Adaptive Harmonic Jammer Canceler

Jean Pierre Delmas

Abstract—**A new adaptive harmonic jammer canceler is proposed. It is based on the use of two sensors that enable an adaptive generation of a reference signal that is uncorrelated with the desired signal. This reference signal is used for the reconstruction of the desired signal by an adaptive subtraction method. This canceler is well suited to radio communications.**

A theoretical analysis of the convergence of the coupled algorithms is presented with the help of the associated ordinary differential equation introduced by L. Ljung. Numerical simulations illustrate the different proposed algorithms.

I. Introduction

IN a number of radio communication applications, a certain desired signal, the bandwidth of which is narrow relative to its center frequency, is corrupted by purely harmonic jammers with unknown frequencies. After complex demodulation, the sampled complex envelope of the received signal $(z(n))_{n\in\mathbb{Z}}$ is given by

$$z(n) = x(n) + y(n) \quad \text{with} \quad y(n) \triangleq \sum_{k=1}^{P} \lambda_k e^{in\omega_k}$$

where x represents the contribution of the desired signal, and it is desirable to eliminate y from z.

This problem was addressed by Hsu and Giordano [7], then extended by Ketchum and Proakis [8] in direct-sequence spread-spectrum communication. In these papers, the desired signal is estimated as the output of a linear prediction error filter (PEF) associated with z. However, this solution has two drawbacks: first, the zeros of this filter cannot be exactly related to the sinusoid frequencies [17], so that the sinusoids are not perfectly eliminated; second, the desired signal contribution is highly distorted by the PEF. However, in direct-sequence spread-spectrum context, these undesirable effects can be overcome by increasing the number of chips per data symbol, because the intersymbol interference can be negligible when the processing gain of spread spectrum is chosen much greater than the length of the PEF.

In this paper, we present a spatio-temporal approach based on the use of two receivers that is composed of two adaptive algorithms. The first algorithm yields a purely harmonic signal that is used by the second algorithm as a reference signal to reconstruct the desired signal by a canceling approach. In particular, since the above mentioned drawbacks of [7] and [8] are eliminated, we believe that our scheme may be applied in a wider context than just the direct-sequence spread-spectrum communications context, because the sinusoids and the distortion of the desired signal can be perfectly removed.

Manuscript received September 30, 1993; revised February 17, 1995. The associate editor coordinating the review of this paper and approving it for publication was Dr. Fuyun Ling.

The author is with the Département Signal et Image, Institut National des Télécommunications, 91011 Evry Cedex, France.

IEEE Log Number 9413861.

This paper is organized as follows. After formulating the problem in Section II, we study in detail the case where the two received signals are not corrupted by noise in Section III. Then, the preceding results are extended to the case where these signals are embedded in additive observation noise in Section IV. Simulations' results are shown in Section V. Finally, a proof of the convergence of the schemes presented in Sections III and IV is reported in the Appendix.

II. Problem Formulation

Let us denote the complex envelope of the signals received on the two receivers by $z_1(n)$ and $z_2(n)$. If the multipath effects are short delays, then the narrow band hypothesis formulated on the desired signal implies that, if $z_1(n)$ is given by[1]

$$z_1(n) = x(n) + \sum_{k=1}^{P} \lambda_k e^{in\omega_k} \tag{1a}$$

then

$$z_2(n) = ax(n) + \sum_{k=1}^{P} \lambda_k a_k e^{in\omega_k} \tag{1b}$$

where a and $(a_k)_{k=1,\cdots,P}$ are unknown complex factors related to the differential time delays and to the attenuations, referenced with respect to the first receiver and corresponding, respectively, to the desired signal and to the jammers impinging on the receiver array. P is also assumed to be unknown.

Our approach is based on the fact that the complex factor a coincides with the first argument of the minimization problem (E denoting the expectation operator)

$$\min_{\alpha\in\mathbb{C},\beta\in\mathbb{C}^N} E\left|z_2(n) - \alpha z_1(n) + \sum_{k=1}^{N} \beta_k [z_2(n-k) - \alpha z_1(n-k)]\right|^2 \tag{2}$$

where N is any integer greater than or equal to P. It is therefore possible to estimate a by any standard stochastic gradient algorithm. If we denote the corresponding sequence of estimates by $a(n)$, the signal $w(n) \triangleq z_2(n) - a(n)z_1(n)$ is, after convergence, a purely harmonic signal whose angular frequencies are $\omega_1, \cdots, \omega_P$. Therefore, x can be reconstructed by

$$x(n) = z_1(n) + [C(z)]w(n) \tag{3}$$

[1] This theoretical model can be applied in practice provided the bandwidths of the jammers are negligible as compared to the inverse of the time of convergence of the adaptive algorithms (because what is just required is that the model be valid during this time of convergence). This is the case if these bandwidths are very negligible compared to the frequencies ω_k.

(the notation [ ] is shorthand for the action of the transversal filter $C(z) \triangleq \Sigma_{k=0}^{M-1} c_k z^{-k}$ on a signal) with $M \geq P$ and where $C(z)$ is found by the minimization of

$$E|z_1(n) + [C(z)]w(n)|^2. \quad (4)$$

It turns out that $w(n)$ plays the role of the reference noise signal in the classic canceling approach presented by Widrow [16], for instance. Consequently, this essentially comes down to the situation studied by Glover [5], where the fixed step-size stochastic algorithm implements a notch filter applied to $z_1(n)$ with a notch located at each of the frequencies of the jammers.

III. The Noise-Free Case

In this section, we study in detail the above mentioned adaptive approach. In particular, we discuss the uniqueness of the solution to the minimization problems (2) and (4). First, we must show that the complex gain a defined by (1b) coincides with the first argument of the minimization problem (2). Let us first remark that $z_2(n) - az_1(n)$ coincides with a purely harmonic signal with angular frequencies $\omega_1, \cdots, \omega_P$. Therefore, if $N \geq P$, there always exist coefficients $b_1, \cdots, b_N$ for which

$$z_2(n) - az_1(n) + \sum_{k=1}^{N} b_k[z_2(n-k) - az_1(n-k)] = 0.$$

It turns out that in the noise-free case, the minimum value of the criterion (2) is zero and that $(a, b_1, \cdots, b_N)$ is a solution of (2).

Now, let us discuss the uniqueness of that solution. For this purpose, we are going to study the structure of the kernel of the covariance matrix $\mathbf{R}_z$ of the random vector

$$\mathbf{z}(n) \triangleq [z_2(n), z_2(n-1), \cdots, z_2(n-N), z_1(n), z_1(n-1), \cdots, z_1(n-N)]^T.$$

Denote by $\mathbf{R}_x$ the covariance matrix of the random vector

$$x(n) \triangleq [x(n), x(n-1), \cdots, x(n-N)]^T$$

and further assume that x is a stationary random process and $(\lambda_k)_{k=1,\cdots P}$ and $(x(n))_{n\in\mathbb{Z}}$ are complex, uncorrelated zero-mean random variables. It is easily seen that

$$\mathbf{R}_z = \mathbf{R}_x \otimes \left(\begin{bmatrix} a^* \\ 1 \end{bmatrix} [a, 1]\right) + \sum_{k=1}^{P} \gamma_k^2 [\mathbf{d}(\omega_k)\mathbf{d}^H(\omega_k)] \otimes \left(\begin{bmatrix} a_k^* \\ 1 \end{bmatrix} [a_k, 1]\right) \quad (5)$$

in which $\gamma_k^2 \triangleq E|\lambda_k|^2$ and where $\mathbf{d}(\omega_k)$ denotes the vector $[1, e^{i\omega_k}, \cdots, e^{iN\omega_k}]^T$ ($\otimes$ stands for the Kronecker matrix product, * and ^{H} denote, respectively, conjugate and conjugate transposition).

Let $\mathbf{h} \triangleq [h_{0,2}, h_{1,2}, \cdots, h_{N,2}, h_{0,1}, \cdots, h_{N,1}]^T$ be a vector of $\mathbb{C}^{2(N+1)}$ such that $h_{0,2} = 1$. Then $\mathbf{h}^H\mathbf{R}_z\mathbf{h} = 0$ iff. $\mathbf{h}$ belongs to the kernel of each of the two positive-definite matrices found in the right hand side of (5).

As the desired signal x is assumed to be nondeterministic, the matrix $\mathbf{R}_x$ is positive definite. From this, we deduce immediately that the Kernel of $\mathbf{R}_x \otimes ([{}^{a^*}_{1}][a, 1])$ coincides with the space $\mathbb{C}^{N+1} \otimes [{}^{1}_{-a}]$; therefore $\mathbf{h}^H\mathbf{R}_z\mathbf{h} = 0$ iff $\mathbf{h}$ can be written as

$$\mathbf{h}^T = [1, b_1, \cdots, b_N] \otimes [1, -a]. \quad (6)$$

On the other hand, it is easily seen that $\mathbf{h}$ belongs to the Kernel of the second matrix of (5) iff $[1, b_1, \cdots, b_N]^T$ is orthogonal to the vectors $\mathbf{d}(\omega_k)_{k=1,\cdots,P}$.

