Unbiased Blind Adaptive Channel Identification and Equalization

David Gesbert and Pierre Duhamel, *Fellow, IEEE*

***Abstract*—The *blind adaptive* equalization and identification of communication channels is a problem of important current theoretical and practical concerns. Recently proposed solutions for this problem exploit the diversity induced by sensor arrays or time oversampling, leading to the so-called second-order algebraic/statistical techniques. The prediction error method is one of them, perhaps the most appealing in practice, due to its inherent robustness to ill-defined channel lengths as well as for its simple adaptive implementation. Unfortunately, the performance of prediction error methods is known to be severely limited in noisy environments, which calls for the development of noise (bias) removal techniques. We present a low-cost algorithm that solves this problem and allows the adaptive estimation of *unbiased* linear predictors in additive noise with arbitrary autocorrelation. This algorithm does not require the knowledge of the noise variance and relies on a new constrained prediction cost function. The technique can be applied in other noisy prediction problems. Global convergence is established analytically. The performance of the denoising technique is evaluated over GSM test channels.**

***Index Terms*—Blind equalization, blind estimation, convergence, denoising, linear prediction.**

I. Introduction

BLIND channel equalization exploiting the channel diversity induced by sensor arrays and/or time oversampling has attracted a lot of research efforts over the last few years. Methods can be found in the literature, based on the optimization of various second-order criteria, which offer fast converging alternatives to the previously reported higher order statistics (HOS)-based techniques [1], [2]. For the most part, algebraic and second-order statistical techniques have been proposed that exploit the structural properties (Hankel, Toeplitz, *et al.*) of the single-input multiple-output (SIMO) channel or data matrices. The information on channel parameters or transmitted data is typically recovered through subspace decompositions of the received data matrix (deterministic methods, e.g., [3], [8], [9], [10]) or that of the received data correlation matrix (stochastic methods, e.g., [4], [6]).

Although very appealing from the conceptual and signal processing points of view, the use of the aforementioned techniques in real-world telecom applications faces serious challenges. First, subspace-based techniques are usually computationally involved and difficult to implement in contexts with strict delay/complexity constraints. To a large extent, adaptive signal processing solutions are better suited to high dynamic conditions commonly encountered in telecom (especially wireless) applications. Note, however, that instances of adaptive second-order equalization methods can be found, including [11] and [16].

Another drawback of subspace-based methods lies in the fact that they rely on the existence of numerically well-defined dimensions of the signal or noise subspaces or other related subspaces. Since these dimensions are obviously closely related to the channel length, subspace-based techniques are extremely sensitive to channel order modeling errors [7].

The prediction error methods (PEM's) offer an alternative to the class of techniques above. PEM's, which were first introduced in [5] and later refined in [7], exploit the i.i.d. property of the transmitted symbols and apply a minimum prediction error filter on the received data. When the PE filter is chosen long enough, the filter output becomes white and corresponds to a signal free of intersymbol-interference (ISI), which in turn can be used to estimate the channel.

The PEM offers great practical advantages over most other proposed techniques. First, channel estimation using the PE method remains consistent in the presence of channel length mismatch.[1] In fact, we can actually improve the reconstruction performance offered by PEM by capturing the energy of the more distant and less significant channel taps (which corresponds to a practical case of channel length overestimation) just by increasing the prediction horizon. This property guarantees the robustness of the technique with respect to the difficult channel length estimation problem. This is unlike subspace-based approaches in which the incorporation of smaller taps in the channel model introduces ambiguities and degrades the identification/equalization performance. Another significant advantage of the PE method is that it lends itself easily to a low-cost adaptive implementation.

Unfortunately the PEM not only has attractive properties,but it is also known to be sensitive to channel additive noise. Additive noise induces residual ISI in the output of the prediction error filter, which in turn causes inconsistency in the channel estimate. For this reason, a denoising procedure must be applied to remove the predictor's bias [7].

In the existing batch implementations of PEM, we solve the noise problem by removing the pre-estimated noise contribution from the signal's second-order statistics. Then, approximately

Manuscript received February 20, 1997; revised January 5, 1999. The associate editor coordinating the review of this paper and approving it for publication was Editor-in-Chief Prof. José M. F. Moura.

D. Gesbert is with Gigabit Wireless, Inc. Mountain View, CA 94043 USA, on leave from the Information Systems Laboratory, Stanford University, Stanford CA 94305 USA (e-mail: gesbert@rascals.stanford.edu).

P. Duhamel is with Telecom Paris/SIG, Paris, France, (e-mail: duhamel@sig.enst.fr).

Publisher Item Identifier S 1053-587X(00)00085-4.

[1]When the assumed channel length is greater than the theoretical channel length.

1053–587X/00$10.00 © 2000 IEEE

noise-free Yule–Walker equations are solved to obtain unbiased predictors. However, since the estimation of noise statistics may not be sufficiently accurate, this technique is not very attractive. Another more serious drawback is that it is not compatible with a low-cost adaptive implementation.

In this paper, we propose a denoising technique that allows the adaptive computation of an unbiased predictor in a noisy environment. The main assumption behind the method is the exploitation of the rank deficiency in the correlation matrix of the noise-free channel outputs. This structural property is found not only in the SIMO equalization problem but in many array processing applications. In a wider context, the same method can also be applied in other nonarray-based prediction problems where the prediction variables are highly correlated [13], including, e.g., the predictive coding of speech signals.

In this paper, we show that the minimization of the prediction error variance, subject to a specific quadratic constraint, permits the derivation of *asymptotically noise-free* prediction coefficients. The results are first obtained in the white noise situation and later extended to noise with arbitrary autocorrelation. The stationary points of the modified PE cost function, along with their stability, are studied, and global convergence of the proposed scheme is finally established. A nice feature of the proposed technique is that it does not require the estimation of the noise variance, although some adjustment of the constraint value is needed. However, a noise variance estimate[2] can be obtained as a by-product if desired. Finally, it can be implemented adaptively at low cost using a gradient projection algorithm. The performance of the proposed denoising algorithm is evaluated using data received over a simulated wireless test channel.

Throughout the paper, the following notations are adopted:

X^t	transpose of X.
X^+	conjugate transpose of X.
$\|\|X\|\|$	Euclidean norm of X.
X^*	complex conjugate of X.
I	identity matrix.
$\mathbf{0}$	all-zero matrix.
$E()$	Statistical expectation.
$M^{\#}$	Moore–Penrose inverse of M.
$\mathfrak{R}(x)$	Real part of x.

II. Problem Statement

Let $x(t)$ be the continuous-time signal at the output of a noisy PAM/QAM communications channel

$$x(t) = \sum_{n=-\infty}^{+\infty} s(n)h(t - nT) + b(t) \tag{1}$$

where

- $s(n)$ transmitted unit-variance symbol sequence with rate $1/T$;
- $b(t)$ additive noise with variance σ_b^2 [assumed to be uncorrelated with $s(n)$];
- $h(t)$ baseband equivalent channel, including the linear distortion caused by equipment filtering.

[2]In the white noise case.

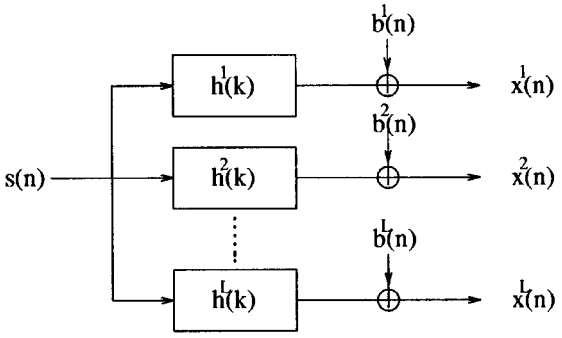

Fig. 1. Equivalent SIMO model with L channels.

In the SIMO model, a number $L > 1$ of measurements are performed for each transmitted symbol. The signal $s(n)$ then passes through L equivalent symbol-rate linear filters, leading to an overdetermined system of equations with respect to the input symbol stream (Fig. 1)

$$x^i(n) = \sum_{k=-\infty}^{+\infty} s(k)h^i(n-k) + b^i(n), \quad i = 1, \cdots, L. \tag{2}$$

Here

- $h^i(k)$ impulse response of the ith channel sampled at rate $1/T$;
- $x^i(n)$ signal measured at the ith channel;
- $b^i(n)$ corresponding noise sequence.

A SIMO model (2) typically arises for digital receivers equipped with an antenna array. A receiver equipped with a single antenna may also yield the desired overdetermination, provided the continuous-time observation is oversampled compared with the transmitted symbol rate, leading to a well-known fractionally spaced scenario. Then, $h^i(k)$, $i = 1, \cdots, L$ correspond to sampled versions (at rate $1/T$) of the same continuous-time channel $h(t)$, with respective sampling phases $(i-1)T/L$, $i = 1, \cdots, L$ [6]. More generally, L denotes the number of sensors times the oversampling factor.

