Second-order blind algorithms for adaptive recursive linear equalizers: Analysis and lattice implementation

Roberto López-Valcarce*

Departamento de Teoría de la Señal y las Comunicaciones, Universidad de Vigo, 36200 Vigo (Pontevedra), Spain

SUMMARY

Adaptive recursive linear equalizers present important advantages in terms of performance and robustness compared to more standard finite impulse response structures, and provide a means for blindly initializing the decision feedback structure. We present an analysis of a pair of algorithms for the adaptation of the recursive part of the equalizer, which are based on the second-order statistics of the received signal, in a multichannel complex-valued setting with spatially coloured noise. When the number of equalizer poles is no less than the channel order, both algorithms enjoy a unique stationary point, which in addition is locally convergent; global convergence properties, on the other hand, can be quite different. When the optimum setting presents poles close to the stability boundary, the lattice structure is preferred for ease of stability monitoring. Lattice versions of the two algorithms are developed and their convergence properties discussed. Copyright © 2006 John Wiley & Sons, Ltd.

KEY WORDS: channel equalization; adaptive IIR filters; adaptive lattice filters

1. INTRODUCTION

Digital data transmission is generally achieved via the modulation of pulses onto the amplitude and phase of a radiofrequency carrier. Due to the dispersive nature of most physical media, the signal observed at the receiver is affected both by noise and intersymbol interference (ISI). With the ever increasing data rates that modern applications demand, the ISI problem becomes more severe due to the reduction of the symbol period with respect to the channel delay spread. The optimum receiver (in terms of error rate) in the presence of ISI is a maximum likelihood sequence estimator [1]; however, its complexity grows exponentially with the channel memory, which makes it unfeasible in most applications. A cheaper alternative is the use of an equalizer followed by a symbol-by-symbol detector. A linear equalizer (LE) is a linear filter whose purpose is to invert the channel transfer function. LE performance may deteriorate considerably

*Correspondence to: R. López-Valcarce, Departamento de Teoría de la Señal y las Comunicaciones, Universidad de Vigo, 36200 Vigo (Pontevedra), Spain.
†E-mail: valcarce@gts.tsc.uvigo.es

Contract/grant sponsor: Spanish Ministry of Education and Science

Received 2 May 2005
Revised 11 November 2005
Accepted 27 December 2005

Copyright © 2006 John Wiley & Sons, Ltd.
if the channel presents deep spectral nulls, due to noise amplification. On the other hand, a decision feedback equalizer (DFE) provides postcursor ISI cancellation with reduced noise enhancement and offers better performance than the LE.

Equalizers are often adaptive in order to be able to cope with unknown channels. In some cases a pilot or training sequence is inserted within the modulated data, and the adaptive algorithm for the equalizer parameters may use this as a reference signal. Alternatively, equalizer adaptation can be performed without training sequences by attempting to restore some of the signal’s statistical properties. This approach is termed blind or unsupervised adaptive equalization, and is often desired for data rate efficiency reasons or in settings in which no training signal is available [2].

Regardless of the structure and adaptive strategy chosen, it is common for the equalizer area to consume a sizable portion of the total chip area in a digital receiver. Thus, alternative approaches to adaptive equalization that allow for a reduction in complexity are of clear interest. One such approach, which has received renewed attention over the past few years, is the use of linear infinite impulse response (IIR) filters for channel equalization. In addition to their potential to outperform more standard FIR structures for the same number of parameters, when blindly adapted IIR LEs may provide a convenient way to start up a DFE by just including the decision device within the feedback loop of the equalizer [3–6]. The use of IIR filters for equalization also improves the convergence properties of blind algorithms, both in terms of avoiding local minima [7] and speeding up convergence [8].

While the topic of adaptive FIR equalization is quite mature, the behaviour of adaptive IIR equalizers is not as well understood yet. Most of the literature on adaptive IIR filtering focuses on the system identification setting [9], in which the adaptive filter is placed in parallel with the unknown system. This is in contrast with the equalization setting, in which one has a series connection of the unknown system (the channel) and the adaptive filter. Under certain conditions, the recursive part of the equalizer can be thought of as a linear predictor [5, 6], so that its adaptation can be based on the minimization of the prediction error variance. Adaptive algorithms for recursive predictors [10, 11] can then be tailored to the equalization problem, as in References [3, 5]. One appealing feature of these schemes is that they operate in an unsupervised manner, and in contrast with more classical blind equalization methods, they are based on second-order statistics (SOS) of the received signal. Blind algorithms based on traditional higher-order criteria, such as the Constant Modulus Algorithm (CMA), may also be applied to the IIR architecture [3].

The goal of this paper is to analyse the behaviour of the IIR SOS-based equalization algorithms in terms of the properties of their stationary points such as existence, uniqueness, and local stability, as well as the relevance of the minimum prediction error variance criterion in the equalization context. With the application of digital communications in mind, we consider a setting in which complex-valued independent symbols are transmitted over a single input, multiple output channel with a complex-valued impulse response. The channel noise is assumed temporally white but may be spatially correlated. In principle, in a multichannel setting it may be possible to achieve perfect equalization using a bank of linear FIR LEs [12] (in the sense that the channel-equalizer combination can be reduced to a pure delay); however two requisites for this are that (i) the subchannels must not have any common roots, and (ii) the equalizer must be at least as long as the channel impulse response. In practice, FIR LEs are not necessarily robust to near violation of these so-called zero & length conditions [13], and therefore the use of IIR LEs in these situations is of interest; especially since the aforementioned strategy of using the IIR LE
as a convenient way to initialize the more powerful DFE in an unsupervised manner carries over to the multichannel case.

In a practical equalization problem, convergence time is also an important issue to consider; however, our main focus is on more fundamental properties of the algorithms such as stability, leaving convergence speed analysis as future work. Given the potential for instability of the adaptive IIR equalizers when implemented in direct form, we also present several approaches for their implementation using the normalized lattice structure. This architecture allows for efficient stability monitoring and is much more reliable in situations where the equalizer has to place poles near the unit circle.

The presentation is organized as follows. The channel equalization problem is presented in Section 2, and the structure of the minimum mean squared error equalizer is reviewed in Section 3. Section 4 discusses the feasibility of a minimum output energy criterion for the adaptation of the equalizer recursive part, and presents the two adaptive algorithms based on this approach. In Section 5 an analysis of these two adaptive algorithms is given, while their lattice implementations are presented in Section 6. Some simulation examples are shown in Section 7; finally, we present our conclusions in Section 8.

The notation adopted in the paper is as follows:

- \( E[\cdot] \) denotes statistical expectation.
- Scalars are denoted in lower case.
- Vectors are denoted in lower case bold.
- Matrices are denoted in upper case bold.
- \((\cdot)^*, (\cdot)^T\) and \((\cdot)^H\) denote, respectively, conjugation, transposition, and conjugate transposition.
- \( \bar{c}(z) \) denotes the ‘paraconjugate’ of the transfer function \( c(z) \), i.e. \( \bar{c}(z) = [c(1/z^*)]^H \).
- The expression \( c(z) s_n \) stands for \( \sum_{k=-\infty}^{\infty} c_k s_{n-k} \). With this notation, there is no need to distinguish between the complex variable and the unit advance operator, both denoted by \( z \).

2. PROBLEM SETTING

Our setting consists of a single input multiple output (SIMO) baseband equivalent model. The scalar symbol sequence \( \{s_n\} \), assumed to be zero mean white with unit variance \( \sigma^2 = 1 \), is transmitted over a linear time invariant channel with FIR impulse response of order \( L \). The \( p \times 1 \) vector-valued channel transfer function is then

\[
e(z) = c_0 + c_1 z^{-1} + \cdots + c_L z^{-L}
\]

This \( p \)-channel model accommodates communication systems with antenna arrays at the receiver, as well as oversampled received signals (fractionally spaced equalizers). The received signal is the \( p \times 1 \) vector

\[
u_n = c(z) s_n + w_n = \sum_{k=0}^{L} c_k s_{n-k} + w_n
\]

where \( \{w_n\} \) is the additive noise which is assumed to be zero mean, independent of \( \{s_n\} \), and temporally white with \( E[w_n w_{n-k}^H] = R \) for \( k = 0 \) and \( 0 \) otherwise. We allow for spatial noise.
correlation in the channel model, since this is the case when the multichannel configuration arises from oversampling the received continuous-time signal [14]. Observe that under these hypotheses the received signal $\{u_n\}$ is a $p$-variate Moving Average process of order $L$, denoted $\text{MA}(L)$. Symbols, channel coefficients and noise are complex-valued in general.