This implies that each argument of the minimization problem

$$\min_{h_{0,2}=1} E|\mathbf{h}^T\mathbf{z}(n)|^2 \quad (7)$$

can be written as $\mathbf{h}^T = [1, b_1, \cdots, b_N] \otimes [1, -a]$ where

$$B(z) \triangleq 1 + \sum_{l=0}^{N} b_l z^{-l}$$

is an FIR filter for which $B(e^{i\omega_k}) = 0$ for $k = 1, \cdots, P$.

Thus, the solutions of the minimization problem (2) are $\alpha = a$ and $(\beta_1, \cdots, b_N)$ such that $1 + \sum_{l=0}^{N} \beta_l e^{-il\omega_k} = 0$ for $k = 1, \cdots, P$.

Let us now consider the adaptive minimization of (2). As is well known [6], in order to be able to minimize (2) in a reliable way by a stochastic gradient algorithm, the functional

$$\phi(\alpha, \beta) \triangleq E\left| z_2(n) - \alpha z_1(n) + \sum_{k=1}^{N} \beta_k[z_2(n-k) - \alpha z_1(n-k)]\right|^2$$

must not have spurious local minima. Unfortunately, this is not the case. In fact, it is easily seen that $\phi(\alpha, \beta)$ has a spurious local minimum iff the function defined by $\Psi(\alpha) = \min_\beta \phi(\alpha, \beta)$ has a spurious local minimum.

We illustrate this fact in Fig. 1, which represents $\Psi(\alpha)$ in a simple situation and demonstrates that $\phi(\alpha, \beta)$ may admit local spurious minima. In this figure, we choose $N = P = 1, x(n)$ is white, with power $\sigma^2 = 1, \gamma_1^2 = 100, a = 1, \omega_1 = 0$ and $a_1 = -1$. $\Psi(\alpha)$ has a global minimum for $\alpha = 1$ associated with $\beta \approx -1$ ($\phi(\alpha, \beta) = 0$) and a spurious minimum $\alpha \approx -0.9$ associated with $\beta \approx -0.217$ ($\phi(\alpha, \beta) \approx 5.26$).

Hence, we cannot work with the variables α and β in order to use stochastic gradient algorithms. However, it is possible to use the Hermitian functional

$$E\left|z_2(n) + \sum_{k=1}^{N} h_{k,2}z_2(n-k) + \sum_{k=0}^{N} h_{k,1}z_1(n-k)\right|^2 \quad (8)$$

in order to estimate a by means of a stochastic gradient algorithm. Indeed, the above functional does not admit any spurious minima. Moreover, as was shown previously, it is minimum when

$$\frac{h_{k,1}}{h_{k,2}} = -a \qquad \forall k = 0, \cdots, N \quad (\text{we put } h_{0,2} = 1). \quad (9)$$

Note that the parameter a that arises via (9) from any vector $\mathbf{h}$ that minimizes the Hermitian form (8) is unique when the number of sinusoids is known or overestimated ($N \geq P$).

To summarize, the following scheme holds for estimating a: adapt $\mathbf{h} = [h_{0,2}, h_{1,2}, \cdots, h_{N,2}, h_{0,1}, \cdots, h_{N,1}]^T$ by a stochastic gradient algorithm corresponding to the Hermitian form $E|\mathbf{h}^T\mathbf{z}(n)|^2$ and estimate a as the L.S. solution to

$\min_a \Sigma_{k=0}^N |h_{k,1} + ah_{k,2}|^2$ given by:

$$a = -\sum_{k=0}^{N} h_{k,1}h_{k,2}^* \bigg/ \sum_{k=0}^{N} |h_{k,2}|^2.$$

This yields a consistent series $a(n)$ of estimates of a.

The signal $w(n) \stackrel{\Delta}{=} z_2(n) - a(n)z_1(n)$ "converges" to $z_2(n) - az_1(n)$ which can be written as

$$w(n) = \sum_{k=1}^{P} \lambda_k(a_k - a)e^{in\omega_k}.$$

Therefore, if $a \neq a_k, \forall k = 1, \cdots, P, w(n)$ contains all the harmonic components of $y(n)$. This suggests that x can be estimated by following the classical canceling approach developed by Widrow [16]. We thus propose to minimize the following with respect to $\mathbf{c} \stackrel{\Delta}{=} (c_0, \cdots, c_{M-1})^T$ (with $M \geq P$).

$$\min_c E\left|z_1(n) + \sum_{k=0}^{M-1} c_k w(n-k)\right|^2. \quad (10)$$

The solution of this minimization is unique only if M is equal to the number P of distinct frequencies.

Finally, the desired signal $x(n) = z_1(n) + \mathbf{c}^T(n)\mathbf{w}(n)$ can be estimated by way of the following adaptive scheme, with $\mathbf{h}^T(n) = [1, \mathbf{h}_0^T(n)]$ and $\mathbf{z}^T(n) = [z_2(n), \mathbf{z}_0^T(n)]$.

The above explanation is only intended to describe a general outline in which the solution is obtained by solving a succession of two steps. These steps, however, can be merged into the following adaptive scheme denoted *Unstructured MSE algorithm.*

$$\mathbf{h}_0^T(n+1) = \mathbf{h}_0^T(n) - \mu_1[z_2(n) + \mathbf{h}_0^T(n)\mathbf{z}_0(n)]\mathbf{z}_0^H(n) \quad (11a)$$

$$a(n) = -\sum_{k=0}^{N} h_{k,1}(n)h_{k,2}^*(n) \bigg/ \sum_{k=0}^{N} |h_{k,2}(n)|^2 \quad (11b)$$

$$w(n-k) = z_2(n-k) - a(n)z_1(n-k), k = 0, \cdots, M-1$$

$$\mathbf{w}(n) = [w(n), w(n-1), \cdots, w(n-M+1)]^T$$

$$\mathbf{c}^T(n+1) = \mathbf{c}^T(n) - \mu_2[z_1(n) + \mathbf{c}^T(n)\mathbf{w}(n)]\mathbf{w}^H(n) \quad (11c)$$

We now discuss the convergence of this scheme. We note that the parameter $a(n)$ is injected from (11a)–(11c) via (11b), which is a function of the parameter $\mathbf{h}$. Therefore, we obtain two coupled stochastic gradient algorithms, and the convergence of the algorithm is not obvious.

The solution cannot be written out analytically. However, in a stationary situation, it is well known that studying the asymptotic behavior of a stochastic algorithm is tantamount to studying the stability of a certain associated ordinary differential equation (ODE), a tool introduced by Ljung [10]. More precisely, results concerning the evolution of the estimates $a(n)$ and $\mathbf{c}(n)$ as well as the convergence of the global stochastic algorithm (11) can be given in case the gain sequences $\mu_1(n)$ and $\mu_2(n)$ tend to zero (and $\Sigma_n\ \mu_i(n) = +\infty$). This technical point is solved in the Appendix.

Unfortunately, these results cannot be applied in a strict sense to our problem at hand, since the necessary hypotheses are unrealistic in this context. The gain sequences must be reduced to constant "small" steps μ_1 and μ_2 if we want the algorithm to be able to track the variations of the parameters in nonstationary environments. In this situation, the algorithm does not converge almost surely any longer. However, it is often conjectured [2], [10] that for n large enough, the adaptive algorithms will oscillate around the theoretical limit of the decreasing step scheme.

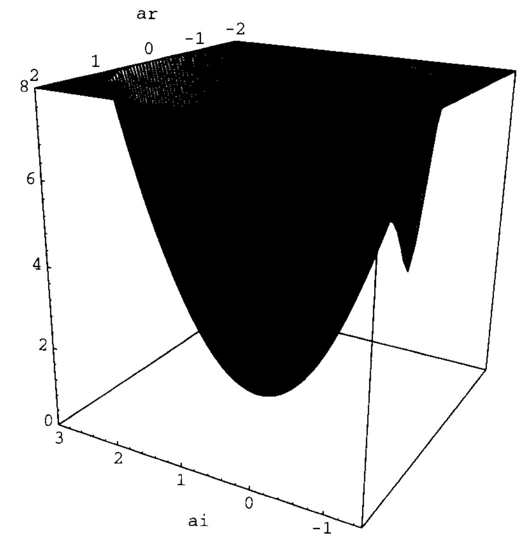

Fig. 1. Spurious local minima of the function $\psi(\alpha) = \sigma^2|a - \alpha|^2[1 + \frac{\gamma_1^4|a_1-\alpha|^4}{\alpha^2|a-\alpha|^2+\gamma_1^2|a_1-\alpha|^2}] + \gamma_1^2|a_1 - \alpha|^2[\frac{\sigma^4|a-\alpha|^4}{\sigma^2|a-\alpha|^2+\gamma_1^2|a_1-\alpha|^2}]$.

