Sufficient Conditions for the Local Convergence of Constant Modulus Algorithms

Philip Schniter and C. Richard Johnson, Jr., *Fellow, IEEE*

***Abstract*—The constant modulus (CM) criterion has become popular in the design of blind linear estimators of sub-Gaussian i.i.d. processes transmitted through unknown linear channels in the presence of unknown additive interference. The existence of multiple CM minima, however, makes it difficult for CM-minimizing schemes to generate estimates of the desired source (as opposed to an interferer) in multiuser environments. In this paper, we present three separate sufficient conditions under which gradient descent (GD) minimization of CM cost will locally converge to an estimator of the desired source at a particular delay. The sufficient conditions are expressed in terms of statistical properties of the initial estimates, specifically, CM cost, kurtosis, and signal-to-interference-plus-noise ratio (SINR). Implications on CM-GD initialization methods are also discussed.**

***Index Terms*—Blind beamforming, blind deconvolution, blind equalization, blind multiuser detection, constant modulus algorithm, Godard algorithm.**

I. INTRODUCTION

CONSIDER the linear estimation problem of Fig. 1, where a desired source sequence $\{s_n^{(0)}\}$ combines linearly with K interfering sources through vector channels $\{\boldsymbol{h}^{(0)}(z), \cdots, \boldsymbol{h}^{(K)}(z)\}$. Our goal is to estimate the desired source using the (vector) linear estimator $\boldsymbol{f}(z)$. The linear estimates $\{y_n\}$ that minimize the mean-squared error (MSE)

$$J_{m,\nu}(y_n) := E\left\{\left|y_n - s_{n-\nu}^{(0)}\right|^2\right\} \tag{1}$$

are generated by the minimum MSE (MMSE) estimator, or Wiener estimator $\boldsymbol{f}_{m,\nu}(z)$. Specification of $\boldsymbol{f}_{m,\nu}(z)$, however, requires knowledge of the joint statistics of the observed sequence $\{\boldsymbol{r}_n\}$ and the desired source $\{s_{n-\nu}^{(0)}\}$, which are typically unavailable when the channel is unknown.

When only the statistics of the observed signal $\{\boldsymbol{r}_n\}$ are known, it may still be possible to estimate $\{s_n^{(0)}\}$ up to unknown magnitude and delay, i.e., $y_n = \sum_i \boldsymbol{f}_i^H \boldsymbol{r}_{n-i} \approx \alpha s_{n-\nu}^{(0)}$ for some $\alpha \in \mathbb{C}$, some $\nu \in \mathbb{Z}$, and all n. The literature refers to this problem as *blind* estimation (or blind deconvolution).

Minimization of the so-called constant modulus (CM) criterion [1], [2] has become perhaps the most studied and implemented means of blind equalization for data communication over dispersive channels (see, e.g., [3] and the references within) and has also been used successfully as a means of blind beamforming (see, e.g., [4]). The CM criterion is defined below in terms of the estimates $\{y_n\}$ and a design parameter γ.

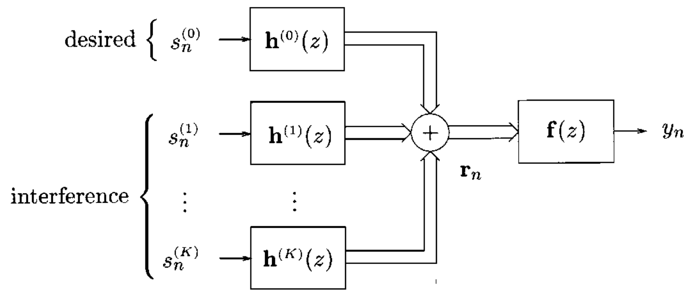


Fig. 1. Linear system model with K sources of interference.

$$J_c(y_n) := E\left\{(|y_n|^2 - \gamma)^2\right\}. \tag{2}$$

The popularity of the CM criterion is usually attributed to 1) the existence of a simple adaptive algorithm (known as the CM algorithm or CMA [1], [2]) for estimation and tracking of the CM-minimizing estimator $\boldsymbol{f}_c(z)$ and 2) the excellent MSE performance of CM-minimizing estimators. The second of these two points was first conjectured in the original works [1], [2] and recently established by the authors for arbitrary linear channels and additive interference [5].

Perhaps the greatest challenge facing successful application of the CM criterion in arbitrary interference environments results from the difficulty in determining CM-minimizing estimates of the desired source (as opposed to mistakenly estimating an interferer). The potential for "interference capture" is a direct consequence of the fact that the CM criterion exhibits multiple local minima in the estimator parameter space, each corresponding to a CM estimator of a particular source at a particular delay. Such multimodality might be suspected from (2); the CM criterion is based on a particular property of the estimates $\{y_n\}$, and one can imagine a case in which this property is satisfied to a similar extent by, e.g., $\{y_n\} \approx \{s_n^{(0)}\}$ and $\{y_n\} \approx \{s_n^{(1)}\}$ when $\{s_n^{(0)}\}$ and $\{s_n^{(1)}\}$ have the same statistics.

Various "multiuser" modifications of the CM criterion have been proposed to jointly estimate all sub-Gaussian sources present in a multisource environment. Some of these techniques add a non-negative term to the CM criterion, which penalizes correlation between any pair of L parallel estimator outputs, forcing the L estimators to generate estimates of L distinct sources [6]–[8]. Other techniques use the CM criterion in a suc-

Manuscript received July 28, 1999; revised April 19, 2000. This work was supported in part by NSF Grants MIP-9509011 and ECS-9811297, an Intel Foundation fellowship, and Applied Signal Technology, Inc. The associate editor coordinating the review of this paper and approving it for publication was Prof. Hideaki Sakai.

P. Schniter is with the Department of Electrical Engineering, The Ohio State University, Columbus, OH 43210 USA.

C. R. Johnson, Jr. is with the Department of Electrical Engineering, Cornell University, Ithaca, NY 14853 USA.

Publisher Item Identifier S 1053-587X(00)07687-X.

cessive interference cancellation scheme, whereby estimates of the n strongest sub-Gaussian sources are used to remove their respective contributions to the received signal before estimation of the $n+1$th source [9]. Both of these approaches, however, require knowledge of the number of (non-Gaussian) sources, result in significant increase in computational complexity when the number of sources is large, and generate estimators with questionable MSE performance. Instead, we focus on the well-known standard CM (or "Godard" [1]) criterion and consider desired-source convergence as an outcome of proper initialization.

Closed-form expressions for CM estimators do not generally exist, and thus, gradient descent (GD) methods provide the typical means of solving for these estimators. Because exact gradient descent requires statistical knowledge of the received process that is not usually available in practical situations, *stochastic* GD algorithms such as CMA are used to estimate and track the (possibly time-varying) CM estimator. It is widely accepted, however, that small step-size stochastic GD algorithms exhibit mean transient and steady-state behaviors very close to those of exact GD under typical operating conditions [10], [11]. Hence, we circumvent the details of stochastic adaptation by restricting our attention to (exact) GD minimization of the CM cost. An important property of GD minimization is that the location of algorithm initialization completely determines the stationary point to which the GD trajectory will eventually converge. The description of the CM-GD regions-of-convergence (ROC) in terms of estimator parameters appears to be a very difficult problem, however, and attempts at finding closed-form expressions for the ROC boundaries have thus far been unsuccessful [12], [13].

In this paper, we derive three sufficient conditions under which CM-GD minimization will generate an estimator for the desired source. The conditions are expressed in terms of statistical properties of the initial estimates, specifically, CM cost, kurtosis, and signal to interference-plus-noise ratio (SINR). Earlier attempts at describing the interference capture or "local convergence" properties of CMA have been made by Treichler and Larimore in [14] and Li and Ding in [15]. Treichler and Larimore constructed a simplifying approximation to the mean behavior of CMA for the case of a constant envelope signal in tonal interference and inferred the roles of initial SINR and initial estimator parameterization on desired convergence. Li and Ding derived a sufficient kurtosis condition for the local convergence of the Shalvi–Weinstein algorithm (SWA) [16] and suggested that the condition applies to small-stepsize CMA as well. Our analysis and simulations suggest that the local convergence behavior of CMA differs from that of SWA, contrasting certain claims of [15].

The organization of the paper is as follows. Section II discusses relevant properties of the system model and of the CM criterion, Section III derives initialization conditions sufficient for CM-GD convergence to desired source estimates, and Section IV discusses the implications of these conditions on choice of CM-GD initialization scheme. Section V presents numerical simulations verifying our analyzes, and Section VI concludes the paper.

II. Background

In this section, we give more detailed information on the linear system model and the CM criterion. The following notation is used throughout:

- $(\cdot)^t$ transpose;
- $(\cdot)^*$ conjugation;
- $(\cdot)^H$ hermitian;
- $E\{\cdot\}$ expectation.

In addition, $\boldsymbol{I}$ denotes the identity matrix, $\mathbb{R}^+$ the field of non-negative real numbers, and $\|\boldsymbol{x}\|_p$ the p-norm defined by $\sqrt[p]{\sum_i |x_i|^p}$. In general, we use boldface lowercase type to denote vector quantities and boldface uppercase type to denote matrix quantities.