The equalizer is a linear IIR filter with transfer function $b(z)/[1 + a(z)]$, with $b(z)$ a $p$-input, 1-output polynomial of degree $N$ and $1 + a(z)$ a monic, scalar polynomial of degree $M$

$$b(z) = b_0^H + b_1^H z^{-1} + \cdots + b_N^H z^{-N}$$

$$1 + a(z) = 1 + a_1^* z^{-1} + \cdots + a_M^* z^{-M}$$

Figure 1 illustrates the equalization problem. Under the minimum mean squared error (MMSE) criterion, the equalizer should be selected in order to minimize the variance of the error $e_n = s_{n-\delta} - y_n$, where $y_n$ is the equalizer output and $\delta \geq 0$ is a suitable delay. As discussed in Reference [4], the optimal realizable (i.e. causal, stable and of finite degree) equalizer for this problem has an IIR transfer function with a numerator of order $\delta$ (the system delay) and a denominator of order $L$ (the channel order). This motivates the use of adaptive IIR equalizers: if the number of feedback coefficients $M$ is no less than the channel order, then optimum performance for the delay $\delta = N$ can be attained in principle.

3. MMSE EQUALIZER

In order to obtain the optimum numerator parameters, let us introduce the process

$$x_n = \frac{1}{1 + a(z)} u_n$$

and the $(N + 1)p \times 1$ vectors

$$x_n = [x_n^T \ x_{n-1}^T \ \cdots \ x_{n-N}^T]^T$$

$$b = [b_0^T \ b_1^T \ \cdots \ b_N^T]^T$$

Then the error can be written as $e_n = s_{n-\delta} - b^H x_n$, so that the optimal value of $b$ is

$$\bar{b}_n = Q_N^{-1} p_\delta, \quad \text{with} \quad Q_N = \mathbb{E}[x_n x_n^H], \quad p_\delta = \mathbb{E}[x_n s_{n-\delta}^*]$$
Hence, assuming that $b(z)$ is optimized as a function of $1 + a(z)$, we obtain a ‘reduced error surface’ which is a function of $\delta$ and the coefficients $a = [a_1 \cdots a_M]^T$ of the recursive part of the equalizer

$$J_{\text{red}}(a) = E[|e|^2]_{b=\hat{b}} = 1 - p_\delta^H Q_N^{-1} p_\delta$$

(8)

$J_{\text{red}}(a)$ is a non-quadratic function since both $Q_N$ and $p_\delta$ are non-linear functions of $a$. As a consequence of this, there is no closed form solution in general for the minimization of $J_{\text{red}}$ in terms of $a$. A notable exception is found in the case in which the delay is selected to be no larger than the numerator order, $\delta \leq N$, and the feedback section satisfies $M \geq L$. In that case, it follows from Reference [4] that the optimum denominator $1 + a_*(z)$ is given by the monic minimum phase (mmp) spectral factor of the received signal power spectral density:

$$\Phi_{uu}(z) = \sum_{k=-\infty}^{\infty} E[u_n u_{n-k}^*] z^{-k} = R + c(z)\tilde{c}(z)$$

(9)

From Reference [4 Equation (12)], one has

$$\det \Phi_{uu}(z) = [1 + \tilde{c}(z) R^{-1} c(z)] \det R$$

(10)

so that $1 + a_*(z)$ is the mmp spectral factor of $1 + \tilde{c}(z) R^{-1} c(z)$. Several observations follow from this fact:

1. As long as $\delta \leq N$ and $M \geq L$ are satisfied, the optimal denominator $1 + a_*(z)$ is independent of the delay $\delta$.
2. In addition, $1 + a_*(z)$ depends only on the received signal $\{u_n\}$, and not (explicitly) on the ‘training’ signal $s_{n-\delta}$. This suggests that blind determination of the optimum denominator should be feasible.
3. Moreover, $1 + a_*(z)$ is completely determined by the second-order statistics (SOS) of the received signal $\{u_n\}$. Hence, blind adaptive criteria based on second-order information may suffice. This is an appealing feature since higher-order equalization methods tend to require a larger number of data points to achieve the same performance level.
4. The adaptation of $1/[1 + a(z)]$ may be decoupled from that of $b(z)$. (If done blindly, the latter must exploit higher order statistical properties of the transmitted symbols.)

In view of these, we will focus on blind SOS-based methods for the adaptation of $1/[1 + a(z)]$. Assuming that $1/[1 + a(z)]$ is placed ahead of $b(z)$ in the receiver chain, adaptation of $b(z)$ can be done using standard blind adaptive schemes such as the Constant Modulus Algorithm (CMA), as in References [5, 7]. We must note that $1/[1 + a(z)]$ could be adapted under CMA as well, as in References [15, 16]. We will not pursue this approach here; the interested reader is referred to References [3, 16, 17] for its analysis.

4. MINIMUM OUTPUT ENERGY APPROACH

4.1. Cost function

Labat et al. [5, 6] proposed an SOS-based criterion for the adaptation of $1 + a(z)$: the minimization of the total energy at the output of the recursive part

$$J_{\text{OE}}(a) = E[x_n^H x_n]$$

(11)
where \( \{x_n\} \) was defined in (4). We refer to this approach as the Minimum Output Energy (MOE) criterion.

Denoting by \( [\Phi_{uu}(z)]_{ij} \) the \((i,j)\) element of the matrix \( \Phi_{uu}(z) \), one has

\[
J_{\text{OE}}(a) = \sum_{k=1}^{p} \frac{1}{2\pi} \int_{-\pi}^{\pi} [\Phi_{uu}(e^{j\omega})]_{kk} d\omega = \frac{1}{2\pi} \int_{-\pi}^{\pi} \text{tr} \Phi_{uu}(e^{j\omega}) d\omega
\]

(12)

Note from (9) that

\[
\text{tr} \Phi_{uu}(z) = \text{tr} R + \tilde{c}(z)c(z)
\]

(13)

which contains only powers of \( z \) ranging from \( z^{-L} \) to \( z^L \). Hence, if \( M \geq L \), (12) is minimized by setting \( 1 + a(z) \) equal to the mmp spectral factor of \( \text{tr} \Phi_{uu}(z) \). By comparing (13) and (10), it is seen that the mmp spectral factors of \( \text{tr} \Phi_{uu}(z) \) and \( \Phi_{uu}(z) \) are in general different. However, if the noise is spatially white with equal variance across the subchannels, then \( R = \sigma^2 I_p \) and

\[
\det \Phi_{uu}(z) = (\sigma^2)^{p-1}[(\sigma^2 + \tilde{c}(z)c(z)) \text{tr} \Phi_{uu}(z)]
\]

(14)

so that their two mmp spectral factors should be close to each other (in fact, they coincide in the single channel case), as recognized by Labat and Laot [6]. On the other hand, they may be far apart for spatially coloured noise.

**Example 1**

With \( p = 2 \), suppose that

\[
c(z) = \begin{bmatrix} 1 \\ -\rho z^{-L} \end{bmatrix}, \quad R = \sigma^2 \begin{bmatrix} 1 & \rho \\ \rho^* & 1 \end{bmatrix}
\]

It is readily checked that in that case,

\[
\det \Phi_{uu}(z) = \sigma^2 [z^* \rho^* z^{-L} + |z|^2 + \sigma^2 (1 - |\rho|^2)] + \rho z^{-L}
\]

\[
\text{tr} \Phi_{uu}(z) = 1 + |z|^2 + 2\sigma^2
\]

Since \( \text{tr} \Phi_{uu}(z) \) is constant, \( J_{\text{OE}} \) is minimized if \( a(z) = 0 \), i.e. all the poles are placed at the origin. However, the mmp spectral factor of \( \det \Phi_{uu}(z) \) may have poles of considerable magnitude depending on the value of \( \rho z \).

This problem can be sidestepped if the noise spatial covariance matrix is known at the receiver. Let \( L^{-1} \) be the Cholesky factor of \( R \), i.e. \( R = L^{-1}L^{-H} \). The received signal \( u_n \) can be preprocessed by the lower triangular matrix \( L \) before entering the recursive filter \( 1/[1 + a(z)] \), to obtain \( v_n = Lu_n \). Thus, the process \( x_n \) is redefined as

\[
x_n = \frac{1}{1 + a(z)}[Lu_n] = \frac{1}{1 + a(z)}v_n
\]

(15)

With this, the cost \( J_{\text{OE}} \) in (11) takes the form

\[
J_{\text{OE}}(a) = \frac{1}{2\pi} \int_{-\pi}^{\pi} \text{tr}[L\Phi_{uu}(e^{j\omega})L^H] d\omega
\]

(16)
In order to adaptively minimize the cost, the single-channel case can be generalized to the $p$-channel setting. We focus now on constant-gain algorithms for the adaptation of the recursive filter $1/[1 + a(z)]$.

