A Unified Approach for Sparsity-Aware and Maximum Correntropy Adaptive Filters

Diego Barreto Haddad
Federal Center for Technological Education, CEFET-RJ
26041-271, Nova Iguaçu
Rio de Janeiro, Brazil
Email: diego@pads.ufrj.br

Mariane Rembold Petraglia and Antonio Petraglia
Federal University of Rio de Janeiro
PEE/COPPE
CP 68504, 21945-970
Rio de Janeiro, Brazil
Email: mariane@pads.ufrj.br, petra@pads.ufrj.br

***Abstract*—Adaptive filters that employ sparse constraints or maximum correntropy criterion (MCC) have been derived from stochastic gradient techniques. This paper provides a deterministic optimization framework which unifies the derivation of such algorithms. The proposed framework has also the ability of providing geometric insights about the adaptive filter updating. New algorithms that exploit both impulse responses sparsity and MCC are proposed, and an estimate of their steady-state MSE is advanced. Simulations show the advantages of the proposed algorithms in the identification of a sparse system with non-Gaussian additive noise.**

***Index Terms*—Adaptive filtering, maximum correntropy criterium, sparse impulse response, mean-square analysis.**

I. Introduction

Adaptive filters have proved useful in several challenging and crucial tasks, such as system identification, echo cancellation and channel equalization [1]. In some sense, much of the recent developments in the area of adaptive filtering lies in the exploitation of some inherent structural properties of the problem, as the sparsity of the impulse response to be identified [2]. Such sparsity occurs, for example, in digital TV transmission channels [3], in echo paths [2], and in underwater acoustic communication channels [4]. Another approach explores nonlinear or non-Gaussian scenarios which revealed the relevance of the maximum correntropy criterion [5]. It is noteworthy that correntropy is a generalized correlation measure that includes information about the distribution and the time structure of a stochastic process [6].

In the adaptive filtering context we are interested in the adaptive vector $\boldsymbol{w}(k) \in \mathbb{R}^N$, which should emulate an ideal and unknown vector $\boldsymbol{w}_o \in \mathbb{R}^N$, by reducing the discrepancy between the output of the filter $y(k) = \boldsymbol{w}^T(k)\boldsymbol{x}(k)$ and the desired (or reference) signal $d(k)$, which is given by

$$d(k) = \boldsymbol{w}_o^T \boldsymbol{x}(k) + \nu(k), \tag{1}$$

where $\boldsymbol{x}(k) \in \mathbb{R}^N$ is the input signal and the perturbation signal $\nu(k)$ accounts for measurement noise and/or modeling errors. In general, the error $e(k) = d(k) - y(k)$ is employed as the discrepancy measure between the adaptive filter output and the reference signal. In practice, $\boldsymbol{w}_o$ may be the acoustic transfer function of a room, and should be identified in order to achieve echo cancellation. Algorithms based on the MCC can undertake non-Gaussian noise $\nu(k)$ in a robust approach, whereas sparsity-aware algorithms are able to take advantage of the concentration of energy of $\boldsymbol{w}_o$ in few coefficients, thereby speeding up the convergence rate of the adaptive filter.

The classic derivation of the least-mean-square (LMS) algorithm is based on the stochastic gradient technique, which is frequently used with the purpose of producing new adaptive non-normalized sparsity-aware or MCC algorithms (e.g., [7], [8]). In their turn, optimization techniques for solving deterministic problems, although also suitable for obtaining non-normalized schemes [9], are usually employed in the design of normalized adaptive filters (e.g., [10]).

This paper provides a deterministic optimization framework, whose solution leads naturally to MCC or sparsity-aware algorithms, such as the MCC [8] and the ℓ_0-LMS algorithm [7], in both normalized and non-normalized versions. Generalized versions of these algorithms are also derived. New algorithms that combine sparseness and MCC are proposed, and an estimate of their steady-state MSE is developed. Some geometric interpretations of the updating process of the proposed algorithm are presented.

II. MCC Algorithms

Under the MCC, the updating equation [8] is given by

$$\boldsymbol{w}(k+1) = \boldsymbol{w}(k) + \beta \exp\left(-\frac{e^2(k)}{2\sigma^2}\right) e(k)\boldsymbol{x}(k), \tag{2}$$

where β is the step-size and σ is a positive parameter that induces a trade-off between convergence rate and steady-state MSE [11].

