Avoiding divergence in the Shalvi-Weinstein Algorithm

Maria D. Miranda, Magno T. M. Silva*, *Member, IEEE,*

and Vítor H. Nascimento, *Member, IEEE*

Abstract

The most popular algorithms for blind equalization are the constant-modulus algorithm (CMA) and the Shalvi-Weinstein algorithm (SWA). It is well-known that SWA presents a higher convergence rate than CMA, at the expense of higher computational complexity. If the forgetting factor is not sufficiently close to one, if the initialization is distant from the optimal solution, or if the signal-to-noise ratio is low, SWA can converge to undesirable local minima or even diverge. In this paper, we show that divergence can be caused by the inconsistency in the nonlinear estimate of the transmitted signal, or (when the algorithm is implemented in finite precision) by the loss of positiveness of the estimate of the autocorrelation matrix, or by a combination of both. In order to avoid the first cause of divergence, we propose a dual-mode SWA (DM-SWA). In the first mode of operation, the new algorithm works as SWA, and in the second mode, it rejects non-consistent estimates of the transmitted signal. Assuming the persistence of excitation condition, we present a deterministic stability analysis of the new algorithm. To avoid the second cause of divergence, we propose a dual-mode lattice SWA (DM-LSWA), which is stable even in finite-precision arithmetic, and has a computational complexity that increases linearly with the number of adjustable equalizer coefficients. The good performance of the proposed algorithms is confirmed through numerical simulations.

Index Terms

Adaptive equalizers, blind equalization, Shalvi-Weinstein Algorithm, nonlinearities, numerical stability, lattice filters.

This work was partly supported by FAPESP under grant No. 2008/00773-1 and by CNPq under grant No. 303361/2004-2.

M. D. Miranda, M. T. M. Silva and V. H. Nascimento are with Escola Politécnica, University of São Paulo, São Paulo, SP, Brazil, e-mails: maria@lcs.poli.usp.br, {magno, vitor}@lps.usp.br, ph. +55-11-3091-5606, fax: +55-11-3091-5718.

SP EDICS: 1.ASP-ANAL, 2.SPC-BLND, 3.ASP-FAST.

*Corresponding author: Magno T. M. Silva.

I. Introduction

Adaptive equalizers are widely used in modern digital communications systems to remove intersymbol interference introduced by dispersive channels. Over the last decades, this has been an area of intense research in the signal processing community (see, e.g., [1]–[4] and the many references therein). The performance of an adaptive equalization algorithm can be evaluated by different factors such as accuracy of the steady-state solution, convergence rate, tracking abilities, computational cost, numerical robustness, stability, etc. [5], [6]. The design of an adaptive algorithm for supervised or blind equalization with a good tradeoff among these factors is a problem of wide interest.

In supervised equalization, the least-mean-squares (LMS) and the recursive least-squares (RLS) algorithms are the most popular for the adaptation of finite impulse response (FIR) equalizers [5]. It is well-known in the literature that these algorithms present a tradeoff between convergence rate and computational cost, which tends to be less critical when fast versions of RLS are compared to LMS [5]. Among the members of the fast RLS family, the fast QR-RLS [7] and the error feedback least-squares lattice (EF-LSL) [8] algorithms are the most attractive in terms of numerical stability and computational cost. The former uses the QR factorization of matrices and is numerically stable as shown analytically in [9] from a backward stability perspective. The latter is numerically well-behaved even in finite precision, although no proof of its numerical stability is known [10]. Moreover, its computational cost is slightly lower than the fast QR-RLS algorithm.

In blind equalization, there is no training data and the algorithms may update the equalizer coefficients using higher-order statistics (HOS) of the transmitted signal. A simplified communications system with a blind equalizer based on HOS is depicted in Figure 1 [3]. The signal $a(n)$, assumed iid (independent and identically distributed) and non Gaussian, is transmitted through an unknown channel, whose model is constituted by an FIR filter $H(z)$ and additive white Gaussian noise $\eta(n)$. From the received signal $u(n)$ and the known HOS of the transmitted signal, the blind equalizer must mitigate the channel effects and recover the signal $a(n)$ for some delay τ_d. The equalizer output is given by $y(n) = \mathbf{w}^H\mathbf{u}(n)$, where $\mathbf{u}(n)$ is the input regressor vector, $\mathbf{w}$ the equalizer weight vector (both column vectors with M coefficients), and the superscript H denotes complex conjugate transposition.

In this context, the constant-modulus algorithm (CMA) [11], [12] and the Shalvi-Weinstein algorithm (SWA) [13] are the most popular. CMA is a stochastic gradient algorithm obtained from the minimization of the CM cost function defined as

$$J_{\mathrm{CM}} = \mathrm{E}\left\{(|y(n)|^2 - r)^2\right\}, \tag{1}$$

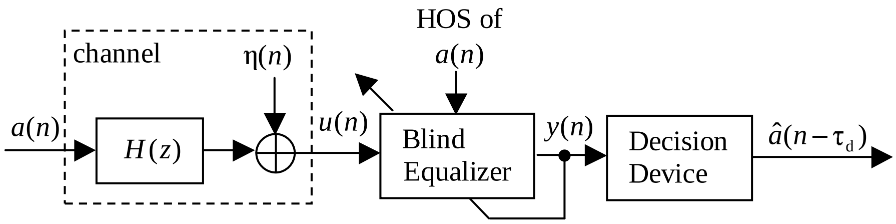

Fig. 1. Schematic representation of a communications system with blind equalizer.

where $\mathrm{E}\{\cdot\}$ stands for the expectation operation and $r = \mathrm{E}\{|a(n)|^4\}/\mathrm{E}\{|a(n)|^2\}$. SWA was originally derived in [13], using empirical cumulants in the minimization of the SW cost function defined as

$$J_{\mathrm{SW}} = \frac{C^y_{2,2}}{(C^y_{1,1})^2}, \tag{2}$$

where

$$C^y_{2,2} \triangleq \mathrm{E}\{|y(n)|^4\} - \beta \mathrm{E}\{|y(n)|^2\}^2$$

and

$$C^y_{1,1} \triangleq \mathrm{E}\{|y(n)|^2\}$$

are the cumulants of y of order $(2,2)$ and $(1,1)$, respectively, and $\beta = 2$ (resp., $\beta = 3$) for complex (resp., real) data. As shown in [14], under certain conditions, the SW cost function reduces to the CM cost function. Thus, CMA and SWA seek to optimize the same criterion. Hence, SWA can also be interpreted as a constant-modulus-based algorithm which uses an approximation for the Hessian matrix, i.e., it can be considered as a quasi-Newton-type algorithm [13], [15].

Based on the link between blind equalization and classical adaptive filtering of [16], CMA and SWA can be viewed as the blind versions of LMS and RLS, respectively. Hence, they also have a tradeoff between convergence rate and computational cost. Thus, SWA has a higher convergence rate than that of CMA, at the expense of higher computational complexity. Due to the multimodality of the SW (resp., CM) cost function, an inadequate choice of the forgetting factor (resp., step-size) of SWA (resp., CMA), an initialization distant from the zero-forcing solution, and a low signal-to-noise ratio are three factors which can lead both algorithms to diverge (i.e., the norm of the weight vector goes to infinity) or to converge to undesirable local minima. Many important results on the convergence and stability of constant-modulus-based algorithms have been obtained in the literature (see, e.g., [3], [17]–[21] and the references therein). However, the major part of these results is focused on analyses of CMA. Little has been done to avoid divergence of these algorithms, mainly in the case of SWA.

In this work, we show that the divergence of SWA has two different causes: (i) inconsistency in the nonlinear estimate of the transmitted signal, and (ii) the loss of numerical compatibility in the update of the inverse of the autocorrelation matrix, as occurs in RLS. We also propose solutions for both problems. First, using a representation equivalent to the state-space representation of the RLS algorithm [9] and temporarily assuming infinite-precision arithmetic (so that (ii) does not occur), we find conditions to ensure the stability of SWA, and we propose a dual-mode algorithm, denoted by DM-SWA. In the first mode of operation, the proposed algorithm works as SWA. In the second mode, it does not use the estimate of the transmitted signal, updating the coefficient vector with an error proportional to the equalizer output. Assuming the persistence of excitation condition, we show that DM-SWA is stable in infinite-precision arithmetic. Second, inspired in [9] and [10], we propose a dual-mode lattice SWA (DM-LSWA), which maintains the convergence rate of SWA, avoids divergence even when implemented in finite precision, and whose computational cost increases linearly with M. As happens for EF-LSL, we do not provide a stability proof for DM-LSWA in finite-precision arithmetic. However, results of exhaustive simulations show that DM-LSWA presents reliable numerical results, avoiding both causes of divergence. Some of our results were published as a conference paper in [22]. In this paper, we extend our previous results by providing a stability proof for the infinite-precision filter and considering complex data.