4.2. Algorithms

We focus now on constant-gain algorithms for the adaptation of the recursive filter $1/[1 + a(z)]$. In order to adaptively minimize the cost $J_{OE}$ in (11), let $v_n^{(k)}$, $x_n^{(k)}$ denote the $k$th components of $v_n$, $x_n$, respectively, so that the input–output mapping of the filter $1/[1 + a(z)]$ reads

$$x_n^{(k)} = v_n^{(k)} - \sum_{j=1}^{M} a_j^* x_{n-j}^{(k)}, \quad 1 \leq k \leq p \quad \iff \quad x_n = v_n - X_n^T a^*$$

(18)

where the $M \times p$ matrix $X_n$ is given by

$$X_n = [x_{n-1} \ x_{n-2} \ \cdots \ x_{n-M}]^T$$

(19)

Writing $J_{OE} = \sum_{k=1}^{p} E[|x_n^{(k)}|^2]$, the stochastic gradient descent presented in Reference [3] for the single-channel case can be generalized to the $p$-channel setting. The derivative signals $\partial x_n^{(k)}/\partial a_i^*$ are given by

$$\frac{\partial x_n^{(k)}}{\partial a_i^*} = -\nabla_n^{(k)} \quad \text{with} \quad \nabla_n^{(k)} = \frac{1}{1 + a(z)} x_n^{(k)}, \quad 1 \leq k \leq p$$

(20)

Therefore, upon defining

$$d_n = [\nabla_n^{(1)} \ \nabla_n^{(2)} \ \cdots \ \nabla_n^{(p)}]^T, \quad D_n = [d_{n-1} \ d_{n-2} \ \cdots \ d_{n-M}]^T$$

(21)

the MOE adaptive algorithm can be written as follows, with $\mu > 0$ a stepsize:

$$a_{n+1} = a_n + \mu D_n x_n^*$$

(22)

Note that the implementation of the MOE algorithm (22) (also known as Recursive Maximum Likelihood [11] or Recursive LMS [18]) requires a total of $2p$ copies of the scalar filter.

---

These derivative signals are well defined whenever the Cauchy–Riemann conditions

$$\frac{\partial \text{Re} x_n^{(k)}}{\partial \text{Re} a_i} = \frac{\partial \text{Im} x_n^{(k)}}{\partial \text{Im} a_i}, \quad \frac{\partial \text{Im} x_n^{(k)}}{\partial \text{Re} a_i} = \frac{\partial \text{Re} x_n^{(k)}}{\partial \text{Im} a_i}$$

hold, as it is the case. Then

$$\frac{\partial^2 x_n^{(k)}}{\partial a_i^2} = \frac{\partial \text{Re} x_n^{(k)}}{\partial \text{Re} a_i} + j \frac{\partial \text{Im} x_n^{(k)}}{\partial \text{Re} a_i}, \quad \frac{\partial^2 x_n^{(k)}}{\partial a_i^2} = \frac{\partial \text{Im} x_n^{(k)}}{\partial \text{Re} a_i} + j \frac{\partial \text{Re} x_n^{(k)}}{\partial \text{Im} a_i}$$

Copyright © 2006 John Wiley & Sons, Ltd.

1/[1 + a(z)]; p of them to generate $x_n$ via (18) and $p$ more to generate the derivative signals $\nabla_n^{(k)}$ as in (20). The resulting computational load may be excessive depending on the application, and thus a common approach is to use the approximations $\nabla_n^{(k)} \approx x_n^{(k)}$ instead of (20). The resulting algorithm reads

$$a_{n+1} = a_n + \mu X_n x_n^*$$

with $X_n$ as in (19). We will refer to (23) as the Pseudolinear Regression (PLR) algorithm, as it can be derived from the assumption that $X_n$, the regressor of the filter $1/[1 + a(z)]$, is independent of the filter coefficients. This is also known in the literature as Feintuch’s approximation, after Feintuch’s [19]; some authors refer to the PLR algorithm as Extended LMS [18]. Figure 2 illustrates the generation of the signals involved for both the MOE and PLR schemes.

The PLR algorithm disposes of the additional $p$ scalar filters that MOE requires for the computation of (20), and therefore becomes an appealing choice in applications in which the channel presents large delay spreads (thus requiring a large order $M$ in the adaptive filter). However, it is important to note that, due to the approximations introduced, PLR is not a stochastic gradient descent of the cost $J_{OE}$. As a consequence, the convergence properties of MOE and PLR are in general different.

5. ALGORITHM ANALYSIS

5.1. Stationary points

Observe that the function

$$\gamma(z) := tr \Phi_n(z) = tr[L \Phi_{nn}(z)L^H]$$

appearing in (16) can be seen as a (scalar) psd itself, corresponding to some (scalar) MA($L$) process. Therefore the $p$-channel versions of MOE and PLR can be seen as single-channel versions operating on an MA($L$) process with psd $\gamma_n(z)$.
The stationary points of the adaptive algorithms are those for which the expected value of the update term vanishes

$$E[D_n x_n^*] = 0_M \quad \text{(MOE)}, \quad E[X_n x_n^*] = 0_M \quad \text{(PLR)}$$

For MOE, these conditions simply state that the derivatives of $J_{OE}$ with respect to the coefficient vector $a$ must be zero; as corresponds to a stochastic gradient descent, the stationary points of MOE will be given by the minima of $J_{OE}(a)$. Although this cost is non-quadratic, it follows from Reference [10] that if $M \geq L$ then $J_{OE}(a)$ has a single minimum, corresponding to the mmp spectral factor of $\gamma(z)$. On the other hand, if $M < L$, i.e. if the adaptive filter is not long enough, then $J_{OE}(a)$ will be multimodal in general.

**Example 2**

Consider a noiseless undermodeled setting in which $p = 1$, $M = 1$, $L = 6$, signals and filters are real-valued, and the channel is given by $c(z) = 1 - 0.25z^{-1} - 1.4z^{-2} + 1.77z^{-3} + 0.3z^{-4} - 1.1z^{-5} + 0.6z^{-6}$. The cost $J_{OE}(a)$ presents two local minima at $a^{(1)} \approx -0.7$ and $a^{(2)} \approx 0.977$, yielding $J_{OE}(a^{(2)})$ 3 dB above $J_{OE}(a^{(1)})$.

For PLR, the characterization of stationary points admits an intuitive physical interpretation. Let $\Phi_{x_n}(z)$ be the psd of the process $\{x_n\}$, and denote its trace by

$$\gamma(z) = \text{tr} \Phi_{x_n}(z) = \sum_{k=-\infty}^{\infty} E[x_{n-k}^H x_n]z^{-k}$$

Then conditions (25) are seen to mean that, at any PLR stationary point, the coefficients of $z^{\pm 1}, \ldots, z^{\pm M}$ in the power series expansion of $\gamma(z)$ given in (26) are zero. Note that if $M \geq L$, the solution to the MOE criterion results in a constant $\gamma(z)$. The PLR approach provides an approximation to this ‘whiteness condition’ by nulling the first $M$ coefficients. It turns out that when $M \geq L$, these requirements are enough to guarantee that $\gamma(z)$ is constant; it follows that the stationary point of PLR is therefore unique. (A proof of this fact can be found in Reference [20]; in Appendix, a more intuitive proof is provided.)

When $M < L$, the first question that arises about the behaviour of the PLR approach is whether any stationary point exists. This issue is not trivial since PLR does not correspond to the minimization of any meaningful cost. The following result is proved in Appendix.

**Theorem 1**

If $\gamma(z)$ is bounded and non-zero for all $|z| = 1$, then the PLR algorithm (23) admits a stationary point $a_*$ corresponding to a minimum phase polynomial $1 + a_*(z)$.

Observe that any small amount of measurement noise in the received signal $\{u_n\}$ is enough to ensure the positivity of $\gamma(z)$. This result does not inform, however, of how many stationary points there are, or whether an attractor point always exists among these. In the case of a first-order filter ($M = 1$), it is shown in Reference [21] that the stationary point is unique even in the $M < L$ case. To the author’s knowledge, no counterexamples have been found in which PLR presents multiple stationary points.
5.2. Relation to the MSE cost

As we have seen, the recursive filter $1/(1 + a(z))$ can be thought of as a linear predictor or prewhitener operating on the psd $\gamma_x(e^{j\omega})$, in order to provide at its output a psd $\gamma_x(e^{j\omega})$ which ideally should be constant. If $M < L$, the prewhitener cannot make $\gamma_x(e^{j\omega})$ perfectly flat, but this 'spectral flatness' notion can still be related to the original cost $J_{\text{red}}$ in (8), which is the true measure of the quality of the equalizer. Observe that the vector $p_\delta$ defined in (7) contains the coefficients of the impulse response of $L_c(z)/[1 + a(z)]$ with indices $d$ through $N$.

