Stochastic Stability Analysis for the Constant-Modulus Algorithm (CMA)

Vítor H. Nascimento*, *Member, IEEE*, and Magno T. M. Silva, *Member, IEEE*

Abstract

We derive an easy-to-compute approximate bound for the range of step-sizes for which the constant-modulus algorithm (CMA) will remain stable if initialized close to a minimum of the CM cost function. Our model highlights the influence of the signal constellation used in the transmission system: for smaller variation in the modulus of the transmitted symbols, the algorithm will be more robust, and the steady-state misadjustment will be smaller. The theoretical results are validated through several simulations, for long and short filters and channels.

Index Terms

Adaptive filters, adaptive equalizers, tracking, least mean square methods, recursive estimation, unsupervised learning.

I. Introduction

When using a gradient-based adaptive filter, such as the constant-modulus algorithm (CMA) or the least-mean squares algorithm (LMS), it is important to have an estimate of the range in which the step-size must remain to guarantee an adequate (stable) behavior of the algorithm [1], [2]. For supervised algorithms, such as LMS, normalized LMS and other variants, approximations for this range are well-known: from the simpler and more practical earlier results, which assume Gaussian, independent regressors [3], to exact results for more general models, but which are only practical for very short filters [4]. For blind algorithms, to the best of our knowledge, available works that provide approximations for the mean-square behavior of blind algorithms either arrive at complex recursions whose stability is not easy to study [5], [6], or rely on linearization arguments in such a way that only stability in the mean is guaranteed [7]. This last kind of analysis can prove that the algorithm will converge

This work was partly supported by CNPq under grant No. 303361/2004-2 and by FAPESP under grant No. 2004/15114-2.

V. H. Nascimento and M. T. M. Silva are with the Dept. of Electronic Systems Eng., University of São Paulo, São Paulo, SP, Brazil, e-mails: {vitor, magno}@lps.usp.br, ph. +55-11-3091-5606, fax: +55-11-3091-5718.

SP EDICS: 1.ASP-ANAL, 2.SPC-BLND.

Corresponding author: Vítor H. Nascimento.

for “sufficiently small” step-sizes, but can provide only very loose bounds for the range of step-sizes that guarantee convergence.

A recent result [8] proves that CMA is in fact always *unstable* when the noise is not bounded (as, e.g., Gaussian noise). It is also proven that the algorithm has useful stability properties for small step-sizes when a finite time horizon is considered. Using a different approach, [9] shows that the least-mean fourth algorithm (LMF) is also always mean-square unstable when the regressor is Gaussian, and [10] computes a range of step-sizes for which the algorithm will be stable in a finite horizon. In this paper, we extend some of these latter results to CMA.

In [11] we proposed a simple model for the real-valued constant-modulus algorithm, that we used to study convex combinations of CMA with the Shalvi-Weinstein algorithm (SWA). In this correspondence we use and extend that model to obtain explicit stability conditions for CMA. The stability of CMA in fact depends on its initial condition (a similar dependence appears in the LMF algorithm, see, e.g., [10]). In the simplified model presented here, this dependence is not described; instead we assume that the algorithm was initialized close to the optimum solution, providing the largest range of the step-size for which stable performance of CMA is possible. Our assumptions regarding initialization are similar to what was used in the first work on the LMF algorithm [12], whose model was latter improved in works such as [9], [10], [13].

II. THE CONSTANT-MODULUS ALGORITHM

The data transmission problem with which we will work is depicted in Fig. 1. A sequence $\{a(n)\}$ of data is transmitted through a channel, which in general will distort and add noise ($\eta(n)$ in the figure) to the transmitted signal. The equalizer should approximately invert the effect of the (mostly linear) distortion, without significantly amplifying the noise. The constant-modulus algorithm attempts to perform this task blindly, i.e., without the help of a training sequence. Although there is more than one variant of CMA, we will focus on the so-called CMA2-2, whose performance was found in previous works to be better than that of other variants [14].

CMA2-2 will update a weight vector $\boldsymbol{w}_n$ through the recursion

$$\boldsymbol{w}_{n+1} = \boldsymbol{w}_n + \mu e(n)\boldsymbol{x}_n, \tag{1}$$

where $\boldsymbol{x}_n \in \mathbb{R}^M$ is a column-vector. The estimation error is $e(n) = (r_a - y^2(n))y(n)$, where $y(n) = \boldsymbol{x}_n^T \boldsymbol{w}_n$ is the filter output, $r_a = \mathrm{E}\, a^4(n)/\,\mathrm{E}\, a^2(n)$ is a positive constant, E denotes expected value and $(\cdot)^T$ denotes transposition. To keep the discussion short, in this correspondence we will consider only the case of real signals. We will assume that the sampling rate at the equalizer is higher than the symbol rate, i.e., we will consider fractionally-spaced equalizers (FSEs).

As Fig. 1 shows for $L = 2$, in an FSE the regressor $\boldsymbol{x}_n$ is a concatenation of L tap-delay lines, where L is the up-sampling factor [14]. For $L = 2$, $\boldsymbol{x}_n = \left[\boldsymbol{x}_{\mathrm{e},n}^T \ \boldsymbol{x}_{\mathrm{o},n}^T\right]^T$, $\boldsymbol{x}_{\mathrm{e},n} = \left[x_{\mathrm{e}}(n) \ x_{\mathrm{e}}(n-1) \ \dots \ x_{\mathrm{e}}(n-M/2+1)\right]^T$ and $\boldsymbol{x}_{\mathrm{o},n} =$

$\left[x_o(n)\ x_o(n-1)\ \ldots\ x_o(n-M/2+1) \right]^T$, where $x_e(n) = x(2n)$ are the even, and $x_o(n)$, the odd samples of $x(n)$. The full weight vector $\boldsymbol{w}_n$ is also a concatenation: $\boldsymbol{w}_n = \left[\boldsymbol{w}_{e,n}^T\ \boldsymbol{w}_{o,n}^T \right]^T$.