A. Vectorization

Assume that ISI is causal with finite extent M. Consider the $L \times 1$ vectorized random processes $\boldsymbol{x}(n) = (x^1(n), \cdots, x^L(n))^t$, $\boldsymbol{b}(n) = (b^1(n) \cdots b^L(n))^t$, and $\boldsymbol{h}(k) = (h^1(k), \cdots, h^L(k))^t$. We may now write

$$\boldsymbol{x}(n) = \sum_{k=0}^{M} s(n-k)\boldsymbol{h}(k) + \boldsymbol{b}(n).$$

Let N be the prediction order. Form the following regression vectors: $\underline{\boldsymbol{x}}(n) = (\boldsymbol{x}(n)^t, \boldsymbol{x}(n-1)^t, \cdots, \boldsymbol{x}(n-N+1)^t)^t$, $\underline{\boldsymbol{b}}(n) = (\boldsymbol{b}(n)^t, \boldsymbol{b}(n-1)^t, \cdots, \boldsymbol{b}(n-N+1)^t)^t$, and $\underline{\boldsymbol{s}}(n) = (s(n), \cdots, s(n-M-N+1))^t$. The following compact linear model holds:

$$\underline{\boldsymbol{x}}(n) = \mathcal{H}\underline{\boldsymbol{s}}(n) + \underline{\boldsymbol{b}}(n) \tag{3}$$

where $\mathcal{H}$ is the so-called $LN \times (M+N)$ Sylvester matrix

$$\mathcal{H} = \begin{pmatrix} \boldsymbol{h}(0) & \cdots & \boldsymbol{h}(M) & \mathbf{0} & \cdots & \mathbf{0} \\ \mathbf{0} & \boldsymbol{h}(0) & \cdots & \boldsymbol{h}(M) & \mathbf{0} & \mathbf{0} \\ & & & \vdots & & \\ \mathbf{0} & \cdots & \mathbf{0} & \boldsymbol{h}(0) & \cdots & \boldsymbol{h}(M) \end{pmatrix}.$$

We assume that the prediction order N is large enough and that there exists a sufficient channel diversity so that

H1) $LN > M+N$ ($\mathcal{H}$ has strictly more rows than columns).

H2) $\mathcal{H}$ has full column rank $M+N$.

The problem is that of estimating the coefficients of $\mathcal{H}$ or, in our case, those of an inverse filter, based on the knowledge of $E(\boldsymbol{x}(n)\boldsymbol{x}(n-k)^+)$, $k = 0, 1, \cdots, N$ only.

B. Prediction Error Methods for Blind Estimation

The idea of PE methods is to find a channel inverse filter through whitening of the observations, using the following *a priori* constraint:

H3) The input sequence $s(n)$ is white, $E(s(n)s(n-k)^*) = \delta_k$.

In the SISO case, the whiteness constraint is too weak to allow phase *and* amplitude equalization of a mixed-phase channel. the prediction criterion then must be augmented with a nonlinear higher order based criterion (CMA or other); see, for instance, [14] and [15]. Fortunately, this limitation disappears in the context of SIMO channels. The main reason for this is that the matrix $\mathcal{H}$ has full column rank [under H1) and H2)]; therefore, there exist exact FIR inverses of the channel. Although the channel output is MA by construction, it is at the same time a finite-order AR signal by nature [under H3)] [7]. It was shown first in [5] that the input sequence $s(n)$ can be obtained from the noise-free prediction error signal, i.e., the error committed in linearly projecting one of the channel outputs, e.g., $x^i(n)$, onto the subspace spanned by the vector variables $\boldsymbol{x}(n-1)$, $\boldsymbol{x}(n-2)$, $\cdots$, $\boldsymbol{x}(n-N)$, as recalled in the following lemma.

Lemma 2.1: Let $\boldsymbol{p}$ be a $LN \times 1$ complex-valued vector of prediction coefficients. $\boldsymbol{p}$ is also called "predictor." Let $\tilde{x}^i(n)$ be the scalar prediction error defined, for i chosen in $1, \cdots, L$, by $\tilde{x}^i(n) = x^i(n) - \boldsymbol{p}^+\underline{\boldsymbol{x}}(n-1)$. In the *noise free case*, and under H1)–H3), the prediction error variance $J(\boldsymbol{p}) = E|\tilde{x}^i(n)|^2$ achieves its minimum if and only if

$$\tilde{x}^i(n) = (\underline{1}, -\boldsymbol{p}^+)\begin{pmatrix} \boldsymbol{x}(n) \\ \underline{\boldsymbol{x}}(n-1) \end{pmatrix} = h^i(0)s(n) \qquad (4)$$

where $\underline{1} = (0, \cdots, 0, 1, 0, \cdots, 0)$ has a 1 in the ith position. Note that the channel degree M need not be known precisely in order to use this approach; only the prediction order N should be chosen such that H1) holds. This makes the method robust toward the overestimation of the channel length.

Based on the result above, a theoretical approach to blind channel equalization was proposed, which consists of i) solving the noise-free prediction problem corresponding to the minimization of $J(\boldsymbol{p})$ and ii) using the prediction error filter $(\underline{1}, -\boldsymbol{p}^+)$ as a particular channel FIR inverse, with a remaining gain ambiguity $h^i(0)$ [5]. Obviously, the performance versus SNR of this theoretical equalization method would critically depend on the realization of the particular coefficient $h^i(0)$ in (4)[3]. For this reason, it is more appropriate to exploit the predictor as a tool to identify the channel. For instance, the prediction error $\tilde{x}^i(n)$ can be used as a training sequence against which one correlates the measured signals in order to estimate the channel coefficients:

$$\boldsymbol{h}(k) = \frac{E(\boldsymbol{x}(n+k)\tilde{x}^i(n)^*)}{h^i(0)^*} \text{ for } k = 0, \cdots, M.$$

Unfortunately, a major limitation of the prediction scheme above lies in the fact that prediction error filtering is not robust with respect to the additive noise, as shown below.

C. Case of Noisy Data

In the presence of noisy data, the minimization of $J(\boldsymbol{p})$ gives rise to modified (noisy) Yule–Walker equations, which are given by

$$R\boldsymbol{p} = \boldsymbol{a}$$

with

$$R \stackrel{\text{def}}{=} E(\underline{\boldsymbol{x}}(n-1)\underline{\boldsymbol{x}}(n-1)^+)$$

and

$$\boldsymbol{a} \stackrel{\text{def}}{=} E(\underline{\boldsymbol{x}}(n-1)x^i(n)^*). \qquad (5)$$

Under the noise/signal decorrelation condition, R and $\boldsymbol{a}$ split into

$$R \stackrel{\text{def}}{=} R_{sig} + \sigma_b^2 R_b$$
$$\boldsymbol{a} \stackrel{\text{def}}{=} \boldsymbol{a}_{sig} + \sigma_b^2 \boldsymbol{a}_b$$

where R_{sig} denotes the "signal only" part of correlation matrix, and $\boldsymbol{a}_{sig}$ denotes the "signal only" part of the correlation vector. R_b denotes the known noise correlation matrix and σ_b^2 the unknown scaling (noise variance) factor. The noise correlation vector $\boldsymbol{a}_b$ is defined in a similar way.

When the noise is temporally and spatially white, we obtain $R_b = \sigma_b^{2^{-1}} E(\underline{\boldsymbol{b}}(n)\underline{\boldsymbol{b}}(n)^+) = I$ and $\boldsymbol{a}_b = \sigma_b^{2^{-1}} E(\underline{\boldsymbol{b}}(n-1)b^i(n)^*) = \mathbf{0}$. Let $\boldsymbol{p}_0$ denote a particular *noise free* predictor, which is defined by

$$(R - \sigma_b^2 I)\boldsymbol{p}_0 = \boldsymbol{a}. \qquad (6)$$

Note that there is always at least one solution to (6) by construction. There are actually an infinite set of them in general, and this will be exploited later.

[3]Refer to [12] and [20] for possible solutions to this particular problem.

The noise induces a "bias" in $\boldsymbol{p}_0$ in the form of

$$\delta\boldsymbol{p} \stackrel{\text{def}}{=} \boldsymbol{p} - \boldsymbol{p}_0 = -\sigma_b^2 R^{-1}\boldsymbol{p}_0.$$

The biased predictor error filter $(\underline{1}, -\boldsymbol{p}^+)$ is no longer an acceptable channel inverse, which causes ISI in the prediction error signal. This in turn gives rise to inconsistent channel estimates, which calls for a bias removal procedure [7]. Traditionally, this problem is solved by first acquiring an estimate of $\sigma_b^2 I$ and then by removing it from the estimated noisy correlation matrix so that the remaining prediction algorithm deals with roughly noise-free statistics. A concern in this approach is that the noise level estimation is not always necessarily feasible or sufficiently accurate. Another drawback is that the procedure does not lend itself easily to an adaptive implementation. In this paper, we derive and analyze a modified prediction scheme allowing the adaptive computation of an *unbiased* predictor $\boldsymbol{p}_0$ in a noisy environment.