A. Linear System Model

First, we formalize the linear time-invariant multichannel model illustrated in Fig. 1. Say that the desired symbol sequence $\{s_n^{(0)}\}$ and K sources of interference $\{s_n^{(1)}\}, \cdots, \{s_n^{(K)}\}$ each pass through separate linear "channels" before being observed at the receiver. The interference processes may correspond, e.g., to interference signals or additive noise processes. In addition, say that the receiver uses a sequence of P-dimensional vector observations $\{\boldsymbol{r}_n\}$ to estimate (a possibly delayed version of) the desired source sequence, where the case $P > 1$ corresponds to a receiver that employs multiple sensors and/or samples at an integer multiple of the symbol rate. The observations can be written $\boldsymbol{r}_n = \sum_{k=0}^{K} \sum_{i=0}^{\infty} \boldsymbol{h}_i^{(k)} s_{n-i}^{(k)}$, where $\{\boldsymbol{h}_n^{(k)}\}$ denote the impulse response coefficients of the linear time-invariant (LTI) channel $\boldsymbol{h}^{(k)}(z)$. We assume that $\boldsymbol{h}^{(k)}(z)$ is causal and bounded-input bounded-output (BIBO) stable. Note that such $\boldsymbol{h}^{(k)}(z)$ admit infinite-duration impulse response (IIR) channel models.

From the vector-valued observation sequence $\{\boldsymbol{r}_n\}$, the receiver generates a sequence of linear estimates $\{y_n\}$ of $\{s_{n-\nu}^{(0)}\}$, where ν is a fixed integer. Using $\{\boldsymbol{f}_n\}$ to denote the impulse response of the linear estimator $\boldsymbol{f}(z)$, the estimates are formed as $y_n = \sum_{i=-\infty}^{\infty} \boldsymbol{f}_i^H \boldsymbol{r}_{n-i}$. We will assume that the linear system $\boldsymbol{f}(z)$ is BIBO stable with *constrained* ARMA structure, i.e., the pth element of $\boldsymbol{f}(z)$ takes the form

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

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

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

$$q_n^{(k)} = \sum_{i=-\infty}^{\infty} \boldsymbol{f}_i^H \boldsymbol{h}_{n-i}^{(k)} \quad (3)$$

allowing the estimates to be written as $y_n = \sum_{k=0}^{K} \sum_{i=-\infty}^{\infty} q_i^{(k)} s_{n-i}^{(k)}$. Adopting the following vector notation helps to streamline the remainder of the paper.

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

For instance, the estimates can be rewritten concisely as

$$y_n = \sum_{k=0}^{K} \boldsymbol{q}^{(k)t} \boldsymbol{s}^{(k)}(n) = \boldsymbol{q}^t \boldsymbol{s}(n). \tag{4}$$

We now point out two important properties of $\boldsymbol{q}$. First, it is important to recognize that placing a particular structure on the channel and/or estimator will restrict the set of *attainable* global responses, which we will denote by $\mathcal{Q}_a$. For example, when the estimator is FIR, (3) implies that $\boldsymbol{q} \in \mathcal{Q}_a = \text{row}(\boldsymbol{\mathcal{H}})$, where

$$\boldsymbol{\mathcal{H}} := \begin{pmatrix} \boldsymbol{h}_0^{(0)} \cdots \boldsymbol{h}_0^{(K)} & \boldsymbol{h}_1^{(0)} \cdots \boldsymbol{h}_1^{(K)} & \boldsymbol{h}_2^{(0)} \cdots \boldsymbol{h}_2^{(K)} & \cdots \\ \boldsymbol{0} \cdots \boldsymbol{0} & \boldsymbol{h}_0^{(0)} \cdots \boldsymbol{h}_0^{(K)} & \boldsymbol{h}_1^{(0)} \cdots \boldsymbol{h}_1^{(K)} & \cdots \\ \vdots \quad \vdots & \vdots \quad \vdots & \vdots \quad \vdots & \vdots \\ \boldsymbol{0} \cdots \boldsymbol{0} & \boldsymbol{0} \cdots \boldsymbol{0} & \boldsymbol{h}_0^{(0)} \cdots \boldsymbol{h}_0^{(K)} & \cdots \end{pmatrix}. \tag{5}$$

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

Throughout the paper, we make the following assumptions on the $K+1$ source processes.

- S1) For all k, $\{s_n^{(k)}\}$ is zero-mean i.i.d.
- S2) $\{s_n^{(0)}\}, \cdots, \{s_n^{(K)}\}$ are jointly statistically independent.
- S3) For all k, $E\{|s_n^{(k)}|^2\} = \sigma_s^2$.
- S4) $\mathcal{K}(s_n^{(0)}) < 0$, where $\mathcal{K}(\cdot)$ denotes kurtosis:

$$\mathcal{K}(s_n) := E\{|s_n|^4\} - 2\,E^2\{|s_n|^2\} - \left|E\{s_n^2\}\right|^2. \tag{6}$$

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

Note that S4) assumes the desired source is "sub-Gaussian," whereas S5) assumes that all sources are "circularly-symmetric" if any of the global responses or sources are complex valued.

B. Signal to Interference-Plus-Noise Ratio

Given global response $\boldsymbol{q}$, we can decompose the estimate into signal and interference terms:

$$y_n = q_\nu^{(0)} s_{n-\nu}^{(0)} + \bar{\boldsymbol{q}}^t \bar{\boldsymbol{s}}(n) \tag{7}$$

where $\bar{\boldsymbol{q}}$ denotes $\boldsymbol{q}$ with the $q_\nu^{(0)}$ term removed, and $\bar{\boldsymbol{s}}(n)$ denotes $\boldsymbol{s}(n)$ with the $s_{n-\nu}^{(0)}$ term removed.

The SINR associated with y_n, which is an estimate of $s_{n-\nu}^{(0)}$, is then defined as

$$\text{SINR}_\nu := \frac{E\left\{\left|q_\nu^{(0)} s_{n-\nu}^{(0)}\right|^2\right\}}{E\left\{\left|\bar{\boldsymbol{q}}^t \bar{\boldsymbol{s}}(n)\right|^2\right\}} = \frac{\left|q_\nu^{(0)}\right|^2}{\|\bar{\boldsymbol{q}}\|_2^2} \tag{8}$$

where the equality invokes assumptions S1)–S3).

C. Constant Modulus Criterion

The constant modulus (CM) criterion, which was introduced independently in [1] and [2], was defined in (2) in terms of the estimates $\{y_n\}$. In (2), γ is a positive parameter known as the "dispersion constant." Although γ is often chosen according to the (marginal) statistics of the desired source process (when they are known), we will see below that the choice of γ does not affect the SINR performance of the CM-minimizing estimator.

For general sources, channels, and estimators fitting the framework of Fig. 1 and S1)–S5), the authors have bounded the MSE performance of the CM-minimizing estimator [5]. To avoid the inherent gain ambiguity[1] of blind estimators, the analysis examined conditionally unbiased MSE (UMSE) performance, which can be directly related to SINR as follows: $\text{UMSE}_\nu = \text{SINR}_\nu^{-1} \sigma_s^2$ [5]. An approximation of one of these bounds is given below. Let $\text{SINR}_{m,\nu}$ denote the maximum (i.e., Wiener) SINR associated with estimation of the desired source at delay ν. Then, the SINR characterizing CM-minimizing estimators of the desired source at the same delay can be written as

$$\text{SINR}_c^{-1} = \text{SINR}_{m,\nu}^{-1} + A \cdot \text{SINR}_{m,\nu}^{-2} + \mathcal{O}(\text{SINR}_{m,\nu}^{-3}) \tag{9}$$

as long as $\text{SINR}_{m,\nu} \geq B$. Here, A and B are constants that depend on the kurtoses of the desired and interfering sources. As an example, $A = 1/2$ and $B = \sqrt{2}+1 = 3.8$ dB when no sources are super-Gaussian, and no sources have kurtosis less than the desired source [i.e., $\mathcal{K}(\boldsymbol{s}_n^{(0)}) \leq \mathcal{K}(\boldsymbol{s}_n^{(k)}) \leq 0\ \forall\, k$].

From (9), it can be seen that the SINR of CM-minimizing estimators approaches infinity as the Wiener SINR approaches infinity. In other words, the conditions leading to perfect Wiener estimators also lead to CM-minimizing estimators, which are "perfect" up to a gain ambiguity [3], [17]. Equation (9) also shows that the SINR performance of CM-minimizing estimators is insensitive to choice of dispersion constant γ.

III. Sufficient Conditions for Local Convergence of CM-GD

A. Main Idea

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

$$\mathcal{Q}_\nu^{(0)} := \left\{ \boldsymbol{q} \quad \text{s.t.}\ \left|q_\nu^{(0)}\right| > \max_{(k,\delta)\neq(0,\nu)} \left|q_\delta^{(k)}\right| \right\}. \tag{10}$$

[1]Gain ambiguity occurs when both the symbol power and channel gain are unknown.