However, a simplified analysis can be made when $N = M = P$ with constant small gains. In that case, if we denote by $[1, \mathbf{h}_{0*}^T]^T$ the unique solution of minimization (7) and (11a) gives:

$$\mathbf{h}_0^T(n+1) - \mathbf{h}_{0*}^T = [\mathbf{h}_0^T(n) - \mathbf{h}_{0*}^T][\mathbf{I}_N - \mu_1\mathbf{z}_0(n)\mathbf{z}_0^H(n)]. \quad (12)$$

Therefore the algorithm (11a) is a homogeneous least mean square (LMS) algorithm and so has only a transient behavior with no asymptotic fluctuation. We can apply the results of Bitmead [3] about the transient behavior of the LMS algorithm: There exists $\lambda > 0$ such that $e^{\lambda n}[\mathbf{h}_0(n) - \mathbf{h}_{0*}] \to 0$ almost surely or in the mean square sense. So we can show from (11b) that there exists $\lambda' > 0$ such that $e^{\lambda' n}[a(n) - a] \to 0$ in the same senses. Unfortunately, we cannot say anything about the constant λ and so the evaluation of the speed of convergence of this first algorithm is not possible. As for the second algorithm (11c), if we make the classical "independence assumption" that here is far from being true but nevertheless in agreement with the simulations, the classical asymptotical excess mean squared error [6] (MSE) (provided μ_2 is small enough) becomes once the first algorithm has "converged":

$$E|\hat{x}(n) - x(n)|^2 \approx \frac{P}{2}\mu_2\sigma^2 \sum_{k=1}^{P} \gamma_k^2|a_k - a|^2 \quad (13)$$

where $\hat{x}(n)$ is the estimation of $x(n)$ given by (3) with the filter $C(z)$ issued from (11c).

IV. The Case of Noisy Observations

Here, we consider the important case where the received signals $z_1(n)$ and $z_2(n)$ are corrupted by two uncorrelated additive white noises $b_1(n)$ and $b_2(n)$ of same variance[2] σ_b^2.

[2] This restriction can be removed if it is possible to have a noise reference by using a calibration of the receivers acting upon the factors a and $(a_k)_{k=1,\cdots,P}$ of the model (1).

In this case, we indicate how it is possible to modify the minimization (2) in order to obtain a series of estimates of a and we present two different criteria in order to reconstruct x from $z_1(n)$ and $w(n) \triangleq z_2(n) - a(n)z_1(n)$.

In order to extend the results of Section III, it is sufficient to replace the vector $\mathbf{h}$, defined in Section III, by the eigenvector associated with the smallest eigenvalue σ_b^2 of $\mathbf{R}_z$. Thus, the vector $\mathbf{h}$ is solution of the minimization problem:

$$\min_h \frac{\mathbf{h}^H \mathbf{R}_z \mathbf{h}}{\|\mathbf{h}\|^2}. \tag{14}$$

Since $\mathbf{h}$ still has the structure (6), the parameter a is determined by (9) from any solution of (14).

Let us now discuss the canceling approach. In this noisy case, the signal $w(n)$ can be written as

$$w(n) = \sum_{k=1}^{P} \lambda_k (a_k - a) e^{in\omega_k} + b_2(n) - ab_1(n). \tag{15}$$

Therefore, $w(n)$ coincides with a purely harmonic signal corrupted by a white noise. It turns out that the signal x cannot be exactly reconstructed by the canceling approach described previously. However, depending on the context, several reconstruction criteria can be considered.

a) When it is desirable to cancel the residual filtered sinusoidal component (sinusoidal canceling criteria) in the estimation $\hat{x}(n)$ of $x(n)$, the vector $\mathbf{c}$ can be estimated as the solution of the minimization problem derived from (4)

$$\min_{\mathbf{c}} \mathbf{h}_{a,c}^H [\mathbf{R}_z - \sigma_b^2 \mathbf{I}_{2M}] \mathbf{h}_{a,c} \tag{16}$$

with $\mathbf{h}_{a,c} \triangleq [c_0, c_1, \cdots, c_{M-1}, 1 - ac_0, -ac_1, \cdots, -ac_{M-1}]^T$. In this context, the solution $\mathbf{c}$ is not unique when the number of sinusoids is overestimated ($M > P$), which is usually the case.

We can deduce an adaptive scheme that we denote *Pisarenko MSE*. It is composed of two stochastic gradient algorithms that are coupled, since $a(n)$ and $\sigma_b^2(n)$ are injected from the first part to the second part of the algorithm.

$$e(n) = \tilde{\mathbf{h}}^T(n)\mathbf{z}(n)$$

$$\mathbf{h}^T(n+1) = \tilde{\mathbf{h}}^T(n) - \mu_1 e(n)[\mathbf{z}^*(n) - e^*(n)\tilde{\mathbf{h}}^T(n)] \tag{17a}$$

$$\tilde{\mathbf{h}}(n+1) = \frac{\mathbf{h}(n+1)}{\|\mathbf{h}(n+1)\|} \tag{17b}$$

$$a(n) = -\sum_{k=0}^{N} h_{k,1}(n)h_{k,2}^*(n) \Big/ \sum_{k=0}^{N} |h_{k,2}(n)|^2 \tag{17c}$$

$$\sigma_b^2(n) = |e(n)|^2 \tag{17d}$$

$$\mathbf{c}^T(n+1) = \mathbf{c}^T(n) - \mu_2 \mathbf{h}_{a,c}^T(n)[\mathbf{z}(n)\mathbf{z}^H(n) - \sigma_b^2(n)\mathbf{I}_{2M}] \begin{bmatrix} \mathbf{I}_M \\ -a^*(n)\mathbf{I}_M \end{bmatrix}$$

$$\hat{x}(n) = z_1(n) + \mathbf{c}^T(n)[\mathbf{z}_2(n) - a(n)\mathbf{z}_1(n)] \tag{17e}$$

where

$$\mathbf{z}_i^T(n) \triangleq [\mathbf{z}_2^T(n), \mathbf{z}_1^T(n)]$$

with

$$\mathbf{z}_i^T(n) \triangleq [z_i(n), z_i(n-1), \cdots, z_i(n-M+1)]^T \quad \text{for} \quad i = 1, 2.$$

To get the normalized eigenvector $\tilde{\mathbf{h}}(n)$, note that we have used in (17a) and (17b) the adaptive estimation method presented in [12]. We can also use an extension to the complex case [4] of a parametrization proposed by Regalia in [13].

The estimate $\hat{x}(n)$ of $x(n)$ can be written as $\hat{x}(n) = x(n) + b_f(n)$, in which $b_f(n)$ represents the part of the input signal that comes from the observation noise.

Although the solution $\mathbf{c}$ is not unique, the solution that minimizes the power of $b_f(n)$ is that of minimum norm. This minimum power can be computed explicitly when $P = 1$:

$$E|b_f(n)|^2 = \sigma_b^2 \left(1 + \frac{2}{M|a - a_1|^2}\right). \tag{18}$$

It is interesting to note that the power of this residual error is independent of the power of x and y. It decreases with the order M of the canceling filter, and also with the distance between a and a_1, i.e., with the difference of DOA of the desired signal and the jammer.

b) On the other hand, if we prefer to use a MSE criteria for the estimation $\hat{x}(n)$ of $x(n)$, we can keep (17a)–(17c) and replace (17e) by (11c). We obtain the adaptive scheme that we denote *Pisarenko Sinusoidal Canceling* algorithm:

$$e(n) = \tilde{\mathbf{h}}^T(n)\mathbf{z}(n)$$

$$\mathbf{h}^T(n+1) = \tilde{\mathbf{h}}^T(n) - \mu_1 e(n)[\mathbf{z}^H(n) - e^*(n)\tilde{\mathbf{h}}^T(n)] \tag{19a}$$

$$\tilde{\mathbf{h}}(n+1) = \frac{\mathbf{h}(n+1)}{\|\mathbf{h}(n+1)\|} \tag{19b}$$

$$a(n) = -\sum_{k=0}^{N} h_{k,1}(n)h_{k,2}^*(n) \Big/ \sum_{k=0}^{N} |h_{k,2}(n)|^2 \tag{19c}$$

$$\mathbf{w}(n) = [w(n), w(n-1), \cdots, w(n-M+1)]^T$$

$$\mathbf{c}^T(n+1) = \mathbf{c}^T(n) - \mu_2[z_1(n) + \mathbf{c}^T(n)\mathbf{w}(n)]\mathbf{w}^H(n). \tag{19d}$$

