A Block-Sparsity Induced NLMS Algorithm with Bias Compensation

Rodrigo M. S. Pimenta[1,2], Leonardo C. Resende[2,3], Lucas P. R. da Silva[4], Newton N. Siqueira[2,5], Jurair R. P. Junior[6], Mariane R. Petraglia[2] and Diego B. Haddad[4,6]

***Abstract*— Adaptive filtering algorithms are flexible mechanisms that adapt themselves to the environment statistics in which they are immersed. It is known that in practice several transfer functions are sparse, in the sense that their energy is concentrated in a few (sometimes clustered) coefficients. In this paper, a new normalized adaptive algorithm tailored to identifying block-sparse systems using a mixed $\ell_{2,0}$-norm of the adaptive coefficients is devised. Since the presence of noise in the input signal may induce an additional asymptotic bias in the estimation procedure, a compensation scheme is also advanced to address such an issue. At last, the computational burden is controlled by the adoption of a selective partial-update strategy. Simulated results indicate that the proposed algorithms present good performance compared to state-of-the-art alternatives, and allows the designer the choice of a convenient point regarding the trade-off between computational cost and convergence rate.**

***Keywords*— Block-Sparsity, Bias Compensation, Selective Partial Update**

I. Introduction

Adaptive filtering algorithms find wide application in areas such as channel equalization, acoustic echo cancellation and noise cancellation [1]. It is widely known that several systems in practice are sparse (or compressible), which means that most of their entries are close to zero (or even zero) and only a small fraction of nonzero or large coefficients exist in the impulse response [2]. Unfortunately, traditional sparsity-agnostic algorithms do not take advantage from this feature. In order to enhance both steady-state and transient abilities, sparsity-aware adaptive schemes are proposed and modelled [3]–[6]. Block-sparse systems (such as in satellite-linked or indoor MIMO) are an important kind of sparse transfer functions, whose impulse response concentrates itself in one or more clusters [7].

This paper advances a deterministic optimization problem whose solution describes the update equation of a normalized adaptive algorithm optimized for the identification of block-sparse plants. The estimated transfer function is partitioned into M equal-length groups and a penalization of solutions with large *mixed* $\ell_{2,0}$-norms is enforced.

[1]Coordenação de Telecomunicações, CEFET/RJ, Rio de Janeiro-RJ, Brazil. [2]Programa de Engenharia Elétrica (PEE), COPPE/UFRJ, Rio de Janeiro-RJ, Brazil. [3]Instituto Federal do Rio de Janeiro, *campus* Paracambi, Paracambi-RJ, Brazil. [4]Programa de Engenharia Elétrica (PPEEL), CEFET/RJ, Rio de Janeiro-RJ, Brazil. [5]Coordenação de Telecomunicações, CEFET/RJ *campus* Nova Iguaçu, Nova Iguaçu-RJ, Brazil. [6]Coordenação de Engenharia de Computação, CEFET/RJ *campus* Petrópolis, Petrópolis-RJ, Brazil. E-mails: rodrigo.pimenta@cefet-rj.br, leonardo.resende@ifrj.edu.br, lucas.paiva92@hotmail.com, newton.siqueira@cefet-rj.br, jurair.junior@cefet-rj.br, mariane@pads.ufrj.br, diego.haddad@cefet-rj.br.

This work is based on [7] and covers the following new contributions: *i*) the derivation methodology exploits the Lagrange multiplier method, instead of the stochastic gradient optimization; *ii*) the possibility of using a normalized update scheme, which facilitates the adoption of a step size β that guarantees convergence [8]; *iii*) the incorporation of a bias compensation mechanism, which takes into account the existence of a measurement noise in the input sequence $x(k)$ and does not impact the reference signal $d(k)$; *iv*) the insertion of a *selective update* procedure, which is able to reduce the required computation burden.

The paper is structured as follows. Section II describes the standard NLMS algorithm. Section III derives the proposed Block-Sparsity Induced NLMS algorithm, which is generalized in Section IV in order to compensate the bias caused by noise at the input of the adaptive filter. A further evolution of the advanced algorithm that reduces the required computational burden is described in Section V. Section VI shows simulation results. Concluding remarks are presented in Section VII.

