Blind Equalization Using the Constant Modulus Criterion: A Review

C. Richard Johnson, Jr., Philip Schniter, Thomas J. Endres, James D. Behm, Donald R. Brown, and Raúl A. Casas

Abstract

This paper provides a tutorial introduction to the constant modulus (CM) criterion for blind fractionally-spaced equalizer (FSE) design via a (stochastic) gradient descent algorithm such as the Constant Modulus Algorithm. The topical divisions utilized in this tutorial can be used to help catalog the emerging literature on the CM criterion and on the behavior of (stochastic) gradient descent algorithms used to minimize it.

Keywords

The constant modulus algorithm (CMA), blind equalization, adaptive equalizers, blind deconvolution, intersymbol interference, equalizers, digital communication, least mean square methods.

Correspondence:

Rick Johnson, Cornell University, 390 Frank Rhodes Hall, Ithaca, NY, 14853 USA, (607) 254-0429, (607) 255-9072 FAX, *johnson@ee.cornell.edu*

Phil Schniter, Cornell University, 397 Frank Rhodes Hall, Ithaca, NY, 14853 USA, (607) 254-8819, (607) 255-9072 FAX, *schniter@ee.cornell.edu*

Thomas Endres, Sarnoff Digital Communications, Suite 100, 6 Penns Trail, Newtown, PA 18940 USA, (215) 579-8419, (215) 579-8482 FAX, *endres@erols.com*

Jim Behm, 9612 Eagle Ct., Ellicott City, MD 21042 USA, (410)750-7382, *jimbehm@toad.net*

Rick Brown, Cornell University, 397 Frank Rhodes Hall, Ithaca, NY, 14853 USA, (607) 254-8819, (607) 255-9072 FAX, *browndr@ee.cornell.edu*

Raúl Casas, Cornell University, 397 Frank Rhodes Hall, Ithaca, NY, 14853 USA, (607) 254-8819, (607) 255-9072 FAX, *raulc@ee.cornell.edu*

I. INTRODUCTION

Information bearing signals transmitted between remote locations often encounter a signal altering physical channel. Examples of common physical channels include coaxial, fiber optic, or twisted-pair cable in wired communications, and the atmosphere or ocean in wireless communications. Each of these physical channels may cause signal distortion, including echoes and frequency-selective filtering of the transmitted signal. In digital communications, a critical manifestation of distortion is inter-symbol interference (ISI), whereby symbols transmitted before and after a given symbol corrupt the detection of that symbol. All physical channels (at high enough data rates) tend to exhibit ISI. The presence of ISI is readily observable in the sampled impulse response of a channel; an impulse response corresponding to a lack of ISI contains a single spike of width less than the time between symbols. An example of a terrestrial microwave channel impulse response (obtained from the SPIB[1] database) is shown in Figure 1.

Linear channel equalization, an approach commonly used to counter the effects of linear channel distortion, can be viewed as the application of a linear filter (i.e. the equalizer) to the received signal. The equalizer attempts to extract the transmitted symbol sequence by counteracting the effects of ISI, thus improving the probability of correct symbol detection.

Since it is common for the channel characteristics to be unknown (e.g. at startup) or to change over time, the preferred embodiment of the equalizer is a structure adaptive in nature. Classical equalization techniques employ a time-slot (recurring periodically for time-varying situations) during which a training signal, known in advance by the receiver, is transmitted. The receiver adapts the equalizer (e.g. via LMS [*Haykin Book 96*], [*Widrow Book 85*]) so that its output closely matches the known reference training signal. As the inclusion of such signals sacrifices valuable channel capacity, adaptation without resort to training, i.e. *blind* adaptation, is preferred. The most studied and implemented blind adaptation algorithm of the 1990s is the Constant Modulus Algorithm (CMA).
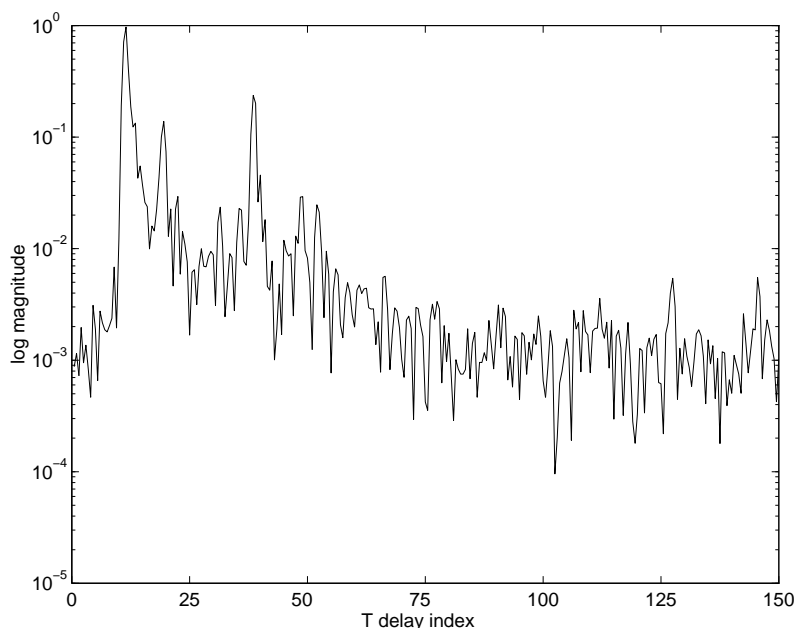

Fig. 1. Terrestrial microwave channel impulse response magnitude, $1/T = 30 \times 10^6$ symbols/sec (SPIB Channel #3).

CMA seeks to minimize a cost defined by the Constant Modulus (CM) criterion. The CM criterion penalizes deviations in the modulus (i.e. magnitude) of the equalized signal away from a fixed value.

[1]The Rice University Signal Processing Information Base (SPIB) microwave channel database resides at `http://spib.rice.edu/spib/microwave.html`

In certain ideal conditions, minimizing the CM cost can be shown to result in perfect (zero-forcing) equalization of the received signal. Remarkably, the CM criterion can successfully equalize signals characterized by source alphabets *not* possessing a constant modulus (e.g. 16-QAM), as well as those possessing a constant modulus (e.g. 8-PSK). (See Figure 2.) This paper attempts to explore the behavior of CMA by consideration of similarities between the CM and mean-squared error (MSE) criteria. This relationship is important because of well-known connections between MSE and the actual quantity we desire to minimize, probability of bit error (see, e.g., the discussion in [*Gitlin Book 92*]).
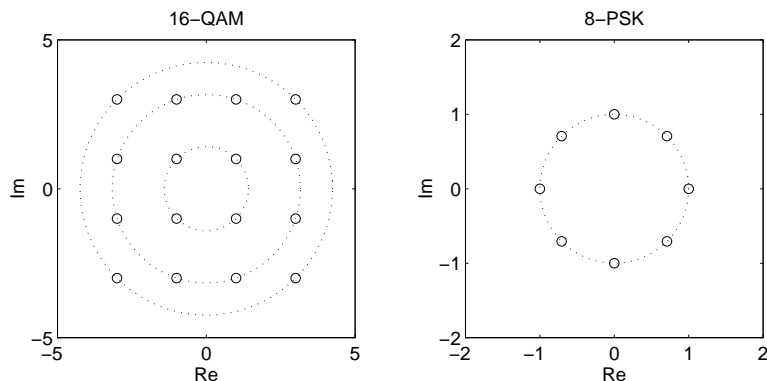

Fig. 2. Non-constant modulus source constellation (16-QAM) versus constant modulus source constellation (8-PSK).

Plotting the CM cost versus the equalizer coefficients results in a surface referred to as the CM cost surface. Stochastic gradient descent algorithms (SGD) [*Luenberger Book 90*], [*Haykin Book 96*] attempt to minimize the CM cost by starting at some location on the surface and following the trajectory of steepest descent. The CM cost surface characteristics are important because they can be used to understand the behavior of any SGD attempting to minimize the CM cost, such as CMA. Specifically, these characteristics lend insight into the channel, equalizer, and source properties which affect SGD behavior.

The success of a stochastic gradient descent equalizer adaptation algorithm is dependent on a certain amount of stationarity in the received process. Thus, throughout the paper, we restrict our focus to stationary source and noise processes, and to channels whose impulse response is fixed or slowly[2] time-varying.

History

In the literature, blind equalization algorithms blossomed in the 1980s. The two principal precursors are Lucky's blind decision-direction algorithm [*Lucky BSTJ 66*] and Sato's algorithm [*Sato TCOM 75*]. What we term the CM criterion was introduced for blind equalization of QAM signals in [Godard TCOM 80] and of PAM and FM signals in [Treichler TASSP 83]. By the end of the 1980s blind equalizers were commercialized for microwave radio [*Larimore ASIL 85*]. By the mid 1990s, blind equalizers were realized in VLSI for HDTV set-top cable demodulators [*Treichler SPM 96*]. The current explosion of interest in the constant modulus (CM) criterion stems from blind processing applications in emerging wireless communication technology (e.g., blind equalization, blind source separation, and blind antenna steering) and from CMA's record of practical success.

Our Mission

This paper is intended to be a resource both to readers experienced in blind equalization as well as those new to the subject. In a tutorial style, Section I-A provides background in fractionally-

---

[2]Here "slow" is considered relative to the tracking speed of the SGD algorithm.

3

spaced equalizer (FSE) modeling and design. (For baud-spaced equalizer (BSE) design, we refer the interested reader to a variety of classical references, e.g. [*Gitlin Book 92*], [*Haykin Book 96*], [*Lee Book 94*], and [*Proakis Book 95*]). Section II then illustrates several low-dimensional examples that help to characterize the behavior of FSEs adapted under the constant modulus criterion.

In Section III, we construct a categorization of literature focusing on the application of the CM criterion to blind equalization. The annotated bibliography in Section V catalogs the existing literature according to the classifications of Section III, providing the reader with a valuable tool for further research. Our attempt to be exhaustive is justified only by the relative infancy of the subfield; evidence of the emerging status of this literature is seen in the wealth of conference papers in the bibliography of Section V.

Following the introductory FSE tutorial, Section I-C presents a novel view of classical non-blind adaptive equalization that illuminates the connection between the MSE and CM criteria. Specifically, the LMS-with-training strategy requires pre-selection of a design variable, namely training sequence delay, that may lead to a potentially suboptimal solution. The delay-optimized MSE, a function of equalizer parameters only, yields a cost surface (see Figure 7) for which a simple LMS-like parameter update algorithm is not known to exist. Remarkably, the CM criterion offers a proxy for this surface for which there exists a (blind) parameter update algorithm, CMA.

A. Fractionally-Spaced Linear Equalization

In this section we describe the fractionally-spaced equalization scenario and present some fundamental results regarding minimum mean squared error (i.e. ~~Wiener [~~*Haykin Book 96*~~]~~) equalizers. This material is primarily intended to provide background and context. For simplicity, our focus is restricted to a $T/2$-spaced FSE, where T denotes the baud, or symbol, duration. All results are extendible to the more general T/N-spaced case. Examples of seminal work on fractionally-spaced equalization include [*Lucky ALL 69*], [*Macchi ANT 75*], [*Ungerboeck TCOM 76*] and [*Gitlin BSTJ 81*], while more comprehensive references are [*Qureshi PROC 85*] and [*Gitlin Book 92*].
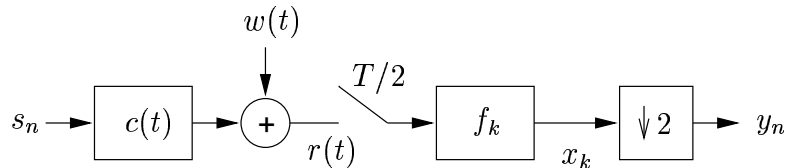

Fig. 3.   Baseband model of single-channel communication system with $T/2$-spaced receiver.
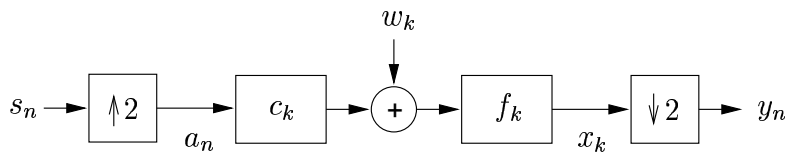
Fig. 4.   Multirate system model.

A.1 Multirate and Multichannel System Models

Consider the single-channel model illustrated in Figure 3. A (possibly complex-valued) T-spaced symbol sequence $\{s_n\}$ is transmitted through a pulse shaping filter, modulated onto a propagation channel, and demodulated. We assume all processing between the transmitter and receiver is linear and time invariant (LTI) and can thus be described by the continuous-time impulse response $c(t)$. The received signal $r(t)$ is also corrupted by additive channel noise, whose baseband equivalent we denote

4

Fig. 5. Multichannel system model.

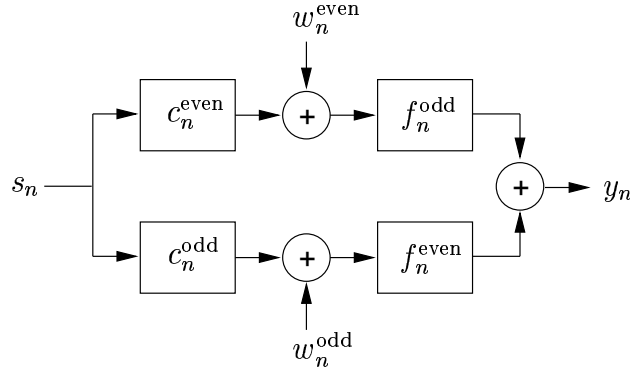by $w(t)$. The received signal is then sampled at $T/2$-spaced intervals and filtered by a $T/2$-spaced finite impulse response (FIR) equalizer of length $2N$. (An even length is chosen for notational simplicity.) This filtering can be regarded as a convolution of the sampled received sequence with the equalizer coefficients f_k. Finally, the FSE output $\{x_k\}$ is decimated by a factor of 2 to create the T-spaced output sequence $\{y_n\}$. Decimation is accomplished by disregarding alternate samples, thus producing the baud-spaced "soft decisions" y_n. We note that, in general, all quantities are complex-valued. For clarity, we reserve the index n for baud-spaced quantities and the index k for fractionally-spaced quantities throughout the paper.

Appendix A derives the equivalence between the continuous-time model in Figure 3 and the discrete time models in Figures 4 and 5, both constructed using $T/2$-spaced samples of $c(t)$ and $w(t)$. Figure 4 depicts the "multirate" model while Figure 5 depicts the "multichannel" model. Though our derivation of the discrete-time models is based on the single-channel system in Fig. 3, the equivalence between the multirate and multichannel models suggests that we could have based our model on a two-sensor (T-sampled) communication system instead. For a concise discussion on the equivalence between temporal and spatial diversity, see [*Moulines TSP 95*].

The multirate model of Figure 4 uses the discrete-time fractionally-spaced channel coefficients $c_k = c(k\frac{T}{2})$ and the discrete-time random process $w_k = w(k\frac{T}{2})$. The multichannel model of Figure 5 subdivides these sample sequences into even and odd baud-spaced counterparts (of relative delay $T/2$), so that $c_n^{\text{even}} = c_{2n}$ and $c_n^{\text{odd}} = c_{2n+1}$ for $n = 0, 1, 2, \ldots$. In a similar manner, the FSE coefficients are partitioned as $f^{\text{even}} = f_{2n}$ and $f^{\text{odd}} = f_{2n+1}$.

Given a fractionally-spaced channel of finite[3] and (even) length $2M$, we can collect the even and odd sets of equalizer and channel coefficients into column vectors

$$
\begin{aligned}
\mathbf{f}_\text{e} &= [f_0, f_2, f_4, \ldots, f_{2N-2}]^t &= [f_0^{\text{even}}, f_1^{\text{even}}, f_2^{\text{even}}, \ldots, f_{N-1}^{\text{even}}]^t, \\
\mathbf{f}_\text{o} &= [f_1, f_3, f_5, \ldots, f_{2N-1}]^t &= [f_0^{\text{odd}}, f_1^{\text{odd}}, f_2^{\text{odd}}, \ldots, f_{N-1}^{\text{odd}}]^t, \\
\mathbf{c}_\text{e} &= [c_0, c_2, c_4, \ldots, c_{2M-2}]^t &= [c_0^{\text{even}}, c_1^{\text{even}}, c_2^{\text{even}}, \ldots, c_{M-1}^{\text{even}}]^t, \\
\mathbf{c}_\text{o} &= [c_1, c_3, c_5, \ldots, c_{2M-1}]^t &= [c_0^{\text{odd}}, c_1^{\text{odd}}, c_2^{\text{odd}}, \ldots, c_{M-1}^{\text{odd}}]^t.
\end{aligned}
\tag{1}
$$

It is possible to form the (baud-spaced) impulse response of the linear system relating s_n to y_n using

---

[3]In practice, we would consider the FS channel to be of "finite length" M if the response magnitude can be said to decay below some sufficiently small threshold for all time $t \geq M\frac{T}{2}$.

a pair of $P \times N$ baud-spaced convolution matrices $\mathbf{C}_\mathrm{e}$ and $\mathbf{C}_\mathrm{o}$, where $P = M + N - 1$.

$$
\mathbf{C}_\mathrm{e} = \begin{bmatrix}
c_0^{\mathrm{even}} & & & \\
c_1^{\mathrm{even}} & c_0^{\mathrm{even}} & & \\
\vdots & c_1^{\mathrm{even}} & & \\
c_{M-1}^{\mathrm{even}} & \vdots & \ddots & c_0^{\mathrm{even}} \\
& c_{M-1}^{\mathrm{even}} & & c_1^{\mathrm{even}} \\
& & & \vdots \\
& & & c_{M-1}^{\mathrm{even}}
\end{bmatrix}, \quad
\mathbf{C}_\mathrm{o} = \begin{bmatrix}
c_0^{\mathrm{odd}} & & & \\
c_1^{\mathrm{odd}} & c_0^{\mathrm{odd}} & & \\
\vdots & c_1^{\mathrm{odd}} & & \\
c_{M-1}^{\mathrm{odd}} & \vdots & \ddots & c_0^{\mathrm{odd}} \\
& c_{M-1}^{\mathrm{odd}} & & c_1^{\mathrm{odd}} \\
& & & \vdots \\
& & & c_{M-1}^{\mathrm{odd}}
\end{bmatrix}.
\tag{2}
$$

Convolution matrices are constructed so that, for example, the vector $\mathbf{C}_\mathrm{o}\mathbf{f}_\mathrm{e}$ is composed of coefficients from the convolution $c_n^{\mathrm{odd}} \star f_n^{\mathrm{even}}$.

Defining the compound matrix and vector quantities

$$
\mathbf{C} = \begin{bmatrix} \mathbf{C}_\mathrm{o} & \mathbf{C}_\mathrm{e} \end{bmatrix}, \quad \mathbf{f} = \begin{bmatrix} \mathbf{f}_\mathrm{e} \\ \mathbf{f}_\mathrm{o} \end{bmatrix}.
\tag{3}
$$

we can rewrite the noise-free multichannel convolution equation (39) compactly in terms of the P (baud-spaced) system impulse response coefficients, $\mathbf{h} = [h_0, h_1, \ldots, h_{P-1}]^t$:

$$
\mathbf{h} = \mathbf{C}\mathbf{f}.
\tag{4}
$$

Equation (4) indicates that $\mathbf{C}$ maps the FSE coefficient vector to its corresponding system response. Note that $\mathbf{C}$ is a Sylvester matrix [*Kailath Book 80*].

