SINR-based Sufficient Conditions for CMA Desired-User-Lock

Philip Schniter and C. Richard Johnson, Jr.
Cornell University, Ithaca, NY 14853.

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 CM local 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 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).**

I. Introduction

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. The CM criterion is defined below in terms of the estimates $\{y_n\}$ and a design parameter γ.

$$J_c(y_n) := \mathrm{E}\{(|y_n|^2 - \gamma)^2\}. \quad (1)$$

The popularity of the CM criterion is usually attributed to (i) 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 $\mathbf{f}_c(z)$, and (ii) the excellent MSE performance of CM-minimizing estimates. The second of these two points was first conjectured in the original works [1], [2] and recently proven by the authors for arbitrary linear channels and additive interference [4].

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.

Various "multiuser" modifications of the CM criterion have been proposed to jointly estimate all sub-Gaussian sources (e.g., [5], [6]). These methods, however, often require knowledge of the number of interfering sources, result in significant increase in computational complexity when the number of sources is large, and generate estimates with questionable MSE performance. In contrast, we focus on the 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 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 [7]. 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.

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 [8] and Li and Ding in [9].

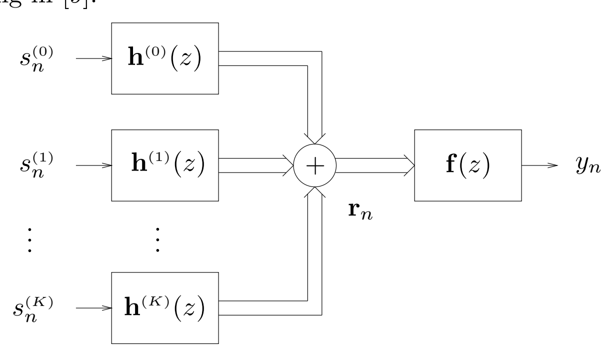

Fig. 1. Linear system model with K sources of interference.

II. Background

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

Supported in part by NSF grants MIP-9509011 and ECS-9811297, an Intel Foundation fellowship, and Applied Signal Technology, Inc.
Direct all correspondence to schniter@ee.cornell.edu.

notation is used throughout: $(\cdot)^t$ denotes transpose, $(\cdot)^H$ hermitian, and $\mathrm{E}\{\cdot\}$ expectation. Also, $\mathbf{I}$ denotes the identity matrix, $\mathbb{R}^+$ the field of non-negative real numbers, and $\|\mathbf{x}\|_p$ the p-norm defined by $\sqrt[p]{\sum_i |x_i|^p}$.

A. Linear System Model

First we formalize the linear time-invariant multi-channel model illustrated in Fig. 1. Say that the desired symbol sequence $\{s_n^{(0)}\}$ and K sources of interference $\{s_n^{(1)}\}, \dots, \{s_n^{(K)}\}$ each pass through separate linear "channels" before being observed at the receiver. In addition, say that the receiver uses a sequence of P-dimensional vector observations 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 $P \times 1$ received signal samples $\mathbf{r}_n$ can be written $\mathbf{r}_n = \sum_{k=0}^{K} \sum_{i=0}^{\infty} \mathbf{h}_i^{(k)} s_{n-i}^{(k)}$, where $\{\mathbf{h}_n^{(k)}\}$ denote the impulse response coefficients of the linear time-invariant (LTI) channel $\mathbf{h}^{(k)}(z)$. The only assumptions placed on $\mathbf{h}^{(k)}(z)$ are causality and bounded-input bounded-output (BIBO) stability. Note that such $\mathbf{h}^{(k)}(z)$ admit infinite impulse response (IIR) channel models.

From the vector-valued observation sequence $\{\mathbf{r}_n\}$, the receiver generates a sequence of linear estimates $\{y_n\}$ of $\{s_{n-\nu}^{(0)}\}$, where ν is a fixed integer. Using $\{\mathbf{f}_n\}$ to denote the impulse response of the linear estimator $\mathbf{f}(z)$, the estimates are formed as $y_n = \sum_{i=-\infty}^{\infty} \mathbf{f}_i^H \mathbf{r}_{n-i}$. We will assume that the linear system $\mathbf{f}(z)$ is BIBO stable, though not necessarily causal nor FIR.

