AVOIDING DIVERGENCE IN THE CONSTANT MODULUS ALGORITHM

Maria D. Miranda, Magno T. M. Silva, and Vítor H. Nascimento

Escola Politécnica, University of São Paulo
Av. Prof. Luciano Gualberto, trav. 3, nº 158, CEP 05508-900 – São Paulo, SP, Brazil
E-mails: maria@lcs.poli.usp.br, {vitor, magno}@lps.usp.br

ABSTRACT

One of the most popular algorithms for blind equalization is the Constant Modulus Algorithm (CMA), due to its simplicity and low computational cost. However, if the step-size is not properly chosen or if the initialization is distant from the optimal solution, CMA can diverge or converge to undesirable local minima. In order to avoid divergence, we propose a dual-mode algorithm, which works as CMA with a time-variant step-size, but rejects non-consistent estimates of the transmitted signal. We present a deterministic analysis of the stability of the new algorithm for scalar filters. In the vector case, the good performance of the new algorithm is confirmed through numerical simulations.

Index Terms— Adaptive filters, blind equalization, Constant Modulus Algorithm, dual-mode algorithms, stability.

1. INTRODUCTION

Blind equalizers are used in modern digital communication systems to remove intersymbol interference introduced by dispersive channels. They avoid the repeated transmission of training signals, optimizing the use of the channel capacity [1]. The Constant Modulus Algorithm (CMA) [2] and the Shalvi-Weinstein Algorithm (SWA) [3] are the most popular for the adaptation of these equalizers. Due to the equivalence between the Constant Modulus and Shalvi-Weinstein cost functions shown in [4], CMA and SWA seek to optimize the same criterion, presenting similar convergence problems. Thus, an inadequate choice of the step-size (resp., forgetting factor) of CMA (resp., SWA) in conjunction with an initialization distant from the zero-forcing solution can lead them to diverge (i.e., the norm of the weight vector goes to infinity) or to converge to undesirable local minima.

The convergence and stability of constant-modulus-based algorithms have been the subject of research for many years. Many important results have been obtained (see, e.g., [1, 5, 6] and the references therein). In particular, [5] analyzed the convergence of a stop-and-go variant of CMA for constant modulus signals, considering a deterministic approach based on the feedback framework of [7]. This analysis is based on the assumption that the equalizer output is uniformly bounded in a finite interval and stability is ensured due to the stop-and-go nature of the algorithm. However, the major drawback of stop-and-go-based algorithms is the operation mode in which the algorithm stops updating the coefficients until a predefined condition is satisfied.

To avoid divergence, we propose a dual-mode algorithm which never stops adjusting the equalizer coefficients, unlike stop-and-go-based algorithms. In the first mode, the algorithm works as a normalized CMA: CMA with a time-variant step-size proportional to the inverse of the squared Euclidean norm of the input regressor vector. In the second mode, the algorithm does not use the estimate (obtained with higher-order statistics) of the transmitted signal, updating the coefficient vector with an error proportional to the equalizer output. The switching rule between the two operation modes is the same used to avoid divergence in SWA, which was recently proposed in [8]. It is based on a rule to check the consistency of the estimate of the transmitted signal. If this consistency rule is satisfied, the algorithm will work in the region of interest and can reach a stationary point of the cost function. Otherwise, the estimate is rejected and a modified update formula is used to make the algorithm return to the region of interest. Furthermore, we give a proof that the proposed algorithm is stable for scalar filters. In spite of the lack of a stability proof in the vector case, results of exhaustive numerical simulations suggest that the algorithm also does not diverge in this case, with a performance similar to that of a well-initialized CMA with the time-variant step-size mentioned above.

The paper is organized as follows. The problem is formulated in Section 2, and the proposed algorithm is introduced in Section 3. In Section 4, we present a deterministic stability analysis for the new algorithm, for scalar filters. Simulation results and conclusions are presented in sections 5 and 6, respectively. In order to simplify the presentation, we assume real data throughout the paper.