A different (though essentially equivalent) construction of $\mathbf{C}$ and $\mathbf{f}$ deserves mention. First consider the fractionally-spaced convolution matrix $\mathbf{C}_\mathrm{FS}$ constructed as in either $\mathbf{C}_\mathrm{e}$ or $\mathbf{C}_\mathrm{o}$ in (2), but from a vector of fractionally-spaced channel coefficients $\mathbf{c}_\mathrm{FS} = [c_0, c_1, c_2, \ldots, c_{2M-1}]^t$:

$$
\mathbf{C}_\mathrm{FS} = \begin{bmatrix}
c_0 & & & \\
c_1 & c_0 & & \\
c_2 & c_1 & \ddots & \\
\vdots & c_2 & \ddots & c_0 \\
c_{2M-1} & \vdots & \ddots & c_1 \\
& c_{2M-1} & & c_2 \\
& & \ddots & \vdots \\
& & & c_{2M-1}
\end{bmatrix}.
\tag{5}
$$

The product of $\mathbf{C}_\mathrm{FS}$ with FSE coefficient vector $\bar{\mathbf{f}} = [f_0, f_1, f_2, \ldots, f_{2N-1}]^t$ yields the fractionally-spaced impulse response between the upsampler and downsampler in Figure 4, i.e. $\mathbf{h}_\mathrm{FS} = \mathbf{C}_\mathrm{FS}\bar{\mathbf{f}}$, just as in (4). (See (47) in Appendix A-B.) Since the baud-spaced impulse response $\mathbf{h}$ is formed using the odd coefficients of $\mathbf{h}_{FS}$, we reason that $\mathbf{h}$ can be constructed from the product of $\bar{\mathbf{f}}$ and a row-decimated

version of $\mathbf{C}_{\mathrm{FS}}$. In other words, $\mathbf{h} = \bar{\mathbf{C}}\bar{\mathbf{f}}$ where $\bar{\mathbf{C}}$ is formed from the odd[4] rows of $\mathbf{C}_{\mathrm{FS}}$:

$$
\bar{\mathbf{C}} \;=\; \begin{bmatrix}
c_1 & c_0 & & & & \\
c_3 & c_2 & c_1 & c_0 & & \\
\vdots & \vdots & c_3 & c_2 & \ddots & \\
c_{2M-1} & c_{2M-2} & \vdots & \vdots & \ddots & c_0 \\
& & c_{2M-1} & c_{2M-2} & & c_2 \\
& & & & \ddots & \vdots \\
& & & & & c_{2M-2}
\end{bmatrix} . \tag{6}
$$

Notice that $\bar{\mathbf{C}}$ is a column reordering of $\mathbf{C}$ and $\bar{\mathbf{f}}$ is a row reordering of $\mathbf{f}$. Thus, we consider the alternate formulation of the "decimated fractionally-spaced convolution matrix" $\bar{\mathbf{C}}$ in (6) as essentially equivalent to $\mathbf{C}$ in (3).

The convention we adopt in constructing $\mathbf{C}$ and $\bar{\mathbf{C}}$, sometimes referred to as "odd-sampled" decimation, connects the odd subchannel output to the even subequalizer input and vice versa (see Figure 5). Appendix A discusses the implications of this choice.

In the baud-spaced equalization context [*Proakis Book 95*], [*Lee Book 94*], the convolution matrix $\mathbf{C}_{\mathrm{BS}}$ relating the equalizer coefficient vector to the baud-spaced impulse response does not have the compound form of (3) or (6). Instead it appears like $\mathbf{C}_{\mathrm{e}}$ (or $\mathbf{C}_{\mathrm{o}}$) in (2), but with columns constructed from the T-spaced samples of the channel response. In the absence of channel noise, this construction of $\mathbf{C}_{\mathrm{BS}}$ yields the BSE design equation

$$
\mathbf{h} \;=\; \mathbf{C}_{\mathrm{BS}}\mathbf{f}_{\mathrm{BS}}, \tag{7}
$$

where $\mathbf{f}_{\mathrm{BS}}$ is the baud-spaced equalizer coefficient vector.

A.2 Requirements for Perfect Source Recovery

Equation (4) leads to what are commonly referred to as the "length and zero" conditions for perfect fractionally-spaced equalization. We use the term perfect equalization interchangeably with perfect source recovery (PSR), i.e., when $y_n = s_{n-\delta}$ for some fixed delay δ and any source sequence $\{s_n\}$. In addition to the absence of noise, PSR requires the "zero-forcing" system impulse response

$$
\mathbf{h}_\delta \;=\; [0 \ldots 0, 1, 0 \ldots 0]^t, \tag{8}
$$

where the nonzero coefficient is in the δth position (and δ must satisfy $0 \le \delta \le P - 1$). This response characterizes a system which merely delays the transmitted symbols by δ baud intervals. In order to achieve this particular response, the system of linear equations described by $\mathbf{h}_\delta = \mathbf{C}\mathbf{f}$ must have a solution. For PSR under arbitrary δ[5], $\mathbf{C}$ must be full row rank [*Tong TIT 95*]. This condition is sometimes referred to as *strong perfect equalization*.

The full-rank requirement implies that $\mathbf{C}$ must have at least as many columns as rows, which, in the $T/2$-spaced case, results in the following equalizer length requirement:

$$
2N \ge M + (N - 1) \;\;\Rightarrow\;\; N \ge M - 1. \tag{9}
$$

---

[4]Throughout, we assume a vector/matrix indexing that starts with zero rather than one, so that the first row is considered "even" and the second "odd."

[5]A necessary and sufficient condition on perfect equalization (in the absence of noise) is that there exist a δ for which $\mathbf{h}_\delta$ lies in the column space of $\mathbf{C}$. Hence, there exist channels that do *not* result in full row-rank convolution matrices but that do satisfy $\mathbf{h}_\delta = \mathbf{C}\mathbf{f}$ for *particular* δ. Though we acknowledge the existence of such channels, we consider them to be trivial in the physical sense.

Applying the same argument to (7) reveals the reason that *no FIR BSE can perfectly equalize a nontrivial FIR channel*: the row dimension of $\mathbf{C}_{\mathrm{BS}}$ always exceeds the column dimension. The $T/2$-spaced full rank requirement also implies that the polynomials specified by the coefficients $\mathbf{c}_{\mathrm{e}}$ and $\mathbf{c}_{\mathrm{o}}$ share no common roots (i.e., the polynomials are coprime). Appendix A-C discusses this common-root condition in more detail.

B. Mean-Square Error Criterion

In the presence of noise, we desire to minimize the expected squared magnitude of the recovery error

$$e_n \;=\; y_n - s_{n-\delta} \tag{10}$$

for a particular choice of delay (δ). We will see that this criterion can be interpreted as the best compromise between inter-symbol interference and noise amplification in a minimum mean-squared error (MMSE) sense.

To formulate this error criterion more precisely, we collect the P previous T-spaced elements of the source sequence into the vector

$$\mathbf{s}(n) \;=\; [s_n,\; s_{n-1},\; s_{n-2}, \ldots, s_{n-(P-1)}]^t, \tag{11}$$

and the last $2N$ fractionally-sampled values of noise into vector $\mathbf{w}(n)$,

$$\mathbf{w}(n) \;=\; [w_{n-1},\; w_{n-3},\; w_{n-5}, \ldots, w_{n-(2N-1)},\; w_n,\; w_{n-2},\; w_{n-4}, \ldots, w_{n-(2N-2)}]^t, \tag{12}$$

where the collection of even noise samples follows the collection of odd noise samples, to be consistent with our definitions of $\mathbf{C}$ and $\mathbf{f}$ in (3). (Note, however, that this particular ordering of samples in the noise vector is inconsequential when assuming an independent identically-distributed (i.i.d.) noise process.) With these quantities, the nth equalizer output, $y_n = y(nT + \frac{T}{2})$, can be written compactly as

$$y_n \;=\; \mathbf{s}^t(n)\mathbf{C}\mathbf{f} + \mathbf{w}^t(n)\mathbf{f}, \tag{13}$$

yielding an expression for the recovery error

$$e_n \;=\; \mathbf{s}^t(n)(\mathbf{C}\mathbf{f} - \mathbf{h}_\delta) + \mathbf{w}^t(n)\mathbf{f}. \tag{14}$$

Under the assumption that the noise and source processes are i.i.d. and jointly uncorrelated, with respective variances σ_w^2 and σ_s^2, the expected value of the magnitude-squared recovery error becomes

$$\mathrm{E}\{|e_n|^2\} \;=\; (\mathbf{C}\mathbf{f} - \mathbf{h}_\delta)^H(\mathbf{C}\mathbf{f} - \mathbf{h}_\delta)\sigma_s^2 + \mathbf{f}^H\mathbf{f}\sigma_w^2. \tag{15}$$

(Appendix A-D discusses the independence assumption regarding fractionally-sampled channel noise.) Note that (15) is proportional to the source-power-normalized MSE cost function:

$$J_{\mathrm{MSE}} \;=\; (\mathbf{C}\mathbf{f} - \mathbf{h}_\delta)^H(\mathbf{C}\mathbf{f} - \mathbf{h}_\delta) + \lambda\mathbf{f}^H\mathbf{f}, \tag{16}$$

where $\lambda = \sigma_w^2/\sigma_s^2$. In terms of $\mathbf{A} = \mathbf{C}^H\mathbf{C} + \lambda\mathbf{I}$, the technique of "completing the square" yields

$$J_{\mathrm{MSE}} \;=\; (\mathbf{f} - \mathbf{A}^{-1}\mathbf{C}^H\mathbf{h}_\delta)^H\mathbf{A}(\mathbf{f} - \mathbf{A}^{-1}\mathbf{C}^H\mathbf{h}_\delta) - \mathbf{h}_\delta^H\mathbf{C}\mathbf{A}^{-1}\mathbf{C}^H\mathbf{h}_\delta + \mathbf{h}_\delta^H\mathbf{h}_\delta. \tag{17}$$

Note that $\mathbf{A}$ is positive definite for $\lambda > 0$.

Equation (17) indicates that the equalizer parameter vector minimizing J_{MSE} is

$$\mathbf{f}^\dagger \;=\; \mathbf{A}^{-1}\mathbf{C}^H\mathbf{h}_\delta, \tag{18}$$

and it follows that the **f**-optimal mean squared error

$$\min_{\mathbf{f}} J_{\mathrm{MSE}}(\mathbf{f}, \delta) \;=\; J_{\mathrm{MSE}}(\mathbf{f}^{\dagger}, \delta) \tag{19}$$

$$=\; \mathbf{h}_{\delta}^{H}\left(\mathbf{I} - \mathbf{C}\mathbf{A}^{-1}\mathbf{C}^{H}\right)\mathbf{h}_{\delta} \tag{20}$$

remains a function of system delay δ. We make this property explicit by adopting the notation $J_{\mathrm{MSE}}(\mathbf{f}, \delta)$. It follows from (20) and (8) that the optimum delay $\delta^{\dagger}$ corresponds to the index of the minimum diagonal element of $\mathbf{I} - \mathbf{C}\mathbf{A}^{-1}\mathbf{C}^{H}$ [*Johnson ASIL 95*]. This is written formally below:

$$\delta^{\dagger} \;=\; \arg\min_{\delta}\left\{\left[\mathbf{I} - \mathbf{C}(\mathbf{C}^{H}\mathbf{C} + \lambda\mathbf{I})^{-1}\mathbf{C}^{H}\right]_{\delta,\delta}\right\}. \tag{21}$$

For a $T/2$-spaced FSE with 300 taps and a SNR ($= 1/\lambda$) of 30 dB Figure 6 plots $J_{\mathrm{MSE}}(\mathbf{f}^{\dagger}, \delta)$ versus δ for the "typical" impulse response of Figure 1. Note the degree to which δ can affect MSE performance.
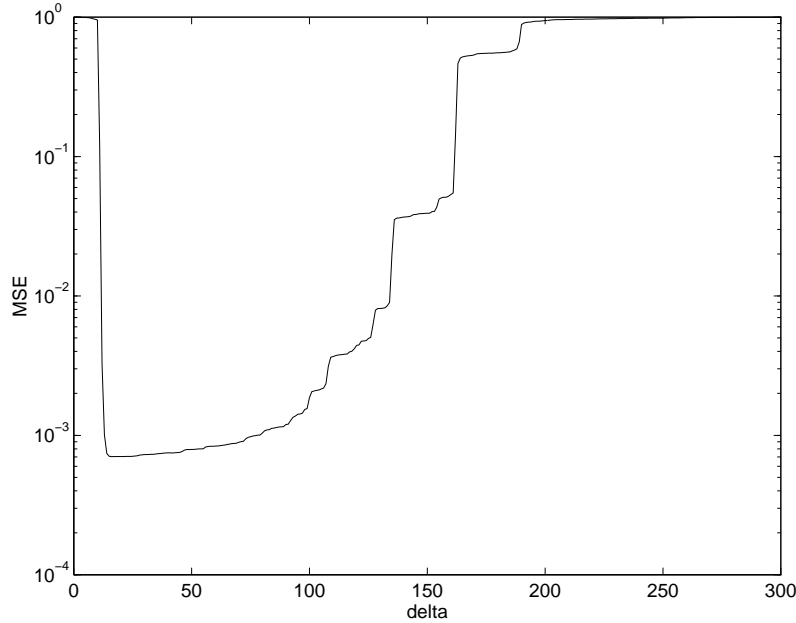

Fig. 6. **f**-optimal MSE, $J_{\mathrm{MSE}}(\mathbf{f}^{\dagger}, \delta)$, versus T-spaced delay δ for the channel of Figure 1 and 30 dB SNR using 300-tap FSE.

We conclude that proper pre-selection of δ is important for equalizer-based minimization of $J_{\mathrm{MSE}}(\mathbf{f}, \delta)$. This idea of fixed-$\delta$ optimization is of particular relevance because it describes the typical adaptive equalization scenario when a training signal is available [*Qureshi TCOM 73*].

C. An Amalgamated MSE Cost Function

When the source is differentially encoded [*Gitlin Book 92*], knowledge of absolute phase is not required for symbol detection. For example, either $y_n = s_{n-\delta}$ or $y_n = -s_{n-\delta}$ (for all n) would form an acceptable output sequence for differentially-encoded BPSK. (For complex-valued source alphabets such as QAM, we allow $y_n = e^{j\frac{\pi}{2}m}s_{n-\delta}$ for fixed $m \in \{0, 1, 2, 3\}$.) Therefore, an acceptable system impulse response can include a fixed phase shift in addition to a bulk system delay δ. With this in mind, we construct a phase- and delay-optimized amalgamated cost function $J_{\mathrm{A}}(\mathbf{f})$:

$$J_{\mathrm{A}}(\mathbf{f}) \;=\; \min_{\delta,\rho}\left\{(\mathbf{C}\mathbf{f} - \rho\mathbf{h}_{\delta})^{H}(\mathbf{C}\mathbf{f} - \rho\mathbf{h}_{\delta}) + \lambda\mathbf{f}^{H}\mathbf{f}\right\}, \tag{22}$$

where ρ is one of the set of allowable phase shifts (e.g., $\{+1, -1\}$ for real-valued PAM).

J_{A} is a multimodal fabrication, bearing similarity to an $(2N+1)$-dimensional egg carton. A surface plot appears in Figure 7 for well-behaved $T/2$-spaced channel $\mathbf{c}_1$ defined in Table I. By "well-behaved" we mean that $\mathbf{c}_1$ has no common or nearly-common subchannel roots. Figure 7 indicates that if we
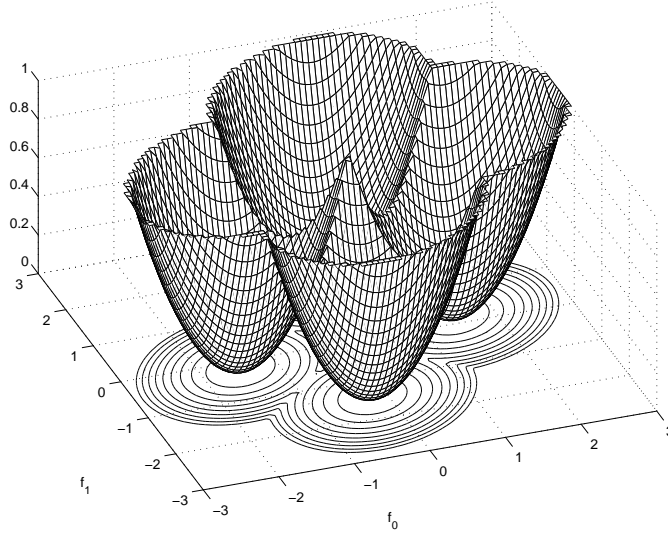

Fig. 7. J_{A} for well-behaved channel $\mathbf{c}_1$ and no noise, in equalizer ($\mathbf{f}$) space.

minimize $J_{\text{A}}(\mathbf{f})$ by a gradient descent strategy, then the initial value of $\mathbf{f}$ will determine the values of δ and ρ to which the descent scheme will asymptotically converge. In other words, optimization of $J_{\text{A}}(\mathbf{f})$ by gradient descent accomplishes pre-selection of δ via choice of $\mathbf{f}$-initialization.

The following section attests to the claim that

The CM criterion serves as a close proxy to J_A which is robust under typical operating conditions.

For a preview, compare the CM cost surface in Figure 8 to the amalgamated MSE surface in Figure 7 for the same channel, $\mathbf{c}_1$. As such, the CM criterion offers a performance metric that bears many similarities to MSE but which is capable of minimization by (stochastic) gradient descent schemes conducted blindly with respect to the transmitted symbols.

With our tutorial orientation, Section II restricts focus to a two-tap FSE design task that permits visualization of equalizer-parameter-space cost-contour plots illustrating various properties of the CM cost function J_{CM}. In particular, we can isolate an "ideal, zero-cost" situation where the stationary points in J_{CM} and J_{A} match exactly and where the minima achieve zero cost. This special case requires several assumptions not often satisfied in practice. We will examine examples of CM-adapted FSE behavior conducted under violations of these requirements for ideal zero-cost equalization. This implicit taxonomy will be used in Section III to provide an overview of the literature citations in the annotated bibliography of Section V.

II. Two-Tap Illustrative Examples

The shape of the cost surface defining a particular stochastic gradient algorithm often lends great insight into the expected behavior of that algorithm. With this in mind, we embark on a tutorial study of the cost surface defined by the CM criterion and descended by CMA. First, however, consider the following list of features characterizing a generic (stochastic) gradient descent algorithm:

• Far from a stationary point, the gradient (i.e. first derivative) of the cost surface determines local convergence rate.

• Near a stationary point, the local curvature (i.e. second derivative) of the cost surface determines local convergence rate.

• Local minima with non-zero cost induce excess steady-state error in *stochastic* gradient descent algorithms with non-vanishing step-sizes.

• Multimodal surfaces may exhibit local minima of varying cost, thus linking initialization to achievable asymptotic performance.

• "Poor" initialization on a multimodal surface can lead a trajectory into temporary capture by (one or more) saddle points, resulting in arbitrarily slow convergence to a minimum.

• Nontrivial deformations of a multimodal surface relocate each saddle point and alter the region of attraction associated with each local minima.

The following sections combine low-dimensional examples with the well-known characteristics above to formulate an intuitive understanding of the CM criterion and its connection to the MSE criterion.