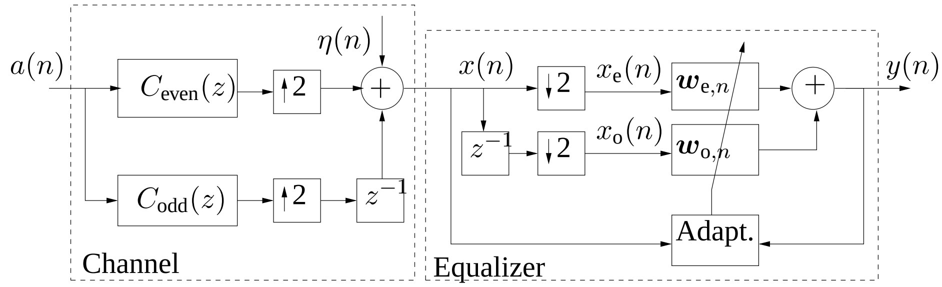

Fig. 1. A fractionally-spaced equalizer (assuming an oversampling by a factor of 2). In this model, the equalizer is adapted once two new samples arrive at the receiver.

The property of FSEs that is of interest to us is that a linear equalizer may achieve zero-forcing equalization in the absence of noise [15], such that there exists an optimum weight vector $\boldsymbol{w}_*$ such that $\boldsymbol{w}_*^T \boldsymbol{x}_n = a(n - \tau_d)$, where τ_d is a delay. Note that other optimum choices would also exist, e.g., $\boldsymbol{w}'_* = -\boldsymbol{w}_*$. Since the performance of the equalizer would be the same for these alternative solutions if an adequate rotation is applied to the output of the filter, we assume, without loss of generality, that the algorithm is initialized so that it converges to $\boldsymbol{w}_*$.

III. A MODEL FOR CMA

In order to find a simple model for CMA which allows determination of the stability range for the algorithm's step-size, we need to make, as is usually the case for the study of adaptive filters, some simplifying assumptions in order to make analysis tractable. In this section we describe and justify the assumptions made in this correspondence.

First, for stability analysis it is easier to consider that the statistics of the input signal to the equalizer is time-invariant, so our first assumption is

A-1 *The channel is time-invariant and* $\{a(n)\}$, $\{\eta(n)\}$ *and* $\{\boldsymbol{x}_n\}$ *are stationary and have zero mean. We also assume that* $\{a(n)\}$ *and* $\{\eta(n)\}$ *are both iid (independent, identically distributed) and independent of each other (see also Assumption A-6 further ahead).*

Second, we want to describe the filter output in terms of the transmitted data $\{a(n)\}$ and of the difference between the estimated parameter vector $\boldsymbol{w}_n$ and the zero-forcing vector $\boldsymbol{w}_*$, as was done in [14]. When the SNR at the receiver is high, the optimum parameter vector will be close to $\boldsymbol{w}_*$, so our next assumption is

A-2 *The noise power* $\sigma_\eta^2 = \mathrm{E}\,\eta^2(n)$ *is small enough for the zero-forcing solution* $\boldsymbol{w}_*$ *to be one of the global minimizers of the CM cost-function* $J(\boldsymbol{w}) = (r_a - (\boldsymbol{w}^T \boldsymbol{x}_n)^2)^2$. *Under this assumption,* $\boldsymbol{w}_*^T \boldsymbol{x}_n \approx a(n - \tau_d)$.

Under Assumptions A-1 and A-2, we can define $\tilde{\boldsymbol{w}}_n = \boldsymbol{w}_* - \boldsymbol{w}_n$, and write the filter output as

$$y(n) = \boldsymbol{w}_n^T \boldsymbol{x}_n = (\boldsymbol{w}_* - \tilde{\boldsymbol{w}}_n)^T \boldsymbol{x}_n \approx a(n - \tau_d) - \tilde{\boldsymbol{w}}_n^T \boldsymbol{x}_n \stackrel{\Delta}{=} a(n - \tau_d) - e_a(n), \tag{2}$$

where we defined the *a-priori* error $e_{\mathrm{a}}(n) = \tilde{\boldsymbol{w}}_n^T \boldsymbol{x}_n$.

Next we need to simplify the CMA recursion (1). Most difficulties arise from the nonlinear error term, $e(n)$. Using Assumption A-2 and (2), we can write

$$\begin{aligned} e(n) &= y(n)\big(r_a - y^2(n)\big) \approx \big(a(n-\tau_{\mathrm{d}}) - e_{\mathrm{a}}(n)\big)\left(r_a - \big(a(n-\tau_{\mathrm{d}}) - e_{\mathrm{a}}(n)\big)^2\right) \\ &= -a^3(n-\tau_{\mathrm{d}}) + e_{\mathrm{a}}^3(n) + 3a^2(n-\tau_{\mathrm{d}})e_{\mathrm{a}}(n) - 3a(n-\tau_{\mathrm{d}})e_{\mathrm{a}}^2(n) + a(n-\tau_{\mathrm{d}})r_a - e_{\mathrm{a}}(n)r_a. \end{aligned} \tag{3}$$

To simplify this expression, we assume that higher powers of $e_{\mathrm{a}}(n)$ can be disregarded, for all $n \geq 0$:

A-3 *The initial condition $\boldsymbol{w}_0$ is close enough to $\boldsymbol{w}_*$ (i.e., $e_a(n)$ is small enough) so that* (3) *reduces to*

$$e(n) \approx \big(3a^2(n-\tau_d) - r_a\big)\, e_a(n) - a^3(n-\tau_d) + a(n-\tau_d) r_a. \tag{4}$$

Defining $\gamma(n) = 3a^2(n-\tau_{\mathrm{d}}) - r_a$ and $\beta(n) = a(n-\tau_{\mathrm{d}})\left(r_a - a^2(n-\tau_{\mathrm{d}})\right)$, we obtain $e(n) \approx \gamma(n)e_{\mathrm{a}}(n) + \beta(n)$. This relation has a number of interesting interpretations, as we show next (some of these properties were shown in [11]). We will need an extra assumption:

A-4 *The constellation used to generate the $a(n)$ has circular symmetry, so that $\mathrm{E}\, a^k(n) = 0$ for all odd integers $k > 0$. This assumption is not restrictive, since these conditions are true for practical constellations.*

Lemma 1 *Under Assumptions A-1 and A-4, $\beta(n)$ has zero mean and is uncorrelated with $\boldsymbol{x}_n$.*

Proof: $\mathrm{E}\,\beta(n) = \mathrm{E}\left(r_a a(n-\tau_{\mathrm{d}}) - a^3(n-\tau_{\mathrm{d}})\right) = 0$, if $\mathrm{E}\, a(n-\tau_{\mathrm{d}}) = \mathrm{E}\, a^3(n-\tau_{\mathrm{d}}) = 0$.