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

$$\{\boldsymbol{q}_r\} := \arg\min_{\boldsymbol{q} \in \text{bndr}(\mathcal{Q}_\nu^{(0)})} J_c(\boldsymbol{q}) \tag{11}$$

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

$$\mathcal{Q}_c(\boldsymbol{q}_r) := \{\boldsymbol{q} \quad \text{s.t. } J_c(\boldsymbol{q}) \le J_c(\boldsymbol{q}_r)\} \cap \mathcal{Q}_\nu^{(0)}.$$

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

Since the size and shape of $\mathcal{Q}_c(\boldsymbol{q}_r)$ are not easily characterizable, we find it more useful to derive sufficient CM-GD initialization conditions in terms of well-known statistical quantities such as kurtosis and SINR. It has been shown that CM cost and kurtosis are closely related [15], and we will see that translation between these two quantities is relatively straightforward. Translation of the initial CM-cost condition into an initial SINR condition is more difficult but can be accomplished through definition of $\text{SINR}_{\min,\nu}$, the SINR above which all $\boldsymbol{q}$ have scaled versions in $\mathcal{Q}_c(\boldsymbol{q}_r)$:

$$\text{SINR}_{\min,\nu} := \min x \text{ s.t. } \left\{ \forall \boldsymbol{q}\text{: } \text{SINR}_\nu(\boldsymbol{q}) \ge x \quad \exists\, a_* \text{ s.t. } \frac{a_* \boldsymbol{q}}{\|\boldsymbol{q}\|_2} \in \mathcal{Q}_c(\boldsymbol{q}_r) \right\}. \tag{12}$$

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

Since $\mathcal{Q}_c(\boldsymbol{q}_r)$ and $\text{SINR}_\nu(\boldsymbol{q})$ are all invariant to phase rotation (i.e., scalar multiplication by $e^{j\phi}$ for $\phi \in \mathbb{R}$) of $\boldsymbol{q}_r$ and $\boldsymbol{q}$, respectively, we can (w.l.o.g.) restrict our attention to the "derotated" set of global responses $\{\boldsymbol{q} \text{ s.t. } q_\nu^{(0)} \in \mathbb{R}^+\}$. Such $\boldsymbol{q}$ allow parameterization in terms of gain $a = \|\boldsymbol{q}\|_2$ and interference response $\bar{\boldsymbol{q}}$ (defined in Section II-B), where $\|\bar{\boldsymbol{q}}\|_2 \le a$. In terms of the pair $(a, \bar{\boldsymbol{q}})$, the SINR (8) can be written

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

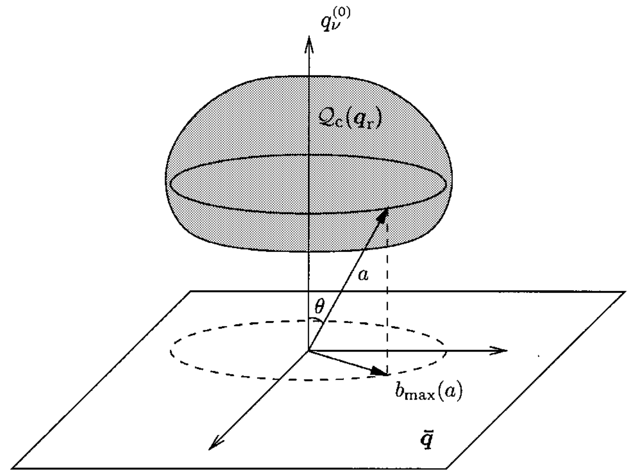


Fig. 2. Illustration of maximum interference gain $b_{\max}(a)$ below which all global responses with gain a are contained in the CM cost region $\mathcal{Q}_c(\boldsymbol{q}_r)$. Note that $\text{SINR}(a, b_{\max}) = \cot^2(\theta)$.

so that (12) becomes

$$\text{SINR}_{\min,\nu} := \min x \text{ s.t. } \left\{ \forall\, (a, \bar{\boldsymbol{q}})\text{: } \frac{a^2 - \|\bar{\boldsymbol{q}}\|_2^2}{\|\bar{\boldsymbol{q}}\|_2^2} \ge x \quad \exists\, a_* \text{ s.t. } \left(a_*, \frac{a_*}{a}\bar{\boldsymbol{q}}\right) \in \mathcal{Q}_c(\boldsymbol{q}_r) \right\}. \tag{13}$$

Under particular conditions on a and $\boldsymbol{q}_r$ (which will be made explicit in Section III-B), there exists a maximum interference gain b, specified as a function of system gain a, below which all $\bar{\boldsymbol{q}}$ are contained in $\mathcal{Q}_c(\boldsymbol{q}_r)$:

$$b_{\max}(a) := \max b(a) \text{ s.t. } \{\forall \bar{\boldsymbol{q}}\text{: } \|\bar{\boldsymbol{q}}\|_2 \le b(a), \quad (a, \bar{\boldsymbol{q}}) \in \mathcal{Q}_c(\boldsymbol{q}_r)\}. \tag{14}$$

For an illustration of a, $b_{\max}(a)$, and $\mathcal{Q}_c(\boldsymbol{q}_r)$, see Fig. 2. Now, consider the quantity

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

Since $\text{SINR}_\nu(a, b_{\max})$ is a decreasing function of $b_{\max}(a)$ (over its valid domain), (14) implies that

$$\frac{a^2 - \|\bar{\boldsymbol{q}}\|_2^2}{\|\bar{\boldsymbol{q}}\|_2^2} \ge \text{SINR}_\nu(a, b_{\max}) \quad \Rightarrow \quad (a, \bar{\boldsymbol{q}}) \in \mathcal{Q}_c(\boldsymbol{q}_r).$$

Using the previous expression to minimize SINR in accordance with (13) yields the key quantities defined in (12):

$$\text{SINR}_{\min,\nu} = \min_a \text{SINR}_\nu(a, b_{\max}) \tag{15}$$

$$a_* = \arg\min_a \text{SINR}_\nu(a, b_{\max}). \tag{16}$$

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

B. Derivation of Sufficient Conditions

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

It is convenient to now define the *normalized* kurtosis [not to be confused with $\mathcal{K}(\cdot)$ in (6)]:

$$\kappa_s^{(k)} := \frac{E\left\{\left|s_n^{(k)}\right|^4\right\}}{E^2\left\{\left|s_n^{(k)}\right|^2\right\}}. \tag{17}$$

Under the following definition of κ_g

$$\kappa_g := \begin{cases} 3, & s_n^{(k)} \in \mathbb{R} \quad \forall k, n \\ 2, & \text{otherwise} \end{cases} \tag{18}$$

our results will hold for both real-valued and complex-valued models. Note that under S1) and S5), κ_g represents the normalized kurtosis of a Gaussian source. It can be shown that the normalized and un-normalized kurtoses are related by $\mathcal{K}(s_n^{(k)}) = (\kappa_s^{(k)} - \kappa_g)\sigma_s^4$ under S3) and S5). Next, we define the minimum and maximum (normalized) interference kurtoses.

$$\kappa_s^{\min} := \begin{cases} \min\limits_{0 \le k \le K} \kappa_s^{(k)}, & \dim\left(\boldsymbol{q}^{(0)}\right) > 1 \\ \min\limits_{1 \le k \le K} \kappa_s^{(k)}, & \dim\left(\boldsymbol{q}^{(0)}\right) = 1 \end{cases} \tag{19}$$

$$\kappa_s^{\max} := \max_{0 \le k \le K} \kappa_s^{(k)} \tag{20}$$

where "dim" denotes the dimension of a vector. Note that the second case in (19) applies only when the desired source contributes zero intersymbol interference (ISI). The following kurtosis-based quantities will also be convenient in the sequel.

$$\rho_{\min} := \frac{\kappa_g - \kappa_s^{\min}}{\kappa_g - \kappa_s^{(0)}} \tag{21}$$

$$\rho_{\max} := \frac{\kappa_g - \kappa_s^{\max}}{\kappa_g - \kappa_s^{(0)}} \tag{22}$$

$$\sigma_y^2\big|_{\text{crit}} := \gamma\left(\frac{4}{\kappa_s^{(0)} + \kappa_s^{\min} + 2\kappa_g}\right). \tag{23}$$

Lemma 1: The CM cost (2) may be written in terms of global response $\boldsymbol{q}$ as

$$\frac{J_c(\boldsymbol{q})}{\sigma_s^4} = \sum_k \left(\kappa_s^{(k)} - \kappa_g\right)\left\|\boldsymbol{q}^{(k)}\right\|_4^4 + \kappa_g\|\boldsymbol{q}\|_2^4 - 2(\gamma/\sigma_s^2)\|\boldsymbol{q}\|_2^2 + (\gamma/\sigma_s^2)^2. \tag{24}$$

Lemma 2: The minimum CM cost on the boundary of $\mathcal{Q}_\nu^{(0)}$ is

$$J_c(\boldsymbol{q}_r) = \min_{\boldsymbol{q} \in \text{bndr}(\mathcal{Q}_\nu^{(0)})} J_c(\boldsymbol{q}) = \gamma^2\left(1 - \frac{4}{\kappa_s^{(0)} + \kappa_s^{\min} + 2\kappa_g}\right). \tag{25}$$

Theorem 1: If $\{y_n\}$ are initial estimates of the desired source at delay ν [i.e., $y_n = \boldsymbol{q}_{\text{init}}^t \boldsymbol{s}(n)$ for $\boldsymbol{q}_{\text{init}} \in \mathcal{Q}_\nu^{(0)} \cap \mathcal{Q}_a$] with CM cost

$$J_c(y_n) < \gamma^2\left(1 - \frac{4}{\kappa_s^{(0)} + \kappa_s^{\min} + 2\kappa_g}\right) \tag{26}$$

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

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

$$\kappa_y < \kappa_y^{\text{crit}} := \tfrac{1}{4}\left(\kappa_s^{(0)} + \kappa_s^{\min} + 2\kappa_g\right) \tag{27}$$

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

Theorem 3: If $\kappa_s^{(0)} \le (\kappa_s^{\min} + 2\kappa_g)/3$, and if $\{y_n\}$ are initial estimates with variance $\sigma_y^2 = \sigma_y^2|_{\text{crit}}$ and $\text{SINR}_\nu(y_n) > \text{SINR}_{\min,\nu}$, where we have (28), shown at the bottom of the page, then estimators resulting from subsequent CM-minimizing gradient descent will also yield estimates of the desired source at delay ν.