In this new context, the solution $\mathbf{c}$ is unique. We obtain an estimate that can be written as $\hat{x}(n) = x(n) + b_f(n) + y_f(n)$, where $b_f(n)$ and $y_f(n)$ are, respectively, a filtered observation noise component and a residual filtered sinusoidal component. Their powers can be computed explicitly when $P = 1$ as

$$E|b_f(n)|^2 = \sigma_b^2 \left(1 + \frac{2M\rho^2|a - a_1|^2}{[2 + M\rho|a - a_1|^2]^2}\right) \tag{20a}$$

and

$$E|y_f(n)|^2 = \frac{2\mu_1^2}{[2 + M\rho|a - a_1|^2]^2} \tag{20b}$$

with $\rho \triangleq \gamma_1^2/\sigma_b^2$. These powers are independent of the power of x. They decrease with the order M of the canceling filter and with the distance between a and a_1. The larger the sinusoidal noise ratio ρ, the weaker the residual filtered sinusoidal component powers, which is typical in adaptive line enhancers.

As in Section III, the two schemes so introduced are also coupled stochastic gradient algorithms. The intermediate quantity that is injected from the first part to the second part of the algorithm is now $a(n)$ and $\sigma_b^2(n)$ (resp. $a(n)$) for the first (resp. second) algorithm.

The problem of their convergence is equivalent to the problem stated in Section III. We obtain the same results in case the gain sequences $\mu_i(n)$ tend to zero (and $\Sigma_n\ \mu_i(n) = +\infty$). This is also proved in the Appendix. For the same reasons, these results cannot be applied in a nonstationary context. In case of constant small gains, the analysis is much more involved and therefore will not be developed in this paper.

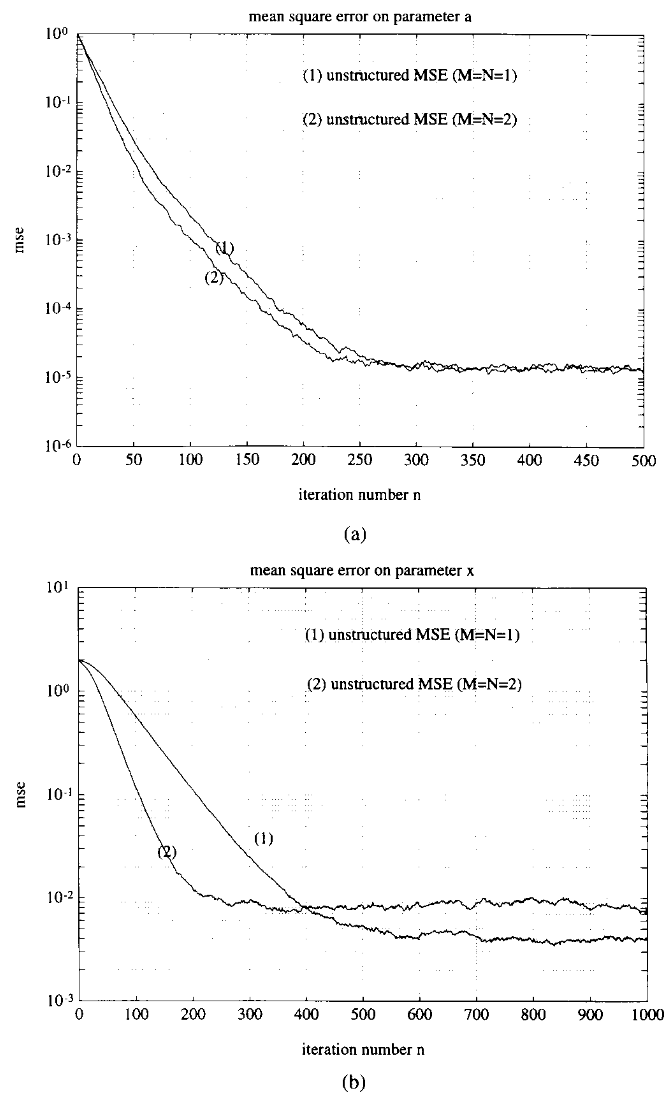

Fig. 2. (a) Learning curves of the first part of the adaptive algorithms, noise free-case. (b) Learning curves of the second part of the adaptive algorithms, noise-free case.

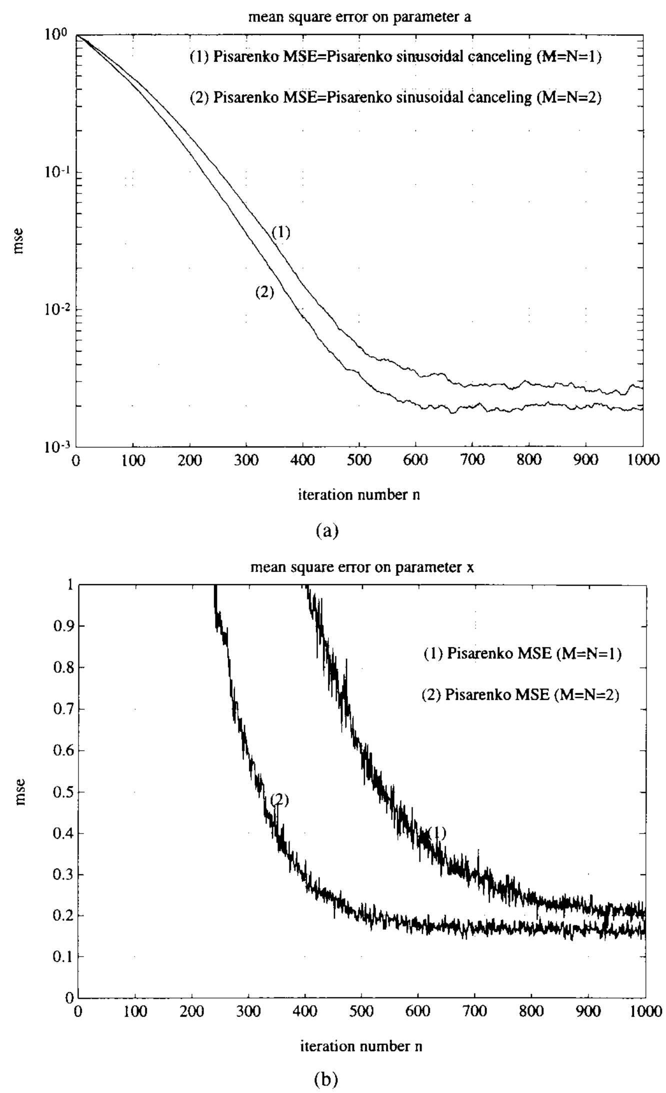

Fig. 3. (a) Learning curves of the first part of the adaptive algorithms, noisy case. (b) Learning curves of the second part of the adaptive algorithms, noisy case.

V. SIMULATION RESULTS

• The first simulation corresponds to the noise-free case where $P = N = M = 1, a = e^{i\theta}$ and $a_1 = e^{i\theta_1}$ with $\theta = -31°$ and $\theta_1 = 61°, f_1 = 0.21, \lambda_1^2 = 2, \sigma^2 = 1$, and $\sigma_b^2 = 10^{-4}$. Fig. 2(a) and (b) shows the learning curves corresponding, respectively, to $E|a(n) - a|^2$ and $E|\hat{x}(n) - x(n)|^2$, averaged over 200 runs. We present the *unstructured MSE* algorithm (11) with $\mu_1 = 0.05$ and $\mu_2 = 0.002$. The asymptotic MSE on the signal x, the value of which is 0.004, is in accordance with those given by (13) despite $\sigma_b^2 = 10^{-4}$.

• The second simulation corresponds to the same values of the parameters of the first simulation but with: $N = M = 2$. We see in Fig. 2(a) and (b) that with the same values of μ_1 and μ_2, the asymptotic MSE on the parameter a and on the signal x are roughly doubled but the speed of convergence is improved.