A. Two-Tap Equalizer Design Equations

As discussed in Section I-A, satisfaction of the "length and zero" conditions ensures an exact solution to the zero-forcing equation $\mathbf{h}_\delta = \mathbf{C}\mathbf{f}$. For a two-tap $T/2$-spaced FSE, the length condition is satisfied for channels with impulse responses $[c_0, c_1, c_2, c_3]$ and shorter. For a length-four channel, the root condition is satisfied when the even and odd subchannel polynomials, $C_{\text{even}}(z^{-1}) = c_0 + c_2 z^{-1}$ and $C_{\text{odd}}(z^{-1}) = c_1 + c_3 z^{-1}$, have distinct roots.

In this case, (3) specifies that the FSE design quantities take the following form:

$$\mathbf{C} = \left[\begin{array}{cc} c_1 & c_0 \\ c_3 & c_2 \end{array} \right], \quad \mathbf{f} = \left[\begin{array}{c} f_0 \\ f_1 \end{array} \right]. \tag{23}$$

Since $\mathbf{h}_\delta$ has one nonzero coefficient, the zero-forcing equalizer will be proportional to either the first or the second column of $\mathbf{C}^{-1}$. Thus, all four channel parameters enter into the design of $\mathbf{f}$; the sub-equalizers of Figure 5 are *not* simply inverses of their respective subchannels.

B. Introduction to the CM Cost Function

The CM cost function can be motivated using the temporary assumption that the source is binary valued (± 1). In this case, s_n has a constant squared-modulus of one ($|s_n|^2 = 1$). Under perfect symbol recovery, we know that the output y_n has the same constant-modulus property, and can thus imagine a cost that penalizes deviations from this output condition. This, in fact, defines the CM cost function for a BPSK source:

$$J_{\text{CM}}|_{\text{BPSK}} = \mathrm{E}\big\{(1 - |y_n|^2)^2\big\}.$$

Appendix B presents more general versions of the CM cost function and derives expressions for J_{CM} in terms of channel parameters, particular source and noise statistics, and equalizer coefficients.

The leap of faith, first espoused by [Godard TCOM 80], is the application of J_{CM} to a multilevel (i.e. *non*-constant modulus) source. [Godard TCOM 80], which addressed baud-spaced blind equalization via minimization of J_{CM}, makes the first observation concerning the proximity of the J_{CM} and J_{A} minima:

"It should also be noted that the equalizer coefficients minimizing the dispersion functions closely approximate those which minimize the mean squared error."

This is remarkable because an approximation of J_{CM} can be formed solely from the equalizer output y_n; no training signal is required to compose an accurate gradient approximation for use in a stochastic gradient minimization algorithm such as CMA [Treichler TASSP 83]. It is worth noting that the phase-independent nature of J_{CM} has its own advantages in modem design [*Treichler PROC 98*].

TABLE I
SUMMARY OF CHANNELS USED FOR TWO-TAP FSE EXAMPLES.

Name	$T/2$-spaced Impulse Response	Classification
$\mathbf{c}_1$	[-0.0901, 0.6853, 0.7170, -0.0901]	well-behaved[7]
$\mathbf{c}_2$	[1.0, -0.5, 0.2, 0.3]	well-behaved
$\mathbf{c}_3$	[-0.0086, 0.0101, 0.9999, -0.0086]	nearly-common subchannel roots
$\mathbf{c}_4$	[1.0, -0.5, 0.2, 0.3, -0.2, -0.15]	undermodelled

C. Illustrative Cost Surface Examples

The following subsections present mesh and contour plots of the CM cost surface for a two-tap $T/2$-spaced FSE under various operating conditions. Refer to Table I for definitions of the various channels used in our experiments. In all contour plots, the asterisks ($*$) indicate the locations of global MSE (i.e. J_A) minima while the crosses ($\times$) indicate the locations of local MSE minima. Recall that different pairs of MSE minima (reflected through the origin) correspond to different values of system delay, while the two elements composing each pair correspond to the two choices of system polarity[6]. Thus, the asterisks mark the MMSE equalizers of optimum system delay. The "MSE ellipse axes" appearing in the upper left corner of each contour plot indicate the orientation and eccentricity of the elliptical MSE contours (see Figure 9).

All quantities in the experiments are real-valued. Unless otherwise noted, the source used was zero-mean and i.i.d. with alphabet $\{-1, 1\}$.
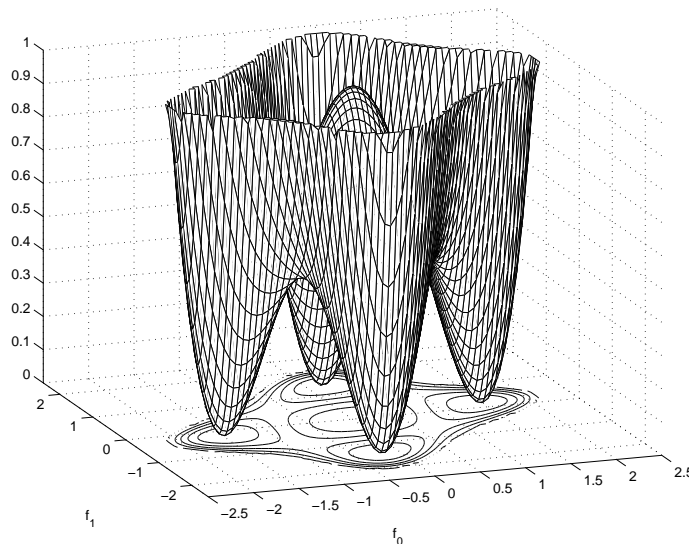

Fig. 8. J_{CM} for well-behaved channel $\mathbf{c}_1$ and no noise, in equalizer ($\mathbf{f}$) space.

C.1 Ideal Zero-Cost Equalization

For well-behaved channel $\mathbf{c}_1$ in the absence of channel noise, Figure 8 plots J_{CM} in equalizer space. Recall that Figure 7 plots J_A for the same noiseless channel. For a different well-behaved and noiseless channel, $\mathbf{c}_2$, Figure 9 superimposes the corresponding J_{CM} and J_A cost contours. Note the symmetry (with respect to the origin) exhibited by both J_{CM} and J_A cost surfaces.

---

[6]We note that in the complex-valued CM criterion, each pair of minima would be replaced by a continuum of minima spanning the full range $(0 - 2\pi)$ of allowable system phase.

[7]"Well-behaved" indicates the absence of common or nearly-common subchannel roots.
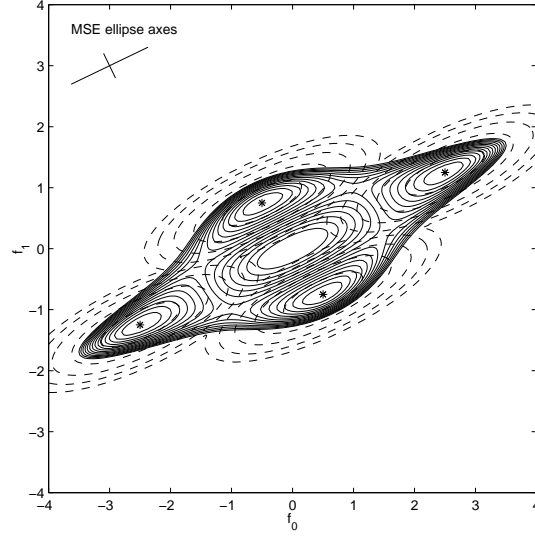
Fig. 9. J_{CM} contours (solid) for well-behaved channel $\mathbf{c}_2$ and no noise, with J_A overlay (dashed) and global MSE minima marked by "$*$", in equalizer ($\mathbf{f}$) space.

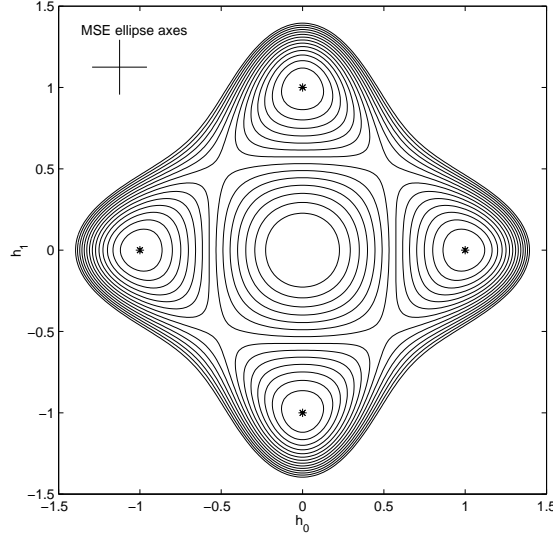Fig. 10. J_{CM} contours for well-behaved channel $\mathbf{c}_2$ and no noise in combined channel-equalizer ($\mathbf{h}$) space.

In these ideal situations, all MSE and CM minima attain costs of zero (see Figures 7 and 8). In addition, it can be seen that the locations of the J_{CM} and J_A minima coincide. (The J_{CM} minima locations can be inferred from the J_{CM} cost contours.) Figure 9 also indicates that the curvatures of CM and MSE cost surfaces in the neighborhoods of local minima are closely related.

C.2 Combined Channel-Equalizer Space

The behavior of a gradient descent of J_{CM} is sometimes studied in the (downsampled) combined channel-equalizer space (i.e. $\mathbf{h}$ from Section I-A). The appeal of studying J_{CM} in $\mathbf{h}$-space follows from the normalization and alignment of J_{CM} with the coordinate axes. These features are clear in a comparison of Figure 10 to Figure 9, both constructed from the same noiseless channel. Equation (4) implies that a unique reversible mapping (i.e. an isomorphism) exists between points on the J_{CM} surfaces in $\mathbf{h}$- and $\mathbf{f}$- spaces when $\mathbf{C}$ is invertible, as it is here in our 2-tap example.
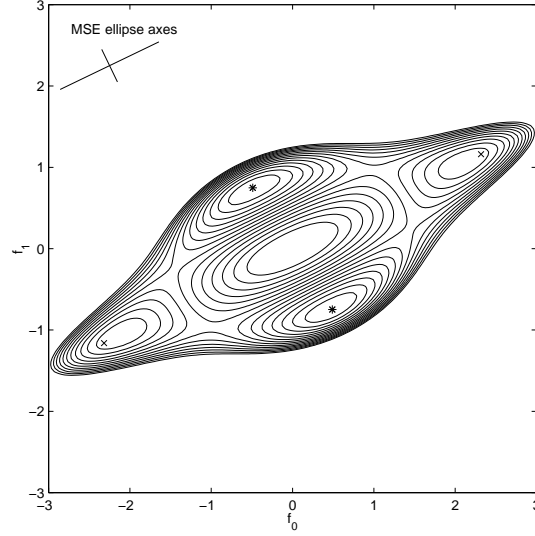
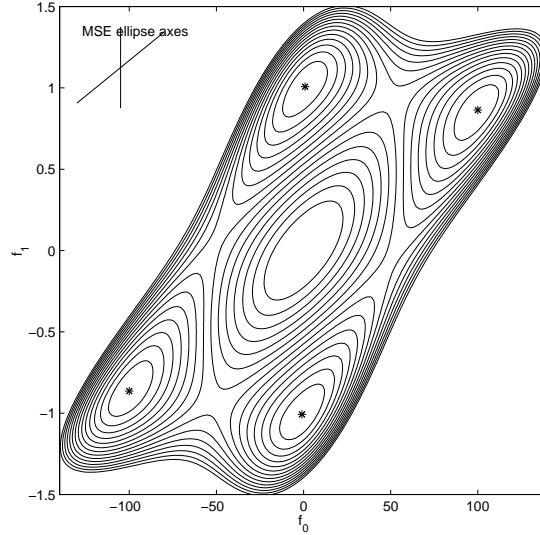Fig. 11. J_{CM} contours for well-behaved channel $\mathbf{c}_2$ and 20 dB SNR.

Fig. 12. J_{CM} contours for nearly-common subchannel-roots channel $\mathbf{c}_3$ and no noise. Note axis scaling.

C.3 Additive White Channel Noise

As channel noise is introduced, Figure 11 indicates that the MSE and CM minima both move towards the origin in $\mathbf{f}$-space. The J_A and J_{CM} minima move by different amounts, though, destroying the equivalence that existed between them in the ideal case of Figure 9. However, the relative proximity of J_A and J_{CM} minima, evident in Figure 11, still prompts consideration of J_{CM} as a close proxy for the amalgamated MSE cost J_A even when in the presence of channel noise.

C.4 Common Subchannel Roots

As evidenced by the expression we derived for MSE minima,

$$\mathbf{f}^\dagger = (\mathbf{C}^H \mathbf{C} + \lambda \mathbf{I})^{-1} \mathbf{C}^H \mathbf{h}_\delta,$$

when $\mathbf{C}^H \mathbf{C}$ has a large condition number, modest values of λ can have significant consequences on $\mathbf{f}^\dagger$ (and thus on the J_A cost surface). If the two subchannels ($c_0 + c_2 z^{-1}$ and $c_1 + c_3 z^{-1}$) have a nearly-
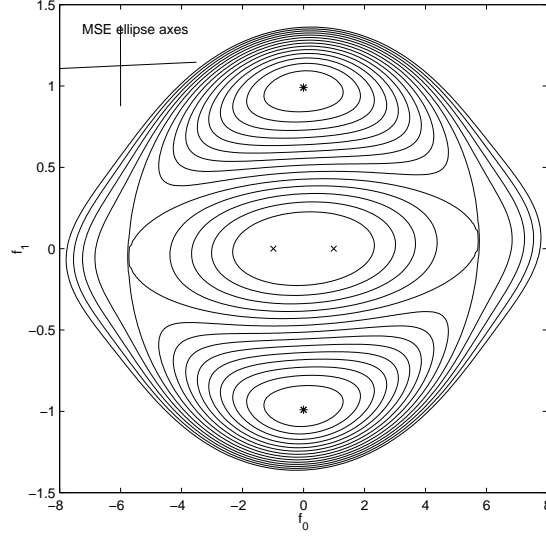
14

Fig. 13. J_{CM} contours for nearly-common subchannel-roots channel $\mathbf{c}_3$ and 20 dB SNR. Note axis scaling.
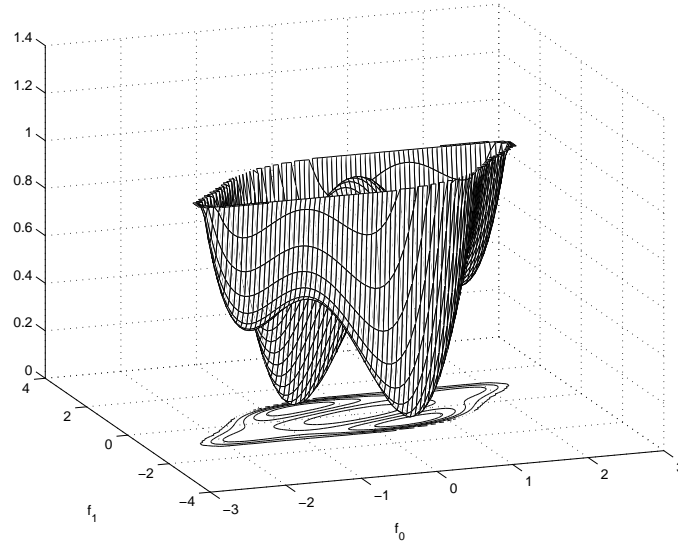

Fig. 14. J_{CM} for undermodelled channel $\mathbf{c}_4$ and no noise.

common root $(c_3/c_1 \approx c_2/c_0)$ then (23) indicates that the column space of $\mathbf{C}$ collapses, and thus we expect that one eigenvalue of $\mathbf{C}^H \mathbf{C}$ will be near zero [*Strang Book 88*]. Figures 12 and 13 use channel $\mathbf{c}_3$ to demonstrate the cost surface sensitivity to noise in the presence of nearly-common subchannel roots. Even under such severe surface deformation, we note that the global J_{CM} minima remain in the vicinity of global J_{A} minima. This further demonstrates the robustness of the relationship between J_{CM} and J_{A}.

C.5 Channel Undermodelling

In general, under violation of the length condition (discussed in Section I-A), no equalizer settings are capable of achieving zero MSE or CM cost. This can be confirmed by extending the length of impulse response $\mathbf{c}_2$ by two samples, thus forming the "undermodelled" channel $\mathbf{c}_4$. (Note that the two extra coefficients forming $\mathbf{c}_4$ are no larger than any of the coefficients in $\mathbf{c}_2$.) Figure 14 shows the CM cost surface for this undermodelled channel. Large differences in the heights of local minima
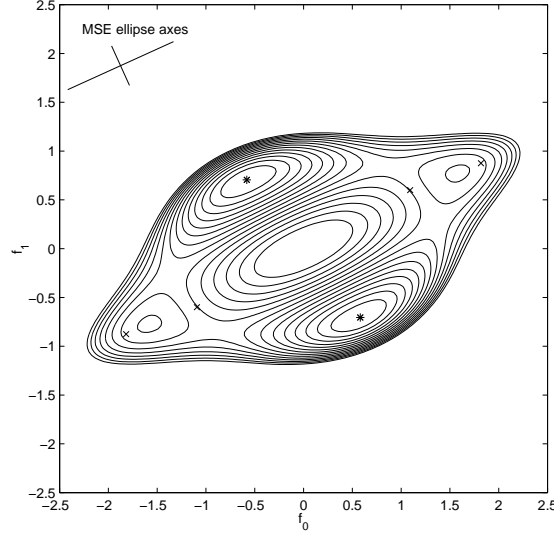
15

Fig. 15. J_{CM} contours for undermodelled channel $\mathbf{c}_4$ and no noise.
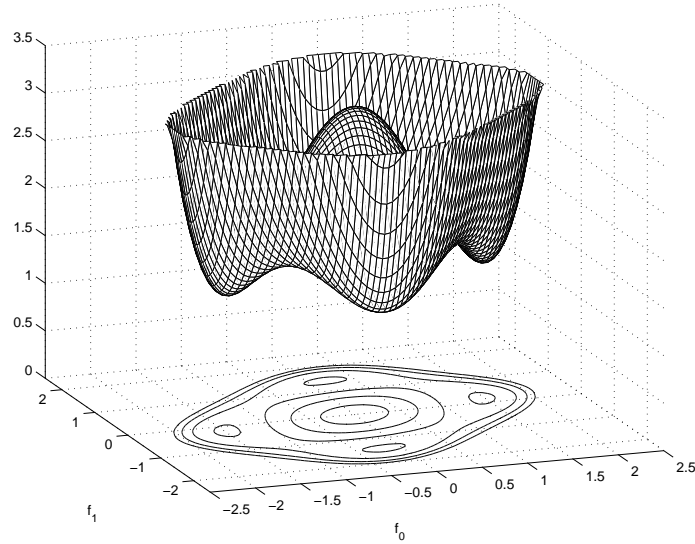

Fig. 16. Effect of source shaping ($\kappa_s = 1.8$) on J_{CM} for channel $\mathbf{c}_1$ in equalizer space with no noise.

demonstrate that the CM cost surface can indeed be significantly multimodal.

Elongating the channel impulse response adds another possibility for the system delay δ and thus increases the number of J_A minima (see Figure 15). Note, however, that the number of CM minima have not changed. More importantly, note that the global CM minima remain close to their MSE counterparts under violations of the length condition.

