ON BIAS REMOVAL AND UNIT NORM CONSTRAINTS IN EQUATION ERROR ADAPTIVE IIR FILTERS

S.C. Douglas[1] *and M. Rupp*[2]

[1]Dept. of Electrical Engineering, University of Utah, Salt Lake City, UT 84112 USA

[2]Wireless Research Laboratory, Lucent Technologies, Holmdel, NJ 07733 USA

ABSTRACT

In this paper, we develop two simple gradient-based algorithms for unbiased adaptive IIR filtering in the presence of zero-mean output noise. The algorithms are derived according to a constrained minimization problem that is known to have a unique solution and are generalizations of bias removal techniques for equation-error-based filters for uncorrelated output noises. We propose simple methods for estimating the noise correlation statistics within the algorithm. Our stochastic analyses of these algorithms yield necessary conditions on the step sizes for the stability of the mean values of the coefficients. In addition, we give a more accurate mean-square analysis of one of the algorithms assuming jointly Gaussian input and desired response signals. Simulations indicate that the algorithms can achieve unbiased parameter estimates at least as accurately as other, more-complex techniques.

1. INTRODUCTION

The potential advantages of an adaptive infinite-impulse-response (IIR) filter over the more-conventional adaptive finite-impulse-response (FIR) filter have long been recognized [1, 2, 3]. However, the convergence and stability properties of the simplest gradient-based adaptive IIR filters often prevent their use in real-world systems. Equation-error adaptive IIR filters have predictable convergence characteristics, but the parameter estimates at convergence are biased in the presence of any output noise. Output-error adaptive IIR filters implemented in direct form can produce unbiased parameter estimates, but their ability to converge to these estimates is hampered by strictly positive real (SPR) constraints on the autoregressive portion of the unknown system. A desirable direct-form adaptive IIR filter would combine the convergence properties of the equation-error-based filters with the ability to produce unbiased coefficient estimates possessed by output-error-based filters.

To this end, researchers have modified the equation error adaptive algorithm using *bias removal* or *constrained minimization* techniques. Bias removal methods compensate for the bias in the autoregressive parameter estimates by introducing a correction term within the coefficient updates. The simplest algorithm employing this concept is [5]

$$\mathbf{a}_{i+1} = \mathbf{a}_i + \mu e(i)\mathbf{d}_{i-1} + \mu\widehat{\sigma_v^2}(i)\mathbf{a}_i \quad (1)$$
$$\mathbf{b}_{i+1} = \mathbf{b}_i + \mu e(i)\mathbf{x}_i \quad (2)$$
$$e(i) = d(i) - \mathbf{a}_i^T\mathbf{d}_{i-1} - \mathbf{b}_i^T\mathbf{x}_i, \quad (3)$$

[0]This research was supported by NSF Grant No. MIP-9501680.

where $\mathbf{a}_i = [a_1(i)\, a_2(i) \cdots a_N]^T$ is the autoregressive coefficient vector, $\mathbf{b}_i = [b_0(i)\, b_1(i) \cdots b_{M-1}(i)]^T$ is the moving average coefficient vector, $\mathbf{d}_i = [d(i) \cdots d(i-N+1)]^T$ and $\mathbf{x}_i = [x(i) \cdots x(i-M+1)]^T$ are the desired response and input signal vectors, respectively, $\widehat{\sigma_v^2}(i)$ is an estimate of the observation noise power, $e(i)$ is the equation error, and μ is the step size. This adaptive filter assumes that the desired response signal is generated from an IIR filter with parameter vectors $\mathbf{a}$ and $\mathbf{b}$ whose output $y(i)$ is corrupted by an additive zero mean uncorrelated noise signal $v(i)$ such that

$$d(i) = y(i) + v(i) \quad (4)$$
$$y(i) = \mathbf{a}^T\mathbf{y}_{i-1} + \mathbf{b}^T\mathbf{x}_i, \quad (5)$$

where $\mathbf{y}_i = [y(i) \cdots y(i-N+1)]^T$. If the estimate of the noise variance $\widehat{\sigma_v^2}(i)$ equals the true noise variance $\sigma_v^2(i)$, the adaptive filter in (1)–(3) can be shown to give unbiased estimates in the mean at convergence if $v(i)$ is uncorrelated. However, in most cases, the noise variance is unknown and thus must be estimated. More recently, other modifications of this idea have been introduced [4], including those that are loosely grouped under the label of instrumental variable (IV) techniques [6]. Although IV techniques can be generalized to include correlated observation noises, these methods require a prewhitening filter to decorrelate the error residuals, thus complicating the system.