At this point, it is interesting to note the analogy and difference between our problem and the one studied (and solved) by Regalia in [18], where he addresses a similar problem of bias removal in the class of equation error methods for system identification. By analogy, $\boldsymbol{x}(n-1), \boldsymbol{x}(n-2), \cdots$ can be viewed as the input of a to-be-identified system (represented by the prediction filter) whose output is $x^i(n)$ in our context; hence, the equation error is here defined as the prediction error. However, a perfect modeling is assumed in [18], i.e., the equation error variance should be zero in the absence of measurement errors. This is clearly not the case here since (4) shows that the noise-free prediction error variance is proportional to the input power. Another difference is that [18] supposes that only the system output is corrupted by noise, which, also, is not applicable here. His model allows Regalia to identify the unbiased system simply by replacing a monic constraint by an *arbitrary* norm constraint in the equation error optimization. The nonperfect modeling makes the problem more difficult to handle here as will now be seen.

III. Adaptive Unbiased Linear Prediction

Unbiased predictors are defined for arbitrary noise autocorrelation by the solutions to the noise-free Yule–Walker equation:

$$(R - \sigma_b^2 R_b)\boldsymbol{p} = \boldsymbol{a} - \sigma_b^2 \boldsymbol{a}_b. \tag{7}$$

The focus of this paper is on the problem of direct and adaptive estimation of $\boldsymbol{p}$ in (7) with knowledge of the noise correlation matrix structure R_b but without knowledge of σ_b^2. In [19], Treichler tackles a similar problem, proposing an adaptive algorithm (so called γ-LMS algorithm) that attempts explicitly to remove the noise contribution from the prediction error instantaneous statistics, but this approach requires the estimation of the noise variance σ_b^2.

A. Constrained Prediction

We propose to use a *constrained* prediction criterion, which was introduced earlier in [21]. The idea is now sketched briefly: The prediction criterion $J(\boldsymbol{p})$ in Lemma 2.1 consists of the noisy case of a "signal" part and a "noise" part:

$$\begin{aligned} J(\boldsymbol{p}) &= J_{sig}(\boldsymbol{p}) + \sigma_b^2 J_b(\boldsymbol{p}) \\ &\cdot \begin{cases} J_{sig}(\boldsymbol{p}) = R_{sig}(i,i) + \boldsymbol{p}^+ R_{sig}\boldsymbol{p} - 2\Re\{\boldsymbol{p}^+\boldsymbol{a}_{sig}\} \\ J_b(\boldsymbol{p}) = 1 + \boldsymbol{p}^+ R_b\boldsymbol{p} - 2\Re\{\boldsymbol{p}^+\boldsymbol{a}_b\}. \end{cases} \end{aligned} \tag{8}$$

We propose to look at the minimization of the overall criterion $J(\boldsymbol{p})$ subject to the constraint that the noise part remains constant $J_b(\boldsymbol{p}) = \gamma^2$, where γ is a real number. Using the Lagrange multiplier technique, this problem leads to a set of stationary points characterized by

$$\frac{\partial J(\boldsymbol{p})}{\partial \boldsymbol{p}} + \mu \frac{\partial J_b(\boldsymbol{p})}{\partial \boldsymbol{p}} = \frac{\partial (J(\boldsymbol{p}) + \mu J_b(\boldsymbol{p}))}{\partial \boldsymbol{p}} = 0 \tag{9}$$

where μ is a real Lagrange multiplier whose value is determined somehow by the choice of the constraint. The idea here is that the stationary points of (9) are also those obtained in the *unconstrained* prediction problem with a new noise variance given by $\sigma_b^2 + \mu$. Clearly, we are interested in selecting the constraint value γ to obtain $\mu \approx -\sigma_b^2$ to achieve the desired denoising.[4] A more challenging task is to prove that the constrained prediction scheme will actually converge toward the desired solutions. These are the issues that are addressed below. The analysis is first carried out for the sake of simplicity in the case of temporally and spatially white noise. The results are later extended to arbitrary noise autocorrelation.

B. Denoising in the White Noise Case

In the white noise case, the constraint shown in Section III-A reduces to $J_b(\boldsymbol{p}) = 1 + \|\boldsymbol{p}\|^2 = \gamma^2$. Without loss of generality, we can redefine our constraint as $\|\boldsymbol{p}\| = \gamma$ (norm constraint), which brings up an interesting connection between the noise variance and the norm of the predictors. Next, we show how the predictor's norm can actually serve as a basis to discriminate between biased and unbiased predictors.

1) Norm of Predictors: In the white noise case, unbiased predictors $\boldsymbol{p}$ are defined by

$$(R - \sigma_b^2 I)\boldsymbol{p} = R_{sig}\boldsymbol{p} = \boldsymbol{a} \tag{10}$$

with $\boldsymbol{a} = \boldsymbol{a}_{sig}$. The noise -free correlation matrix is given by $R_{sig} = \mathcal{H}\mathcal{H}^+$, which, under H1), makes R_{sig} a rank-deficient matrix. For this reason, the solutions to the noise free Yule–Walker system in (10) belong to an infinite set. The singularity of the noise free Yule–Walker system constitutes an remarkable property. The main idea behind the denoising method here consists of restoring this property, in the noisy case, by acting on the norm of the predictor. In what follows, we give a useful characterization of the unbiased predictors in terms of their norms.

[4]This trick can be used to denoise other second-order statistics criteria, another example of which was shown in [17].

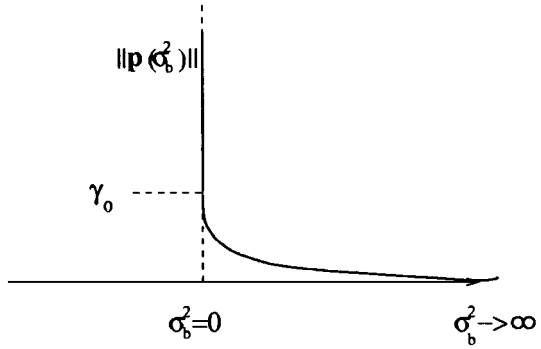

Fig. 2. $\|\boldsymbol{p}(\sigma_b^2)\|$ as a function of parameter $\sigma_b^2 > 0$, where $\boldsymbol{p}(\sigma_b^2) = (R_{sig} + \sigma_b^2 I)^{-1}\boldsymbol{a}$. The norm is monotonously decreasing as the noise level increases.

The solutions to (10) span an affine subspace that can be parameterized by

$$\begin{aligned} \boldsymbol{p} &= \boldsymbol{p}_0 + \boldsymbol{q} \\ \boldsymbol{p}_0 &\stackrel{\text{def}}{=} R_{sig}^{\#}\boldsymbol{a} \\ \boldsymbol{q} &\in \text{null}(R_{sig}) \end{aligned}$$

where "#" corresponds to the Moore–Penrose matrix inversion operator. $\boldsymbol{p}_0$ then denotes a particular unbiased predictor that is characterized by the fact that it has the minimum norm in the class of all unbiased predictors. Let us denote by γ_0 the norm of this particular predictor

$$\gamma_0 \stackrel{\text{def}}{=} \|\boldsymbol{p}_0\| = \|R_{sig}^{\#}a\|. \tag{11}$$

By construction of γ_0, all unbiased predictors $\boldsymbol{p}$ are such that $\|\boldsymbol{p}\| \geq \gamma_0$. In addition, for any positive value γ such that $\gamma_0 \leq \gamma \leq +\infty$, there exists another unbiased predictor $\boldsymbol{p}$ such that $\gamma = \|\boldsymbol{p}\|$. In other words, a predictor is unbiased if and only if its norm is lower bounded but not upper bounded due to the rank-deficiency of R_{sig}.

Conversely, the nice and remarkable result is shown below, in which *all* biased predictors, whatever the noise variance σ_b^2 actually is, have norms lower than γ_0. See Fig. 2 for an illustration.

Lemma 3.1: Let $\boldsymbol{p}(\sigma_b^2)$ be the solution to the noised Yule–Walker (5), where σ_b^2 is now a positive real-valued free parameter, i.e., $(R_{sig} + \sigma_b^2 I)\boldsymbol{p}(\sigma_b^2) = \boldsymbol{a}$. Then, for all $\sigma_b^2 > 0$, $\|\boldsymbol{p}(\sigma_b^2)\| < \gamma_0$.