Therefore, neglecting the effect of noise and assuming that $d = N$ and that $N$ is large enough, it is found that

$$
\|p_\delta\|^2 \approx E[x_n^H x_n] = \frac{1}{2\pi} \int_{-\pi}^{\pi} \gamma_x(e^{j\omega}) \, d\omega
$$

On the other hand, $J_{\text{red}}$ can be upper bounded as

$$
J_{\text{red}} = 1 - p_\delta^H Q_N^{-1} p_\delta
\leq 1 - \|p_\delta\|^2 \lambda_{\text{min}} \{Q_N^{-1}\}
= 1 - \frac{\|p_\delta\|^2}{\lambda_{\text{max}} \{Q_N\}}
$$

Also, $\lambda_{\text{max}} \{Q_N\}$ satisfies (see Reference [22])

$$
\lambda_{\text{max}} \{Q_N\} \leq \max_{\omega} \lambda_{\text{max}} \{\Phi_{x x}(e^{j\omega})\} \leq \max_{\omega} \text{tr} \Phi_{x x}(e^{j\omega}) = \max_{\omega} \gamma_x(e^{j\omega})
$$

and thus from (27) to (29), we get the approximate bound

$$
J_{\text{red}} \leq 1 - \frac{\frac{1}{2\pi} \int_{-\pi}^{\pi} \gamma_x(e^{j\omega}) \, d\omega}{\max_{\omega} \gamma_x(e^{j\omega})}
$$

Note that the quotient in the right-hand side of (30) is less than or equal to one, and constitutes a measure of the spectral flatness of $\gamma_x(e^{j\omega})$. Therefore, the flatter this function becomes, the smaller the resulting cost $J_{\text{red}}$ will be. Notice that a monic linear predictor minimizing $J_{OE}$ (as $1/[1 + a(z)]$ with the MOE criterion, in our case) maximizes a different spectral flatness measure, defined as the ratio of the geometric to arithmetic means of $\gamma_x(e^{j\omega})$ [23]. This is so because the geometric mean of a psd does not change after filtering with a causal monic impulse response.

On the other hand, it is possible to obtain an expression for the output energy $J_{OE}$ obtained at a stationary point $a_*$ of the PLR algorithm. First, we note from (18) and (25) that $a_*$ must satisfy

$$
E[X_n X_n^H] a_* = E[X_n v_n^*]
$$

and that the conditions (25) for PLR imply that $E[X_n X_n^H]$ is a multiple of the identity

$$
E[X_n X_n^H] = E[x_n^H x_n] I_M
$$
Hence from (31), $\mathbf{a}_*$ is seen to be a scaled version of the input–output crosscorrelation vector

$$
\mathbf{a}_* = \frac{1}{E[\mathbf{x}_n^H \mathbf{x}_n]} E[\mathbf{x}_n^H \mathbf{v}_n^*]
$$

(33)

Now premultiplying (18) by $\mathbf{x}_n^H$, taking expectations, and using (18) and (33)

$$
E[\mathbf{x}_n^H \mathbf{x}_n] = E[\mathbf{x}_n^H \mathbf{v}_n] - E[\mathbf{x}_n^H \mathbf{X}_n^*] \mathbf{a}_*
$$

$$
= E[\mathbf{x}_n^H \mathbf{v}_n] - \mathbf{a}_*^T E[\mathbf{X}_n^* \mathbf{v}_n]
$$

$$
= E[\mathbf{x}_n^H \mathbf{v}_n] - \mathbf{a}_*^H \mathbf{a}_* E[\mathbf{x}_n^H \mathbf{x}_n]
$$

(34)

from which we finally obtain

$$
E[\mathbf{x}_n^H \mathbf{x}_n] = \frac{E[\mathbf{x}_n^H \mathbf{v}_n]}{1 + \mathbf{a}_*^H \mathbf{a}_*} \leq E[\mathbf{x}_n^H \mathbf{v}_n]
$$

(35)

This directly relates the output energy to the input energy via the filter coefficients, and shows that the filter obtained with the PLR approach cannot be worse (in terms of the output energy achieved) than the trivial filter $\mathbf{a} = \mathbf{0}$.

5.3. Local convergence

Having established the existence of stationary points, in this section we explore the convergence properties of the MOE and PLR algorithms. To do so, we resort to the ordinary differential equation (ODE) method. Note that both algorithms can be written in the general form

$$
\mathbf{a}_{n+1} = \mathbf{a}_n + \mu \mathbf{A}_n \mathbf{x}_n^* \quad \text{with} \quad \mathbf{A}_n = \begin{cases} 
\mathbf{D}_n & \text{for MOE} \\
\mathbf{X}_n & \text{for PLR}
\end{cases}
$$

(36)

where $\mathbf{D}_n$ and $\mathbf{X}_n$ were defined in (21) and (19), respectively. Then, under some general conditions, the trajectories of (36) converge in some probabilistic sense to the solution of the following ODE as $\mu \to 0$ (see, e.g. Reference [24]):

$$
\frac{d \text{Re} \mathbf{a}(t)}{dt} = \text{Re} E[\mathbf{A}_n \mathbf{x}_n^*]_{\mathbf{a} = \mathbf{a}(t)}, \quad \frac{d \text{Im} \mathbf{a}(t)}{dt} = \text{Im} E[\mathbf{A}_n \mathbf{x}_n^*]_{\mathbf{a} = \mathbf{a}(t)}
$$

(37)

If $\mathbf{a}_*$ is a stationary point of (36), then the ODE (37) can be linearized in a neighbourhood of $\mathbf{a}_*$

$$
\frac{d}{dt} \begin{bmatrix} 
\text{Re} \mathbf{a}(t) \\
\text{Im} \mathbf{a}(t)
\end{bmatrix} = \begin{bmatrix} 
\frac{d \text{Re} E[\mathbf{A}_n \mathbf{x}_n^*]}{d \text{Re} \mathbf{a}} & \frac{d \text{Re} E[\mathbf{A}_n \mathbf{x}_n^*]}{d \text{Im} \mathbf{a}} \\
\frac{d \text{Im} E[\mathbf{A}_n \mathbf{x}_n^*]}{d \text{Re} \mathbf{a}} & \frac{d \text{Im} E[\mathbf{A}_n \mathbf{x}_n^*]}{d \text{Im} \mathbf{a}}
\end{bmatrix} \begin{bmatrix} 
\text{Re} \mathbf{a}(t) - \mathbf{a}_* \\
\text{Im} \mathbf{a}(t) - \mathbf{a}_*
\end{bmatrix}
$$

(38)

We will denote the feedback matrix in the right-hand side of (38) by $\mathbf{S}_d(\mathbf{a}_*)$. Then $\mathbf{a}_*$ will be a locally convergent stationary point of the corresponding algorithm if and only if all eigenvalues of $\mathbf{S}_d(\mathbf{a}_*)$ have negative real parts.

Since MOE is a stochastic gradient descent of the cost $J_{OE}$, the set of its stationary points is the set of singular points of $J_{OE}$ (minima, maxima and saddle points). The feedback matrix

Copyright © 2006 John Wiley & Sons, Ltd.
is readily found that the Hessian of $J_{OE}(a)$ (seen as a function of Rea, Ima) evaluated at $a_*$, so that the locally convergent stationary points are the local minima of $J_{OE}$. In undermodelled settings ($M < L$), multiple minima can exist as shown in Example 2, and therefore MOE could converge to a shallow local minimum. This emphasizes the need for a sufficiently long filter $1/\left[1 + a(z)\right]$ in order to ensure unimodality of $J_{OE}$. In that case ($M \geq L$), it is readily found that the Hessian matrix of $J_{OE}(a)$ evaluated at the unique minimum $a_*$ (for which $\gamma_x(e^{\Theta})/\left[1 + a_*(e^{\Theta})\right]^2$ = constant) reduces to the $M \times M$ autocorrelation matrix of a process having psd $1/\gamma_x(z)$. Thus the larger the variation on $|z| = 1$ of this psd, the larger the eigenvalue spread of the Hessian matrix, resulting in slower convergence of the MOE adaptive algorithm.