The normalized MCC (NMCC) updating equation can be written as [11]:

$$\boldsymbol{w}(k+1) = \boldsymbol{w}(k) + \beta \frac{\exp\left(-\frac{e^2(k)}{2\sigma^2}\right) e(k)\boldsymbol{x}(k)}{\|\boldsymbol{x}(k)\|^2}. \tag{3}$$

Through the demonstration of Theorem 1 below, it is shown how MCC and NMCC updates could be derived from the *exact*

978-0-9928-6265-7/16/$31.00 ©2016 IEEE

solution of a local (as opposed to a global) deterministic optimization problem (see [9] for a discussion on such approach for algorithms such as LMS or NLMS).

Theorem 1. The MCC updating equations (2) and (3) are the solutions of the following problem:

$$\min \frac{1}{2}\|\boldsymbol{w}(k+1)-\boldsymbol{w}(k)\|^2, \tag{4}$$

$$\text{s.t. } \tilde{e}(k)=\left[1-\gamma \exp\left(-\frac{e^2(k)}{2\sigma^2}\right)\right]e(k),$$

where $\tilde{e}(k) \triangleq d(k)-\boldsymbol{w}^T(k+1)\boldsymbol{x}(k)$, $\|\cdot\|^2$ denotes Euclidean vector norm, $\gamma=\beta\|\boldsymbol{x}(k)\|^2$ for obtaining (2) and $\gamma=\beta$ in order to get (3).

Proof. The constrained problem (4) can be transformed into an unconstrained one by using the Lagrange multiplier technique, thereby yielding the cost function

$$\mathcal{F}_{\mathrm{M}}\left[\boldsymbol{w}(k+1)\right]=\frac{1}{2}\|\boldsymbol{w}(k+1)-\boldsymbol{w}(k)\|^2$$
$$+\lambda\left(\tilde{e}(k)-\left[1-\gamma \exp\left(-\frac{e^2(k)}{2\sigma^2}\right)\right]e(k)\right), \tag{5}$$

where λ is the Lagrange multiplier. The minimization of $\mathcal{F}_{\mathrm{M}}\left[\boldsymbol{w}(k+1)\right]$ can be carried out by zeroing its gradient w.r.t. $\boldsymbol{w}(k+1)$, from which results

$$\boldsymbol{w}(k+1)=\boldsymbol{w}(k)+\lambda\boldsymbol{x}(k). \tag{6}$$

Applying (6) in the constraint (4) gives

$$\lambda=\gamma\frac{\exp\left(-\frac{e^2(k)}{2\sigma^2}\right)e(k)}{\|\boldsymbol{x}(k)\|^2}. \tag{7}$$

Using $\gamma=\beta\|\boldsymbol{x}(k)\|^2$ in (7) and inserting the result in (6) leads to (2), whereas using $\gamma=\beta$ yields (3). □

Note that the exponential term in (4) eliminates large errors, which turns the algorithm robust against impulsive-like noises.

III. Sparsity-Aware Adaptive Algorithms

The framework developed for the optimization problem (4) can be readily applied to sparsity-aware adaptive filters. Accordingly, let us consider the problem

$$\min \frac{1}{2}\|\boldsymbol{w}(k+1)-\boldsymbol{w}(k)\|^2+\alpha\|\boldsymbol{w}(k+1)\|_0, \tag{8}$$

$$\text{s.t. } \tilde{e}(k)=(1-\gamma)\,e(k),$$

where $\|\cdot\|_0$ denotes the ℓ_0 norm (or, in practice, ℓ_0 *pseudo*-norm) and α is a positive parameter that penalizes solutions that present low sparsity.

The direct solution of (8) is not an adequate strategy, since it leads to a computationally expensive NP-hard problem. An alternative approach is the use of *almost everywhere* differentiable approximation $F_\rho\left[\boldsymbol{w}(k+1)\right]$ of the ℓ_0 norm[1] [7], [12], so that (8) can be *approximately* solved by the following unconstrained cost function, whose updating equation is similar to the ℓ_0-LMS (or ℓ_0-NLMS) algorithms [7]:

$$\mathcal{F}_{\mathrm{S}}\left[\boldsymbol{w}(k+1)\right]=\frac{1}{2}\|\boldsymbol{w}(k+1)-\boldsymbol{w}(k)\|^2+\alpha F_\rho\left[\boldsymbol{w}(k+1)\right]$$
$$+\lambda\left[\tilde{e}(k)-(1-\gamma)\,e(k)\right]. \tag{9}$$