2. PROBLEM FORMULATION

A simplified communications system is depicted in Fig. 1. The signal $a(n)$, assumed independent, identically distributed, and non Gaussian, is transmitted through an unknown channel, whose model is constituted by an FIR (Finite Impulse Response) filter $H(z)$ and additive white Gaussian noise $\eta(n)$. From the received signal $u(n)$ and the known statistical properties of the transmitted signal, the blind equalizer must mitigate the channel effects and recover the signal $a(n)$ for some delay τ_d. The output of the equalizer is given by $y(n) = \mathbf{u}^T(n)\mathbf{w}(n-1)$, where $\mathbf{u}(n)$ is the input regressor vector, $\mathbf{w}(n-1)$ the equalizer weight vector (both column vectors with M coefficients), and the superscript T denotes the transpose of a vector.

The equations of CMA and SWA are given respectively by

$$\mathbf{w}(n) = \mathbf{w}(n-1) + \mu e(n)\mathbf{u}(n) \tag{1}$$

and

$$\mathbf{w}(n) = \mathbf{w}(n-1) + \frac{1}{3\sigma_a^2 - r} e(n)\widehat{\mathbf{R}}^{-1}(n)\mathbf{u}(n), \tag{2}$$

where $e(n) = y(n)\left[r - y^2(n)\right]$, $r = \mathrm{E}\{a^4(n)\}/\mathrm{E}\{a^2(n)\}$, $\sigma_a^2 = \mathrm{E}\{a^2(n)\}$, $\mathrm{E}\{\cdot\}$ denotes the expectation operation, and $(1-\lambda)\widehat{\mathbf{R}}(n)$ is an estimate (with forgetting factor λ) of the autocorrelation matrix

This work was supported in part by FAPESP under grant 04/15114-2, and CNPq under grant 303.361/2004-2.

of the input signal, i.e., $\mathbf{R} \triangleq \mathrm{E}\{\mathbf{u}(n)\mathbf{u}^T(n)\}$. Note that the constellation $a(n)$ in practice is sub-Gaussian, ensuring that the denominator $(3\sigma_a^2 - r)$ in (2) is always positive [3, 7].

It is well-known that SWA and CMA present a tradeoff between convergence rate and computational cost [3, 9]. This tradeoff tend to be less critical when "normalized" versions of CMA are compared to SWA. Different versions of normalized CMA have been proposed in the literature (see, e.g., [10, 6] and the references therein). To the best of our knowledge, they are based on variants of CMA other than (1) and do not present a mechanism to avoid divergence. Thus, the design of a stable blind equalization algorithm, along with a faster convergence for CMA, are problems of wide interest. With these goals, we propose a stable dual-mode algorithm based on a novel normalized CMA, as shown in the next section.

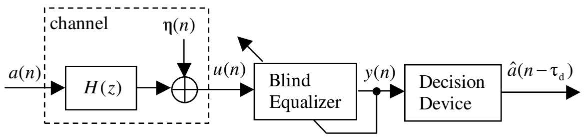

Fig. 1. Schematic representation of a communications system.

3. A DUAL-MODE ALGORITHM

SWA can be interpreted as a quasi-Newton algorithm, where $\widehat{\mathbf{R}}^{-1}(n)$ is an estimate of the Hessian of the constant-modulus cost function [9]. Thus, we will start the derivation of our normalized version of CMA from the following regularized Newton-type recursion, inspired in the recursion (2) for SWA,

$$\mathbf{w}(n) = \mathbf{w}(n-1) + \frac{\tilde{\mu}}{3\sigma_a^2 - r}[\delta\mathbf{I} + \mathbf{R}]^{-1}\mathrm{E}\{e(n)\mathbf{u}(n)\},$$

where $\tilde{\mu}$ is a step-size and δ is a small positive parameter. As in the derivation of the NLMS (Normalized Least-Mean-Square) algorithm [7, p. 225], the quantities $[\delta\mathbf{I} + \mathbf{R}]$ and $\mathrm{E}\{e(n)\mathbf{u}(n)\}$ should be replaced by instantaneous approximations, which leads to the stochastic recursion

$$\mathbf{w}(n) = \mathbf{w}(n-1) + \frac{\tilde{\mu}}{3\sigma_a^2 - r}[\delta\mathbf{I} + \mathbf{u}(n)\mathbf{u}^T(n)]^{-1}e(n)\mathbf{u}(n).$$

Then, using the matrix inversion lemma [7, p. 67], we get

$$\mathbf{w}(n) = \mathbf{w}(n-1) + \frac{\tilde{\mu}}{(3\sigma_a^2 - r)(\delta + \|\mathbf{u}(n)\|^2)}e(n)\mathbf{u}(n), \quad (3)$$

which can be interpreted as a normalized CMA (NCMA).

