborhood* $\mathcal{B}(\mathbf{q}_{\mathrm{m}}, \rho_{\mathrm{U}}, \theta_{\mathrm{L}}, \theta_{\mathrm{U}})$ *defined in (1.84), define*

$$D(\rho) \;\triangleq\; c_1(\rho)^2 - 4c_2(\rho)c_0, \tag{1.88}$$

where

$$c_0 \;=\; \frac{1}{3 - 2\theta_{\mathrm{m}}^2 - 2\theta_{\mathrm{m}}^2\|\mathbf{q}_{\mathrm{mI}}\|_4^4}, \tag{1.89}$$

$$c_1(\rho) \;=\; -2(\rho^2 + \frac{1}{\theta_{\mathrm{m}}}), \tag{1.90}$$

$$c_2(\rho) \;=\; 3(\rho^2 + \frac{1}{\theta_{\mathrm{m}}})^2 - 2(1 + (\rho + \|\mathbf{q}_{\mathrm{mI}}\|_4)^4). \tag{1.91}$$

Under the conditions that MMSE $< \frac{1}{3}$ *and* $D(\|\mathbf{q}_{\mathrm{mI}}\|_2) < 0$*, a CM local minimum exists in the neighborhood* $\mathcal{B}_p(\mathbf{q}_{\mathrm{m}}, \rho_{\mathrm{U}}^*, \theta_{\mathrm{L}}^*, \theta_{\mathrm{U}}^*)$*. Here,* ρ_{U}^* *is the smallest positive root of* $D(\rho)$*, and*

$$\theta_{\mathrm{L}}^* \;=\; \min_{0\le\rho\le\rho_{\mathrm{U}}^*} \sqrt{\frac{-c_1(\rho) - \sqrt{c_1(\rho)^2 - 4c_2(\rho)c_0}}{2c_2(\rho)}}, \tag{1.92}$$

$$\theta_{\mathrm{U}}^* \;=\; \max_{0\le\rho\le\rho_{\mathrm{U}}^*} \sqrt{\frac{-c_1(\rho) + \sqrt{c_1(\rho)^2 - 4c_2(\rho)c_0}}{2c_2(\rho)}}. \tag{1.93}$$

Furthermore, as the noise power decreases

$$\lim_{\sigma_w\to 0} \mathcal{B}_p(\mathbf{q}_{\mathrm{m}}, \rho_{\mathrm{U}}^*, \theta_{\mathrm{L}}^*, \theta_{\mathrm{U}}^*) \;=\; \lim_{\sigma_w\to 0} \mathbf{q}_{\mathrm{c}} \;=\; \lim_{\sigma_w\to 0} \mathbf{q}_{\mathrm{m}} \;=\; \mathbf{q}_{\mathrm{z}}. \tag{1.94}$$

Proof: See [Zeng TIT 98].

<u>*Remarks on Theorem 1.2*</u>. Some comments on the implications of Theorem 1.2 follow:

- Given $\mathbf{q}_{\mathrm{m}} = \theta_{\mathrm{m}}\binom{1}{\mathbf{q}_{\mathrm{mI}}}$, this theorem can be used (i) to detect the existence of a CM local minimum by checking the sufficient condition $D(\|\mathbf{q}_{\mathrm{mI}}\|_2) < 0$, and (ii) to determine the region $\mathcal{B}_p(\mathbf{q}_{\mathrm{m}}, \rho_{\mathrm{U}}^*, \theta_{\mathrm{L}}^*, \theta_{\mathrm{U}}^*)$.
- Equation (1.94) states that the region $\mathcal{B}_p(\mathbf{q}_{\mathrm{m}}, \rho_{\mathrm{U}}^*, \theta_{\mathrm{L}}^*, \theta_{\mathrm{U}}^*)$ shrinks to the Wiener/ZF receiver as noise vanishes, which is intuitively satisfying.

<u>**MSE of CM Receivers**</u>. Theorem 1.2 also enables us to give performance bounds on the MSE of CM receivers:

Theorem 1.3 *Given the Wiener receiver* $\mathbf{q}_{\mathrm{m}} = \theta_{\mathrm{m}}\binom{1}{\mathbf{q}_{\mathrm{mI}}}$*, suppose that the MMSE is less than* $\frac{1}{3}$*, the condition in Theorem 1.2 is satisfied, and* $\mathbf{q}_{\mathrm{c}} \in \mathcal{B}_p(\mathbf{q}_{\mathrm{m}}, \rho_{\mathrm{U}}^*, \theta_{\mathrm{L}}^*, \theta_{\mathrm{U}}^*)$ *is a CM minimum. Then, letting* $\Delta\mathcal{E}$ *be the extra MSE of* $\mathbf{q}_{\mathrm{c}}$ *defined by* $\Delta\mathcal{E} \triangleq J_{\mathrm{m}}((\mathbf{H}^T)^\dagger \mathbf{q}_{\mathrm{c}}) - J_{\mathrm{m}}((\mathbf{H}^T)^\dagger \mathbf{q}_{\mathrm{m}})$,

$$\underbrace{\frac{(\theta_{\mathrm{U}}^* - \theta_{\mathrm{m}})^2}{\theta_{\mathrm{m}}}}_{\Delta\mathcal{E}_{\mathrm{L}}} \;\le\; \Delta\mathcal{E} \;\le\; \underbrace{\frac{(\theta_{\mathrm{L}}^* - \theta_{\mathrm{m}})^2}{\theta_{\mathrm{m}}} + (\theta_{\mathrm{U}}^*\rho_{\mathrm{U}}^*)^2}_{\Delta\mathcal{E}_{\mathrm{U}}}. \tag{1.95}$$

Proof: See [Zeng TIT 98]. The bounds in (1.95) involve the computation of $\theta_{\mathrm{L}}^*, \theta_{\mathrm{U}}^*$, and ρ_{U}^*. Note that if the $\mathcal{B}$ is small in size, then the MSE of the CM receiver can be approximated by the MSE of the reference, and the location of the CM receiver $\mathbf{q}_{\mathrm{c}}$ can by approximated by the reference location $\mathbf{q}_{\mathrm{r}}$ (which, we recall, is a scaled version of the Wiener receiver).

Behavior of the Constant Modulus Algorithm

Having established the robustness of the CM criterion—that is, the close relationship between CM and MSE minima under violation of the perfect conditions **(A1)** to **(A4)**—we now turn to examination of the behavior of the Constant Modulus Algorithm (CMA), which accomplishes a stochastic gradient descent of the CM cost surface. CMA can be described by the update equation

$$\begin{aligned}\mathbf{f}(n+1) &= \mathbf{f}(n) + \mu\,\mathbf{f}^T(n)\mathbf{r}(n)(\gamma - |\mathbf{f}^T(n)\mathbf{r}(n)|^2)\mathbf{r}^*(n) \\ &= \mathbf{f}(n) + \mu\underbrace{y_n(\gamma - |y_n|^2)}_{\text{CM error function}}\mathbf{r}^*(n).\end{aligned} \tag{1.96}$$

Recall that γ is the dispersion constant typically chosen equal to $\frac{\mathrm{E}\{|s_n|^4\}}{\sigma_s^2}$, $\mathbf{f}(n)$ is the equalizer impulse response vector at time n, and μ is a small step-size. Though (1.96) presents the complex-valued CMA update equation, we focus the remainder of the section on strictly real-valued quantities. Averaging theory [Benveniste Book 90, Ljung Book 83] predicts that a SGD algorithm such as CMA converges almost surely and in mean to the local minima of its cost surface (i.e., the CM receivers).

The performance of CMA can be categorized by its transient and steady-state behaviors. The multimodal nature of the CM cost surface makes analysis of CMA transient behavior (e.g., the rate of convergence to steady-state operation) more difficult than with unimodal linear-equalizer update algorithms such as LMS. CMA's steady-state behavior is characterized by parameter "jitter" around the (local) CM receiver, resulting in non-zero excess CM-cost. In certain cases, the excess CM-cost can be translated directly into excess MSE, a popular measure of adaptive equalizer performance. These topics are discussed in more detail in the following subsections.

Transient Behavior: Convergence Rate. From a given initialization, CMA updates the equalizer by descending the CM cost surface according to estimates of the direction of steepest descent. Due to the multi-modality of J_{cm}, the convergence time (viewed as the number of iterations to reach "steady-state" behavior) is greatly affected by equalizer initialization. Consequently, very few studies exist on CMA convergence rate, [Larimore ICASSP 83], [Touzni EUSIPCO 96], and [Lambotharan SP 97] being exceptions.

To analyze CMA convergence, we consider the average CMA trajectories, or in other words, the trajectories following an (exact) gradient descent of the CM cost surface in equalizer space. For a very small step-size μ, a SGD algorithm of the form

$$\mathbf{f}(n+1) \quad = \quad \mathbf{f}(n) + \mu F\big(\mathbf{f}(n), \mathbf{r}(n)\big)$$

closely follows the average system (see [Benveniste Book 90, Ljung Book 83]) described by the ordinary differential equation (ODE):

$$\frac{d\mathbf{f}(t)}{dt} \quad = \quad \bar{F}(\mathbf{f}),$$

where $\bar{F}(\mathbf{f}) = \mathrm{E}\{F\big(\mathbf{f}, \mathbf{r}(n)\big)\}$ represents the mean update term of the algorithm and where t is a continuous variable proportional to n.

In the absence of noise, CMA's ODE can be derived from the mean CMA update term $\mathrm{E}\{\mathbf{r}(n)\mathbf{f}^t\mathbf{r}(n)(\gamma - |\mathbf{f}^t\mathbf{r}(n)|^2)\} = \sigma_s^4 \mathbf{H}\boldsymbol{\Delta}_{\mathbf{q}}\mathbf{q}$ (recall (1.70)). Transforming to the system space, the ODE becomes

$$\frac{d\mathbf{q}}{dt} = \mathbf{H}^t\frac{d\mathbf{f}}{dt} \quad = \quad \sigma_s^4 \mathbf{H}^t\mathbf{H}\boldsymbol{\Delta}_{\mathbf{q}}\mathbf{q}. \tag{1.97}$$

The trajectories resulting from the solution of (1.97) represent average paths for the evolution of $\mathbf{q}(n)$ when updated by small step-size CMA. A display of some typical ODE trajectories is provided in Fig. 1.18.

Local Behavior. Under the PBE conditions and in the vicinity of a CM minimum (i.e., $\mathbf{q} \approx \mathbf{e}_\delta$), the exact gradient can be approximated by the first term of its Taylor series expansion [Touzni EUSIPCO 96] using

$$\boldsymbol{\Delta}_{\mathbf{q}}\mathbf{q} \quad = \quad \underbrace{\boldsymbol{\Delta}_{\mathbf{e}_\delta}\mathbf{e}_\delta}_{=0} - \Psi_{\mathbf{e}_\delta}(\mathbf{q} - \mathbf{e}_\delta) + o(\|\mathbf{q} - \mathbf{e}_\delta\|_2), \tag{1.98}$$

where it follows from (1.73) that, in the real-valued case,

$$\Psi_{\mathbf{e}_\delta} \quad = \quad (\kappa_g - \kappa_s)\mathbf{I} + 3(\kappa_s - 1)\mathbf{e}_\delta\mathbf{e}_\delta^T. \tag{1.99}$$

Applying this approximation to (1.97), the exact gradient trajectories in the vicinity of $\mathbf{e}_\delta$ are well described by

$$\mathbf{q}(t) \quad = \quad \mathbf{e}_\delta + \exp(-t\sigma_s^4\mathbf{H}^t\mathbf{H}\Psi_{\mathbf{e}_\delta})\mathbf{q}(0). \tag{1.100}$$

The time constant τ_{cma} associated with the local convergence of $\mathbf{q}(n)$ in the neighborhood of the global minimum $\mathbf{e}_\delta$ (i.e., $\mathbf{q}(n) = \mathbf{e}_\delta + e^{-n/\tau_{\text{cma}}}\mathbf{q}(0)$) can be roughly bounded using a quadratic approximation to the CM cost in the neighborhoods of local minima [Lambotharan SP 97]. This approach is equivalent to a further approximation of (1.98) in which case the second term of $\Psi_{\mathbf{e}_\delta}$ in (1.99) disappears (as occurs naturally with BPSK). The trajectory resulting from the application of this $\Psi_{\mathbf{e}_\delta}$ approximation to (1.100) with $\sigma_s^2 = 1$ is characterized by a local-convergence time constant

$$\frac{1}{\mu(\kappa_g - \kappa_s)\lambda_{\max}(\mathbf{H}^T\mathbf{H})} \;\le\; \tau_{\text{cma}} \;\le\; \frac{1}{\mu(\kappa_g - \kappa_s)\lambda_{\min}(\mathbf{H}^T\mathbf{H})}.$$

Here $\lambda_{\min}(\mathbf{H}^T\mathbf{H})$ and $\lambda_{\max}(\mathbf{H}^T\mathbf{H})$ indicate minimum and maximum eigenvalues of $\mathbf{H}^T\mathbf{H}$, respectively. Note that the term $(\kappa_g - \kappa_s)$ relates the steepness of the the locally-approximated CMA cost surface to that of the LMS cost surface. For example, with a BPSK source, the local CMA surface curvature is twice that of LMS. Observe, then, that an increase in source kurtosis corresponds to a decrease in (local) convergence rate, where a Gaussian source provides the limiting case: zero convergence rate.