By contrast, constrained minimization techniques employ a cost function that is designed to have a single unique minimum at the optimum parameter estimates. These techniques include Steiglitz-McBride methods [3] and unit-norm constraint methods [7], in which the error signal $e_c(i)$ is defined as

$$e_c(i) = \alpha_0(i)d(i) + \underline{\alpha}_i^T\mathbf{d}_{i-1} + \underline{\beta}_i^T\mathbf{x}_i, \quad (6)$$

where $\alpha_0(i)$, $\underline{\alpha}_i = [\alpha_1(i)\, \alpha_2(i) \cdots \alpha_N(i)]^T$, and $\underline{\beta}_i = [\beta_0(i)\, \beta_1(i) \cdots \beta_{M-1}(i)]^T$ contain the parameters of the system. These parameters are updated using a gradient search on the mean-squared error $E[e_c^2(i)]$ whereby a unit-norm constraint on $[\alpha_0(i)\ \underline{\alpha}_i^T]$ is imposed [7]. The original parameters can then be recovered via the relations

$$\underline{\alpha}_i = -\alpha_0(i)\mathbf{a}_i \quad (7)$$
$$\underline{\beta}_i = -\alpha_0(i)\mathbf{b}_i. \quad (8)$$

In [8], a stochastic gradient algorithm is given for adjusting the parameters $\alpha_0(i)$, $\underline{\alpha}_i$, and $\underline{\beta}_i$ in which the constraint is imposed by dividing $\alpha_0(i)$ and $\underline{\alpha}_i$ by $\sqrt{\alpha_0^2(i) + \underline{\alpha}_i^T\underline{\alpha}_i}$ at infrequent intervals.

1058-6393/97 $10.00 © 1997 IEEE

Recently, adaptive IIR filters based on total least squares (TLS) concepts have been proposed [9, 10]. Because the algorithm in [9] is a true TLS formulation, it is guaranteed to converge, but its overall complexity is much greater than those of gradient-based techniques. The approach in [10] is not guaranteed to converge, and its stability behavior has not been analyzed.

In this paper, we develop two new stochastic gradient adaptive filters based on a constrained minimization cost function. Unlike the algorithms described in [4, 5, 8, 9, 10], our adaptive filters can be employed in cases where the observation noise signal $v(i)$ is correlated. Since our algorithms adapt the coefficient vectors $\mathbf{a}_i$ and $\mathbf{b}_i$ directly, we avoid the costly square roots and numerous divisions inherent in the algorithm of [8]. We show how the resulting algorithms can be related to the bias removal technique of [5] and to the anti-Hebbian adaptation algorithm in [10] in the white noise case. As the correlation statistics of $v(i)$ may not be known, we provide simple methods for estimating these statistics. Our stochastic analyses of the algorithms indicate that they produce unbiased estimates of the true parameter values for system identification tasks. Simulations show that these adaptive filters perform as designed, producing unbiased parameter estimates at convergence.

2. ALGORITHM DERIVATIONS

For our derivation, we define the equation error cost function as

$$J_e(i) = \frac{1}{2}E\left[\left(d(i) - \mathbf{a}_i^T\mathbf{d}_{i-1} - \mathbf{b}_i^T\mathbf{x}_i\right)^2\right]. \tag{9}$$

For the constrained minimization problem, define $J(i)$ as

$$J(i) = \frac{1}{2}E\left[\left(\alpha_0(i)d(i) + \underline{\alpha}_i^T\mathbf{d}_{i-1} + \underline{\beta}_i^T\mathbf{x}_i\right)^2\right]. \tag{10}$$

By substituting (7)–(8) into (10), we have that

$$J(i) = \alpha_0^2(i)J_e(i). \tag{11}$$

So long as $\alpha_0(i) \neq 0$, minimization of $J_e(i)$ with respect to $\mathbf{a}_i$ and $\mathbf{b}_i$ is equivalent to minimization of $J(i)$ with respect to $\underline{\alpha}_i$ and $\underline{\beta}_i$, but the resulting solutions are biased due to the noise signal $v(i)$ in $d(i)$. If the noiseless observation signal $y(i)$ were available, minimizing the criterion

$$J_o(i) = \frac{1}{2}E\left[\left(d(i) - \mathbf{a}_i^T\mathbf{y}_{i-1} - \mathbf{b}_i^T\mathbf{x}_i.\right)^2\right] \tag{12}$$

with respect to $\mathbf{a}_i$ and $\mathbf{b}_i$ produces unbiased estimates.