Following the steps of the demonstration of Theorem 1, one can derive the sparsity-aware adaptive update

$$\boldsymbol{w}(k+1)=\boldsymbol{w}(k)+\gamma\overbrace{\frac{e(k)\boldsymbol{x}(k)}{\|\boldsymbol{x}(k)\|^2}}^{\triangleq \boldsymbol{t}_1}-\alpha\boldsymbol{f}_\rho\left[\boldsymbol{w}(k)\right] \tag{10}$$
$$+\alpha\overbrace{\frac{\boldsymbol{x}^T(k)\boldsymbol{f}_\rho\left[\boldsymbol{w}(k)\right]\boldsymbol{x}(k)}{\|\boldsymbol{x}(k)\|^2}}^{\triangleq \boldsymbol{t}_2},$$

where $\frac{\nabla F_\rho[\boldsymbol{w}(k+1)]}{\nabla \boldsymbol{w}(k+1)} \triangleq \boldsymbol{f}_\rho\left[\boldsymbol{w}(k+1)\right]$ and $\boldsymbol{f}_\rho\left[\boldsymbol{w}(k+1)\right]$ was replaced by $\boldsymbol{f}_\rho\left[\boldsymbol{w}(k)\right]$ to produce a valid recursion [12]. Note that $\boldsymbol{t}_1$ and $\boldsymbol{t}_2$ have the same direction, and hence $\boldsymbol{t}_2$ can be neglected. Indeed, in our simulations, the insertion of $\boldsymbol{t}_2$ makes virtually no difference in the transient and steady-state performances of the ℓ_0-LMS (or ℓ_0-NLMS) algorithms.

Equation (10) establishes a family of sparsity-aware algorithms. If, among the several feasible choices for $F_\rho\left[\boldsymbol{w}(k)\right]$, we select

$$F_\rho\left[\boldsymbol{w}(k)\right] \triangleq \sum_{n=0}^{N-1}\left[1-\exp\left(-\rho|w_i(k)|\right)\right], \tag{11}$$

adopt $\gamma=\beta\|\boldsymbol{x}(k)\|^2$ and discard $\boldsymbol{t}_2$ we arrive, after some simplifications, at the ℓ_0-LMS. For the same choice of $F_\rho\left[\boldsymbol{w}(k)\right]$, adopting $\gamma=\beta$ and discarding $\boldsymbol{t}_2$ produces the ℓ_0-NLMS. Note that the derivation of ℓ_0-NLMS in [7] was accomplished by the heuristic insertion of the denominator $\|\boldsymbol{x}(k)\|^2$. In the approach proposed here, on the other hand, ℓ_0-NLMS is obtained straightforwardly from the deterministic optimization problem.

IV. MCC-based Sparsity-Aware Algorithms

So far, we have considered algorithms that employ the MCC or the sparsity of the impulse response to be identified. The advantages of both approaches can be exploited simultaneously by using updates that take into account both properties. The framework proposed in this paper is suitable for performing this task. Applying steps similar to the ones of the previous sections, we obtain from the cost function

$$\mathcal{F}_{\mathrm{MS}}\left[\boldsymbol{w}(k+1)\right]=\frac{1}{2}\|\boldsymbol{w}(k+1)-\boldsymbol{w}(k)\|^2+\alpha F_\rho\left[\boldsymbol{w}(k+1)\right]$$
$$+\lambda\left(\tilde{e}(k)-\left[1-\gamma \exp\left(-\frac{e^2(k)}{2\sigma^2}\right)\right]e(k)\right) \tag{12}$$

the proposed MCC-based sparsity-aware update

$$\boldsymbol{w}(k+1)=\boldsymbol{w}(k)+\gamma\overbrace{\frac{\exp\left(-\frac{e^2(k)}{2\sigma^2}\right)e(k)\boldsymbol{x}(k)}{\|\boldsymbol{x}(k)\|^2}}^{\triangleq \boldsymbol{p}_1}$$

[1]Frequently, the function $F_\rho[\cdot]$ is dependent on a parameter $\rho \in \mathbb{R}_+$, which can be arbitrated by the designer.

$$\overbrace{-\alpha \boldsymbol{f}_{\rho}\left[\boldsymbol{w}(k)\right]}^{\triangleq \boldsymbol{p}_2} + \overbrace{\alpha \frac{\boldsymbol{x}^T(k)\boldsymbol{f}_{\rho}\left[\boldsymbol{w}(k)\right]\boldsymbol{x}(k)}{\|\boldsymbol{x}(k)\|^2}}^{\triangleq \boldsymbol{p}_3 = \boldsymbol{t}_2}. \quad (13)$$