The paper is organized as follows. In Section II, SWA is derived from the minimization of a deterministic cost function, reinforcing the link with the RLS algorithm. Based on this link and on the well-founded results on the RLS algorithm, DM-SWA is introduced and a stability analysis is presented in Section III. Then, in Section IV, we propose DM-LSWA, a fast and numerically stable version of DM-SWA. In Section V, we present some simulation results, comparing the proposed algorithms to the conventional SWA in different situations. The main conclusions of the paper are presented in Section VI.

II. Revisiting SWA

SWA was originally obtained in [13] through the minimization of (2), using empirical approximations for the cumulants. However, in order to reinforce the link between SWA and the RLS algorithm, SWA is revisited in this section from a least-squares perspective. This link is obtained by deriving SWA from the minimization of a deterministic cost-function (much as RLS solves a deterministic least-squares problem [5]), and will help us extend to SWA some of the well-known methods to avoid numerical instability in RLS.

It was shown in [14] that the SW cost function reduces to the CM cost function under certain conditions. In a different approach, we show below that SW may be obtained from a deterministic version of the CM cost function, that is,

$$J(n) = \sum_{\ell=0}^{n} \lambda^{n-\ell} (|y_{n,\ell}|^2 - r)^2, \tag{3}$$

where $y_{n,\ell} \triangleq \mathbf{w}^H(n)\mathbf{u}(\ell)$ and $0 \ll \lambda < 1$ is a forgetting factor.

Equating the gradient of $J(n)$ with respect to $\mathbf{w}(n)$ to the null vector, we get the normal equations

$$r\widehat{\mathbf{R}}(n)\mathbf{w}(n) = \mathbf{p}(n), \tag{4}$$

where

$$\widehat{\mathbf{R}}(n) = \sum_{\ell=0}^{n} \lambda^{n-\ell} \mathbf{u}(\ell)\mathbf{u}^H(\ell), \tag{5}$$

$$\mathbf{p}(n) = \sum_{\ell=0}^{n} \lambda^{n-\ell} |y_{n,\ell}|^2 y_{n,\ell}^* \mathbf{u}(\ell), \tag{6}$$

and $*$ stands for complex conjugate.

In order to obtain an update equation for $\mathbf{w}$, we define

$$\mathbf{\Delta w}(n) = \mathbf{w}(n) - \mathbf{w}(n-1) \tag{7}$$

and use the updating of $\widehat{\mathbf{R}}(n)$, which is given by

$$\widehat{\mathbf{R}}(n) = \lambda\widehat{\mathbf{R}}(n-1) + \mathbf{u}(n)\mathbf{u}^H(n). \tag{8}$$

The matrix $\widehat{\mathbf{R}}$ is initialized as $\widehat{\mathbf{R}}(-1) = \delta\mathbf{I}$, where $\mathbf{I}$ is the M-by-M identity matrix, and δ is a small positive constant. Thus, using (7) and (8), the left-hand side of (4) can be rewritten as

$$r\widehat{\mathbf{R}}(n)\mathbf{w}(n) = r\widehat{\mathbf{R}}(n)\mathbf{\Delta w}(n) + \lambda r\widehat{\mathbf{R}}(n-1)\mathbf{w}(n-1) + ry^*(n)\mathbf{u}(n), \tag{9}$$

where $y(n) \triangleq y_{n-1,n} = \mathbf{w}^H(n-1)\mathbf{u}(n)$.

For convenience, we define

$$\Delta y_{n,\ell} \triangleq |y_{n,\ell}|^2 y_{n,\ell} - |y_{n-1,\ell}|^2 y_{n-1,\ell} \tag{10}$$

and rewrite $\mathbf{p}(n)$ as

$$\mathbf{p}(n) = \sum_{\ell=0}^{n} \lambda^{n-\ell} \left[\Delta y_{n,\ell}^* + |y_{n-1,\ell}|^2 y_{n-1,\ell}^*\right] \mathbf{u}(\ell). \tag{11}$$

After some algebraic manipulations in (11), we get

$$\mathbf{p}(n) = \lambda\mathbf{p}(n-1) + |y(n)|^2 y^*(n)\mathbf{u}(n) + \boldsymbol{\rho}(n), \tag{12}$$

where

$$\boldsymbol{\rho}(n) \triangleq \sum_{\ell=0}^{n} \lambda^{n-\ell} \Delta y_{n,\ell}^{*} \mathbf{u}(\ell). \tag{13}$$

Using (9), (12), (13), and assuming that $\mathbf{w}(n-1)$ satisfies (4) at time instant $(n-1)$, we rewrite (4) as

$$r\widehat{\mathbf{R}}(n)\boldsymbol{\Delta}\mathbf{w}(n) - \boldsymbol{\rho}(n) = e^{*}(n)\mathbf{u}(n), \tag{14}$$

where

$$e(n) = \left[|y(n)|^2 - r\right] y(n). \tag{15}$$

To proceed, we obtain an approximation for $\boldsymbol{\rho}(n)$, similar to the approximations used in [13] and [23]. To this end, we assume that the cumulant

$$\mathbf{c} \triangleq \mathrm{E}\{|y(n)|^2 y^{*}(n)\mathbf{u}(n)\} - \beta C_{1,1}^{y} \mathrm{E}\{y^{*}(n)\mathbf{u}(n)\} \tag{16}$$

can be approximated at the time instant n, replacing expectations with empirical averages, i.e.,

$$\widehat{\mathbf{c}}_{n,k} = \sum_{\ell=0}^{n} \lambda^{n-\ell} |y_{k,\ell}|^2 y_{k,\ell}^{*} \mathbf{u}(\ell) - \beta C_{1,1}^{a} \widehat{\mathbf{R}}(n)\mathbf{w}(k). \tag{17}$$

As usual, we replaced $C_{1,1}^{y}$ by $C_{1,1}^{a} = \mathrm{E}\{|a(n)|^2\}$. The difference between (17) and the corresponding approximations in [13] and [23] is that we use an exponential window instead of a rectangular window. Using $k = n$ and $k = n-1$ in (17) and assuming that n is large enough such that $\widehat{\mathbf{c}}_{n,n} \approx \widehat{\mathbf{c}}_{n,n-1}$, we obtain

$$\boldsymbol{\rho}(n) \approx \beta C_{1,1}^{a} \widehat{\mathbf{R}}(n)\boldsymbol{\Delta}\mathbf{w}(n). \tag{18}$$

Then, replacing (18) in (14), we arrive at

$$\mathbf{w}(n) = \mathbf{w}(n-1) + \frac{1}{r - \beta C_{1,1}^{a}} e^{*}(n)\widehat{\mathbf{R}}^{-1}(n)\mathbf{u}(n). \tag{19}$$

It is important to observe that:

1) Eq. (19) characterizes SWA and was originally obtained in [13] through the minimization of (2), using empirical approximations for cumulants. Some of these approximations are similar to $\mathbf{c} \approx (1-\lambda)\widehat{\mathbf{c}}_{n,k}$, with $k = n-1$ or $k = n$;
2) as in RLS, the inverse matrix $\widehat{\mathbf{R}}^{-1}(n)$ is obtained via the matrix inversion lemma applied to (8) [6, p. 67];
3) the constellation $a(n)$ in practice is sub-Gaussian, ensuring that the denominator $(r - \beta C_{1,1}^{a})$ in (19) is always negative [6], [13];
4) the link between SWA and RLS is reinforced through the deterministic cost function (3) and the derivation of SWA presented here. Different from RLS, which provides an exact solution for

the least-squares criterion, (19) is not an exact solution for the minimization of (3) due to the approximation (18).

III. A DUAL-MODE SWA

In order to avoid divergence in SWA and derive a robust and dual-mode version of the algorithm, we rewrite (19) in the form of a supervised algorithm, i.e.,

$$\mathbf{w}(n) = \mathbf{w}(n-1) + [d(n) - y(n)]^* \,\widehat{\mathbf{R}}^{-1}(n)\mathbf{u}(n), \tag{20}$$

defining $d(n) \triangleq x(n)y(n)$ and

$$x(n) \triangleq \frac{|y(n)|^2 - \beta C_{1,1}^a}{r - \beta C_{1,1}^a}. \tag{21}$$

Note that (20) has the same structure as that of the RLS algorithm. Thus, using the state-space representation of the RLS algorithm [9], after some algebra, (20) can be rewritten as

$$\begin{bmatrix} \mathbf{w}(n) \\ d^*(n) - y^*(n) \end{bmatrix} = \mathbf{\Gamma}(n) \begin{bmatrix} \lambda \mathbf{w}(n-1) \\ d^*(n) \end{bmatrix}, \tag{22}$$

where

$$\mathbf{\Gamma}(n) = \begin{bmatrix} \widehat{\mathbf{R}}^{-1}(n)\widehat{\mathbf{R}}(n-1) & \widehat{\mathbf{R}}^{-1}(n)\mathbf{u}(n) \\ -\lambda^{-1}\mathbf{u}^H(n) & 1 \end{bmatrix}. \tag{23}$$

From (22), we can observe that:

1) $\mathbf{\Gamma}(n)$ is the same as the state-transition matrix of the RLS algorithm [9].
2) $d(n)$ can be interpreted as an estimate of the desired response (transmitted sequence). Different from the RLS algorithm where $d(n)$ does not depend on the filter output $y(n)$, here, it is obtained from a nonlinear function that depends on $y(n)$ and on the HOS of the transmitted sequence $a(n)$.