II. The NLMS Algorithm

The normalized least mean squares (NLMS) algorithm consists of an adaptive scheme that solves the following optimization problem:

$$\min_{\boldsymbol{w}(k+1)} \mathcal{F}[\boldsymbol{w}(k+1)] \triangleq \frac{1}{2}\|\boldsymbol{w}(k+1)-\boldsymbol{w}(k)\|^2$$
$$\text{s.t. } e_{\rm p}(k)=(1-\beta)e(k), \tag{1}$$

in which[1] $\boldsymbol{w}(k)\in\mathbb{R}^N$ denotes the adaptive coefficient vector,

$$e(k) \triangleq d(k)-\boldsymbol{w}^T(k)\boldsymbol{x}(k), \tag{2}$$
$$e_{\rm p}(k) \triangleq d(k)-\boldsymbol{w}^T(k+1)\boldsymbol{x}(k), \tag{3}$$

with the current excitation data concatenated in the vector

$$\boldsymbol{x}(k)\triangleq\begin{bmatrix}x(k) & x(k-1) & \dots & x(k-N+1)\end{bmatrix}^T. \tag{4}$$

It is noteworthy that $\mathcal{F}[\boldsymbol{w}(k+1)]$ penalizes solutions that are distant from the current estimated parameters $\boldsymbol{w}(k)$, which is an application of the conservative minimum disturbance principle [9]. The use of the Lagrange multipliers technique to solve the constrained optimization problem (1) leads to the NLMS algorithm [10]

$$\boldsymbol{w}(k+1)=\boldsymbol{w}(k)+\beta\frac{\boldsymbol{x}(k)e(k)}{\|\boldsymbol{x}(k)\|^2}, \tag{5}$$

[1]All vectors in this paper are of column type.

which does not take advantage from the block-sparsity of the system it intends to emulate. In the next section, a new block-sparsity-aware normalized algorithm is devised, in order to circumvent such an issue.

III. BS-NLMS Algorithm

In order to improve NLMS performance on block-sparse system identification, in this paper a mixed $\ell_{2,0}$-norm regularization term is inserted into the optimization problem (1). Such a regularization is applied to a partition of M equal-size groups, so that $\boldsymbol{w}(k)$ is decomposed into M blocks $\boldsymbol{w}_i(k)$, for $i \in \{0, 1, \dots, M-1\}$,

$$\boldsymbol{w}(k) = \begin{bmatrix} \boldsymbol{w}_0^T(k) & \boldsymbol{w}_1^T(k) & \dots & \boldsymbol{w}_{M-1}^T(k) \end{bmatrix}^T, \tag{6}$$

where the length L of each block is given by $L = N/M$, where[2] $L \in \mathbb{N}$.

Assuming that the unknown large coefficients are clustered (rather than being spread in an arbitrary manner), it is expected that the following *mixed* $\ell_{2,0}$-norm does not present large values in practice [7]:

$$\|\boldsymbol{w}(k)\|_{2,0} \triangleq \left\| \begin{bmatrix} \|\boldsymbol{w}_0(k)\|_2 \\ \|\boldsymbol{w}_1(k)\|_2 \\ \vdots \\ \|\boldsymbol{w}_{M-1}(k)\|_2 \end{bmatrix} \right\|_0, \tag{7}$$

where the ℓ_0-norm (actually, a *pseudo-norm*) is commonly approximated in order to make the mathematics tractable [11].

By using a regularization term related to $\ell_{2,0}$-norm penalization into (1), one obtains the following *proposed* optimization problem:

$$\min_{\boldsymbol{w}(k+1)} \mathcal{F}_{\text{BS}}[\boldsymbol{w}(k+1)] \triangleq \frac{1}{2}\|\boldsymbol{w}(k+1) - \boldsymbol{w}(k)\|^2 + \gamma\|\boldsymbol{w}(k+1)\|_{2,0}$$
$$\text{s.t. } e_{\text{p}}(k) = (1-\beta)e(k), \tag{8}$$

where $\gamma \in \mathbb{R}_+$ is an adjustable parameter that controls the regularization intensity.

Remarks: It should be emphasized that (8) consists in a new paradigm for the derivation of block-sparsity-induced algorithms, since it employs a deterministic local optimization problem, instead of the stochastic gradient adopted by [7]. Furthermore, as will be seen, it engineers a normalized update scheme, which is not the case of the algorithm advanced by [7].

A solution for (8) can be encountered by the Lagrange multiplier technique, which provides the *equivalent* unconstrained problem

$$\min_{\boldsymbol{w}(k+1)} \mathcal{G}_{\text{BS}}\left[\boldsymbol{w}(k+1)\right] \triangleq \mathcal{F}_{\text{BS}}[\boldsymbol{w}(k+1)] + \lambda\left[e_{\text{p}}(k) - (1-\beta)e(k)\right], \tag{9}$$

whose solution can be obtained by zeroing its gradient w.r.t. $\boldsymbol{w}(k+1)$:

$$\begin{aligned} \nabla_{\boldsymbol{w}(k+1)}\mathcal{G}_{\text{BS}}\left[\boldsymbol{w}(k+1)\right] &= \boldsymbol{w}(k+1) - \boldsymbol{w}(k) - \lambda\boldsymbol{x}(k) \\ &\quad -\gamma\boldsymbol{g}[\boldsymbol{w}(k+1)] = \boldsymbol{0}, \end{aligned} \tag{10}$$

[2]For simplicity, it is assumed that the ratio N/M is an integer.

where the i-th element of $\boldsymbol{g}[\boldsymbol{w}]$ is defined by

$$g_i(k) \triangleq \begin{cases} 2\rho^2 w_i(k) - \frac{2\rho w_i(k)}{\|\boldsymbol{w}_{\lfloor i/L \rfloor}\|_2}, & 0 < \|\boldsymbol{w}_{\lfloor i/L \rfloor}\|_2 < 1/\rho \\ 0, & \text{otherwise.} \end{cases},$$

where $\rho \in \mathbb{R}_+$ is a user-defined parameter that influences the approximation of the ℓ_0-norm [12]. Fig. 1 depicts $g_0(k)$ in the bi-dimensional case (*i.e.*, $N = 2$).