Discarding $\boldsymbol{p}_3$ and choosing $\gamma = \beta\|\boldsymbol{x}(k)\|^2$ yields a solution that can be derived from a stochastic gradient technique whose step-size is β, and applied on the following cost function:

$$\mathcal{F}_G\left[\boldsymbol{w}(k)\right] = \mathbb{E}\left[\exp\left(-\frac{e^2(k)}{2\sigma^2}\right)\right] + \frac{\alpha}{\beta} F_{\rho}\left[\boldsymbol{w}(k)\right]. \quad (14)$$

It should be observed that (13) represents a *family* of MCC sparsity-aware adaptive algorithms, of which several previously proposed algorithms are particular cases. If $\alpha \neq 0$ and $\gamma = \beta\|\boldsymbol{x}(k)\|^2$, we have a new algorithm, which will henceforth be called ℓ_0-MCC. If $\alpha \neq 0$ and $\gamma = \beta$, we have a normalized version of the ℓ_0-MCC algorithm, which we call ℓ_0-NMCC.

A. Geometric Interpretation

The deterministic optimization problem in (12) could provide a geometric interpretation of the updating equation (13) [12]. In Fig. 1, which shows the 2-D case, we observe that $\boldsymbol{p}_1$ approximates $\boldsymbol{w}(k)$ to the hyperplane Π defined $\tilde{e}(k) = 0$. The directions of both $\boldsymbol{p}_1$ and $\boldsymbol{p}_3$ are orthogonal to the hyperplane. The term $\boldsymbol{p}_2$ is a zero-attracting one, i.e., it forces near-zero coefficients of $\boldsymbol{w}(k)$ in the direction of zero. Note that in the case of large $e(k)$, the exponential term $\exp\left(-\frac{e^2(k)}{2\sigma^2}\right)$ approximates the hyperplane Π to $\boldsymbol{w}(k+1)$, which reduces the update magnitude, providing robustness to the algorithm against impulsive noise.

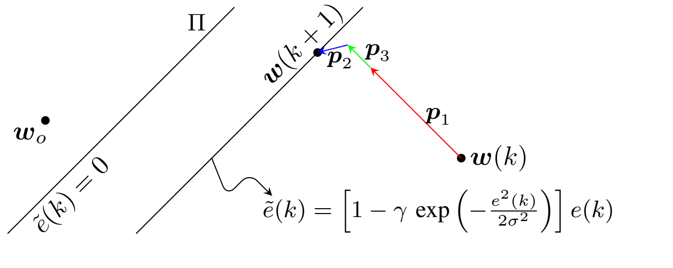

Fig. 1. Geometric interpretation of the general proposed algorithm (13) updating process.

V. Energy Conservation Relationship

In this section, we derive the energy conservation relationship [13], [14] for the proposed family of algorithms (13). Let

$$\overline{f}\left[e(k)\right] \triangleq \exp\left(-\frac{e^2(k)}{2\sigma^2}\right) e(k) \quad (15)$$

and $\tilde{\boldsymbol{w}}(k) \triangleq \boldsymbol{w}_o - \boldsymbol{w}(k)$, so that the updating equation (13) can be rewritten as:

$$\begin{aligned}\tilde{\boldsymbol{w}}(k+1) = \tilde{\boldsymbol{w}}(k) - \gamma \frac{\overline{f}[e(k)]\boldsymbol{x}(k)}{\|\boldsymbol{x}(k)\|^2} + \alpha \boldsymbol{f}_{\rho}\left[\boldsymbol{w}(k)\right] \\ -\alpha \frac{\boldsymbol{x}^T(k)\boldsymbol{f}_{\rho}\left[\boldsymbol{w}(k)\right]\boldsymbol{x}(k)}{\|\boldsymbol{x}(k)\|^2}. \quad (16)\end{aligned}$$

By left-multiplying both sides of (16) by $\boldsymbol{x}^T(k)$, we find

$$e_p(k) = e_a(k) - \gamma \overline{f}\left[e(k)\right], \quad (17)$$

where $e_p(k) \triangleq \boldsymbol{x}^T(k)\tilde{\boldsymbol{w}}(k+1)$ and $e_a(k) \triangleq \boldsymbol{x}^T(k)\tilde{\boldsymbol{w}}(k)$. Note that, although a common result, (17) does not hold true for the ℓ_0-LMS algorithm (whose updating equation does not contain $\boldsymbol{t}_2$ of (13)).