Since numerical instability in RLS arises in the recursion for $\widehat{\mathbf{R}}^{-1}(n)$ (obtained by applying the matrix inversion lemma to (8)) [9], and this recursion is the same for SWA and RLS, SWA may also diverge in finite-precision arithmetic. This observation is reinforced by the similarity of recursion (22) to its equivalent for RLS (the state-transition matrix $\mathbf{\Gamma}(n)$ is the same in both cases). In addition, SWA can also diverge because of the nonlinear nature of its recursion, since $d(n)$ depends on $y(n)|y(n)|^2$. This effect is similar to the instability problems of CMA and the least mean fourth algorithm (LMF), in which $y(n)$ is also fed back through a cubic function [21], [24].

Thus, SWA can diverge due to (i) the nonlinear feedback of the filter output (as LMF and CMA); and (ii) finite arithmetic problems (as RLS). A combination of both causes of divergence can also occur.

This will become clearer from the analysis further ahead, and also from the simulations, in which we use histograms of the time-before-divergence in several runs of SWA and of the modified versions (DM-SWA, LSWA, and DM-LSWA) to separate the two causes of divergence.

In the literature, there are different techniques to ensure the numerical consistency in the update of $\widehat{\mathbf{R}}^{-1}(n)$ as is the case of some versions of RLS algorithms that use the QR decomposition and guarantee implicitly the existence of $\widehat{\mathbf{R}}^{-1}(n)$ even for poorly exciting input sequences [5], [9]. In the remainder of this section, we assume that (ii) does not occur and focus on the first cause of divergence. A blind equalization algorithm that ensures the positiveness of $\widehat{\mathbf{R}}^{-1}(n)$, even when implemented in finite-precision arithmetic, is considered in Section IV.

Note that both $d(n)$ and $y(n)$ represent estimates of the desired response. Thus, it is reasonable to assume that these two estimates will be consistent only if they 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 negative, $x(n) \geq 0$ occurs when $|y(n)|^2 \leq \beta C_{1,1}^a$. Our proposal to remove the first cause of instability in SWA is to restrict $x(n)$ to be positive. To this end, we define

$$\bar{d}(n) = \begin{cases} d(n), & |y(n)|^2 \leq \beta C_{1,1}^a \\ 0, & \text{otherwise} \end{cases} \tag{24}$$

and we use $\bar{d}(n)$ instead of $d(n)$ in (20). Thus, when $d(n)$ and $y(n)$ are consistent, we use (20) unmodified with $\bar{d}(n) = d(n) = x(n)y(n)$. In this case, we say that the algorithm is in the *region of interest* (ROI). On the other hand, when $x(n) < 0$ ($|y(n)|^2 > \beta C_{1,1}^a$) the estimate $d(n)$ is rejected, i.e., $\bar{d}(n) = 0$ and (20) reduces to

$$\mathbf{w}(n) = \mathbf{w}(n-1) - y^*(n)\widehat{\mathbf{R}}^{-1}(n)\mathbf{u}(n). \tag{25}$$

The proposed dual-mode algorithm, denoted by DM-SWA, is summarized in Table I, where $\mathbf{P}(n) \triangleq \widehat{\mathbf{R}}^{-1}(n)$. Its computational cost per iteration is shown in Table III (see Section IV) for real and complex-valued data, and considering the number of real multiplications, real additions, real divisions, and comparisons ($\mathcal{C}$). The estimated number of operations per iteration of DM-SWA depends on the manner in which the calculations are performed. To obtain the computational cost of Table III, we assumed that the DM-SWA calculations are performed as for the RLS algorithm described in [6, Sec. 5.9].

A stability proof for DM-SWA is presented in the sequel. It is relevant to notice that it may converge to a good or a poor stationary point [3], [17], [18], that is, it still presents the problems of possible convergence to local minima, common to constant-modulus-based algorithms. The advantage is that the new algorithm will keep the filter weights bounded in infinite-precision arithmetic.

A stability analysis for DM-SWA

We show that the Euclidean norm of the coefficient vector $\|\mathbf{w}(n)\|$ of DM-SWA is bounded for all initial conditions if the input $u(n)$ satisfies a persistence of excitation condition. We also show that if $y(n)$ leaves the ROI for any reason (poor initial condition or large noise sample, for example), DM-SWA is guaranteed to return to the ROI after a finite-time interval. Our only assumption is the persistence of excitation condition below, which is a widely-used condition for deterministic stability of adaptive filters [5], [9], [25], [26].

A-1 *The input sequence $\{u(n)\}$ is persistently exciting if there exist α and ζ with $0 < \alpha \leq \zeta < \infty$, such that* [1],

$$\alpha \mathbf{I} \leq \widehat{\mathbf{R}}(n) \leq \zeta \mathbf{I}, \quad \forall n \geq -1.$$

As consequences of A-1, we have that

1) the spectral norms of $\widehat{\mathbf{R}}^{-1}(n)$ and of $\widehat{\mathbf{R}}(n)$ satisfy

$$\|\widehat{\mathbf{R}}^{-1}(n)\| \leq \alpha^{-1} \quad \text{and} \quad \|\widehat{\mathbf{R}}(n)\| \leq \zeta; \tag{26}$$

TABLE I

SUMMARY OF DM-SWA.

Initialization:
$\mathbf{w}(-1) = [0 \;\cdots\; 0\; 1\; 0 \;\cdots 0]^T, \quad 0 \ll \lambda < 1,$
$\mathbf{P}(-1) = \delta^{-1}\mathbf{I}, \quad \delta$: small positive constant
for $n = 0, 1, 2, 3, \ldots$ do:
$y(n) = \mathbf{w}^H(n-1)\mathbf{u}(n)$
$x(n) = \dfrac{\lvert y(n)\rvert^2 - \beta C_{1,1}^a}{r - \beta C_{1,1}^a}$
if $x(n) \geq 0$,
$\bar{d}(n) = x(n)y(n)$
else
$\bar{d}(n) = 0$
end
$\bar{e}(n) = \bar{d}(n) - y(n)$
$\mathbf{P}(n) = \dfrac{1}{\lambda}\left[\mathbf{P}(n-1) - \dfrac{\mathbf{P}(n-1)\mathbf{u}(n)\mathbf{u}^H(n)\mathbf{P}(n-1)}{\lambda + \mathbf{u}^H(n)\mathbf{P}(n-1)\mathbf{u}(n)}\right]$
$\mathbf{w}(n) = \mathbf{w}(n-1) + \bar{e}^*(n)\mathbf{P}(n)\mathbf{u}(n)$
end

[1]Given two matrices $\mathbf{A}$ and $\mathbf{B}$ of dimensions $M \times M$, the inequality $\mathbf{A} \geq \mathbf{B}$ means that the matrix difference $(\mathbf{A} - \mathbf{B})$ is positive semi-definite.

2) the Euclidean norm of the regressor vetor $\mathbf{u}(n)$ is bounded from above, i.e.,

$$0 \le \|\mathbf{u}(n)\| \le B_u < \infty, \quad B_u > 0. \tag{27}$$

From (24), under Assumption A-1 the absolute value of $\bar{d}(n)$ is bounded from above, i.e.,

$$0 \le |\bar{d}(n)| < \frac{\beta C^a_{1,1}}{\beta C^a_{1,1} - r}\sqrt{\beta C^a_{1,1}} = B_d < \infty. \tag{28}$$

The bound (28) is obtained from the maximum values that $x(n)$ and $|y(n)|$ can assume in the ROI. In this case, we have $\max\{x(n)\} = \beta C^a_{1,1}/(\beta C^a_{1,1} - r)$ and $\max\{|y(n)|\} = \sqrt{\beta C^a_{1,1}}$. Note that the maximum value of $|y(n)|$ corresponds to $x(n) = 0$, and consequently this is a very conservative bound. Furthermore, when the algorithm is outside the ROI, $\bar{d}(n) = 0$. Hence, (28) is valid independently of the operation mode of DM-SWA.

Theorem 1 *Under the persistence of excitation condition (Assumption A-1), the Euclidean norm of the coefficient vector of DM-SWA has the following upper bound*

$$\|\mathbf{w}(n)\| < \frac{1}{\alpha}\left[\lambda^{n+1}\zeta\,\|\mathbf{w}(-1)\| + \frac{B_d\,B_u}{1-\lambda}\right] < \infty, \tag{29}$$

which ensures the stability of the algorithm, independently of the operation mode.