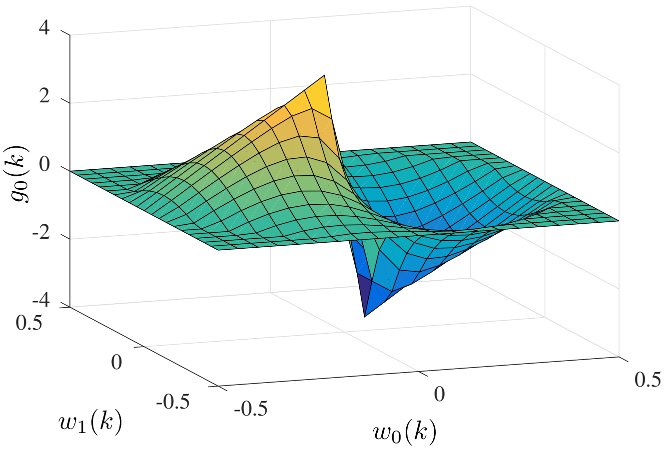

Fig. 1. Function $g_0(k)$ w.r.t. $w_0(k)$ and $w_1(k)$, for $\rho = 2$.

Using the approximation $\boldsymbol{g}\left[\boldsymbol{w}(k+1)\right] \approx \boldsymbol{g}\left[\boldsymbol{w}(k)\right]$ in order to obtain a proper recursion [2], (10) can be rewritten as

$$\boldsymbol{w}(k+1) = \boldsymbol{w}(k) + \gamma\boldsymbol{g}[\boldsymbol{w}(k)] + \lambda\boldsymbol{x}(k), \tag{11}$$

where λ can be computed by applying the affine constraint of (8) into (11), which leads to

$$\lambda\|\boldsymbol{x}(k)\|^2 + \underbrace{\gamma\boldsymbol{g}^T[\boldsymbol{w}(k)]\boldsymbol{x}(k)}_{\approx\boldsymbol{0}} = \beta e(k) \Rightarrow \lambda = \frac{\beta e(k)}{\|\boldsymbol{x}(k)\|^2}, \tag{12}$$

where the approximation in (12) (which reduces the computational burden required by the algorithm) is motivated by [13]. From (12) and (11), the update equation of the *proposed* BS-NLMS (block-sparsity-induced NLMS) algorithm is given by

$$\boldsymbol{w}(k+1) = \boldsymbol{w}(k) + \gamma\boldsymbol{g}[\boldsymbol{w}(k)] + \beta\frac{\boldsymbol{x}(k)e(k)}{\|\boldsymbol{x}(k)\|^2}. \tag{13}$$

Remark: it is noteworthy that term $\gamma\boldsymbol{g}[\boldsymbol{w}(k)]$ in (13) is responsible for inserting a bias in the steady-state solution obtained by the normalized algorithm. Such a bias leads to enhanced mean square performance when the unknown plant indeed presents a sparse characteristic [11], [14].

IV. BC-BS-NLMS

Recent papers bring attention to a bias introduced in the adaptive filtering approach by the presence of a noise $\eta(k)$ at the input of the adaptive filter [15], [16], as depicted in Fig. 2.

An additional contribution of this paper in order to attenuate the impact of $\eta(k)$ is the insertion of a *bias compensation term* $\boldsymbol{\psi}(k)$ in (13), given by

$$\boldsymbol{w}(k+1) = \boldsymbol{w}(k) + \gamma\boldsymbol{g}[\boldsymbol{w}(k)] + \beta\frac{\boldsymbol{x}(k)e(k)}{\|\boldsymbol{x}(k)\|^2} + \boldsymbol{\psi}(k), \tag{14}$$

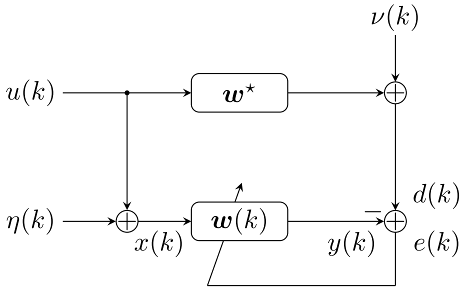

Fig. 2. Block diagram of a system identification adaptive filtering system, where the adaptive filter input is corrupted by noise $\eta(k)$.

where the compensation vector $\boldsymbol{\psi}(k)$ should *not* impose an asymptotic unbiased estimation[3], but aims at reducing the component of the resulting bias derived from the noise $\eta(k)$.