Proof: Assume that $R_{sig} = \sum_{k=1,\,LN} \lambda_k \boldsymbol{w}_k \boldsymbol{w}_k^+$ represents an orthonormal eigen-decomposition of R_{sig}. Since R has a "signal" subspace of dimension $M+N$, we have classically $\lambda_1 \geq \cdots > \lambda_{M+N+1} = \cdots = \lambda_{LN} = 0$. A biased predictor $\boldsymbol{p}(\sigma_b^2)$ is given by

$$\begin{aligned} (R_{sig} + \sigma_b^2 I)^{-1}\boldsymbol{a} &= \sum_{k=1}^{LN} (\lambda_k + \sigma_b^2)^{-1} \boldsymbol{w}_k \boldsymbol{w}_k^+ \boldsymbol{a}, \\ &= \sum_{k=1}^{M+N} (\lambda_k + \sigma_b^2)^{-1} \boldsymbol{w}_k \boldsymbol{w}_k^+ \boldsymbol{a} \end{aligned}$$

where we use the fact that $\boldsymbol{a}$ lies in the signal subspace (i.e., $\boldsymbol{w}_k^+\boldsymbol{a} = 0$, $k > M+N$). The norm of $\boldsymbol{p}(\sigma_b^2)$ is then given by

$$\begin{aligned} \|\boldsymbol{p}(\sigma_b^2)\|^2 &= \sum_{k=1}^{M+N} (\lambda_k + \sigma_b^2)^{-2} \|\boldsymbol{w}_k^+\boldsymbol{a}\|^2, \\ &< \sum_{k=1}^{M+N} \lambda_k^{-2} \|\boldsymbol{w}_k^+\boldsymbol{a}\|^2 = \gamma_0^2. \end{aligned} \tag{12}$$

□

In addition, we see from (12) that $\|\boldsymbol{p}(\sigma_b^2)\|^2$ monotonously decreases as σ_b^2 increases, as we could have expected. The predictor on noisy data tries to achieve the incompatible tasks of predicting the channel output as well as the noise samples. Due to the whiteness property, the noise samples are unpredictable, which forces the predictor to decrease its gain (or norm) to minimize noise amplification. In the limit, we obtain $\boldsymbol{p} \to 0$ as $\sigma_b^2 \to \infty$.

The analysis above is useful because it allows us to characterize the bias/noise in a given predictor on the basis of its norm being below or above a threshold γ_0.

C. The Proposed Algorithm

Based on what is said in the earlier sections, the following criterion for the tracking of an unbiased predictor arises naturally:

$$\text{minimize } J(\boldsymbol{p}) = E|x^i(n) - \boldsymbol{p}^+ \underline{\boldsymbol{x}}(n-1)|^2 \tag{13}$$

$$\text{subject to } \|\boldsymbol{p}\| = \gamma. \tag{14}$$

Note that the problem of choosing γ remains open at this point. Remarkably, this simple technique consists of *augmenting* the existing monic constraint corresponding to the "$\underline{1}$" in the prediction error filter $(\underline{1}, -\boldsymbol{p}^+)$ with a *specific* norm constraint on $\boldsymbol{p}$, whereas Regalia's problem could be solved by only *replacing* the monic constraint by an *arbitrary* norm constraint.

D. Convergence Study

1) Stationary Points: Let $\mathcal{S}_\gamma$ denote the set of stationary points for the quadratically constrained criterion in (13) and (14). The lemma below provides a simple characterization of $\mathcal{S}_\gamma$ and gives a mathematical ground for the statements in Section III-A.

Lemma 3.2: The elements of $\mathcal{S}_\gamma$ are the vectors $\boldsymbol{p}(\xi)$ defined by

$$(R_{sig} + \xi I)\boldsymbol{p}(\xi) = \boldsymbol{a} \tag{15}$$

such that $\|\boldsymbol{p}(\xi)\| = \gamma$ for some arbitrary (not necessarily positive) real number ξ.

Proof: The elements of $\mathcal{S}_\gamma$ coincide with the stationary points of the auxiliary function $\mathcal{L}(\boldsymbol{p}, \mu) = J(\boldsymbol{p}) + \mu(\boldsymbol{p}^+\boldsymbol{p} - \gamma^2)$,

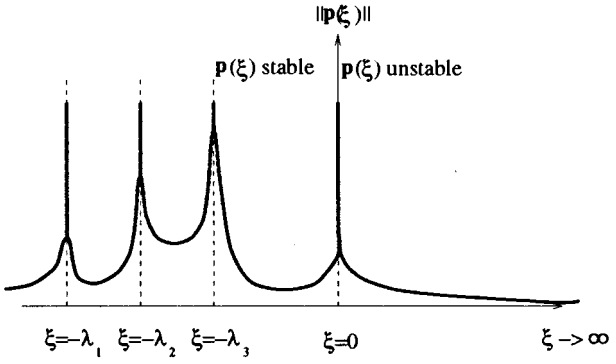

Fig. 3. $\|\boldsymbol{p}(\xi)\|$ as a function of the real-valued parameter ξ. The stability region is characterized by $\xi \geq 0$. λ_1, λ_2, λ_3 are nonzero eigenvalues of R_{sig}.

where μ is a real-valued Lagrange multiplier. The noisy prediction criterion (13) and the first-order derivative of $\mathcal{L}()$ with respect to $\boldsymbol{p}$ can then be developed according to

$$\begin{aligned} J(\boldsymbol{p}) &= E|x^i(n)|^2 + \boldsymbol{p}^+ R\boldsymbol{p} - 2\Re\{\boldsymbol{p}^+\boldsymbol{a}\} \\ \frac{\partial \mathcal{L}}{\partial \boldsymbol{p}} &= \boldsymbol{p}^+ R - \boldsymbol{a}^+ + \mu \boldsymbol{p}^+ \\ &= \boldsymbol{p}^+ R_{sig} - \boldsymbol{a}^+ + (\mu + \sigma_b^2)\boldsymbol{p}^+. \end{aligned} \tag{16}$$

where (15) simply comes from setting the first derivative $\partial\mathcal{L}/\partial\boldsymbol{p}$ to zero and denoting $\xi \stackrel{\text{def}}{=} \mu + \sigma_b^2$. □

Note the nice effect of ξ in (15), which plays the role of a virtual noise variance in the constrained prediction problem. In particular, denoising is ideally achieved for the choices of γ that result in $\xi = 0$ since (15) boils down to (10). Important remaining questions are i) do such choices of γ always exist, and ii) how can such values be predicted? Answers to these questions are given below. Based on the examination of the stability of the stationary points of (13) and (14), we characterize all possible limits for the proposed algorithm.

2) Stability Study: Let A_γ be a gradient-descent algorithm for the minimization of $J(\boldsymbol{p})$ under constraint $\|\boldsymbol{p}\| = \gamma$. The set of stationary points for A_γ is $\mathcal{S}_\gamma$ by definition. Among these, only the subset of the *stable* stationary points of A_γ is of interest here since an actual gradient algorithm will not converge to the other ones. A perturbation analysis conducted on the constrained criterion yields the following result (see also an illustration in Fig. 3).

Lemma 3.3: Let $\boldsymbol{p}(\xi)$, with norm γ, be a stationary point of algorithm A_γ defined as in (15). Then, $\boldsymbol{p}$ is a *stable* stationary point of A_γ if and only if i) the matrix $R_{sig} + \xi I \geq 0$, or, equivalently, ii) $\xi \geq 0$.

Proof:: Assume a small increment $\epsilon\boldsymbol{d}$ with respect to $\boldsymbol{p}(\xi)$. $\boldsymbol{d}$ here denotes an arbitrary unit-norm direction vector, where ϵ is the scalar amplitude of the increment. Note that ϵ is chosen to keep $\boldsymbol{p}(\xi) + \epsilon\boldsymbol{d}$ on the surface of constraint, i.e.

$$\|\boldsymbol{p}(\xi)\| = \|\boldsymbol{p}(\xi) + \epsilon\boldsymbol{d}\| = \gamma. \tag{17}$$

Due to (16), the resulting increment in $J(\boldsymbol{p}(\xi))$ is given by

$$\begin{aligned} \Delta J &\stackrel{\text{def}}{=} J(\boldsymbol{p}(\xi) + \epsilon\boldsymbol{d}) - J(\boldsymbol{p}(\xi)) \\ &= \|\epsilon^2\|\boldsymbol{d}^+ R\boldsymbol{d} + 2\Re\{\boldsymbol{p}(\xi)^+ R\boldsymbol{d}\epsilon\} - 2\Re\{\boldsymbol{a}^+\boldsymbol{d}\epsilon\} \\ &= \|\epsilon^2\|\boldsymbol{d}^+ R\boldsymbol{d} + 2\Re\{\boldsymbol{p}(\xi)^+ R_{sig}\boldsymbol{d}\epsilon\} + 2\sigma_b^2\Re\{\boldsymbol{p}(\xi)^+\boldsymbol{d}\epsilon\} \\ &\quad - 2\Re\{\boldsymbol{a}^+\boldsymbol{d}\epsilon\}. \end{aligned}$$

Since $\boldsymbol{p}(\xi)$ is a stationary point, (15) can be used, yielding

$$\begin{aligned} \boldsymbol{p}(\xi)^+ R_{sig} - \boldsymbol{a}^+ &= -\xi\boldsymbol{p}(\xi)^+ \\ \Delta J &= \|\epsilon^2\|\boldsymbol{d}^+ R\boldsymbol{d} - 2(\xi - \sigma_b^2)\Re\{\boldsymbol{p}(\xi)^+\boldsymbol{d}\epsilon\}. \end{aligned} \tag{18}$$

Finally, (17) implies that $2\Re\{\boldsymbol{p}(\xi)^+\epsilon\boldsymbol{d}\} = -\|\epsilon\|^2\boldsymbol{d}^+\boldsymbol{d}$. Then, the second-order expansion of the constrained criterion is found to be

$$\Delta J = \|\epsilon^2\|\boldsymbol{d}^+(R_{sig} + \xi)\boldsymbol{d}. \tag{19}$$

Clearly, $\boldsymbol{p}(\xi)$ is a *stable* stationary point if and only if ΔJ has a positive value, irrespective of the increment direction $\boldsymbol{d}$. Then, we should have $R_{sig} + \xi \geq 0$, which proves i). Then, ii) immediately follows from the fact that R_{sig} is positive nondefinite.