The locally-quadratic approximation implies the following CMA step-size guideline (for $\sigma_s^2 = 1$), analogous to the LMS result in (1.60):

$$0 < \mu \le \frac{2}{(\kappa_g - \kappa_s)\lambda_{\max}(\mathbf{H}^T\mathbf{H})}. \tag{1.101}$$

Equation (1.101) gives a useful step-size guideline when considering trajectories in the vicinity of CM local minima, that is, near the point of convergence. In the quest for more general step-size guidelines, the quartic nature of J_{cm} (see (1.68)) suggests that a reasonable step-size upper bound should decrease for trajectories far from the origin. Specifically, [Larimore ICASSP 83] suggests that the upper bound should be equalizer dependent: inversely proportional to $\|\mathbf{f}\|_2^2$.

Global Behavior. The multi-modality of the CM cost function prevents a simple, direct comparison of time-to-convergence between CMA and LMS. An approximate understanding, however, may be obtained from the following two-stage convergence model for CMA:

1. Fast convergence to a region $\mathcal{S}$ near the unit sphere in system space (i.e., $\|\mathbf{q}\|_2 = 1$).
2. Within $\mathcal{S}$, slower convergence to a minimum.

[Touzni EUSIPCO 96] shows that the norm of the mean gradient, $\|\mathbf{H}\boldsymbol{\Delta}_{\mathbf{q}}\mathbf{q}\|_2$, is very large when $\|\mathbf{q}\|_2 \gg 1$ (i.e., far from the unit sphere). Furthermore, in this region, the mean gradient points towards the origin. This implies that from an initialization $\mathbf{q}(0)$ far from the unit sphere, there is fast convergence towards the origin until $\mathbf{q}(n)$ penetrates the vicinity of the unit sphere. A similar effect occurs when $0 < \|\mathbf{q}(0)\|_2 \ll 1$. This time, the gradient points away from the origin so that there is fast convergence to the vicinity of the unit sphere from inside. These effects can be observed in Figs. 1.18 and 1.24.

A more precise definition of the aforementioned "region near the unit-sphere" is the hyper-annulus $\mathcal{S}$ that includes all CM minima and saddle points. From Theorem 1.1, the CM power constraint [Zeng TIT 98] implies

$$\mathcal{S} = \left\{\mathbf{q} \,:\, \frac{\kappa_s}{3} \leq \|\mathbf{q}\|_2^2 \leq 1\right\}.$$

Figure 1.18 shows this region for the simplest case ($L_h + L_f + 1 = 2$ and $\kappa_s = 1$) where the borders of $\mathcal{S}$ are denoted by the dotted circles. For $\mathbf{q}$ in $\mathcal{S}$, the norm of the gradient is small and thus the convergence rate is slowed.

In Fig. 1.18, notice that as the trajectory $\mathbf{q}(t)$ approaches $\mathcal{S}$, it becomes attracted to the nearest stationary point. If the nearest attractor is a minimum, convergence heads towards the minimum at the rate determined by the approximately quadratic character of the region. If the nearest stationary point is a saddle point, the trajectory is first attracted to the saddle before converging to a stable minimum, consequently increasing convergence time [Lambotharan SP 97] as demonstrated by Fig. 1.24. (In higher dimensional cases, the trajectory can be attracted by several saddles before reaching a minimum—see Fig. 1.25.) Our brief observations about CMA's transient behavior highlight the importance of equalizer initialization, since it can be seen that convergence time is highly dependent on the trajectory's proximity to saddle points.

Asymptotic Behavior: Performance at Steady-state. We now analyze the steady-state scenario in which the CMA trajectory has converged to the vicinity of a CM minimum. In general, the CMA-adapted parameters continue to "jitter" around the true CM minimum due to a non-vanishing CMA update. This non-vanishing update may be a result of violations in the Perfect Blind Equalization (PBE) conditions and/or the use of a non-constant-modulus source (shown in Fig. 1.19).

We may consider quantifying the amount of parameter jitter in terms of *excess CM-cost*, defined as the steady-state CM cost above that achieved by the fixed local CM minimum. The excess CM cost induced by CMA is intuitively similar to the excess MSE induced by LMS (see Section 1.2) in that both stem from the adapted parameters "bouncing around the bottom of the bowl," as illustrated by Fig. 1.20. An alternate and perhaps more useful method of quantifying CMA's parameter jitter is in terms of excess MSE. We define the EMSE of CMA as the difference between CMA's steady-state MSE and the MSE of the (fixed) local CM receiver[16].

[16]Note that the extra MSE of CMA is different from the "extra MSE" of CMA used in Theorem 1.3.

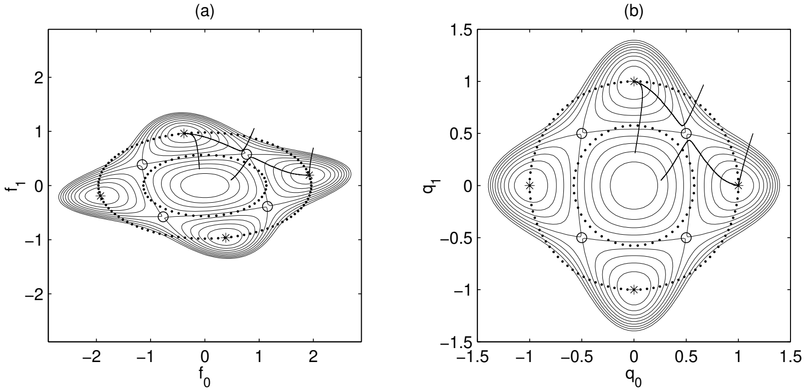

Figure 1.18: Exact gradient trajectories in (a) equalizer and (b) system space, with $L_f = 0$, $L_h = 1$ and BPSK. Both figures show dotted hyper-annulus $\mathcal{S}$ boundaries and denote saddle points by '$\circ$' .

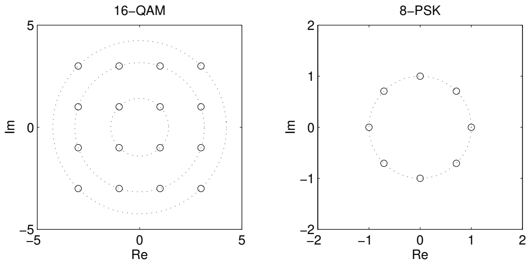

Figure 1.19: Non-constant modulus source constellation (16-QAM) versus constant modulus source constellation (8-PSK).

When the PBE conditions are satisfied, the CM receiver achieves perfect symbol recovery (i.e., $y_n = s_{n-\delta}$) and hence zero MSE. In this case, the excess MSE of CMA is equal to the total MSE of CMA and, thus, is relatively straightforward to calculate (see Fig. 1.20(a)). When the locations of the Wiener and CM receivers are distinctly different, however, the steady-state EMSE of CMA is more difficult to calculate (see Fig. 1.20(b)). Fortunately, Godard's conjecture suggests that mild violations of the PBE conditions preserve the proximity between the CM and Wiener solutions, implying that EMSE derivations assuming PBE conditions are expected to hold suitably well in realistic (non-PBE) situations.

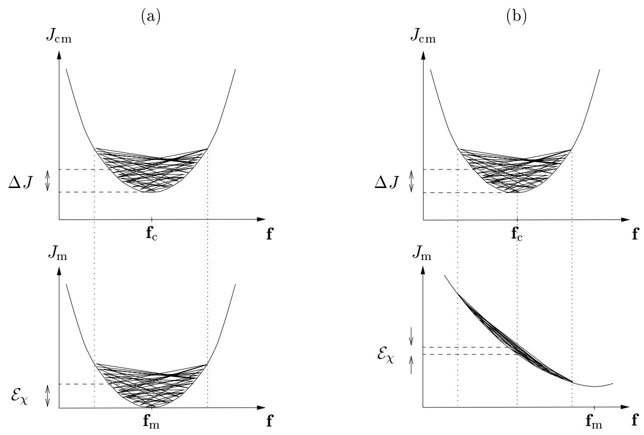

Figure 1.20: Illustration of the relationship between excess CM-cost ΔJ resulting from CMA adaptation and excess MSE $\mathcal{E}_\chi$ of the same CMA-adapted trajectory under (a) small and (b) larger violations of the PBE conditions.

__EMSE due to a non-CM Source__. As stated earlier, the CM receiver is capable of perfect symbol recovery when the PBE conditions are satisfied. Even so, the use of a non-CM source and a non-vanishing step-size μ allows only mean-convergence of $\mathbf{q}(n)$ to $\mathbf{e}_\delta$. As evident from (1.96), a non-zero update will result from a fixed equalizer's inability to force the instantaneous squared modulus of a non-CM source to a constant (γ). Thus we expect the amount of parameter jitter to depend, in part, on particular source properties. Below, we approximate the steady-state excess MSE of CMA under satisfaction of the PBE conditions.

Following the approach adopted by [Widrow Book 85], the EMSE at time index n can be calculated as:

$$\begin{aligned}
\mathcal{E}_\chi &= \mathrm{E}\{|y_n - s_{n-\delta}|^2\} \\
&= \mathrm{E}\{|(\mathbf{q}(n) - \mathbf{e}_\delta)^T \mathbf{s}(n)|^2\} \\
&= \mathrm{tr}\left(\mathrm{E}\{(\mathbf{q}(n) - \mathbf{e}_\delta)(\mathbf{q}(n) - \mathbf{e}_\delta)^T \mathbf{s}(n)\mathbf{s}(n)^T\}\right) \\
&= \sigma_s^2\, \mathrm{tr}\left(\mathrm{E}\{(\mathbf{q}(n) - \mathbf{e}_\delta)(\mathbf{q}(n) - \mathbf{e}_\delta)^T\}\right).
\end{aligned}$$

The last step above is justified by the usual independence assumption between source symbols and parameter errors [Widrow Book 85]. For small μ, [Fijalkow TSP 98] uses averaging theory to approximate the matrix $\mathbf{P} \triangleq \mathrm{E}\{(\mathbf{q}(n) - \mathbf{e}_\delta)(\mathbf{q}(n) - \mathbf{e}_\delta)^T\}$. At steady state, $\mathbf{P}$ is the unique positive definite solution of the Lyapunov equation $\mathbf{QP} + \mathbf{PQ} = \mathbf{M}$, where $\mathbf{Q} = \Psi_{\mathbf{e}_\delta}$ and

$$\mathbf{M} = \sum_{n=-\infty}^{+\infty} \mathrm{E}\{y_n(\gamma - |y_n|^2)\mathbf{r}(n)\left(y_0(\gamma - |y_0|^2)\mathbf{r}(0)\right)^T\}$$

(see [Benveniste Book 90], p. 107). Under the PBE conditions and at steady-state, the output y_n takes the form $y_n = \mathbf{e}_\delta^T \mathbf{s}(n)$, which, when used above, yields the following approximation for the EMSE of CMA [Fijalkow TSP 98]:

$$\mathcal{E}_\chi \approx \frac{\mu\, P(L_f+1)}{2} \left(\frac{\frac{\mathrm{E}\{s_n^6\}}{\sigma_s^6} - {\kappa_s}^2}{\kappa_g - \kappa_s}\, \sigma_s^4 \right) \sigma_r^2. \tag{1.102}$$

In (1.102), $\sigma_r^2 \triangleq \mathrm{E}\{|r_k|^2\}$ denotes the power of the received signal.