In order to obtain a feasible vector $\boldsymbol{\psi}(k)$, consider

$$\tilde{\boldsymbol{w}}(k) \triangleq \boldsymbol{w}^\star - \boldsymbol{w}(k) \tag{15}$$

as the deviation vector, so that the error signal can be written as

$$e(k) = \tilde{\boldsymbol{w}}^T(k)\boldsymbol{u}(k) + \nu(k) - \boldsymbol{w}^T(k)\boldsymbol{\eta}(k), \tag{16}$$

where

$$\boldsymbol{u}(k) \triangleq \begin{bmatrix} u(k) & u(k-1) & \dots & u(k-N+1) \end{bmatrix}, \tag{17}$$
$$\boldsymbol{\eta}(k) \triangleq \begin{bmatrix} \eta(k) & \eta(k-1) & \dots & \eta(k-N+1) \end{bmatrix}. \tag{18}$$

Using (14)-(16) and applying the expectation operator $\mathbb{E}[\cdot]$, the following average recursion on $\tilde{\boldsymbol{w}}(k)$ can be established:

$$\begin{aligned}\mathbb{E}[\tilde{\boldsymbol{w}}(k+1)] = \mathbb{E}[\tilde{\boldsymbol{w}}(k)] &- \beta\mathbb{E}\left[\frac{\boldsymbol{x}(k)\boldsymbol{u}^T(k)\tilde{\boldsymbol{w}}(k)}{\|\boldsymbol{x}(k)\|^2}\right] \\ &-\beta\mathbb{E}\left[\frac{\boldsymbol{x}(k)\nu(k)}{\|\boldsymbol{x}(k)\|^2}\right] + \beta\mathbb{E}\left[\frac{\boldsymbol{x}(k)\boldsymbol{\eta}^T(k)\boldsymbol{w}(k)}{\|\boldsymbol{x}(k)\|^2}\right] \\ &-\mathbb{E}[\boldsymbol{\psi}(k)] - \gamma\mathbb{E}\{\boldsymbol{g}[\boldsymbol{w}(k)]\},\end{aligned} \tag{19}$$

which can be simplified by the usage of the following stochastic hypotheses:

⋆ **H1**: zero-mean sequences $\nu(k)$, $u(k)$ and $\eta(k)$ are statistically independent;

⋆ **H2**: $\nu(k)$ and $\eta(k)$ are white processes;

⋆ **H3**: excitation vector $\boldsymbol{x}(k)$ and weight vector $\boldsymbol{w}(k)$ are statistically independent.

Remarks: whereas **H1** and **H2** are popular in the open literature and often valid, the *independence assumption* **H3** is clearly violated in practice, since the tapped-delay adaptive structure imposes a *deterministic* coherence between consecutive vectors $\boldsymbol{x}(k)$ [17]. Such an assumption, popular in the field of stochastic approximations, turns the mathematics tractable, and thereby more accurate when the step size is not large. It is noteworthy that this presumption can be circumvented by the *exact expectation analysis* method, which requires a cumbersome number of algebraic manipulations, even for small-length filters [18]–[20].

[3]It should be noted that the original BS-NLMS imposes a bias in order to enhance mean square performance of the identification of block-sparse plants.

Consider $\boldsymbol{b}_{\text{ss}} \triangleq \lim_{k\to\infty}\mathbb{E}[\tilde{\boldsymbol{w}}(k)]$ as the steady-state bias implied by (14). Using **H1**-**H3**, it can be shown that such a bias can be written as

$$\boldsymbol{b}_{\text{ss}} = \lim_{k\to\infty}\left\{\beta\sigma_\eta^2\mathbb{E}[\boldsymbol{w}(k)] - \mathbb{E}[\boldsymbol{\psi}(k)] + \boldsymbol{\chi}(k)\right\}, \tag{20}$$

where $\boldsymbol{R}_u \triangleq \mathbb{E}\left[\boldsymbol{u}(k)\boldsymbol{u}^T(k)\right]$, σ_η^2 denotes the variance of $\eta(k)$ and

$$\boldsymbol{\chi}(k) \triangleq \beta\boldsymbol{R}_u\mathbb{E}[\tilde{\boldsymbol{w}}(k)] - \gamma\mathbb{E}\{\boldsymbol{g}[\boldsymbol{w}(k)]\} \tag{21}$$

depends on the block-sparsity induced penalization. Note that component $\beta\sigma_\eta^2\mathbb{E}[\boldsymbol{w}(k)]$ is due to $\eta(k)$, and that its cancelling leads to

$$\mathbb{E}[\boldsymbol{\psi}(k)] = \beta\sigma_\eta^2\mathbb{E}[\boldsymbol{w}(k)]. \tag{22}$$