We now make a few comments on the theorems. First, notice the stringent gain condition $\sigma_y^2 = \sigma_y^2|_{\text{crit}}$ in Theorems 2 and 3. Is this a necessary component of our sufficient conditions? The answer is a qualified yes. It is possible to construct situations

$$\text{SINR}_{\min,\nu} = \begin{cases} \dfrac{\sqrt{1+\rho_{\min}}}{2 - \sqrt{1+\rho_{\min}}}, & \kappa_s^{\max} \le \kappa_g, \\ \left\{\begin{array}{ll} \dfrac{\rho_{\max} + \sqrt{1 - (1+\rho_{\max})(3-\rho_{\min})/4}}{1 - \sqrt{1 - (1+\rho_{\max})(3-\rho_{\min})/4}}, & \rho_{\max} \ne -1 \\ \dfrac{5+\rho_{\min}}{3-\rho_{\min}}, & \rho_{\max} = -1 \end{array}\right\}, & \kappa_s^{\max} > \kappa_g \end{cases} \tag{28}$$

in which the critical values of kurtosis from (27) or SINR from (28) are satisfied, yet misconvergence occurs because the gain condition $\sigma_y^2 = \sigma_y^2|_{\rm crit}$ is not satisfied.[2] Fortunately, it appears that such scenarios are quite rare unless σ_y^2 is far from $\sigma_y^2|_{\rm crit}$ or unless the SINR and/or kurtosis conditions are themselves near violation. Thus, in practice, successful CM-GD convergence should be quite robust to small violations in the gain condition. We mention that it *is* possible to rederive Theorems 2 and 3 so that they guarantee convergence for initial σ_y^2 in a bounded interval around $\sigma_y^2|_{\rm crit}$ and appropriately adjusted kurtosis/SINR requirements. This should be evident from the proofs in the Appendix.

Finally, it should be pointed out that the relatively complicated expressions in Theorem 3 simplify under the operating conditions commonly encountered in, e.g., data communication. When the sources of interference are nonsuper-Gaussian (i.e., $\kappa_s^{\max} \le \kappa_g$) and none have kurtosis less than the desired source [i.e., $\kappa_s^{(0)} \le \kappa_s^{(k)}$], we find that $\rho_{\min} = 1$, and thus, $\text{SINR}_{\min,\nu} = 1 + \sqrt{2}$ or 3.8 dB.

IV. Implications for CM-GD Initialization Schemes

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

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

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

A. "Single-Spike" Initialization

The so-called single-spike initialization, which was first proposed in [1], is quite popular in single-user environments. Single-spike initializations for single-sensor baud-spaced equalizers (i.e., $P = 1$) are characterized by impulse responses with a single nonzero coefficient, i.e., $\boldsymbol{f}(z) = z^{-\delta}$. There exists a straightforward extension to multirate/multichannel (i.e., $P > 1$) estimators: $\boldsymbol{f}(z) = \mathbf{1} z^{-\delta}$ for $\mathbf{1} := (\sqrt{1/P}, \cdots, \sqrt{1/P})^t \in \mathbb{R}^P$. For $P = 2$, this has been called the "double-spike" initialization [17]. The spike position is often an important design parameter, as we explain below.

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

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

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

$$\kappa_y = \left(\kappa_s^{(0)} - \kappa_g\right) \frac{\sum_i \left|\mathbf{1}^t \boldsymbol{h}_i^{(0)}\right|^4}{\left(\sum_i \left|\mathbf{1}^t \boldsymbol{h}_i^{(0)}\right|^2 + \sigma_w^2/\sigma_s^2\right)^2} + \kappa_g. \quad (29)$$

TABLE I
Single-Spike Kurtoses for SPIB Microwave Channel Models and 20 dB SNR

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

Table I shows initial kurtoses κ_y from (29) for Signal Processing Information Base[3] (SPIB) microwave channel models in AWGN (resulting in 20 dB SNR at channel output), along with the critical kurtosis $\kappa_y^{\rm crit}$ from (27). From Table I, we see that the single-spike initialization procedure generates estimates with kurtosis less than the critical value for all SPIB channels. The implication is that *the CM gradient descent from a single-spike initialization with magnitude chosen in accordance with Theorem 2 typically preserves the estimation delay of the initial estimates*. Similar conjectures have been made in [15] and [17].

Since MMSE performance is known to vary (significantly) with estimation delay, the recently established connection between Wiener and CM performance [5] implies that the MSE performance of CM-minimizing estimators should also vary with estimation delay. Thus, from our observations on the local convergence of single-spike initializations, we conclude that the asymptotic MSE performance of CM-GD estimators can be directly linked to the choice of initial spike delay.

B. Initialization Using Partial Information

Although the single-spike scheme has desirable properties in (noisy) single-user applications, one would not expect it to yield reliable estimates of the desired source when in the presence

[2]Thus, initial kurtosis cannot be the sole indicator of CM-GD convergence, as claimed in [15].

[3]The SPIB microwave channel database resides at http://spib.rice.edu/spib/microwave.html.

of significant sub-Gaussian interference since single-spike initialized CM-GD might lock onto a sub-Gaussian interferer instead of the desired source. With partial knowledge of the desired user's channel, however, it may be possible to construct rough guesses of the desired estimator that are good enough for use as CM-GD initializations. Then, if the initialization satisfies the sufficient conditions in the previous section, we know that CM-GD can be used to design an estimator with nearly optimal MSE performance (as discussed in Section II-C). The "partial knowledge" may come in various forms, for example, short training records in semi-blind equalization applications, rough direction-of-arrival knowledge in array applications, spreading sequences in code-division multiple access (CDMA) applications, or desired polarization angle in cross-pole interference cancellation.

We have seen that various criteria could be used to design and evaluate initialization hypotheses. Since reliable evaluation of higher order statistics typically require a larger sample size than second-order statistics, the design and/or evaluation of SINR-based initializations might be advantageous when sample size is an issue. For this reason, SINR-based methods will be considered for the remainder of this section. Still, good results have been reported for kurtosis-based CM-GD initialization schemes for CDMA applications when sample size is not an issue [18].

The SINR-maximizing linear estimator is given by Wiener estimator $\boldsymbol{f}_{m,\nu}(z)$. It can be shown that the Wiener estimator has the form [19]

$$\boldsymbol{f}_{m,\nu}(z) = z^{-\nu}\sigma_s^2 \left(E\left\{ \boldsymbol{r}(z)\boldsymbol{r}^H\left(\frac{1}{z^*}\right)\right\}\right)^{\dagger} \boldsymbol{h}^{(0)}\left(\frac{1}{z^*}\right) \quad (30)$$

where $(\cdot)^\dagger$ denotes pseudo-inverse. As evident from (30), design of $\boldsymbol{f}_{m,\nu}(z)$ requires knowledge of the desired channel $\boldsymbol{h}^{(0)}(z)$ in addition to the autocorrelation of the received signal. Although various methods exist for the design of blind SINR-maximizing (i.e., MSE-minimizing) estimators based on partial knowledge of $\boldsymbol{h}^{(0)}(z)$, the Wiener expression (30) suggests the following CM initialization when given only a channel estimate $\hat{\boldsymbol{h}}^{(0)}(z)$ and knowledge of $E\{\boldsymbol{r}(z)\boldsymbol{r}^H(1/z^*)\}$.

$$\boldsymbol{f}_{\text{init}}(z) = z^{-\nu} \left(E\left\{ \boldsymbol{r}(z)\boldsymbol{r}^H\left(\frac{1}{z^*}\right)\right\}\right)^{\dagger} \hat{\boldsymbol{h}}^{(0)}\left(\frac{1}{z^*}\right). \quad (31)$$

Note that (31) may require additional scaling to satisfy the σ_y^2-requirements of Theorems 2 and 3.