To see that $\beta(n)$ and $\boldsymbol{x}_n$ are uncorrelated, recall that each entry of $\boldsymbol{x}_n$, $x(n-k)$, $0 \leq k \leq M-1$, is a noisy linear combination of the transmitted symbols $a(n)$: $x(n) = \sum_{k=0}^{K} c_i(k) a(n-k) + \eta(n)$, where the $c_i(k)$ are the coefficients of the impulse response of one of the sub-channels. If n is even, the even sub-channel is used ($i=0$), if n is odd, $i=1$ and the coefficients of the odd sub-channel are used. Both sub-channels are assumed to have effective length $K+1$. From the iid part of Assumption A-1, we obtain

$$\mathrm{E}\, x(m)\beta(n) = \sum_{k=0}^{K} \mathrm{E}\left\{\left(r_a a(n-\tau_{\mathrm{d}}) - a^3(n-\tau_{\mathrm{d}})\right) c_i(k) a(m-k)\right\} + \mathrm{E}\left\{\left(r_a a(n-\tau_{\mathrm{d}}) - a^3(n-\tau_{\mathrm{d}})\right)\eta(m)\right\}.$$

From assumption A-1, the last term in the right-hand side is zero. Assumptions A-1 and A-4 guarantee that the expected values in the sum are zero whenever $n - \tau_{\mathrm{d}} \neq m - k$. On the other hand,

$$\mathrm{E}\left\{\left(r_a a(n-\tau_{\mathrm{d}}) - a^3(n-\tau_{\mathrm{d}})\right) c_i(m+\tau_{\mathrm{d}}-n) a(n-\tau_{\mathrm{d}})\right\} = c_i(m+\tau_{\mathrm{d}}-n)\left(r_a\, \mathrm{E}\, a^2(n-\tau_{\mathrm{d}}) - \mathrm{E}\, a^4(n-\tau_{\mathrm{d}})\right) = 0,$$

since the channel is constant and $r_a = \mathrm{E}\, a^4(n) / \mathrm{E}\, a^2(n)$. ■

Using a similar argument, and the symmetry of $a(n)$ (Assumption A-4), it can be shown that $\mathrm{E}\, x^2(m)\beta(n) = \mathrm{E}\, x(m)\beta^2(n) = 0$ for all m, n.

Next, note that if $a(n)$ has indeed a constant modulus, then $\beta(n) \equiv 0$. We can say that (when the SNR is high) $\beta(n)$ plays a role similar to that of the measurement noise in system identification. We should expect CMA to present a very low variance in steady-state about the optimum solution when the constellation has constant modulus, and a larger variance as $|a(n)|$ is allowed to vary (see also [11]).

In the sequel, we will also need to evaluate the expected values $\mathrm{E}\,\beta^2(n)x(n-m_1)x(n-m_2)$, $\mathrm{E}\,\gamma(n)x(n-m_1)x(n-m_2)$ and $\mathrm{E}\,\gamma^2(n)x(n-m_1)x(n-m_2)x(n-m_3)x(n-m_4)$, for integer $m_1 \dots m_4$. Consider the second of these quantities (Below, $i=0,1$, depending if $n-m_1$ is even or odd, and similarly for j and $n-m_2$. We used Assumption (A-1) to simplify terms involving the noise):

$$\mathrm{E}\left\{\gamma(n)x(n-m_1)x(n-m_2)\right\} = \mathrm{E}\left\{ \left(3a^2(n-\tau_{\mathrm{d}})-r_a\right)\cdot\left[\sum_{k=0}^{K} c_i(k)a(n-m_1-k)\right]\left[\sum_{\ell=0}^{K} c_j(\ell)a(n-m_2-\ell)\right]\right\}+ \\ +\bar{\gamma}\,\mathrm{E}\{\eta(n-m_1)\eta(n-m_2)\}, \tag{5}$$

where we defined $\bar{\gamma}=\mathrm{E}\{\gamma(n)\}$. Recalling A-1 and A-4, the expected value of $\mathrm{E}\,a^{(2\ell+1)}(m)a(k)=0$ if $m\neq k$ for any integer $\ell>0$. Defining $\sigma_a^2=\mathrm{E}\,a^2(n)$, $\sigma_p=\mathrm{E}\,a^p(n)$ and $\Delta=m_1-m_2$, the only nonzero terms of (5) are

$$\mathrm{E}\left\{\gamma(n)x(n-m_1)x(n-m_2)\right\} = \bar{\gamma}\sigma_\eta^2\delta(\Delta) - r_a\sigma_a^2\sum_{\substack{k=0\\0\le k+\Delta\le K}}^{K} c_i(k)c_j(k+\Delta) + 3\sigma_a^4\sum_{\substack{k=0\\k\neq\tau_{\mathrm{d}}-m_1\\0\le k+\Delta\le K}}^{K} c_i(k)c_j(k+\Delta) \\ +3\sigma_4 c_i(\tau_{\mathrm{d}}-m_1)c_j(\tau_{\mathrm{d}}-m_2), \tag{6}$$

where $\delta(n)$ is the Kronecker delta. On the other hand, expanding $(\mathrm{E}\,\gamma(n))\,(\mathrm{E}\{x(n-m_1)x(n-m_2)\})$, we obtain

$$(\mathrm{E}\,\gamma(n))\,(\mathrm{E}\,x(n-m_1)x(n-m_2)) = \bar{\gamma}\sigma_\eta^2\delta(\Delta) - r_a\sigma_a^2\sum_{\substack{k=0\\0\le k+\Delta\le K}}^{K} c_i(k)c_j(k+\Delta) + 3\sigma_a^4\sum_{\substack{k=0\\0\le k+\Delta\le K}}^{K} c_i(k)c_j(k+\Delta). \tag{7}$$