In order to derive a robust form for NCMA, we rewrite (3) in the form of a supervised algorithm, i.e.,

$$\mathbf{w}(n) = \mathbf{w}(n-1) + \frac{\tilde{\mu}}{\delta + \|\mathbf{u}(n)\|^2}\left[d(n) - y(n)\right]\mathbf{u}(n), \quad (4)$$

where $d(n) = x(n)y(n)$ and

$$x(n) = \frac{3\sigma_a^2 - y^2(n)}{3\sigma_a^2 - r}.$$

Note that (4) has the same structure as the NLMS algorithm. The difference is that $d(n)$, as well as $y(n)$, is an estimate of the desired response. We conjecture that the consistency between these two estimates will be ensured if $d(n)$ and $y(n)$ have the same sign, which is equivalent to requiring the correction factor $x(n)$ to be always positive. Since the denominator of $x(n)$ is always positive, $x(n) \geq 0$ occurs when $y^2(n) \leq 3\sigma_a^2$. On the other hand, if $y^2(n) > 3\sigma_a^2$, the algorithm leaves what we call the *region of interest* and the estimate $d(n)$ is simply rejected, i.e., we make $d(n) = 0$ and (4) reduces to

$$\mathbf{w}(n) = \mathbf{w}(n-1) - \frac{\tilde{\mu}}{\delta + \|\mathbf{u}(n)\|^2}y(n)\mathbf{u}(n). \quad (5)$$

To complete the derivation of the algorithm, the step-size $\tilde{\mu}$ should be chosen. If, at a certain iteration, $\|\mathbf{w}(n-1)\|$ is so large that $y^2(n) > 3\sigma_a^2$, we can guarantee that $\|\mathbf{w}(n)\| \leq \|\mathbf{w}(n-1)\|$, by choosing $\tilde{\mu}$ as follows. Expanding $y(n) = \mathbf{u}^T(n)\mathbf{w}(n-1)$ in (5), we obtain

$$\mathbf{w}(n) = \left[\mathbf{I} - \frac{\tilde{\mu}}{\delta + \|\mathbf{u}(n)\|^2}\mathbf{u}(n)\mathbf{u}^T(n)\right]\mathbf{w}(n-1), \quad (6)$$

and since $\mathbf{u}(n)\mathbf{u}^T(n)$ has one eigenvalue equal to $\|\mathbf{u}(n)\|^2$, and $M-1$ zero eigenvalues, the matrix between brackets has all eigenvalues with absolute values less than one if $\tilde{\mu} < 2$. The proposed dual-mode algorithm is summarized in Table 1 and denoted by DM-CMA. Using a recursion to update the successive squared-norms of the regressors [7, p. 227], each iteration of DM-CMA requires $2M + 6$ multiplications, $2M + 3$ additions, 1 division, and 1 comparison, which represents a computational cost slightly higher than that of CMA.

We should notice that $\tilde{\mu} < 2$ by itself does not imply in $y^2(n+1) \leq y^2(n)$, i.e, this condition does not guarantee that the algorithm returns to the region of interest, but only that the Euclidean norm of the coefficient vector does not increase with time. To prove stability, we need a persistence of excitation condition, as we show next for the scalar case.

Table 1. Summary of DM-CMA.