• The third simulation corresponds to the case of noisy observations with the same values of the parameters of the first simulation but with $\lambda_1^2 = 4$ and $\sigma_b^2 = 0.1$. Fig. 3(a) and (b) shows the learning curves corresponding, respectively, to $E|a(n) - a|^2$ and $E|\hat{x}(n) - x(n)|^2$, averaged over 200 runs. We use the *Pisarenko MSE* algorithm (17) and *Pisarenko Sinusoidal Canceling* algorithm (19) with $\mu_1 = 0.004$ and $\mu_2 = 0.004$, which in this case give very close results. The sinusoidal component in $\hat{x}(n)$ is negligible, as the 512-point periodogram spectral estimate of $z_1(n)$ and $\hat{x}(n)$ shows (see Fig. 4). These results are in accordance with relation (18), which gives $E|b_f(n)|^2 = 0.20$ and with relation (20a), (20b), which gives $E|b_f(n)|^2 = 0.20$ and $E|y_f(n)|^2 = 1.2 \times 10^{-4}$. The difference between 0.20 and the observed asymptotic mean square value of 0.22 is explained by an excess MSE induced by the constant gain adaptive algorithm.

• The fourth simulation corresponds to the same values of the parameters of the third simulation but with $N = M = 2$. We see in Fig. 3(a) and (b) that with the same values of μ_1 and μ_2, the speed of convergence and the asymptotic MSE on the parameter a and the signal x are improved. This latter result is also in accordance with relation (18), which gives $E|b_f(n)|^2 = 0.15$ and with relation (20a), (20b), which gives $E|b_f(n)|^2 = 0.15$ and $E|y_f(n)|^2 = 3.1 \times 10^{-4}$. The difference between 0.15 and the observed asymptotic mean square value of 0.17 is also explained by an excess MSE induced by the use of a constant adaptive gain.

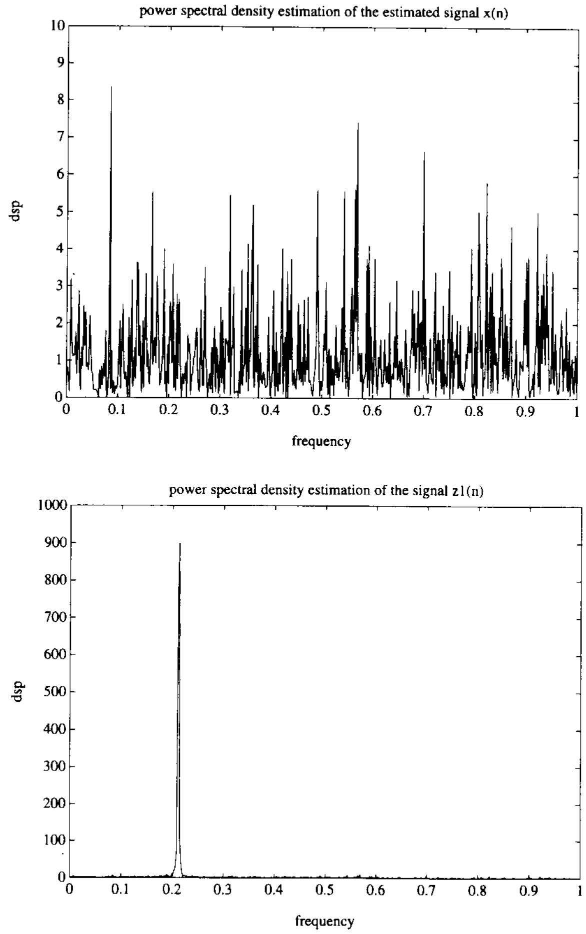

Fig. 4. 512-point periodogram spectral estimate of $z_1(n)$ and $\hat{x}(n)$.

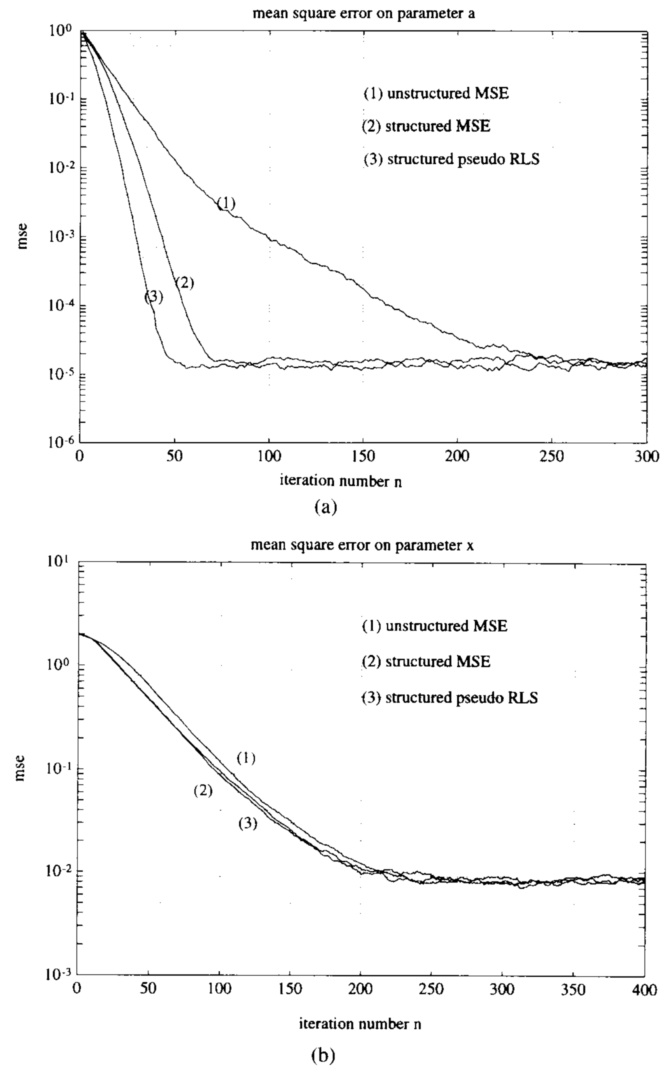

Fig. 5. (a) Learning curves of the first part of the adaptive algorithms, noise free-case. (b) Learning curves of the second part of the adaptive algorithms, noise free-case.

Although the functional (2) can have local spurious minima, we also propose two other algorithms (in the noise-free case, as well as in the case of noisy observations) that use the parameterization (6) of $\mathbf{h}$ because the speed of convergence of these algorithms is faster.

• In the noise-free case, the first algorithm, *denoted structured MSE,* is a stochastic gradient algorithm associated with the minimization of $E|[(1, b_1, \cdots, b_N)\otimes(1, -a)]\mathbf{z}(n)|^2$, where (11a) and (11b) are replaced by

$$\begin{bmatrix} a \\ \mathbf{b} \end{bmatrix}_{n+1} = \begin{bmatrix} a \\ \mathbf{b} \end{bmatrix}_{n} - \mu_1 \mathbf{H}_{a,\mathbf{b}}(n)\mathbf{z}^*(n)\mathbf{z}^T(n)\mathbf{h}_{a,\mathbf{b}}(n)$$

where

$$\begin{aligned} \mathbf{b} &\triangleq (1, b_1, \cdots, b_N)^T, \\ \mathbf{h}_{a,\mathbf{b}} &\triangleq [1, b_1, \cdots, b_N]^T \otimes [1, -a]^T \\ \mathbf{H}_{a,\mathbf{b}} &\triangleq \begin{bmatrix} 0 & \mathbf{0}^T & 1 & \mathbf{b}^H \\ \mathbf{0} & \mathbf{I}_N & \mathbf{0} & -a^*\mathbf{I}_N \end{bmatrix}. \end{aligned}$$

With the parameters of the second simulation, we note in Fig. 5(a) and (b) that if we choose $\mu_1 = 0.15$ and $\mu_2 = 0.002$ in order to have the same MSE on the parameter a and on the signal x, the speed of convergence is improved when compared to the previous algorithm, but the speed of convergence of the second part of the algorithm is hardly modified because the precision given by the first part is adequate.