Although these equations can be further simplified by replacing the value of r_a, we can already see that the difference between (6) and (7) is only the term with σ_4 that had to be separated from the second sum in (6). Since $\mathrm{E}\,a^4(n)=(1+\epsilon_2)\left(\mathrm{E}\,a^2(n)\right)^2$ with $\epsilon_2\ge 0$, we conclude that (6) and (7) will be approximately equal if the channel is long enough and ϵ_2 is small enough for the term $3\epsilon_2\sigma_a^4 c_i(\tau_{\mathrm{d}}-m_1)c_j(\tau_{\mathrm{d}}-m_2)$ to be small next to the two sums in (6). Similar considerations can be made for $\mathrm{E}\,\beta^2(n)x(n-m_1)x(n-m_2)$ and $\mathrm{E}\,\gamma^2(n)x(n-m_1)x(n-m_2)x(n-m_3)x(n-m_4)$. We thus assume

A-5 *The channel is long enough, and the constellation has small enough constants* ϵ_p, *where* $\mathrm{E}\,a^{2p}(n)=(1+\epsilon_p)\left(\mathrm{E}\,a^2(n)\right)^p$, $p=1\dots4$, *so that*

$$\mathrm{E}\beta^2(n)x(n-m_1)x(n-m_2)\approx\mathrm{E}\beta^2(n)\,\mathrm{E}x(n-m_1)x(n-m_2),$$
$$\mathrm{E}\,\gamma(n)x(n-m_1)x(n-m_2)\approx\mathrm{E}\,\gamma(n)\,\mathrm{E}\,x(n-m_1)x(n-m_2),$$
$$\mathrm{E}\,\gamma^2(n)x(n-m_1)x(n-m_2)x(n-m_3)x(n-m_4)\approx\mathrm{E}\,\gamma^2(n)\,\mathrm{E}\,x(n-m_1)x(n-m_2)x(n-m_3)x(n-m_4).$$

A. Convergence in the mean

Subtracting the zero-forcing solution from both sides in (1) and using (4), we obtain (below, $\boldsymbol{I}$ is the identity matrix and $\tilde{\boldsymbol{w}}_n = \boldsymbol{w}_* - \boldsymbol{w}_n$)

$$\tilde{\boldsymbol{w}}_{n+1} = \tilde{\boldsymbol{w}}_n - \mu e(n)\boldsymbol{x}_n \approx \left(\boldsymbol{I} - \mu\gamma(n)\boldsymbol{x}_n\boldsymbol{x}_n^T\right)\tilde{\boldsymbol{w}}_n - \mu\beta(n)\boldsymbol{x}_n. \tag{8}$$

This expression is very similar to the weight-error vector recursion for LMS, only in the case of LMS we would have $\gamma(n) \equiv 1$, and instead of $\beta(n)$ we would have the measurement noise. Since $\beta(n)$ is uncorrelated with $\boldsymbol{x}_n$, from this point on the analysis of CMA is similar to well-known results about LMS [2], [3], [16].

Taking the expected value of this expression, we find a recursion for $\mathrm{E}\,\tilde{\boldsymbol{w}}_n$ if we make the next assumption

A-6 *The vectors $\boldsymbol{x}_n$ and $\tilde{\boldsymbol{w}}_n$ are independent.*

This assumption would be true if the vector sequence $\{\boldsymbol{x}_n\}$ were iid. Although this is never true for an equalizer, it is well-known that approximations of the mean $\mathrm{E}\,\tilde{\boldsymbol{w}}_n$ and covariance $\mathrm{E}\,\tilde{\boldsymbol{w}}_n\tilde{\boldsymbol{w}}_n^T$ of the weight-error vector are good for small step-sizes. Approximations for the range of step-sizes obtained using this assumption tend to overestimate the range of allowed step-sizes, but not so much as to render the estimated range useless [1].

Using Assumptions A-5 and A-6 and Lemma 1, the expected value of (8) reduces to

$$\mathrm{E}\,\tilde{\boldsymbol{w}}_{n+1} \approx \left(\boldsymbol{I} - \mu\,\mathrm{E}\{\gamma(n)\}\,\mathrm{E}\{\boldsymbol{x}_n\boldsymbol{x}_n^T\}\right)\mathrm{E}\,\tilde{\boldsymbol{w}}_n.$$

Defining the autocorrelation matrix $\boldsymbol{R} = \mathrm{E}\{\boldsymbol{x}_n\boldsymbol{x}_n^T\}$ and recalling that $\bar{\gamma} = \mathrm{E}\{\gamma(n)\}$, we see that CMA converges in the mean if $0 < \mu < 1/(\bar{\gamma}\lambda_1)$, where λ_1 is the largest eigenvalue of $\boldsymbol{R}$.