Initialization:
$\mathbf{w}(-1) = [0\ \cdots\ 0\ 1\ 0\ \cdots 0]^T$
$0 < \tilde{\mu} < 2$, δ: small positive constant
for $n = 0, 1, 2, 3, \ldots$ do:
$y(n) = \mathbf{u}^T(n)\mathbf{w}(n-1)$
$x(n) = \dfrac{3\sigma_a^2 - y^2(n)}{3\sigma_a^2 - r}$
if $x(n) \geq 0$,
$d(n) = x(n)y(n)$
else
$d(n) = 0$
end
$\bar{e}(n) = d(n) - y(n)$
$\mathbf{w}(n) = \mathbf{w}(n-1) + \dfrac{\tilde{\mu}}{\delta + \|\mathbf{u}(n)\|^2}\bar{e}(n)\mathbf{u}(n)$
end

4. A STABILITY ANALYSIS FOR THE SCALAR CASE

The difficulty of finding a stability proof arises from the fact that we want to ensure that $\mathbf{w}(n)$ is bounded, but, in order to maintain a good performance, we only switch to a different mode of operation when $y^2(n) = (\mathbf{u}^T(n)\mathbf{w}(n-1))^2$ is large. For a rigorous proof, one needs to make sure that there is no way for $\mathbf{w}(n)$ to diverge to infinity without $\mathbf{u}^T(n)\mathbf{w}(n-1)$ also diverging to infinity. In this section, we present a simple, but rigorous, stability proof for the scalar ($M = 1$) case, using a simplified persistence of excitation condition[1].

[1]A less restrictive and more usual persistence of excitation condition is $0 < b \leq \sum_{k=0}^{N} u^2(k) \leq B < \infty$ [11]. However, the use of this condition is considerably involved, even for the simpler case of the LMS algorithm, and we believe that the simpler condition is sufficient to describe the stability properties of the proposed algorithm.

In order to simplify the arguments, we assume that $u(n)$ is bounded from above and below, i.e., $0 < b \leq u^2(n) \leq B < \infty$. In this case, we also have $y^2(n) = u^2(n)w^2(n-1)$. We prove now that $w^2(n)$ always enters and stays inside a ball $w^2(n) \leq 3\sigma_a^2/b$. By contradiction, we assume that this condition is not true. It then follows that

$$y^2(n+1) = u^2(n+1)w^2(n) \geq b(3\sigma_a^2/b) = 3\sigma_a^2,$$

so $w(n+1)$ will be computed from (6), and

$$w^2(n+1) = \left[1 - \frac{\tilde{\mu}u^2(n+1)}{\delta + u^2(n+1)}\right]^2 w^2(n) \leq \alpha w^2(n),$$

where $\alpha < 1$ for $\tilde{\mu} < 2$ (e.g., $\alpha = [1-\tilde{\mu}b/(\delta+b)]^2 < 1$, if $\tilde{\mu} \leq 1$). Since $w^2(n+1)$ will be strictly smaller than $w^2(n)$, we conclude that $w^2(n)$ will decrease until at some time $n+k$, $w^2(n+k) \leq 3\sigma_a^2/b$.

It is relevant to notice that, when the algorithm is in the region of interest, it may converge to a good or a poor stationary point, that is, it presents the same problems of constant-modulus-based algorithms, except the divergence.

A similar analysis in the vector case ($M \geq 2$) should use persistence of excitation conditions [11], and will be presented in a further work. However, we tested the algorithm extensively for longer filters, and never observed divergence. Some of these simulations are presented in the next section.

5. SIMULATION RESULTS

The proposed DM-CMA is compared to NCMA and CMA, assuming 2-PAM (Pulse Amplitude Modulation) with symbols $\{\pm1\}$ or 4-PAM with symbols $\{\pm1; \pm3\}$, a channel $H(z) = h_0 + z^{-1} + h_0 z^{-2}$, and an equalizer with $M = 11$ coefficients in all simulations. If not mentioned otherwise, the equalizer is initialized with a the typical center spike [1] and the adaptation parameters μ and $\tilde{\mu}$ are adjusted to make the algorithms converge to the same steady-state EMSE.

Fig. 2 shows the squared error for each algorithm, the sign of $x(n)$, and the equalizer output $y(n)$ for $h_0 = 0.3$ and a signal-to-noise ratio (SNR) of 20 dB. To facilitate visualization, the squared error curve was filtered by a moving-average filter with 16 coefficients. CMA diverges after 200 iterations and NCMA after 6250 iterations. On the other hand, DM-CMA has a stable and adequate behavior. We can observe from the sign of $x(n)$ that there are points in which DM-CMA leaves the region of interest, but quickly returns to it. The figure shows that, little before NCMA became unstable, DM-CMA used (5) to update $\mathbf{w}(n)$, thereby avoiding divergence.