C.6 Non-CM Source

The constant-modulus source property leading to the ideal zero-cost situation in Figures 8, 9, and 10 is violated in constructing the cost surface in Figure 16. Here, the source is real-valued 32-PAM, which is far from constant-modulus. The non-CM property increases the source kurtosis κ_s (defined in (50)) and increases the minimum CM cost relative to that of a CM source. Notice also that the CM cost surface has become "flattened" in the parameter plane. However, as the CM surface deforms due to a non-CM source, the minima locations remain unchanged.

16

D. Summary

Our investigations of low-dimensional examples under the following "ideal, zero-cost" conditions:
- no channel noise (i.e. $\lambda = 0$),
- no common subchannel roots (i.e. avoidance of $c_2/c_0 = c_3/c_1$),
- sufficient equalizer length (i.e. $N \geq M - 1$ for $T/2$-spaced FSEs),
- i.i.d., zero-mean, constant-modulus source (circularly-symmetric when complex),

showed that, under such conditions, the J_{A} and J_{CM} minima coincide and achieve zero respective cost. Our other examples suggest that modest deviations from the ideal conditions can be tolerated in the following sense: under suitable choice of initialization, a stochastic-gradient minimization of J_{CM} will approximate the performance achieved by the same minimization of J_{A}. We did find, however, that the deformations caused by various violations of the ideal zero-cost conditions are different. In fact, substantial effort has been expended to characterize the performance robustness properties of the CM criterion (as descended by popular gradient descent strategies). Section III catalogs much of this effort.

The previous examples can be used to illustrate and interpret the following observations:
- *Channel noise*: CMA-based blind equalization is typically successful in common noise environments (i.e., $\sigma_s^2 > \sigma_n^2 > 0$). Under modest noise levels, relocation of global minima toward the origin is typically more severe than changes in surface curvature around such minima.
- *Undermodelling of channel length*: Given hardware constraints on equalizer length, residual ISI is unavoidable in practice. Mild contributions from uncompensated portion of channel response typically result in mild surface deformation.
- *Nearly-common subchannel roots*: These seem quite likely as channel length increases. (See Figure 18.) Nearly common subchannel roots increase sensitivity to other violations from ideal conditions, but only for sub-optimal CM solutions; global CM minima still exhibit robust performance.
- *Source kurtosis*: Non-uniform (i.e. shaped) symbol distributions often leads to increased source kurtosis. As source kurtosis approaches Gaussian[8], the surface lifts and flattens. Lifting increases the excess error of stochastic adaptation (e.g. CMA), while flattening reduces its convergence rate. If the source exhibits a Gaussian kurtosis, the minima and saddle points vanish along a *rim* of the CM surface so that the gradient has solely a radial component. In this case, convergence to desirable settings is practically impossible.
- *Source correlation*: This may occur, e.g., as a result of differential encoding. Small amounts result in slight cost surface deformation. Large amounts cause major problems, such as additional local minima with terrible performance.
- *Non-CM source*: This property is unavoidable in communication systems using multi-level constellations. Though non-CM sources do not alter the minima locations, they raise and flatten the CM surface (as a consequence of increased source kurtosis — see above).
- *Initialization*: The CM surface is unavoidably multimodal. Choice of initialization affects both time-to-convergence and steady-state performance. One approach referred to in the literature suggests initializing the equalizer with a single spike[9] time-aligned with the channel response's center of mass. In this way, crude knowledge of the channel impulse response envelope can be used to aid initialization.
- *Channel time-variation*: We proceed under the global assumption that the channel varies slowly enough in time to be tracked by the CM-minimizing gradient descent algorithm. In the vicinity of a local minimum, the tracking capabilities of any gradient descent scheme can be related to the local curvature.
- *Equalizer tap-spacing*: Fractionally-spaced equalizers have the ability to perfectly cancel ISI caused by a finite-length channel impulse response. In contrast, a baud-spaced equalizer requires an infinite number of taps for the same capability. Though we admit that this noiseless FIR channel model is

---

[8]Table II presents the values of normalized kurtosis for various sources.

[9]The single spike initialization has its origins in baud-spaced equalization. Fractionally-spaced counterparts are discussed in Section III-B.3.

rather academic, practical experience offers much evidence for the superiority of fractionally-spaced equalization [*Gitlin Book 92*].

• *Transient versus steady-state performance*: Dynamic system design is often a tradeoff between transient and steady-state performance. Convergence rate is a transient behavior descriptor; slow convergence is undesired. Excess error (due to a non-vanishing step-size and a nonzero local minimum) is a steady-state feature; abundance of excess error is undesired.

III. CM-Minimizing Equalization Literature Categorization

The last section presented a tutorial view of the linear equalizer design task and related the minimization of the delay-optimized and phase-indifferent mean-squared recovery error (J_A) to minimization of the CM criterion (J_{CM}). Section V presents a bibliography of the literature dealing with the CM criterion and its optimization via steepest gradient descent (such as with CMA). Each entry in the bibliography is annotated with boldface letters that indicate the classification of its content. The purpose of this section is to describe our classification scheme in terms of the problem formulation and the examples of the preceding section. We also take this opportunity to cite certain papers as recommended reading on particular topics.

In addition to the birth of the CM criterion in the early 1980s, highlights in its analytical history include:

• establishment of "perfect" conditions under which a gradient descent of the CM cost surface results in asymptotically perfect symbol recovery, i.e. "global convergence,"

• confirmation that, under slightly imperfect conditions, the CM minima remain in the vicinity of the MSE minima for various choices of delay and sign,

• recognition that, due to performance differences between CM minima under less-than-perfect conditions, initialization may be critical to acceptable transient and steady-state behavior.

The "perfect" global-convergence conditions referred to in these statements differ in detail between the baud- and fractionally-spaced cases. As discussed in Section I-A, achievement of perfect source recovery devolves into exact solution of a set of simultaneous linear equations when channel noise is absent. Solution of these equations ensures that the transfer function characterizing the baud-spaced system (relating source symbols to equalized soft decisions) achieves that of a pure delay. One requirement on the existence of this perfectly-equalizing solution is that the equalizer must have enough degrees of freedom. For a baud-spaced equalizer and a FIR channel, this latter requirement necessitates an equalizer with infinite impulse response (IIR) [Foschini ATT 85]. For $T/2$-spaced FSEs, on the other hand, an equalizer response length matching (or exceeding) that of the channel proves sufficient [*Tong CISS 92*]. The other requirement for the existence of a perfectly-equalizing solution is that the system of equations be well-posed. We mean, in an algebraic sense, that the matrix characterizing the linear system of equations must be non-singular. For baud-spaced equalizers, this non-singularity condition prohibits nulls in the channel frequency response (which implies, for example, that no FIR channel zeros are tolerated on the unit circle). We henceforth refer to satisfaction of this baud-spaced condition as "invertibility." For $T/2$-spaced FSEs, this non-singularity translates into a lack of common subchannel roots (see Appendix A-C) and is commonly referred to as "subchannel disparity."

If conditions on the source (e.g., zero-mean, circularly-symmetric, white, and sub-Gaussian) are added onto the perfect equalization requirements described in the last paragraph, a gradient descent of the CM criterion will provide asymptotically perfect source recovery from any baud- or fractionally-spaced equalizer initialization. In this case, the multiple CM minima all have the same depth — like an egg carton. The distinctions in global convergence conditions between the baud- and fractionally-spaced cases prompt our separation of these two cases. We note that, while analysis of CM-minimizing baud-spaced equalizers has been published since their introduction in 1980, very little analysis of CM-minimizing fractionally-spaced equalizers was published before 1990.

The stringency of the global convergence requirements has prompted theoreticians to examine the impact of their violation. For example, what if the FSE length is less than the total channel response but greater than the "significant" portion of the channel response? How are prominent features of the CM cost surface (e.g., stationary point locations, regions of attraction, and heights of local minima) altered as the source is shaped or correlated and/or channel noise power increases and/or channel disparity is lost? While engineering practice desires answers about simultaneous dissatisfaction of all global convergence conditions, theoretical analysis is more likely to move forward by studying individual (or possibly pairwise) violation of these conditions. Therefore, we are encouraged to adopt a set of literature categorizations concerning studies of robustness to violations in each of the four global-convergence conditions (i.e., absence of channel noise, sufficient length, adequate disparity, and use of a zero-mean, white, circular, sub-Gaussian source process).

In Section II-C we noted that the CM and MSE error surfaces are quite similar in the vicinity of the CM local minima. This relationship implies that the local behaviors of their stochastic gradient descent minimizers (e.g. CMA and LMS, respectively) should be closely related. As a result, we are encouraged to use key behavioral descriptors associated with "classical" trained-LMS equalization theory as further categories for our literature classification. In particular, we borrow excess mean-squared error (i.e. misadjustment[10]) and convergence rate.

While the CM and MSE criterion are comparable in a local context, their global characteristics are strikingly different. Recall the multimodality of the CM cost surface (see, e.g., Figures 8 and 14). As noted earlier, a good gradient-descent initialization may be necessary to ensure convergence to a "good" local minimum as well as to avoid temporary local capture by saddle points. In contrast, consider the trained-LMS cost surface: a unimodal elliptical hyper-paraboloid. Its unimodality obviates the need for a clever initialization strategy (assuming the training delay has been chosen). In fact, the LMS equalizer is often initialized by zeroing the parameters[11]. If we consider delay-selection as part of the "initialization" of trained LMS, however, we find many similarities with the equalizer parameter initialization of CMA. Specifically, the choice of training delay bounds asymptotic LMS performance, and, in conjunction with the equalizer initialization, LMS time-to-convergence. Conversely, CMA equalizer initialization determines (asymptotic) system delay. With these thoughts in mind, we add surface topology and initialization strategy as literature categories under the heading of gradient descent behavior.

Summarizing, the classification scheme we adopt for our literature review uses a total of 11 labels within the three main categories discussed above:

1. Equalizer Tap-Spacing:
- (**B**) Baud-spaced
- (**F**) Fractionally-spaced

2. Global Convergence Criteria Dissatisfaction:
- (**P**) Perfect: no noise, sufficient length, adequate disparity/invertibility, and zero-mean, white, circular sub-Gaussian source
- (**N**) Noise present
- (**L**) Equalizer length inadequate
- (**D**) Disparity/Invertibility lost or threatened
- (**S**) Source shaped or correlated

3. Gradient Descent Algorithm Behavior:
- (**E**) Excess error (due to non-vanishing step-size)
- (**R**) Rate of convergence
- (**T**) Topology of cost surface
- (**I**) Initialization strategy

---

[10]Misadjustment is defined as the ratio of excess MSE to minimum MSE.
[11]Initializing CMA at the origin proves bovine (i.e., slow and unwise) due to the zero-valued CM-cost gradient there.

The remainder of this section is organized by the categorization above; each of the 11 labels is discussed using selected citations drawn from the bibliography.

Because the focus of this paper is the CM criterion in a blind linear equalizer application, we have not considered work that

1. principally deals with algorithm modifications (e.g. normalized, least-squares, Newton-based, block, anchored, or signed CMA) that may alter the (effective) cost function surface shape,

2. infers behavior principally from simulation studies with no connection made to the CM cost function, or

3. principally addresses applications other than linear equalization (e.g. beamforming, source separation, interference cancellation, channel identification, depolarization, or decision-feedback equalization).

Though some of our citations do involve the categories above, we have chosen to include them because they contain a substantial amount of directly relevant material as well.

We do not provide a synopsis of each citation in the bibliography. Rather, we propose the abstracts of each paper as a source for synopses and provide a postscript bibliography that includes abstracts at `http://backhoe.ee.cornell.edu/BERG/bib/CM_bib.ps`.

A. Equalizer Tap-Spacing

Practically speaking, the equalizer tap-spacing refers to the rate at which the received signal is sampled and processed by the equalizer. In creating a discrete linear system model, the tap-spacing determines the delay time of the equalizer difference equation. Using T to denote the source symbol interval, baud- or T-spaced FIR equalizers use a unit delay of T seconds in their tapped delay line. Fractionally-spaced equalizers use a tap-spacing less than T. The most common fractional tap-spacing is $T/2$ seconds. In the bibliography in Section V, approximately two-thirds of the citations cover baud-spaced equalization, while the remaining one-third cover fractionally-spaced equalization.

A.1 Baud-Spaced Equalization

The pioneering paper introducing the CM criterion for a complex-valued source [Godard TCOM 80] considers baud-spaced equalization only.

Conditions assuring global convergence of a baud-spaced equalizer updated via CMA: (i) no channel noise, (ii) infinite impulse response equalizer, (iii) no nulls in channel frequency response (i.e. no FIR channel zeros on the unit circle), and (iv) a zero-mean, independent (and circularly-symmetric if complex-valued) finite-alphabet source with sub-Gaussian kurtosis.

The first proof of global convergence for CMA in adapting a baud-spaced equalizer relied on a doubly-infinite equalizer parameterization which allowed any combined channel-equalizer impulse response [Foschini ATT 85]. This allows convergence study in the combined channel-equalizer space, which has analytical advantages.

A.2 Fractionally-Spaced Equalization

Original motivations for the use of fractional- rather than baud-spacing included: insensitivity to sampling phase, ability to function as a matched filter, ability to compensate for severe band-edge delay distortion, and reduced noise enhancement [*Gitlin Book 92*]. Fractionally-spaced equalizers have nearly dominated practice since the 1980s [*Wolff MIL 88*]. One feature of fractionally-spaced equalizers — virtually unnoticed until the 1990s — was the possibility that under ideal conditions a fractionally-spaced equalizer of finite time-span could perfectly equalize a FIR channel [*Bergmans PJR 87*]. As noted in [*Tong CISS 92*], this suggests the same connection of equalizer parameters to the combined channel-equalizer parameters exploited in [Foschini ATT 85] and therefore confirms the potential for global convergence of a CM-minimizing fractionally-spaced equalizer.

*Conditions assuring global convergence of a T/2-spaced FSE updated each baud interval via CMA: (i)
no channel noise, (ii) equalizer time span matching or exceeding that of the FIR channel, (iii) no
reflected zeros in the T/2-sampled FIR channel transfer function, and (iv) a zero-mean, independent
(and circularly-symmetric if complex-valued) finite-alphabet source with sub-Gaussian kurtosis.*
These global convergence inducing conditions do not include restriction to a constant modulus source,
which was included among the "ideal zero-cost" conditions of Section II-D.

The first global convergence proofs for fractionally-spaced CMA not simply relying on the extension
of the baud-spaced arguments in [Foschini ATT 85] appears in [Li TSP 96a].

B. Gradient Descent Algorithm Behavior Theory

The algorithm that performs a stochastic gradient descent of J_{CM} is often referred to as the Constant
Modulus Algorithm or CMA:

$$\mathbf{f}_{n+1} = \mathbf{f}_n + \mu\,\mathbf{r}_n^*\,y_n(\gamma - |y_n|^2). \tag{24}$$

Equation (24) is written in terms of the (fractionally-sampled) regressor vector at time n:

$$\mathbf{r}_n = \left[r_n^{\mathrm{odd}}, \ldots, r_{n-(N-1)}^{\mathrm{odd}}, r_n^{\mathrm{even}}, \ldots, r_{n-(N-1)}^{\mathrm{even}}\right]^t, \tag{25}$$

the equalizer parameter vector $\mathbf{f}_n$ at time index n, the equalizer output y_n, a step-size μ, and the
squared source-modulus γ (also referred to as the dispersion constant).

The study of dynamic systems, such as CMA, is often divided into transient and steady-state
stages. Convergence rate is the dominant transient performance descriptor in classical LMS theory.
Minimum MSE and excess MSE (and their dimensionless ratio, misadjustment = EMSE/MMSE) are
the dominant steady-state performance descriptors. Therefore, we consider their CM counterparts
here.

Though initialization is not a major concern for the unimodal cost functions of MSE-minimizing
equalizers (with preselected delay and phase), it is an unavoidable issue for CM-minimizing equalizers
due to the multimodal topology of their associated cost surface. Though initialization strategies exist,
none have been proven 100% successful in practice.

B.1 Convergence Rate

For trained LMS, the convergence rate (or geometric decay factor) of the sum-squared parameter
error (and squared recovery error) is approximately bounded above and below by one minus twice
the product of the step-size and the smallest and largest eigenvalues, respectively, of the received-
signal's autocovariance matrix (i.e., $1 - 2\mu\lambda_{\min} > \frac{1}{\tau} > 1 - 2\mu\lambda_{\max}$). This arises because the underlying
quadratic cost function has the same Hessian, or curvature, across its entire surface. In contrast,
the multimodal CM cost function has a Hessian that varies across its surface. Early convergence
rate studies addressed this variation in convergence rate across the CM cost surface by focusing on
convergence rate descriptors in various regions, such as far from minima and near minima [Larimore
ICASSP 83].

Referring to Figure 9, initialization near $[f_0, f_1] = [2.5, 0]$ will lead to a small-stepsize gradient-
descent trajectory that passes through the neighborhood of a saddle point. An example displaying
multiple temporary saddle-captures appears in [Lambotharan SP 97]. We believe this saddle capture
phenomenon to be the source of the folklore that considers CMA to be "slow converging."
*A lower bound on the initialization-independent convergence rate is impossible with the multimodal CM
surface due to potential of indefinite-term capture by saddle points.*

In the neighborhood of a local minimum, the curvature of CMA's cost surface can be directly related
to that of trained-LMS [Touzni EUSIPCO 96]. Thus, the LMS convergence rate expression can be

used in a traditional manner (e.g. [*Treichler SPM 96*]) to provide limits on the channel tracking[12] capabilities of CMA.

B.2 Excess Cost at Convergence

In realistic situations, it is impossible to zero the update of a non-vanishing-stepsize stochastic gradient descent algorithm, even at the optimum solution. With trained LMS or CMA, this undying perturbation may be a result of channel noise or residual ISI. With CMA, the non-zero update may also be the result of a non-CM source. The effect of a non-vanishing equalizer update is an asymptotic mean-squared error level higher than that attained by the optimum fixed equalizer. This is directly related to the lifting effect that a non-CM source has on the CM cost surface, evident in Figure 16. *In addition to the factors determining the excess MSE of trained LMS (i.e., stepsize, minimum achievable cost, equalizer length, and received signal power) CMA also has a term dependent on the source kurtosis.* Excess MSE of fixed (small) step-size CMA due to a non-CM source is analyzed in [Fijalkow TSP 98].

Figures 8 and 16 show the effect of changing the source from constant to non-constant modulus while simultaneously satisfying all of the global convergence conditions. Though the CM minima rise in height, they remain in the same locations in the equalizer parameter plane. As long as the source is kept sub-Gaussian, a (pure) gradient descent algorithm would be still able to asymptotically achieve perfect symbol recovery.

B.3 Initialization

As noted in the examples of Section II-C and illustrated in Figures 11 and 15, the presence of noise or channel undermodelling causes some CM minima to achieve better performance than others. *Under violation of the conditions ensuring global convergence, choice of initialization determines asymptotic performance.* Two initialization strategies are common in the literature and in practice: spike-based or matched filter. The "single-spike" initialization, promoted in [Godard TCOM 80] for baud-spaced CMA, is characterized by one non-zero equalizer tap, usually located somewhere in the central portion of the equalizer tapped-delay line. For $T/2$-spaced CMA, a suitable extension of the single-spike idea might be a "double-spike" initialization, whereby two adjacent taps are initialized non-zero. In the frequency domain, double-spike initialization has a lowpass characteristic, a property also shared by the transmitter's pulse-shaping filter. In a mild-ISI environment, one might even consider initializing the FSE with an impulse response matching the pulse-shaping filter itself, as (in this mild case) this response is close to the expected steady-state equalizer solution (assuming that the FSE is used to accomplish matched filtering at the receiver.)