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

$$q_n^{(k)} \;=\; \sum_{i=-\infty}^{\infty} \mathbf{f}_i^H \mathbf{h}_{n-i}^{(k)}, \tag{2}$$

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

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

For instance, the estimates can be rewritten concisely as

$$y_n \;=\; \sum_{k=0}^{K} \boldsymbol{q}^{(k)t} \boldsymbol{s}^{(k)}(n) \;=\; \boldsymbol{q}^t \boldsymbol{s}(n). \tag{3}$$

We now point out two important properties of $\boldsymbol{q}$. Let $\ell_1(\mathbb{C}^P)$ denote the set of absolutely-summable vector sequences, defined as follows

$$\{\mathbf{f}_i\} \in \ell_1(\mathbb{C}^P) \quad \Leftrightarrow \quad \sum_i \|\mathbf{f}_i\|_2 < \infty, \quad \mathbf{f}_i \in \mathbb{C}^P. \tag{4}$$

BIBO stable $\mathbf{f}(z)$ and $\mathbf{h}^{(k)}(z)$ imply that $\{\mathbf{f}_i\}, \{\mathbf{h}_i^{(k)}\} \in \ell_1(\mathbb{C}^P)$, which ensures that $\{q_i^{(k)}\} \in \ell_1(\mathbb{C})$. This, in turn, implies that $\|\boldsymbol{q}\|_a < \infty$ (for any $a \geq 1$).

Next, it is important to recognize that placing a particular structure on the channel and/or estimator will restrict the set of *admissible* channel-estimator responses, which we will denote by $\mathcal{Q}_{\mathrm{a}}$. For example, when the estimator is FIR, (2) implies that $\boldsymbol{q} \in \mathcal{Q}_{\mathrm{a}} = \mathrm{row}(\boldsymbol{\mathcal{H}})$, where

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

Restricting the estimator to be causal IIR, for example, would generate a different admissible set $\mathcal{Q}_{\mathrm{a}}$. In general, we allow any channel/estimator restrictions which ensure that $\mathcal{Q}_{\mathrm{a}}$ is a linear subspace of $\ell_1(\mathbb{C})$.

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) For $k \neq \ell$, $\{s_n^{(k)}\}$ is statistically independent of $\{s_n^{(\ell)}\}$.
S3) For all k, $\mathrm{E}\{|s_n^{(k)}|^2\} = \sigma_s^2$.
S4) $\mathcal{K}(s_n^{(0)}) < 0$, where $\mathcal{K}(\cdot)$ denotes kurtosis:

$$\mathcal{K}(s_n) \;:=\; \mathrm{E}\{|s_n|^4\} - 2\,\mathrm{E}^2\{|s_n|^2\} - \big|\mathrm{E}\{s_n^2\}\big|^2. \tag{5}$$

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

B. Signal to Interference-plus-Noise Ratio

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

$$y_n = q_\nu^{(0)} s_{n-\nu}^{(0)} + \bar{\boldsymbol{q}}^t \bar{\boldsymbol{s}}(n), \tag{6}$$

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 signal to interference-plus-noise ratio (SINR) associated with y_n, an estimate of $s_{n-\nu}^{(0)}$, is then defined

$$\mathrm{SINR}(y_n) \;:=\; \frac{\mathrm{E}\big\{|q_\nu^{(0)} s_{n-\nu}^{(0)}|^2\big\}}{\mathrm{E}\big\{|\bar{\boldsymbol{q}}^t \bar{\boldsymbol{s}}(n)|^2\big\}} \;=\; \frac{|q_\nu^{(0)}|^2}{\|\bar{\boldsymbol{q}}\|_2^2}, \tag{7}$$

where the equality invokes assumptions S1)–S3). Note that $\mathrm{SINR} = \sigma_s^2/\mathrm{UMSE}$, where UMSE denotes (conditionally) unbiased mean-squared error.

C. The Constant Modulus Criterion

The CM criterion, introduced independently in [1] and [2], was defined in (1) in terms of the estimates $\{y_n\}$. For sources, channels, and estimators allowed in Section II-A, the authors have shown [4] that

$$\mathrm{UMSE_c} \;\leq\; \mathrm{UMSE_m} + A \cdot \mathrm{UMSE_m^2} + \mathcal{O}(\mathrm{UMSE_m^3}),$$

where $\mathrm{UMSE_m}$ denotes to Wiener estimates of the desired source at a particular delay, $\mathrm{UMSE_c}$ denotes to the CM-minimizing estimates of the desired source at the same delay, and A is a constant that depends on source kurtoses. For example, $A = 1/2\sigma_s^2$ when $\mathcal{K}(\mathbf{s}_n^{(0)}) \leq \mathcal{K}(\mathbf{s}_n^{(k)}) \leq 0\ \forall k$.