In the sequel, the algorithms are compared in terms of the probability of divergence P_d [12] and of the steady-state excess mean-square error (EMSE). The probability of divergence is obtained from L repetitions of each experiment, starting from the same initial condition $\mathbf{w}(0)$. As in [12], we assume that a sample function is labeled as "diverging", if $\|\mathbf{w}(N)\| \geq 10^4$ after N iterations. Then, we compute the probability of divergence as $P_d = (\text{Number of curves diverging})/L$. The EMSE, defined as $\mathrm{E}\{e_a^2(n)\} = \mathrm{E}\{[\hat{a}(n-\tau_d) - y(n)]^2\}$, is estimated through ensemble-averages corresponding to $(1-P_d)L$ independent runs for each algorithm. Thus, each EMSE value is associated to a P_d value.

Fig. 3 shows EMSE and P_d as a function of SNR for a channel with little intersymbol interference, i.e., $h_0 = 0.1$. For SNR $\geq$ 20 dB, the algorithms have similar performance, since $P_d = 0$ and the reached EMSE is the same. For SNR $\leq$ 20 dB, although DM-CMA and NCMA have also similar steady-state performance, the probability of divergence for the latter increases with increased noise,

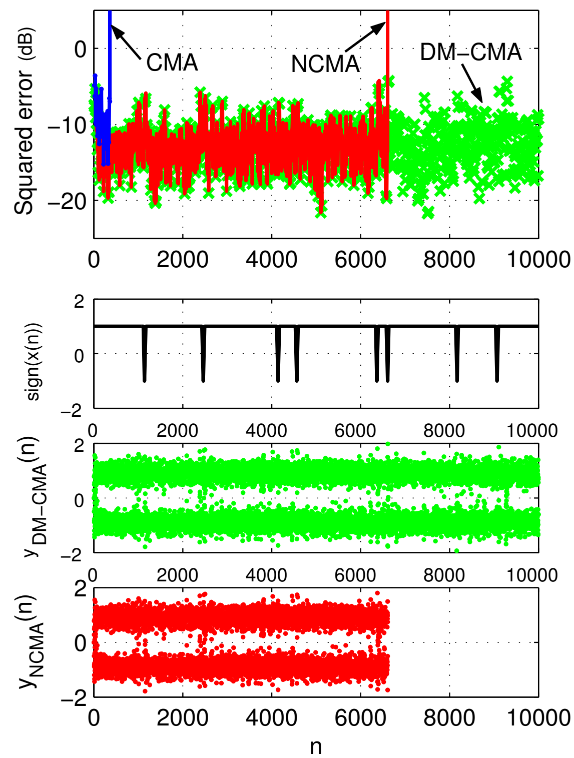

Fig. 2. Squared error in dB ($e^2(n)$ for CMA and $\bar{e}^2(n)$ for NCMA and DM-CMA), sign of $x(n)$ for DM-CMA and output of the equalizer for DM-CMA and NCMA; 2-PAM, $h_0 = 0.3$, SNR $= 20$ dB, $\mu = 0.044$, $\tilde{\mu} = 1$, $\delta = 10^{-5}$.

reaching $P_d \approx 1$ for SNR ≤ 2.5 dB, while DM-CMA does not diverge for the range of SNR considered. CMA presents a more unstable behavior than NCMA, diverging with $P_d = 1$ for SNR < 10 dB.

Fig. 4 shows EMSE and P_d as a function of the channel $H(z) = h_0 + z^{-1} + h_0 z^{-2}$ in the absence of noise. The larger the value of h_0, the greater is the intersymbol interference introduced by the channel. For $0.1 \leq h_0 \leq 0.3$, the algorithms behave in a similar manner, since divergence is not observed. However, for $0.5 \leq h_0 \leq 1$ CMA always diverges and NCMA has $P_d \approx 1$, converging only in a small number of times (e.g., 24 times in 10^3 runs for $h_0 = 0.7$). Again, DM-CMA does not diverge, presenting the same steady-state performance of the ensemble-average of $(1 - P_d)L$ runs of CMA and NCMA. Thus, besides avoiding divergence, (5) does not cause meaningful changes in the performance of the algorithm since a quick return to the region of interest was always observed.