Since (22) imposes a constraint in the mean value of random vector $\boldsymbol{\psi}(k)$, it is necessary to approximate it by using information observable in practice, which leads to the following choice:

$$\boldsymbol{\psi}(k) = \beta\hat{\sigma}_\eta^2\boldsymbol{w}(k), \tag{23}$$

where $\hat{\sigma}_\eta^2$ is the estimate of σ_η^2. Methods for estimating σ_η^2 are not addressed in this paper. Identity (23) leads to the update equation of the devised BC-BS-NLMS algorithm

$$\boldsymbol{w}(k+1) = \boldsymbol{w}(k) + \gamma\boldsymbol{g}[\boldsymbol{w}(k)] + \beta\frac{\boldsymbol{x}(k)e(k)}{\|\boldsymbol{x}(k)\|^2} + \beta\hat{\sigma}_\eta^2\frac{\boldsymbol{w}(k)}{\|\boldsymbol{x}(k)\|^2}. \tag{24}$$

Note that by rewriting optimization problem (8) as

$$\begin{aligned}\min_{\boldsymbol{w}(k+1)} \mathcal{F}_{\text{BS}}[\boldsymbol{w}(k+1)] &\triangleq \frac{1}{2}\|\boldsymbol{w}(k+1) - \boldsymbol{w}(k)\|^2 + \gamma\|\boldsymbol{w}(k+1)\|_{2,0} \\ \text{s.t. } e_{\text{p}}(k) &= (1 - \beta\|\boldsymbol{x}(k)\|^2)e(k),\end{aligned} \tag{25}$$

and using similar steps than those that have led to (24), one may obtain the following (proposed) BC-BS-LMS algorithm:

$$\boldsymbol{w}(k+1) = \boldsymbol{w}(k) + \gamma\boldsymbol{g}[\boldsymbol{w}(k)] + \beta\boldsymbol{x}(k)e(k) + \beta\hat{\sigma}_\eta^2\boldsymbol{w}(k). \tag{26}$$

V. SPU-BC-BS-NLMS

Applications that require large-length adaptive filters may demand a prohibitively high computational burden. In such critical cases, a *selective partial update* (SPU) method can be performed. The use of a SPU strategy implies that only a fraction of the adaptive coefficients is updated at each iteration.

In order to motivate the SPU-BC-BS-NLMS, consider the following partition of the regressor vector into M equal-length blocks[4]:

$$\boldsymbol{x}(k) \triangleq \begin{bmatrix}\boldsymbol{x}_0^T(k) & \boldsymbol{x}_1^T(k) & \dots & \boldsymbol{x}_{M-1}^T(k)\end{bmatrix}^T, \tag{27}$$

where in each iteration the devised algorithm updates B blocks of the adaptive weight vector (see (6)).

The SPU method can be obtained by the resulting solution of the constrained optimization problem

$$\begin{aligned}\min_{\boldsymbol{w}_i(k+1)} \ &\frac{1}{2}\|\boldsymbol{w}_i(k+1) - \boldsymbol{w}_i(k)\|^2 + \gamma\|\boldsymbol{w}_i(k+1)\|_{2,0} \\ &\text{s.t. } e_{\text{p}}(k) = (1-\beta)e(k),\end{aligned} \tag{28}$$

[4]Note that it is not necessary that the number of partitions of the SPU strategy be the same as that of the BS method. Such an equality was enforced in order to simplify the equations.

which leads to[5]

$$\boldsymbol{w}_i(k+1) = \boldsymbol{w}_i(k) + \gamma \boldsymbol{g}[\boldsymbol{w}_i(k)] + \beta \frac{\boldsymbol{x}_i(k)e(k)}{\|\boldsymbol{x}_i(k)\|^2}, \quad (29)$$

whereas the remaining blocks of $\boldsymbol{w}(k)$ last unaltered. Note that (29) does not indicate which block should be updated. In this paper, the following criteria is proposed:

$$\begin{aligned} i &= \arg \min_{1\le j\le M} \|\boldsymbol{w}_j(k+1) - \boldsymbol{w}_j(k)\|^2 \\ &\approx \arg \min_{1\le j\le M} \left[\frac{\beta^2 e^2(k)}{\|\boldsymbol{x}_i(k)\|^2}\right] = \arg \max_{1\le j\le M} \|\boldsymbol{x}_j(k)\|^2, \quad (30) \end{aligned}$$

whose approximation neglects the component $\gamma \boldsymbol{g}[\boldsymbol{w}_i(k)]$ in (29), in order to reduce the computational effort.