Equation (1.102), approximating the EMSE of CMA, looks similar to Equation (1.63) which approximated the EMSE of LMS. Note the quantities present in each: step-size μ, total number of equalizer taps $P(L_f+1)$, and received power σ_r^2. The EMSE component unique to CMA pertains to the source distribution, from which we make the following observations:

- EMSE is 0 for a CM source, that is, when $\kappa_s = 1$, when the PBE conditions are satisfied.
- EMSE increases as the source distribution approaches Gaussianity, that is, $\kappa_s \to \kappa_g$.
- EMSE is proportional to a sixth-order moment of the source, implying high sensitivity to non-CM source characteristics.

Our observations emphasize that nearly Gaussian source processes are particularly difficult for CMA-adapted blind equalization, which is in agreement with a similar observation made in previous chapters of the book. Specifically, as $(\kappa_g - \kappa_s) \to 0$, it can be shown that $\left(\frac{\mathrm{E}\{s_n^6\}}{\sigma_s^6} - \kappa_s^2\right)$ remains non-zero (see, e.g., [Rosenblatt Book 85]) so that EMSE tends to infinity.

__EMSE due to Violations in PBE Conditions__. As stated earlier, a globally non-zero CM-cost implies a non-zero CMA update at all points in the equalizer space. This, in turn, results in steady-state parameter jitter around CM local minima and yields non-zero steady-state excess CM-cost. Recall that violation of any of the PBE conditions **(A1)-(A3)** yields the following:

- CM minima have CM-cost greater than zero.
- Not all CM minima yield the same CM-cost.
- CM receivers are no longer equivalent to MMSE or ZF receivers.

Hence, we conclude that any of the following: channel undermodelling, lack of subchannel disparity, the presence of channel noise, or the use of a temporally correlated source will increase the steady-state level of excess CMA. An example is presented in Section 1.4, where the difference between the dashed and solid lines in Figs. 1.23(d)-(f) reveals EMSE due to noise, undermodelling, and nearly-common subchannel roots.

The fact that the CM cost surface is nearly quadratic in the vicinity of local minima suggests that the excess CM-cost for CMA should be proportional to the same factors determining EMSE for LMS: step-size, equalizer length, received signal power, and CM-cost at the minimum. Unfortunately, translating this excess CM-cost into excess MSE is not always straightforward, especially for significant violations in the PBE conditions (recall Fig. 1.20). This is due in part to the fact that we lack closed-form expressions for the CM receivers when any of the conditions **(A1)-(A3)** is violated. To summarize, a method of quantifying the EMSE of CMA under arbitrary violation of the PBE conditions remains an open problem.

1.4 Design Issues with Illustrative Examples

This section addresses design issues regarding fractionally spaced equalizers adapted by means of the constant modulus algorithm. In this context, equalizer design may be reduced to the selection of three key parameters: equalizer length, step-size, and initialization. It will be shown that equalizer performance (e.g., MSE or convergence rate) is directly linked to each of these design parameters.

With this motivation, we seek to establish design guidelines for each parameter. When possible, we consider each parameter independently and follow an example-driven tutorial approach in developing these guidelines. The reader may notice that in many cases there are no clear answers to the optimal design choice. Despite the ambiguities, we are able to provide reasonable design guidelines supported by analysis and illustrative examples.

For the remainder of this section, we assume that the designer has a rough knowledge of channel characteristics in order to effectively choose the equalizer design parameters (i.e., blind equalization does not imply "blind" equalizer *design*). We focus on the use of $T/2$-spaced CMA as a blind start-up for eventual transfer to decision-directed (DD) equalization. Finally, we assume that the signalling format is known by the receiver, as is usually the case in cooperative communication.

Equalizer Initialization

Recall that the MSE cost (1.21) is dependent on the choice of system delay. Hence, the objective of MMSE equalization may be considered the minimization of MSE over both the equalizer parameters *and* the system delay, δ. In systems that employ training, the system delay is directly linked to the designer's choice of training sequence delay (relative to a data-derived synchronization interval). In blind receivers employing CMA, *the system delay achieved by the steady-state equalizer is a direct result of the equalizer initialization.*

Take for example the simple two-tap system of Fig. 1.21. This figure shows two CMA equalizer tap trajectories superimposed on the CMA cost contours that converge to different solutions for the channel $\mathbf{h} = (0.2, 0.5, 1.0, -0.1)^T$ at 10dB SNR. Such convergence behavior can be understood by the equalizer-space regions of convergence, whose boundaries are plotted in Fig. 1.21 with dotted lines. The trajectory initialized within the region corresponding to CM minimum "A" achieved MSE= 0.103, while the trajectory initialized within the region corresponding to minimum "B" achieved MSE= 0.294. The CM minimum "A" is near the Wiener solution for delay $\delta = 1$, while "B" neighbors the Wiener solution for $\delta = 0$. Thus, for this channel/noise combination, we conclude that the system delay $\delta = 1$ is optimal. (The minima corresponding to the reflections of "A" and "B" through the origin correspond to sign-inversions of the system responses at "A" and "B," respectively, as discussed in Section 1.3. Allowing for sign ambiguity, these reflected minima have equal performance.)

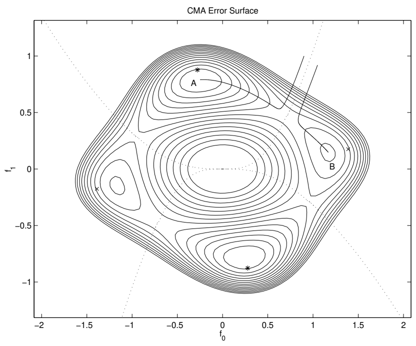

Figure 1.21: Two initializations of CMA with convergence to minima with different MSEs. Convergence region boundaries denoted by dotted lines, and Wiener solutions for $\delta = 0$ and $\delta = 1$ represented by $(\times)$ and $(*)$, respectively.

To understand the effects of initialization on equalizers with more than two taps, we need to consider the MMSE performance achieved at various system delays. Consider the delay-dependent performances in Figs. 1.22(b)-(e), where minimum MSE is plotted versus system delay for FSE lengths of 16, 32, 64, and 299. The channel used for these experiments was SPIB[17] Microwave Channel 3, whose length-300 impulse response magnitude appears in Fig. 1.22(a). The MMSE traces demonstrate a behavior common to nearly all channel-equalizer combinations: system delays far from the channel cursor exhibit relatively poor performance. (The channel *cursor* is defined as the location of peak magnitude in the channel impulse response.) A reasonable conclusion is that, if possible, *CMA should be initialized within the region of attraction associated with the MMSE-optimal system delay.* In practice, it is usually sufficient to initialize within the regions of attraction of minima having MSE below a particular application-dependent threshold.

When the equalizer length is sufficient to satisfy the Perfect Blind Equalization condition **(A1)**, the number of CMA minima (of a given phase) is equal to the number of achievable system delays, and so all delays are attainable through proper choice of initialization. As the length of the FSE decreases below the length required for perfect equalization, the number of CMA minima typically decreases as well, in which case fewer system delays remain reachable. This situation nearly always arises in practical environments, and the literature refers to such equalizers as being *undermodelled* with respect to the channel. As an example, Fig. 1.11 shows one CMA solution near two sub-optimal Wiener solutions, while the other CMA solution remains near the Wiener receiver of optimal delay. Ideally, we would like the limited set of reachable delays

[17]The SPIB database resides at `http://spib.rice.edu/spib/select_comm.html`.

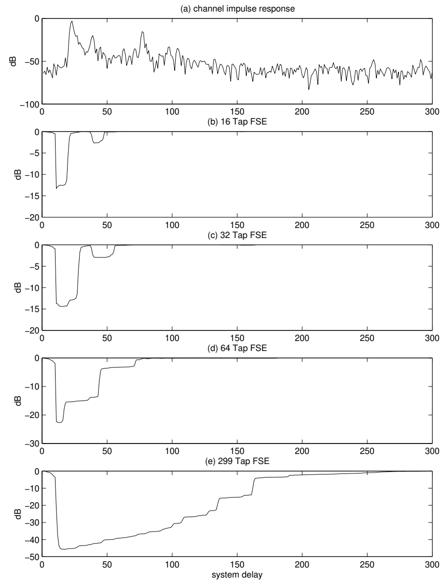

Figure 1.22: (a) SPIB Microwave Channel 3 impulse response magnitude and (b)-(e) achieved MMSE for various system delays and four equalizer lengths (SNR=50dB).

to be those that achieve the best MMSE performance. Our claim is that undermodelled FSE-CMA typically provides this desired behavior, at least using popular initialization strategies. Before substantiating this claim, however, we will discuss these initialization strategies.

Single-Spike Initialization. The most common strategy for initialization of CMA is termed *single-spike* initialization and derives its name from the baud-spaced strategy first suggested in [Godard TCOM 80]. The single-spike technique specifies that all but one of the BSE coefficients are set to zero, though the location and amplitude of the nonzero tap has long been a subject of debate. Recently, [Chung ASIL 98] made the observation that initializations with small norm result in almost certain CMA convergence to "good" local minima, that is, CM minima associated with relatively good MSE performance. Figure 1.21 illustrates this phenomenon for a 2-tap FSE: note how the regions of convergence corresponding to the better minima completely dominate the area near the origin. Unfortunately, the fact that the origin is a local maximum of the CM cost surface implies that initializations with extremely small norm will result in slow (initial) convergence. It has been found that, in practical[18] situations, an initialization norm $\|\mathbf{f}_{\text{init}}\|_2 \approx 1$ is usually a good compromise between convergence rate and region-of-convergence selection.

The single-spike strategy as described above is not always appropriate for fractionally-spaced equalizers. For example, consider the trivial $T/2$-spaced channel $\mathbf{h} = (0, 1, 0, 0)^T$ with the single-spike FSE initialization $\mathbf{f} = (0, 1)^T$. If the equalizer output is decimated such that the odd-indexed samples are retained, then the equalizer output (and hence the CMA update) will be zero! In other words, the subchannel providing the signal energy will be nulled by the initial equalizer. What might be a more appropriate $T/2$-spaced initialization strategy is known as *double-spike* initialization and contains non-zero components in both subequalizers. Note that choosing the non-zero tap magnitudes $\approx 1/\sqrt{2}$ will result in an equalizer norm ≈ 1. It is customary to place the non-zero fractionally-spaced taps adjacent to one another. This practice may be justified in the frequency domain, as adjacent positioning results in an initialization spectrum similar to a raised-cosine-pulse matched filter. The relative placement of the adjacent tap pair will be the focus of an upcoming subsection.

Relationship between FSE-Length and Initialization Sensitivity. At this point we return to our discussion of initialization for undermodelled equalizers. (Recall that undermodelling nearly always exists in practice.) Previously we claimed that, under a reasonable initialization, FSE-CMA achieves a "good" range of reachable system delays. Using the unit-norm double-spike initialization strategy, we now offer example evidence for this claim[19]. Figure 1.23 shows steady-state MSE performance achieved by CMA for the full range of (adjacent) double-spike locations. For ease of comparison, the same channel and equalizer lengths used in Fig. 1.22 are used. Other details include the use of QPSK, a SNR of 50 dB, step-sizes $\{0.005, 0.005, 0.0025\}$

[18]We assume that the received signal power is unit-normalized by automatic gain control prior to equalization.