The algorithms were also simulated with different initializations as shown in Fig. 5. They were initialized with a vector in the same direction as the optimal solution $\mathbf{w}_o$ but with different magnitudes, i.e., $\mathbf{w}(0) = p\mathbf{w}_o$, $0.2 \leq p \leq 3.8$. For $p \geq 1.8$, CMA and NCMA have $P_d \approx 1$. In this case, an initialization far from $\mathbf{w}_o$ seems to be more critical for NCMA, since we did not observe convergence for $p \geq 2.2$. For an initialization distant from the optimal solution but with a smaller magnitude (e.g., $p = 0.2$), although $P_d \approx 0$, the algorithms did not always converge to the optimal solution (EMSE $= 50$ dB), staying several times in local minima. For all the initial conditions considered, DM-CMA did not diverge. Note that CMA presents an EMSE value a little smaller than that of DM-CMA for $p = 3$ and $p = 3.4$. However, CMA converged only once ($P_d = 0.9990$), which does not indicate an acceptable performance of the algorithm, while DM-CMA converged in all the 10^3 runs.

Fig. 6 shows EMSE and P_d as a function of step-sizes $\tilde{\mu}$ for NCMA and DM-CMA, and μ (different scale) for CMA. We consider a non-constant modulus signal (4-PAM) and the channel $h_0 = 0.3$ in the absence of noise. The probability of divergence is close to 1 for $\mu \geq 8 \times 10^{-4}$ for CMA and $\tilde{\mu} > 0.4$ for NCMA. We can also observe that DM-CMA does not diverge for the considered range of step-sizes, i.e., $0.01 \leq \tilde{\mu} \leq 1.8$ and presents performance close to the stable runs of CMA and NCMA. Thus, DM-CMA avoids divergence and maintains an adequate behavior even for non-constant modulus signals.

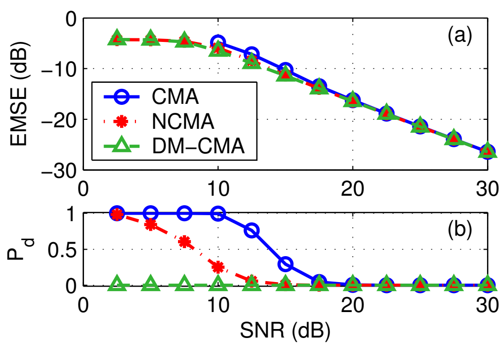

Fig. 3. (a) EMSE and (b) P_d as a function of SNR; 2-PAM, $h_0 = 0.1$, $N = 10^4$, $L = 10^3$, $\mu = 0.044$, $\tilde{\mu} = 1$, $\delta = 10^{-5}$.

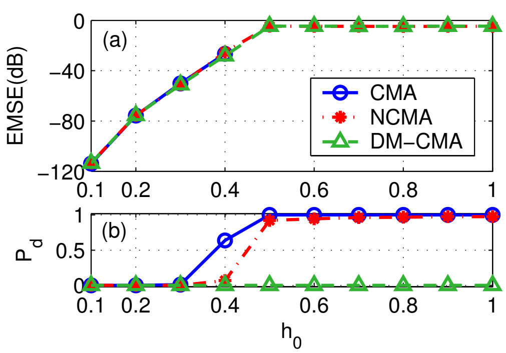

Fig. 4. (a) EMSE and (b) P_d as a function of channel $H(z) = h_0 + z^{-1} + h_0 z^{-2}$ in the absence of noise; 2-PAM, $N = 10^4$, $L = 10^3$, $\mu = 0.04$, $\tilde{\mu} = 1$, $\delta = 10^{-5}$.