The update of one block per iteration can be a very restricted method. Consider that the designer intends to update B blocks in each iteration, whose indices are denoted by $\mathcal{I}_B = \{i_0, i_1, \ldots, i_{B-1}\}$, which are a subset of $S = \{0, 1, \ldots, M-1\}$. In order to address such a configuration, one may generalize (28), which leads to

$$\begin{aligned} \min_{\boldsymbol{w}_{\mathcal{I}_B}(k+1)} & \frac{1}{2}\|\boldsymbol{w}_{\mathcal{I}_B}(k+1) - \boldsymbol{w}_{\mathcal{I}_B}(k)\|^2 + \gamma \|\boldsymbol{w}_{\mathcal{I}_B}(k+1)\|_{2,0} \\ & \text{s.t. } e_{\mathrm{p}}(k) = (1-\beta)e(k), \quad (31) \end{aligned}$$

whose solution (following the same steps that gave rise to (29)) is

$$\boldsymbol{w}_{\mathcal{I}_B}(k+1) = \boldsymbol{w}_{\mathcal{I}_B}(k) + \gamma \boldsymbol{g}[\boldsymbol{w}_{\mathcal{I}_B}(k)] + \beta \frac{\boldsymbol{x}_{\mathcal{I}_B}(k)e(k)}{\|\boldsymbol{x}_{\mathcal{I}_B}(k)\|^2}, \quad (32)$$

where

$$\mathcal{I}_B = \arg \max_{J_B \in S} \sum_{j\in J_B} \|\boldsymbol{x}_j(k)\|^2 \quad (33)$$

is a selection criterion that requires $N+BL+2$ multiplications. Using the same steps of Section IV, Eq. (32) can be generalized in order to obtain the proposed SPU-BC-BS-NLMS:

$$\begin{aligned} \boldsymbol{w}_{\mathcal{I}_B}(k+1) &= \boldsymbol{w}_{\mathcal{I}_B}(k) + \gamma \boldsymbol{g}[\boldsymbol{w}_{\mathcal{I}_B}(k)] + \beta \frac{\boldsymbol{x}_{\mathcal{I}_B}(k)e(k)}{\|\boldsymbol{x}_{\mathcal{I}_B}(k)\|^2} \\ &+ \beta \hat{\sigma}_\eta^2 \frac{\boldsymbol{w}_{\mathcal{I}_B}(k)}{\|\boldsymbol{x}_{\mathcal{I}_B}(k)\|^2}. \quad (34) \end{aligned}$$

VI. Results

In the following simulations, the transfer functions to be estimated are sampled from a Markov-Gaussian (M-G) model $\mathcal{M}(N, p_1, p_2, \sigma_s^2)$, which engineers a wide range of block-sparse systems [7]. In such a model, the impulse response $\boldsymbol{w}^\star$ is computed in two steps. In the first one, a first-order Markov process is responsible for producing the sets which contain the index of nonzero and zero coefficients (see Fig. 3), according to the following rule:

$$P\{s_j = 0 | s_{j-1} = 0\} = p_1, \quad (35)$$
$$P\{s_j \neq 0 | s_{j-1} \neq 0\} = p_2, \quad (36)$$

where p_1 and p_2 are adjustable parameters. Note that $(1-p_2)$ should be far larger than $(1-p_1)$ in order to guarantee a block-sparse system response [7].

After the determination of the index set of nonzero coefficients, the amplitudes of the nonzero coefficients are sampled

[5]The derivation is omitted here due to lack of space.

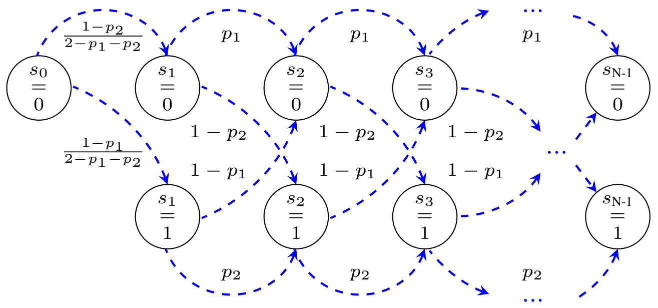

Fig. 3. Diagram for block-sparse model with impulse-response generation.

from a zero-mean Gaussian distribution with variance σ_s^2. Mathematically, one may write

$$w_i^\star = \begin{cases} 0, & \text{if } s_i = 0 \\ r, & \text{if } s_i = 1 \end{cases}, \quad (37)$$

where r is a Gaussian random variable with variance σ_s^2. In the following simulations, the configuration $(p_1, p_2, \sigma_s^2) = (0.999, 0.9, 1)$ was adopted.

Fig. 4 depicts the evolution of the mean square deviation (MSD) of LMS, BC-LMS and BC-BS-LMS (proposed) algorithms, where the MSD is defined by

$$\text{MSD}(k) \triangleq \mathbb{E}\left[\|\boldsymbol{w}^\star - \boldsymbol{w}(k)\|^2\right]. \quad (38)$$

The following parameters were employed in Fig. 4: $N = 800$, $\rho = 1$, $\sigma_\nu^2 = 2\cdot 10^{-2}$, $\sigma_u^2 = 1$, $\hat{\sigma}_\eta^2 = \sigma_\eta^2 = 10^{-1}$, $P = 4$, $\delta = 10^{-8}$, $\beta_{\text{LMS}} = \frac{0.5}{N\cdot\sigma_x^2}$, $\beta_{\text{BSLMS}} = \frac{0.5}{N\cdot\sigma_x^2}$, $\beta_{\ell_0\text{BSNLMS}} = 0.6$ and $\kappa = 1.5\cdot 10^{-6}$. All results were computed from the average of 200 independent Monte Carlo trials. From Fig. 4, one may notice that the advanced BC-BS-LMS algorithm outperforms the other ones.