V. Numerical Examples

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

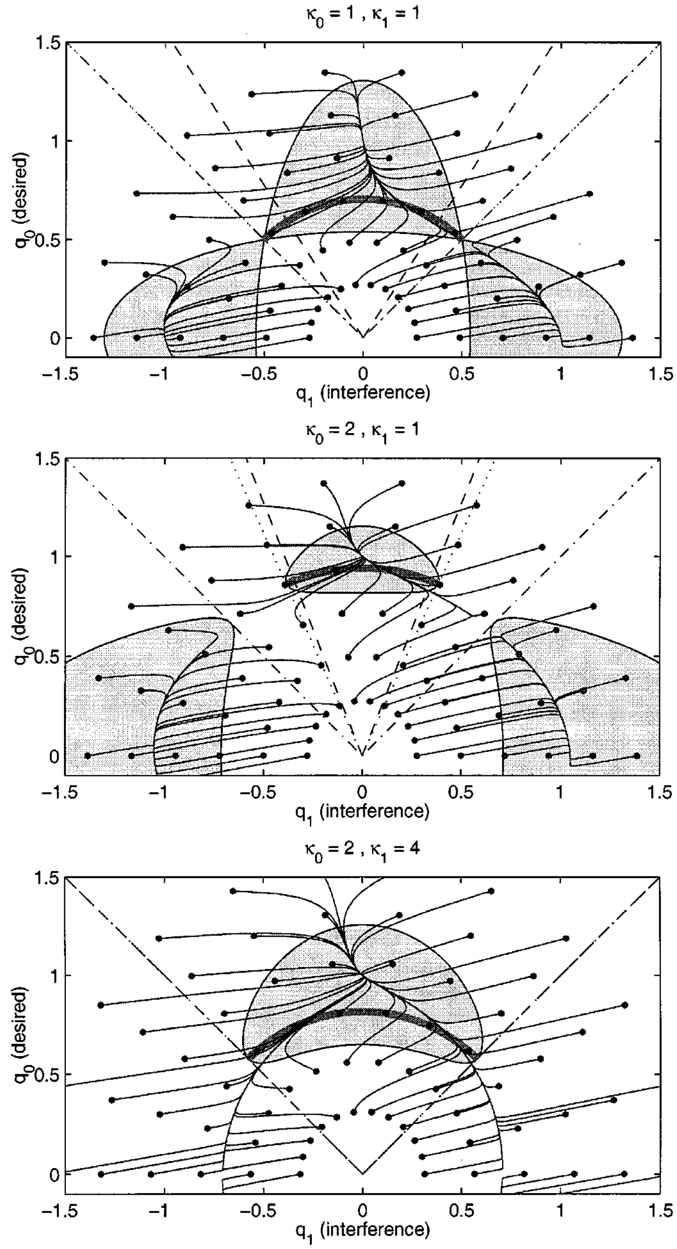


Fig. 3. CM-GD trajectories in channel-plus-estimator space ($\boldsymbol{q} \in \mathbb{R}^2$) for (a) $\kappa_s^{(0)} = 1$ and $\kappa_s^{(1)} = 1$, (b) $\kappa_s^{(0)} = 2$ and $\kappa_s^{(1)} = 1$, and (c) $\kappa_s^{(0)} = 2$ and $\kappa_s^{(1)} = 4$. $\mathcal{Q}_\nu^{(0)}$ boundaries (dash-dotted), $\text{SINR}_{\min,\nu}$ boundaries (dashed), κ_y^{crit} boundaries (dotted), and $J_c(\boldsymbol{q}) < J_c(\boldsymbol{q}_r)$ regions (shaded) are also shown. Channel estimators resulting in σ_y^2 of (23) are shown by the fat shaded arcs. Note that dotted and dash-dotted lines are coincident in (a), whereas dotted, dash-dotted, and dashed lines are coincident in (c).

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

Fig. 3(a)–(c) depicts the $\mathcal{Q}_\nu^{(0)}$ boundaries as dash-dotted lines, the $\text{SINR}_{\min,\nu}$ boundaries as dashed lines, and the κ_y^{crit} boundaries as dotted lines. Note that in Fig. 3(a), the dash-dotted and dotted lines are coincident, whereas in Fig. 3(c), the dash-dotted, dashed, and dotted lines are coincident. The three subplots in Fig. 3 differ only in the kurtosis of the desired source: In

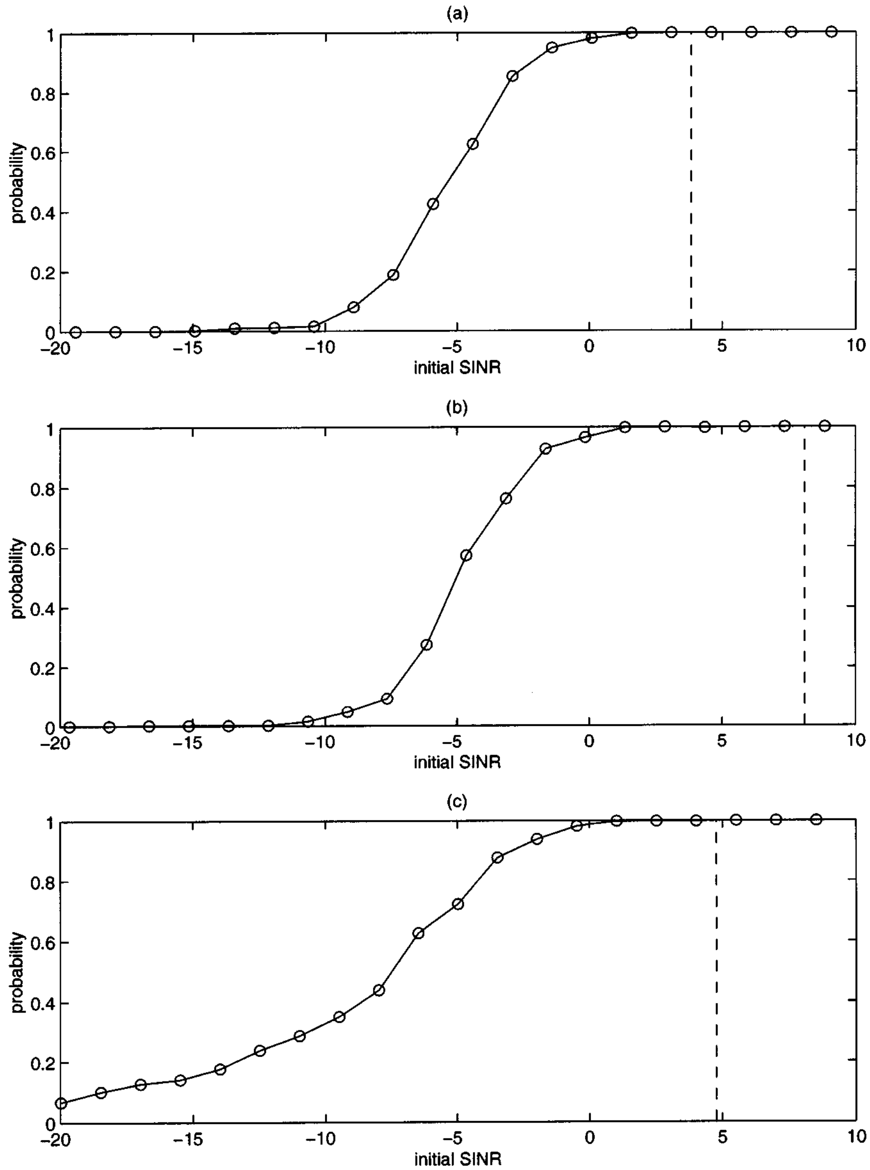


Fig. 4. Estimated probability of convergence to desired source/delay for random channels and random initializations scaled according to Theorem 3 as a function of initialization SINR. (a) $\kappa_s^{(0)} = 1$ with interfering $\kappa_s^{(k)} \in \{1, 3\}$. (b) $\kappa_s^{(0)} = 2$ with interfering $\kappa_s^{(k)} \in \{1, 2\}$. (c) $\kappa_s^{(0)} = 2$ with interfering $\kappa_s^{(k)} \in \{2, 4\}$. $\mathrm{SINR}_{\min,\,\nu}$ from (28) shown by dashed lines.

Fig. 3(a), the sources have $\kappa_s^{(0)} = 1$ and $\kappa_s^{(1)} = 1$, in Fig. 3(b), they have $\kappa_s^{(0)} = 2$ and $\kappa_s^{(1)} = 1$, whereas in Fig. 3(c), they have $\kappa_s^{(0)} = 2$ and $\kappa_s^{(1)} = 4$.

The following three behaviors can be observed in every subplot of Fig. 3. First, all trajectories entering into $\mathcal{Q}_c(\boldsymbol{q}_r)$ (which is denoted by the shaded region between the dash-dotted lines) converge to an estimator for the desired source, confirming Theorem 1. Next, all trajectories initialized with small enough kurtosis (indicated by the region between the dotted lines) and proper gain (indicated by the fat shaded arc) converge to an estimator for the desired source, thus confirming Theorem 2. Finally, all trajectories initialized with high enough SINR (indicated by the region between the dashed lines) and proper gain (again indicated by the fat shaded arc) converge to estimators for the desired source, confirming Theorem 3.

Fig. 3 suggests that the sufficient-SINR condition of Theorem 3 is more restrictive than the sufficient-kurtosis condition of Theorem 2, which in turn is more restrictive than the J_c-based condition of Theorem 1: The sufficient-SINR region (between the dashed lines) is contained by the sufficient-kurtosis region (between the dotted lines), which is contained by the sufficient-J_c region (between the dash-dotted lines). The relative ordering of these three conditions is, in fact, formally implied by the proofs in the Appendix.

We stress again that initial kurtosis or SINR is not sufficient for desired local convergence; initial estimator gain plays an important role. This is demonstrated by Fig. 3(a) and (b), wherein some trajectories initialized within the $\mathrm{SINR}_\nu(\boldsymbol{q}) > \mathrm{SINR}_{\min,\,\nu}$ region (between the dashed lines), but with insufficient initial gain, converge to the undesired equilibria $\boldsymbol{q} = (0, \pm 1)^t$. Although recognized in [14], this fact was overlooked in [15], resulting in some overly strong claims about the convergence behavior of CMA.

In Fig. 4, we examine probability of CM-GD convergence to desired {source, delay} versus SINR for higher dimensional estimators. CM gradient descents randomly initialized in a ball around $\boldsymbol{f}_{m,\nu}$ (and subsequently normalized according to Theorem 3) were conducted using random channel matrices $\{\boldsymbol{\mathcal{H}}\} \in \mathbb{R}^{10\times 11}$ with zero-mean Gaussian elements. Every data point in Fig. 4 represents an average of 500 CM-GD simulations. Fig. 4(a) demonstrates $\kappa_s^{(0)} = 1$ and ten interfering sources with $\kappa_s^{(k)} = 1$; Fig. 4(b) demonstrates $\kappa_s^{(0)} = 2$, five interfering sources with $\kappa_s^{(k)} = 1$, and five interfering sources with $\kappa_s^{(k)} = 2$; Fig. 4(c) demonstrates $\kappa_s^{(0)} = 2$, five interfering sources with $\kappa_s^{(k)} = 2$, and five interfering sources with $\kappa_s^{(k)} = 4$.

Fig. 4 also confirms the claim of Theorem 3: All properly-scaled CM-GD initializations with SINR_ν greater than $\text{SINR}_{\min,\nu}$ converge to the desired source. Recalling that the SINR condition is sufficient, but not necessary, for desired convergence, it is interesting to note that both Figs. 3 and 4 suggest that the sufficiency of our SINR condition becomes "looser" as the kurtosis of the desired source rises above the minimum interference kurtosis (i.e., as $\rho_{\min}$ increases).