III. Sufficient Conditions for Local Convergence of CM-GD

A. The Main Idea

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

$$\mathcal{Q}_\nu^{(0)} \;:=\; \Big\{ \boldsymbol{q} \;\text{ s.t. }\; |q_\nu^{(0)}| > \max_{(k,\delta)\neq(0,\nu)} |q_\delta^{(k)}| \Big\}. \tag{8}$$

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 channel-estimator responses on the boundary of $\mathcal{Q}_\nu^{(0)}$ with minimum CM cost,

$$\{\boldsymbol{q}_\mathrm{r}\} \;:=\; \arg \min_{\boldsymbol{q}\in \mathrm{bndr}(\mathcal{Q}_\nu^{(0)})} J_\mathrm{c}(\boldsymbol{q}), \tag{9}$$

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

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

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

Since the size and shape of $\mathcal{Q}_\mathrm{c}(\boldsymbol{q}_\mathrm{r})$ are not easily characterizable, we find it more useful to derive sufficient CM-GD initialization conditions in terms of more well-known statistical quantities such as kurtosis or SINR. It has been shown that CM cost and kurtosis are closely related [9], thus we expect that translation between these two quantities will be relatively straightforward. Translation of the initial CM-cost condition into an initial SINR condition is more difficult, but can be accomplished with the definition

$$\mathrm{SINR}_\mathrm{min} := \min_x \;\text{ s.t. }\; \Big\{ \forall \boldsymbol{q} : \mathrm{SINR}(\boldsymbol{q}) \ge x,\; \exists\, a_* \text{ s.t. } \tfrac{a_* \boldsymbol{q}}{\|\boldsymbol{q}\|_2} \in \mathcal{Q}_\mathrm{c}(\boldsymbol{q}_\mathrm{r}) \Big\}, \tag{10}$$

which says that an initialization in the set $\{\boldsymbol{q} : \mathrm{SINR}(\boldsymbol{q}) \ge \mathrm{SINR}_\mathrm{min}\}$ can be scaled so that the resulting CM-GD trajectory remains within $\mathcal{Q}_\mathrm{c}(\boldsymbol{q}_\mathrm{r})$ and hence within $\mathcal{Q}_\nu^{(0)}$. In other words, when a particular channel-estimator $\boldsymbol{q}$ yields sufficiently high SINR, CM-GD initialized at a properly scaled version of $\boldsymbol{q}$ will preserve its {source, delay} combination. The SINR condition is formalized below.

Since $\mathcal{Q}_\mathrm{c}(\boldsymbol{q})$ and $\mathrm{SINR}(\boldsymbol{q})$ are all invariant to phase rotation of $\boldsymbol{q}$ (i.e., scalar multiplication by $e^{j\phi}$ for $\phi \in \mathbb{R}$), we can (w.l.o.g.) restrict the our attention to the "de-rotated" set of channel-estimator 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). In terms of the pair $(a, \bar{\boldsymbol{q}})$, the SINR (7) can be written

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

so that (10) becomes

$$\mathrm{SINR}_\mathrm{min} \;:=\; \min_x \;\text{ s.t. }\; \Big\{ \forall (a, \bar{\boldsymbol{q}}) : \tfrac{a^2 - \|\bar{\boldsymbol{q}}\|_2^2}{\|\bar{\boldsymbol{q}}\|_2^2} \ge x,\; \exists\, a_* \text{ s.t. } (a_*, \tfrac{a_*}{a}\bar{\boldsymbol{q}}) \in \mathcal{Q}_\mathrm{c}(\boldsymbol{q}_\mathrm{r}) \Big\}. \tag{11}$$

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

$$b_\mathrm{max}(a) := \max_{b(a)} \;\text{ s.t. }\; \Big\{ \forall\, \bar{\boldsymbol{q}} : \|\bar{\boldsymbol{q}}\|_2 \le b(a),\; (a, \bar{\boldsymbol{q}}) \in \mathcal{Q}_\mathrm{c}(\boldsymbol{q}_\mathrm{r}) \Big\}. \tag{12}$$

For an illustration of a, $b_\mathrm{max}(a)$, and $\mathcal{Q}_\mathrm{c}(\boldsymbol{q}_\mathrm{r})$, see Fig. 2.