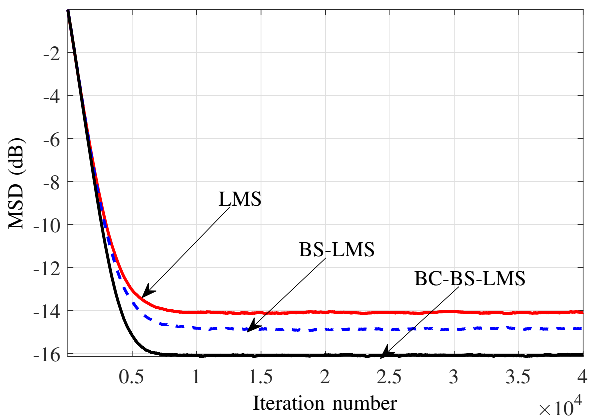

Fig. 4. MSD evolution (in dB) for the LMS, BS-LMS and BC-BS-LMS algorithms.

Fig. 5 shows the evolution of the NLMS, BS-NLMS and BC-BS-NLMS algorithms as a function of the number of iterations, also demonstrating the superior learning capability of the proposed BC-BS-NLMS algorithm. For generating Fig. 5, the following parameters were employed: $\beta_{\text{NLMS}} = 0.5$, $\beta_{\text{BSNLMS}} = 0.4$, $\beta_{\text{BCBSNLMS}} = 0.4$ $N = 800$, $\rho = 1$,

$\sigma_\nu^2 = 2 \cdot 10^{-2}$, $\sigma_u^2 = 1$, $\hat{\sigma}_\eta^2 = \sigma_\eta^2 = 10^{-1}$, $P = 4$, $\delta = 10^{-8}$. All results were computed from 100 independent Monte Carlo trials.

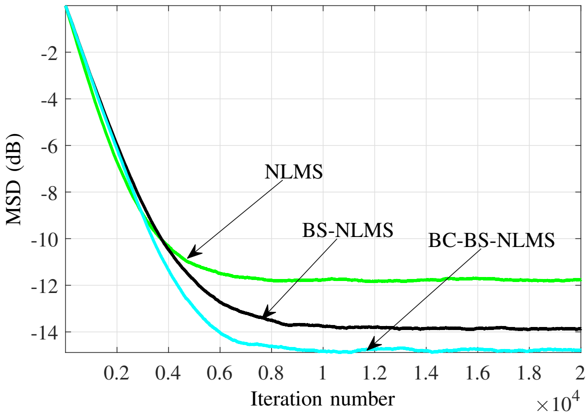

Fig. 5. MSD evolution (in dB) for the NLMS, BS-NLMS and BC-BS-NLMS algorithms.

Fig. 6 presents the MSD as a function of the number of iterations for four values of $B \in 1, 2, 3, 4$ and $M = 4$, using the same parameters as those of the experiment that led to Fig. 5. When $B = 4$, the algorithm degenerates into the proposed BC-BS-NLMS and that the SPU versions allow one to exchange convergence rate by computational complexity. All results were computed from the average of 100 independent Monte Carlo trials.

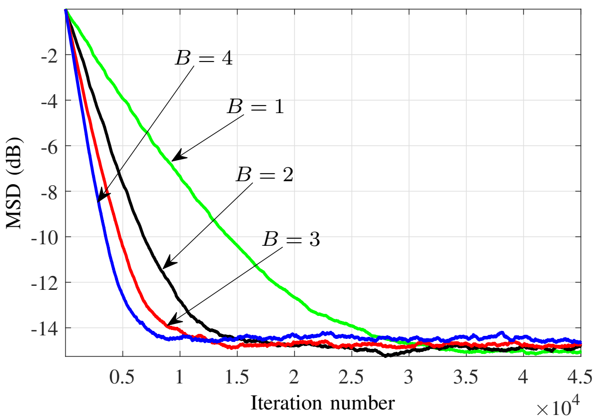

Fig. 6. MSD evolution (in dB) of the SPU-BC-BS-NLMS algorithm for different values of B.

VII. Conclusions

Recently, adaptive filtering algorithms tailored to the identification block-sparse transfer functions have been proposed. The first contribution of this paper is their extension to normalized schemes, which has the advantage of presenting a stability upper bound on the step size that is less dependent on the input signal statistics. Furthermore, the steady-state bias induced by the presence of additive noise in the excitation data is mitigated by a novel bias compensation strategy. The last contribution is the adoption of a partial update method that can be used to reduce the computational burden of the devised methods. The results have confirmed the performance enhancement obtained by the advanced algorithms.