B. Mean-square stability

Multiplying (8) by its transpose, taking the expected value, and using Assumptions A-5 and A-6, we obtain a recursion for the autocorrelation $\boldsymbol{S}_n \triangleq \mathrm{E}\,\tilde{\boldsymbol{w}}_n\tilde{\boldsymbol{w}}_n^T$:

$$\boldsymbol{S}_{n+1} = \boldsymbol{S}_n - \mu\bar{\gamma}\boldsymbol{R}\boldsymbol{S}_n - \mu\bar{\gamma}\boldsymbol{S}_n\boldsymbol{R} + \mu^2 r_\gamma^2\,\mathrm{E}\left\{\boldsymbol{x}_n\boldsymbol{x}_n^T\boldsymbol{S}_n\boldsymbol{x}_n\boldsymbol{x}_n^T\right\} + \mu^2\sigma_\beta^2\boldsymbol{R}, \tag{9}$$

where we defined $\sigma_\beta^2 \triangleq \mathrm{E}\,\beta^2(n)$, $r_\gamma^2 \triangleq \mathrm{E}\,\gamma^2(n)$. We obtained a similar recursion in [11], but there we assumed that μ is small enough so that the fourth term on the right-hand side could be ignored. This approximation is good for a steady-state analysis, as was the case in [11]. However, all terms in (9) must be considered when the goal is to estimate the range of step-sizes for stable filter behavior. For that, we need to make a further assumption on the statistics of $\boldsymbol{x}_n$. Recalling that the entries of $\boldsymbol{x}_n$ are $x(m) = \sum_{k=0}^{K} c_i(k)a(m-k)$, where the $a(m-k)$ are iid, we see that the distribution of $x(m)$ will approach the Gaussian if the channel is long. The assumption below is thus reasonable for long channels:

A-7 *The channel is long enough for the fourth-order moments of $\boldsymbol{x}_n$ to be well approximated by those of a Gaussian vector, so that [3] ($\mathrm{Tr}(\boldsymbol{A})$ is the trace of matrix $\boldsymbol{A}$):*

$$\mathrm{E}\left\{\boldsymbol{x}_n\boldsymbol{x}_n^T\boldsymbol{S}_n\boldsymbol{x}_n\boldsymbol{x}_n^T\right\} \approx 2\boldsymbol{R}\boldsymbol{S}_n\boldsymbol{R} + \boldsymbol{R}\,\mathrm{Tr}\{\boldsymbol{R}\boldsymbol{S}_n\}.$$

The recursion for $\boldsymbol{S}_n$ then reads

$$\boldsymbol{S}_{n+1} = \boldsymbol{S}_n - \mu\bar{\gamma}\boldsymbol{R}\boldsymbol{S}_n - \mu\bar{\gamma}\boldsymbol{S}_n\boldsymbol{R} + 2\mu^2 r_\gamma^2\boldsymbol{R}\boldsymbol{S}_n\boldsymbol{R} + \mu^2 r_\gamma^2\,\mathrm{Tr}\{\boldsymbol{R}\boldsymbol{S}_n\}\boldsymbol{R} + \mu^2\sigma_\beta^2\boldsymbol{R}. \tag{10}$$

Since $\boldsymbol{S}_n$ is nonnegative-definite, the stability of (10) is determined through a recursion for the diagonal of the rotated $\boldsymbol{U}^T\boldsymbol{S}_n\boldsymbol{U}$, where $\boldsymbol{U}$ is the orthogonal transformation that diagonalizes $\boldsymbol{R}$, i.e., $\boldsymbol{U}^T\boldsymbol{U} = \boldsymbol{I}$, $\boldsymbol{U}^T\boldsymbol{R}\boldsymbol{U} = \boldsymbol{\Lambda}$, where $\boldsymbol{\Lambda} = \mathrm{diag}(\lambda_1, \dots, \lambda_M)$ and $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_M$ are the eigenvalues of $\boldsymbol{R}$. Defining $\boldsymbol{s}_n = \mathrm{diag}(\boldsymbol{U}^T\boldsymbol{S}_n\boldsymbol{U})$ (the diagonal elements of $\boldsymbol{U}^T\boldsymbol{S}_n\boldsymbol{U}$) and $\boldsymbol{l} = [\,\lambda_1\ \dots\ \lambda_M\,]^T$, (10) reduces to

$$\boldsymbol{s}_{n+1} = \left[\boldsymbol{I} - 2\mu\bar{\gamma}\boldsymbol{\Lambda} + \mu^2 r_\gamma^2\left(2\boldsymbol{\Lambda}^2 + \boldsymbol{l}\boldsymbol{l}^T\right)\right]\boldsymbol{s}_n + \mu^2\sigma_\beta^2\boldsymbol{l}. \tag{11}$$

An exact condition for the stability of (11) can be obtained as follows [16]. First, note that the system matrix $\boldsymbol{A}$ of (11) is positive-definite for all $\mu > 0$, since

$$\boldsymbol{A} = \boldsymbol{I} - 2\mu\bar{\gamma}\boldsymbol{\Lambda} + \mu^2 r_\gamma^2\left(2\boldsymbol{\Lambda}^2 + \boldsymbol{l}\boldsymbol{l}^T\right) = (\boldsymbol{I} - \mu\bar{\gamma}\boldsymbol{\Lambda})^2 + \mu^2(2r_\gamma^2 - \bar{\gamma}^2)\boldsymbol{\Lambda}^2 + \mu^2 r_\gamma^2\boldsymbol{l}\boldsymbol{l}^T,$$

and $r_\gamma^2 = \mathrm{E}\,\gamma^2(n) \geq \{\mathrm{E}\,\gamma(n)\}^2 = \bar{\gamma}^2$, so we need to guarantee only that the largest eigenvalue of $\boldsymbol{A}$ is less than 1. This, on the other hand, is equivalent to $\boldsymbol{B} = \boldsymbol{I} - \boldsymbol{A}$ being positive-definite. A necessary and sufficient condition for a matrix to be positive-definite is that all its principal minors (i.e., all the determinants of the principal sub-matrices) are positive [17]. This requires that (i) the diagonal entries of $\boldsymbol{B}$ be positive, i.e.,

$$2\mu\bar{\gamma}\lambda_k - 3\mu^2 r_\gamma^2\lambda_k^2 > 0 \Leftrightarrow 0 < \mu < \frac{2\bar{\gamma}}{3r_\gamma^2\lambda_k}, \quad k = 1\dots M, \tag{12}$$

and (ii) the determinants of all $k \times k$ principal sub-matrices are positive. These determinants can be evaluated using the relation $\det(\boldsymbol{I} - \boldsymbol{a}\boldsymbol{b}^T) = 1 - \boldsymbol{b}^T\boldsymbol{a}$, leading to the conditions

$$2^i\prod_{j=1}^{i}(\bar{\gamma}\lambda_j - \mu r_\gamma^2\lambda_j^2)\left[1 - \frac{\mu r_\gamma^2}{2}\sum_{k=1}^{i}\frac{\lambda_k}{\bar{\gamma} - \mu r_\gamma^2\lambda_k}\right] > 0, \quad 1 \leq i \leq M.$$