Proof: Using (22), we get

$$\mathbf{w}(n) = \lambda\widehat{\mathbf{R}}^{-1}(n)\widehat{\mathbf{R}}(n-1)\mathbf{w}(n-1) + d^*(n)\widehat{\mathbf{R}}^{-1}(n)\mathbf{u}(n). \tag{30}$$

Multiplying both sides of (30) on the left by $\widehat{\mathbf{R}}(n)$, we obtain

$$\widehat{\mathbf{R}}(n)\mathbf{w}(n) = \lambda\widehat{\mathbf{R}}(n-1)\mathbf{w}(n-1) + d^*(n)\mathbf{u}(n). \tag{31}$$

Defining $\mathbf{q}(n) \triangleq \widehat{\mathbf{R}}(n)\mathbf{w}(n)$, (31) can be rewritten more compactly as

$$\mathbf{q}(n) = \lambda\mathbf{q}(n-1) + d^*(n)\mathbf{u}(n). \tag{32}$$

Considering the initial condition $\mathbf{q}(-1)$, (32) can be rewritten as

$$\mathbf{q}(n) = \lambda^{n+1}\mathbf{q}(-1) + \sum_{\ell=0}^{n}\lambda^{n-\ell}d^*(\ell)\mathbf{u}(\ell). \tag{33}$$

Multiplying both sides of (33) by $\widehat{\mathbf{R}}^{-1}(n)$, i.e., returning to $\mathbf{w}(n)$, we obtain

$$\mathbf{w}(n) = \lambda^{n+1}\widehat{\mathbf{R}}^{-1}(n)\widehat{\mathbf{R}}(-1)\mathbf{w}(-1) + \widehat{\mathbf{R}}^{-1}(n)\sum_{\ell=0}^{n}\lambda^{n-\ell}d^*(\ell)\mathbf{u}(\ell). \tag{34}$$

Applying the triangle inequality to (34), we arrive at

$$\|\mathbf{w}(n)\| \le \lambda^{n+1}\left\|\widehat{\mathbf{R}}^{-1}(n)\right\|\left\|\widehat{\mathbf{R}}(-1)\right\|\|\mathbf{w}(-1)\| + \left\|\widehat{\mathbf{R}}^{-1}(n)\right\|\sum_{\ell=0}^{n}\lambda^{n-\ell}|d^*(\ell)|\|\mathbf{u}(\ell)\|. \tag{35}$$

Using (26), (27), and (28) in (35), we obtain (29), which completes the proof. ■

Suppose that the algorithm leaves the ROI at the iteration n_0. While it remains outside the ROI, since $\bar{d}(n) = 0$, (35) reduces to

$$0 \leq \|\mathbf{w}(n)\| < \frac{\zeta}{\alpha}\lambda^{n-n_0+1}\|\mathbf{w}(n_0-1)\| < \infty, \quad n \geq n_0. \tag{36}$$

In this case, the upper bound of the norm of the coefficient vector of DM-SWA will decrease exponentially with time. This leads to a reduction of the norm of the coefficient vector after a sufficiently large number of iterations, forcing the algorithm to return to the ROI, as shown in the next theorem.

Theorem 2 *Under the persistence of excitation condition (Assumption A-1), if DM-SWA leaves the ROI at time instant n_0, it will return to it, **at most** after*

$$k_{\max} = \left\lceil \frac{1}{\ln(\lambda)} \ln\left(\frac{\alpha}{\zeta}\frac{\sqrt{\beta C^a_{1,1}}}{B_u\,\|\mathbf{w}(n_0-1)\|}\right)\right\rceil < \infty \tag{37}$$

iterations, where $\lceil\cdot\rceil$ is the ceiling function.

Proof: The absolute value of the equalizer output has the following upper bound

$$|y(n)| \leq \|\mathbf{w}(n-1)\|\|\mathbf{u}(n)\| \leq \|\mathbf{w}(n-1)\|B_u. \tag{38}$$

Since DM-SWA left the ROI at $n = n_0$, we can replace (36) in (38) to obtain

$$|y(n)| \leq B_u\frac{\zeta}{\alpha}\lambda^{n-n_0}\|\mathbf{w}(n_0-1)\|, \tag{39}$$

which is an upper bound for the equalizer output in this case.

The algorithm returns to the ROI at instant n_1, for which $|y(n_1)| \leq \sqrt{\beta C^a_{1,1}}$. From the upper-bound (39), we see that DM-SWA returns to the ROI at most at n_1 such that

$$B_u\frac{\zeta}{\alpha}\lambda^{n_1-n_0}\|\mathbf{w}(n_0-1)\| \leq \sqrt{\beta C^a_{1,1}}. \tag{40}$$

Solving (40) for n_1 and taking the next higher integer, we arrive at

$$n_1 \leq n_0 + k_{\max}, \tag{41}$$

where $k_{\max}$ is defined in (37), which completes the proof. ■

Note that the coefficient vector $\mathbf{w}(n)$ of DM-SWA could be updated in different manners outside the ROI, by choosing different values for $\bar{d}(n)$. However, the choice $\bar{d}(n) = 0$ leads to the minimum upper bound, given by (36). If we made $|\bar{d}(n)| > 0$, (35) would provide an upper bound higher than (36) and the algorithm could, in the worst case, spend more iterations to return to the ROI. Furthermore, it is important to emphasize that no *approximations* have been used to establish Theorems 1 and 2. Assuming the persistence of excitation condition, Theorem 1 provides a deterministic upper bound for the norm of the coefficient vector of DM-SWA, showing that the algorithm is stable independently of the operation mode. Theorem 2 shows that DM-SWA may stay outside the ROI only for a finite-time interval. However, the upper bounds in (29) and in (41) are very conservative due to A-1, mainly in the case of $\zeta \gg \alpha$.

We see from (39) that the rate of reduction of the worst-case bound (36) for $\|\mathbf{w}(n)\|$, when DM-SWA is outside the ROI, is λ^n. Using some simplifying assumptions (A-2 and A-3 below), we can show that this rate of reduction applies, on average, also to $\mathbf{w}(n)$, not only to the worst-case bound. It is important to notice that these assumptions are valid only in the steady-state. A-2 is a part of the widely used independence assumptions in adaptive filter theory, and A-3 is a reasonable steady-state assumption, mainly in the case of $\lambda \approx 1$ [27], [28].

A-2 *The coefficient vector* $\mathbf{w}(n-1)$ *is independent of* $\widehat{\mathbf{R}}^{-1}(n)\mathbf{u}(n)\mathbf{u}^H(n)$ *in the steady-state.*

A-3 *Using* (8), *we consider valid the approximation*

$$\mathrm{E}\{\widehat{\mathbf{R}}^{-1}(n)\mathbf{u}(n)\mathbf{u}^H(n)\} \approx (1-\lambda)\mathbf{I}. \tag{42}$$

Then, to obtain a model for the reduction of the coefficient vector in the mean, outside the ROI, we first rewrite (25) as

$$\mathbf{w}(n) = \left[\mathbf{I} - \widehat{\mathbf{R}}^{-1}(n)\mathbf{u}(n)\mathbf{u}^H(n)\right]\mathbf{w}(n-1). \tag{43}$$

Taking the expectations of both sides of (43), using A-2 and A-3, we obtain

$$\mathrm{E}\{\mathbf{w}(n)\} \approx \lambda \mathrm{E}\{\mathbf{w}(n-1)\}. \tag{44}$$

This approximation shows that the mean of the coefficient vector of DM-SWA decreases exponentially with time outside the ROI, with rate λ^n.

When the persistence excitation condition is satisfied and DM-SWA is implemented in infinite-precision arithmetic, Theorem 1 establishes that the equalizer coefficients are bounded. However, if DM-SWA is implemented in finite precision, it can diverge due to numerical problems. Moreover, its computational cost increases linearly with M^2, which is a disadvantage when compared to fast algorithms. To solve these problems, in the next section, we obtain a fast version of DM-SWA. The new algorithm is numerically well-behaved even when implemented in finite-precision arithmetic and has a computational cost which increases linearly with M. Thus, it avoids the two causes of divergence of SWA and has the advantage of being fast.

IV. A DUAL-MODE LATTICE SWA

It is well-known in the literature that one of the problems of the conventional RLS algorithm is its numerical instability, which can arise because of the loss of the numerical compatibility in the update of $\widehat{\mathbf{R}}^{-1}(n)$ [5]. Through changes of coordinates, the state-space representation of the conventional RLS algorithm can be transformed to an unlimited number of systems with the same input-output relation, and hence solving the same least-squares problem. Although all these realizations are equivalent in infinite precision, the numerical behavior will vary from one coordinate system to another [9]. Thus, the numerical instability of the conventional RLS can be avoided by choosing a convenient transformation on the state-space representation.