To study the convergence properties of PLR, let us introduce the matrices

$$T(a) = E[X_nD_n^H], \quad H(a) = E[D_nX_{n+M+1}^H]J$$

where $J$ is the $M \times M$ exchange matrix with ones in the antidiagonal and zeros elsewhere. Let $a_*$ be a stationary point of PLR. Then the PLR feedback matrix $S_d(a_*)$ can be written in terms of $T$ and $H$ as

$$S_d(a_*) = \begin{bmatrix}
\text{Re} T(a_*) + \text{Re} H(a_*) & -\text{Im} T(a_*) + \text{Im} H(a_*) \\
\text{Im} T(a_*) + \text{Im} H(a_*) & \text{Re} T(a_*) - \text{Re} H(a_*)
\end{bmatrix}$$

so that the linearized ODE can be compactly written as

$$\frac{da(t)}{dt} = -T(a_*)(a(t) - a_*) - H(a_*)(a(t) - a_*)^*$$

Note that the $(i,j)$ elements of these matrices are $T(a)_{ij} = E[d_{n-j}^H x_{n-i}]$, $H(a)_{ij} = E[x_n^H d_{n+j}]$, with $1 \leq i, j \leq M$. Then, due to stationarity of the processes $\{x_n\}$, $\{d_n\}$ for fixed $a$, it follows that $T(a)$ is Toeplitz and that $H(a)$ is Hankel.

Assume now that $M \geq L$. As we have seen, the only stationary point $a_*$ of PLR in that case is such that $1 + a_*(z)$ corresponds to the mmp spectral factor of $\gamma_x(z)$, and then $\gamma_x(z) = \gamma_x$, a positive constant. The elements of the matrices $T(a_*)$, $H(a_*)$ become

$$T(a_*)_{ij} = \text{tr}E[X_{n-i}^H d_{n+j}]|_{\gamma_x(z)\text{=constant}} = \gamma_x \cdot I_{i-j}$$

$$H(a_*)_{ij} = \text{tr}E[d_{n-i}^H X_{n+j}^H]|_{\gamma_x(z)\text{=constant}} = \gamma_x \cdot I_{i-j}^*$$

where we have used the series expansion $1/[1 + a(z)] = t_0^* + t_1^* z^{-1} + t_2^* z^{-2} + \cdots$, with $t_0^* = 1$. Since $t_n^* = 0$ for $n < 0$, it follows that the Hankel term $H(a_*)$ is zero, whereas the Toeplitz term $T(a_*)$ is lower triangular with diagonal elements equal to $\gamma_x > 0$. Observe that if $\{\lambda_1, \ldots, \lambda_M\}$ denote the eigenvalues of $T(a_*)$, then the eigenvalues of a matrix with the structure of (40) with $H(a_*) = 0$ are given by $\{\lambda_1, \ldots, -\lambda_M, -\lambda_1^*, \ldots, -\lambda_M^*\}$. Thus these eigenvalues are all real and negative, and the stationary point $a_*$ is locally stable. This was observed in Reference [25] for the scalar real-valued case.

It is important to note that this convergence analysis is local; it does not provide information about the global convergence properties of the PLR algorithm. Casas et al. presented in Reference [3] a real-valued, single channel example for a third-order filter with a noiseless
mmp third-order channel \( c(z) \). In such setting the only stationary point of PLR is given by \( 1 + a_n(z) = c(z) \). It was observed in Reference [3] that the algorithm failed to converge when initialized in a neighbourhood of \( a_n \), which led the authors to suggest that in some cases PLR could be locally unstable. However, the preceding analysis has shown local stability of PLR in a setting like the one in Reference [3]. As it will be shown in Section 7, what actually happens is that the domain of attraction of the (locally convergent) stationary point can be extremely small, so that a stochastic algorithm such as (23), even if correctly initialized, may eventually escape this domain due to unavoidable adaptation noise unless the stepsize is sufficiently small. On the other hand, no convergence problems appear for the MOE algorithm in the same setting.

5.4. PLR stability

Macchi has analysed in Reference [18] the behaviour of a PLR-updated pole-zero predictor with narrowband inputs. In such setting the poles of the optimal predictor are located on the unit circle \( \mathcal{C} = \{ z : |z| = 1 \} \). Adaptation noise will inevitably push the poles outside \( \mathcal{C} \), so one could expect the adaptive filter coefficients to blow up. Surprisingly, Macchi [18] revealed a ‘self-stabilization’ mechanism by which excursions outside \( \mathcal{C} \) actually make the PLR algorithm push the offending poles back into the stability region \( \mathcal{C} = \{ z : |z| < 1 \} \). A related result is Lemma 4.2 in Reference [11], establishing the boundedness of the adaptive filter output with a vanishing stepsize \( \mu = \bar{\mu}/n \) (which is not well suited for tracking time variations of the channel.)

In our problem we have a non-vanishing stepsize and a broadband stochastic input, so it makes sense to ask about the behaviour of PLR whenever some of the filter poles get too close to \( \mathcal{C} \). To study this question, we resort again to the ODE (37) and consider the function

\[
W(t) = \frac{1}{2} \| a(t) \|^2 = \Re \left[ a^H(t) \frac{\mathrm{d} a(t)}{\mathrm{d} t} \right]
\]  

(44)

Using the fact that \( \frac{\mathrm{d} a(t)}{\mathrm{d} t} = \mathbb{E}[X_n x^*_n] \), we can write

\[
\frac{\mathrm{d} W(t)}{\mathrm{d} t} = \Re \{ a^H \mathbb{E}[X_n x^*_n] \}
\]

\[
= \Re \{ \mathbb{E}[(v_n - x_n)^T x^*_n] \}
\]

\[
= \Re \{ \mathbb{E}[x^H_n v_n] - \mathbb{E}[x^H_n x_n] \}
\]

(45)

The first term in (45) can be written as

\[
\mathbb{E}[x^H_n v_n] = \frac{1}{2\pi j} \oint_{|z| = 1} \frac{\gamma_n(z)}{z} \frac{\mathrm{d} z}{z}
\]

(46)

If all roots of \( 1 + a(z) \) are inside \( \mathcal{C} \), then the only poles of the integrand inside \( \mathcal{C} \) are those of \( \gamma_n(z)/z \). Therefore, using the residue theorem to evaluate (46), one finds that \( \mathbb{E}[x^H_n v_n] \) remains finite even as some of the roots of \( 1 + a(z) \) approach \( \partial \mathcal{C} \). On the other hand, in such situation a sufficient condition for the output power term \( \mathbb{E}[x^H_n x_n] \) in (45) to grow unbounded is that \( \gamma_n(e^{j\omega}) > 0 \) for all \( \omega \). Therefore, the time derivative of \( W(t) \) will eventually become negative, so that the adaptive filter coefficient vector norm will decrease. Although this does not necessarily imply that the adaptive filter poles will return to \( \mathcal{C} \), it shows that PLR enjoys some robustness.
to occasional pole excursions. In most cases the burst produced by the sudden increase in the output power is sufficient to return the filter to the stability region.

6. LATTICE IMPLEMENTATIONS

So far we have discussed direct form implementations of the recursive filter $\frac{1}{1 + a(z)}$. When making this filter adaptive (and therefore time varying) filter stability may be hard to ensure, especially if the desired poles are close to the unit circle. For this reason it is preferable to consider the normalized lattice structure, which guarantees filter stability even in the time varying case, provided that the magnitudes of the reflection coefficients are kept bounded away from one at all times [9].

As a step in this direction, Nandi and Anfinsen [26] presented a single-channel implementation of the PLR algorithm on a two-multiplier lattice structure, which can be readily adapted to the normalized lattice architecture. By analogy with the direct form version, the driving vector of this lattice implementation was directly taken as the state vector of the lattice structure. Similarly, a lattice version of the MOE algorithm can be devised, in which the driving vector is taken as the state vector of a copy of the lattice filter driven by the adaptive filter output. We refer to these approaches as lattice-state lattice algorithms. Another possibility is to use in the adaptation of the reflection coefficients the same driving vector as in that of the direct form parameters: we will refer to these schemes as direct form-state lattice algorithms.

As shown in Reference [27] in the context of adaptive system identification, although both the lattice- and direct form-state lattice algorithms present the same stationary points as their direct form counterparts, local stability is not necessarily preserved. In order to avoid this problem we will consider the framework of Reference [27], which allows a systematic way to devise lattice algorithms from direct form versions. In the following we derive an extension of such framework to include multichannel complex-valued signals.