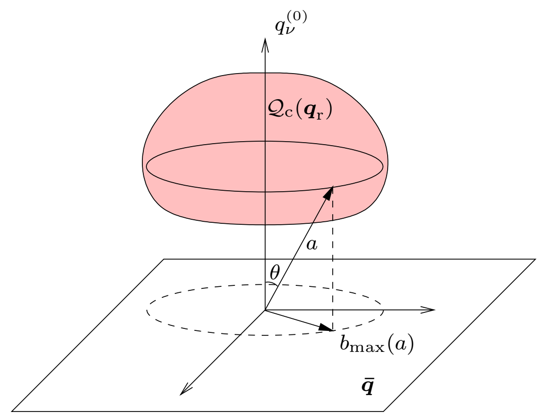

Fig. 2. Illustration of maximum interference gain $b_\mathrm{max}(a)$ below which all channel-estimator responses of gain a are contained in the CM cost region $\mathcal{Q}_\mathrm{c}(\boldsymbol{q}_\mathrm{r})$. Note that $\mathrm{SINR}(a, b_\mathrm{max}) = \cot^2(\theta)$.

Now, consider the quantity

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

Since $\mathrm{SINR}(a, b_\mathrm{max})$ is a decreasing function of $b_\mathrm{max}(a)$ (over its valid domain), definition (12) implies that

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

Using the previous expression to minimize SINR in accordance with (11) yields

$$\mathrm{SINR}_\mathrm{min} \;=\; \min_a \mathrm{SINR}(a, b_\mathrm{max}) \tag{13}$$

$$a_* \;=\; \arg\min_a \mathrm{SINR}(a, b_\mathrm{max}). \tag{14}$$

B. Statement of Sufficient Conditions

In this section we formalize the previously-described initialization conditions for CM-GD desired-source convergence. The main results are presented as theorems whose proofs can be found in [10].

It is convenient to now define the *normalized* kurtosis (not to be confused with $\mathcal{K}(\cdot)$ in (5)):

$$\kappa_s^{(k)} \;:=\; \frac{\mathrm{E}\{|s_n^{(k)}|^4\}}{\mathrm{E}^2\{|s_n^{(k)}|^2\}}. \tag{15}$$

Under the following definition of κ_g,

$$\kappa_g \;:=\; \begin{cases} 3, & \forall k,n:\ \mathbf{h}^{(k)}(z)\in\mathbb{R}^P \text{ and } s_n^{(k)}\in\mathbb{R} \\ 2, & \text{otherwise,} \end{cases} \tag{16}$$

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.

$$\kappa_s^{\min} \;:=\; \begin{cases} \min_{0\le k\le K} \kappa_s^{(k)}, & \dim(\boldsymbol{q}^{(0)}) > 1, \\ \min_{1\le k\le K} \kappa_s^{(k)}, & \dim(\boldsymbol{q}^{(0)}) = 1, \end{cases} \tag{17}$$

$$\kappa_s^{\max} \;:=\; \max_{0\le k\le K} \kappa_s^{(k)}. \tag{18}$$

Note that the second case of (17) applies only when the desired source contributes zero intersymbol interference (ISI).

Before statement of the main results, we introduce three more kurtosis-based quantities that will appear later.

$$\rho_{\min} \;:=\; \frac{\kappa_g - \kappa_s^{\min}}{\kappa_g - \kappa_s^{(0)}} \tag{19}$$

$$\rho_{\max} \;:=\; \frac{\kappa_g - \kappa_s^{\max}}{\kappa_g - \kappa_s^{(0)}} \tag{20}$$

$$\sigma_y^2\big|_{\text{crit}} \;:=\; \gamma\left(\frac{4}{\kappa_s^{(0)} + \kappa_s^{\min} + 2\kappa_g}\right). \tag{21}$$

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 $\boldsymbol{q}_{\text{init}} \in \mathcal{Q}_\nu^{(0)} \cap \mathcal{Q}_{\text{a}}$) with CM cost

$$J_{\text{c}}(y_n) \;<\; \gamma^2\left(1 - \frac{4}{\kappa_s^{(0)} + \kappa_s^{\min} + 2\kappa_g}\right), \tag{22}$$