Since (22) is equivalent to the state-space representation of the conventional RLS, the changes of coordinates usually applied to obtain numerically stable versions of RLS can also be used to obtain numerically stable versions of SWA and DM-SWA. Since a coordinate system based on the Cholesky decomposition of $\widehat{\mathbf{R}}^{-1}(n)$ leads to a fast and stable version of RLS [9], we chose it to obtain a fast and stable version of DM-SWA. Thus, we use the following transformation matrix

$$\mathbf{L}(n) = \begin{bmatrix} \mathbf{K}^{-H}(n) & 0 \\ \mathbf{0}^T & 1 \end{bmatrix}.$$

The matrix $\mathbf{K}(n)$ appears in the Cholesky factorization of $\widehat{\mathbf{R}}^{-1}(n)$ and is a lower triangular matrix with ones along its main diagonal and zeros above the main diagonal; the nonzero elements of each row, except for complex conjugation, are equal to the weights of a backward prediction-error filter whose order corresponds to the position of that row in the matrix [5]. Defining $\mathbf{v}(n) \triangleq \mathbf{K}^{-H}(n)\mathbf{w}(n)$, (22) can be rewritten as

$$\begin{bmatrix} \mathbf{v}(n) \\ d^*(n) - y^*(n) \end{bmatrix} = \boldsymbol{\Upsilon}(n) \begin{bmatrix} \lambda \mathbf{v}(n-1) \\ d^*(n) \end{bmatrix}, \tag{45}$$

where $\boldsymbol{\Upsilon}(n) = \mathbf{L}(n)\boldsymbol{\Gamma}(n)\mathbf{L}^{-1}(n-1)$.

The realization (45) is equivalent to (22) and can be used to implement DM-SWA using a lattice structure. The resulting algorithm is named dual-mode lattice SWA (DM-LSWA). Each lattice stage provides prediction errors in its output [5]. The literature contains different versions of algorithms to obtain prediction errors from the observed sequence $\{u(n)\}$. The modified EF-LSL presents reliable numerical properties, even in the absence of persistent excitation and when implemented in finite-precision arithmetic [10].

Thus, DM-LSWA, summarized in Table II, uses the modified EF-LSL algorithm of [10] for the prediction section. The variables $(E_i^f(n),\ \vartheta_i,\ k_i^f(n))$ and $(E_i^b(n),\ \psi_i(n),\ k_i^b(n))$ represent respectively, energies, *a priori* prediction errors and reflection coefficients of the forward and backward predictions. The conversion factors are the $\gamma_i(n)$. The variables $(b, \bar{b}, f, \bar{f})$ were introduced to reduce the computational complexity of the algorithm [10]. To ensure robust numerical behavior in the prediction section, it is necessary to avoid divisions by values close to zero in their computations. To this end, we add a small positive constant ϵ to the denominators, whose value depends on the implementation precision. In general, $\epsilon = 2^{k-b}$ should be employed for input signals satisfying $-2^{k/2} \leq u(n) \leq 2^{k/2}$, with k being an integer and b the mantissa wordlength to which the energies are quantized [10].

For the joint estimation section, $y(n) = \mathbf{w}^H(n-1)\mathbf{u}(n)$ can be rewritten as $y(n) = \mathbf{v}^H(n-1)\boldsymbol{\psi}(n)$, where $\boldsymbol{\psi}(n) = \mathbf{K}(n-1)\mathbf{u}(n)$ is the *a priori* backward prediction error vector. The estimation errors ξ_i, $i = 1, 2, \cdots, M-1$ are obtained from the backward prediction errors and the coefficients $v_i(n-1)$. The zero-order estimation error is $\xi_0 = d(n)$.

The variables, which are initialized with non null values, are listed at the top of Table II. Using the same initialization for the vector $\mathbf{v}(-1)$ as for $\mathbf{w}(-1)$ (the center-tap initialization method), and choosing the initial energies $E_i^f(-1)$ and $E_i^b(-1)$, $i = 0, \ldots, M-1$, equal to the same small positive constant δ as for $\mathbf{P}(-1)$, DM-LSWA and DM-SWA will present close performance in infinite-precision arithmetic. The computational cost per iteration of DM-LSWA is shown in Table III for real and complex-valued data, and considering the number of real multiplications, real additions, real divisions, and comparisons ($\mathcal{C}$). As its computational cost increases linearly with M, DM-LSWA can be interpreted as a fast version of DM-SWA. Furthermore, as the modified EF-LSL algorithm of [10], DM-LSWA has an inherent parallelism that can be advantageously exploited for fast implementations.

TABLE II

SUMMARY OF DM-LSWA

Initialization:
$\mathbf{v}(-1) = [0 \ \cdots \ 0 \ 1 \ 0 \ \cdots 0]^T$
$E_i^f(-1) = E_i^b(-1) = \delta, \ i = 0, \ldots, M-1$

for $n = 0, 1, 2, 3, \ldots$ do:
$\vartheta_0 = \psi_0(n) = u(n)$
$\xi_0 = \bar{d}(n-1)$
$\gamma_0 = 1$
for $i = 0 \colon M-1$,
$b = \psi_i(n-1)\ \gamma_i$
$f = \vartheta_i\ \gamma_i$
$E_i^b(n-1) = \lambda E_i^b(n-2) + b\,\psi_i^*(n-1)$
$E_i^f(n) = \lambda E_i^f(n-1) + f\,\vartheta_i^*$
$\bar{b} = b/(\epsilon + E_i^b(n-1))$
$\bar{f} = f/(\epsilon + E_i^f(n))$
$\gamma_{i+1} = \gamma_i - \bar{b}\ b^*$
Lattice:
$\psi_{i+1}(n) = \psi_i(n-1) - k_i^{b*}(n-1)\vartheta_i$
$\vartheta_{i+1} = \vartheta_i - k_i^{f*}(n-1)\psi_i(n-1)$
$k_i^b(n) = k_i^b(n-1) + \bar{f}\,\psi_{i+1}^*(n)$
$k_i^f(n) = k_i^f(n-1) + \bar{b}\,\vartheta_{i+1}^*$
Joint estimation:
$\xi_{i+1} = \xi_i - \psi_i(n-1)\ v_i^*(n-1)$
$v_i(n-1) = v_i(n-2) + \bar{b}\ \xi_{i+1}^*$
end
$y(n) = \mathbf{v}^H(n-1)\boldsymbol{\psi}(n)$
$x(n) = (|y(n)|^2 - \beta C_{1,1}^a)/(r - \beta C_{1,1}^a)$
if $x(n) \geq 0$
$\bar{d}(n) = x(n)\ y(n)$
else
$\bar{d}(n) = 0$
end
end

TABLE III

COMPUTATIONAL COST PER ITERATION OF DM-SWA AND DM-LSWA.

Real-valued data				
Algorithm	$\times$	$+$	$\div$	$\mathcal{C}$
DM-SWA	M^2+5M+4	M^2+3M+1	1	1
DM-LSWA	$14M+3$	$12M$	$2M$	1
Complex-valued data				
Algorithm	$\times$	$+$	$\div$	$\mathcal{C}$
DM-SWA	$4M^2+16M+8$	$4M^2+12M+4$	1	1
DM-LSWA	$46M+7$	$39M+1$	$4M$	1

V. SIMULATION RESULTS

To verify the influence of the arithmetic precision, we implemented the algorithms in floating point, in Matlab with 64 or 32 bits. The precision is indicated as subscripts (e.g., SWA_{32}). We label a given run of the algorithms as "diverging" if $\bar{e}(n)$ overflows (we check for NaNs). For a given algorithm, the mean-squared error (MSE) is estimated from an ensemble-average of 10^3 independent runs of $\bar{e}(n)$. On the other hand, the squared error (SE) corresponds to only one run of $\bar{e}(n)$ filtered by a moving-average filter. For complex data, since constant-modulus-based algorithms are insensible to random phase rotations, we include a phase correction algorithm. For simplicity, we use the phase tracking algorithm [29], [30], which provides the following phase update equation

$$\varphi(n+1) = \varphi(n) + \mu_p \text{Im}\{\bar{y}(n)\hat{a}^*(n-\tau_d)\}, \tag{46}$$

where $\bar{y}(n) = y(n)e^{-j\varphi(n)}$, μ_p is the step-size, and $\text{Im}\{\cdot\}$ stands for the imaginary part of a complex number.