These conditions reduce to

$$\frac{\mu r_\gamma^2}{2}\sum_{k=1}^{M}\frac{\lambda_k}{\bar{\gamma} - \mu r_\gamma^2\lambda_k} < 1. \tag{13}$$

A simpler, sufficient but not necessary condition to guarantee stability of (11) can be obtained as follows: since $\mathrm{Tr}\{\boldsymbol{R}\} \geq \lambda_k$, it follows that

$$\frac{\mu r_\gamma^2}{2}\frac{\mathrm{Tr}\{\boldsymbol{R}\}}{\bar{\gamma} - \mu r_\gamma^2\,\mathrm{Tr}\{\boldsymbol{R}\}} = \frac{\mu r_\gamma^2}{2}\sum_{k=1}^{M}\frac{\lambda_k}{\bar{\gamma} - \mu r_\gamma^2\,\mathrm{Tr}\{\boldsymbol{R}\}} \geq \frac{\mu r_\gamma^2}{2}\sum_{k=1}^{M}\frac{\lambda_k}{\bar{\gamma} - \mu r_\gamma^2\lambda_k}.$$

If μ is small enough to make the left-hand side of the above expression less than one, then (13) will also be satisfied. A range of step-sizes that guarantees stability of (11) is thus

$$0 < \mu < \frac{2\bar{\gamma}}{3r_{\gamma}^{2}\,\mathrm{Tr}\{\boldsymbol{R}\}}. \tag{14}$$

IV. Simulations

Due to the cubic term in $y^3(n)$ appearing in the CMA recursion (1), we can expect the overall behavior of CMA to be similar to that of LMF, which also has a cubic nonlinearity [9]. Both algorithms present dependence on the initial condition, and a probability of divergence that is larger for large step-sizes and poor initial conditions. We say that a given run of the algorithm diverged when the Euclidean norm $\|\boldsymbol{w}_n\|$ of the weight vector becomes unbounded. In the simulations, we label a given run of the algorithm as "diverging" if $e(n)$ overflows (we check for NaNs in Matlab). The probability of divergence P_{div} is estimated from N_{r} runs of the algorithm by

$$P_{\text{div}} = \frac{\text{Number of curves diverging}}{N_{\text{r}}}.$$

In fact, [8] shows that, if the regressor $\boldsymbol{x}_n$ were indeed Gaussian, CMA would be mean-square unstable no matter how small the step-size (a similar result for LMF was proven in [9] using an alternative approach.) However, the mean-square stability analysis provided here predicts accurately the range of step-sizes for which the probability of divergence of a realization of CMA is small (see also [10] for LMF). When the regressor is not exactly Gaussian (which is the case in practical applications) what happens is the following:

1) For very small step-sizes and an initial condition close enough to a local minimum, CMA will converge and stay close to that local minimum;
2) As the step-size is increased, still with a good initial condition, the algorithm may switch between local minima (see Figure 3), and occasionally diverge (i.e., the weights will become unbounded), with a small probability that increases with the step-size;
3) For large step-sizes, the algorithm will always diverge;
4) Whether a given step-size should be considered "small" or "large" depends on how close the initial condition is to a local optimum of the CM cost function;
5) If the initial condition is far enough from a local optimum so that approximation (4) for $e(n)$ is not valid, then, for a fixed step-size, the algorithm will be much more likely to diverge in the first iterations, while the weights are far from the optimum. In this case, the probability of divergence is virtually independent of the number of iterations in a given run (see [9], [10] for equivalent results for LMF);
6) If the initial condition is good (close to a minimum) *and the step-size is large enough for divergence to occur*, the probability of divergence will depend on the number of iterations considered. If there are no local minima

close together, the probability of divergence $P_{\text{div, }N}$ in N iterations will be equal to $1-(1-P_{\text{div, }N/k})^k$, where $P_{\text{div, }N/k}$ is the probability of divergence in N/k iterations (see Figure 2).

We observed in the many simulations that we performed, that the probability of divergence starts being significative (e.g., larger than 0.1% for 10^4 iterations) for a step-size between the two bounds, (14) and (13). A step-size as large as half of (14) will make the probability of divergence virtually zero (no divergence observed in 10^3 runs of 2×10^5 iterations, or 10^5 runs of 10^4 iterations.)

In the following we present a few of our simulations. In all cases, we use pulse-amplitude modulation (PAM) constellations, either PAM-4 ($a(n) = \{\pm\alpha, \pm 3\alpha\}$) or PAM-6 ($a(n) = \{\pm\alpha, \pm 3\alpha, \pm 5\alpha\}$), with α chosen so that $\mathrm{E}\, a^2(n) = 1$, and FSEs with $L = 2$. We tested both long and short channels, and long and short adaptive filters. Except for Figure 4, the algorithm was initialized close to a minimum of the CM cost-function, as required by Assumption A-3. The initial condition was obtained by running the algorithm once with a small step-size (less than half the approximate bound), and saving the final weight vector after the algorithm converged to a good solution (with an open eye). In the following, the step-size predicted by the approximate bound (14) is denoted μ_{a}, and μ_{e} for (13). In almost all plots, the step-sizes are normalized by μ_{a} (i.e., we plot the probabilities of divergence as functions of μ/μ_{a}.)

The first examples are for channel C_1, with impulse response $\{0.1, 0.3, 1.0, -0.1, 0.5, 0.2\}$ in the absence of noise, with PAM-4 and an FSE with a total of $M = 4$ coefficients. Figure 2 shows the probability of divergence (P_{div}) as a function of the step-size. One of the curves is the observed P_{div} for $N_{\text{r}} = 10^3$ runs of $N_{\text{it}} = 10^4$ iterations each, the other for $N_{\text{it}} = 3 \times 10^4$. We included a plot of the probability predicted for the longer simulation, using $P_{\text{div},3\times 10^4} \approx 1 - (1 - P_{\text{div},10^4})^3$. The observed and estimated plots are close to each other – they are not identical, because in this simulation CMA has a few local minima close together, and with medium-large step-sizes, the algorithm tends to switch between these minima. The probability of divergence depends a little on how deep a minimum is. In Figure 3 we provide a plot of the coefficients of $\boldsymbol{w}_n$, with $\mu = 1.4\mu_{\text{a}}$, showing the filter switching between two states.