[19]As no rigorous proofs of optimality exist for the single/double-spike initialization scheme, we are not claiming that the reachable set of delays is *always* optimal. However, practitioners have found these initialization strategies to work well.

for respective FSE lengths $\{16, 32, 64\}$, and simulation lengths of 10^6 iterations. Figure 1.23 has two

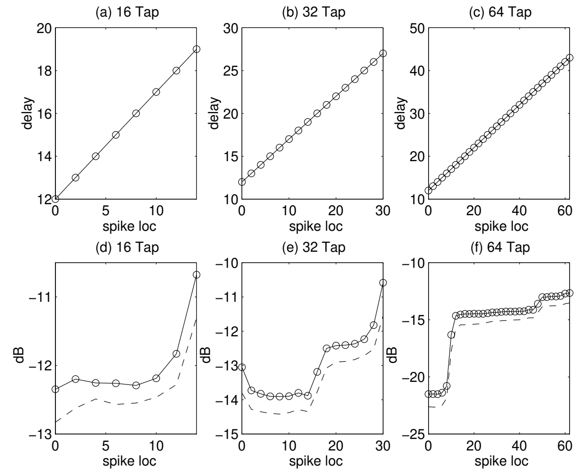

Figure 1.23: (a)-(c) Achieved system delay as a function of double-spike location and (d)-(f) CMA's corresponding MSE performance for the channel response in Fig. 1.22(a). The dashed lines indicate Wiener (i.e., MMSE) performance at the respective system delays.

important messages:

- First, the only system delays reached by FSE-CMA under double-spike initialization were indeed those corresponding to the low-MMSE "troughs" in Fig. 1.22; FSE-CMA avoided convergence to system delays characterized by poor MMSE. (Notice that the dashed lines in Figs. 1.23(d)-(f) correspond to the lowest sections of the MMSE traces in Fig. 1.22.)

- Second, the relative MSE performance versus delay seems to "even out" as the degree of undermodelling increases. In other words, initialization sensitivity *decreases* as equalizer length decreases. Note, however, that for the same (nonzero) noise level, the MMSE performance achieved by shorter equalizers is typically[20] worse. In fact, the best MSE achieved by the 16-tap CMA-adapted equalizer in Fig. 1.23 is not as good as the worst MSE achieved by the 64-tap equalizer.

The shape of the typical "trough" in the MMSE-versus-delay profile [Larimore ICASSP 97] implies that,

[20]Due to the effects of EMSE, longer adaptive equalizers will not always result in lower steady state EMSE, as discussed in Section 1.4 and illustrated by Fig. 1.28.

even for long equalizers, good performance is assured when the initialization spike is chosen anywhere within the region corresponding to the lowest-MMSE delay.

Spike-Delay Selection. We note from Figs. 1.23(a)-(c) that there seems to be a direct correspondence between initial spike location and achieved system delay. A similar observation was made in [Li TSP 95] for single-spike initialized BSEs. Thus, knowing something about the channel impulse response shape can help with the selection of initialization. For example, if the channel impulse response is reasonably symmetric, a center-spike is most appropriate. If, on the other hand, the channel impulse response is asymmetric (i.e., the center of mass is skewed towards the left or right), then the initial spike location should be moved in the direction of the center of mass. An example is the channel response in Fig. 1.22(a), which is skewed very much towards the left. The achieved MSE in Figs. 1.23(d)-(f) confirms that, for this channel, the initialization spike locations toward the left yield better performance. Finally, when little or nothing is known about the channel, the width of the MSE troughs in Figs. 1.23(d)-(f) suggests that as few as three double-spike initializations (e.g., "left", "center", and "right") might be sufficient for repeated attempts of a successful initialization.

Attraction by Saddles. Unfortunately, location within a good region of convergence is not the only requirement of an ideal CMA initialization. As it turns out, initialization can have a profound effect on time-to-convergence. The potential for arbitrarily slow convergence can be explained by the existence of saddle-points on the CM cost surface, that is, points with zero gradient that are concave along certain directions and convex along others [Johnson IJACSP 95]. The convergence-slowing mechanism of saddle points can be explained as follows: concavity attracts trajectories from particular directions, lack of local gradient stalls the trajectories in the vicinity of the saddles, and convexity eventually ejects the trajectories in other directions[21]. Figure 1.24 shows examples of this behavior for different two-tap FSE initializations using the noiseless $T/2$-spaced channel response $(0.2, 0.5, 1, -0.1)^T$ and a BPSK source. Note that in this two-dimensional example, the CMA trajectories are only capable of passing through one saddle before converging to a stable minimum. Figure 1.25 plots the CM-cost evolution for a higher-dimensional case in which the CMA trajectory passes near multiple saddles [Lambotharan SP 97]. Here the $T/2$-spaced channel was $\mathbf{h} = (0.7571, -0.2175, 0.1010, 0.4185, 0.4038, 0.1762)^T$ and the four-tap FSE was double-spike initialized at $\mathbf{f} = (1, 1, 0, 0)^T/\sqrt{2}$. Notice that even the popular double-spike initialization is susceptible to attraction by saddles.

Indirect Initialization Techniques. If needed, the information gained from convergence to one CMA solution can be used to better initialize a second attempt. For instance, [Tong SPL 97] proposes an iterative scheme whereby a delay-shifted version of a channel estimate derived from one CMA-derived equalizer solution is used to compute a new CMA initialization. The hope is that the new trajectory converges to a

[21] Technically speaking, the directions of attraction correspond to the eigenvectors of the Hessian associated with positive eigenvalues, while the directions of repulsion correspond to the eigenvectors associated with negative eigenvalues.

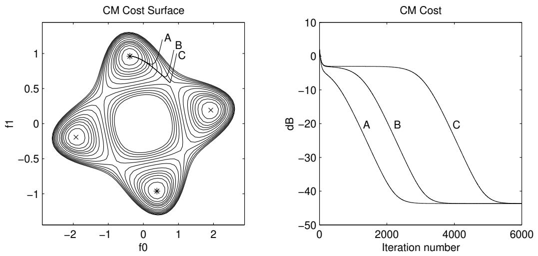

Figure 1.24: Two-tap FSE-CMA trajectories and CM-cost history for several initializations showing the effect of a saddle point on convergence speed.

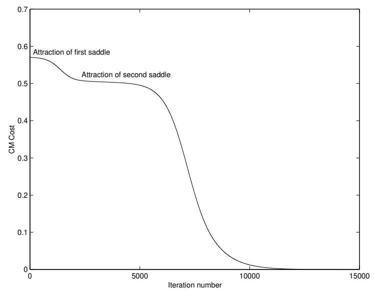

Figure 1.25: Slow convergence of a double-spike-initialized 4-tap FSE due to multiple saddle point attractions.

different (and perhaps better) CMA minimum. This process can be repeated until the performance of all CMA minima have been investigated or until satisfactory performance has been obtained.

Algorithm Step-size

CMA step-size plays a critical role with both time-varying and time-invariant channels. As is the case with LMS (discussed in Section 1.2), increasing the CMA step-size has the advantage of faster convergence but the disadvantage of increased steady-state error. As discussed in Section 1.3, this asymptotic performance may be quantified in terms of either excess CM-cost or excess MSE. Figure 1.26 depicts, for a range of step-sizes, the convergence time and the averaged steady-state values of CM-cost and MSE. The experiments were conducted using BPSK, the channel $\mathbf{h} = (0.2, 0.5, 1.0, -0.1)^T$ at 20 dB SNR, and a two-tap FSE initialized at $(1,1)^T$. Evident is the fundamental tradeoff between convergence rate and excess steady-state error for a time-invariant channel, as discussed in Section 1.3.

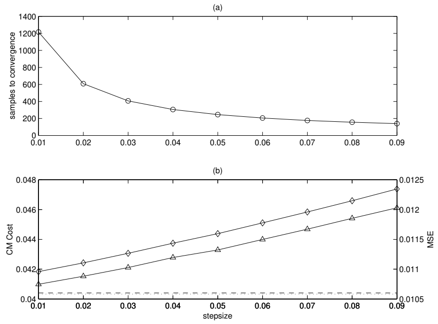

Figure 1.26: Step-size dependence on (a) number of samples required for CMA convergence to steady-state from a particular initialization, and (b) the resulting steady-state CM-cost (◇) and MSE (△). Dashed line indicates (locally) minimum CM-cost, while dotted line indicates (locally) minimum MSE.

When the channel response is time varying, we would like the equalizer to adapt quickly enough to track the channel variations. The amount by which the adapting equalizer lags behind the optimal solution leads to a proportional increase in excess CM-cost (and, we expect, in excess MSE). As a larger step-size leads to faster adaptation, we expect that it decreases the lag contribution to CM-cost. However, we have also seen that a larger step-size increases the excess CM-cost resulting from a non-vanishing parameter update term (i.e., "adaptation noise"). Hence, we anticipate a compromise between the effects of tracking lag and gradient noise.

In practice, the guidelines for CMA step-size may be obtained by slight modification of LMS step-size guidelines. First, CMA convergence rate in the vicinity of a "good" local minimum is known to be well-characterized by quadratic convergence models and be $\kappa_g - \kappa_s$ times as fast as LMS convergence rate (recall the material presented in Section 1.3). Note that, for uniformly-distributed PAM and QAM constellations, we may show $0.6 < \kappa_g - \kappa_s \leq 2$. Second, the EMSE of CMA and LMS differ (recall (1.63) and (1.102)). Though both are linearly proportional to step-size, equalizer length, and received signal power, a non-CM source will increase CMA's EMSE significantly above that of LMS (all else being equal). However, recalling that the goal of CMA is typically the achievement of MSE below the DD transfer level (see Section 1.4), particularly low values of steady-state EMSE might not be necessary. We conclude that CMA's step-size may often be chosen within a factor of two of the recommended LMS step-size (e.g., the values in [Treichler SPM 96]) so long as the anticipated value of the (source-dependent) EMSE is deemed adequate.

In order to demonstrate the step-size tradeoff for FSE-CMA, we use a random walk channel with impulse response $\mathbf{h}(k+1) = \mathbf{h}(k) + \mathbf{n}(k)$ at time $\frac{kT}{2}$. The vector process $\{\mathbf{n}(k)\}$ is composed of zero-mean uncorrelated Gaussian elements with variance 0.0005. With an additive channel noise of $\sigma_w^2 = 0.05$, averaged steady-state CM-cost is plotted in Fig. 1.27 for various values of step-size.

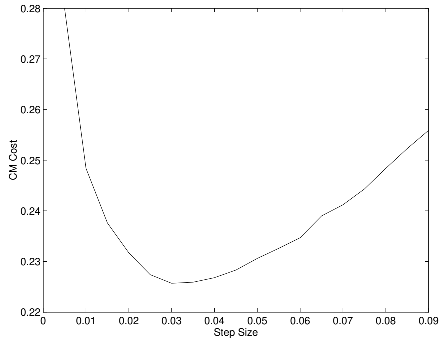

Figure 1.27: Averaged steady-state CM-cost versus step-size for random-walk channel $\mathbf{h}(k)$.

Equalizer Length

Equalizer length is one of the most important variables in communication receiver system design. It is directly related to both optimum receiver performance and implementation cost. For example, approximately 80% of Applied Signal Technology's VLSI chip described in [Treichler SPM 96] is devoted to the FSE.

A common myth regarding equalizer length selection is that a longer FSE always achieves better MSE performance than a shorter FSE. This principle is false for equalizers adapted by a stochastic gradient descent algorithm such as CMA. In the presence of channel noise, a longer FSE is better at mitigating the effects of ISI and noise than a shorter FSE, but a longer FSE will incur the penalty of additional MSE due to increased adaptation noise [Widrow PROC 76]. Although CMA and LMS both exhibit this undesirable excess MSE (EMSE) due to adaptation noise (see Section 1.2), CMA suffers additional EMSE when used with a non-CM source. (Recall from Section 1.3 that a non-CM source makes it impossible to achieve zero CM cost, and it is this non-zero CM cost that causes the CMA-adapted FSE coefficients to "rattle around" the locally-optimal CM solution when driven by a non-vanishing step-size.) The resulting EMSE under the perfect blind equalization conditions was approximated in (1.102) and, as with LMS in (1.63), is proportional to both step-size and FSE length. Herein lies the classical compromise: the FSE length should be long enough to mitigate channel-induced noise and ISI but short enough to prevent the MSE due to adaptation noise from dominating.

Combining the approximate effects of equalizer length and source kurtosis on MSE from [Endres ICASSP 97] and (1.102), respectively, Fig. 1.28 presents steady-state MSE performance of CMA for various equalizer lengths and step-sizes. The calculated data assumes a 16-QAM source over SPIB Microwave Channel 2 in the absence of noise. Our goal for the blind start-up mode is to choose a length/step-size combination capable of achieving a MSE below the decision-directed (DD) transfer level. The dashed line in Fig. 1.28 indicates this transfer level for 16-QAM. (Choosing the transfer level will be the subject of the next section.) Figure 1.28 demonstrates that a longer equalizer does not necessarily provide better steady-state performance than a shorter one. In fact, with certain channel/step-size combinations, an FSE-length *less* than that required for perfect equalization may be required for CMA to reach the DD transfer level. A similar observation was made for BSE-CMA in [Li TSP 96c].