In fact, stationary points corresponding to $\xi < 0$ are characterized by $R_{sig} + \xi$ being negative definite or having negative and positive eigenvalues simultaneously, thus resulting in local maxima or saddle points. These should not affect the asymptotic behavior of the algorithm, although saddle points may imply slower convergence in their vicinity.

A global result can now be established.

Lemma 3.4: Let γ be an arbitrary positive value such that $\gamma \geq \gamma_0$, where γ_0 is the minimum norm in the class of unbiased predictors defined as in (11). Then, all stable stationary points of A_γ are solutions to (10) and, hence, are unbiased predictors.

Proof: Let $\gamma \geq \gamma_0$. Let $\boldsymbol{p}(\xi)$ be a stationary point for A_γ, i.e., a solution to (15). By application of Lemma 3.3, $\boldsymbol{p}(\xi)$ is stable and is a possible limit for A_γ iff $\xi \geq 0$, i.e. in either one of the following cases.

$\underline{\xi > 0}$*:* Equation (15) is not singular, and predictor $\boldsymbol{p}(\xi)$ is still noisy, with virtual noise variance given by ξ, but then, we know that $\|\boldsymbol{p}(\xi)\| < \gamma_0$, which is contradictory with our assumption $\gamma \geq \gamma_0$.

$\underline{\xi = 0}$*:* Equation (15) *is* singular, which results in $\|\boldsymbol{p}(\xi)\| = \gamma \geq \gamma_0$. Consequently, a stationary point with norm $\gamma \geq \gamma_0$ is stable if and only if $\xi = 0$. □

What Lemma 3.4 indicates is that, provided that the norm constraint is chosen to satisfy $\gamma \geq \gamma_0$, *any gradient-based algorithm for the minimization of the prediction error variance augmented with the constraint* $\|\boldsymbol{p}\| = \gamma$ *will converge to a solution to the noise free Yule–Walker equation* (10), regardless of initialization. This constitutes the main result of the paper. A few remarks need to be made at this point.

- Since the threshold γ_0 is the norm of a minimal unbiased predictor $\boldsymbol{p}_0$, it is unknown *a priori*. However, γ_0 does not depend on the noise level, which makes the condition $\gamma \geq \gamma_0$ not difficult to achieve in practice. The analysis above shows remarkably that any value above γ_0 will be

a good choice. Hence, there is no need for an accurate knowledge of the parameter γ_0. Guidelines are given in Section III-F on how to obtain a coarse estimate of it.

- Several solutions to (10) may have the same norm γ. All of them are equally good with respect to the denoised criterion, but they may have different performance in the required application.

Generalization: Since the proposed denoising technique essentially relies on the rank-deficiency of the correlation matrix R_{sig}, the field of application is not limited to blind equalization and can be developed in a more general context [13]. In particular, any noisy linear regression problem in which the actual (non-noisy) variables of regression are highly correlated (hence, causing the singularity of R_{sig}) constitutes a potential application.

E. Computation of σ_b^2

The knowledge of σ_b^2 is not required in the above technique, but it can be obtained as a by product if wanted. Here, we assume that the denoising algorithm has converged ideally to one of the unbiased predictors $\boldsymbol{p}$,which is a solution to (10), and we write

$$E(\underline{\boldsymbol{x}}(n-1)\underline{\boldsymbol{x}}(n-1)^+)\boldsymbol{p}-\sigma_b^2\boldsymbol{p}=E(\underline{\boldsymbol{x}}(n-1)x^i(n)^*)$$
$$E(\underline{\boldsymbol{x}}(n-1)(\underline{\boldsymbol{x}}(n-1)^+\boldsymbol{p}-x^i(n)^*))=\sigma_b^2\boldsymbol{p},$$

then

$$\frac{||E(\underline{\boldsymbol{x}}(n-1)e^i(n)^*)||}{\gamma}=\sigma_b^2 \tag{20}$$

where $e^i(n) = x^i(n) - \boldsymbol{p}^+\underline{\boldsymbol{x}}(n-1)$ denotes the prediction error. In practice, the correlation term $E(\underline{\boldsymbol{x}}(n-1)e^i(n)^*)$ may be estimated adaptively using standard time averaging techniques.

F. Implementation

The adaptive implementation of quadratically constrained filters has been widely studied in the literature and the optimization of adaptive structures for the criterion in (13) and (14) is not addressed here. The most direct solution is given by the gradient projection technique as described for instance in the work by Cox [23]. This algorithm has a very low computational cost and shows desirable robustness properties. The gradient projection algorithm is classically used in adaptive beamforming problems. It consists of a standard LMS update, here based on an expression for the gradient of $J(\boldsymbol{p})$, followed by a projection stage onto the constraint surface (here $||\boldsymbol{p}|| = \gamma$). Since all vectors are orthogonal to a sphere centered at the origin, the projection stage boils down to a simple scaling. The algorithm goes as follows:

$$\begin{aligned}
e^i(n)&=x^i(n)-\boldsymbol{p}(n-1)^+\underline{\boldsymbol{x}}(n-1)\\
\boldsymbol{u}^i(n)&=\underline{\boldsymbol{x}}(n-1)e^i(n)^*\\
\tilde{\boldsymbol{p}}(n)&=\boldsymbol{p}(n-1)+\alpha\boldsymbol{u}^i(n)\\
\boldsymbol{p}(n)&=\frac{\tilde{\boldsymbol{p}}(n)\gamma}{||\tilde{\boldsymbol{p}}(n)||}
\end{aligned}$$

where $\boldsymbol{p}(n)$ is the estimated predictor at time n, and α is a small stepsize. An estimate of the noise variance can be obtained from

$$\begin{aligned}
\boldsymbol{r}(n)&=\lambda\boldsymbol{r}(n-1)+(1-\lambda)\boldsymbol{u}^i(n)\\
\sigma_b^2(n)&=||\boldsymbol{r}(n)||/\gamma
\end{aligned} \tag{21}$$

where λ is a forgetting factor ($0 < \lambda < 1$).

Although the analysis shows that any value γ above the true γ_0 leads to unbiased predictors, the choice of γ has an influence on the mean-square-error behavior of this particular algorithm. For this reason, we should try not to overestimate too much γ_0. A coarse guess of γ_0 can be obtained in the following way: The constrained gradient algorithm first starts with γ being equal to the norm of the biased predictor (this norm can be obtained from the noisy statistics) and then slowly increases γ. Increasing γ triggers the denoising effect of the constraint. For each value of γ, the estimate $\sigma_b^2(n)$ in (21) indicates the amount of noise currently removed from the signal's statistics. At full denoising, eventually, $\sigma_b^2(n)$ reaches the actual noise variance $\sigma_b^2(n) \approx \sigma_b^2$ and saturates. At this point, γ is close to γ_0. Another, perhaps less elegant, way to guess γ_0 consists iof solving the Yule–Walker problem using a coarse estimate of R_{sig} obtained from a few tens of samples in the manner shown in [7]. This provides an estimate $\hat{\gamma}_0$ to which a small margin can be added.

G. Case of Arbitrary Noise

The channel additive noise now has an arbitrary correlation matrix $\sigma_b^2 R_b$, where R_b is a known positive definite matrix. σ_b^2 remains, as before, unknown and can be defined without loss of generality by $\sigma_b^2 = E|b^i(n)|^2$, where i is the same channel index as in (13). The right-hand term of (5) now has a nonzero noise term. The unbiased solutions to our prediction problem are, again, given by

$$(R-\sigma_b^2 R_b)\boldsymbol{p}=\boldsymbol{a}-\sigma_b^2\boldsymbol{a}_b. \tag{22}$$

Following the idea suggested in Section III-A, the following constraint is used:

$$J_b(\boldsymbol{p})=1+\boldsymbol{p}^+R_b\boldsymbol{p}-2\Re\{\boldsymbol{p}^+\boldsymbol{a}_b\}=\gamma^2. \tag{23}$$

$J_b(\boldsymbol{p})$ can be viewed as a generalization of the Euclidean norm used in the white noise case, only in this case, it takes into account the noise second-order statistics. $J_b(\boldsymbol{p})$ is also given by

$$J_b(\boldsymbol{p})=\lim_{\sigma_b^2\to\infty}\frac{J(\boldsymbol{p})}{\sigma_b^2} \tag{24}$$

which is the normalized prediction error variance in the noise-only case.

The results of Lemma 3.4 are extended as follows:

Lemma 3.5: Let A_γ be an adaptive algorithm for the minimization of $J(\boldsymbol{p})$ subject to $J_b(\boldsymbol{p}) = \gamma^2$. Let $\gamma \geq \gamma_0$, where γ_0^2 is defined as the maximum value of $J_b(\boldsymbol{p})$ for any biased predictor, i.e.,

$$\begin{aligned}
\gamma_0^2 \stackrel{\text{def}}{=} \sup\{&J_b(\boldsymbol{p});\boldsymbol{p}\text{ such that }(R_{sig}+\xi R_b)^{-1}(\boldsymbol{a}_{sig}+\xi\boldsymbol{a}_b)\\
&\text{with } 0<\xi<+\infty\}.
\end{aligned} \tag{25}$$

Then, all stable stationary points of A_γ are solutions to (22) (and, hence, are unbiased predictors).