From (16) and (17) we obtain

$$\begin{aligned}\tilde{\boldsymbol{w}}(k+1) + \frac{e_a(k)\boldsymbol{x}(k)}{\|\boldsymbol{x}(k)\|^2} + \alpha \frac{\boldsymbol{x}^T(k)\boldsymbol{f}_{\rho}\left[\boldsymbol{w}(k)\right]}{\|\boldsymbol{x}(k)\|^2} \\ = \tilde{\boldsymbol{w}}(k) + \frac{e_p(k)\boldsymbol{x}(k)}{\|\boldsymbol{x}(k)\|^2} + \alpha \boldsymbol{f}_{\rho}\left[\boldsymbol{w}(k)\right], \quad (18)\end{aligned}$$

The evaluation of the energy on both sides of (18) allows us to write the following *exact* energy relation:

$$\begin{aligned}\|\tilde{\boldsymbol{w}}(k+1)\|^2 + \frac{e_a^2(k)}{\|\boldsymbol{x}(k)\|} + 2\alpha \frac{e_a(k)\boldsymbol{x}^T(k)\boldsymbol{f}_{\rho}\left[\boldsymbol{w}(k)\right]}{\|\boldsymbol{x}(k)\|^2} \\ +\alpha^2 \frac{\left[\boldsymbol{x}^T(k)\boldsymbol{f}_{\rho}\left[\boldsymbol{w}(k)\right]\right]^2}{\|\boldsymbol{x}(k)\|^2} = \|\tilde{\boldsymbol{w}}(k)\|^2 + \frac{e_p^2(k)}{\|\boldsymbol{x}(k)\|^2} \quad (19) \\ +2\alpha \tilde{\boldsymbol{w}}^T(k)\boldsymbol{f}_{\rho}\left[\boldsymbol{w}(k)\right] + \alpha^2 \|\boldsymbol{f}_{\rho}\left[\boldsymbol{w}(k)\right]\|^2,\end{aligned}$$

which is a general result that can be used for obtaining steady-state MSE estimates of several algorithms. In the next section, we focus on the NMCC[2].

VI. NMCC Steady-State MSE

In order to demonstrate the powerfulness of the general energy conservation (19), we consider the NMCC algorithm, which corresponds to the choices $\alpha = 0$ and $\gamma = \beta$. Here, our objective is to derive the steady-state MSE of this algorithm. The following assumptions, which are frequently employed in the literature [13], [15], [16], are in order:
A1: the adaptive filter converges.
A2: the noise $\nu(k)$, whose variance is σ_ν^2, is i.i.d., zero-mean, Gaussian and independent of $\boldsymbol{x}(k)$.
A3: in steady-state, $\|\boldsymbol{x}(k)\|^2$ is independent of $e_a(k)$ and $e(k)$.
Henceforth, we assume steady-state regime. Using *A1*, we can state that

$$\lim_{k\to\infty} \mathbb{E}\left[\|\tilde{\boldsymbol{w}}(k+1)\|^2\right] = \mathbb{E}\left[\|\tilde{\boldsymbol{w}}(k)\|^2\right]. \quad (20)$$

Applying the expectation operator in (19) and using (20) with assumptions *A2*-*A3*, we obtain

$$\mathbb{E}\left[\frac{e_a^2(k)}{\|\boldsymbol{x}(k)\|^2}\right] = \mathbb{E}\left[\frac{e_p^2(k)}{\|\boldsymbol{x}(k)\|^2}\right]. \quad (21)$$

From (17) and assumptions *A2*-*A3*, (21) can be rewritten as

$$2\mathbb{E}\left[e_a(k)\overline{f}[e(k)]\right] = \beta \mathbb{E}\left[\overline{f}^2[e(k)]\right]. \quad (22)$$

Using *A2*, we obtain [17]

$$\mathbb{E}\left[e_a(k)\overline{f}\left[e(k)\right]\right] = \frac{\sigma^3\Psi}{(\sigma^2 + \sigma_\nu^2 + \Psi)^{\frac{3}{2}}}, \quad (23)$$

[2]The analysis of algorithms that employ ℓ_0 penalizations is out of the scope of this paper.

$$\mathbb{E}\left[\overline{f}^2\left[e(k)\right]\right] = \frac{\sigma^3\left(\Psi+\sigma_\nu^2\right)}{\left(2\Psi+2\sigma_\nu^2+\sigma^2\right)^{\frac{3}{2}}}, \qquad (24)$$

where $\Psi \triangleq \lim_{k\to\infty}\mathbb{E}\left[e_a^2(k)\right]$ is the excess MSE (EMSE) of the algorithm [13]. From (22), (23) and (24), we can estimate Ψ by solving the fixed-point equation

$$\Psi = \beta\frac{\left(\Psi+\sigma_\nu^2\right)\left(\sigma^2+\sigma_\nu^2+\Psi\right)^{3/2}}{2\left(2\Psi+2\sigma_\nu^2+\sigma^2\right)^{3/2}}. \qquad (25)$$

Whereas in theory (25) could have more than one solution, our simulations showed only the desired one.