Figure 4 shows the probability of divergence for a fixed step-size (equal to μ_{a}), as a function of the initial condition: the initial condition for each point in the curve was chosen as $p\boldsymbol{w}_*$, where p is a scalar, and $\boldsymbol{w}_*$ is the optimum weight vector, evaluated experimentally. The plot shows that the requirement that the algorithm be initialized close to a minimum is not too restrictive: the probability of divergence stays small in a reasonably-sized ball around the minimum.

Additive noise at the receiver will increase the probability of divergence, but our results are still valid if the signal-to-noise ratio is not too low, as can be seen in Figure 5.

Figure 6 shows the results of simulations under different conditions: The solid line is a plot of P_{div} as a function

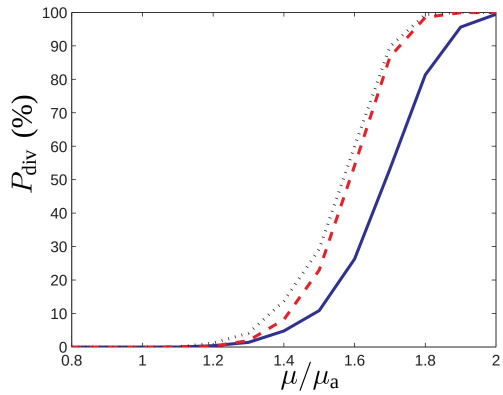

Fig. 2. Probability of divergence as a function of the step-size for PAM-4, for channel C_1. Equalizer with $M = 4$ coefficients, $\mu_{\text{a}} = 0.0418$, $\mu_{\text{e}} = 0.0814$. Solid line (–): 10^3 runs of 10^4 iterations each; broken line (- -): 3×10^4 iterations; dotted line ($\cdots$): P_{div} for 3×10^4 iterations, estimated as described in the text from the results for 10^4 iterations.

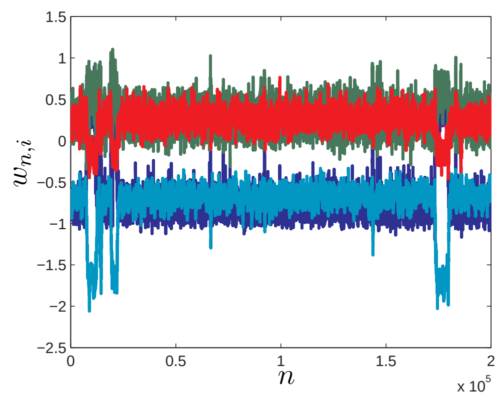

Fig. 3. Filter weights $w_{n,i}$ for PAM-4, for channel C_1, $\mu = 1.4\mu_{\text{a}}$, $M = 4$, $\mu = \mu_{\text{a}}$.

of the step-size for the same conditions as in Figure 2, but with PAM-6 instead of PAM-4. The other two curves are examples with longer channels: the real part of the microwave channels from [18]. The files have 300 coefficients, although the effective length of the channels is much shorter (e.g., for `chan7.mat`, 99% of the energy is between samples 20 and 55). The curve in broken line is P_{div} for the channel in file `chan7.mat` (equalizer with $M = 6$ coefficients, PAM-4), and the last curve uses the channel from `chan9.mat`, and an equalizer with $M = 50$

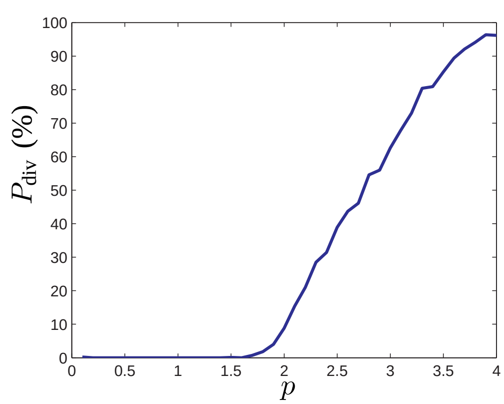

Fig. 4. Probability of divergence as a function of the initial condition for PAM-4, for channel C_1. The filter had $M = 4$ coefficients, with $\mu = \mu_{\text{a}} = 0.0418$, $\mu_{\text{e}} = 0.0814$. $N_{\text{r}} = 10^3$ runs of $N_{\text{it}} = 10^4$ iterations each.

coefficients and PAM-4. In all cases, the simulations were performed with $N_{\text{it}} = 10^4$ iterations, $N_{\text{r}} = 10^3$ runs, and no noise. In this figure, we plot the probabilities of divergence normalizing the step-size by the approximate bound μ_{a} (top) and by the more precise bound μ_{e} (bottom). It can be seen that the approximate bound (14) does guarantee a reasonable performance for the filter.

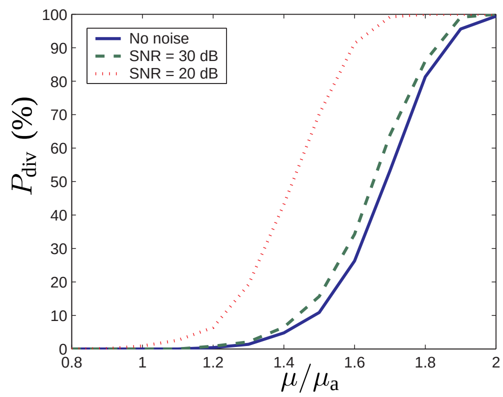

Fig. 5. Probability of divergence in the same conditions as in Figure 2, but with additive noise at the receiver. For SNR = 30 dB, $\mu_{\text{a}} = 0.0425$, $\mu_{\text{e}} = 0.0813$; for SNR = 20 dB, $\mu_{\text{a}} = 0.0417$, $\mu_{\text{e}} = 0.0804$. Solid line (–): no noise; broken line (- -): SNR = 30 dB; dotted line ($\cdots$): SNR = 20 dB.