All of the initialization techniques above still require a selection of delay, i.e. spike positioning within the equalizer time span. This delay choice is intimately connected to the delay-choice in trained-LMS equalization in the following way: *CMA tends to converge to minima with the same group delay as its initialization.* Figure 17 provides evidence for this claim using double-spike initializations of $T/2$-spaced CMA on the SPIB microwave channel shown in Figure 1 under 50 dB SNR and a QPSK source. Note the (affine) linear correspondence between double-spike position and asymptotically achieved system delay. Another interesting characteristic of Figure 17, seen after comparing sub-plots (d)-(f) to Figure 6, is its suggestion that *the set of system delays reachable by CMA are best in an MMSE sense.* We offer these last two statements as educated conjectures, as no theoretical proofs yet exist to verify them.

---

[12]In many practical implementations, such as those with low ambient noise levels, CMA lowers the symbol error rate to level suitable for decision-directed LMS (DD-LMS) to take over. Due to its lower excess error, DD-LMS is preferred for tracking the slow channel variations. In low-SNR situations, however, such as those that may arise with a coded system, the tracking ability of CMA might prove important due to the potential infeasibility of DD-LMS.
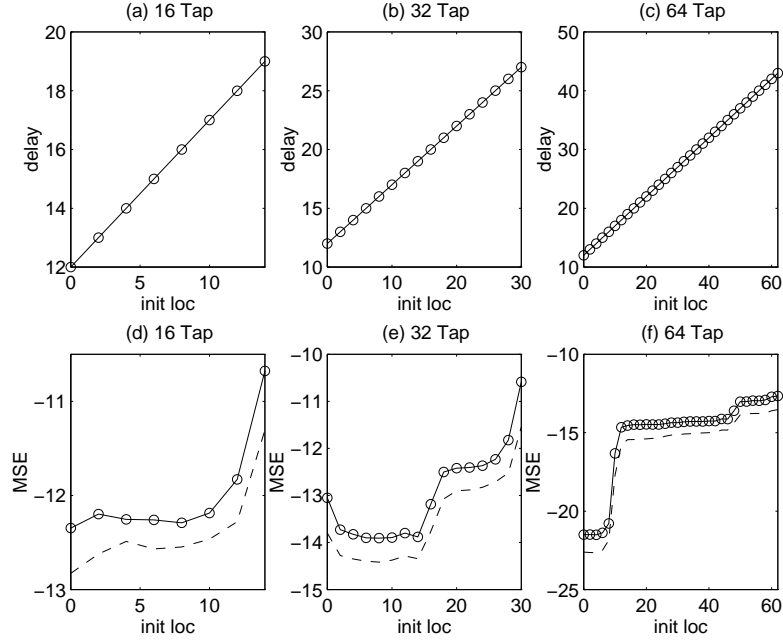
Fig. 17. (a)-(c) CMA's achieved system delay as a function of double-spike location and (d)-(f) CMA's asymptotic MSE performance (solid) compared to same-delay MMSE performance (dashed).

The aforementioned relationship between initialization and channel group delays suggests that *a priori* information about the channel may aid in selection of initialization delay choice. The bibliography notes the existence of other, more complicated, off-line initialization schemes that leverage such notions.

B.4 Surface Topology

Refer to Figures 8 and 14. The "molar" shape of the CM cost surface in two-tap real-valued equalizer space was used in Section II-C to aid in an understanding of CMA's transient and asymptotic performance as well as to motivate the importance of initialization. Section II-C also described how deformation of this molar shape occurs with violation of the various ideal zero-cost conditions, and it used this surface-centric view to predict the pertinent effects of these violations.

The three-dimensional "molar" shape typical of the real-valued 2-tap-equalizer CM cost surface offers a compact visualization of virtually all of the major features of CMA behavior theory, applicable even to longer equalizers.

Surface characterization via gradient and Hessian formulas is provided in [Johnson IJACSP 95] for baud-spaced equalizers. [LeBlanc IJACSP 98] offers a more developed topological study of the fractionally-spaced CM criterion.

C. Violation of Conditions Ensuring Global Convergence

C.1 Perfect: All Conditions Satisfied

While Sections III-A.1 and III-A.2 listed conditions ensuring the global convergence of CMA, their violation is unavoidable in practice.

There exists a set of conditions under which an arbitrarily-initialized gradient-descent minimization of the CM criterion results in perfect symbol recovery. These "global convergence" inducing conditions, however, are unconditionally violated — if only modestly — in practice.

Our claim is that modest violation of the global convergence conditions does not destroy the utility of the CM criterion.

23

C.2 Channel Noise Present

CM-based blind equalization typically remains successful in common noise environments (i.e., $\sigma_s^2 > \sigma_n^2 > 0$). To recall the cost surface deformations due to noise, compare Figure 11 to Figure 9. *When the presence of (modest) channel noise is the only violation of the global convergence conditions, the locations of global CM minima shift toward the origin in equalizer parameter space and the minimum achievable CM cost is increased.* This behavior is strikingly similar to the behavior of the MSE criterion in the presence of channel noise. In fact, under modest amounts of noise, the CM minima remain near the MSE minima [Zeng TIT 98], [Fijalkow TSP 97].

At extremely high noise levels (i.e. $\sigma_n^2 > \sigma_s^2$), the two criteria differ in the following manner: the MSE minima continue to move towards the origin, while the CM minima remain within an annulus outside the origin. This behavior is attributed to the so-called "CMA power constraint" [Zeng TIT 98].

We have also observed the disappearance of *local* minima under modest-to-high noise levels [Chung ICASSP 98], especially for channels without much disparity (see Figure 13).

C.3 Insufficient Equalizer Length

In order to completely cancel the ISI induced by an arbitrary FIR channel, one requires an IIR baud-spaced equalizer or a sufficiently long FIR fractionally-spaced equalizer. In the presence of channel noise, the MSE-optimal equalizer makes a compromise between ISI cancellation and noise gain, and the resulting equalizer impulse response is no longer finite-length, even for fractionally-spaced equalizers [*Gitlin Book 92*]. *In the presence of noise, the (baud- and fractionally-spaced) MMSE equalizers have an infinite impulse response, implying that the length of an FIR equalizer should be chosen to capture "enough" of the desired response.*

Studies on the effect of violations in the equalizer length condition include [Li TSP 96c] in a baud-spaced context, and [Endres ICASSP 97], [Endres SPAWC 97] in a fractionally-spaced context. The latter provide evidence of CMA robustness to modest channel undermodelling and include approximate bounds on performance.

As hardware advances permit increased baud-rate, yet physical channel delay-spreads remain unchanged, the relative length of the channel impulse response grows proportionally. To combat ISI, there is a corresponding need to increase equalizer length. Therefore, the desire for higher communication rates will always stress the equalization task. This is a primary justification for the continued development/study of truly simple adaptive equalization algorithms like LMS and CMA.

C.4 Disparity/Invertibility Lost

As discussed earlier, the set of zero-forcing equalizer design equations becomes poorly conditioned in the presence of deep spectral nulls for baud-spaced equalizers or the presence of nearly-common subchannel roots for fractionally-spaced equalizers. Poor conditioning implies an increased *parameter* sensitivity to noise and other violations of the global convergence conditions. Fortunately, this parameter sensitivity does not imply a *performance* sensitivity. In other words, global CMA minima remain robust under a loss of disparity. We note that the same is true for the delay-optimal MMSE solutions. *A near-loss of disparity (for FSEs) or invertibility (for BSEs) dramatically increases the sensitivity of suboptimal CM (and MSE) minima to other violations in the global convergence conditions. However, global CM (and MSE) minima remain robust under these conditions.*

The behavior of fractionally-spaced CM (and MSE) minima under loss of disparity is explained through the following design procedure. For simplicity, let us assume the absence of noise. (1) Factor the common root(s) out of the subchannels in Figure 5 and form a new system composed of the common root(s) component and what remains of the multichannel component, connected in series. (2)

Design the subequalizers so that the remaining multichannel component approximates the inverse of the common root(s) component. At this point, the cascaded system should approximate a pure delay. This procedure closely describes the construction of the MMSE or CM-optimal equalizers under a loss of disparity [Fijalkow TSP 97]. We describe this idea more formally in Appendix A-C.

There are a number of reasons that we expect the presence of nearly-common subchannel roots, i.e. nearly-reflected[13] $T/2$-spaced roots, in realistic situations. Looking at Figure 18, which portrays the roots of the length 300 $T/2$-sampled SPIB channel whose impulse response appears in Figure 1 and whose response we consider to be "typical," one notices the apparent plethora of nearly-reflected roots. Similarly, one might realize that a long FIR approximation to a pole[14] in the physical channel would also generate nearly-reflected roots. All of these reasons suggest the likelihood of nearly-common subchannel roots in realistic situations. See [Ding SPL 96] for further discussion on the existence of reflected roots in physical systems (and its negative implications on second-order-statistics based blind equalization).
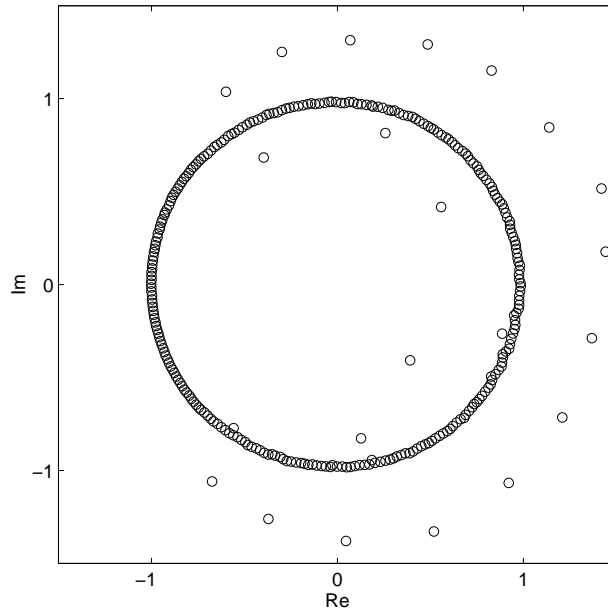

Fig. 18.   Roots of $T/2$-sampled SPIB terrestrial microwave Channel #3.

C.5 Shaped or Correlated Source

Source shaping, encouraged by a potential increase in coding gain (see, e.g. [Forney CM 96]), has the effect of making the source symbol distribution more Gaussian. As far as our problem is concerned, it has the practical effect of raising the kurtosis. Increases in source kurtosis, as long as they remain sub-Gaussian, do not affect the locations of CM local minima. However, they are known to flatten the CM cost surface in all but the radial direction, making CMA's convergence to the minima slower (and in the limiting Gaussian case, impossible). In addition, increases in source kurtosis have been shown to raise the CM surface (see Figure 16), thus increasing the excess asymptotic error levels achieved by non-vanishing-step-size stochastic gradient algorithms.

Recall that non-CM sources also have kurtoses greater than one. To put source shaping in perspective, Table II presents the kurtosis of popular source alphabets along with the limiting Gaussian

---

[13]Common subchannel roots have been shown to be identical to $T/2$-spaced channel roots reflected across the origin [Tugnait TIT 95].

[14]A degree-N polynomial forming a close approximation to a single pole can be constructed using N roots on a ring in the complex plane with a radius equal to the pole magnitude. The roots are spaced at $N + 1$ equal intervals on the ring with the exception that there exists no root at the location of the approximated pole.

TABLE II

NORMALIZED KURTOSES FOR VARIOUS SOURCE DISTRIBUTIONS.

real-valued alphabet	kurtosis	complex-valued alphabet	kurtosis
uniform BPSK	1	uniform M-PSK	1
uniform 4-PAM	1.64	uniform 16-QAM	1.32
uniform 8-PAM	1.762	uniform 64-QAM	1.381
uniform 16-PAM	1.791	uniform 256-QAM	1.395
uniform 32-PAM	1.798	uniform 1024-QAM	1.399
Gaussian	3	Gaussian	2

values. Note that a shaped source has the potential for raising the kurtosis far past that of a dense (uniform) constellation like 1024-QAM.

For shaped sources with near-Gaussian kurtoses, the CM cost surface is raised and flattened, therefore unsuited to stochastic gradient descent.

Source correlation results from the use of certain types of coding (e.g. differential encoding) or under particular operational circumstances [Treichler ASIL 91], [Axford TSP 98]. Moderate amounts of source correlation may shift the locations of local minima. Large amounts of correlation may even cause additional (false) minima to appear in the CM cost surface. Recall that any amount of source correlation violates the CM global convergence requirements. The most thorough studies on the effects of shaped and/or correlated sources appear in the works of LeBlanc, e.g. [LeBlanc Thesis 95] and [LeBlanc IJACSP 98].

As a final note, we point out that the global convergence conditions for complex-valued implementations of the CM criterion specify a circularly-symmetric source, i.e. $\mathrm{E}\{s_n^2\} = 0$. Studies have shown that violations of this requirement (e.g. from the use of a real-valued source with a complex-valued channel and/or equalizer) can result in the appearance of undesired CM minima [Papadias ICASSP 97].

D. Enjoy

With these descriptions of the literature categorization, you are now equipped to utilize the annotated bibliography in Section V to guide your own descent into the constant modulus literature. A postscript file containing the abstracts of papers in this list is provided at http://backhoe.ee.cornell.edu/BERG/bib/CM_bib.ps.

We take this opportunity to advertise THE BERGULATOR, a public-domain MATLAB-5 based software environment allowing experimentation with the CM criterion and various implementations of CMA. It can be used, for example, to generate all contour plots in this paper. THE BERGULATOR was written by Phil Schniter of Cornell University's Blind Equalization Research Group (CU-BERG), one of the primary authors of this paper. It is available from our web site: http://backhoe.ee.cornell.edu/BERG/.

IV. ACKNOWLEDGEMENTS

TABLE III

ANNOTATIONS USED IN THE BIBLIOGRAPHY AND THEIR INTPRETATIONS

Symbol	Meaning	Comments
P	Perfect Equalization	Addresses the case where any pure delay is achievable; global convergence
N	Noise	Addresses the effects of noise
L	Length	Addresses the effects of equalizer length
D	Disparity	Addresses the effects of channel disparity
S	Source	Addresses the effects of sources that are shaped, non-constant modulus or correlated
F	FSE	Fractionally-spaced equalization context
B	BSE	Baud-spaced equalization context
I	Initialization	Discusses initialization procedures for adaptive implementations
E	Excess error	Discusses sources of excess error in adaptive implementations due to nonvanishing stepsize
R	Convergence Rate	Discusses convergence rate of adaptive implementations
T	Surface Topology	Studies topology of error surface

V. ANNOTATED BIBLIOGRAPHY

The symbols in Table III are used in the annotated bibliography to classify papers consistent with our partitioning of the literature. Tables IV and V list the abbreviations used to reference journal and conference publications throughout this document.

Note that two reference lists are provided. The first list consists of the categorized CM-criterion literature, while the second list contains supportive material not directly focused on blind CM equalization. Citations from the first list in the text are typeset in Roman font, while citations from the second list appear in *Italics*.

A. CM-Minimizing Equalization Literature

REFERENCES

[Axford MIL 94] R.A. Axford, Jr., L.B. Milstein, and J.R. Zeidler, "On the misconvergence of CMA blind equalizers in the reception of PN sequences," in *Proc. IEEE Military Communications Conference* (Fort Monmouth, NJ), pp. 281-6, 2-5 Oct. 1994.
Categories: **B**, **S**, **R**

[Axford Thesis 95] R.A. Axford, Jr., "Refined techniques for blind equalization of phase-shift keyed (PSK) and quadrature amplitude modulated QAM digital communications signals," *Ph.D. dissertation*, University of California, San Diego, CA, 1995.
Categories: **B**, **S**, **E**

[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. 519-523, Feb. 1998.
Categories: **B**, **S**, **R**

[Bershad ICASSP 90] N.J. Bershad and S. Roy, "Performance of the 2-2 constant modulus (CM) adaptive algorithm for Rayleigh fading sinusoids in Gaussian noise," in *Proc. IEEE International Conference on Acoustics, Speech and Signal Processing* (Albuquerque, NM), pp. 1675-1678, Apr. 1990.
Categories: **B**, **N**

[Chan TASSP 90] C.K. Chan and J.J. Shynk, "Stationary points of the constant modulus algorithm for real Gaussian signals," *IEEE Transactions on Acoustics, Speech, and Signal Processing*, vol. 38 no. 12, pp. 2176-2181, Dec. 1990.
Categories: **B**, **S**, **T**

[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, May 1998.
Categories: **F**, **T**, **N**

[Cusani ETTRT 95] R. Cusani and A. Laurenti, "Evaluation of the constant modulus algorithm in blind equalization of three ray multipath fading channels," *European Transactions on Telecommunications and Related Technologies*, vol. 6, no.2, pp. 187-90, Feb.-Apr. 1995.