Equation (25) can be expressed as

$$\Psi = \frac{\beta\left(\Psi+\sigma_\nu^2\right)}{2}\overbrace{\left(\frac{\sigma^2+\sigma_\nu^2+\Psi}{2\Psi+2\sigma_\nu^2+\sigma^2}\right)^{\frac{3}{2}}}^{\triangleq\Phi}, \qquad (26)$$

from which the EMSE of the NMCC algorithm becomes

$$\Psi = \frac{\beta\sigma_\nu^2\Phi}{2-\beta\Phi}. \qquad (27)$$

Since $\Phi < 1$, we conclude that the EMSE of the NMCC algorithm is always lower than the EMSE of the NLMS algorithm, which is given by [13]:

$$\Psi_{\text{NLMS}} = \frac{\beta\sigma_\nu^2}{2-\beta}. \qquad (28)$$

In the limit $\sigma\to\infty$ the EMSE of both algorithms (NMCC and NLMS) coincide[3].

VII. Simulations

A. Performance Evaluation

1) Artificial Signals: In this experiment, we evaluate the mean-square deviation (MSD), given by

$$\text{MSD}(k) \triangleq \|\boldsymbol{w}(k)-\boldsymbol{w}_o\|^2, \qquad (29)$$

of the non-normalized and normalized versions of the LMS, MCC, ℓ_0-LMS and ℓ_0-MCC algorithms. Similarly to [7], the unknown system comprises 128 coefficients, of which eight (whose locations are randomly selected) are sampled from a zero-mean and unitary-variance Gaussian variable. The input signal is a zero-mean white Gaussian noise with variance $\sigma_x^2 = 4$. The noise $\nu(k)$ is sampled from a zero-mean uniform distribution with variance $\sigma_\nu^2 = \frac{4}{3}$, to simulate a non-Gaussian noise with large perturbations (a challenging scenario). For a fair comparison, equal parameter values were used in all algorithms. In ℓ_0-based algorithms, $F_\rho\left[\boldsymbol{w}(k)\right]$ is given by (11) with $\rho = 5$ and $\alpha = 3\times10^{-5}$, whereas in MCC-based algorithms $\sigma = 2$. We employed $\beta = 1.5\times10^{-3}$ and $\beta = 0.6$, respectively, for non-normalized (Fig. 2) and normalized (Fig. 3) algorithms. As indicated in Figs. 2 and 3, the proposed ℓ_0-MCC and ℓ_0-NMCC produced the lowest steady-state MSD among all tested algorithms, with a faster convergence rate than those of their MCC and NMCC counterparts. It can be noted, however, that ℓ_0-LMS and ℓ_0-NLMS converge faster than the proposed algorithms, so that their employment poses a classical trade-off between convergence rate and steady-state performance.

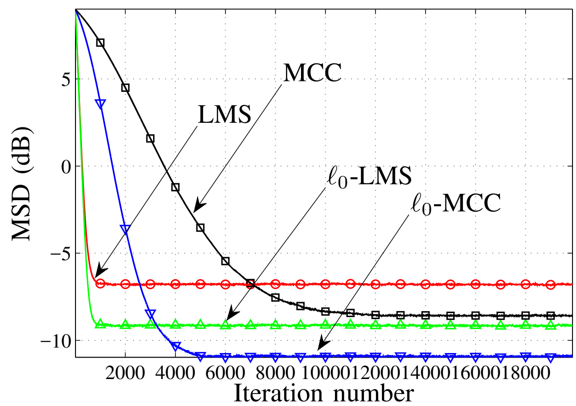

Fig. 2. MSD (in dB) of non-normalized algorithms ($\gamma = \beta\|\boldsymbol{x}(k)\|^2$), with 1000 Monte Carlo averages.

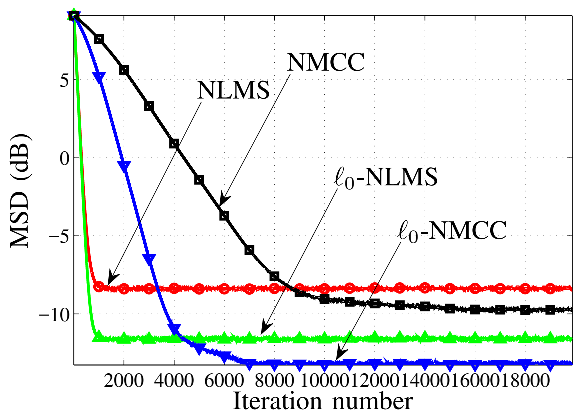

Fig. 3. MSD (in dB) of normalized algorithms ($\gamma = \beta$), with 1000 Monte Carlo averages.