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 $\boldsymbol{q}_{\text{init}} \in \mathcal{Q}_\nu^{(0)} \cap \mathcal{Q}_{\text{a}}$) with variance $\sigma_y^2 = \sigma_y^2\big|_{\text{crit}}$ and normalized kurtosis

$$\kappa_y \;<\; \kappa_y^{\text{crit}} \;:=\; \frac{1}{4}\big(\kappa_s^{(0)} + \kappa_s^{\min} + 2\kappa_g\big), \tag{23}$$

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 $\sigma_y^2 = \sigma_y^2\big|_{\text{crit}}$ and $\mathrm{SINR}(y_n) > \mathrm{SINR}_{\min}$,

$$\mathrm{SINR}_{\min} = \begin{cases} \frac{\sqrt{1+\rho_{\min}}}{2-\sqrt{1+\rho_{\min}}}, & \kappa_s^{\max} \le \kappa_g, \\ \left\{ \begin{array}{l} \frac{5+\rho_{\min}}{3-\rho_{\min}} \text{ when } \rho_{\max} = -1, \text{ else} \\ \frac{\rho_{\max}+\sqrt{1-(1+\rho_{\max})(3-\rho_{\min})/4}}{1-\sqrt{1-(1+\rho_{\max})(3-\rho_{\min})/4}} \end{array} \right\}, & \kappa_s^{\max} > \kappa_g, \end{cases} \tag{24}$$

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

IV. Numerical Examples

In Fig. 3, CM-GD minimization trajectories conducted in estimator space are plotted in channel-estimator space ($\boldsymbol{q} \in \mathbb{R}^2$) to demonstrate the key results of this paper. In all experiments, a two-tap FIR channel with $P = 2$ was chosen, corresponding to channel matrix $\boldsymbol{\mathcal{H}} = \left(\begin{smallmatrix} 0.2940 & -0.0596 \\ 0.1987 & 0.9801 \end{smallmatrix}\right)$ of condition number 3. The subplots of Fig. 3 correspond to different source kurtoses, as described in the caption.

First observe that all trajectories entering into $\mathcal{Q}_{\text{c}}(\boldsymbol{q}_{\text{r}})$ (denoted by the shaded region between the dash-dotted lines) converge to an estimator for the desired source, confirming Theorem 1. Next, observe that 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, observe that 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.

From Fig. 3 it is evident that initial kurtosis or SINR is not sufficient for desired local convergence; initial estimator gain plays an important role. Though recognized in [8], this fact was overlooked in the work of Li and Ding [9], rendering incorrect some of their 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 $\mathbf{f}_{\text{m}}$ (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 with $\kappa_s^{(k)} = 2$; while Fig. 4(c) demonstrates $\kappa_s^{(0)} = 2$, five interfering sources with $\kappa_s^{(k)} = 2$, and five with $\kappa_s^{(k)} = 4$. Fig. 4

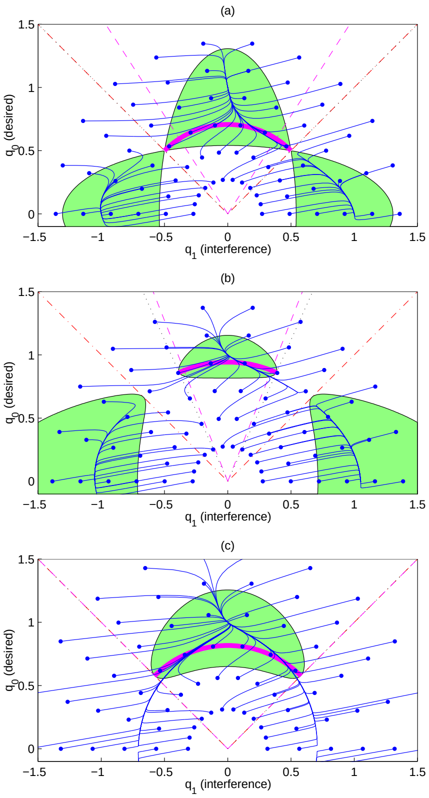

Fig. 3. CM-GD trajectories in channel-estimator space ($\boldsymbol{q} \in \mathbb{R}^2$) for (a) $\kappa_s^{(0)} = 1$ and $\kappa_s^{(1)} = 1$, (b) $\kappa_s^{(0)} = 2$ and $\kappa_s^{(1)} = 1$, and (c) $\kappa_s^{(0)} = 2$ and $\kappa_s^{(1)} = 4$. Also shown are $\mathcal{Q}_\nu^{(0)}$ boundaries (dash-dotted), $\mathrm{SINR}_{\min}$ boundaries (dashed), κ_y^{crit} boundaries (dotted), and $J_c(\boldsymbol{q}) < J_c(\boldsymbol{q}_r)$ regions (shaded). Channel-estimators resulting in σ_y^2 of (21) shown by the fat shaded arcs.