VI. Conclusions

In this paper, we have derived, under the general linear model of Fig. 1, three sufficient conditions for the convergence of CM-minimizing gradient descent to an estimator for a particular source at a particular delay. The sufficient conditions are expressed in terms of statistical properties of initial estimates, i.e., estimates generated by an estimator parameterization from which the gradient descent procedure is initialized. More specifically, we have proven that when initial estimators result in sufficiently low CM cost, or in sufficiently low kurtosis *and* a particular variance, CM-GD will preserve the source/delay combination associated with the initial estimator. In addition, we have proven that when the SINR of the initial estimators (with respect to a particular source/delay combination) is above a prescribed threshold and the estimates have a particular variance, CM-GD will converge to an estimator of the same source/delay. These results suggest ways in which *a priori* channel knowledge may be used to predict and control the convergence behavior of CMA and are of particular importance in multiuser applications.

Appendix
Derivation Details for Local Convergence Conditions

This Appendix contains the proofs of the theorems and lemmas found in Section III-B.

A. Proof of Lemma 1

See [5] or [19].

B. Proof of Lemma 2

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

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

If (ℓ, i) represents the {source, delay} pair of minimum interference kurtosis [recall the definition of $\kappa_s^{\min}$ in (19)], we henceforth use $\breve{\boldsymbol{q}}^{(k)}$ to denote $\boldsymbol{q}^{(k)}$ with the terms $q_\nu^{(0)}$ and $q_i^{(\ell)}$ removed. Then, we have

$$\begin{aligned}&\min_{(\ell,i)\neq(0,\nu)} \sum_k \left(\kappa_s^{(k)} - \kappa_g\right) \left\|\boldsymbol{q}^{(k)}\right\|_4^4 \Big|_{q_\nu^{(0)} = q_i^{(\ell)}} \\ &\quad = \left(\kappa_s^{(0)} + \kappa_s^{\min} - 2\kappa_g\right) \left|q_\nu^{(0)}\right|^4 + \sum_k \left(\kappa_s^{(k)} - \kappa_g\right) \left\|\breve{\boldsymbol{q}}^{(k)}\right\|_4^4\end{aligned}$$

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

Plugging the two previous equations into (24), we find that

$$\min_{(\ell,i)\neq(0,\nu)} \min_{q_\nu^{(0)} = q_i^{(\ell)}} J_c(\boldsymbol{q}) \quad \Leftrightarrow \quad \min_{q_\nu^{(0)},\, \breve{\boldsymbol{q}}} J_c\left(q_\nu^{(0)}, \breve{\boldsymbol{q}}\right)$$

where

$$\begin{aligned}&\frac{J_c\left(q_\nu^{(0)}, \breve{\boldsymbol{q}}\right)}{\sigma_s^4} \\ &\quad := \left(\kappa_s^{(0)} + \kappa_s^{\min} - 2\kappa_g\right) \left|q_\nu^{(0)}\right|^4 + \sum_k \left(\kappa_s^{(k)} - \kappa_g\right) \left\|\breve{\boldsymbol{q}}^{(k)}\right\|_4^4 \\ &\qquad + \kappa_g \left(2\left|q_\nu^{(0)}\right|^2 + \|\breve{\boldsymbol{q}}\|_2^2\right)^2 \\ &\qquad - 2(\gamma/\sigma_s^2)\left(2\left|q_\nu^{(0)}\right|^2 + \|\breve{\boldsymbol{q}}\|_2^2\right) + (\gamma/\sigma_s^2)^2 \\ &\quad = \left(\kappa_s^{(0)} + \kappa_s^{\min} + 2\kappa_g\right)\left|q_\nu^{(0)}\right|^4 \\ &\qquad - 4\left((\gamma/\sigma_s^2) - \kappa_g \|\breve{\boldsymbol{q}}\|_2^2\right)\left|q_\nu^{(0)}\right|^2 + J_c(\breve{\boldsymbol{q}})/\sigma_s^4. \end{aligned} \tag{32}$$

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

$$\arg\min_{|q_\nu^{(0)}|^2} J_c\left(q_\nu^{(0)}, \breve{\boldsymbol{q}}\right) = 2\frac{(\gamma/\sigma_s^2) - \kappa_g\|\breve{\boldsymbol{q}}\|_2^2}{\kappa_s^{(0)} + \kappa_s^{\min} + 2\kappa_g} \tag{33}$$

and thus

$$\begin{aligned}&\frac{J_c\left(q_\nu^{(0)}\Big|_{\min}, \breve{\boldsymbol{q}}\right)}{\sigma_s^4} \\ &\quad = \frac{J_c(\breve{\boldsymbol{q}})}{\sigma_s^4} - 4\frac{\kappa_g^2\|\breve{\boldsymbol{q}}\|_2^4 - 2\kappa_g(\gamma/\sigma_s^2)\|\breve{\boldsymbol{q}}\|_2^2 + (\gamma/\sigma_s^2)^2}{\kappa_s^{(0)} + \kappa_s^{\min} + 2\kappa_g} \\ &\quad = \sum_k \left(\kappa_s^{(k)} - \kappa_g\right)\left\|\breve{\boldsymbol{q}}^{(k)}\right\|_4^4 + \kappa_g\, d\|\breve{\boldsymbol{q}}\|_2^4 - 2(\gamma/\sigma_s^2)\, d\|\breve{\boldsymbol{q}}\|_2^2 \\ &\qquad + (\gamma/\sigma_s^2)^2\left(1 - 4\left(\kappa_s^{(0)} + \kappa_s^{\min} + 2\kappa_g\right)^{-1}\right)\end{aligned} \tag{34}$$

using the abbreviation

$$d := \frac{\kappa_s^{(0)} + \kappa_s^{\min} - 2\kappa_g}{\kappa_s^{(0)} + \kappa_s^{\min} + 2\kappa_g}. \tag{35}$$

Gradient and Hessian analysis [19] reveals that when $\kappa_s^{\min} \leq 2\kappa_g - \kappa_s^{(0)}$, the (unique) global minimum of $J_c(q_\nu^{(0)}|_{\min}, \breve{\boldsymbol{q}})$ occurs at $\breve{\boldsymbol{q}} = \mathbf{0}$, implying [via (34)] that

$$\min_{\boldsymbol{q} \in \mathrm{bndr}(\mathcal{Q}_\nu^{(0)})} \frac{J_c(\boldsymbol{q})}{\sigma_s^4} = \min_{\breve{\boldsymbol{q}}} \frac{J_c\left(q_\nu^{(0)}\Big|_{\min}, \breve{\boldsymbol{q}}\right)}{\sigma_s^4} = \left(\frac{\gamma}{\sigma_s^2}\right)^2 \left(1 - \frac{4}{\kappa_s^{(0)} + \kappa_s^{\min} + 2\kappa_g}\right). \tag{36}$$

C. Proof of Theorem 1

If $\boldsymbol{q}_{\text{init}} \in \mathcal{Q}_\nu^{(0)}$ satisfies $J_c(\boldsymbol{q}_{\text{init}}) \leq J_c(\boldsymbol{q}_r)$, then by definition, $\boldsymbol{q}_{\text{init}} \in \mathcal{Q}_c(\boldsymbol{q}_r)$. (Note that for $\boldsymbol{q}_{\text{init}}$ to be meaningful, we also require that it be an attainable global response, i.e., $\boldsymbol{q}_{\text{init}} \in \mathcal{Q}_a$.) Combining $\mathcal{Q}_c(\boldsymbol{q}_r) \subset \mathcal{Q}_\nu^{(0)}$ with the fact that a CM-GD trajectory initialized within $\mathcal{Q}_c(\boldsymbol{q}_r)$ remains entirely within $\mathcal{Q}_c(\boldsymbol{q}_r)$, we conclude that a CM-GD trajectory initialized at $\boldsymbol{q}_{\text{init}}$ remains entirely within $\mathcal{Q}_\nu^{(0)}$. Using the $J_c(\boldsymbol{q}_r)$ expression (25) appearing in Lemma 2, we arrive at (26).

D. Proof of Theorem 2

Continuing the arguments used in the proof of Lemma 1, the CM cost expression (24) can be restated as follows [19].

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

From Theorem 1, a CM cost satisfying (26) suffices to guarantee the desired CM-GD property. Normalizing (26) by σ_s^4 and plugging in the previous expression, we obtain the equivalent sufficient conditions

$$0 > \frac{J_c(\boldsymbol{q})}{\sigma_s^4} - (\gamma/\sigma_s^2)^2 \left(1 - \frac{4}{\kappa_s^{(0)} + \kappa_s^{\min} + 2\kappa_g}\right)$$
$$0 > \kappa_y \|\boldsymbol{q}\|_2^4 - 2(\gamma/\sigma_s^2)\|\boldsymbol{q}\|_2^2 + (\gamma/\sigma_s^2)^2 \cdot \left(\frac{4}{\kappa_s^{(0)} + \kappa_s^{\min} + 2\kappa_g}\right)$$
$$\kappa_y < 2(\gamma/\sigma_s^2)\|\boldsymbol{q}\|_2^{-2} - (\gamma/\sigma_s^2)^2 \left(\frac{4}{\kappa_s^{(0)} + \kappa_s^{\min} + 2\kappa_g}\right) \|\boldsymbol{q}\|_2^{-4}.$$

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

$$2(\gamma/\sigma_s^2)(a^2)^{-1} - (\gamma/\sigma_s^2)^2 \left(\frac{4}{\kappa_s^{(0)} + \kappa_s^{\min} + 2\kappa_g}\right)(a^2)^{-2}$$

w.r.t. a^2. Straightforward calculus reveals that the maximizing value of $\|\boldsymbol{q}\|_2^2$ is

$$\|\boldsymbol{q}\|_2^2\big|_{\max} = \left(\frac{\gamma}{\sigma_s^2}\right)\left(\frac{4}{\kappa_s^{(0)} + \kappa_s^{\min} + 2\kappa_g}\right)$$

which implies that the maximum critical kurtosis is

$$\kappa_y = \left(\kappa_s^{(0)} + \kappa_s^{\min} + 2\kappa_g\right)/4.$$

Since S1)–S3) imply that $\sigma_y^2 = \|\boldsymbol{q}\|_2^2 \sigma_s^2$, the expression for $\|\boldsymbol{q}\|_2^2|_{\max}$ above is easily rewritten in terms of estimate variance σ_y^2.