Then, in order to further increase the speed of convergence we use a pseudo-recursive least squares algorithm denoted *structured pseudo RLS.* In this approach, we minimize the classical expression with respect to a and $\mathbf{b}$:

$$\sum_{k=1}^{n} \lambda^{n-k} \left| z_2(n) - a z_1(n) + \sum_{i=1}^{N} b_i[z_2(n-i) - a z_1(n-i)] \right|^2$$

and this expression with respect to $\mathbf{c}$:

$$\sum_{k=1}^{n} \lambda^{n-k} \left| z_1(n) + \sum_{i=0}^{M-1} c_i[z_2(n-i) - a z_1(n-i)] \right|^2.$$

Using the classical matrix inversion lemma for the parameters $a, \mathbf{b}$ and then for $\mathbf{c}$, we can develop three coupled recursive algorithms by introducing internal variables $p_a, k_a, \mathbf{P}_\mathbf{b}, \mathbf{k}_\mathbf{b}, \mathbf{P}_\mathbf{c}, \mathbf{k}_\mathbf{c}$ ([6, p. 480]). Then, we can derive a global suboptimal recursive algorithm:

$$\begin{aligned} k_a(n) &= \frac{\lambda^{-1} p_a(n-1) u_\mathbf{b}(n)}{1 + \lambda^{-1} u_\mathbf{b}^*(n) p_a(n-1) u_\mathbf{b}(n)} \\ p_a(n) &= \lambda^{-1} p_a(n-1) - \lambda^{-1} k_a(n) u_\mathbf{b}^*(n) p_a(n-1) \\ a(n) &= a(n-1) + k_a(n)[g_\mathbf{b}(n) - u_\mathbf{b}^*(n) a(n-1)] \end{aligned}$$

where

$$u_{\mathbf{b}}(n) \triangleq [1, \mathbf{b}^H(n)]\mathbf{z}_1^*(n), \quad g_{\mathbf{b}}(n) \triangleq [1, \mathbf{b}^T(n)]\mathbf{z}_2(n)$$

and

$$\mathbf{z}(n) \triangleq \begin{bmatrix} \mathbf{z}_2(n) \\ \mathbf{z}_1(n) \end{bmatrix}.$$

$$\mathbf{k}_{\mathbf{b}}(n) = \frac{\lambda^{-1}\mathbf{P}_{\mathbf{b}}(n-1)\mathbf{u}_a(n)}{1+\lambda^{-1}\mathbf{u}_a^H(n)\mathbf{P}_{\mathbf{b}}(n-1)\mathbf{u}_a(n)}$$
$$\mathbf{P}_{\mathbf{b}}(n) = \lambda^{-1}\mathbf{P}_{\mathbf{b}}(n-1) - \lambda^{-1}\mathbf{k}_{\mathbf{b}}(n)\mathbf{u}_a^H(n)\mathbf{P}_{\mathbf{b}}(n-1)$$
$$\mathbf{b}(n) = \mathbf{b}(n-1) - \mathbf{k}_{\mathbf{b}}(n)[w(n) + \mathbf{u}_a^H(n)\mathbf{b}(n-1)]$$

where $\mathbf{u}_a(n) \triangleq \mathbf{w}^*(n-1)$.

$$\mathbf{k}_c(n) = \frac{\lambda^{-1}\mathbf{P}_c(n-1)\mathbf{u}_a(n)}{1+\lambda^{-1}\mathbf{u}_a^H(n)\mathbf{P}_c(n-1)\mathbf{u}_a(n)}$$
$$\mathbf{P}_c(n) = \lambda^{-1}\mathbf{P}_c(n-1) - \lambda^{-1}\mathbf{k}_c(n)\mathbf{u}_a^H(n)\mathbf{P}_c(n-1)$$
$$\mathbf{c}(n) = \mathbf{c}(n-1) - \mathbf{k}_c(n)[z_1(n) + \mathbf{u}_a^H(n)\mathbf{c}(n-1)].$$

In the simulations, the choice $\lambda = 1, p_a = p_{\mathbf{b}} = p_c = 100$ was made in order to yield roughly the same MSE on the parameter a and on the signal x. We see in Fig. 5(a) and (b) that the speed of convergence is improved with respect to the previous two algorithms but, just as in the case of structured MSE algorithm, the speed of convergence of the second part of the algorithm is hardly modified because the precision furnished by the first part is adequate.

• In the case of noisy observations, we can use the parametrization (6) of $\mathbf{h}$ in two other algorithms. We use a stochastic gradient algorithm associated with the minimization of the Rayleigh quotient

$$\frac{E|[(1, b_1, \cdots, b_N) \otimes (1, -a)]\mathbf{z}(n)|^2}{\|(1, b_1, \cdots, b_N) \otimes (1, -a)\|^2}.$$

We derive readily an algorithm denoted *structured Pisarenko MSE:*

$$\begin{bmatrix} a \\ \mathbf{b} \end{bmatrix}_{n+1} = \begin{bmatrix} a \\ \mathbf{b} \end{bmatrix}_n - \mu_1 \mathbf{H}_{a,\mathbf{b}}(n) \frac{e(n)}{\|\mathbf{h}_{a,\mathbf{b}}(n)\|} \cdot \left\{ \mathbf{z}^*(n) - \frac{e^*(n)\mathbf{h}_{a,\mathbf{b}}(n)}{\|\mathbf{h}_{a,\mathbf{b}}(n)\|} \right\}$$

where

$$e(n) \triangleq \frac{\mathbf{h}_{a,\mathbf{b}}^T(n)\mathbf{z}(n)}{\|\mathbf{h}_{a,\mathbf{b}}(n)\|}.$$

With the parameters of the fourth simulation, for the MSE criteria with $\mu_1 = 0.015$ and $\mu_2 = 0.004$, we obtain the same MSE on the parameter a and on the signal x. We note in Fig. 6(a) and (b) that the speed of convergence is improved when compared to the *Pisarenko MSE* algorithm.

Then, in order to simplify the previous algorithm, and in particular to avoid divisions, we can write the Rayleigh ratio under the normalized form: $E|(\mathbf{h}_{\mathbf{b}} \otimes \mathbf{h}_a)\mathbf{z}(n)|^2$ where the unit norm vectors

$$\mathbf{h}_{\mathbf{b}} \triangleq \frac{(1, b_1, \cdots, b_N)^T}{\|(1, b_1, \cdots, b_N)\|} \quad \text{and} \quad \mathbf{h}_a \triangleq \frac{(1, -a)^T}{\|(1, -a)\|}$$

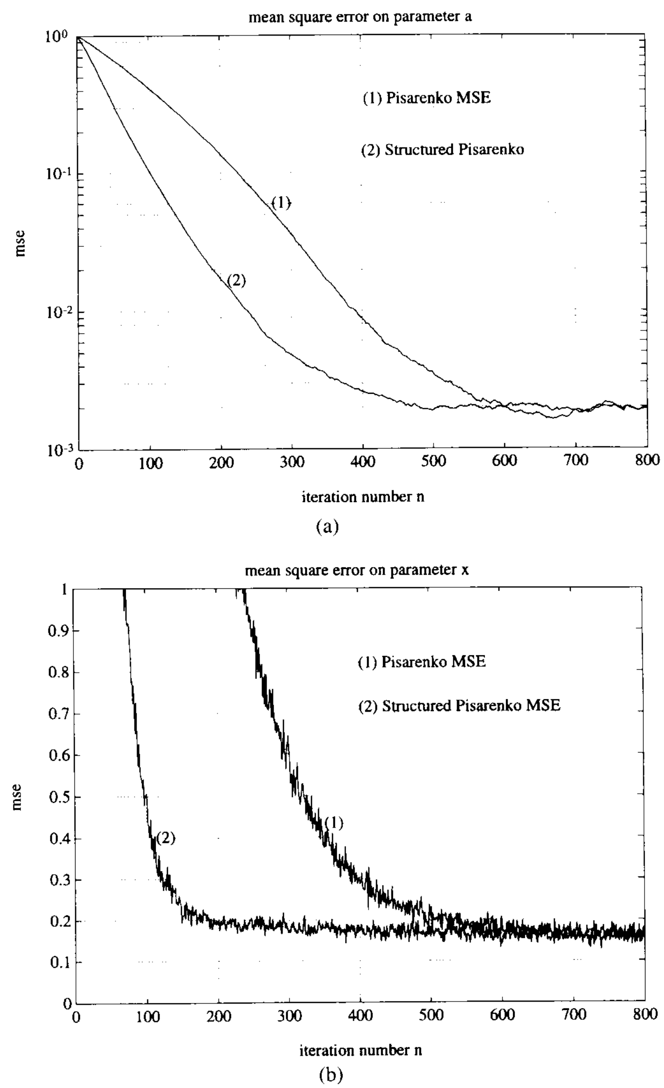

Fig. 6. (a) Learning curves of the first part of the adaptive algorithms, noisy case. (b) Learning curves of the second part of the adaptive algorithms, noisy case.

are taken as the last column of an unitary matrix parameterized by complex Givens rotations [4]. We can deduce a stochastic gradient algorithm [4] that we denote *Rotational structured Pisarenko MSE* with the MSE criteria. With $\mu_1 = 0.01$ and $\mu_2 = 0.004$ we obtain the same MSE on the parameter a and on the signal x. This algorithm presents the same speed of convergence as the *structured Pisarenko MSE* algorithm.

VI. Conclusion

In this paper, we present a new adaptive harmonic jammer canceler, based on the use of two receivers, in order to reconstruct by an adaptive algorithm a desired signal corrupted by purely harmonic jammers with unknown frequencies. We propose some adaptive schemes that reconstruct the desired signal in the noise-free case and in the case of noisy observations. In the latter case, we consider several reconstruction criteria. We discuss the convergence of the coupled introduced algorithms for small constant gains and for decreasing gains. Furthermore, in spite of possible local minima, we present several structural schemes that possess an improved speed of convergence.