TABLE IV
JOURNAL ABBREVIATIONS USED IN THE BIBLIOGRAPHY

Abbreviation	Journal Name
ANT	Annales des Telecommunications
ATT	AT&T Technical Journal
AUTOM	Automatica
BSTJ	Bell System Technical Journal
CM	IEEE Communications Magazine
CSM	IEEE Control Systems Magazine
ETTRT	European Transactions on Telecommunications and Related Technologies
TASSP	IEEE Transactions on Acoustics, Speech, and Signal Processing
TCS	IEEE Transactions on Circuits and Systems
TCOM	IEEE Transactions on Communications
TIT	IEEE Transactions on Information Theory
TSP	IEEE Transactions on Signal Processing
IJACSP	International Journal of Adaptive Control & Signal Processing
MAROC	Journal Marocain d' Automatique, d'Informatique et de Traitement du Signal
PJR	Philips Journal of Research
PROC	Proceedings of the IEEE
SP	Signal Processing
SPM	IEEE Signal Processing Magazine
SPL	IEEE Signal Processing Letters

TABLE V
CONFERENCE ABBREVIATIONS USED IN THE BIBLIOGRAPHY

Abbreviation	Conference Name
ALL	Allerton Conference on Ciruits and System Theory
ASIL	Asilomar Conference on Signals, Systems and Computers
CDC	IEEE Conference on Decision and Control
CISS	Conference on Information Science and Systems
COST	COST 229 Workshop on Adaptive Algorithms in Communications
EUSIPCO	European Signal Processing Conference
ISCS	IEEE International Symposium on Circuits and Systems
GLOBE	IEEE Global Telecommunications Conference
GRETSI	Colloque GRETSI sur le Traitement du Signal et des Images
MIL	IEEE Military Communications Conference
SPW	IEEE Signal Processing Workshop
ICASSP	IEEE International Conference on Acoustics, Speech and Signal Processing
ICC	IEEE International Conference on Communications
ICDSP	International Conference on Digital Signal Processing
SPAWC	IEEE Signal Processing Workshop on Signal Processing Advances in Wireless Communications
SPWSSAP	IEEE Signal Processing Workshop on Statistical Signal and Array Processing
SSST	Southeastern Symposium on System Theory
SPIE	The International Society for Optical Engineering

[Cusani TCOM 95] R. Cusani and A. Laurenti, "Convergence analysis of the CMA blind equalizer," *IEEE Transactions on Communications*, vol. 43, no.2-4, pp. 1304-7, Feb.-Apr. 1995.
   Categories: **B**

[Ding CISS 89] Z. Ding, C.R. Johnson, Jr., R.A. Kennedy, and B.D.O. Anderson, "On the ill-convergence of Godard blind equalizers in data communication systems," in *Proc. Conference on Information Science and Systems* (Baltimore, MD), pp. 538-543, Mar. 1989.
   Categories: **B, L, T, I**

[Ding ICASSP 90] Z. Ding, C.R. Johnson, Jr. and R.A. Kennedy, "Admissibility in blind adaptive equalizers," in *Proc. IEEE International Conference on Acoustics, Speech and Signal Processing* (Albuquerque, NM), pp. 1707-1710, Apr. 1990.
   Categories: **B**

[Ding Thesis 90] Z. Ding, "Application aspects of blind adaptive equalizers in QAM data communcations," *Ph.D. dissertation*, Cornell University, Ithaca NY, 1990.
   Categories: **B, L, T, I**

[Ding ICASSP 91] Z. Ding, C.R. Johnson, Jr., and R.A. Kennedy, "Local convergence of 'globally convergent' blind adaptive equalization algorithms," in *Proc. IEEE International Conference on Acoustics, Speech and Signal Processing* (Toronto, Ont., Canada), pp. 1533-6, 14-17 May 1991.
   Categories: **B, L, T**

[Ding SPIE 91] Z. Ding and C.R. Johnson, Jr., "Existing gap between theory and application of blind equalization," in *Proc. The International Society for Optical Engineering* (San Diego, CA), pp. 154-165, July 1991.
   Categories: **B, L, T**

[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. 1313-1327, Sep. 1991.
   Categories: **B, L, T**

[Ding TSP 92] Z. Ding, C.R. Johnson, Jr., and R.A. Kennedy, "On the (non)existence of undesirable equilibria of Godard blind equalizers," *IEEE Transactions on Signal Processing*, vol. 40, no.10, pp. 2425-2432, Oct. 1992.
   Categories: **B, L, T**

[Ding TCS 92] Z. Ding, and R.A. Kennedy, "On the whereabouts of local minima for blind adaptive equalizers," *IEEE Transactions on Circuits and Systems II*, vol. 39, no. 2, pp. 119-123, Feb. 1992.
   Categories: **B, L, T**

[Ding TSP 93] Z. Ding, and C.R.Johnson, Jr., "On the nonvanishing stability of undesirable equilibria for FIR Godard blind equalizers," *IEEE Transactions on Signal Processing*, vol. 41, no. 5, pp. 1940-1944, May 1993.
   Categories: **B, L, T**

[Ding (Haykin) 94] Z. Ding, C.R. Johnson, Jr., and R.A. Kennedy, "Global convergence issues with linear blind adaptive equalizers," in *Blind Deconvolution*, Englewood Cliffs, NJ: Prentice-Hall, 1994, pp. 60-120.
   Categories: **B, L, T**

[Ding TSP 97] Z. Ding, "On convergence analysis of fractionally spaced adaptive blind equalizers," *IEEE Transactions on Signal Processing*, vol. 45, no.3, pp. 650-57, Mar. 1997.
   Categories: **F, P, T**

[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.
   Categories: **F, L, T**

[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. 3605-3608, 20-24 Apr. 1997.
   Categories: **F, L, T, E**

[Endres Thesis 97] T.J. Endres, "Equalizing with fractionally-spaced constant modulus and second-order-statistics blind receivers," *Ph.D. dissertation*, Cornell University, Ithaca, NY, 1997.
   Categories: **F, L, T, E**

[Evans ASIL 97] S. Evans and L. Tong, "Adaptive channel surfing re-initialization of the constant modulus algorithm," to appear in *Proc. Asilomar Conference on Signals, Systems and Computers*, (Pacific Grove, CA) Nov. 1997.
   Categories: **B, I**

[Fijalkow SPW 94] I. Fijalkow, F.L. 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.
   Categories: **F, P, T**

[Fijalkow ICASSP 95] I. Fijalkow, J.R. Treichler and C.R. Johnson., Jr, "Fractionally spaced blind equalization: Loss of channel disparity," in *Proc. IEEE International Conference on Acoustics, Speech and Signal Processing* (Detroit, MI), pp. 1988-91, 9-12 May 1995.
   Categories: **F, D**

[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.
   Categories: **F, N, D**

[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.
   Categories: **F, N, D, T**

[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. 227-231, Jan. 1998.
   Categories: **F, P, E**

[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.
   Categories: **B**, **P**, **T**, **R**

[Frater CDC 92] M.R. Frater, R.R. Bitmead, C.R. Johnson, Jr., "Escape from stable equilibria in blind adaptive equalizers," in *Proc. IEEE Conference on Decision and Control* (Tucson, AZ), pp. 1756-1761, Dec. 1992.
   Categories: **B**, **T**, **R**

[Frater ICASSP 94] M.R. Frater and C.R. Johnson, Jr., "Local minima escape transients of CMA," in *Proc. IEEE International Conference on Acoustics, Speech and Signal Processing* (Adelaide, SA, Australia), pp. 37-40, 19-22 Apr. 1994.
   Categories: **B**, **T**, **R**

[Frater AUTOM 95] M.R. Frater, R.R. Bitmead, and C.R. Johnson, Jr., "Local minima escape transients by stochastic gradient descent algorithms in blind adaptive equalizers," *Automatica*, vol. 31, no.4, pp. 637-41, Apr. 1995.
   Categories: **B**, **T**, **R**

[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. 1867-1875, Nov. 1980.
   Categories: **B**, **T**, **I**, **P**

[Gu ICASSP 93] Z. Gu and W.A. Sethares, "A geometrical view of blind equalization," in *Proc. IEEE International Conference on Acoustics, Speech and Signal Processing* (Mineapolis, MN), pp. 551-4 vol.3, 27-30 Apr. 1993.
   Categories: **B**, **T**

[Hilal ICASSP 92] K. Hilal, and P. Duhamel, "A general form for recursive adaptive algorithms leading to an exact recursive CMA," in *Proc. IEEE International Conference on Acoustics, Speech and Signal Processing* (San Francisco, CA), pp. 17-20, Mar. 1992.
   Categories: **B**, **R**

[Hilal EUSIPCO 92] K. Hilal and P. Duhamel, "A convergence study of the constant modulus algorithm leading to a normalized-CMA and a block-normalized-CMA," in *Proc. European Signal Processing Conference* (Brussels, Belgium), pp. 135-8 vol.1, 24-27 Aug. 1992.
   Categories: **B**, **R**

[Hilal EUSIPCO 94] K. Hilal and P. Duhamel, "Stability and convergence analysis of the constant modulus algorithm. Comments on finite equalization schemes and the stability of the normalized CMA," in *Proc. European Signal Processing Conference* (Edinburgh, UK), pp. 724-7, 13-16 Sep. 1994.

[Jamali ASIL 90] H. Jamali and S.L. Wood, "Error surface analysis for the complex constant modulus adaptive algorithm," in *Proc. Asilomar Conference on Signals, Systems and Computers* (Pacific Grove, CA), pp. 248-252, Nov. 1990.
   Categories: **B**, **T**

[Jamali ASIL 92] H. Jamali, S.L. Wood, and R. Cristi, "Experimental validation of the Kronecker product Godard blind adaptive algorithms," in *Proc. Asilomar Conference on Signals, Systems and Computers* (Pacific Grove, CA), pp. 1-5 vol.1, 26-28 Oct. 1992.
   Categories: **B**, **T**, **N**, **L**

[Johnson TASSP 88] C.R. Johnson, Jr., S. Dasgupta, and W.A. Sethares, "Averaging analysis of local stability of a real constant modulus algorithm adaptive filter," *IEEE Transactions on Acoustics, Speech, and Signal Processing*, vol. 36, no.6, pp. 900-10, June 1988.
   Categories: **B**

[Johnson CSM 91] C.R. Johnson, Jr., "Admissibility in blind adaptive channel equalization," *IEEE Control Systems Magazine*, vol. 11, pp. 3-15, Jan. 1991.
   Categories: **B**, **T**, **I**

[Johnson CISS 92] C.R. Johnson, Jr., P.C.E. Bennet, J.P. LeBlanc, and V. Krishnamurthy, "Blind adaptive equalizer average stationary point stability analysis with admissibility consequences," in *Proc. Conference on Information Science and Systems* (Princeton, NJ), pp. 727-732, Mar. 1992.
   Categories: **B**, **S**, **L**, **I**

[Johnson COST 92] C.R. Johnson, Jr. and J.P. LeBlanc, "Towards operational guidelines for memoryless-error-function-style blind equalizers," in *Proc. COST 229 Workshop on Adaptive Algorithms in Communications* (Bordeaux-Technopolis, France), pp. 5-17, Sep. 1992.
   Categories: **B**, **S**

[Johnson MAROC 93] C.R. Johnson, Jr., J.P. LeBlanc and V. Krishnamurthy, "Godard blind equalizer misbehavior with correlated sources," *Journal Marocain d' Automatique, d'Informatique et de Traitment du Signal*, vol. 2, pp. 1-39, June 1993.
   Categories: **B**, **S**

[Johnson IJACSP 95] C.R. Johnson, Jr. and B.D.O. Anderson, "Godard blind equalizer error surface charateristics: White, zero-mean, binary case," *International Journal of Adaptive Control & Signal Processing*, vol. 9, pp. 301-324, July-Aug. 1995.
   Categories: **B**, **T**, **L**

[Kwon SP 92] O.W. Kwon, C.K. Un, and J.C. Lee, "Performance of constant modulus adaptive digital filters for interference cancellation," *Signal Processing*, vol. 26, no. 2, pp. 185-196, Feb. 1992.
   Categories: **B**, **R**, **N**

[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. 335-340, June 1997.
   Categories: **F**, **R**, **T**, **P**

[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.
   Categories: **B**, **R**

[Larimore ASIL 84] M.G. Larimore and J.R. Treichler, "The capture properties of CMA-based interference cancellers," in *Proc. Asilomar Conference on Signals, Systems and Computers* (Pacific Grove, CA), pp. 49-52, 5-7 Nov. 1984.
   Categories: **B**, **S**, **I**, **N**

[Larimore ICASSP 85] M.G. Larimore and J.R. Treichler, "Noise capture properties of constant modulus algorithm," in *Proc. IEEE International Conference on Acoustics, Speech and Signal Processing* (Tampa, FL), pp. 1165-8, 26-29 Mar. 1985.
    Categories: **B, N, S**

[LeBlanc ICASSP 94] J.P. LeBlanc, K. Dogancay, R.A. Kennedy and C.R. Johnson, Jr., "Effects of input data correlation on the convergence of blind adaptive equalizers," in *Proc. IEEE International Conference on Acoustics, Speech and Signal Processing* (Adelaide, SA, Australia), pp. 313-16, 19-22 Apr. 1994.
    Categories: **B, R, S**

[LeBlanc ICASSP 95] J.P. LeBlanc, I. Fijalkow, B. Huber, and C.R. Johnson, Jr., "Fractionally spaced CMA equalizers under periodic and correlated inputs," in *Proc. IEEE International Conference on Acoustics, Speech and Signal Processing* (Detroit, MI), pp. 1041-4 vol.2, 9-12 May 1995.
    Categories: **F, S, T**

[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, 1995.
    Categories: **F, S, T, R**

[LeBlanc CISS 96] J.P. LeBlanc and S.W. McLaughlin, "Non-equiprobable constellation shaping and blind constant modulus algorithm equalization," in *Proc. Conference on Information Science and Systems* (Princeton, NJ), pp. 901-903, Mar. 1996.
    Categories: **F, S**

[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, 7-9 May 1996.
    Categories: **F, S, T, R**

[LeBlanc ICDSP 97] J.P. LeBlanc and C.R. Johnson, Jr., "Global CMA error surface characteristics, source statistic effects: Polytopes and manifolds," to appear in *International Conference on Digital Signal Processing*, (Santorini, Greece), 2-4 July 1997.
    Categories: **F, S, T, P**

[LeBlanc IJACSP 98] J.P. LeBlanc, I. Fijalkow, and C. Richard 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.
    Categories: **F, S, T**

[Li ASIL 94] Y. Li and Z. Ding, "Global convergence of fractionally spaced Godard equalizers," in *Proc. Asilomar Conference on Signals, Systems and Computers* (Pacific Grove, CA), pp. 617-21 vol.1, 31 Oct.-2 Nov. 1994.
    Categories: **F, P, T**

[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. 2120-2129, Sep. 1995.
    Categories: **B, T, L, I**

[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.
    Categories: **F, P, T**

[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.
    Categories: **B, T, L**

[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.
    Categories: **B, L, E**

[Lopez de V. SPW 94] F. Lopez de Victoria, A. Bosser, I. Fijalkow, C.R. Johnson, Jr., and J.R. Treichler, "Observed (mis)behavior of CMA with periodic sources: Assessment and guidelines," in *Proc. IEEE Signal Processing Workshop* (Yosemite National Park, CA), pp. 261-4, 2-5 Oct. 1994.
    Categories: **B, S**

[Meyer CISS 96] W.E. Meyer and J.P. LeBlanc, "Blind adaptive fractionally-spaced CMA in the presence of channel noise," in *Proc. Conference on Information Science and Systems* (Princeton, NJ), pp. 373-374, Mar. 1996.
    Categories: **F, N**

[Papadias ICASSP 97] C. 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. 3937-3940, 20-24 Apr. 1997.
    Categories: **F, T, S, P**

[Shynk ICASSP 90] J.J. Shynk and C.K. Chan, "A comparative analysis of the stationary points of the constant modulus algorithm based on Gaussian assumptions," in *Proc. IEEE International Conference on Acoustics, Speech and Signal Processing* (Albuquerque, NM), pp. 1249-1252, Apr. 1990.
    Categories: **B, S, T**

[Shynk ISCS 90] J.J. Shynk and C.K. Chan, "Error surfaces of the constant modulus algorithm," in *Proc. IEEE International Symposium on Circuits and Systems* (New Orleans, LA), pp. 1335-1338, May 1990.
    Categories: **B, T, S**

[Shynk TSP 93] J.J. Shynk and C.K. Chan, "Performance surfaces of the constant modulus algorithm based on a conditional Gaussian model," *IEEE Transactions on Signal Processing*, vol. 41, no. 5, pp. 1965-1969, May 1993.
    Categories: **B, S, T**

[Smith ICASSP 85] J.O. Smith and B. Freidlander, "Global convergence of the constant modulus algorithm," in *Proc. IEEE International Conference on Acoustics, Speech and Signal Processing* (Tampa, FL), pp. 1161-4, 26-29 Mar. 1985.
    Categories: **B, P, T**

[Swaminathan ICASSP 93] R. Swaminathan and J.K. Tugnait, "On improving the convergence of constant modulus algorithm adaptive filters," in *Proc. IEEE International Conference on Acoustics, Speech and Signal Processing* (Minneapolis, MN), pp. 340-3, 27-30 Apr. 1993.
   Categories: B, I

[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. 85-87, Mar. 1997.
   Categories: F, I

[Touzni ICASSP 96] A. Touzni, I. Fijalkow, and J.R. Treichler, "Fractionally-spaced CMA under channel noise," in *Proc. IEEE International Conference on Acoustics, Speech and Signal Processing* (Atlanta, GA), pp. 2674, 7-9 May 1996.
   Categories: F, N, T

[Touzni SPWSSAP 96] A. Touzni, I. Fijalkow and J.R. Treichler, "Robustness of fractionally-spaced equalization by CMA to lack of channel disparity," in *Proc. IEEE Signal Processing Workshop on Statistical Signal and Array Processing* (Corfu, Greece), pp. 144-147, 24-26 June 1996.
   Categories: F, D, T

[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. 1227-1230, Sep. 1996.
   Categories: F, R, T, P

[Touzni SPAWC 97] A. Touzni and I. Fijalkow, "Channel robust blind fractionally-spaced equalization," in *Proc. IEEE Signal Processing Workshop on Signal Processing Advances in Wireless Comm.* (Paris, France), pp. 33-36, Apr. 1997.
   Categories: F, D, I

[Touzni ICASSP 97] A. Touzni and I. Fijalkow, "Robustness of blind fractionally-spaced identification/equalization to loss of channel disparity," in *Proc. IEEE International Conference on Acoustics, Speech and Signal Processing* (Munich, Germany), pp. 3937-3940, Apr. 1997.
   Categories: F, D, T

[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.
   Categories: B, T, I, P

[Treichler ICASSP 85] J.R. Treicher and M.G. Larimore, "Convergence rates for the constant modulus algorithm with sinusoidal inputs," in *Proc. IEEE International Conference on Acoustics, Speech and Signal Processing* (Tampa, FL), pp. 1157-60, 26-29 Mar. 1985.
   Categories: B, R, S

[Treichler TASSP 85a] J.R. Treichler and M.G. Larimore, "The tone capture properties of CMA-based interference suppressors," *IEEE Transactions on Acoustics, Speech, and Signal Processing*, vol. ASSP-33, no.4, pp. 946-58, Aug. 1985.
   Categories: B, S

[Treichler TASSP 85b] J.R. Treichler and M.G. Larimore, "New processing techniques based on the constant modulus adaptive algorithm," *IEEE Transactions on Acoustics, Speech, and Signal Processing*, vol. ASSP-33, no.2, pp. 420-31, Apr. 1985.
   Categories: B, F

[Treichler ASIL 91] J.R. Treichler, V. Wolff, and C.R. Johnson, Jr., "Observed misconvergence in the constant modulus adaptive algorithm," in *Proc. Asilomar Conference on Signals, Systems and Computers* (Pacific Grove, CA), pp. 663-667, Nov. 1991.
   Categories: S

[Treichler SPAWC 97] J.R. Treichler, L. Tong, I. Fijalkow, C.R. Johnson, Jr., and C.U. Berg, "On the current shape of FSE-CMA behavior theory," in *Proc. IEEE Signal Processing Workshop on Signal Processing Advances in Wireless Comm.* (Paris, France), pp. 105-108, Apr. 1997.
   Categories: F, N, L

[Tugnait ICC 94] J.K. Tugnait, "A parallel multimodel: CMA/Godard adaptive filter bank approach to fractionally-spaced blind adaptive equalization," in *Proc. IEEE International Conference on Communications* (New Orleans, LA), pp. 549-53 vol.1, 1-5 May 1994.
   Categories: F

[Tugnait ICASSP 95] J.K. Tugnait, "On fractionally-spaced blind adaptive equalization under symbol timing offsets using Godard and related equalizers," in *Proc. IEEE International Conference on Acoustics, Speech and Signal Processing* (Detroit, MI), pp. 1976-1979 vol.3, 9-12 May 1995.
   Categories: F

[Tugnait TSP 96] J.K. Tugnait, "On fractionally spaced blind adaptive equalization under symbol timing offsets using Godard and related equalizers," *IEEE Transactions on Signal Processing*, vol. 44, no.7, pp. 1817-21, July 1996.
   Categories: B, F

[Vembu TSP 94] S. Vembu, S. Verdu, R.A. Kennedy, and W.A. Sethares, "Convex cost functions in blind equalization," *IEEE Transactions on Signal Processing*, vol. 42, no. 8, pp. 1952-1960, Aug. 1994.
   Categories: B

[Wu SSST 95] M. Wu and F. Cornett, "Discrete-time and continuous-time constant modulus algorithm analysis," in *Proc. Southeastern Symposium on System Theory* (Starkville, MS), pp. 504-8, 12-14 Mar. 1995.
   Categories: B, R, E