E. Proof of Theorem 3

Section III-A established that estimators yielding gain a_* and producing estimates of SINR_ν greater than $\mathrm{SINR}_{\min,\nu}$ are contained within the set $\mathcal{Q}_c(\boldsymbol{q}_r)$, and thus, further CM-GD adaptation of these estimates will guarantee estimation of the desired source. In this section, we will derive explicit formulas for the quantities $\mathrm{SINR}_{\min,\nu}$ and a_*. This will be accomplished through (15) and (16) after first solving for $b_{\max}(a)$ defined in (14).

To find $b_{\max}(a)$, (14) may be translated as

$$b_{\max}(a) = \max\ b(a) \text{ s.t. } \left\{\|\bar{\boldsymbol{q}}\|_2 \leq b(a) \ \Rightarrow\ J_c(a, \bar{\boldsymbol{q}}) < J_c(\boldsymbol{q}_r)\right\}. \tag{37}$$

To proceed further, the CM cost expression (24) must be rewritten in terms of gain $a = \|\boldsymbol{q}\|_2$ and interference response $\bar{\boldsymbol{q}}$ (which was defined in Section II-B). Using the fact that $|q_\nu^{(0)}|^2 = a^2 - \|\bar{\boldsymbol{q}}\|_2^2$

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

Plugging the previous expression into (24), we find that

$$\begin{aligned}\frac{J_c(a, \bar{\boldsymbol{q}})}{\sigma_s^4} = \sum_k \left(\kappa_s^{(k)} - \kappa_g\right)\left\|\bar{\boldsymbol{q}}^{(k)}\right\|_4^4 + \kappa_s^{(0)} a^4 \\ - 2\left(\kappa_s^{(0)} - \kappa_g\right)a^2\|\bar{\boldsymbol{q}}\|_2^2 + \left(\kappa_s^{(0)} - \kappa_g\right)\|\bar{\boldsymbol{q}}\|_2^4 \\ - 2(\gamma/\sigma_s^2)a^2 + (\gamma/\sigma_s^2)^2.\end{aligned} \tag{38}$$

From (25) and (38), the following statements are equivalent, as shown in (39) at the bottom of the next page. The reversal of inequality in (39) occurs because $\kappa_s^{(0)} - \kappa_g < 0$ [as implied by S4)]. Using the definition of $\kappa_s^{\max}$ in (20), $0 \leq (\|\bar{\boldsymbol{q}}^{(k)}\|_4^4 / \|\bar{\boldsymbol{q}}^{(k)}\|_2^4) \leq 1$ implies that

$$\begin{aligned}\sum_k \left(\kappa_s^{(k)} - \kappa_g\right)\left\|\bar{\boldsymbol{q}}^{(k)}\right\|_4^4 &\leq (\kappa_s^{\max} - \kappa_g)\|\bar{\boldsymbol{q}}\|_4^4 \\ &\leq \begin{cases} 0, & \kappa_s^{\max} \leq \kappa_g \\ (\kappa_s^{\max} - \kappa_g)\|\bar{\boldsymbol{q}}\|_2^4, & \kappa_s^{\max} > \kappa_g. \end{cases}\end{aligned} \tag{40}$$

Thus, with $\rho_{\max}$ defined in (22), the following becomes a sufficient condition for (39).

$$0 > \begin{cases} \|\bar{\boldsymbol{q}}\|_2^4 - 2a^2\|\bar{\boldsymbol{q}}\|_2^2 + C(a, \boldsymbol{q}_r), & \kappa_s^{\max} \le \kappa_g \\ (1+\rho_{\max})\|\bar{\boldsymbol{q}}\|_2^4 - 2a^2\|\bar{\boldsymbol{q}}\|_2^2 + C(a, \boldsymbol{q}_r), & \kappa_s^{\max} > \kappa_g. \end{cases} \tag{41}$$

Focusing first on the super-Gaussian case ($\kappa_s^{\max} > \kappa_g$), we see from (41) that valid $b^2_{\max}(a)$ satisfying (37) can be determined by solving for the roots of

$$P_1(x) = (1+\rho_{\max})x^2 - 2a^2x + C(a, \boldsymbol{q}_r).$$

Specifically, we are interested in the smaller root when $1 + \rho_{\max} > 0$ and the larger root when $1 + \rho_{\max} < 0$. In either of these two cases, the appropriate root has the form

$$b^2_{\max}(a) = a^2 \left(\frac{1 - \sqrt{1 - (\rho_{\max}+1)\dfrac{C(a, \boldsymbol{q}_r)}{a^4}}}{\rho_{\max}+1} \right). \tag{42}$$

When $1 + \rho_{\max} = 0$ instead, $P_1(x)$ becomes linear, and

$$b^2_{\max}(a) = \frac{C(a, \boldsymbol{q}_r)}{2a^2}. \tag{43}$$

As a valid interference power, we require that $b^2_{\max}(a) \in [0, a^2]$. Straightforward manipulations show that for all valid super-Gaussian values of $\rho_{\max}$ (i.e., $\rho_{\max} < 0$)

$$b^2_{\max}(a) \in [0, a^2] \quad \Leftrightarrow \quad 0 \le \frac{C(a, \boldsymbol{q}_r)}{a^4} \le 1 - \rho_{\max}. \tag{44}$$

From (41), it can be seen that the same arguments may be applied to the nonsuper-Gaussian case ($\kappa_s^{\max} \le \kappa_g$) by setting $\rho_{\max}$ to zero. This yields

$$b^2_{\max}(a) = a^2 \left(1 - \sqrt{1 - \frac{C(a, \boldsymbol{q}_r)}{a^4}}\right) \tag{45}$$

with the requirement that $0 \le C(a, \boldsymbol{q}_r)/a^4 \le 1$.

The expressions for $b^2_{\max}(a)$ in (42), (43), and (45) can now be used to calculate $\mathrm{SINR}_{\min,\nu}$ and a_* given in (15) and (16).

First, we tackle the super-Gaussian case ($\kappa_s^{\max} > \kappa_g$). Assuming for the moment that $\rho_{\max} \ne -1$, we plug (42) into (15) to obtain

$$\mathrm{SINR}_{\min,\nu} = \min_a \frac{\rho_{\max} + \sqrt{1 - (\rho_{\max}+1)\dfrac{C(a, \boldsymbol{q}_r)}{a^4}}}{1 - \sqrt{1 - (\rho_{\max}+1)\dfrac{C(a, \boldsymbol{q}_r)}{a^4}}}. \tag{46}$$

Since the fraction on the right of (46) is non-negative and strictly decreasing in $C(a, \boldsymbol{q}_r)/a^4$ over the valid range $C(a, \boldsymbol{q}_r)/a^4 \in [0, 1 - \rho_{\max}]$ identified by (44), finding a that minimizes this expression [in accordance with (16)] can be accomplished by finding a that maximizes $C(a, \boldsymbol{q}_r)/a^4$. To find these maxima, we first write $C(a, \boldsymbol{q}_r)/a^4$ using (39):

$$\frac{C(a, \boldsymbol{q}_r)}{a^4} = C_0 + C_1 \cdot (a^2)^{-1} + C_2 \cdot (a^2)^{-2}$$

where C_0, C_1, and C_2 are independent of a. Computing the partial derivative with respect to the quantity a^2 and setting it equal to zero, we find that

$$a_*^2 = -2\frac{C_2}{C_1} = \left(\frac{\gamma}{\sigma_s^2}\right)\left(\frac{4}{\kappa_s^{(0)} + \kappa_s^{\min} + 2\kappa_g}\right). \tag{47}$$

Plugging a_*^2 into (39) and using the definition of $\rho_{\min}$ in (21) gives the simple result $C(a_*, \boldsymbol{q}_r)/a_*^4 = (3 - \rho_{\min})/4$. With this value of a_*, requirement (44) translates into

$$\kappa_s^{(0)} \le \min\left\{\frac{\kappa_s^{\min} + 2\kappa_g}{3}, \quad 4\kappa_s^{\max} - \kappa_s^{\min} - 2\kappa_g\right\}. \tag{48}$$