So far we have focused on choosing equalizer length suited to the requirements of CMA in the blind start-up mode. The behavior and requirements of LMS-adapted DD equalization, however, may lead to a slightly different equalizer length tradeoff. For example, we have already noticed that LMS equalizers are not affected by non-CM sources in the way that CMA equalizers are. In addition, the goal of the DD equalization mode may be to attain a symbol error rate (SER) of $\leq 10^{-6}$, while the goal of CMA equalization might only be to attain SER $< 10^{-2}$. The differences between the start-up and DD modes suggest that it may be useful to allow different equalizer lengths for the two modes. In general, the DD-adapted equalizer should be longer than the CMA-adapted start-up equalizer.

There are two common methods for determining a reasonable guess at FSE length [Treichler PROC 98]. If representative signal snapshots are available, one approach is to run simulations of various length equalizers on these signal snapshots in both CMA and DD modes to determine an optimal equalizer length. In the absence of signal snapshots, a good rule-of-thumb is to choose an FSE length of twice the channel delay

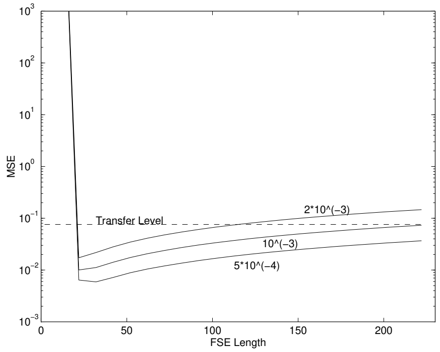

Figure 1.28: MSE performance of CMA as a function of FSE length for various step-sizes: SPIB Microwave Channel 2, 16-QAM, no noise.

spread, measuring delay spread by the duration of the channel response region containing 99.9% of the energy in the total channel response [Treichler SPM 96]. This rule of thumb generally applies to communication systems with moderate complexity constellations (e.g., 64-QAM). Higher complexity constellations may require a longer equalizer to achieve reasonable performance.

DD Transfer Level

Used for blind start-up, the role of CMA becomes one of reducing symbol-error rate (SER) to a level sufficient for successful decision-directed (DD) equalization. Practitioners usually consider SER between 10^{-1} and 10^{-2} as reasonable for initiating DD equalization. As we often measure CMA performance in terms of MSE, we would like a means of translating the SER requirement to a MSE requirement.

In general, the exact relationship between SER and MSE is not simple to describe or compute since it depends on the pdf of the interference [Gitlin Book 92]. One approximation considers the ISI as an additional source of additive white Gaussian noise. Though not a good assumption in general, it is adequate for obtaining design guidelines based on DD transfer level. Recall the expression for the steady-state MSE of LMS given in (1.62). If the signal-to-interference (SIR) ratio is given as $\Upsilon = \sigma_s^2 / \mathrm{E}\{|e_n|^2\}$, then under the approximation of Gaussian interference the SER may be expressed as

$$SER = 1 - \left[1 - \left(1 - \frac{1}{\sqrt{M}}\right) \operatorname{erfc}\left(\sqrt{\frac{3\Upsilon}{2(M-1)}}\right)\right]^2 \qquad (1.103)$$

for an M-QAM system and as

$$SER = \frac{(M-1)}{M}\mathrm{erfc}\left(\sqrt{\frac{3\Upsilon}{M^2-1}}\right) \tag{1.104}$$

for an M-PAM system. Figure 1.29 illustrates the SER versus SIR approximation for various source alphabets.

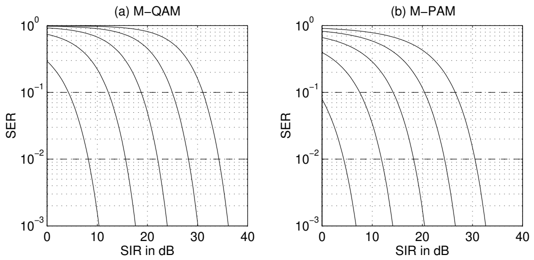

Figure 1.29: Approximate SER versus SIR for (a) M-QAM alphabets, $M = \{4, 16, 64, 256, 1024\}$ from left to right, and (b) M-PAM alphabets, $M = \{2, 4, 8, 16, 32\}$ from left to right.

Conclusions

It should be clear from the examples and analysis of this section that the design choices for initialization, step-size, and equalizer length all impact receiver performance. Most of these design choices rely on knowledge of the prototypical channel impulse response, or at least a rough notion of the channel class.

1.5 Case Studies

To illustrate design methodology, we consider three actual communication scenarios addressing different classes of channel impairments to digital transmissions. The first, the *voiceband telephone channel*, typically has a high SNR and is dominated by transmission line effects and passive audio shaping. The second example involves the *coaxial cable channel* carrying residential TV service. Over the normal channel allocation, this broadband medium suffers both from noise/interference effects and dispersion due to analog mismatching. Lastly, we consider the terrestrial *microwave radio channel*, with its dynamic reflective multi-path environment.

For each case we examine the specification of equalizer settings, such as step-size, FSE-length and initialization, from the varying degrees of a priori information available. Then, using actual received signal snapshots, we examine CMA's blind equalization performance in each case. Use of experimentally acquired received data from actual communications channels also requires us to consider other issues such as carrier frequency offset, noise, and channel time-variations. The effects of carrier frequency offset, to which CMA is immune because of its independence of phase, will be illustrated in the voiceband modem example. The digital cable television example will illustrate the effects of noise on blind equalization of large signal constellations. The last example, mobile microwave communication, is characterized by a time-varying channel. We will show how the three signal environments impact our equalizer parameter selection and equalizer performance.

Good design practice requires us to define the equalizer using any information known about the particular signal environment. For example, for a particular communications link, we may be able to characterize a suitable range of equalizer lengths, communication channel delays, degree of time variation in the channel, and range of SNR. To minimize hardware complexity and maximize agility, we would like to simultaneously choose the shortest equalizer length, largest step-size, and most broadly successful initialization. While our objective is to rapidly reduce the ISI to allow carrier tracking and switching to decision-directed equalization in the least amount of time, practically we try to select the parameters to provide a solution allowing steady-state decision-directed equalization within the constraints of the software and/or hardware being used to process the data.

In the following examples, all of the channels are derived from empirical measurements by Applied Signal Technology, Inc., and reside in the SPIB database at `http://spib.rice.edu/spib/select_comm.html`.

Voiceband Modem: The Dial-up Telephone Channel

We first consider a class of channels that exhibits a wide range of challenging characteristics. Today, data transmission over a dial-up telephone channel is taken for granted as an integral part of personal telecommunications, enabling, for example, the individual's access to the Internet as well as widespread use of dialup facsimile machines. The channel effects in such a transmission medium are due mainly to the analog "local subscriber loop" originally engineered for speech service. This passive connection runs from the subscriber's premises to the telephone service provider's local office facility. The signal must be carried over this copper line, up to 5 km in length, subject not only to passive transmission line dispersion, but also to a variety of non-ideal effects such as spliced wire gauges, non-terminated shunts, and dielectric variations. Once connected to the public switched telephone network (PSTN), the signal occupies the voice band in the range from 300 to 3400 Hz, with band-limiting sufficient to support the 8-kHz sampling operation used for TDM in the digital switching and transmission system. The channel typically exhibits phase irregularities near band edges, with a sloping or rippling magnitude loss on the upper half of the passband; Fig. 1.30(a) shows the magnitude response for a typical voiceband channel.

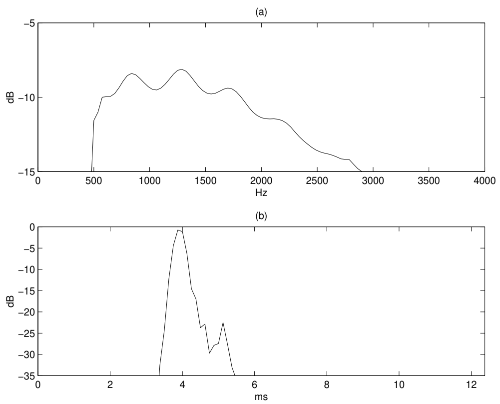

Figure 1.30: Prototypical voiceband channel: (a) frequency response magnitude, (b) impulse response magnitude.

Assuring reliable transmission of data under such conditions has proven very challenging and has motivated some of the earliest endeavors in adaptive equalization and coded modulation. The nature of the basic channel limits symbol rate to on the order of 3400 symbols/sec, with modulating carrier frequency between 1700 and 1900 Hz. Furthermore, the complexity of QAM modulation, hence the actual data rate, is limited mainly by severe ISI rather than additive noise, since the SNR typically exceeds 30 dB. Because the channel characteristics may vary widely from connection to connection, equalization of so-called "dial-up" service by the receiving terminal must be highly robust. This often involves automatic training options that not only initialize equalizer parameters but also allow adjustment of modulation symbol rate and constellation

complexity when conditions dictate.

Currently available voice channel modems rely on training at the start of a connection. For many years, the premier modulation technique for voice channel service was V.29, formalized in a published recommendation [ITU V29 88] by the International Telecommunication Union (ITU). It uses a differentially-encoded QAM constellation transmitted at 2400 symbols/sec. At this highest rate, the 16-point constellation depicted in Fig. 1.31 supports a 9600 bps data rate. To assure reliability, the V.29 specification details a multi-tiered setup sequence, including timing synchronization and training of an MSE-minimizing adaptive equalizer. When adequate performance is achieved, the receiver switches to its decision-directed (DD) mode, and the information transfer portion of the session begins. Upon failure, the procedure calls for repeated attempts at "fall-back" modulation modes of 7200 bps and 4800 bps.

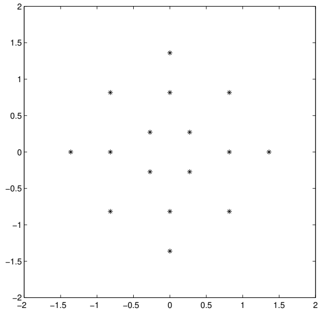

Figure 1.31: V.29 constellation.

When originally defined, the V.29 Recommendation was intended for 4-wire or leased service, where high-quality lines are selected and specially conditioned for long-term point-to-point connections. Over the last twenty years, however, the quality of switched telephone connections has improved to the point that the majority of subscriber loops can now support the full 9600 bps rate without special conditioning. This improvement in channel capacity has fostered an evolution toward higher point-to-point service data rates using a variety of innovative coding and modulation techniques.

While the present point-to-point mode of operation is adequately supported using trained MMSE equalization, future flexibility in telephone connectivity promises to support other network configurations. For example, in a multi-point or broadcast connection, users may connect to an ongoing session through a broadcast hub. Ideally, each user's terminal will establish receiver lock autonomously so as not to unduly impact the overall data rate (as would occur with the inclusion of training).

Below we consider the design implications for a blind equalizer suitable for this type of broadcast connection over existing telephone transmission facilities. The goal for such a blind equalizer is to reduce constellation dispersion to the point where a decision-directed scheme can reliably converge to the MMSE equalizer solution. Note that, while we illustrate using the V.29 modulation format, this by no means represents a limit on the data rate.

The typical voice channel has the frequency response magnitude shown in Fig. 1.30(a) and has the impulse response magnitude seen in Fig. 1.30(b), where the bulk of the energy is confined to a time duration on the order of 2-3 ms. For our design, we recall the equalizer length discussion in Section 1.4 and select an equalizer duration of 30 $T/2$-spaced taps (spanning 6.25 ms, or about twice the channel spread). For adequate agility, we specify a step-size $\mu = 0.001$, providing a basic response time in the millisecond range.