2) Speech Input: Frequently (e.g., in echo cancellation applications) the adaptive filter is driven by speech signal. In this simulation, the identification performances of normalized algorithms (i.e., NLMS, NMCC, ℓ_0-NLMS and ℓ_0-NMCC) are compared. The input signal is a speech signal (extracted from the file `03a01Wa.wav` of the `Berlin Database of Emotional Speech`[4]). The employed transfer function is the first model of [18] (with zero-padding up to a 512 length). The noise $n(k)$ is sampled according to a Gaussian mixture model, whose probability density function (pdf) is:

$$f_\nu(\nu) = p_i\mathcal{N}(0,\sigma_1^2) + (1-p_i)\mathcal{N}(0,\sigma_2^2), \qquad (30)$$

where $\mathcal{N}(\mu,\sigma^2)$ is the Gaussian pdf whose mean is μ and variance is σ^2, $p_i = 0.98$, $\sigma_1^2 = 10^{-6}$ and $\sigma_2^2 = 1$. In order that the algorithms behave properly, a regularization constant $\delta = 0.1$ was added to the denominator $\|\boldsymbol{x}(k)\|^2$ of the normalized

[3] See [17] for similar conclusions related to the MCC and LMS algorithms.

[4] `http://www.expressive-speech.net`

step-sizes. In ℓ_0-based algorithms, $F_\rho[\boldsymbol{w}(k)]$ is given by (11) with $\rho = 1000$ and $\alpha = 5 \times 10^{-9}$, whereas in MCC-based algorithms $\sigma = 0.01$. In all algorithms, the choice $\beta = 0.1$ is used. The results are presented in Fig. 4. One can note that only correntropy-based adaptive algorithms converge, and the proposed ℓ_0-NMCC performs better than the NMCC.

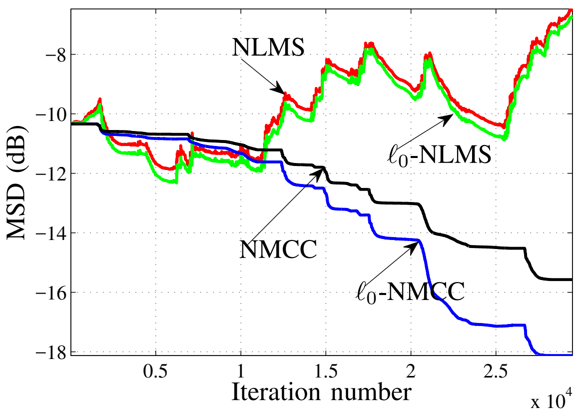

Fig. 4. MSD (in dB) of normalized algorithms ($\gamma = \beta$), with speech signal at the input of the adaptive filter.

B. Steady-State MSE

In this section, we verify the accuracy of (25) in estimating the steady-state MSE of the NMCC algorithm. The impulse response to be identified is the 8th model of [18]. A white Gaussian noise having variance $\sigma_\nu^2 = 10^{-3}$ was added to the reference signal. The input signal is a unitary-variance white Gaussian noise. The NMCC algorithm employed $\sigma = 2$. The experimental MSE was obtained by averaging 50000 consecutive steady-state quadratic errors. Figure 5 shows that the experimental and theoretical results are in close agreement for a large range of values of β.

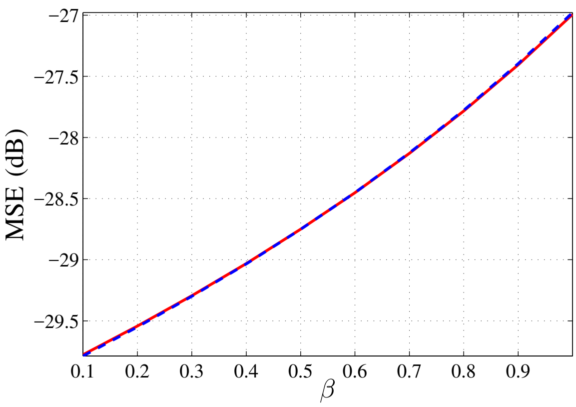

Fig. 5. Theoretical (red) and experimental (blue) steady-state MSE (in dB) of the NMCC algorithm.