APPENDIX

The convergence of the stochastic algorithms presented previously are derived with the aid of the following Lemma:

A. Lemma

Consider the system of differential equations:

$$\frac{d\mathbf{h}}{dt} = f[\mathbf{h}(t)] \quad (21a)$$

$$\frac{d\mathbf{c}}{dt} = \mathbf{A}[\mathbf{h}(t)]\mathbf{c}(t) + \mathbf{b}[\mathbf{h}(t)] \quad (21b)$$

where we assume that:

- any solution $\mathbf{h}(t)$ of (21a) admits a limit $\mathbf{h}_*$ depending on the initial conditions.
- $\mathbf{A}$ and $\mathbf{b}$ are analytic functions of $\mathbf{h}$.
- $\mathbf{A}[\mathbf{h}_*]$ and $\mathbf{b}[\mathbf{h}_*]$ are invariant if $\mathbf{h}_*$ belongs to the set of attractors of (21a).
- there exists a positive real-valued function $W(\mathbf{h}, \mathbf{c})$ for which

$$\mathbf{A}(\mathbf{h})\mathbf{c} + \mathbf{b}(\mathbf{h}) = -\nabla_c W(\mathbf{h}, \mathbf{c}).$$

Then, the stationary points $\mathbf{c}_*$ of $d\mathbf{c}/dt = \mathbf{A}[\mathbf{h}_*]\mathbf{c}(t)+\mathbf{b}[\mathbf{h}_*]$ are globally asymptotically stable for (21a) and (21b).

B. Proof of the Lemma

Since $\mathbf{h}_*$ belongs to the set of attractors of (18a), we have by linearizing (21a) around $\mathbf{h}_*$ for t large enough:

$$\|\mathbf{h}(t) - \mathbf{h}_*\| \leq k_1 e^{-\lambda_1 t} \text{ and } \left\|\frac{d\mathbf{h}}{dt}\right\| \leq k_2 e^{\lambda_2 t} \text{ with } \lambda_1 \text{ and } \lambda_2 \geq 0. \quad (22)$$

Consider the Lyapounov function $\phi(t) \triangleq W(\mathbf{h}, \mathbf{c}) \geq 0$.

$$\frac{d\phi(t)}{dt} = \frac{d\mathbf{h}^T}{dt}\nabla_{\mathbf{h}} W(\mathbf{h}(t), \mathbf{c}(t)) + \frac{d\mathbf{c}^T}{dt}\nabla_{\mathbf{c}} W(\mathbf{h}(t), \mathbf{c}(t)).$$

By hypothesis of (21b) we have

$$\frac{d\mathbf{c}^T}{dt}[\nabla_{\mathbf{c}} W(\mathbf{h}(t), \mathbf{c}(t))] = -\|\nabla_{\mathbf{c}} W(\mathbf{h}(t), \mathbf{c}(t))\|^2$$

and if we prove that $\mathbf{c}(t)$ is bounded $\|\nabla_{\mathbf{h}} W(\mathbf{h}(t), \mathbf{c}(t)\|$ is also bounded because this gradient is analytic in $\mathbf{h}$ and quadratic in $\mathbf{c}$. So

$$\left|\frac{d\mathbf{h}^T}{dt}\nabla_{\mathbf{h}} W(\mathbf{h}(t), \mathbf{c}(t))\right| \leq \left\|\frac{d\mathbf{h}}{dt}\right\| \|\nabla_{\mathbf{h}} W(\mathbf{h}(t), \mathbf{c}(t))\| \leq k_3 e^{-\lambda_2 t}$$

thanks to (22).

Consequently

$$\frac{d\phi(t)}{dt} \leq k_3 e^{-\lambda_2 t} - \|\nabla_{\mathbf{c}} W(\mathbf{h}(t), \mathbf{c}(t))\|^2.$$

Then, $\phi(t) + (k_3/\lambda_2)e^{-\lambda_2 t}$ is a decreasing function of t, so $\lim_{t\to\infty} \phi(t)$ exists, which in turn implies that

$$\lim_{t\to\infty} \frac{d\phi(t)}{dt} = 0 \quad \text{and} \quad \lim_{t\to\infty} \|\nabla_{\mathbf{c}} W(\mathbf{h}(t), \mathbf{c}(t))\|^2 = 0.$$

Therefore, the stationary points $\mathbf{c}_*$ of (21b) are globally asymptotically stable for (21a) and (21b).

We must now prove that $\mathbf{c}(t)$ is bounded. We can consider (21b) as an almost-constant linear equation that we can rewrite as:

$$\frac{d\mathbf{c}}{dt} = [\mathbf{A} + \mathbf{A}'(t)]\mathbf{c}(t) + [\mathbf{b} + \mathbf{b}'(t)] \quad (23)$$

with $\mathbf{A} \triangleq \mathbf{A}(\mathbf{h}_*), \mathbf{b} \triangleq \mathbf{b}(\mathbf{h}_*)$. We have for t large enough

$$\|\mathbf{A}'(t)]\| \leq k_1' e^{-\lambda_1' t} \text{ and } \|\mathbf{b}'(t)]\| \leq k_1'' e^{-\lambda_1'' t} \text{ with } \lambda_1' \text{ and } \lambda_1'' \geq 0. \quad (24)$$

The differential equation $d\mathbf{c}/dt = \mathbf{A}\mathbf{c}(t) + \mathbf{b}$ can be written $d\mathbf{c}/dt = -\nabla_{\mathbf{c}} W[\mathbf{h}_*, \mathbf{c}(t)]$, since $\mathbf{A}\mathbf{c}(t) + \mathbf{b} = -\nabla_{\mathbf{c}} W[\mathbf{h}_*, \mathbf{c}(t)]$. Then, $\mathbf{A}$ is stable and their solutions $\mathbf{c}_1(t)$ converge to the stationary points of this equation. These solutions are bounded; so that the solutions of $d\mathbf{c}/dt = \mathbf{A}\mathbf{c}(t)$ are also bounded.

Let $\mathbf{c}_2(t)$ the solution of $d\mathbf{c}/dt = \mathbf{A}\mathbf{c}(t)$, so that $\mathbf{c}_2(0) = \mathbf{c}(0) - \mathbf{c}_1(0)$, where $\mathbf{c}(t)$ is a solution of (23), and let $\mathbf{u}(t) \triangleq \mathbf{c}(t) - \mathbf{c}_1(t)$. $\mathbf{u}(t)$ is solution of:

$$\frac{d\mathbf{u}}{dt} = \mathbf{A}\mathbf{u}(t) + [\mathbf{A}'(t)\mathbf{c}(t) + \mathbf{b}'(t)].$$

Due to a result of [1, relation (9), p. 14], we have

$$\mathbf{u}(t) = \mathbf{c}_2(t) + \int_0^t \mathbf{C}_2(t-\tau)[\mathbf{A}'(\tau)\mathbf{c}(\tau) + \mathbf{b}'(\tau)]\, d\tau$$

in which $\mathbf{C}_2(t)$ is solution of $d\mathbf{C}_2/dt = \mathbf{A}\mathbf{C}_2(t)$ with $\mathbf{C}_2(0) = \mathbf{I}$. Then

$$\|\mathbf{u}(t)\| \leq \|\mathbf{c}_2(t)\| + \int_0^t \|\mathbf{C}_2(t-\tau)\|\|\|\mathbf{A}'(\tau)\|\|\|\mathbf{u}(\tau) + \mathbf{c}_1(t)\|\, d\tau + \int_0^t \|\mathbf{C}_2(t-\tau)\|\|\|\mathbf{b}'(t)\|\, d\tau.$$

Since $\int_0^t \|\mathbf{A}'(\tau)\|\, d\tau$ and $\int_0^t \|\mathbf{b}'(\tau)\|\, d\tau$ are bounded by (24), we have:

$$\|\mathbf{u}(t)\| \leq k_1 + k_2 \int_0^t \|\mathbf{A}'(\tau)\|\|\mathbf{u}(\tau)\|\, d\tau.$$

By the Gronwall Lemma ([1, p. 35]), we have

$$\|\mathbf{u}(t)\| \leq k_1 \exp\left[k_2 \int_0^t \|\mathbf{A}'(\tau)\|\, d\tau\right]$$

and therefore, $\mathbf{u}(t)$, and then, $\mathbf{c}(t)$ are bounded.

C. Convergence of the Stochastic Algorithms

Now, we show that the three different versions of coupled stochastic gradient algorithms that we have presented in Sections III and IV have an associated ODE of the form (21a), (21b).