Proof: This result is proved in Appendix A. □

TABLE I
PATH PROFILE FOR CHANNEL EQUALIZATION TESTS (COST-GSM MODEL)

path	1	2	3	4	5	6
delay (μs)	0.0	2.2	4.4	6.6	8.8	11.0
attenuation (db)	0	0	0	0	0	0

This gives a practical denoising algorithm for additive noise with arbitrary known autocorrelation. Unfortunately, the noise variance estimation scheme shown in Section III-E does not extend easily to this case.

IV. SIMULATIONS

We evaluate the performance of the proposed techniques in a digital wireless communications situation at 900 MHz, with signals impinging on a two-sensor array ($L = 2$). On sensor 1, a six-path propagation channel is simulated according to the model of Clarke [22] as

$$k^{(1)}(t) = \sum_{p=1}^{6} A_p \delta(t - \tau_p) \sum_{n=1}^{r} C_{n,p} e^{j\phi_{n,p}} \qquad (26)$$

where $r = 20$ is the number of rays impinging on the sensor within each path. $C_{n,p}$ (resp. $\phi_{n,p}$) denotes the amplitude (resp. the phase) of the nth ray within the pth path. $C_{n,p} = 1/\sqrt{r}$ is a constant. $\phi_{n,p}$ is an i.i.d. uniform process ranging from 0–2π. Each path p has amplitude A_p and delay τ_p. The path profile is shown in Table I. Note that all paths have equal power here, causing a rather severe ISI. Assuming planar wavefronts and a distance of a half wavelength ($\lambda/2 = 16$ cm) between the two sensors, we take

$$k^{(2)}(t) = \sum_{p=1}^{12} A_p \delta(t - \tau_p) \sum_{n=1}^{r} C_{n,p} e^{j(\phi_{n,p} + \Delta\phi_{n,p})} \qquad (27)$$

where $\Delta\phi_{n,p} = \pi \cos(\theta_{n,p})$ is computed according to the (random) angular incidence $\theta_{n,p}$ of the (n, p) ray with respect to the array. The symbols are unit-variance white QPSK with duration 3.7 μs as in the GSM context. Root–Nyquist filters are used for pulse shaping and reception filters (0.5 rolloff). As seen in Fig. 4, the overall channel impulse responses $h^{(1)}(t)$ and $h^{(2)}(t)$ span over a little more than four symbols. Then, the symbol-spaced channels have degree $M = 3$. Note that the symbol-spaced channels are normalized to have unit-gain.

Figs. 5–8 illustrate the behavior of the denoised PE algorithm, in a white Gaussian noise situation, based on the norm-constrained LMS implementation shown in Section III-F. The output SNR is defined by SNR $= -10 \log(\sigma_b^2)$ (since the channels have unit-gain) and is set to 10 dB. The temporal window size is chosen as $N = 4$. Then, the noiseless correlation matrix R_{sig} is 8 × 8, whereas its rank is only $M + N = 7$. Prediction is performed on the signal of sensor 2 ($i = 2$ in Section III-F). It turned out that the norm of the minimal unbiased predictor is $\gamma_0 = 1.38$. The convergence and residual bias of the adaptive

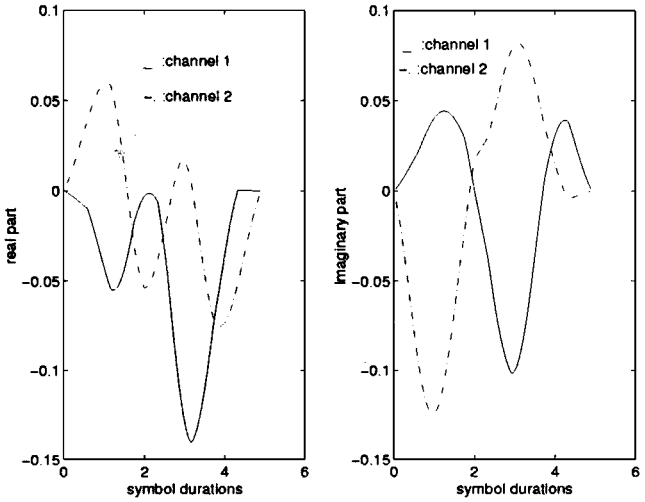

Fig. 4. Continuous-time channel impulse responses.

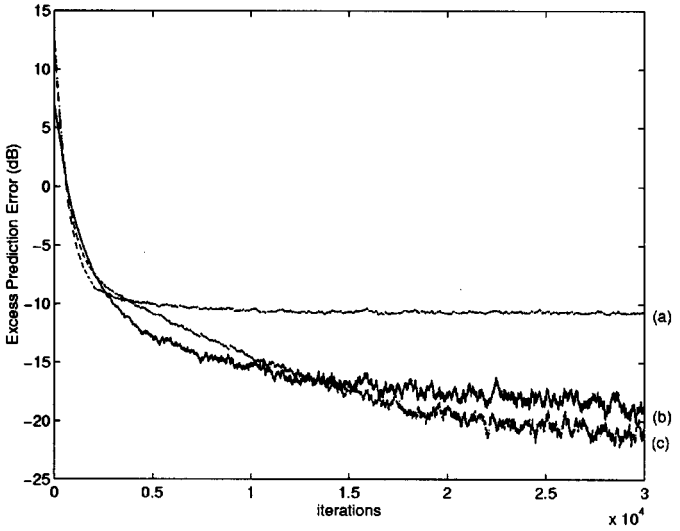

Fig. 5. Performance in the white noise case. Constraint: $\gamma = 1.5$. (a) Noisy prediction. (b) Denoised prediction. (c) Noiseless prediction.

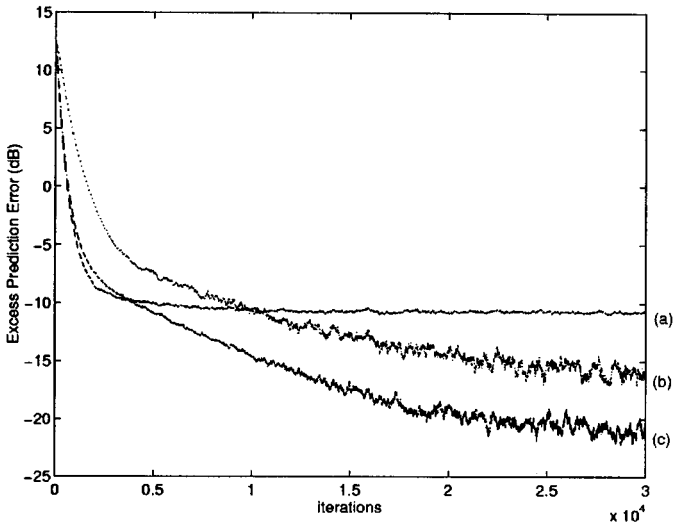

Fig. 6. Performance in the white noise case. Constraint: $\gamma = 3$. (a) Noisy prediction. (b) Denoised prediction. (c) Noiseless prediction.

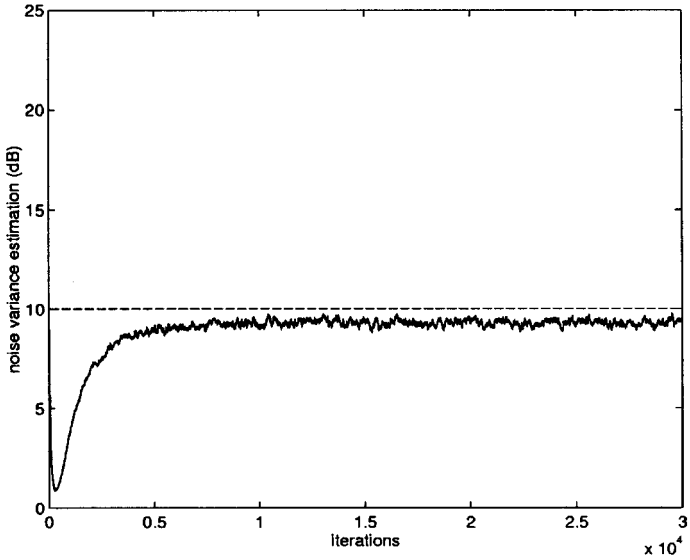

Fig. 7. Noise level estimation. Constraint: $\gamma = 1.5$.