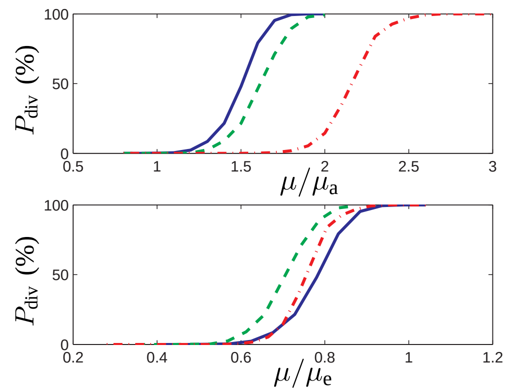

Fig. 6. Probability of divergence as a function of the step-size. Above (a), the step-size is normalized by μ_{a}; below (b), by μ_{e}. $N_{\text{r}} = 10^3$ runs of $N_{\text{it}} = 10^4$ iterations each. Solid line (–): PAM-6, channel C_1, $\mu_{\text{a}} = 0.0367$, $\mu_{\text{e}} = 0.0706$. Broken line (- -): PAM-4, $M = 6$, the channel is the real part of `chan7.mat` from [18], $\mu_{\text{a}} = 0.0288$, $\mu_{\text{e}} = 0.0658$. Line-dot (-·): PAM-4, $M = 50$, real part of `chan9.mat` from [18], $\mu_{\text{a}} = 0.0024$, $\mu_{\text{e}} = 0.0068$.

V. Conclusions

In this paper we derive a simple expression for the range of step-sizes for which the constant-modulus algorithm remains stable. To our knowledge, this is the first easy-to-compute bound for CMA, which should be a valuable help for designers. We validate our theoretical model with several simulations, for long and short filters and channels. The simulations show that the behavior of CMA is qualitatively different from that of LMS and similar to that of LMF: for a range of step-sizes, the algorithm may diverge or not in a given run, with a probability of divergence that depends on how close the initial condition is to a local minimum, the step-size, and the noise level.

REFERENCES

[1] A. H. Sayed, *Fundamentals of Adaptive Filtering*. Wiley-Interscience, 2003.

[2] P. S. R. Diniz, *Adaptive Filtering: Algorithms and Practical Implementation*, 2nd ed. Kluwer Academic Publishers, 2002.

[3] A. Feuer and E. Weinstein, "Convergence analysis of LMS filters with uncorrelated Gaussian data," *IEEE Trans. Acoust., Speech, Signal Process.*, vol. ASSP-33, no. 1, pp. 222–229, Feb. 1985.

[4] S. Florian and A. Feuer, "Performance analysis of the LMS algorithm with a tapped delay line (two-dimensional case)," *IEEE Trans. Acoust., Speech, Signal Process.*, vol. ASSP-34, no. 6, pp. 1542–1549, Dec. 1986.

[5] R. Cusani and A. Laurenti, "Convergence analysis of the CMA blind equalizer," *IEEE Trans. Commun.*, vol. 43, no. 2/3/4, pp. 1304–1307, Feb./Mar./Apr. 1995.

[6] L. M. Garth, "A dynamic convergence analysis of blind equalization algorithms," *IEEE Trans. Commun.*, vol. 49, no. 4, pp. 624–634, Apr. 2001.

[7] Y. Li and K. J. R. Liu, "Static and dynamic behavior of adaptive blind equalizers," *IEEE Trans. Signal Process.*, vol. 44, no. 11, pp. 2736–2745, Nov. 1996.

[8] O. Dabeer and E. Masry, "Convergence analysis of the constant modulus algorithm," *IEEE Trans. Inf. Theory*, vol. 49, no. 6, pp. 1447–1464, Jun. 2003.

[9] V. H. Nascimento and J. C. M. Bermudez, "Probability of divergence for the least-mean fourth (LMF) algorithm," *IEEE Trans. Signal Process.*, vol. 54, no. 4, pp. 1376–1385, Apr. 2006.

[10] J. C. M. Bermudez, P. I. Hübscher, and V. H. Nascimento, "A mean-square stability analysis of the least mean fourth (LMF) adaptive algorithm," *IEEE Trans. Signal Process.*, vol. 55, no. 8, pp. 4018–4028, Aug. 2007.

[11] M. T. M. Silva and V. H. Nascimento, "Improving the tracking capability of adaptive filters via convex combination," 2008, to appear, IEEE Trans. Signal Process.

[12] E. Walach and B. Widrow, "The least mean fourth (LMF) adaptive algorithm and its family," *IEEE Trans. Inf. Theory*, vol. IT-30, no. 2, pp. 275–283, Mar. 1984.

[13] S. Koike, "Stability conditions for adaptive algorithms with non-quadratic error criteria," in *EUSIPCO2000*, Tampere, Finland, Sep. 2000, pp. 131–134.

[14] J. Y. Mai and A. H. Sayed, "A feedback approach to the steady-state performance of fractionally spaced blind adaptive equalizers," *IEEE Trans. Signal Process.*, vol. 48, no. 1, pp. 80–91, Jan. 2000.

[15] J. R. Treichler, I. Fijalkow, and C. R. Johnson Jr., "Fractionally spaced equalizers," *IEEE Signal Process. Mag.*, vol. 13, pp. 65–81, May 1996.

[16] V. Solo and X. Kong, *Adaptive Signal Processing Algorithms*. Englewood Cliffs, NJ: Prentice Hall, 1995.

[17] R. A. Horn and C. R. Johnson, *Matrix Analysis*. Cambridge University Press, 1987.

[18] "Signal processing information base," IEEE Signal Processing Society, 2007. [Online]. Available: http://spib.rice.edu/spib/microwave.html