The performance of this prototype design can be demonstrated using a record of actual V.29 signal captured at a subscriber's premises (as found in the SPIB database). The waveform is quadrature down-converted by the known carrier frequency of 1700 Hz and sampled at a $T/2$-spaced rate of 4800 samples/sec. The resulting complex baseband samples are plotted in Fig. 1.32, where the indistinct cloud indicates complete closure of the eye of the data waveform. Following convergence of the blind equalizer, the samples take the form shown in Fig. 1.33, clearly exhibiting clustering consistent with the V.29 constellation. At this point, a phase-tracking loop is used to estimate bulk phase error and properly align the constellation with the decision regions. Figure 1.34 shows the end result following switch-over to the decision-directed mode.

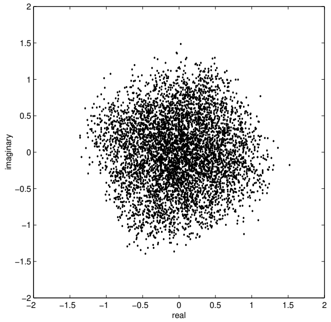

Figure 1.32: $T/2$-spaced input in V.29 demodulation example.

It is interesting to examine Fig. 1.33, the constellation following blind equalization, more closely. The elongated shape of the individual cluster points is symptomatic of "phase roll" due to error in the estimation of carrier frequency. In this case, the accumulated phase change amounts to only about 20° over the observation window. In many cases the carrier offset may be several Hz (e.g., in analog FDM transmission), resulting in a spinning of the equalizer output constellation, as illustrated in Fig. 1.35. Because its adaptation

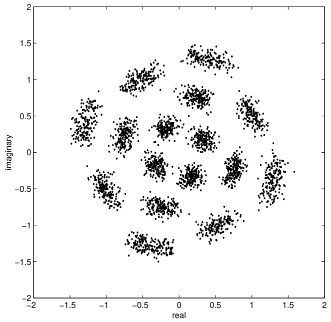

Figure 1.33: Output of CMA-adapted equalizer in V.29 demodulation example.

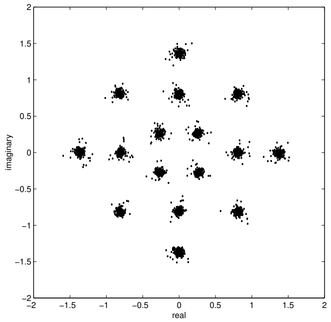

Figure 1.34: Output after DD-equalization and carrier tracking in V.29 demodulation example.

depends only on modulus error, CMA-based blind equalization is largely immune to such phase drift and frequency error. By decoupling mitigation of channel effects from carrier phase tracking circuitry in this manner, accurate carrier phase recovery is greatly simplified.

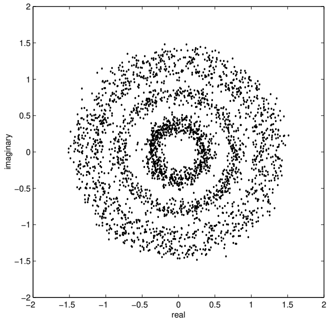

Figure 1.35: Output of CMA-adapted equalizer under 2 Hz of carrier frequency offset in V.29 demodulation example.

The specifics of this modem design are based on dispersion exhibited by a typical voice channel and serve as a feasibility baseline for the extension of traditional voice channel technology. The benefits possible through blind recovery enable modem usage in the broadcast environment of the future.

Digital TV: The Broadband Coaxial Cable Channel

Next, we focus on the use of coaxial cable as a transmission medium for broadcast distribution of digital television signals for residential coverage. Cable service providers operate regional "head end" distribution centers, where national and international feeds are combined with local programming into a composite signal and conditioned for transmission to local subscribers. Much of the physical cable currently in place was originally intended for analog transmission, that is, 6-MHz TV signals frequency-multiplexed to channel assignments of up to 700 MHz. Relative to broadcast transmission, distribution via cable provides a potentially higher-quality signal, unaffected by atmospheric attenuation, RF interference and multipath, and where frequency-dependent losses are normally small throughout any given 6-MHz band. Practical cable systems, however, do suffer from non-ideal effects related to reflective mismatched terminations, imperfect isolation, and crosstalk. In addition, shielding breaks or weak connections may allow so-called "ingress" interference to enter the band (from atmospheric EMI and/or RF broadcast signals).

To provide better signal quality while improving spectral efficiency, digital transmission of TV signals

offers an attractive replacement of traditional analog VSB modulation, as depicted in Fig. 1.36. Drawing on high-speed compression technology, image signals as well as other innovative services can be digitized and multiplexed into a single high-rate bit-stream, then modulated to a standard 6-MHz channel allocation. With QAM techniques, this can provide a six-to-one improvement in bandwidth efficiency, suitable for use with older coaxial systems as well as newer fiber systems.

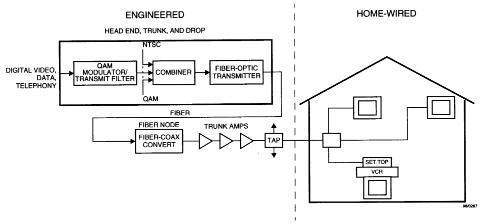

Figure 1.36: Digital cable television distribution.

This type of digital broadcast dictates that blind equalization be a vital component of the subscriber's set-top receiver. The alternative, a conventional trained equalizer, requires valuable bandwidth to provide a re-training signal sufficiently often to facilitate appropriately rapid acquisition time. For example, an adequate response for the "channel surfing" subscriber might call for acquisition within a few hundred milliseconds of a tuning operation.

Because the blind equalizer allows training on the broadcast data itself, no periodic transmission of a training signal is necessary. The receiver response to re-tuning is then basically limited by CM equalizer agility, which, as noted in Section 1.4, is determined by the choice of step-size and equalizer length. Our goal in this section is to apply the rules for blind FSE design to this class of cable channels.

To characterize the nature of the cable channel, tests were conducted using typical residential segments; models of the dispersive and interference effects were obtained after exciting selected bands with digital test signals. Throughout this discussion, we refer to the cable channel time response provided by the SPIB dataset, entitled `chan1.mat`. Analysis of an actual test signal provides this channel's effective response, as shown in Fig. 1.37. The magnitude of the complex-valued time response is shown in Fig. 1.37(b), revealing modest time dispersion (i.e., micro-reflections) with an essentially symmetric distribution, and with the bulk of the energy limited to around $\pm 0.5\mu s$ about its peak. In Fig. 1.37(a), we see its frequency response behavior: channel passband flat to within a couple of dB, as is typical of properly maintained cable systems.

In the current application, the TV signal is a 64-QAM symbol stream with a 5-MHz symbol rate, thus confined to the required 6-MHz allocation. The $T/2$ sampling rate of 10 MHz implies that the 1 μs dispersion noted in Fig. 1.37(b) spans about 10-tap intervals of the fractionally-spaced equalizer. Due to the nature of

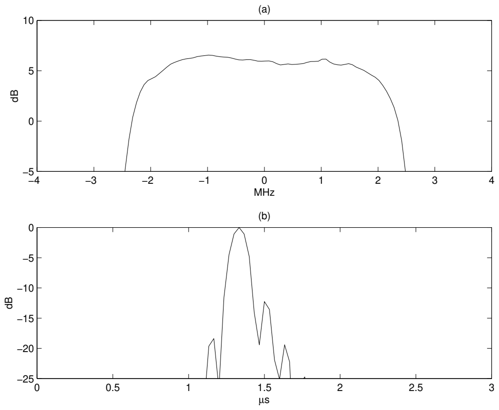

Figure 1.37: Prototypical cable TV channel: (a) frequency response magnitude, (b) impulse response magnitude.

the cable channel, the noise environment can be assumed to be benign with SNR in excess of 30 dB.

Using this scenario as representative for the cable TV environment, we consider the requirements for the receiver's equalizer implementation, that is, adequate asymptotic error performance and sufficient agility. Drawing on the CMA undermodelling analysis and excess MSE analysis presented in Section 1.3, Fig. 1.38 depicts the estimated MSE performance of the CM receiver as a function of equalizer length, parameterized by three values of step-size. Shown as a dashed line is the required MSE for reliable switch-over to DD mode given the 64-QAM signal constellation. A step-size selection of $\mu = 5 \times 10^{-4}$ provides a time constant on the order of a millisecond, presumed adequate in our design. Finally, choosing an equalizer length of 22, we expect a MSE sufficiently low (upon transition to the DD mode) to ensure convergence.

It was noted in Section 1.4 that proper initialization of CMA is essential to assure appropriate convergence. When initializing the FSE coefficients, the position of the impulsive double tap must be specified. Given a typical channel response, selection of the spike location is aided by the MMSE-versus-delay plot that can be calculated for a particular combination of equalizer length and SNR, as shown in Fig. 1.39(b). As discussed in Section 1.4, the double-spike location is closely related to the system delay achieved by CMA (and, likewise, to the choice of training sequence delay in the trained-LMS scenario). For our example, the symmetry of the channel's time response leads to a fairly constant minimum, symmetric about the center, spanning ± 5 baud intervals. The symmetric nature of the typical cable channel indicates a double-spike initialization at taps 11 and 12 should be suitable in most situations.

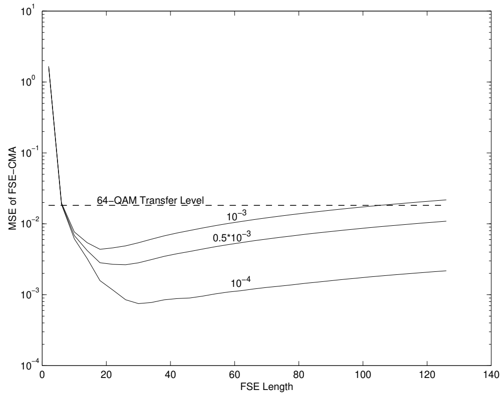

Figure 1.38: Prototypical cable TV channel: MSE due to under-modeling and excess MSE.

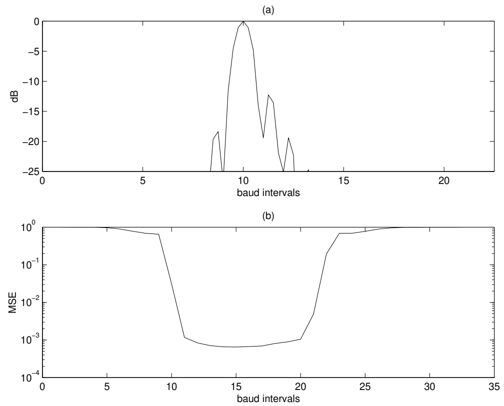

Figure 1.39: Sensitivity of equalizer initialization for a prototypical cable TV channel: (a) channel impulse response magnitude (as in Fig. 1.37) and (b) MMSE versus system delay for 22-tap $T/2$ FSE and 30 dB SNR.

To validate our design choices, Fig. 1.40 shows an ensemble-averaged MSE trajectory characterizing CMA adaptation. Ten sequences of 64-QAM symbols were passed through our typical channel model and the result applied to the CMA-adapted blind FSE. The plot shows transient behavior of the average mean-squared error, where the acceptable "open-eye" condition necessary for switch-over to DD adaptation occurs in about a millisecond.

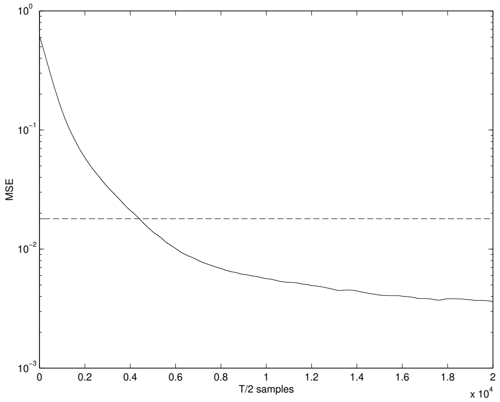

Figure 1.40: The MSE trajectory of CMA as applied to the prototypical cable TV channel.

Based on these observations, we can see that reliable blind reception of digital television signals is possible with reasonably simple and low-cost hardware at the receiver, as is critical for commercial acceptance. Note that our design approach has been purposely conservative in order to accommodate extremes in channel behavior that might be encountered in less ideal situations.