Fig. 2 shows the MSE curves for SWA_{64}, DM-SWA_{32}, and DM-LSWA_{32}, considering the transmission of binary signals (2-PAM - pulse amplitude modulation) through the non-minimum phase channel $H(z) = 0.1 + z^{-1} + 0.1z^{-2}$ with a signal-to-noise ratio (SNR) of 50 dB. The equalizer has $M = 11$ coefficients initialized with only one non-null element (equal to one) in the sixth position. The channel model and the length of the equalizer were obtained from the computer experiment of [5, p. 455] with $W = 2.5152$. Although $H(z)$ is a non-minimum phase channel, its amplitude distortion is not large. In Figures 2-(a), 2-(b), and 2-(c), we vary the value of the forgetting factor and add a constant at a specific sample of the input signal $u(n)$. This added constant causes divergence in SWA_{64} and makes DM-SWA_{32} and DM-LSWA_{32} leave the ROI. However, as established by Theorem 2 and Eq. (39), DM-SWA_{32} and DM-LSWA_{32} return to the ROI with a rate that depends on the forgetting factor. Since there is no divergence due to numerical problems, these algorithms present the same performance. In Fig. 2-(d), we show a histogram of the MSE for the final 10^3 iterations of DM-SWA_{32} and DM-LSWA_{32} for the case of Fig. 2-(a). One can observe that, after the perturbation at $n = 2500$, the algorithms may converge back to different minima: DM-SWA_{32} and DM-LSWA_{32} reached the minimum with $\text{MSE} \approx -41$ dB in 767 of the 1000 independent runs, the minimum with $\text{MSE} \approx -31$ dB in 123 of the 1000 runs, and the minimum with $\text{MSE} \approx -12$ dB in 110 of the 1000 runs. This behavior leads to the steady-state MSE of approximately -20 dB after the perturbation, as observed in Fig. 2-(a). Note that the perturbation is equivalent to a new initialization for DM-SWA_{32} and DM-LSWA_{32}. This convergence to different minima is a common problem of constant-modulus-based algorithms, and occurred only for $\lambda = 0.9$. For $\lambda \geq 0.99$, the new algorithms always returned to the same steady-state performance of the ensemble-average of SWA_{64} before the perturbation. Thus, besides avoiding divergence, (24) does not cause meaningful changes in the performance of the algorithm since a quick return to the region of interest was always observed.

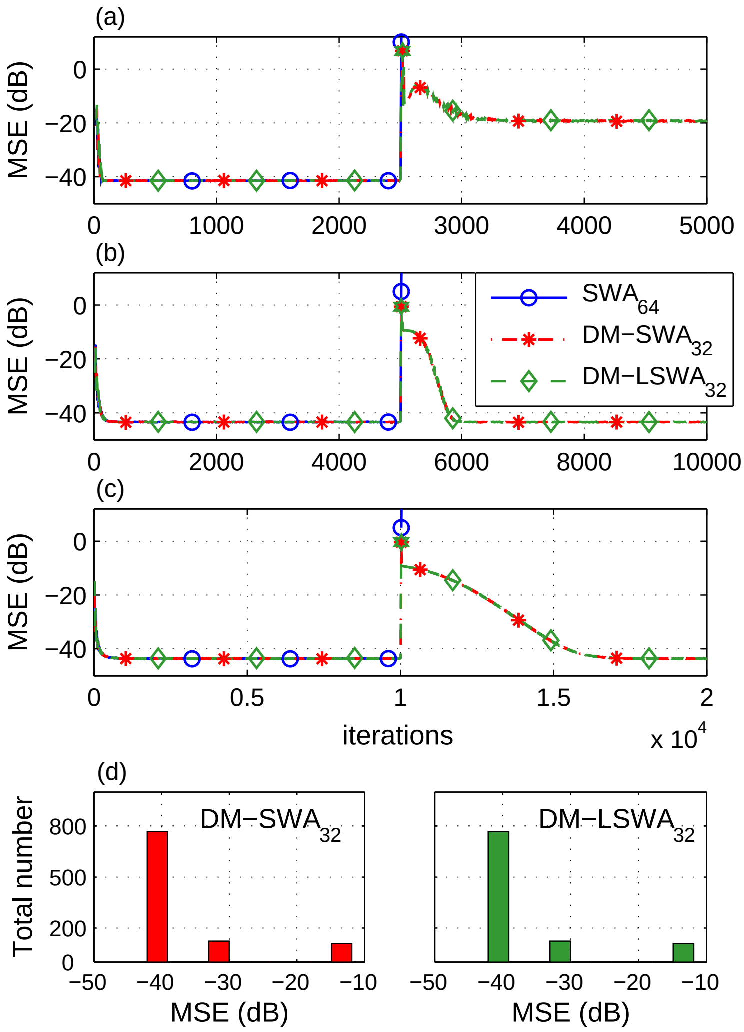

Fig. 2. MSE for SWA_{64}, DM-SWA_{32}, and DM-LSWA_{32}, assuming (a) $\lambda = 0.9$, $u(2500) = 10$; (b) $\lambda = 0.99$, $u(5000) = 10$; (c) $\lambda = 0.999$, $u(10000) = 20$; (d) Histogram of MSE for the final 10^3 iterations of DM-SWA_{32} and DM-LSWA_{32} for the case (a); $\delta = 1$, $\epsilon = 1.2 \times 10^{-5}$; mean of 10^3 independent runs; $H(z) = 0.1 + z^{-1} + 0.1z^{-2}$, SNR=50 dB; $M = 11$, 2-PAM .

Fig. 3 shows a histogram of the divergence time for 10^3 independent runs of SWA, DM-SWA, LSWA[2], and DM-LSWA, all implemented with 64 bits of arithmetic precision. We assume the transmission of 4-QAM (quadrature amplitude modulation) signals through the channel $H(z) = 0.1 + z^{-1} + 0.1z^{-2}$ with $\text{SNR} = 30$ dB. Again, the equalizer has $M = 11$ coefficients initialized with only one non-null element in the sixth position, but the non-null element is equal to 1.6. This simulation scenario was carefully

[2]We call LSWA the algorithm of Table II without the conditional statement (if), i.e., LSWA always uses $\bar{d}(n) = d(n) = x(n)y(n)$ independently of the sign of $x(n)$.

chosen to highlight the two causes of divergence of SWA. The divergence due to the inconsistency in the nonlinear estimate of the transmitted signal is more likely to occur at the first iterations, when the algorithm is still far from a minimum. This cause of divergence leads SWA and LSWA to diverge in 594 of the 1000 independent runs, but it is avoided by DM-SWA and DM-LSWA through the use of (24). The divergence due to the loss of the numerical compatibility in the update of $\widehat{\mathbf{R}}^{-1}(n)$ occurs around $n = 3 \times 10^4$. This numerical problem causes divergence in SWA and DM-SWA. When SWA "survives" to the divergence due to the nonlinearity, it diverges due to the numerical problem, which happens in 406 of the 1000 runs. DM-SWA also diverges in all the 1000 independent runs but only due to the loss of the numerical compatibility. DM-LSWA never diverges since it combines the numerical robustness of LSWA with the capability of DM-SWA of rejecting inconsistent estimates of the transmitted signal. This behavior also occurs when the algorithms are implemented with 32 bits of arithmetic precision. The only difference is that the divergences due to the numerical problem occur earlier, around $n = 10^4$.

Again, Fig. 4 shows a histogram of the divergence time for 10^3 independent runs of SWA, DM-SWA, LSWA, and DM-LSWA, but now implemented with 32 bits. We also choose a different simulation scenario: transmission of non-constant modulus signals (16-QAM) through the non-minimum phase channel $H(z) = (0.37 - j0.06) + (0.47 + j0.70)z^{-1} + (0.37 - j0.06)z^{-2}$ in the absence of noise. The equalizer has $M = 23$ coefficients initialized with only one non-null element (equal to 0.05) in the 12th position. The two causes of divergence of SWA are again highlighted. Only DM-LSWA avoids both causes of divergence. Due to the lower-precision arithmetic, the divergence due to the numerical problem occurs around $n = 8 \times 10^3$. Using 64 bits, this kind of divergence occurs later, around $n = 2.8 \times 10^4$.

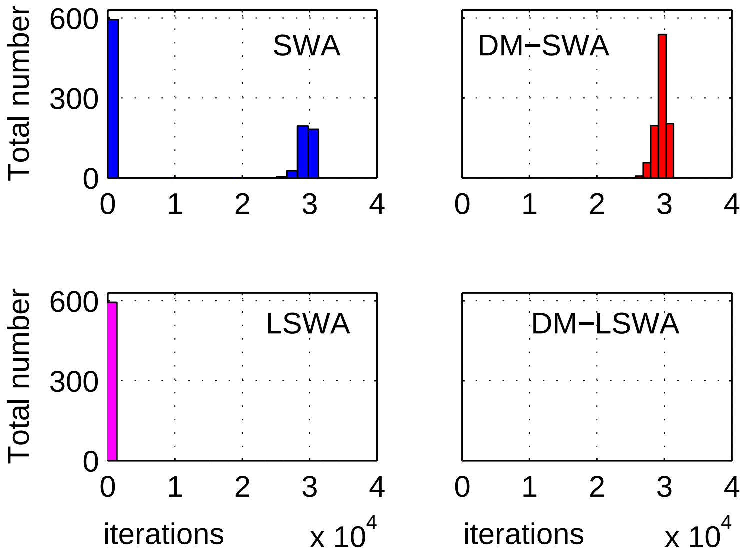

Fig. 3. Histogram of the divergence time for 10^3 independent runs of SWA, DM-SWA, LSWA, and DM-LSWA; 64 bits, $\lambda = 0.999$, $\delta = 1$, $\epsilon = 6.5 \times 10^{-16}$, $\mu_p = 10^{-3}$; $M = 11$; 4-QAM; $H(z) = 0.1 + z^{-1} + 0.1z^{-2}$, SNR=30 dB.