[Zeng CISS 96] H.H. Zeng and L. Tong, "On the performance of CMA in the presence of noise — Some new results on blind channel estimation: Performance and algorithms," in *Proc. Conference on Information Science and Systems* (Princeton, NJ), pp. 890-894, Mar. 1996.
   Categories: F, N, T

[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. 305-309, Nov. 1996.
   Categories: F, N, T, D, L

[Zeng ICASSP 97] H.H. Zeng and L. Tong, "The MSE performance of constant modulus receivers," in *Proc. IEEE International Conference on Acoustics, Speech and Signal Processing* (Munich, Germany), pp. 3577-3580, Apr. 1997.
    Categories: **B**, **F**, **N**, **T**
[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.
    Categories: **F**, **N**, **T**, **D**, **L**
[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.
[Zervas SPIE 91] E. Zervas, J.G. Proakis, and V. Eyuboglu, "Effect of constellation shaping on blind equalization," in *Proc. The International Society for Optical Engineering* (San Diego, CA), pp. 1393-1397, 1991.
    Categories: **B**, **S**

B. Related Material

REFERENCES

[*Bergmans PJR 87*] J.W.M. Bergmans and A.J.E.M. Janssen, "Robust data equalization, fractional tap spacing and the Zak transform," *Philips Journal of Research*, vol. 42, no. 4, pp. 351-398, 1987.
[*Ding SPL 96*] Z. Ding, "Characteristics of band-limited channels unidentifiable from second-order cyclostationary statistics," *IEEE Signal Processing Letters*, vol. 3, no. 5, pp. 150-152, May 1996.
[*Forney CM 96*] G.D. Forney, L. Brown, M.V. Eyuboglu, J.L. Moran III, "The V.34 high-speed modem standard," *IEEE Communications Magazine*, vol. 34, no. 12, pp. 28-33, Dec. 1996.
[*Gitlin BSTJ 81*] R.D. Gitlin, "Fractionally-spaced equalization: An improved digital transversal equalizer," *Bell System Technical Journal*, vol. 60, pp. 275-296, Feb. 1981.
[*Gitlin Book 92*] R.D. Gitlin, J.F. Hayes, and S.B. Weinstein, *Data Communications Principles*, New York, NY: Plenum Press, 1992.
[*Haykin Book 96*] S. Haykin, *Adaptive Filter Theory*, 3rd ed., Englewood Cliffs, NJ: Prentice-Hall, 1996.
[*Johnson ASIL 95*] C.R. Johnson, Jr. et al., "On fractionally-spaced equalizer design for digital microwave radio channels," in *Proc. Asilomar Conference on Signals, Systems and Computers* (Pacific Grove, CA), pp. 290-294, Oct. 1995.
[*Kailath Book 80*] T. Kailath, *Linear Systems*, Englewood Cliffs, NJ: Prentice-Hall, 1980.
[*Larimore ASIL 85*] M.G. Larimore and M.J. Goodman, "Implementation of the constant modulus algorithm at RF bandwidths," in *Proc. Asilomar Conference on Signals, Systems and Computers* (Pacific Grove, CA), pp. 626-630, Nov. 1985.
[*Lee Book 94*] E.A. Lee and D.G. Messerschmitt, *Digital Communication*, 2nd ed., Boston, MA: Kluwer Academic Publishers, 1994.
[*Lucky BSTJ 66*] R.W. Lucky, "Techniques for adative equalization of digital communication systems," *Bell System Technical Journal*, vol. 45, pp. 255-286, Feb. 1966.
[*Lucky ALL 69*] R.W. Lucky, "Signal filtering with the transversal equalizer," in *Proc. Allerton Conference on Ciruits and System Theory* (Allerton, IL), pp. 792, Oct. 1969.
[*Luenberger Book 90*] D.G. Luenberger, *Optimization by Vector Space Methods*, 2nd ed., New York, NY: Wiley, 1990.
[*Macchi ANT 75*] O. Macchi and L. Guidoux, "A new equalizer: the double-sampling equalizer," *Annales des Telecommunications*, vol. 30, no. 9-10, pp. 331-8, Sep. 1975.
[*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. 516-525, Feb. 1995.
[*Proakis Book 95*] J.G. Proakis, *Digital Communications*, 3rd ed., New York, NY: McGraw-Hill, 1995.
[*Qureshi TCOM 73*] S.U.H. Qureshi, "Adjustment of the position of the reference tap of an adaptive equalizer," *IEEE Transactions on Communications*, vol. xxx, pp. 1046-1052, Sep. 1973.
[*Qureshi PROC 85*] S.U.H. Qureshi, "Adaptive equalization," *Proceedings of the IEEE*, vol. 73, no. 9, pp. 1349-1387, Sep. 1985.
[*Sato TCOM 75*] Y. Sato, "A method of self-recovering equalization for multilevel amplitude-modulations systems," *IEEE Transactions on Communications*, vol. 23, pp. 679-682, June 1975.
[*Strang Book 88*] G. Strang, *Linear Algebra and its Applications*, 3rd ed., Fort Worth, TX: Harcourt Brace Jovanovich, 1988.
[*Tong CISS 92*] L. Tong, "A fractionally spaced adaptive blind equalizer," in *Proc. Conference on Information Science and Systems* (Princeton, NJ), pp. 711-716, Mar. 1992.
[*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. 329-334, Jan. 1995.
[*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," to appear in *Proceedings of the IEEE*, Sep. 1998.
[*Tugnait TIT 95*] J.K. Tugnait, "On blind identifiability of multipath channels using fractional sampling and 2nd-order cyclostationary statistics," *IEEE Transactions on Information Theory*, vol. 41, no. 1, pp. 308-311, Jan. 1995.
[*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. 856-864, Aug. 1976.
[*Widrow Book 85*] B. Widrow and S.D. Stearns, *Adaptive Signal Processing*, Engelwood Cliffs, NJ: Prentice Hall, Inc., 1985.
[*Wolff MIL 88*] V.G. Wolff, R.P. Gooch, and J.R.Treichler, "Specification and development of an equalizer-demodulator for wideband digital microwave radio signals," in *Proc. IEEE Military Communications Conference* (San Diego, CA), pp. 461-467, Oct. 1988.

I. FRACTIONALLY-SPACED SYSTEM MODEL

We denote the combined (LTI) channel and pulse-shape impulse response by $c(t)$ and the baseband additive channel noise process by $w(t)$. The continuous-time baseband representation of the waveform seen by the receiver can then be described by

$$r(t) = \sum_{n=-\infty}^{\infty} s_n c(t - nT - t_0) + w(t) \tag{26}$$

for symbol sequence $\{s_n\}$, baud interval T, and arbitrary time delay t_0. Sampling[15] the received signal every $T/2$ seconds at the receiver, we denote the sampled received sequence by

$$r(k\tfrac{T}{2}) = \sum_{n=-\infty}^{\infty} s_n c(k\tfrac{T}{2} - nT - t_0) + w(k\tfrac{T}{2}). \tag{27}$$

The output x_k of a length $2N$ FIR fractionally-spaced equalizer (FSE) with tap spacing of $T/2$ can be written as a $T/2$-rate convolution with the sampled received sequence:

$$x_k = \sum_{i=0}^{2N-1} f_i \, r((k-i)\tfrac{T}{2}). \tag{28}$$

The choice of an even number of equalizer taps is chosen for notational simplicity. Now suppose that only the "odd" fractionally-spaced equalizer output samples are retained in a decimation by two (i.e. $k = 2n+1$ for $n = 0, 1, 2, \ldots$). The decimated equalized output sequence y_n^{odd} then becomes

$$y_n^{\mathrm{odd}} = x_{2n+1} \tag{29}$$

$$= \sum_{i=0}^{2N-1} f_i \, r(nT - i\tfrac{T}{2} + \tfrac{T}{2}) \tag{30}$$

$$= \sum_{i=0}^{N-1} \left(f_{2i} \, r((n-i)T + \tfrac{T}{2}) + f_{2i+1} \, r((n-i)T) \right). \tag{31}$$

Note that a similar procedure can be carried out for even-indexed output sampling (i.e. $k = 2n$ and $y_n^{\mathrm{even}} = x_{2n}$). An illustration of the setup described above appears in Figure 3.

A. Multichannel Model

From (31) we observe that the decimated output y_n^{odd} can be considered the sum of two baud-spaced convolutions:

$$y_n^{\mathrm{odd}} = \sum_{i=0}^{N-1} \left(f_i^{\mathrm{even}} \, r_{n-i}^{\mathrm{odd}} + f_i^{\mathrm{odd}} \, r_{n-i}^{\mathrm{even}} \right), \tag{32}$$

where

$$f_n^{\mathrm{even}} = f_{2n}, \quad f_n^{\mathrm{odd}} = f_{2n+1}, \quad r_n^{\mathrm{even}} = r(nT), \quad \text{and} \quad r_n^{\mathrm{odd}} = r(nT + \tfrac{T}{2}). \tag{33}$$

---

[15]The noise and channel are considered band-limited assuming antialias filtering is done prior to $T/2$-spaced sampling at the receiver.

We refer to r_n^{even} and r_n^{odd} as the "even" and "odd" received sequences and to f_n^{even} and f_n^{odd} as the "even" and "odd" sub-equalizers.

Defining the even and odd baud-rate channel response samples

$$c_n^{\text{even}} = c(nT - t_0) \quad \text{and} \quad c_n^{\text{odd}} = c(nT + \tfrac{T}{2} - t_0), \tag{34}$$

and channel noise samples $w_n^{\text{even}} = w(2n\tfrac{T}{2})$ and $w_n^{\text{odd}} = w((2n+1)\tfrac{T}{2})$ (for non-negative integers n), we can confirm that they are related to the received subsequences in a straightforward manner:

$$r_n^{\text{even}} = \sum_l s_l\, c_{n-l}^{\text{even}} + w_n^{\text{even}}, \tag{35}$$

$$r_n^{\text{odd}} = \sum_l s_l\, c_{n-l}^{\text{odd}} + w_n^{\text{odd}}. \tag{36}$$

These expressions allow us to rewrite the decimated equalizer output in terms of the baud-spaced symbol sequence.

It is important to note that the arbitrary delay t_0 has been incorporated into our definitions of the channel response samples. This implies that the "even" and "odd" subchannel classifications are merely notational and have no real physical significance. Furthermore, the inclusion of arbitrary delay implies that our convention of retaining the odd-indexed (as opposed to the even-indexed) decimated equalizer output samples also lacks practical significance. In this spirit, we drop the "odd" notation on y_n^{odd} and simply refer to the baud-spaced system output samples as y_n. Here we are seeing evidence for the inherent baud-synchronization capabilities of a FSE (not characteristic of BSEs).

Substituting the received subsequence expressions (35) and (36) into (32),

$$y_n = \sum_{i=0}^{N-1} f_i^{\text{even}} \left(\sum_l s_l\, c_{n-i-l}^{\text{odd}} + w_{n-i}^{\text{odd}} \right) + \sum_{i=0}^{N-1} f_i^{\text{odd}} \left(\sum_l s_l\, c_{n-i-l}^{\text{even}} + w_{n-i}^{\text{even}} \right) \tag{37}$$

$$= s_n \star \left(f_n^{\text{even}} \star c_n^{\text{odd}} + f_n^{\text{odd}} \star c_n^{\text{even}} \right) + f_n^{\text{even}} \star w_n^{\text{odd}} + f_n^{\text{odd}} \star w_n^{\text{even}}, \tag{38}$$

where the "$\star$" indicates convolution. The relationships between the source, noise, sub-equalizers, and subchannels described above appears in the multichannel model of Figure 5.

Consider for a moment the noiseless case. The impulse response h_n from transmitted source to baud-spaced equalizer output follows immediately from consideration of s_n as the Kronecker delta sequence δ_n. Thus we conclude that

$$h_n = f_n^{\text{even}} \star c_n^{\text{odd}} + f_n^{\text{odd}} \star c_n^{\text{even}}. \tag{39}$$

This impulse response leads directly to a transfer function $H(z^{-1})$ with unit delay (z^{-1}) of duration T:

$$H(z^{-1}) = F_{\text{even}}(z^{-1})C_{\text{odd}}(z^{-1}) + F_{\text{odd}}(z^{-1})C_{\text{even}}(z^{-1}). \tag{40}$$

Note that the perfect zero-forcing system $H(z) = z^{-\delta}$ (with non-negative integer delay δ), leads to the Bezout relationship [*Kailath Book 80*]:

$$z^{-\delta} = F_{\text{even}}(z^{-1})C_{\text{odd}}(z^{-1}) + F_{\text{odd}}(z^{-1})C_{\text{even}}(z^{-1}). \tag{41}$$

B. Multirate Model

To show that the multirate model of Figure 4 also originates from the fractionally-spaced communication system of Figure 3, we show that the fractionally-spaced equalizer output $\{x_k\}$ in (28) can be written in terms of a zero-filled version of the source sequence, $\{a_k\}$,

$$a_k = \begin{cases} s_{\frac{k}{2}} & \text{for } k \text{ even,} \\ 0 & \text{for } k \text{ odd.} \end{cases} \tag{42}$$

as depicted in Figure 4. Rewriting (27) as

$$r(k\tfrac{T}{2}) = \sum_{l=-\infty}^{\infty} a_l\, c((k-l)\tfrac{T}{2} - t_0) + w_k, \qquad (43)$$

we see upon its substitution into (28) that

$$x_k = \sum_{i=0}^{2N-1} f_i \left(\sum_l a_l\, c((k-i-l)\tfrac{T}{2} - t_0) + w_{k-i} \right) \qquad (44)$$

$$= \sum_{i=0}^{2N-1} f_i \left(\sum_l a_l\, c_{k-i-l} + w_{k-i} \right) \qquad (45)$$

$$= f_k \star (a_k \star c_k + w_k), \qquad (46)$$

where the fractionally-spaced channel response samples c_k are defined such that $c_k = c(k\tfrac{T}{2} - t_0)$.

At this point we can observe that, in the noiseless case, the fractionally-spaced system impulse response h_k^{FS} becomes

$$h_k^{\mathrm{FS}} = f_k \star c_k. \qquad (47)$$

Note from (39) that only half of the terms in the fractionally-spaced impulse response (47) are directly relevant to the system output since the fractionally-spaced output, $\{x_k\}$, is later decimated by two.

C. The Subchannel Disparity Condition

The Bezout equation (41) leads directly to the perfect equalization requirement concerning subchannel roots. Specifically, for the existence of a (finite-length) zero-forcing equalizer, the subchannel polynomials, $C_{\mathrm{even}}(z^{-1})$ and $C_{\mathrm{odd}}(z^{-1})$, must not share a common root.

The existence of perfectly equalizing sub-equalizer polynomials $F_{\mathrm{even}}(z^{-1})$ and $F_{\mathrm{odd}}(z^{-1})$ implies that (41) can be satisfied. For example, if the subchannels share one root, a common polynomial $G(z^{-1}) = g_0 + g_1 z^{-1}$ can be factored out of both $C_{\mathrm{even}}(z^{-1})$ and $C_{\mathrm{odd}}(z^{-1})$, leaving $\bar{C}_{\mathrm{even}}(z^{-1})$ and $\bar{C}_{\mathrm{odd}}(z^{-1})$, respectively. The perfect equalization relationship would then become

$$z^{-\delta} = G(z^{-1}) \left(F_{\mathrm{even}}(z^{-1})\bar{C}_{\mathrm{odd}}(z^{-1}) + F_{\mathrm{odd}}(z^{-1})\bar{C}_{\mathrm{even}}(z^{-1}) \right), \qquad (48)$$

but this is contradicted by the fact that there is no finite-length polynomial $F_{\mathrm{even}}(z^{-1})\bar{C}_{\mathrm{odd}}(z^{-1}) + F_{\mathrm{odd}}(z^{-1})\bar{C}_{\mathrm{even}}(z^{-1})$ that when multiplied by $G(z^{-1})$ results in the delay operator $z^{-\delta}$.

However, $F_{\mathrm{even}}(z^{-1})$ and $F_{\mathrm{odd}}(z^{-1})$ can be chosen so that (48) is approximated, in which case the following relationship is satisfied:

$$z^{-\delta} G^{-1}(z^{-1}) \approx F_{\mathrm{even}}(z^{-1})\bar{C}_{\mathrm{odd}}(z^{-1}) + F_{\mathrm{odd}}(z^{-1})\bar{C}_{\mathrm{even}}(z^{-1}). \qquad (49)$$

In other words, the FSE combines with the non-common-root component of the channel to approximate the (IIR) inverse of the (T-spaced) common root component.

D. On The Independence of Fractionally-Sampled Channel Noise

A typical assumption on the (baseband equivalent) channel noise $w(t)$ is that it is well modeled by a zero-mean, circularly-symmetric Gaussian process [*Lee Book 94*]. In many situations, $w(t)$ is also assumed to have a flat wideband power spectrum. Does this imply that the fractionally-sampled noise process $\{w_k\}$ will also be white? Under these conditions, $\{w_k\}$ will only be white when the anti-alias filters prior to $T/2$-spaced sampling satisfy a rate $2/T$ Nyquist criterion. In practice, this criterion is satisfied by anti-alias filters that are power-symmetric about the frequency $1/T$ Hz. If, for example, the filtering prior to equalization is matched to the pulse shape of the transmitted signal, then $\{w_k\}$ will not be white.

II. The Constant Modulus Cost Function

Below, we provide the general formulation of the CM cost function for a complex i.i.d. zero-mean source and complex baseband channel in additive white zero-mean noise. We will assume that each member of the symbol alphabet is equi-probable in the source sequence. Furthermore, we also assume that the receiver sampling clock is frequency synchronous (a fixed time offset is allowed) with the source symbol clock. In practice, this is a reasonable assumption since the symbol clock can often be extracted by computing the square magnitude of the received signal (commonly known as envelope detection). Given these assumptions, we follow the general formulation of the CM cost function with expressions for the specific cases of PAM, PSK, and QAM input signals.