VIII. Conclusions

A deterministic framework derivation for sparsity-aware and MCC-based algorithms was provided. Pre-existing algorithms were obtained in a novel way and new ones were proposed. A geometric interpretation of their updating process was provided, as well as a steady-state MSE expression for the NMCC algorithm, derived from a general energy conservation relationship. Computer simulations showed that the algorithms advanced in this paper present lower steady-state MSD than previously proposed algorithms when modeling sparse impulse responses in non-Gaussian high-noise environments.

Acknowledgment

The authors would like to thank CNPq and FAPERJ, Brazil, for providing partial support.

References

[1] S. O. Haykin, *Adaptive Filter Theory*, Prentice Hall, Englewood Cliffs, NJ, 4 edition, 2001.
[2] D. L. Duttweiler, "Proportionate normalized least-mean-squares adaptation in echo cancelers," *IEEE Trans. on Speech and Audio Processing*, vol. 8, no. 5, pp. 508–518, Sep. 2000.
[3] William F Schreiber, "Advanced television systems for terrestrial broadcasting: Some problems and some proposed solutions," *Proceedings of the IEEE*, vol. 83, no. 6, pp. 958–981, 1995.
[4] Milica Stojanovic and James Preisig, "Underwater acoustic communication channels: Propagation models and statistical characterization," *Communications Magazine, IEEE*, vol. 47, no. 1, pp. 84–89, 2009.
[5] W. Liu, P. P. Pokharel, and J. C. Príncipe, "Correntropy: Properties, and applications in non-Gaussian signal processing," *IEEE Trans. Signal Process.*, vol. 55, no. 11, pp. 5286–5298, Nov. 2007.
[6] I. Santamaría, P. P. Pokharel, and J. C. Principe, "Generalized correlation function: Definition, properties, and application to blind equalization," *IEEE Trans. on Signal Processing*, vol. 54, no. 6, pp. 2187–2197, Jun. 2006.
[7] Gu. Y, J. Jin, and S. Mei, "l_0 norm constraint LMS algorithm for sparse system identification," *IEEE Signal Processing Letters*, vol. 16, no. 9, pp. 774–777, Sep. 2009.
[8] A. Singh and J. C. Príncipe, "Using correntropy as a cost function in linear adaptive filters," *Proc. Int. Joint Conf. Neural Networks*, pp. 2950–2955, Jun. 2009.
[9] A. H. Sayed, *Adaptive Filters*, John Wiley & Sons, New Jersey, 1 edition, 2008.
[10] K. A. Lee and W. S. Gan, "Improving convergence of the NLMS algorithm using constrained subband updates," *IEEE Signal Process. Lett.*, vol. 11, no. 9, pp. 736–739, Sep. 2004.
[11] W. Liu, J. C. Príncipe, and S. Haykin, *Kernel Adaptive Filtering*, John Wiley & Sons, New Jersey, 1 edition, 2010.
[12] M. V. S. Lima, T. N. Ferreira, W. A. Martins, and P. S. R. Diniz, "Sparsity-aware data-selective adaptive filters," *IEEE Trans. Signal Process.*, vol. 62, no. 17, pp. 4557–4572, Sep. 2014.
[13] N. R. Yousef and A. H. Sayed, "A unified approach to the steady-state and tracking analysis of adaptive filters," *IEEE Trans. Signal Process.*, vol. 49, no. 2, pp. 314–324, Feb. 2001.
[14] T. Y. Al-Naffouri and A. H. Sayed, "Adaptive filters with error nonlinearities: Mean-square analysis and optimum design," *EURASIP T. Appl. Signal Process.*, , no. 4, pp. 192–205, Dec. 2001.
[15] Bin Lin, Rongxi He, Liming Song, and Baisuo Wang, "Steady-state performance analysis for adaptive filters with error nonlinearities," in *Acoustics, Speech and Signal Processing, 2009. ICASSP 2009. IEEE International Conference on*. IEEE, 2009, pp. 3093–3096.
[16] Kong-Aik Lee, Woon-Seng Gan, and Sen-Maw Kuo, "Mean-square performance analysis of the normalized subband adaptive filter," in *Signals, Systems and Computers, 2006. ACSSC'06. Fortieth Asilomar Conference on*. IEEE, 2006, pp. 248–252.
[17] B. Chen, L. Xing, J. Liang, N. Zheng, and J. C. Príncipe, "Steady-state mean-square error analysis for adaptive filtering under the maximum correntropy criterion," *IEEE Signal Processing Letters*, vol. 21, no. 7, pp. 880–884, Jul. 2014.
[18] Study Group n. 15, "Digital networks echo cancellers," Tech. Rep. G.168, International Telecommunication Union (ITU-T), Geneva, 2004.