In the super-Gaussian case, we know that $\kappa_s^{(0)} < 4\kappa_s^{\max} - \kappa_s^{\min} - 2\kappa_g$, and hence, (48) simplifies to

$$\kappa_s^{(0)} \le \frac{\kappa_s^{\min} + 2\kappa_g}{3}.$$

Finally, plugging a_* into (46) gives

$$\mathrm{SINR}_{\min,\nu} = \frac{\rho_{\max} + \sqrt{1 - (1+\rho_{\max})(3-\rho_{\min})/4}}{1 - \sqrt{1 - (1+\rho_{\max})(3-\rho_{\min})/4}}. \tag{49}$$

$$\begin{aligned} J_c(\boldsymbol{q}_r) &> J_c(a, \bar{\boldsymbol{q}}) \\ 0 &> \sum_k \left(\kappa_s^{(k)} - \kappa_g\right)\left\|\bar{\boldsymbol{q}}^{(k)}\right\|_4^4 + \left(\kappa_s^{(0)} - \kappa_g\right)\left(-2a^2\|\bar{\boldsymbol{q}}\|_2^2 + \|\bar{\boldsymbol{q}}\|_2^4\right) \\ &\quad + \kappa_s^{(0)}a^4 - 2\left(\frac{\gamma}{\sigma_s^2}\right)a^2 + \left(\frac{\gamma}{\sigma_s^2}\right)^2\left(\frac{4}{\kappa_s^{(0)} + \kappa_s^{\min} + 2\kappa_g}\right) \\ 0 &< \frac{1}{\kappa_s^{(0)} - \kappa_g}\sum_k \left(\kappa_s^{(k)} - \kappa_g\right)\left\|\bar{\boldsymbol{q}}^{(k)}\right\|_4^4 - 2a^2\|\bar{\boldsymbol{q}}\|_2^2 + \|\bar{\boldsymbol{q}}\|_2^4 \\ &\quad + \underbrace{\frac{1}{\kappa_s^{(0)} - \kappa_g}\left(\kappa_s^{(0)}a^4 - 2\left(\frac{\gamma}{\sigma_s^2}\right)a^2 + \left(\frac{\gamma}{\sigma_s^2}\right)^2\left(\frac{4}{\kappa_s^{(0)} + \kappa_s^{\min} + 2\kappa_g}\right)\right)}_{C(a, \boldsymbol{q}_r)}. \end{aligned} \tag{39}$$

Revisiting the super-Gaussian case with $\rho_{\max} = -1$, we plug (43) into (15) and get

$$\mathrm{SINR}_{\min,\,\nu} = \min_a 2\left(\frac{C(a,\,\boldsymbol{q}_r)}{a^4}\right)^{-1} - 1.$$

Again, the quantity to be minimized is strictly decreasing in $C(a,\,\boldsymbol{q}_r)/a^4$ over $[0,\,1-\rho_{\max}]$. As above, maximization of $C(a,\,\boldsymbol{q}_r)/a^4$ yields the a_* of (47) and the same condition on $\kappa_s^{(0)}$. Applying these to the previous equation

$$\mathrm{SINR}_{\min,\,\nu} = \frac{5+\rho_{\min}}{3-\rho_{\min}}. \tag{50}$$

For the nonsuper-Gaussian case ($\kappa_s^{\max} \le \kappa_g$), we plug (45) into (15) and obtain

$$\mathrm{SINR}_{\min,\,\nu} = \min_a \frac{\sqrt{1-\dfrac{C(a,\,\boldsymbol{q}_r)}{a^4}}}{1-\sqrt{1-\dfrac{C(a,\,\boldsymbol{q}_r)}{a^4}}}. \tag{51}$$

Since (51) equals (46) with $\rho_{\max} = 0$ (i.e., when $\kappa_s^{\max} = \kappa_g$), the nonsuper-Gaussian will have the same a_* as (47) and the same translation of (44) given by (48). After substituting $\kappa_s^{\max} = \kappa_g$ into (48), the nonsuper-Gaussian property implies that (48) simplifies again to

$$\kappa_s^{(0)} \le \frac{\kappa_s^{\min} + 2\kappa_g}{3}.$$

Plugging a_* from (47) into (51), the nonsuper-Gaussian $\mathrm{SINR}_{\min,\,\nu}$ becomes

$$\mathrm{SINR}_{\min,\,\nu} = \frac{\sqrt{1+\rho_{\min}}}{2-\sqrt{1+\rho_{\min}}}. \tag{52}$$

Finally, S1)–S3) imply that $a = \|\boldsymbol{q}\|_2^2 = \sigma_y^2/\sigma_s^2$, linking the critical gain a_* in (47) to the critical estimate variance $\sigma_y^2|_{\mathrm{crit}}$ in (23), yielding Theorem 3.

References

[1] D. N. Godard, "Self-recovering equalization and carrier tracking in two-dimensional data communication systems," *IEEE Trans. Commun.*, vol. COMM-28, pp. 1867–1875, Nov. 1980.

[2] J. R. Treichler and B. G. Agee, "A new approach to multipath correction of constant modulus signals," *IEEE Trans. Acoust., Speech, Signal Processing*, vol. ASSP-31, pp. 459–472, Apr. 1983.

[3] C. R. Johnson, Jr., P. Schniter, T. J. Endres, J. D. Behm, D. R. Brown, and R. A. Casas, "Blind equalization using the constant modulus criterion: A review," *Proc. IEEE, Special Issue on Blind System Identification and Estimation*, vol. 86, pp. 1927–1950, Oct. 1998.

[4] J. J. Shynk and R. P. Gooch, "The constant modulus array for cochannel signal copy and direction finding," *IEEE Trans. Signal Processing*, vol. 44, pp. 652–660, Mar. 1996.

[5] P. Schniter and C. R. Johnson Jr., "Bounds for the MSE performance of constant modulus estimators," *IEEE Trans. Inform. Theory*, to be published.

[6] A. Batra and J. R. Barry, "Blind cancellation of co-channel interference," in *Proc. IEEE Global Telecommun. Conf.*, Singapore, Nov. 13–17, 1995, pp. 157–162.

[7] C. B. Papadias and A. J. Paulraj, "A constant modulus algorithm for multiuser signal separation in presence of delay spread using antenna arrays," *IEEE Signal Processing Lett.*, vol. 4, pp. 178–181, June 1997.

[8] A. Touzni, I. Fijalkow, M. Larimore, and J. R. Treichler, "A globally convergent approach for blind MIMO adaptive deconvolution," in *Proc. IEEE Int. Conf. Acoust., Speech , Signal Process.*, Seattle, WA, May 1998, pp. 2385–2388.

[9] J. R. Treichler and M. G. Larimore, "New processing techniques based on the constant modulus adaptive algorithm," *IEEE Trans. Acoust., Speech, Signal Processing*, vol. ASSP-33, pp. 420–431, Apr. 1985.

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

[11] L. Ljung, *System Identification: Theory for the User*, 2nd ed. Englewood Cliffs, NJ: Prentice Hall, 1999.

[12] W. Chung, "Geometrical understanding of the constant modulus algorithm: Adaptive blind equalization and cross-polarized source separation," M.S. thesis, Cornell Univ., Ithaca, NY, 1999.

[13] M. Gu and L. Tong, "Geometrical characterizations of constant modulus receivers," *IEEE Trans. Signal Processing*, vol. 47, pp. 2745–2756, Oct. 1999.

[14] J. R. Treichler and M. G. Larimore, "The tone capture properties of CMA-based interference suppressors," *IEEE Trans. Acoust., Speech, Signal Processing*, vol. ASSP-33, pp. 946–958, Aug. 1985.

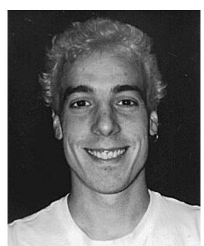
[15] Y. Li and Z. Ding, "Convergence analysis of finite length blind adaptive equalizers," *IEEE Trans. Signal Processing*, vol. 43, pp. 2120–2129, Sep. 1995.

[16] O. Shalvi and E. Weinstein, "New criteria for blind deconvolution of nonminimum phase systems (channels)," *IEEE Trans. Inform. Theory*, vol. 36, pp. 312–321, Mar. 1990.

[17] C. R. Johnson Jr., P. Schniter, I. Fijalkow, L. Tong, , and J. R. Treichler *et al.*, "The core of FSE-CMA behavior theory," in *Unsupervised Adaptive Filtering, Volume 2: Blind Deconvolution*, S. Haykin *et al.*, Eds. New York: Wiley, 2000.

[18] P. Schniter and C. R. Johnson Jr., "Minimum-entropy blind acquisition/equalization for uplink DS-CDMA," in *Proc. Allerton Conf. Commun., Contr., Comput.*, Monticello, IL, Oct. 1998, pp. 401–410.

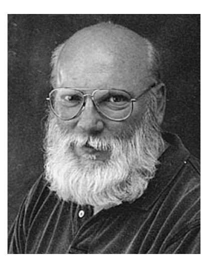
[19] P. Schniter, "Blind estimation without priors: Performance, convergence, and efficient implementation," Ph.D. dissertation, Cornell Univ., Ithaca, NY, May 2000.



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, Urbana-Champaign, in 1992 and 1993, respectively. In 2000, he received the Ph.D. degree in electrical engineering from Cornell University, Ithaca, NY.

From 1993 to 1996, he was with Tektronix Inc., 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. He is currently an Assistant Professor with the Department of Electrical Engineering, The Ohio State University, Columbus. His research interest is signal processing for communication systems, especially blind adaptive equalization.

Dr. Schniter received the 1998 Schlumberger Fellowship and the 1998–1999 Intel Foundation Fellowship with pursuing the Ph.D. degree. He received the 1999 Prize Paper Award from the IEEE Energy Development and Power Generation Committee for work relating to his M.S. thesis.



C. Richard Johnson, Jr. (F'89) 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, Stanford, CA, 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 signal processing, communication systems, system identification, and digital control 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, Lucent Technologies, and Applied Signal Technology. His current research interest is in adaptive parameter estimation theory that is 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.