In addition to the previously introduced notation we will use the following definitions.

$$\kappa_s = \frac{\mathrm{E}\{|s_n|^4\}}{(\mathrm{E}\{|s_n|^2\})^2}, \text{ the normalized kurtosis of } \{s_n\}, \tag{50}$$

$$\gamma = \frac{\mathrm{E}\{|s_n|^4\}}{\mathrm{E}\{|s_n|^2\}}, \quad \text{the dispersion constant of } \{s_n\}, \tag{51}$$

$$\|\mathbf{h}\|_2^2 = \sum_{n=0}^{P-1} |h_n|^2, \quad \text{the squared } \ell_2\text{-norm of } \mathbf{h}. \tag{52}$$

Note that $\gamma = \sigma_s^2 \kappa_s$. Following the presentation of the FS system model in Section I-A and Appendix A, we can redefine the equalizer output using (4) and (13). This results in

$$y_n = \mathbf{h}^t \mathbf{s}(n) + \mathbf{f}^t \mathbf{w}(n). \tag{53}$$

The Constant Modulus (CM) cost function is

$$
\begin{aligned}
J_{\mathrm{CM}} &= \mathrm{E}\{(\gamma - |y_n|^2)^2\} \tag{54}\\
&= \mathrm{E}\{|y_n|^4\} - 2\gamma \mathrm{E}\{|y_n|^2\} + \gamma^2 \\
&= \mathrm{E}\{|y_n|^4\} - 2\sigma_s^2 \kappa_s \mathrm{E}\{|y_n|^2\} + \sigma_s^4 \kappa_s^2. \tag{55}
\end{aligned}
$$

In order to analyze J_{CM}, we will first expand $|y_n|^2$, using equation (53). For convenience we will temporarily let $A_n = \mathbf{h}^t \mathbf{s}(n)$ and $B_n = \mathbf{f}^t \mathbf{w}(n)$, where $y_n = A_n + B_n$. Using the assumptions of mutually independent zero-mean noise and source sequences, we note that A_n and B_n are also independent and zero-mean, i.e.:

$$\mathrm{E}\{A_n\} = \mathbf{h}^t \mathrm{E}\{\mathbf{s}(n)\} = 0, \quad \mathrm{E}\{B_n\} = \mathbf{f}^t \mathrm{E}\{\mathbf{w}(n)\} = 0, \quad \text{and } \mathrm{E}\{A_n B_n\} = \mathrm{E}\{A_n\}\mathrm{E}\{B_n\}.$$

With these assumptions, we arrive at:

$$
\begin{aligned}
\mathrm{E}\{|y_n|^2\} &= \mathrm{E}\{|A_n|^2\} + \mathrm{E}\{A_n\}\mathrm{E}\{B_n^*\} + \mathrm{E}\{A_n^*\}\mathrm{E}\{B_n\} + \mathrm{E}\{|B_n|^2\} \\
&= \mathrm{E}\{|A_n|^2\} + \mathrm{E}\{|B_n|^2\}. \tag{56}
\end{aligned}
$$

Expanding $|A_n|^2$ and $|B_n|^2$, we have that

$$\mathrm{E}\{|y_n|^2\} = \sigma_s^2 \|\mathbf{h}\|_2^2 + \sigma_w^2 \|\mathbf{f}\|_2^2. \tag{57}$$

The same approach can be used to examine $\mathrm{E}\{|y_n|^4\}$, which leads to the following equation.

$$
\begin{aligned}
\mathrm{E}\{|y_n|^4\} &= \mathrm{E}\{|A_n|^4\} + \mathrm{E}\{A_n^2\}\mathrm{E}\{(B_n^*)^2\} + 4\mathrm{E}\{|A_n|^2\}\mathrm{E}\{|B_n|^2\} \tag{58}\\
&\quad + \mathrm{E}\{B_n^2\}\mathrm{E}\{(A_n^*)^2\} + \mathrm{E}\{|B_n|^4\}. \tag{59}
\end{aligned}
$$

Due to space limitations, we omit the details of the derivation of $E\{|y_n|^4\}$ but mention the following properties used in the derivation.

• The second order terms are relatively easy to compute; they involve summations of source (and noise) terms of the form $E\{s_{n-i}s_{n-l}\}$, $E\{s_{n-i}s_{n-l}^*\}$, or $E\{s_{n-i}^*s_{n-l}^*\}$.

• The fourth order terms are more difficult to compute, but each of the source (and noise) terms are of the form $E\{s_{n-i}s_{n-l}^*s_{n-m}s_{n-j}^*\}$.

Any of the expectations not involving an even power (2 or 4) will vanish because the source and noise are both zero-mean and white. After a considerable amount of algebra we arrive at the following expression for $E\{|y_n|^4\}$. Noting that $E\{s_n^2\}$ and $E\{w_n^2\}$ are independent of n, we will denote expectations of this form by $E\{s^2\}$ and $E\{w^2\}$, respectively.

$$
\begin{aligned}
E\{|y_n|^4\} &= \kappa_s \sigma_s^4 \sum_{i=0}^{P-1} |h_i|^4 + 2\sigma_s^4 \sum_{i=0}^{P-1} \sum_{m=0,m\neq i}^{P-1} |h_i|^2|h_m|^2 + |E\{s^2\}|^2 \sum_{i=0}^{P-1} \sum_{j=0,j\neq i}^{P-1} h_i^2(h_j^*)^2 \\
&\quad + \kappa_w \sigma_w^4 \sum_{i=0}^{2N-1} |f_i|^4 + 2\sigma_w^4 \sum_{i=0}^{2N-1} \sum_{m=0,m\neq i}^{2N-1} |f_i|^2|f_m|^2 + |E\{w^2\}|^2 \sum_{i=0}^{2N-1} \sum_{j=0,j\neq i}^{2N-1} f_i^2(f_j^*)^2 \\
&\quad + (E\{s^2\} \sum_{i=0}^{P-1} h_i^2)(E\{w^2\} \sum_{i=0}^{2N-1} f_i^2)^* + 4\sigma_s^2\sigma_w^2 \|\mathbf{h}\|_2^2 \|\mathbf{f}\|_2^2 \\
&\quad + (E\{s^2\} \sum_{i=0}^{P-1} h_i^2)^*(E\{w^2\} \sum_{i=0}^{2N-1} f_i^2).
\end{aligned}
\tag{60}
$$

We define the noise kurtosis κ_w analogous to the source kurtosis κ_s in (50). Substituting (57) and (60) into (55) we have the final expansion of the cost function.

$$
\begin{aligned}
J_{\text{CM}} &= \kappa_s \sigma_s^4 \sum_{i=0}^{P-1} |h_i|^4 + 2\sigma_s^4 \sum_{i=0}^{P-1} \sum_{m=0,m\neq i}^{P-1} |h_i|^2|h_m|^2 + |E\{s^2\}|^2 \sum_{i=0}^{P-1} \sum_{j=0,j\neq i}^{P-1} h_i^2(h_j^*)^2 \\
&\quad + \kappa_w \sigma_w^4 \sum_{i=0}^{2N-1} |f_i|^4 + 2\sigma_w^4 \sum_{i=0}^{2N-1} \sum_{m=0,m\neq i}^{2N-1} |f_i|^2|f_m|^2 + |E\{w^2\}|^2 \sum_{i=0}^{2N-1} \sum_{j=0,j\neq i}^{2N-1} f_i^2(f_j^*)^2 \\
&\quad + (E\{s^2\} \sum_{i=0}^{P-1} h_i^2)(E\{w^2\} \sum_{i=0}^{2N-1} f_i^2)^* + 4\sigma_s^2\sigma_w^2 \|\mathbf{h}\|_2^2 \|\mathbf{f}\|_2^2 \\
&\quad + (E\{s^2\} \sum_{i=0}^{P-1} h_i^2)^*(E\{w^2\} \sum_{i=0}^{2N-1} f_i^2) - 2\sigma_s^2\kappa_s(\sigma_s^2 \|\mathbf{h}\|_2^2 + \sigma_w^2 \|\mathbf{f}\|_2^2) + \sigma_s^4\kappa_s^2.
\end{aligned}
\tag{61}
$$

We will now consider how various restrictions on the source and noise simplify this equation.

A. PAM source, real-valued channel

For PAM, the source symbols s_n are real-valued, so that $E\{|s_n|^2\} = E\{s_n^2\} = \sigma_s^2$. Furthermore, if w_n, $\mathbf{f}$ and $\mathbf{h}$ are real-valued, we have $E\{|w_n|^2\} = E\{w_n^2\} = \sigma_w^2$, $f_i^2 = (f_i^*)^2 = |f_i|^2$ and $h_i^2 = (h_i^*)^2 = |h_i|^2$.

Thus, we have that, for a real-valued source and real-valued channel, Equation (61) reduces to

$$
\begin{aligned}
J_{\mathrm{CM}}|_{\mathrm{PAM}} &= \kappa_s \sigma_s^4 \sum_{i=0}^{P-1} h_i^4 + 3\sigma_s^4 \sum_{i=0}^{P-1} \sum_{m=0, m\neq i}^{P-1} h_i^2 h_m^2 \\
&+ \kappa_w \sigma_w^4 \sum_{i=0}^{2N-1} f_i^4 + 3\sigma_w^4 \sum_{i=0}^{2N-1} \sum_{m=0, m\neq i}^{2N-1} f_i^2 f_m^2 \\
&+ \sigma_s^2 \sigma_w^2 \|\mathbf{h}\|_2^2 \|\mathbf{f}\|_2^2 + 4\sigma_s^2 \sigma_w^2 \|\mathbf{h}\|_2^2 \|\mathbf{f}\|_2^2 + \sigma_s^2 \sigma_w^2 \|\mathbf{h}\|_2^2 \|\mathbf{f}\|_2^2 \\
&- 2\sigma_s^2 \kappa_s (\sigma_s^2 \|\mathbf{h}\|_2^2 + \sigma_w^2 \|\mathbf{f}\|_2^2) + \sigma_s^4 \kappa_s{}^2.
\end{aligned}
$$

Noting that

$$
\sum_{i=0}^{P-1} \sum_{m=0, m\neq i}^{P-1} h_i^2 h_m^2 = \|\mathbf{h}\|_2^4 - \sum_{i=0}^{P-1} h_i^4,
$$

and summing like terms, we arrive at:

$$
\begin{aligned}
J_{\mathrm{CM}}|_{\mathrm{PAM}} &= \sigma_s^4(\kappa_s - 3) \sum_{i=0}^{P-1} h_i^4 + 3\sigma_s^4 \|\mathbf{h}\|_2^4 + \sigma_w^4(\kappa_w - 3) \sum_{i=0}^{2N-1} f_i^4 + 3\sigma_w^4 \|\mathbf{f}\|_2^4 \\
&+ 6\sigma_s^2 \sigma_w^2 \|\mathbf{h}\|_2^2 \|\mathbf{f}\|_2^2 - 2\sigma_s^2 \kappa_s (\sigma_s^2 \|\mathbf{h}\|_2^2 + \sigma_w^2 \|\mathbf{f}\|_2^2) + \sigma_s^4 \kappa_s{}^2. \qquad (62)
\end{aligned}
$$

Note that if the noise is Gaussian, $\kappa_w = 3$, and the third term in (62) is zero.

A.1 BPSK Source, real-valued channel

Considering the sub-case of a BPSK source in a real-valued channel results in further simplifications. For BPSK, $\kappa_s = \sigma_s^4 = 1$, which implies that equation (62) reduces to:

$$
\begin{aligned}
J_{\mathrm{CM}}|_{\mathrm{BPSK}} &= -2 \sum_{i=0}^{P-1} h_i^4 + 3 \|\mathbf{h}\|_2^4 + \sigma_w^4(\kappa_w - 3) \sum_{i=0}^{2N-1} f_i^4 + 3\sigma_w^4 \|\mathbf{f}\|_2^4 \\
&+ 6\sigma_w^2 \|\mathbf{h}\|_2^2 \|\mathbf{f}\|_2^2 - 2(\|\mathbf{h}\|_2^2 + \sigma_w^2 \|\mathbf{f}\|_2^2) + 1. \qquad (63)
\end{aligned}
$$

In the absence of noise, (63) is the equation given in [Johnson IJACSP 95].

A.2 Complex-valued rotationally invariant noise

If we make the assumption that the (complex) noise is rotationally invariant, i.e. $p(w = \rho e^{j\theta}) = p(|w| = \rho)/2\pi$ for all $\theta \in [0, 2\pi]$, then we have that $\mathrm{E}\{w^n\} = \mathrm{E}\{\rho^n\}\mathrm{E}\{e^{jn\theta}\} = 0$, for $n = 1, 2, \ldots$. Using this assumption the cost function reduces to (64).

$$
\begin{aligned}
J_{\mathrm{CM}}|_{\text{r.i. noise}} &= \kappa_s \sigma_s^4 \sum_{i=0}^{P-1} |h_i|^4 + 2\sigma_s^4 \sum_{i=0}^{P-1} \sum_{m=0, m\neq i}^{P-1} |h_i|^2 |h_m|^2 + |\mathrm{E}\{s^2\}|^2 \sum_{i=0}^{P-1} \sum_{j=0, j\neq i}^{P-1} h_i^2 (h_j^*)^2 \\
&+ \kappa_w \sigma_w^4 \sum_{i=0}^{2N-1} |f_i|^4 + 2\sigma_w^4 \sum_{i=0}^{2N-1} \sum_{m=0, m\neq i}^{2N-1} |f_i|^2 |f_m|^2 \\
&+ 4\sigma_s^2 \sigma_w^2 \|\mathbf{h}\|_2^2 \|\mathbf{f}\|_2^2 - 2\sigma_s^2 \kappa_s (\sigma_s^2 \|\mathbf{h}\|_2^2 + \sigma_w^2 \|\mathbf{f}\|_2^2) + \sigma_s^4 \kappa_s{}^2. \qquad (64)
\end{aligned}
$$

For the remaining derivations we will make the assumption of rotationally invariant noise.
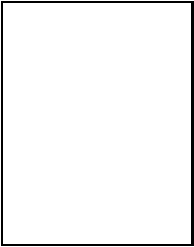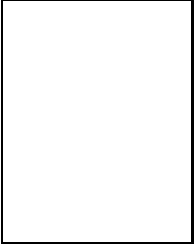
A.3 PSK Source

For PSK symbols, $s_n \in \{e^{j2\pi m/2^M}\}$, $m \in \{0, 1, \ldots, 2^M - 1\}$ (where $j = \sqrt{-1}$), we note that $\sigma_s^2 = \mathrm{E}\{|s|^4\} = \kappa_s = \sigma_s^4 = 1$. Thus, (64) simplifies to

$$
\begin{aligned}
J_{\mathrm{CM}}|_{\mathrm{PSK}} \;=\;& \sum_{i=0}^{P-1} |h_i|^4 + 2 \sum_{i=0}^{P-1} \sum_{m=0,m\neq i}^{P-1} |h_i|^2 |h_m|^2 + |\mathrm{E}\{s^2\}|^2 \sum_{i=0}^{P-1} \sum_{j=0,j\neq i}^{P-1} h_i^2 (h_j^*)^2 \\
&+ \kappa_w \sigma_w^4 \sum_{i=0}^{2N-1} |f_i|^4 + 2\sigma_w^4 \sum_{i=0}^{2N-1} \sum_{m=0,m\neq i}^{2N-1} |f_i|^2 |f_m|^2 \\
&+ 4\sigma_w^2 \|\mathbf{h}\|_2^2 \|\mathbf{f}\|_2^2 - 2(\ \|\mathbf{h}\|_2^2 + \sigma_w^2 \|\mathbf{f}\|_2^2) + 1.
\end{aligned}
\tag{65}
$$

A.4 QAM Source

For 90 degree rotationally invariant QAM (i.e. for every member q_m in the QAM alphabet, $\{jq_m, -q_m, -jq_m\}$ are equally likely members of the alphabet), we have that $\mathrm{E}\{s^2\} = 0$ and equation (64) reduces to

$$
\begin{aligned}
J_{\mathrm{CM}}|_{\mathrm{QAM}} \;=\;& \kappa_s \sigma_s^4 \sum_{i=0}^{P-1} |h_i|^4 + 2\sigma_s^4 \sum_{i=0}^{P-1} \sum_{m=0,m\neq i}^{P-1} |h_i|^2 |h_m|^2 \\
&+ \kappa_w \sigma_w^4 \sum_{i=0}^{2N-1} |f_i|^4 + 2\sigma_w^4 \sum_{i=0}^{2N-1} \sum_{m=0,m\neq i}^{2N-1} |f_i|^2 |f_m|^2 \\
&+ 4\sigma_s^2 \sigma_w^2 \|\mathbf{h}\|_2^2 \|\mathbf{f}\|_2^2 - 2\sigma_s^2 \kappa_s (\sigma_s^2 \|\mathbf{h}\|_2^2 + \sigma_w^2 \|\mathbf{f}\|_2^2) + \sigma_s^4 \kappa_s^2.
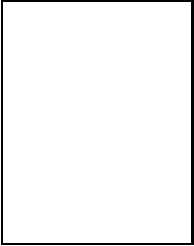\end{aligned}
\tag{66}
$$

C. Richard Johnson, Jr. was born in Macon, GA in 1950. He received the Ph.D. in electrical engineering with minors in engineering-economic systems and art history from Stanford University in 1977. He is currently a Professor of Electrical Engineering and a member of the Graduate Field of Applied Mathematics at Cornell University, Ithaca, NY. His research in adaptive parameter estimation theory with applications in digital control and signal processing has been supported by the National Science Foundation, the Engineering Foundation, the National Aeronautics and Space Administration, Tellabs Research Laboratory, MOOG Technology Center, United Technologies Research Center, and Applied Signal Technology. Dr. Johnson's current research interest is in adaptive parameter estimation theory useful in applications of digital signal processing to telecommunication systems. His principal focus in the 1990s has been blind linear equalization for intersymbol interference removal from received QAM sources.
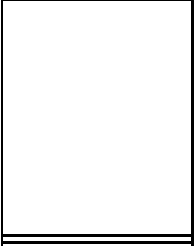
Philip Schniter was born in Evanston, IL in 1970. He received the B.S. and M.S. degrees in electrical and computer engineering from the University of Illinois at Urbana-Champaign in 1992 and 1993, respectively, and in 1993 he formed the band "Backhoe." From 1993 to 1996 he was employed by Tektronix, Inc., in Beaverton, OR, as a systems engineer. There he worked on signal processing aspects of video and communications instrumentation design, including algorithms, software, and hardware architectures.

Since 1996 he has been working toward the Ph.D. degree in electrical engineering at Cornell University, Ithaca, NY, and in 1998 he received the Schlumberger Fellowship. His research interests are in signal processing, communications, and control, and include blind adaptive equalization.
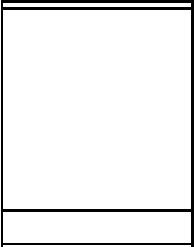
Thomas J. Endres received the B.S. from Cornell University (Ithaca, NY) in 1990, M.S. from the University of Southern California in (Los Angeles, CA) 1994, and the Ph.D. from Cornell University in 1997, all in Electrical Engineering. From 1990 to 1994 he was employed by Hughes Space and Communications, Los Angeles, CA.

Since 1997 he has been a founding member of Sarnoff Digital Communications, Newtown, PA. His current research interests include blind equalization, adaptive systems, and fast Harley-Davidsons with big flywheels.
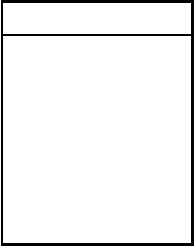
James D. Behm received a M.S. in Mathematics in 1975 from the University of Arizona, Tucson, AZ and a M.S. in Computer Science in 1986 from Johns Hopkins Whiting School of Engineering, Laurel, MD. Since 1975 he has been a mathematician with the Department of Defense, Ft. Meade, MD.

His current research interests include signal processing and communications, and he spent the spring 1997 semester as a visiting researcher at Cornell University, Ithaca, NY.

Donald R. Brown received B.S. and M.S. degrees in electrical engineering from the University of Connecticut, Storrs, CT, in 1992 and 1996. While pursuing his Masters degree, he worked for the General Electric Company from 1992 to 1997 as a Development Engineer. He is currently a graduate student at Cornell University and his research interests include adaptive signal processing and communications.

Raúl A. Casas was born in Denver, CO in 1972. He received a B.S. degree (1994) in electrical engineering from Cornell University. Continuing at the Cornell University electrical engineering department with the support from a National Science Foundation Minorities Fellowship, he received a M.S. degree (1996) and is now working towards a Ph.D. degree. His thesis work is a result of collaborative research with the United Technologies Research Center on identification and control of nonlinear combustion chamber dynamics.

His current research interests include the analysis of signal processing, communications and control related applications from a nonlinear dynamical systems perspective. Right on!