The direct form adaptive algorithms (22)–(23) can be written as in (36). Let us denote $k = \mathcal{L}(a)$ the vector of reflection coefficients corresponding to the direct form vector $a$; here $\mathcal{L}(\cdot)$ denotes the direct form to lattice transformation. Consider the following adaptation rule for $k$:

$$k_{n+1} = k_n + \mu[B_n x_n^* + C_n^* x_n]$$

(47)

where the $M \times p$ matrices $B_n$, $C_n$ are given by

$$B_n = F(k_n)A_n, \quad C_n = G(k_n)A_n$$

(48)

with $F(k)$, $G(k)$ suitable $M \times M$ complex matrices. Now (47) can be rewritten as

$$\begin{bmatrix} \text{Re} k_{n+1} \\ \text{Im} k_{n+1} \end{bmatrix} = \begin{bmatrix} \text{Re} k_n \\ \text{Im} k_n \end{bmatrix} + \mu W(k_n) \begin{bmatrix} \text{Re}(A_n x_n^*) \\ \text{Im}(A_n x_n^*) \end{bmatrix}$$

(49)

where

$$W(k) = \begin{bmatrix} \text{Re} F(k) + \text{Re} G(k) & -\text{Im} F(k) - \text{Im} G(k) \\ \text{Im} F(k) - \text{Im} G(k) & \text{Re} F(k) - \text{Re} G(k) \end{bmatrix}$$

(50)
Our goal is to determine a transformation \( W(k) \) such that (i) the lattice algorithm presents the same stationary points as the direct form version, (ii) the stability properties of these stationary points are likewise preserved, and (iii) the resulting lattice algorithms admit a computationally efficient implementation.

As for the first requirement, note that as long as \( W(k) \) is non-singular for all \( k \) in the stability domain (i.e. the set \( \{ k : |k_i| < 1 \text{ for } 1 \leq i \leq M \} \) ), the stationary points of the lattice algorithm (47) are the same (in transfer function space) as those of the original direct form scheme (36).

As for stability, let \( k_* \) be a stationary point of (51). Then the associated ODE can be linearized in a neighbourhood of \( k_* \) to obtain

\[
\begin{align*}
\frac{d}{dt} \begin{bmatrix} \text{Re} k(t) \\ \text{Im} k(t) \end{bmatrix} &= \begin{bmatrix} \frac{d}{d k} \text{Re} E[B_n x_n^*] & \frac{d}{d k} \text{Re} E[B_n x_n^*] \\ \frac{d}{d k} \text{Im} E[B_n x_n^*] & \frac{d}{d k} \text{Im} E[B_n x_n^*] \end{bmatrix} \begin{bmatrix} \text{Re}[k(t) - k_*] \\ \text{Im}[k(t) - k_*] \end{bmatrix} \\
\end{align*}
\]

(51)

Let \( S_l(k_*) \) denote the feedback matrix in (51), and let \( a_* = \mathcal{L}^{-1}(k_*) \), which is a stationary point of the direct form version assuming \( W(k_*) \) non-singular. Using the techniques of Reference [27], \( S_l(k_*) \) can be related to the corresponding feedback matrix \( S_i(a_*) \) of the direct form version as per

\[
S_l(k_*) = W(k_*) S_i(a_*) J(a_*) \quad \text{with} \quad J(a_*) = \begin{bmatrix} \frac{d}{d a} \text{Re a} & \frac{d}{d a} \text{Im a} \\ \frac{d}{d k} \text{Re k} & \frac{d}{d k} \text{Im k} \end{bmatrix} \left|_{a=a_*} \right.
\]

(52)

It is seen that a choice for \( W(k) \) that always preserves the local convergence properties of the stationary points is the Jacobian inverse \( J^{-1}(a) \). However, an efficient implementation of the resulting lattice algorithm is not known. For this reason, following Reference [27] we will consider instead the choice of the Jacobian transpose: \( W(k) = J^T(a) \). This ensures invertibility of\( W(k) \) since the lattice-to-direct form transformation \( \mathcal{L}^{-1} \) is one-to-one on the stability domain. We refer to the resulting adaptive schemes as congruent lattice algorithms since the feedback matrices \( S_l(k_*) \), \( S_i(a_*) \) are related by a congruence transformation.

Note that the use of the conjugated term \( c_n x_n \) in (47) is necessary in order to be able to set \( W(k) = J^T(a) \). This is because the transformation \( \mathcal{L}^{-1} \) does not satisfy the Cauchy–Riemann conditions: except for the case \( M = 1 \), one has

\[
\begin{align*}
\frac{d}{d k} \text{Re a} & \neq \frac{d}{d k} \text{Im a} , & \frac{d}{d k} \text{Im a} & \neq \frac{d}{d k} \text{Re a} \\
\frac{d}{d k} \text{Re k} & \neq \frac{d}{d k} \text{Im k} , & \frac{d}{d k} \text{Im k} & \neq \frac{d}{d k} \text{Re k} \\
\end{align*}
\]

(53)

With the choice \( W(k) = J^T(a) \), it is straightforward to check that the \( k \)th columns of \( F(k) A_n \), \( G(k) A_n \) are obtained, respectively, as the outputs of the \( 1 \)-input \( M \)-output filters

\[
\begin{align*}
f(z) & = \frac{1}{1 + a(z)} \cdot \frac{1}{2} \left( \frac{d a(z)}{d \text{Re k}} + j \frac{d a(z)}{d \text{Im k}} \right) \quad , \quad g(z) = \frac{1}{1 + a(z)} \cdot \frac{1}{2} \left( \frac{d a(z)}{d \text{Re k}} - j \frac{d a(z)}{d \text{Im k}} \right)
\end{align*}
\]

(54)

when driven by \( x_n^{(k)} \) (for MOE) or \( v_n^{(k)} \) (for PLR). Then, following the guidelines of Reference [9, Section 7.6.2] it is possible to derive an auxiliary lattice network that implements these transfer functions with complexity proportional to \( M \). The resulting flowgraph is shown in Figure 3. Figure 4 shows the signal generation for the two congruent lattice algorithms.
Since by construction we have $S_l(k_*) = J^T(a_*)S_d(a_*)J(a_*)$, it follows that a sufficient condition for local stability of the congruent lattice algorithm at $k_*$ is that the direct form feedback matrix satisfy $S_d(a_*) + S_d^T(a_*) < 0$, see Reference [27]. For the MOE algorithm, $S_d(a_*)$ is the negative of the Hessian of the cost $J_{OE}(a)$ (which is symmetric), evaluated at the stationary point $a_*$. Thus, if $a_*$ corresponds to a (local) minimum of this cost, then $S_d(a_*) = S_d^T(a_*) < 0$ and it follows that $k_* = \mathcal{L}(a_*)$ is a locally stable stationary point of the congruent lattice MOE algorithm.

In general, the feedback matrix of the direct form PLR algorithm evaluated at a locally stable stationary point $a_*$ need not satisfy $S_d(a_*) + S_d^T(a_*) < 0$. The following result gives a sufficient condition for this. (The proof is given in Appendix.)
Theorem 2
If the adaptive filter order satisfies $M \geq L$, and the transfer function corresponding to the stationary point of the PLR algorithm is strictly positive real (SPR), i.e. if $\text{Re}[1 + a_\ast(z^{\omega})] > 0$ for all $\omega$, then the feedback matrix satisfies $S_d(a_\ast) + S_d^T(a_\ast) = 0$.

Thus, if the stationary point transfer function is SPR, then the congruent lattice PLR algorithm is locally convergent. Observe that local stability of the direct form PLR algorithm holds regardless of this SPR condition. In summary, the congruent lattice implementation of the PLR algorithm has two drawbacks: (i) stability is not guaranteed for non-SPR stationary points, and (ii) its complexity is the same as that of the congruent lattice MOE algorithm (see Figure 4), in contrast with the direct form version. This is due to the fact that there is no immediate way to extract the transfer function $1/[1 + a(z)]$ from the auxiliary lattice network of Figure 3, and therefore it is not possible to use the same filter to calculate the output signal $x^{(d)}_n$ and the signals driving the adaptation, as was done in the direct form implementation.

7. SIMULATION RESULTS
Consider Casas’ real-valued, single-channel example from Reference [3], in which the third-order channel is given by $c_0(z) = 1 + 2.2z^{-1} + 1.7z^{-2} + 0.464z^{-3}$ and the additive noise is absent. If the adaptive filter has order three, it follows from the previous analysis that the PLR algorithm (either in direct or lattice form) presents a single stationary point, which is $1 + a_\ast(z) = c_0(z)$ since $c_0(z)$ is mmp. In Reference [3] it was suggested that this point could be unstable for the direct form version of PLR. However, our analysis of the ODE has shown that local stability must hold. This is confirmed by simulating the associated ODE; the coefficient trajectories that result when initializing the ODE solver in a neighbourhood of the stationary point (deviations of $0.01$ about the true values) are shown in Figure 5.

Figure 6 shows the parameter evolution of the direct form and lattice (lattice-state, direct form-state and congruent) implementations of PLR. In all cases the all-zero initialization was used. Very small stepsize values were chosen on purpose in order to ensure that local stability properties are not affected by adaptation noise. The channel input was a unit variance white Gaussian process.