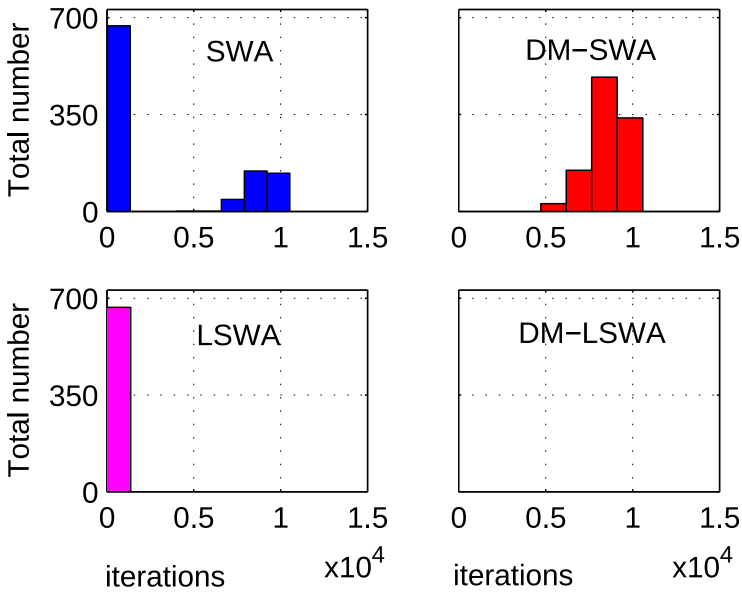

Fig. 4. Histogram of the divergence time for 10^3 independent runs of SWA, DM-SWA, LSWA, and DM-LSWA; 32 bits, $\lambda = 0.999$, $\delta = 5$, $\epsilon = 5.2 \times 10^{-6}$, $\mu_p = 10^{-3}$; $M = 23$; 16-QAM; $H(z) = (0.37 - j0.06) + (0.47 + j0.70)z^{-1} + (0.37 - j0.06)z^{-2}$, absence of noise.

In Fig. 5, we show the squared error for one run of the SWA and DM-SWA, implemented with 64 and 32 bits, and DM-LSWA, implemented with 32 bits. We assume the transmission of 4-QAM signals through the channel $H(z) = 0.1 + z^{-1} + 0.1z^{-2}$ with $\text{SNR} = 50$ dB. The equalizer has $M = 11$ coefficients initialized with only one non-null element (equal to 1) in the sixth position. This simulation scenario was chosen to avoid the divergence at the first iterations due to the nonlinearity and to emphasize the divergence due to the numerical problem. To facilitate the visualization, the SE curves were filtered by a moving-average filter with 1024 coefficients. We can observe that SWA$_{64}$ and DM-SWA$_{64}$ present the same numerical behavior, diverging around $n = 3 \times 10^4$. The same occurs with SWA$_{32}$ and DM-SWA$_{32}$, but the divergence happens around $n = 10^4$. DM-LSWA$_{32}$ does not diverge and maintains its numerical robustness, even implemented with 32 bits.

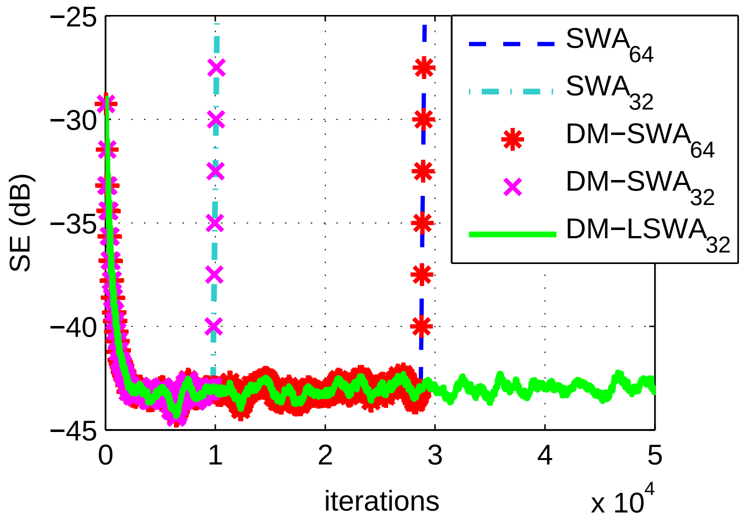

Fig. 5. SE for SWA (64 and 32 bits), DM-SWA (64 and 32 bits), and DM-LSWA (32 bits); $\lambda = 0.999$, $\delta = 1$, $\epsilon_{64} = 6.5 \times 10^{-16}$, $\epsilon_{32} = 3.5 \times 10^{-7}$, $\mu_p = 10^{-3}$; $M = 11$; 4-QAM; $H(z) = 0.1 + z^{-1} + 0.1z^{-2}$, SNR=50 dB.

Fig. 6 shows simulation results considering the transmission of a binary signal (2-PAM) through a linear and time-variant channel $H(z, n) = h_0(n) + h_1(n)z^{-1} + h_2(n)z^{-2}$, with $h_0^2(n) + h_1^2(n) + h_2^2(n) = 1$. We assume a Rayleigh fading channel with fast variation (maximum Doppler spread $f_D = 80$ Hz) and $\mathrm{SNR} = 25$ dB [6, p. 401]. The absolute values of the roots of $h_0(n)z^2 + h_1(n)z + h_2(n)$ are shown in Fig. 6-(a) so that error bursts can be associated with rapid changes of these roots or deep spectral nulls, which occur when both roots are on the unit circle (absolute value equal to one is indicated by a straight line). We compare the performance of SWA$_{64}$ with that of DM-LSWA$_{32}$. Fig. 6-(b) shows the squared error filtered by a moving-average filter with 512 coefficients for each algorithm. The equalizer outputs are shown in Figures 6-(c) and 6-(d) for SWA$_{64}$ and DM-LSWA$_{32}$, respectively. We can observe from the sign of $x(n)$, shown in Fig. 6-(e), that there are iterations in which DM-LSWA$_{32}$ leaves the ROI, but quickly returns to it. Generally, these iterations coincide with critical situations associated with error bursts. The figure also shows that, little before SWA$_{64}$ became unstable, DM-LSWA$_{32}$ makes $\bar{d}(n) = 0$, thereby avoiding divergence. Although the channel is time-variant, it has unit norm for all time instant, which is a favorable condition for SWA$_{64}$.