The coupled stochastic gradient algorithm can be written under this form:

$$\begin{bmatrix}\mathbf{h}^T \\ \mathbf{c}^T\end{bmatrix}_{n+1} = \begin{bmatrix}\mathbf{h}^T \\ \mathbf{c}^T\end{bmatrix}_n - \mu_n \begin{bmatrix}\nabla_{\mathbf{h}}^H V_{\mathbf{h}}(n) \\ \nabla_{\mathbf{c}}^H V_{\mathbf{h},\mathbf{c}}(n)\end{bmatrix} \quad (25)$$

and the associate ODE is

$$\frac{d\mathbf{h}^T}{dt} = -E[\nabla_{\mathbf{h}}^H V_{\mathbf{h}}(t)] \quad (26a)$$

$$\frac{d\mathbf{c}^T}{dt} = -E[\nabla_{\mathbf{c}}^h V_{\mathbf{h},\mathbf{c}}(t)] \quad (26b)$$

where $\nabla_z V$ denotes the vector of components $1/2[(\partial V/\partial x) - i(\partial V/\partial y)]$ with $z \triangleq x + iy$.

We have, according to the algorithms

$$V_{\mathbf{h}}(n) = |z_2(n) + \mathbf{h}_0^T(n)\mathbf{z}_0(n)|^2 \quad \text{or} \quad \frac{|\mathbf{h}^T(n)\mathbf{z}(n)|^2}{|\mathbf{h}(n)|^2}$$

and

$$V_{\mathbf{h},\mathbf{c}}(n) = |z_1(n) + \mathbf{c}^T(n)\mathbf{w}(n)|^2 \quad \text{or}$$
$$\mathbf{h}_{a,\mathbf{c}}^H(n)[\mathbf{z}(n)\mathbf{z}^H(n) - |\tilde{\mathbf{h}}^T(n)\mathbf{z}(n)|^2\mathbf{I}_{2M}]\mathbf{h}_{a,\mathbf{c}}(n).$$

• Since $\nabla_{\mathbf{h}} V_{\mathbf{h}}(t)$ is the derivative of a positive gradient field, the set of all stationary points of (26a) are globally asymptotically stable for that equation. When $N = P$, there is an unique stationary point $\mathbf{h}_*$ that is globally asymptotically stable, but when $N > P$, we can prove that any solution $\mathbf{h}(t)$ of (26a) converges to one point $\mathbf{h}_*$ among the stationary points of (26a).

In the noise-free case, this is based on the following property:

Let $\mathbf{R}$ be a $n \times n$ singular covariance matrix and of rank strictly less than $n-1$. Consider the ODE associated with the stochastic gradient algorithm derived from the minimization of $\mathbf{h}^H\mathbf{R}\mathbf{h}$ with respect to $\mathbf{h}$, with the constraint that the first component of $\mathbf{h}$ is one. The solution of this ODE converges exponentially to one point $\mathbf{h}_*$ among the set of the stationary points.

In addition, in the noisy case, this is based on the following property. Consider the same covariance matrix $\mathbf{R}$ and the ODE associated with the stochastic gradient algorithm derived from the minimization of $\mathbf{h}^H\mathbf{R}\mathbf{h}$ w.r.t $\mathbf{h}$, with the constraint $\|\mathbf{h}\| = 1$.

$$\frac{d\mathbf{h}}{dt} = -[\mathbf{R} - (\mathbf{h}^H\mathbf{R}\mathbf{h})\mathbf{I}]\mathbf{h} \text{ with } \|\mathbf{h}(0)\| = 1.$$

The solution of this ODE also converges exponentially to one point $\mathbf{h}_*$ among the set of the stationary points.

• Equation (26b) is a first-order linear inhomogeneous differential equation where the coefficients depend on $\mathbf{h}(t)$ in an analytical manner.

• Since

$$a(t) = -\sum_{k=0}^{N} h_{k,1}(t)h_{k,2}^*(t) \Big/ \sum_{k=0}^{N} |h_{k,2}(t)|^2$$

and $|\tilde{\mathbf{h}}^T(t)\mathbf{z}(t)|^2 = \sigma_b^2(t), \mathbf{A}[\mathbf{h}_*]$ and $\mathbf{b}[\mathbf{h}_*]$ are invariant if $\mathbf{h}_*$ belongs to the set of attractors of (26a).

• Equation (21b) is of the form $d\mathbf{c}/dt = -\nabla_{\mathbf{c}} W[\mathbf{h}(t), \mathbf{c}(t)]$ with $W(\mathbf{h}, \mathbf{c}) = E(V_{\mathbf{h},\mathbf{c}})$.

Since the associated ODE admits a globally asymptotically stable set of attractors $\mathbf{c}_*$, the stochastic algorithms (11), (17), and (19) converge almost surely to one of its points, provided we are in the stationary situation with decreasing sequence gains [9].

ACKNOWLEDGMENT

The author is grateful to P. Loubaton for very useful discussions.

REFERENCES

[1] R. Bellman, *Stability of Differential Equations.* New York: McGraw-Hill, 1953.
[2] A. Benveniste, M. Metivier, and P. Priouret, *Adaptive Algorithms and Stochastic Approximation.* New York: Springer-Verlag, 1990.
[3] R. R. Bitmead, "Convergence in distribution of LMS-Type adaptive parameter estimates," *IEEE Trans. Automat. Contr.*, vol. AC-28, no. 1, pp. 54–60, Jan. 1983.
[4] J. P. Delmas, "A complex adaptive eigensubspace algorithm for DOA or frequency estimation and tracking," in *EUSIPCO* Sept. 1992, pp. 657–660.
[5] J. R. Glover, "Adaptive noise canceling applied to sinusoidal interferences," *IEEE Trans. Acoust., Speech, Signal Processing*, vol. ASSP-25, no. 6, pp. 484–491, Dec. 1977.
[6] S. Haykin, *Adaptive Filter Theory.* Englewood Cliffs, NJ: Prentice-Hall, 1991.
[7] F. M. Hsu and A. A. Giordano, "Digital whitening techniques for improving spread spectrum communications performance in the presence of narrow band jamming and interference," *IEEE Trans. Commun.*, vol. COM-26, pp. 209–216, Feb. 1978.
[8] J. W. Ketchum and J. G. Proakis, "Adaptive algorithms for estimating and suppressing narrow band interference in PN spread spectrum systems," *IEEE Trans. Commun.*, vol. COM-30, pp. 913–924, May 1982.
[9] H. J. Kushner and D. S. Clark, "Stochastic approximation methods for constrained and unconstrained systems," *Applied Mathematical Science.* New York: Springer Verlag, 1978.
[10] L. Ljung and T. Söderström, *Theory and Practice of Recursive Identification.* Cambridge, MA: MIT Press, 1983.
[11] V. F. Pisarenko, "The retrieval of harmonics from a covariance function," *Geophys. J. Roy. Astron. Soc.*, 1973, pp. 347–366.
[12] V. U. Reddy, B. Egardt, and T. Kailath, "Least squares type algorithm for adaptive implementation of Pisarenko's harmonic retrieval method," *IEEE Trans. Acoust., Speech, Signal Processing*, vol. ASSP-30, no. 3, pp. 399–405, June 1982.
[13] P. A. Regalia, "An adaptive unit norm filter with applications to signal analysis and Karhunen–Love transformations," *IEEE Trans. Circuits Syst.*, vol. 37, no. 5, pp. 646–649, May 1990.
[14] D. C. Rife and R. R. Boorstyn, "Single tone parameter estimation from discrete time observation," *IEEE Trans. Inform. Theory*, vol. IT-20, pp. 591–598, Sept. 1974.
[15] M. Wax, T.-J. Shan, and T. Kailath, "Spatio-temporal spectral analysis by eigenstructure methods," *IEEE Acoust., Speech, Signal Processing*, vol. 32, no. 4, pp. 817–827, Aug. 1984.
[16] B. Widrow *et al.*, "Adaptive noise canceling: Principles and applications," *Proc. IEEE*, vol. 63, pp. 1692–1716, Dec. 1975.
[17] Y. Yoganoindam, V. U. Reddy, and T. Kailath, "Performance analysis of the adaptive line enhancer for sinusoidal signals in broad band noise," *IEEE Trans. Acoust., Speech, Signal Processing*, vol. 36, no. 11, Nov. 1988.

Jean Pierre Delmas was born in France in 1950. He received the Engineer degree from the Ecole Centrale de Lyon in 1973 and the Certificat d'études supérieures from the Ecole Nationale Supérieure des Télécommunications in 1982.

Since 1980, he has been with the Institut National des Télécommunications, where he is currently Maitre de conférences. His research interests are in statistical signal processing.