Mobile Radio: The Terrestrial Microwave Channel

In recent years, point-to-point microwave transmission, which traditionally used analog frequency modulation, has been aggressively converted to higher efficiency digital modulation techniques. This transition has been motivated by increased demands for bandwidth, and has been made possible by advances in high-speed processing technology over the last decade. A traditional point-to-point link deployment involves fixed antenna installations with narrow collinear beams, carefully adjusted to minimize reflective interference. This type of system has traditionally served the needs of remote long-haul telephone service as well as intra-urban private voice, video, and data feeds. In this section we consider equalization requirements for digital radio service involving a terrestrial channel.

Introduction of digital modulation in the microwave band has brought new exacting requirements for mitigation of propagation and interference effects. The characteristics of the point-to-point channel tend to be slowly varying, e.g., broad attenuation due to atmospheric and meteorological conditions as well as frequency selective losses resulting from reflective energy. Such multipath reflections result in correlated interference that generates a frequency-dependent passband ripple, that is, alternating constructive and destructive components. The period of the rippling is a function of differential path delay, and the amplitude is dependent on reflective attenuation. Figure 1.41 shows an actual RF channel subject to significant multipath; note how the delayed arrivals visible in the time domain result in a serious rippling of the passband magnitude.

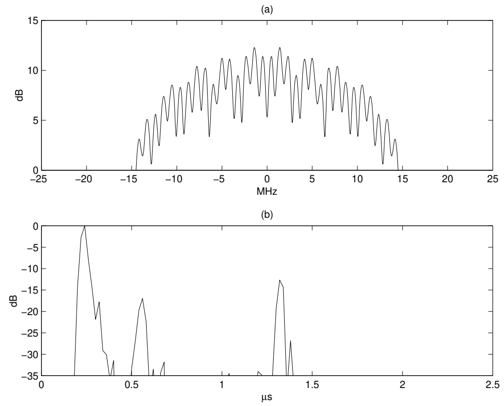

Figure 1.41: Terrestrial microwave channel: (a) frequency response magnitude and (b) impulse response magnitude.

The multipath environment is often a major factor in antenna design and placement, with the goal of minimizing the reflections from buildings and ground surface features illuminated by the main beam. With the introduction of digital modulation schemes using large signal constellations, the use of active digital processing for equalization and polarization restoration has become a necessity. In the point-to-point communication environment, blind adaptation can be performed as a step in outage recovery, with switch-over to decision-directed adaptation conducted during information exchange. Such an implementation is adequate to combat the normal slow channel variations and can often handle faster contributions due to dynamic reflectors, e.g., vehicles and aircraft.

The situation is made considerably more complex with relative motion of receiver and transmitter. In a dynamic scenario, e.g., a mobile network, vehicle motion, coupled with less directive RF transmission, induces variations in the reflective field responsible for interferometric behavior in the channel response (i.e., an in-band "wobbling" fade). This is further complicated by other distributed and motion-induced reflections, e.g., water motion or shimmering foliage. These features call for significant improvement in receiver tracking ability in order to adequately mitigate inter-symbol interference. Furthermore, the nature of the mobile networked environment precludes extensive training. Instead, for the multi-point configuration, it is far preferable that blind adaptation be used to adjust equalizer coefficients, in order to avoid suspension of productive transmission to other receivers in the network.

For this design case we focus on one example of a channel seen by a moving receiver, that is, one node of a networked configuration. The nature of the expected channel then determines the design parameters of the blind equalizer. Specifically, we consider a channel measured as part of an RF survey study Applied Signal Technology conducted in 1996 in rural Northern California [Behm ASIL 97]. A sample dataset from this survey is available from the SPIB database, entitled `hillshadow.otm.4GHz.snp`. Snapshots of a digital transmission at 25 Msymbol/s, down-converted to IF from 4.5 GHz, were taken from a vehicle traveling at modest speed in the vicinity of a fixed transmitter site. By analyzing snapshots taken at 200 ms intervals, the channel characteristics can be estimated using the blind recovery technique of [Gooch ICC 88] and the nature of their variation determined. Figure 1.42 shows that time-variations in the multipath field can be significant, with radical differences exhibited in the impulse response.

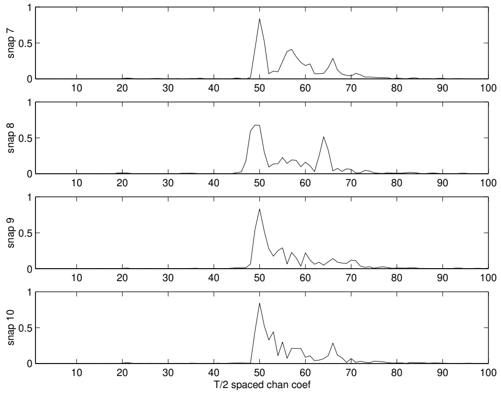

Figure 1.42: Mobile microwave: estimates of channel impulse responses derived from data records separated by 200 ms intervals.

This case represents conflicting design goals: the channel delay spread exhibited in Fig. 1.41(b), typically spanning 10 to 25 symbols (400 to 1000 ns), calls for a relatively long equalizer for adequate ISI recovery, and thus to maintain suitable steady-state performance, the step-size must be kept small. However, the

time-varying nature of the mobile channel requires a step-size large enough to track the typical multipath variations exhibited by Fig. 1.42.

In Sidebar C of [Treichler SPM 96], the minimum recommended $T/2$-spaced equalizer length for the QPSK source used here is 2 taps for each symbol period of delay spread. With a noted maximum delay spread of approximately 25 symbols, and using an additional factor-of-two margin-of-safety, we choose 100 equalizer taps.

From Sidebar D of [Treichler SPM 96], with a unit-variance received signal and an equalizer of up to 100 taps, the recommended upper bound on the step-size (of LMS[22]) is 2 divided by the number of taps, that is, 0.02 in our design. The minimum step-size is related to the variability of the channels of interest. For example, Fig. 1.42 exhibits substantial variation every 200 ms. Presuming that we wish to have a convergence time constant less than 100 times this 0.2 second interval, the average time constant formula for LMS suggests (as discussed in Sidebar D of [Treichler SPM 96]) that for a $T/2$-spaced FSE,

$$\mu > \frac{1}{(\#\text{ of taps}) \cdot (\#\text{ of baud intervals per second}) \cdot (\text{minimum convergence time constant})}.$$

For 25 Mbaud/sec and 100 equalizer taps, this implies that μ should be chosen greater than 2×10^{-5}. Hence, we choose $\mu = 10^{-3}$.

Our choices of length and step-size allowed us to equalize and accurately demodulate the QPSK signal corresponding to the channel shown in Fig. 1.42. This is demonstrated by the equalized constellation in Fig. 1.43 for the source constellation of Fig. 1.44.

This discussion has illustrated use of blind techniques for digital transmission over three very different channels. In each case, the motivation is blind adaptation of the equalizer to avoid disruption of the transmitter's information transfer. And in each case, once adjustment is sufficient to restore constellation integrity, the conventional decision-directed mode is begun. One should note that, for each scenario, future applications involving more elaborate network topologies hinge on the success of blind equalization.

[22] In our application of LMS step-size and length selection guidelines to CMA, we exploit the similarities between LMS and CMA convergence behavior discussed in Section 1.4 (and more technically in Section 1.3). Though CMA and LMS convergence rates may differ by a factor of two or so, we regard such differences to be subsumed by our wide safety margins.

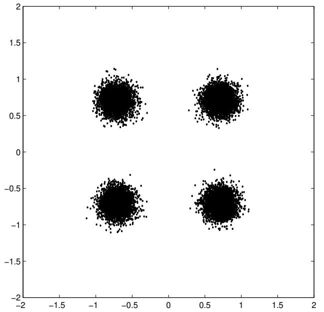

Figure 1.43: Equalized constellation resulting from the design parameters chosen in the mobile radio example, exhibiting a cluster variance of -16.3 dB.

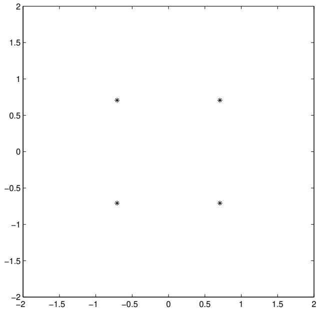

Figure 1.44: QPSK constellation used in the mobile radio example.

1.6 Conclusions

The most basic of adaptive filter configurations in practice and pedagogy is a tapped-delay-line with impulse response coefficients updated via the LMS algorithm in the presence of a replica of the desired filter output sequence (e.g., a training sequence). Thirty years of experience have taught us how to use LMS even though the theory describing its behavior in all pragmatic circumstances is still not complete. We are in an earlier period in a similar pattern of growing utilization and understanding of the constant modulus algorithm (CMA), which has proven to be the baseline algorithm for a variety of applications where the replica of the desired filter output is unavailable, that is, the blind scenario.

The movement from point-to-point, session-oriented communications, for which cooperative equalizer training was acceptable, to digital multi-point and broadcast systems, for which training is unacceptable, mandates blind equalization. Because CMA is the one blind algorithm for adaptive equalization that has seen widespread success in operational communication systems, we have assembled a set of design guidelines for it in this book chapter.

The basic pedagogy employed in this chapter for organizing a behavior theory for a blind fractionally-spaced equalizer is similar to that used in a variety of widely cited textbooks for describing trained, baud-spaced equalizer behavior theory. This approach is bolstered by recent results—described in previous sections of this book chapter—on the similarity of CM and MSE cost function minima. This prompts the adoption of the taxonomy associated with LMS-adapted equalizer design rules. Indeed, the content of this chapter can be characterized in these terms:

- Establish (*as in Sections 1.2 and 1.3*) a relationship between the CM-cost minimizing receiver and the MSE minimizing receiver, and therefore between their respective stochastic gradient descent schemes CMA and LMS.
- Collect (*in Section 1.3*) current understanding of effects of frequency-selective propagation channel response characteristics, source sequence properties, and relative level of additive noise on appropriate adaptive filter length, CMA step-size, and coefficient initialization.
- Extract and illustrate design guidelines with academic examples (*in Section 1.4*) and case studies (*in Section 1.5*).

This book chapter, as a chronicle of the current state-of-the-art in converting an understanding of the behavior of CMA into useful design guidelines, indicates that there is still much more to do, both in terms of fundamental advances in understanding and in terms of a variety of specific details. However, we believe that this chapter—and the almost 20 years of engineering development, practice and basic research in blind equalization via CMA upon which this chapter is based—also establishes CMA as ready for prime time.

Bibliography

[AST Report 96] C. Prettie, R. Yu, J. Trollman, J. Shabel, V. Procopio, and K. Wong, *Results of Laboratory and Field Measurements of Propagation Effects for On-The-Move High Capacity Trunk Radio*, Scientific and Technical Report B001, Applied Signal Technology, Inc., Sunnyvale, CA, Nov. 1996.

[Axford TSP 98] R.A. Axford, Jr., L.B. Milstein, and J.R. Zeidler, "The effects on PN sequences on the misconvergence of the constant modulus algorithm," *IEEE Transactions on Signal Processing*, vol. 46, no. 2, pp. 519-23, Feb. 1998.

[Behm ASIL 97] J.D. Behm, T.J. Endres, P. Schniter, C.R. Johnson, Jr., C. Prettie, M.L. Alberi, and I. Fijalkow, "Characterization of an empirically-derived database of time-varying microwave channel responses," in *Proc. Asilomar Conference on Signals, Systems and Computers* (Pacific Grove, CA), pp. 1549-53, 2-5 Nov. 1997.

[Bellini (Haykin) 94] S. Bellini, "Bussgang techniques for blind deconvolution and equalization," in *Blind Deconvolution*, ed. S. Haykin, Englewood Cliffs, NJ: Prentice-Hall, 1994, pp. 60-120.

[Benveniste Book 90] A. Benveniste, M. M'etivier, and P. Priouret, *Adaptive Algorithms and Stochastic Approximations*, Paris, France: Springer-Verlag, 1990.

[Bussgang MIT 52] J.J. Bussgang, *Crosscorrelation Functions of Amplitude-Distorted Gaussian Signals*, Technical Report 216, MIT Research Laboratory of Electronics, Cambridge, MA, 1952.

[Caines Book 88] P.E. Caines, *Linear Stochastic Systems*, New York, NY: John Wiley & Sons, Inc., 1988.