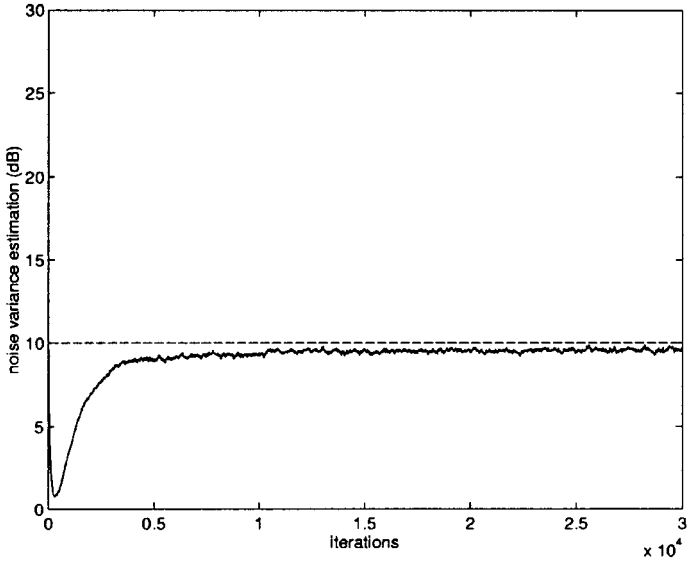

Fig. 8. Noise level estimation. Constraint: $\gamma = 3$.

predictor $\boldsymbol{p}(n)$ is measured by the excess prediction error (EPE), which is defined by

$$\text{EPE}(\boldsymbol{p}(n)) = \frac{J_{sig}(\boldsymbol{p}(n)) - \min(J_{sig})}{\min(J_{sig})}$$

where $J_{sig}()$ denotes the prediction error variance with noiseless statistics defined as in (8), and $\min(J_{sig})$ denotes the minimum of $J_{sig}()$. We can show that the EPE coincides with the level of ISI at the prediction error filter output; therefore, we also measure in this way the "distance" between the prediction error filter and a channel inverse. Clearly, for an unbiased predictor $\boldsymbol{p}$, $J_{sig}(\boldsymbol{p}) = \min(J_{sig})$, and $\text{EPE}(\boldsymbol{p}) = 0$. Learning curves are averaged over 10 independent noise/signal realizations. The stepsize is $\alpha = 0.002$.

- Fig. 5 plots $\text{EPE}(\boldsymbol{p}(n))$ versus the iteration number n. Fig. 5(a) shows the unconstrained linear prediction algorithm. Fig. 5(b) shows the new (denoised) linear prediction algorithm with $\gamma = 1.5 > \gamma_0$. We provide in Fig. 5(c), for reference, the behavior of the unconstrained linear prediction algorithm run with nonnoisy signals (ideal case). Clearly, the proposed algorithm achieves a much lower level of EPE than the initial PE algorithm, showing the efficiency of the denoising technique. The behavior of the denoised prediction scheme is close to that obtained with the ideal noise free data.
- Fig. 6 shows the result of a similar experiment, where the value of the constraint in Fig. 6(b) has been increased: $\gamma = 3 > \gamma_0$. Again, the denoising technique is successful, and it supports the theoretical claims of Lemma 3.4. Note that increasing γ results in increasing the adaptation noise in the gradient projection algorithm but does not change the mean asymptotic behavior.
- Fig. 7 shows the performance of the adaptive noise variance estimation scheme given in Section III-F with $\gamma = 1.5$. λ is set to 0.995. Note that $\sigma_b^2(n)$ converges faster than the predictor. In addition, note that the small bias in noise level is due the nonlinearity in estimating $\sigma_b^2(n)$.
- Fig. 8 shows the performance of the adaptive noise variance estimation scheme given in Section III-F with $\gamma = 3$. λ is set to 0.995. $\sigma_b^2(n)$ still converges to the right value.

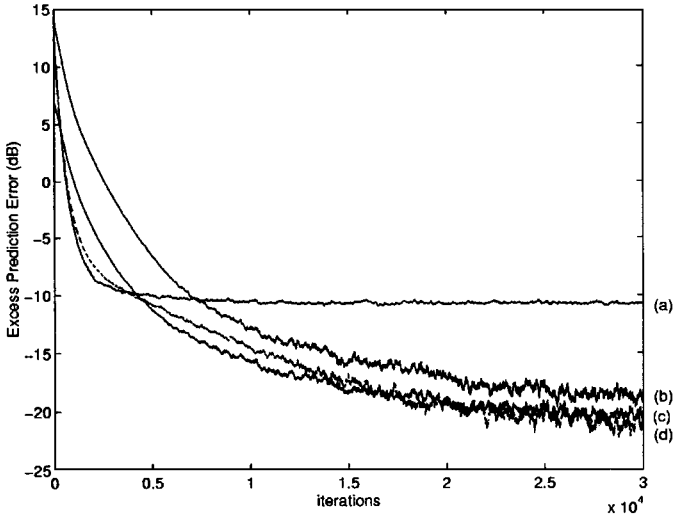

Fig. 9. Performance with an AR(1) noise model.

In the next experiment, we introduce time correlation in the noise process and try the generalized constraint shown in (23). An AR(1) model is used: $\boldsymbol{b}(n) = \beta \boldsymbol{b}(n-1) + \boldsymbol{w}(n)$, where $\boldsymbol{w}(n)$ is white temporally and spatially. We set $\beta = 0.5$. We have $E(\boldsymbol{b}(n)\boldsymbol{b}(n)^+) = \sigma_b^2 I$ with $\sigma_b^2 = 0.1$ (10 dB SNR). A preprocessing of the data showed that $\gamma_0 = 1.4$. To implement the new constraint, we perform noise prewhitening, and we use the linearly quadratically constrained algorithm shown in [23] with stepsize 0.001.

- Fig. 9 plots $\text{EPE}(\boldsymbol{p}(n))$ versus the iteration number. Fig. 9(a) shows the unconstrained linear prediction algorithm. Fig. 9(b) shows the new (denoised) linear prediction algorithm with $\gamma = 3 > \gamma_0$. Fig. 9(c) shows the new linear prediction algorithm with $\gamma = 1.5 > \gamma_0$. Fig. 9(d) shows the unconstrained linear prediction algorithm run with nonnoisy signals. The performance of the denoised algorithm

is close to that obtained in the noise free situation, with a slightly worse MSE behavior.

V. Conclusion

In a noise-free situation, the linear prediction of the outputs of a SIMO system gives rise to a singular Yule–Walker matrix equation. An adaptive prediction denoising method that restores this singularity property in the noisy case and relies on the *a priori* characterization of the norms of predictors as a function of the noise level is proposed. In the general case, the method introduces a quadratic constraint, corresponding to the noise part of the prediction error variance, which is used in the minimization of the overall prediction criterion. The analysis shows that there exists a certain threshold such that any choice of a constraint value above this threshold will result in the convergence of the proposed algorithm toward the desired unbiased prediction vectors. The same method can be used for more general noisy regression problems, provided the regression variables are sufficiently correlated.

Appendix
Convergence with Arbitrary Noise Autocorrelation

Proof of Lemma 3.5: This proof generalizes the proof of convergence given in the white noise case. The demonstration splits into three subresults allowing the characterization of

1) *stationary points*;
2) *stable points*;
3) *values taken by the constraint $J_b(\boldsymbol{p})$ for stable stationary points*.

1) Characterizations of the Stationary Points::

Lemma A.1: The stationary points of $J(\boldsymbol{p})$ subject to $J_b(\boldsymbol{p}) = \gamma^2$ are such that for some arbitrary real number ξ

$$(R_{sig} + \xi R_b)\boldsymbol{p} = \boldsymbol{a}_{sig} + \xi \boldsymbol{a}_b. \tag{28}$$

Proof: The proof of this result uses Lagrange multiplier and is a direct extension of the proof of Lemma 3.2. It is not detailed here due to lack of space.

2) Characterization of the Stable Points::

Lemma A.2: The stable stationary points of $J(\boldsymbol{p})$ subject to $J_b(\boldsymbol{p}) = \gamma^2$ are the solutions of (28) where i) $R_{sig} + \xi R_b \geq 0$, or, equivalently, ii) $\xi \geq 0$.

Proof: Using a perturbation technique similar to that of Lemma 3.3 and recalling Lemma A.1, we find for $\boldsymbol{p} \rightarrow \boldsymbol{p} + \epsilon \boldsymbol{d}$

$$\begin{aligned}\Delta J &\stackrel{\text{def}}{=} J(\boldsymbol{p} + \epsilon \boldsymbol{d}) - J(\boldsymbol{p}) \\ &= ||\epsilon^2||\boldsymbol{d}^+ R \boldsymbol{d} + 2\Re\{\boldsymbol{p}^+ R \boldsymbol{d}\epsilon\} - 2\Re\{a^+ \boldsymbol{d}\epsilon\} \\ &= ||\epsilon^2||\boldsymbol{d}^+ R \boldsymbol{d} - 2(\xi - \sigma_b^2)\Re\{\boldsymbol{p}^+ R_b \boldsymbol{d}\epsilon\} + 2(\xi - \sigma_b^2) \\ &\quad \cdot \Re\{\boldsymbol{a}_b^+ \boldsymbol{d}\epsilon\}.\end{aligned}$$

Since both $\boldsymbol{p}$ and $\boldsymbol{p} + \epsilon \boldsymbol{d}$ should belong to the surface of constraint, we have, in addition, $2\Re\{(\epsilon \boldsymbol{d})^+ (R_b \boldsymbol{p} - \boldsymbol{a}_b\} = -||\epsilon||^2 \boldsymbol{d}^+ R_b \boldsymbol{d}$. Then, it is found that

$$\Delta J = ||\epsilon^2||\boldsymbol{d}^+ (R_{sig} + \xi R_b)\boldsymbol{d}. \tag{29}$$

Again, $R_{sig} + \xi R_b$ is positive iff $\xi \geq 0$ since $R_{sig} \geq 0$ and is rank deficient. □

3) Characterization of $J_b(\boldsymbol{p})$ for Stable Stationary Points:

Lemma A.3: Let $f(\xi)$ be defined over the set $]0; +\infty[$ by $f(\xi) = J_b((R_{sig} + \xi R_b)^{-1}(\boldsymbol{a}_{sig} + \xi \boldsymbol{a}_b))$. Then, i) $f(\xi)$ is bounded from above, which guarantees the existence of γ_0 as defined in (25). ii) For $\gamma \geq \gamma_0$, the only stable stationary points of the extended algorithm in Lemma 3.5 are solutions to the unbiased system (22).