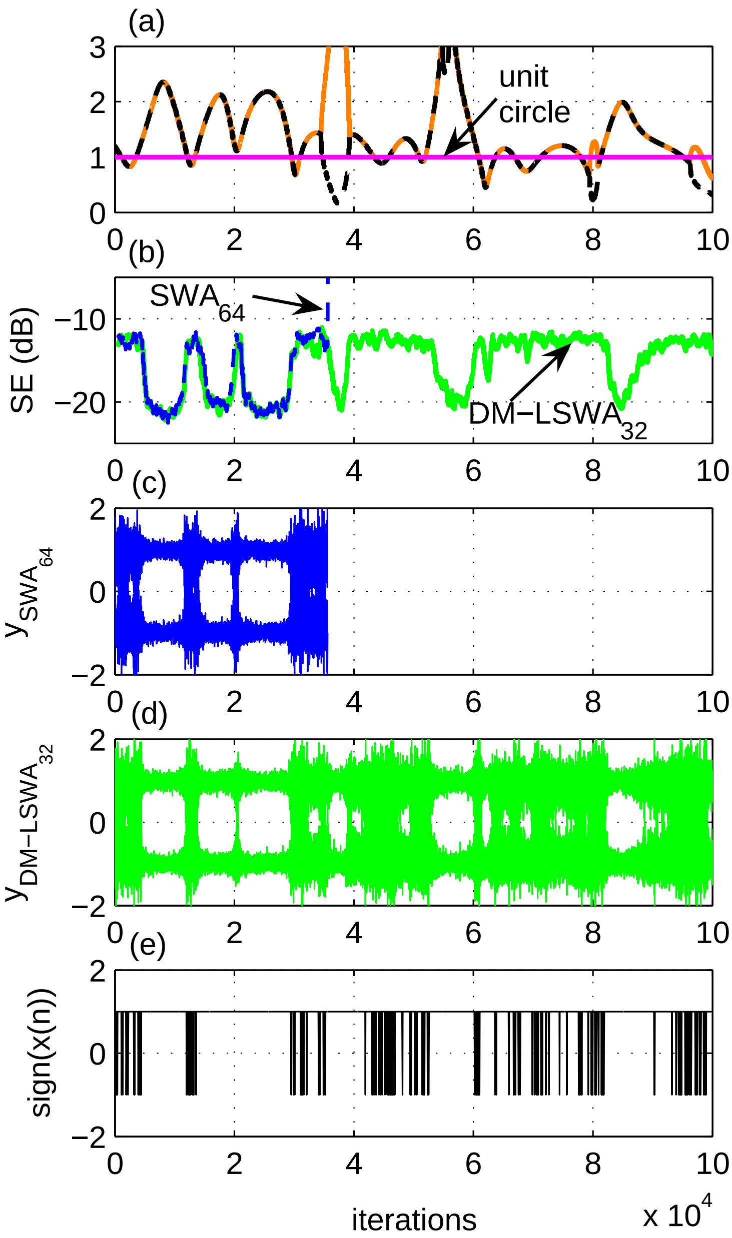

Fig. 6. (a) Absolute root values of a Rayleigh channel (3 coefficients, symbol period $T = 0.8\mu$s, maximum Doppler spread $f_D = 80$ Hz, $\text{SNR} = 25$ dB); (b) SE in dB for SWA$_{64}$ and DM-LSWA$_{32}$; Output of the equalizer for (c) SWA$_{64}$ and (d) DM-LSWA$_{32}$; (e) Sign of $x(n)$ for DM-LSWA$_{32}$; 2-PAM; $\lambda = 0.85$, $\delta = 0.1$, $\epsilon = 4.4 \times 10^{-7}$, $M = 11$.

The simulation results presented here are also valid for different channels, transmitted-signal constellations, and values of the forgetting factor. In all these situations, DM-LSWA remained stable and did not break down, although it did not present useful performance for low λ values due to large estimation errors.

VI. Conclusion

We show that the divergence of SWA can be caused by (i) inconsistency in the nonlinear estimate of the transmitted signal, or (ii) loss of numerical compatibility in the update of the inverse of the autocorrelation matrix. The divergence of SWA due to (ii) has the same origin as in the RLS algorithm. Thus, the proposed solutions in the literature to solve this problem in the conventional and fast RLS algorithms can be used directly to solve (ii). To avoid divergence due to (i), we proposed DM-SWA, which works as the conventional SWA in the first mode of operation and rejects non-consistent estimates of the transmitted signal in the second mode. Although DM-SWA does not diverge due to (i), it can still diverge due to (ii). Assuming the persistence of excitation condition, we proved, through a deterministic analysis, that DM-SWA is stable in infinite-precision arithmetic. Furthermore, if the algorithm leaves the first mode of operation, it will return to it in finite time. This property guarantees that the new algorithm is not trading off performance for robustness, as our simulations confirm. To solve (ii) and to obtain an algorithm with reduced computational cost, we proposed DM-LSWA, which avoids divergence even when implemented in finite precision, and maintains the convergence rate of DM-SWA. In spite of the lack of a proof for the numerical stability for the prediction section of DM-LSWA, the algorithm never diverges when implemented as in Section IV. It is relevant to notice that other methods for ensuring numerical stability in RLS algorithms could also be employed here. For example, we could use the QR-based methods, whose stability proofs are available in the literature. The procedure used to remedy the problem (i) can also be extended to other constant-modulus-based algorithms. Recently, we extended this idea in [31] to avoid divergence in a normalized version of CMA. However, in this case, the stability analysis of the algorithm is not a straightforward extension of that presented here.

References

[1] S. H. Qureshi, "Adaptive equalization," *Proceedings of the IEEE*, vol. 73, pp. 1349–1387, Sept. 1985.

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

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

[4] R. K. Martin and C. R. Johnson, "Adaptive equalization: transitioning from single-carrier to multicarrier systems," *IEEE Signal Processing Magazine*, vol. 22, pp. 108–122, Nov. 2005.

[5] S. Haykin, *Adaptive Filter Theory*, Prentice Hall, Upper Saddle River, 4th edition, 2001.

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

[7] P. A. Regalia and M. G. Bellanger, "On the duality between fast QR methods and lattice methods in least-squares adaptive filtering," *IEEE Transactions on Signal Processing*, vol. 39, pp. 879–891, Apr. 1991.

[8] F. Ling, D. Monolakis, and J. G. Proakis, "Numerically robust least-squares lattice-ladder algorithms with direct updating of the reflection coefficients," *IEEE Trans. on Acoutics, Speech and Signal Process.*, vol. 34, pp. 837–845, Aug. 1986.

[9] P. A. Regalia, "Numerical stability properties of a QR-based fast least squares algorithm," *IEEE Transactions on Signal Processing*, vol. 41, pp. 2096–2109, Jun. 1993.

[10] M. D. Miranda, M. Gerken, and M. T. M. Silva, "Efficient implementation of the a priori error-feedback LSL algorithm," *Electronics Letters*, vol. 35, pp. 1308–1309, Aug. 1999.

[11] D. N. Godard, "Self-recovering equalization and carrier tracking in two dimensional data communication system," *IEEE Transactions on Communications*, vol. 28, pp. 1867–1875, Nov. 1980.

[12] J. R. Treichler and B. Agee, "A new approach to multipath correction of constant modulus signals," *IEEE Transactions on Acoustics Speech and Signal Processing*, vol. ASSP-28, pp. 334–358, Apr. 1983.

[13] O. Shalvi and E. Weinstein, "Super-exponential methods for blind deconvolution," *IEEE Transactions on Information Theory*, vol. 39, pp. 504–519, Mar. 1993.

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

[15] M. T. M. Silva and M. D. Miranda, "Tracking issues of some blind equalization algorithms," *IEEE Signal Processing Letters*, vol. 11, pp. 760–763, Sept. 2004.

[16] C. B. Papadias and D. T. M. Slock, "Normalized sliding window constant modulus and decision-direct algorithms: a link between blind equalization and classical adaptive filtering," *IEEE Transactions on Signal Processing*, vol. 45, pp. 231–235, Jan. 1997.

[17] H. H. Zeng, L. Tong, and C. R. Johnson, "Relationships between the Constant Modulus and Wiener receivers," *IEEE Transactions on Information Theory*, vol. 44, pp. 1523–1538, Jul. 1998.

[18] P. A. Regalia and M. Mboup, "Undermodeled equalization: a characterization of stationary points for a family of blind criteria," *IEEE Transactions on Signal Processing*, vol. 47, pp. 760–770, Mar. 1999.

[19] J. Mai and A. H. Sayed, "A feedback approach to the steady-state performance of fractionally spaced blind adaptive equalizers," *IEEE Transactions on Signal Processing*, vol. 48, pp. 80–91, Jan. 2000.

[20] P. A Regalia, "A finite-interval constant modulus algorithm," in *Proc. of ICASSP'02*. IEEE, 2002, vol. III, pp. 2285–2288.

[21] O. Dabeer and E. Masry, "Convergence analysis of the Constant Modulus Algorithm," *IEEE Transactions on Information Theory*, vol. 49, pp. 1447–1464, Jun. 2003.

[22] 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.

[23] D. Dembélé and Favier G., "Recursive estimation of fourth-order cumulants with application to identification," *Signal Processing*, vol. 68, pp. 127–139, Mar. 1998.

[24] P. I. Hübscher, J. C. M. Bermudez, and V. H. Nascimento, "A mean-square stability analysis of the least mean fourth adaptive algorithm," *IEEE Transactions on Signal Processing*, vol. 55, pp. 4018–4028, Aug. 2007.

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

[26] S. Ljung and L. Ljung, "Error propagation properties of recursive least-squares adaptation algorithms," *Automatica*, vol. 21, pp. 157–167, 1985.

[27] B. Farhang-Boroujeny, *Adaptive Filters - Theory and Applications*, John Wiley & Sons, West Sussex, 1998.

[28] J. Minkoff, "Comment: on the unnecessary assumption of statistical independence between reference signal and filter weights in feedforward adaptive systems," *IEEE Transactions on Signal Processing*, vol. 49, pp. 1109, May 2001.

[29] J. Proakis, *Digital Communications*, McGraw-Hill, NY, 4th edition, 2001.

[30] L. L. Szczecinski and A. Gei, "Blind decision feedback equalisers, how to avoid degenerated solutions," *Signal Processing*, vol. 82, pp. 1675–1693, Nov. 2002.

[31] M. D. Miranda, M. T. M. Silva, and V. H. Nascimento, "Avoiding divergence in the Constant Modulus Algorithm," in *Proc. of ICASSP'08*. IEEE, 2008.