The direct form version exhibits a bursty behaviour that prevents convergence to the stationary point. By examining the time evolution of the adaptive filter poles, one finds that the bursts occur whenever a conjugate pair of these poles reaches the unit circle. This is a consequence of the stabilizing mechanism discussed in Section 5.4. Then the coefficients enter a quasiperiodic regime of adaptation and bursts in which they remain locked. This is not the case for the lattice-state lattice variant, which correctly converges to the true channel parameters. The direct form-state and congruent lattice versions present convergence problems in this setting, and in fact an eigenvalue calculation reveals that the stationary point is unstable for either of these two lattice variants. Note that $c_0(z)$ does not satisfy the SPR condition of Theorem 2.

The behaviour of the different MOE versions with Casas’ channel is shown in Figure 7. All of them correctly converge to the true parameter values. Their trajectories are much smoother than those of the lattice state-lattice version of PLR in Figure 6 (the only PLR variant which managed to converge to the true values).
However, as mentioned in Section 6, the lattice-state and direct form-state versions of the lattice algorithms need not preserve the stability of the stationary point. To illustrate this fact, Figure 8 shows the evolution of the magnitude of the adaptive filter parameters, for the different variants of the MOE algorithm, and with a noiseless seventh-order complex-valued channel given by

\[
c(z) = 1 + (-0.5697 + j1.4513)z^{-1} + (0.4829 - j0.2019)z^{-2} + (-1.9854 + j1.4341)z^{-3} \\
+ (0.1892 - j2.0387)z^{-4} + (-0.3212 + j0.3898)z^{-5} + (0.7549 - j1.0419)z^{-6} \\
+ (0.2730 + j0.5358)z^{-7}
\] (55)

This channel is minimum phase but some of its zeros are very close to the unit circle. As a consequence, the direct form implementation of the adaptive filter presents important convergence problems, as can be seen in Figure 8; reducing the stepsize does not improve this behaviour. The lattice structure is expected to be more robust in this situation, although the importance of the choice of the adaptive algorithm should be stressed. The lattice-state lattice algorithm fails to converge, since its stationary point turns out to be unstable, locking some of the reflection coefficients on the stability boundary. The direct form-state lattice scheme
manages to reach a neighbourhood of the stationary point, but the coefficients oscillate around the true values without settling due also to the unstable nature of this point. The congruent lattice version, on the other hand, shows no convergence problems.

8. CONCLUSIONS

Although the use of an IIR structure for the adaptive equalizer presents great potential in terms of performance and robustness, care must be taken when choosing both the filter structure and the adaptive algorithm. We have analysed two blind SOS-based adaptive schemes: MOE and PLR. Both present a single stationary point, enjoying local stability, provided the number of equalizer poles is no less than the channel order. However, due to the fact that PLR is not a gradient descent, its domain of attraction may be quite small. In undermodelled scenarios, we have shown that at least one stationary point always exists for PLR, which always provides

Figure 6. Parameter evolution of the different PLR variants with Casas’ channel. Dashed lines mark the true values.
some reduction of the output variance. The self-stabilization mechanism of PLR has also been described.

Should the adaptive equalizer be required to place some of its poles close to the unit circle, stability monitoring may become a serious problem. For this reason it is preferable to adopt a normalized lattice rather than a direct form structure. By including an auxiliary lattice network to compute the adaptation signals, we have shown how the MOE and PLR algorithms can be fitted to the lattice architecture. The resulting MOE and PLR lattice versions present the same computational complexity, in contrast with the direct form variants which in this sense favoured the PLR approach. Also, the PLR lattice version is not locally convergent in general, although it has been shown that a strict positive real condition on the stationary point suffices for this. In view of these facts, the lattice MOE algorithm seems to come first in terms of robustness and convergence properties. Further work should address important issues such as the behaviour of these algorithms with practical communication channels, especially in terms of convergence speed and undermodelled settings.

Figure 7. Parameter evolution of the different MOE variants with Casas’ channel. Dashed lines mark the true values.
APPENDIX A: UNIQUENESS OF THE PLR SOLUTION FOR $M \geq L$

Observe that the functions $\gamma_v(z)$, $\gamma_x(z)$ in (24), (26) are related by

$$\gamma_x(z) = \frac{\gamma_v(z)}{[1 + a(z)][1 + \bar{a}(z)]} \quad (A1)$$

Let $\gamma_v(z) = g(z)\bar{g}(z)$ constitute a spectral factorization of $\gamma_v(z)$, with $g(z) = g_0^* + g_1^*z^{-1} + \cdots + g_L^*z^{-L}$ a minimum phase polynomial. If $M \geq L$, we can consider an $M$-dimensional state-space representation $(A, b, c, d)$ of the dynamical system $g(z)/[1 + a(z)]$ having as input a zero mean white process $\{\epsilon_n\}$

$$z_{n+1} = Az_n + b\epsilon_n, \quad y_n = cz_n + d\epsilon_n \quad (A2)$$

Figure 8. Evolution of the absolute value of the filter coefficients for the different MOE variants with a complex channel. Dashed lines mark the true values.
where \( \mathbf{z}_n \) is the state vector and \( y_n \) is the output signal. Observe that \( \gamma_s(z) \) can be written as
\[
\gamma_s(z) = \sum_{k=-\infty}^{\infty} \mathbb{E}[y_n y_{n-k}^*] z^{-k}
\tag{A3}
\]

Let \( \mathbf{K} = \mathbb{E}[\mathbf{z}_n \mathbf{z}_{n}^H] \) be the asymptotic state covariance matrix. As shown in Reference [9, Prob. 2.13], the autocorrelation coefficients satisfy
\[
\mathbb{E}[y_n y_{n-k}^*] = \begin{cases} 
\mathbf{cK}^H + |d|^2, & k = 0 \\
\mathbf{cA}^{k-1} (\mathbf{AKc}^H + \mathbf{bd}^*), & k \geq 1
\end{cases}
\tag{A4}
\]

Since \( \det(z \mathbf{I} - \mathbf{A}) = z^M[1 + a(z)] \), from the Cayley–Hamilton theorem it follows that
\[
\mathbf{A}^M + a_1^* \mathbf{A}^{M-1} + \cdots + a_{M-1}^* \mathbf{A} + a_M^* \mathbf{I} = \mathbf{0}
\tag{A5}
\]
from which one has \( \mathbf{A}^{M+k-1} = -\sum_{i=1}^{M} a_i^* \mathbf{A}^{M+k-i-1} \) for all \( k \geq 1 \). Therefore, from (A4), one has that for \( k \geq 1 \)
\[
\mathbb{E}[y_n y_{n-M-k}^*] = \mathbf{cA}^{M+k-1} [\mathbf{AKc}^H + \mathbf{bd}^*]
\]
\[
= - \sum_{i=1}^{M} a_i^* \mathbf{cA}^{M+k-i-1} [\mathbf{AKc}^H + \mathbf{bd}^*]
\]
\[
= - \sum_{i=1}^{M} a_i^* \mathbb{E}[y_n y_{n-M-k+i}^*]
\tag{A6}
\]
which equals zero if \( 1/[1 + a(z)] \) is a stationary point of the PLR algorithm, since in that case \( \mathbb{E}[y_n y_{n-k}^*] = 0 \) for \( 1 \leq k \leq M \) holds. Therefore \( \gamma_s(z) \) must reduce to a constant, and \( 1 + a(z) \) must be the mmf spectral factor of \( \gamma_s(z) \), which is a scaled version of \( g(z) \).

**APPENDIX B: PROOF OF THEOREM 1**

Given a parameter vector \( \mathbf{a} \) and its corresponding polynomial \( 1 + a(z) \), let us introduce the mapping \( \mathcal{F} : \mathbb{C}^M \rightarrow \mathbb{C}^M \) as
\[
\mathcal{F}(\mathbf{a}) = \mathbb{E}[\mathbf{X}_n \mathbf{X}_{n}^H]^{-1} \mathbb{E}[\mathbf{X}_n \mathbf{v}_n^*]
\tag{B1}
\]
where \( \mathbf{X}_n \) is given by (19). Note that \( \mathcal{F} \) is indeed a function of \( \mathbf{a} \) because the signal \( \mathbf{x}_n \), given by (15), can be written as \( \mathbf{x}_n = \mathbf{v}_n - \mathbf{X}_n^H \mathbf{a}^* \). Also note that \( \mathcal{F}(\mathbf{a}) \) comprises the coefficients of the MMSE linear filter for the estimation of \( \mathbf{v}_n \), given \( \mathbf{X}_n \), and that it can be rewritten as \( \mathcal{F}(\mathbf{a}) = \mathbf{a} + \mathbf{f}(\mathbf{a}) \), where \( \mathbf{f}(\mathbf{a}) = \mathbb{E}[\mathbf{X}_n \mathbf{X}_{n}^H]^{-1} \mathbb{E}[\mathbf{X}_n \mathbf{v}_n^*] \) is the forward prediction error filter of order \( M \) for the process \( \{\mathbf{x}_n\} \).