also confirms the claim of Theorem 3: all properly-scaled CM-GD initializations with SINR greater than $\mathrm{SINR}_{\min}$ converge to the desired source.

V. Conclusions

In this paper we have derived, under a general linear model and general source properties, three sufficient conditions for the convergence of CM-minimizing gradient descent to an linear 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 under a parameterization from which CM-GD is initialized. More specifically, we have shown that when initial estimates result in sufficiently low CM cost, or sufficiently low kurtosis *and* a prescribed variance, CM-GD will preserve the source/delay combination associated with the initial estimator. In addition, we have shown that when the SINR of the initial estimates (with respect to a particular source/delay combination) is sufficiently high *and* the estimates have a prescribed variance, CM-GD will converge to an estimator for that source/delay.

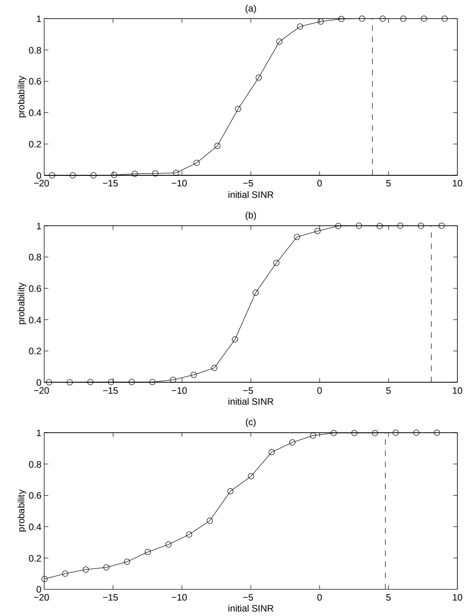

Fig. 4. Estimated probability of convergence to desired source/delay for random channels and random initializations scaled according to (21) as a function of initialization SINR. In (a) $\kappa_s^{(0)} = 1$ with interfering $\kappa_s^{(k)} \in \{1, 3\}$, in (b) $\kappa_s^{(0)} = 2$ with interfering $\kappa_s^{(k)} \in \{1, 2\}$, and in (c) $\kappa_s^{(0)} = 2$ with interfering $\kappa_s^{(k)} \in \{2, 4\}$. $\mathrm{SINR}_{\min}$ from (24) shown by dashed lines.

References

[1] D.N. Godard, "Self-recovering equalization and carrier tracking in two-dimensional data communication systems," *IEEE Trans. Comm.*, vol. 28, no.11, 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, and Signal Processing*, vol. ASSP-31, no.2, pp. 459-72, 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*, vol. 86, no. 10, pp. 1927-50, Oct. 1998.

[4] P. Schniter and C.R. Johnson, Jr., "Bounds for the MSE performance of constant modulus estimators," submitted to *IEEE Trans. Inform. Theory*, June 1999.

[5] 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, no. 6, pp. 178-81, June 1997.

[6] J.R. Treichler and M.G. Larimore, "New processing techniques based on the constant modulus adaptive algorithm," *IEEE Trans. Acoust., Speech, and Signal Processing*, vol. ASSP-33, no.2, pp. 420-31, Apr. 1985.

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

[8] J.R. Treichler and M.G. Larimore, "The tone capture properties of CMA-based interference suppressors," *IEEE Trans. Acoust., Speech, and Signal Processing*, vol. ASSP-33, no.4, pp. 946-58, Aug. 1985.

[9] Y. Li and Z. Ding, "Convergence analysis of finite length blind adaptive equalizers," *IEEE Trans. on Signal Processing*, vol. 43, no. 9, pp. 2120-9, Sep. 1995.

[10] P. Schniter and C.R. Johnson, Jr., "Sufficient conditions for the local convergence of constant modulus algorithms," submitted to *IEEE Trans. on Signal Processing*, June 1999.