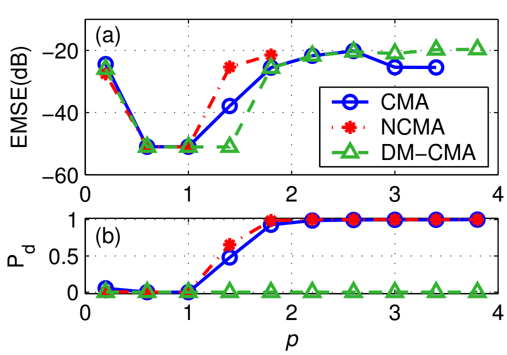

Fig. 5. (a) EMSE and (b) P_d as a function of the initialization $p\mathbf{w}_o$, where $\mathbf{w}_o^T$=[−0.005 0.016 −0.047 0.140 −0.417 1.250 −0.417 0.138 −0.046 0.016 −0.005]; $h_0 = 0.3$ in the absence of noise, 2-PAM, $N = 10^4$, $L = 10^3$, $\mu = 0.037$, $\tilde{\mu} = 1$, $M = 11$, $\delta = 10^{-5}$.

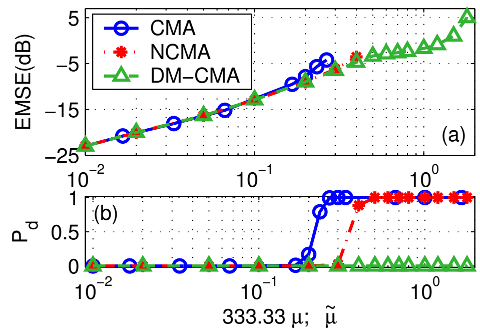

Fig. 6. (a) EMSE and (b) P_d as a function of the step-size μ for CMA and $\tilde{\mu}$ for NCMA and DM-CMA; $h_0 = 0.3$ in the absence of noise, 4-PAM, $N = 10^4$, $L = 10^3$, $\delta = 10^{-5}$.

6. CONCLUSION

In order to avoid divergence in CMA, we propose a dual-mode algorithm. In the region of interest, it works as a normalized CMA. Outside the region of interest, it rejects the estimate of the transmitted signal. Using a deterministic argument, we show for scalar filters that the proposed algorithm is stable. We performed a large number of simulations also for filters with several taps, and never observed divergence of the new algorithm. In a future work we intend to extend the stability analysis, using a less restrictive persistence of excitation condition, to the vector case.

7. REFERENCES

[1] C. R. Johnson Jr. *et al.*, “Blind equalization using the constant modulus criterion: a review,” *Proc. IEEE*, vol. 86, pp. 1927–1950, Oct. 1998.

[2] D. N. Godard, “Self-recovering equalization and carrier tracking in two dimensional data communication system,” *IEEE Trans. Commun.*, vol. 28, pp. 1867–1875, Nov. 1980.

[3] O. Shalvi and E. Weinstein, “Super-exponential methods for blind deconvolution,” *IEEE Trans. Inf. Theory*, vol. 39, pp. 504–519, Mar. 1993.

[4] P. A. Regalia, “On the equivalence between the Godard and Shalvi-Weinstein schemes of blind equalization,” *Signal Process.*, vol. 73, pp. 185–190, 1999.

[5] M. Rupp and A. H. Sayed, “On the convergence of blind adaptive equalizers for constant modulus signals,” *IEEE Trans. Commun.*, vol. 48, pp. 795–803, May 2000.

[6] J. 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, pp. 80–91, Jan. 2000.

[7] A. H. Sayed, *Fundamentals of Adaptive Filtering*, John Wiley & Sons, NJ, 2003.

[8] D. B. Bernardes, M. D. Miranda, and M. T. M. Silva, “A Lattice Shalvi-Weinstein Algorithm for blind equalization,” in *Proc. of ICASSP’07*. IEEE, 2007, vol. III, pp. 1369–1372.

[9] M. T. M. Silva and M. D. Miranda, “Tracking issues of some blind equalization algorithms,” *IEEE Signal Process. Lett.*, vol. 11, pp. 760–763, Sep. 2004.

[10] C. B. Papadias and D. T. M. Slock, “Normalized sliding window constant modulus and decision-direct algorithms,” *IEEE Trans. Signal Process.*, vol. 45, pp. 231–235, Jan. 1997.

[11] P. A. Ioannou and J. Sun, *Robust Adaptive Control*, PTR Prentice Hall, NJ, 1996.

[12] P. I. Hübscher, J. C. M. Bermudez, and V. H. Nascimento, “A mean-square stability analysis of the least mean fourth adaptive algorithm,” *IEEE Trans. Signal Process.*, vol. 55, pp. 4018–4028, Aug. 2007.