Proof: $f(\xi)$ is a continuous function. Clearly, in order to prove this, we have the following.

i) We simply have to show that $f(\xi)$ has finite limits in $\xi \rightarrow 0$ and $\xi \rightarrow \infty$. It is straightforward that $\lim_{\xi \rightarrow 0} f(\xi) = J_b(R_{sig}^{\#} \boldsymbol{a}_{sig})$ and, hence, is clearly finite. In the $\xi \rightarrow \infty$ case, $\boldsymbol{p}(\xi)$ is simply found to be the optimal noise predictor; hence, $J_b(\boldsymbol{p})$ tends to its minimum (finite) value: $\lim_{\xi \rightarrow +\infty} f(\xi) = 1 - \boldsymbol{a}_b^+ R_b^{-1} \boldsymbol{a}_b$.
ii) This follows immediately from the fact that equation (28) degenerates for $\xi = 0$; then, for any choice $\gamma \geq \gamma_0$, the only stable stationary predictors are the solution to (22).

□

Acknowledgment

Fruitful discussions with M. Unser (National Institutes of Health, Bethesda, MD) are gratefully acknowledged. The authors would also like to thank the anonymous reviewers for their comments.

References

[1] S. Haykin, *Blind Deconvolution*. Englewood Cliffs, NJ: Prentice-Hall, 1994.
[2] J. A. Cadzow, "Blind deconvolution via cumulant extrema," *IEEE Signal Processing Mag.*, pp. 24–42, May 1996.
[3] H. Liu, G. Xu, and L. Tong, "A deterministic approach to blind identification of multichannel FIR systems," in *Proc. ICASSP*, vol. 4, 1994, pp. 581–584.
[4] E. Moulines, P. Duhamel, J. F. Cardoso, and S. Mayrargue, "Subspace methods for the blind identification of multichannel FIR filters," *IEEE Trans. Signal Processing*, vol. 43, pp. 516–526, Feb. 1995.
[5] D. Slock, "Blind fractionally-spaced equalization, perfect reconstruction filter-banks and multichannel linear prediction," in *Proc. ICASSP*, vol. 4, 1994, pp. 585–588.
[6] L. Tong, G. Xu, and T. Kailath, "Blind identification and equalization based on second-order statistics: A time-domain approach," *IEEE Trans. Inform. Theory*, vol. 40, pp. 340–349, Mar. 1994.
[7] K. Abed Meraim, E. Moulines, and P. Loubaton, "Prediction error methods for second-order blind identification," *IEEE Trans. Signal Processing*, vol. 45, pp. 694–705, Mar. 1997.
[8] H. Liu and G. Xu, "A deterministic approach to blind symbol estimation," *IEEE Signsl Processing Lett.*, vol. 1, Dec. 1994.
[9] A. J. van der Veen, "Algebraic methods for deterministic blind beamforming," *Proc. IEEE*, 1998.
[10] A. J. Van der Veen, S. Talwar, and A. Paulraj, "Blind estimation of multiple digital signals transmitted over FIR channels," *IEEE Signal Processing Lett.*, vol. 2, pp. 99–102, 1995.
[11] G. Giannakis and S. Halford, "Blind fractionally-spaced equalization of noisy FIR channels: Adaptive and optimal solutions," *IEEE Trans. Signal Processing*, vol. 45, pp. 2277–2292, Sept. 1997.
[12] D. Gesbert and P. Duhamel, "Robust blind channel identification and equalization based on multistep predictors," in *Proc. ICASSP*, Apr. 1997.
[13] ——, "Adaptive and unbiased linear prediction of noisy signals," , submitted for publication.
[14] O. Macchi and A. Hachicha, "Self-adaptive equalization based on a prediction principle," in *Proc. GLOBECOM*, 1986, pp. 1641–1645.
[15] J. Labat, O. Macchi, C. Laot, and N. Le Squin, "Is training of adaptive equalizers still useful?," in *Proc. GLOBECOM*, 1996.
[16] D. Gesbert, P. Duhamel, and S. Mayrargue, "On-line blind multichannel equalization based on mutually referenced filters," *IEEE Trans. Acoust., Speech, Signasl Processing*, vol. 45, pp. 2307–2317, Sept. 1997.

[17] ——, "Blind multichannel adaptive MMSE equalization with controlled delay," in *Proc. Eighth IEEE Signal Processing Workshop SSAP*, 1996.
[18] P. A. Regalia, "An unbiased equation error identifier and reduced-order approximations," *IEEE Trans. Signal Processing*, vol. 42, June 1994.
[19] J. R. Treichler, "γ-LMS and its use in a noise compensating adaptive spectral analysis technique," in *Proc. ICASSP*, 1979.
[20] L. Tong and Q. Zhao, "Blind channel estimation by least-squares smoothing," in *Proc. ICASSP Conf.*, 1998, pp. IV-2121–IV-2124.
[21] D. Gesbert, P. Duhamel, and S. Mayrargue, "A bias removal technique for the prediction-based blind adaptive multichannel deconvolution," in *Proc. 29th Asilomar Conf. Signals, Syst., Comput.*, 1995.
[22] R. H. Clarke, "A statistical theory of mobile radio reception," *Bell Syst. Tech. J.*, vol. 47, pp. 987–1000, 1968.
[23] H. Cox, R. Zeskind, and M. Owen, "Robust adaptive beamforming," *IEEE Trans. Acoust., Speech, Signal Processing*, vol. ASSP–35, Oct. 1987.

David Gesbert was born in Mantes-La-Jolie, France, in 1969. He received the M.Sc. degree in electrical engineering from the National Institute for Telecommunications (INT), Evry, France, in 1993 and the Ph.D. degree from Ecole Nationale Superieure des Telecommunications (Telecom Paris), Paris, France, in 1997.

From 1993 to 1997, he was with France Telecom, CNET, where he was involved in the development and study of receiver algorithms for digital radio communications systems, with emphasis on blind signal detection. Since April 1997, he has been a postdoctoral fellow at the Smart Antenna Research Group Information Systems Laboratory, Stanford University, Stanford, CA. From April 1997 to April 1998, he was the recipient of a French Defense DGA/DRET Postdoctoral Fellowship. In October 1998, he took part in the team of founding employees of Gigabit Wireless, Inc., Mountain View, CA, a startup company promoting high-speed wireless data networks using smart antennas. His research interests are in the area of signal processing for digital communications, blind array processing, smart antennas, multiuser communications, and high-speed wireless networks.

Pierre Duhamel (F'98) was born in France in 1953. He received the Ingenieur degree in electrical engineering from the National Institute for Applied Sciences (INSA), Rennes, France, in 1975, the Dr. Ing. degree in 1978, and the Doctoratés sciences degree in 1986, both from Orsay University, Orsay, France.

From 1975 to 1980, he was with Thomson-CSF, Paris, France, where his research interests were in circuit theory and signal processing, including digital filtering and analog fault diagnosis. In 1980, he joined the National Research Center in Telecommunications (CNET), Issy les Moulineaux, France, where his research activities were first concerned with the design of recursive CCD filters. Later, he worked on fast Fourier transforms and convolution algorithms and applied similar techniques to adaptive filtering, spectral analysis, and wavelet transforms. He is now developing studies in channel equalization (including joint source/channel coding). Since June 1993, he has been Professor at the National School of Engineering in Telecommunications (ENST), Paris, with research activities focused on signal processing for communications. He is currently Head of the Signal and Image Processing Department.

Dr. Duhamel was Chairman of the DSP Committee of the IEEE Signal Processing Society from 1996 to 1998 and was an Associate Editor of the IEEE SIGNAL PROCESSING LETTERS. He was a Guest Editor for the Special Issue of the IEEE TRANSACTIONS ON SIGNAL PROCESSING on wavelets. He is now a member of the Committee for Signal Processing for Communications. He is the IEEE Distinguished Lecturer for 1999. The paper on subspace-based methods for blind equalization, which he co-authored, received the "Best Paper Award" from the IEEE TRANSACTIONS ON SIGNAL PROCESSING in 1998.