[Casas DSP 97] R.A. Casas, F. Lopez de Victoria, I. Fijalkow, P. Schniter, T.J. Endres, and C.R. Johnson, Jr., "On MMSE fractionally-spaced equalizer design," in *Proc. International Conference on Digital Signal Processing* (Santorini, Greece), pp. 395-8, 2-4 July 1997.

[Cioffi TCOM 95] J.M. Cioffi, G.P. Dudevoir, M.V. Eyuboglu, and G.D. Forney, Jr., "MMSE decision feedback equalization and coding – Part I and II," *IEEE Transactions on Communications*, vol. 43, no. 10, pp. 2582-604, Oct. 1995.

[Chung ICASSP 98] W. Chung and J.P. LeBlanc, "The local minima of fractionally-spaced CMA blind equalizer cost function in the presence of channel noise," in *Proc. IEEE International Conference on Acoustics, Speech, and Signal Processing* (Seattle, WA), pp. 3345-8, 12-15 May 1998.

[Chung ASIL 98] W. Chung and C.R. Johnson, Jr., "Characterization of the regions of convergence of CMA adaptive blind fractionally spaced equalizers," to appear in *Proc. Asilomar Conference on Signals, Systems and Computers* (Pacific Grove, CA), 2-4 Nov. 1998.

[Ding TCOM 91] Z. Ding, R.A. Kennedy, B.D.O. Anderson, and C.R. Johnson, Jr., "Ill-convergence of Godard blind equalizers in data communication systems," *IEEE Transactions on Communications*, vol. 39, no. 9, pp. 1313-27, Sep. 1991.

[Endres SPAWC 97] T. J. Endres, B. D. O. Anderson, C. R. Johnson, Jr., and M. Green, "On the robustness of the fractionally-spaced constant modulus criterion to channel order undermodeling: Part I," in *Proc. IEEE Signal Processing Workshop on Signal Processing Advances in Wireless Comm.* (Paris, France), pp. 37-40, 16-18 Apr. 1997.

[Endres ICASSP 97] T. J. Endres, B. D. O. Anderson, C. R. Johnson, Jr., and M. Green, "On the robustness of the fractionally-spaced constant modulus criterion to channel order undermodeling: Part II," in *Proc. IEEE International Conference on Acoustics, Speech, and Signal Processing* (Munich, Germany), pp. 3605-8, 20-24 Apr. 1997.

[Ericson TIT 71] T. Ericson, "Structure of optimum receiving filters in data transmission systems," *IEEE Transactions on Information Theory*, vol. 17, no. 3, pp. 352-3, May 1971.

[Fijalkow SPW 94] I. Fijalkow, F. Lopez de Victoria, and C.R. Johnson, Jr., "Adaptive fractionally spaced blind equalization," in *Proc. IEEE Signal Processing Workshop* (Yosemite National Park, CA), pp. 257-60, 2-5 Oct. 1994.

[Fijalkow GRETSI 95] I. Fijalkow, A. Touzni and C.R. Johnson Jr., "Spatio-temporal equalizability under channel noise and loss of disparity," in *Proc. Colloque GRETSI sur le Traitement du Signal et des Images* (Sophia Antipolis, France), pp. 293-6, Sep. 1995.

[Fijalkow SPWSSAP 96] I. Fijalkow, "Multichannel equalization lower bound: A function of channel noise and disparity," in *Proc. IEEE Signal Processing Workshop on Statistical Signal and Array Processing* (Corfu, Greece), pp. 344-7, June 24-26, 1996.

[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 Transactions 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 equalization: Excess MSE," *IEEE Transactions 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.

[Fuhrmann Book 96] P.A. Fuhrmann, *A Polynomial Approach to Linear Algebra*, New York, NY: Springer-Verlag, 1996.

[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 Transactions on Communications*, vol. 28, no. 11, pp. 1867-75, Nov. 1980.

[Gooch ICC 88] R.P. Gooch and J.C. Harp, "Blind channel identification using the constant modulus adaptive algorithm," in *Proc. IEEE International Conference on Communications* (Philadelphia, PA), pp. 75-9, June 12-15, 1988.

[Gray TIT 72] R.M. Gray, "On the asymptotic eigenvalue distribution of Toeplitz matrices," *IEEE Transactions on Information Theory*, vol. 23, pp. 357-68, Nov. 1980.

[Gu CISS 98] M. Gu and L. Tong, "Geometrical characterizations of constant modulus receivers," to appear in *Proc. Conference on Information Science and Systems* (Princeton, NJ), Mar. 1998.

[Haykin Book 94] S. Haykin (Editor), *Blind Deconvolution*, Englewood Cliffs, NJ: Prentice-Hall, 1994.

[Haykin Book 96] S. Haykin, *Adaptive Filter Theory*, 3rd ed., Englewood Cliffs, NJ: Prentice-Hall, 1996.

[ITU V29 88] International Telecommunication Union, *Data communication over the telephone network: 9600 bits per second modem standardized for use on point-to-point 4-wire leased telephone-type circuits*, ITU-T Recommendation V.29, Blue Book, vol. VIII.I, Geneva, Switzerland, 1988.

[Johnson IJACSP 95] C.R. Johnson, Jr. and B.D.O. Anderson, "Godard blind equalizer error surface characteristics: White, zero-mean, binary case," *International Journal of Adaptive Control & Signal Processing*, vol. 9, pp. 301-24, 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. 1998.

[Kailath Book 80] T. Kailath, *Linear Systems*, Englewood Cliffs, NJ: Prentice-Hall, 1980.

[Lambotharan SP 97] S. Lambotharan, J. Chambers, and C.R. Johnson, Jr., "Attraction of saddles and slow convergence in CMA adaptation," *Signal Processing*, vol. 59, no. 2, pp. 335-40, June 1997.

[Larimore ICASSP 83] M.G. Larimore and J.R. Treichler, "Convergence behavior of the constant modulus algorithm," in *Proc. IEEE International Conference on Acoustics, Speech, and Signal Processing* (Boston, MA), pp. 13-16, 14-16 Apr. 1983.

[Larimore ICASSP 97] M.G. Larimore, S.L. Wood, and J.R. Treichler, "Performance costs for theoretical minimum-length equalizers," in *Proc. IEEE International Conference on Acoustics, Speech, and Signal Processing* (Munich, Germany), pp. 2477-80, 21-24 Apr. 1997.

[LeBlanc Thesis 95] J.P. LeBlanc, "Effects of source distributions and correlation on fractionally spaced blind constant modulus algorithm equalizers," *Ph.D. dissertation*, Cornell University, Ithaca, NY, Aug. 1995.

[LeBlanc ICASSP 96] J.P. LeBlanc, I. Fijalkow, and C.R. Johnson Jr., "Fractionally-spaced constant modulus algorithm blind equalizer error surface characterization: Effects of source distributions," in *Proc. IEEE International Conference on Acoustics, Speech, and Signal Processing* (Atlanta, GA), pp. 2944-7, 7-9 May 1996.

[LeBlanc IJACSP 98] J.P. LeBlanc, I. Fijalkow, and C.R. Johnson, Jr., "CMA fractionally spaced equalizers: Stationary points and stability under IID and temporally correlated sources," *International Journal of Adaptive Control & Signal Processing*, vol. 12, no. 2, pp. 135-55, Mar. 1998.

[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 Transactions 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 Transactions 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 Transactions on Signal Processing*, vol. 44, no. 11, pp. 2726-35, Nov. 1996.

[Li TSP 96c] Y. Li and K.J.R. Liu, "Static and dynamic convergence behavior of adaptive blind equalizers," *IEEE Transactions on Signal Processing*, vol. 44, no. 11, pp. 2736-45, Nov. 1996.

[Liu PROC 98] R.W. Liu, editor, *Proceedings of the IEEE: Special Issue on Blind System Identification and Estimation*, Piscataway, NJ: Institute of Electrical and Electronic Engineers, vol. 86, no. 10, Oct. 1998.

[Ljung Book 83] L. Ljung and T. Soderstrom, *Theory and Practice of Recursive Identification*, Cambridge, MA: M.I.T. Press, 1983.

[Lucky BSTJ 66] R.W. Lucky, "Techniques for adaptive equalization of digital communication systems," *Bell System Technical Journal*, vol. 45, pp. 255-86, Feb. 1966.

[Luenberger Book 90] D.G. Luenberger, *Optimization by Vector Space Methods*, 2nd ed., New York, NY: Wiley, 1990.

[Macchi Book 95] O. Macchi, *Adaptive Processing: The Least Mean Squares Approach with Applications in Transmission*, Chichester, NY: Wiley, 1995.

[Moulines TSP 95] E. Moulines, P. Duhamel, J. Cardoso, and S. Mayrargue, "Subspace methods for blind identification of multichannel FIR filters," *IEEE Transactions on Signal Processing*, vol. 43, no. 2, pp. 516-25, Feb. 1995.

[Papadias ICASSP 97] C.B. Papadias, "On the existence of undesired global minima of Godard equalizers," in *Proc. IEEE International Conference on Acoustics, Speech, and Signal Processing* (Munich, Germany), pp. 3937-40, 20-24 Apr. 1997.

[Paulraj SPM 97] A.J. Paulraj and C.B. Papadias, "Space-time processing for wireless communications," *IEEE Signal Processing Magazine*, vol. 14, no. 6, pp. 49-83, Nov. 1997.

[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.

[Rosenblatt Book 85] M. Rosenblatt, *Stationary Sequences and Random Fields*, Boston, MA: Birkhäuser, 1985.

[Strang Book 88] G. Strang, *Linear Algebra and its Applications*, 3rd ed., Fort Worth, TX: Harcourt Brace Jovanovich, 1988.

[Tong TIT 95] L. Tong, G. Xu, B. Hassibi, and T. Kailath, "Blind identification and equalization based on second-order statistics: A frequency-domain approach," *IEEE Transactions on Information Theory*, vol. 41, no. 1, pp. 329-34, Jan. 1995.

[Tong SPL 97] L. Tong and H. Zeng, "Channel surfing re-initialization for the constant modulus algorithm," *IEEE Signal Processing Letters*, vol. 4, no. 3, pp. 85-7, Mar. 1997.

[Touzni EUSIPCO 96] A. Touzni and I. Fijalkow, "Does fractionally-spaced CMA converge faster than LMS?," in *Proc. European Signal Processing Conference* (Trieste, Italy), pp. 1227-30, Sep. 1996.

[Treichler TASSP 83] J.R. Treichler and B.G. Agee, "A new approach to multipath correction of constant modulus signals," *IEEE Transactions on Acoustics, Speech, and Signal Processing*, vol. ASSP-31, no. 2, pp. 459-72, Apr. 1983.

[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 IEEEspecial issue on Blind System Identification and Estimation*, vol. 86, no. 10, pp. 1907-26, Oct. 1998.

[Ungerboeck TCOM 76] G. Ungerboeck, "Fractional tap-spacing equalizer and consequences for clock recovery in data modems," *IEEE Transactions on Communications*, vol. 24, no. 8, pp. 856-64, Aug. 1976.

[Widrow PROC 76] B. Widrow, J.M. McCool, M.G. Larimore and C.R. Johnson, Jr., "Stationary and nonstationary learning characteristics of the LMS adaptive filter," *Proceedings of the IEEE*, vol. 64, no. 8, pp. 1151-62, Aug. 1976.

[Widrow Book 85] B. Widrow and S.D. Stearns, *Adaptive Signal Processing*, Englewood Cliffs, NJ: Prentice Hall, Inc., 1985.

[Zeng ASIL 96] H.H. Zeng, L. Tong and C.R. Johnson, Jr., "Behavior of fractionally-spaced constant modulus algorithm: Mean square error, robustness and local minima," in *Proc. Asilomar Conference on Signals, Systems and Computers* (Pacific Grove, CA), pp. 305-9, Nov. 1996.

[Zeng TIT 98] H.H. Zeng, L. Tong and C.R. Johnson, Jr., "Relationships between the constant modulus and Wiener receivers," *IEEE Transactions on Information Theory*, vol. 44, no. 4, pp. 1523-38, July 1998.

[Zeng ICC 96] S. Zeng, H.H. Zeng, and L. Tong, "Blind equalization using CMA: Performance analysis and a new algorithm," in *Proc. IEEE International Conference on Communications* (Dallas, TX), pp. 847-51, June 1996.