Consider an iterative procedure given by \( \mathbf{a}_{i+1} = \mathcal{F}(\mathbf{a}_i) = \mathbf{a}_i + \mathbf{f}(\mathbf{a}_i) \). At any fixed point \( \mathbf{a}^* \), one must have \( \mathbf{f}(\mathbf{a}^*) = \mathbf{0}_M \), that is
\[
\mathbb{E}[\mathbf{X}_n \mathbf{X}_{n}^H]^{-1} \mathbb{E}[\mathbf{X}_n \mathbf{v}_n^*] = \mathbf{0}_M \Rightarrow \mathbb{E}[\mathbf{X}_n \mathbf{v}_n^*] = \mathbf{0}_M
\tag{B2}
\]
so that \( \mathbf{a}^* \) is a stationary point of the PLR algorithm (see (25)). We will show that under the conditions of Theorem 1, the iteration \( \mathbf{a}_{i+1} = \mathcal{F}(\mathbf{a}_i) \) admits a fixed point corresponding to a minimum phase polynomial. To do so, let us introduce the map
\( \mathbf{k} = \mathcal{L}(\mathbf{a}) \), where \( \mathbf{k} = [k_1 \; \cdots \; k_M]^{\mathsf{T}} \) is the vector of reflection coefficients (lattice parameters) associated to the polynomial \( 1 + a(z) \). The stability domain becomes the following convex open subset of \( \mathbb{C}^M \):

\[
\mathcal{D} = \{ \mathbf{k} : |k_i| < 1 \quad \text{for} \quad i = 1, \ldots, M \} \quad (B3)
\]

We can define the function \( \mathcal{G} : \mathcal{D} \rightarrow \mathbb{C}^M \) as the composition \( \mathcal{G} = \mathcal{L} \circ \mathcal{F} \circ \mathcal{L}^{-1} \). Thus the iteration can be reparameterized in lattice co-ordinates as \( \mathbf{k}_{i+1} = \mathcal{G}(\mathbf{k}_i) \). Observe that the iteration may break if \( \mathbf{k}_i \) is outside \( \mathcal{D} \).

Let \( \partial \mathcal{D} \) denote the boundary of \( \mathcal{D} \). As the vector of lattice parameters \( \mathbf{k} \) approaches \( \partial \mathcal{D} \), at least one root of the corresponding polynomial \( 1 + a(z) \) approaches the unit circle. In that case the diagonal entries of the matrix \( \mathbb{E}[\mathbf{x}_n \mathbf{x}_n^H] \) become

\[
\mathbb{E}[\mathbf{x}_n^H \mathbf{x}_n] = \frac{1}{2\pi} \int_{-\pi}^{\pi} \frac{\gamma_i(e^{j\omega})}{|1 + a(e^{j\omega})|^2} d\omega \to \infty \quad (B4)
\]

because \( \gamma_i(z) > 0 \) on the unit circle is assumed. On the other hand, the conjugated components of the vector \( \mathbb{E}[\mathbf{x}_n \mathbf{x}_n^*] \) are given by

\[
\mathbb{E}[\mathbf{x}_n^H \mathbf{x}_n] = \frac{1}{2\pi j} \oint_{|z|=1} \frac{\gamma_i(z) z^{-k-1}}{1 + \bar{a}(z)} dz \quad (B5)
\]

for \( 1 \leq k \leq M \). For \( 1 + a(z) \) minimum phase, the only poles of the integrand inside the unit circle are those of \( \gamma_i(z) \); therefore, using the residue theorem to evaluate (B5), we see that these quantities remain finite even as one or more roots of \( 1 + a(z) \) approach \( |z| = 1 \). This implies that as \( \mathbf{k} \to \partial \mathcal{D} \) from inside \( \mathcal{D} \)

\[
\mathcal{F}(\mathcal{L}^{-1}(\mathbf{k})) \to 0 \quad (B6)
\]

in view of (B4), (B5) and the definition of \( \mathcal{F} \). Since \( \mathcal{L}(\mathbf{0}) = \mathbf{0} \), we conclude that the domain of \( \mathcal{G} \) can be extended in order to include \( \partial \mathcal{D} \) by defining \( \mathcal{G}(\mathbf{k}) = \mathbf{0} \) for all \( \mathbf{k} \in \partial \mathcal{D} \), and in this way \( \mathcal{G} : \mathcal{D} \cup \partial \mathcal{D} \rightarrow \mathbb{C}^M \) remains continuous. We now invoke the following Borsuk fixed point theorem [28, p. 46].

**Theorem 3**

Let \( \mathcal{D} \) be a closed, bounded, symmetric and convex subset of \( \mathbb{C}^M \), and let \( \mathcal{G} \) be a continuous mapping from \( \mathcal{D} \) to \( \mathbb{C}^M \). If \( \mathcal{G} \) is odd along the boundary, i.e.

\[
\mathcal{G}(-\mathbf{k}) = -\mathcal{G}(\mathbf{k}) \quad \text{for all} \quad \mathbf{k} \in \partial \mathcal{D} \quad (B7)
\]

then \( \mathcal{G} \) admits a fixed point in \( \mathcal{D} \): there exists \( \mathbf{k}_* \in \mathcal{D} \) such that \( \mathcal{G}(\mathbf{k}_*) = \mathbf{k}_* \).

We can apply this result with \( \mathcal{D} = \mathcal{D} \cup \partial \mathcal{D} \); since \( \mathcal{G}(\mathbf{k}) = \mathbf{0} \) for all \( \mathbf{k} \in \partial \mathcal{D} \), (B7) is clearly satisfied. Thus \( \mathcal{G}(\mathbf{k}_*) = \mathbf{k}_* \) for some \( \mathbf{k}_* \in \mathcal{D} \). Moreover \( \mathbf{k}_* \) cannot belong in \( \partial \mathcal{D} \) since all \( \partial \mathcal{D} \) is mapped onto \( \mathbf{0} \) which is not in \( \partial \mathcal{D} \). Thus the fixed point lies inside the stability domain \( \mathcal{D} \).
APPENDIX C: PROOF OF THEOREM 2

Under the hypotheses of the theorem, the feedback matrix $S_d(a_\ast)$ is obtained by taking $H(a_\ast) = 0$ in (40). In that case we note that for any complex $M \times 1$ vector $v$

$$\begin{bmatrix} \text{Re } v^T & \text{Im } v^T \end{bmatrix} \begin{bmatrix} S_d(a_\ast) + S_d^T(a_\ast) \end{bmatrix} \begin{bmatrix} \text{Re } v \\ \text{Im } v \end{bmatrix} = -2 \text{Re}[v^H T(a_\ast)v]$$  \hfill (C1)

where $T(a_\ast)$ is lower triangular Toeplitz with the (conjugated) impulse response of the transfer function of the stationary point, as stated in (42). Let

$$v = [v_0 \cdots v_{M-1}]^T, \quad w = T(a_\ast)v = [w_0 \cdots w_{M-1}]^T$$  \hfill (C2)

Consider the sequence $\{v_n\}$ such that $v_n = 0$ for $n < 0$ and $n \geq M$, and observe that the vector $w$ contains the samples 0 through $M - 1$ of the sequence $\{w_n\}$ defined as the output of the filter $1/[1 + a^\ast(z^\ast)]$ when driven by $\{v_n\}$. Then we can write

$$\text{Re}[v^H T(a_\ast)v] = \text{Re} \sum_{n=-\infty}^{\infty} v_n^* w_n$$

$$= \text{Re} \frac{1}{2\pi} \int_{-\pi}^{\pi} \left| v(e^{j\omega}) \right|^2 \frac{1}{1 + a^\ast(e^{-j\omega})} d\omega$$

$$= \frac{1}{2\pi} \int_{-\pi}^{\pi} \left| v(e^{j\omega}) \right|^2 \text{Re} \left[ \frac{1}{1 + a^\ast(e^{-j\omega})} \right] d\omega$$  \hfill (C3)

where $v(z)$ is the $z$-transform of $\{v_n\}$. If $v \neq 0$, then the integrand in (C3) is positive provided $1 + a^\ast(z)$ is SPR. Then $\text{Re}[v^H T(a_\ast)v] > 0$, proving the theorem.

ACKNOWLEDGEMENTS

This work was supported by the Ramón y Cajal program of the Spanish Ministry of Education and Science.

REFERENCES