References

[1] Z. Zheng, Z. Liu, H. Zhao, Y. Yu, and L. Lu, "Robust set-membership normalized subband adaptive filtering algorithms and their application to acoustic echo cancellation," *IEEE Transactions on Circuits and Systems I: Regular Papers*, vol. 64, pp. 2098–2111, Aug 2017.

[2] M. V. Lima, T. N. Ferreira, W. A. Martins, and P. S. Diniz, "Sparsity-aware data-selective adaptive filters," *IEEE Transactions on Signal Processing*, vol. 62, no. 17, pp. 4557–4572, 2014.

[3] D. L. Duttweiler, "Proportionate normalized least-mean-squares adaptation in echo cancelers," *IEEE Transactions on Speech and Audio Processing*, vol. 8, no. 5, pp. 508–518, 2000.

[4] M. R. Petraglia and D. B. Haddad, "New adaptive algorithms for identification of sparse impulse responses - analysis and comparisons," in *2010 7th International Symposium on Wireless Communication Systems*, pp. 384–388, Sept. 2010.

[5] R. A. d. Prado, F. d. R. Henriques, and D. B. Haddad, "Sparsity-aware distributed adaptive filtering algorithms for nonlinear system identification," in *2018 International Joint Conference on Neural Networks (IJCNN)*, pp. 1–8, July 2018.

[6] M. R. Petraglia and D. B. Haddad, "Mean-square error and stability analysis of a subband structure for the rapid identification of sparse impulse responses," *Digital Signal Processing*, vol. 22, no. 6, pp. 1068 – 1072, 2012.

[7] S. Jiang and Y. Gu, "Block-sparsity-induced adaptive filter for multi-clustering system identification," *IEEE Transactions on Signal Processing*, vol. 63, pp. 5318–5330, Oct 2015.

[8] D. T. M. Slock, "On the convergence behavior of the LMS and the normalized LMS algorithms," *IEEE Transactions on Signal Processing*, vol. 41, pp. 2811–2825, Sep. 1993.

[9] S. S. Haykin, *Adaptive filter theory*. Pearson Education India, 2005.

[10] A. H. Sayed, *Fundamentals of adaptive filtering*. John Wiley & Sons, 2003.

[11] G. Su, J. Jin, Y. Gu, and J. Wang, "Performance analysis of ℓ_0-norm constraint least mean square algorithm," *IEEE Transactions on Signal Processing*, vol. 60, pp. 2223–2235, May 2012.

[12] Y. Gu, J. Jin, and S. Mei, "ℓ_0-norm constraint LMS algorithm for sparse system identification," *IEEE Signal Processing Letters*, vol. 16, pp. 774–777, Sep. 2009.

[13] D. B. Haddad, M. R. Petraglia, and A. Petraglia, "A unified approach for sparsity-aware and maximum correntropy adaptive filters," in *Signal Processing Conference (EUSIPCO), 2016 24th European*, pp. 170–174, IEEE, 2016.

[14] K. da S. Olinto, D. B. Haddad, and M. R. Petraglia, "Transient analysis of ℓ_0-LMS and ℓ_0-NLMS algorithms," *Signal Processing*, vol. 127, pp. 217–226, 2016.

[15] Z. Zheng and H. Zhao, "Bias-compensated normalized subband adaptive filter algorithm," *IEEE Signal Processing Letters*, vol. 23, pp. 809–813, June 2016.

[16] R. M. S. Pimenta, L. C. Resende, N. N. Siqueira, D. B. Haddad, and M. R. Petraglia, "Algoritmo NLMS com reúso de coeficientes e compensação de viés para entradas ruidosas," in *XXXVI Simpósio Brasileiro de Telecomunicações e Processamento de Sinais*, pp. 1–5, Sept. 2018.

[17] P. Lara, L. D. Tarrataca, and D. B. Haddad, "Exact expectation analysis of the deficient-length LMS algorithm," *Signal Processing*, vol. 162, pp. 54 – 64, 2019.

[18] P. Lara, K. d. S. Olinto, F. R. Petraglia, and D. B. Haddad, "Exact analysis of the least-mean-square algorithm with coloured measurement noise," *Electronics Letters*, vol. 54, no. 24, pp. 1401–1403, 2018.

[19] P. Lara, D. B. Haddad, and L. Tarrataca, "Advances on the analysis of the LMS algorithm with a colored measurement noise," *Signal, Image and Video Processing*, vol. 14, no. 3, pp. 529–536, 2020.

[20] P. Lara, F. Igreja, L. D. T. J. Tarrataca, D. B. Haddad, and M. R. Petraglia, "Exact expectation evaluation and design of variable step-size adaptive algorithms," *IEEE Signal Processing Letters*, vol. 26, no. 1, pp. 74–78, 2019.