If the second-order joint statistics of $x(i)$ and $y(i)$ are well-defined, it can be shown via (9) and (12) that [7]

$$J(i) = \alpha_0^2(i)J_o(i) + \frac{1}{2}\sigma_v^2(i)C(i), \tag{13}$$

where we have defined $C(i)$ as

$$\begin{aligned} C(i) &\triangleq \alpha_0^2(i) + 2\alpha_0(i)\mathbf{p}_{vv}^T\underline{\alpha}_i + \underline{\alpha}_i^T\mathbf{R}_{vv}\underline{\alpha}_i & (14)\\ &= \alpha_0^2(i)\left(1 - 2\mathbf{p}_{vv}^T\mathbf{a}_i + \mathbf{a}_i^T\mathbf{R}_{vv}\mathbf{a}_i\right) & (15)\end{aligned}$$

and $\mathbf{p}_{vv}$ and $\mathbf{R}_{vv}$ are the normalized autocorrelation vector and matrix of the noise defined as

$$\begin{aligned} [\mathbf{p}_{vv}]_m &= E[v(i)v(i-m)]/\sigma_v^2(i) & (16)\\ [\mathbf{R}_{vv}]_{mn} &= E[v(i-m)v(i-n)]/\sigma_v^2(i) & (17)\end{aligned}$$

for $\{m, n\} = \{1, 2, \ldots, N\}$, respectively. From (13), $J(i)$ and $J_o(i)$ have the same global minimum if $C(i)$ is held constant. Thus, we adapt $\mathbf{a}_i$ and $\mathbf{b}_i$ to solve:

$$\text{minimize } J(i) \qquad \text{subject to } C(i) = 1. \tag{18}$$

The constraint in (18) ensures that the algorithm achieves the proper estimates of the unknown system. Moreover, an algorithm that solves (18) does not require knowledge of the noise variance $\sigma_v^2(i)$ for implementation purposes. For white noise, (18) simplifies to the constraint used in [8].

Substituting (15) into (18), we can derive the relationship between $\alpha_0(i)$ and $\mathbf{a}_i$ as

$$\alpha_0^2(i) = \frac{1}{1 - 2\mathbf{p}_{vv}^T\mathbf{a}_i + \mathbf{a}_i^T\mathbf{R}_{vv}\mathbf{a}_i}. \tag{19}$$

2.1. Algorithm #1

To derive our first gradient descent algorithm, we can differentiate $J(i)$ in (11) directly with respect to $\mathbf{a}_i$ and $\mathbf{b}_i$ given (19). Differentiating $J(i)$ with respect to $\mathbf{b}_i$ gives

$$\frac{\partial J(i)}{\partial \mathbf{b}_i} = -E\left[\alpha_0^2(i)e(i)\mathbf{x}_i\right]. \tag{20}$$

Differentiating $J(i)$ with respect to $\mathbf{a}_i$ gives

$$\frac{\partial J(i)}{\partial \mathbf{a}_i} = E\left[\alpha_0^2(i)e(i)\mathbf{d}_{i-1} + e^2(i)\frac{1}{2}\frac{\partial \alpha_0^2(i)}{\partial \mathbf{a}_i}\right]. \tag{21}$$

Using (19), the derivative within the second term is

$$\frac{\partial \alpha_0^2(i)}{\partial \mathbf{a}_i} = \frac{2\alpha_0^2(i)}{1 - 2\mathbf{p}_{vv}^T\mathbf{a}_i + \mathbf{a}_i^T\mathbf{R}_{vv}\mathbf{a}_i}(\mathbf{p}_{vv} - \mathbf{R}_{vv}\mathbf{a}_i) \tag{22}$$

Substituting (22) into (21) and using the standard instantaneous gradient approximation, we arrive at the following parameter updates for $\mathbf{a}_i$ and $\mathbf{b}_i$:

$$\begin{aligned} \mathbf{a}_{i+1} &= \mathbf{a}_i + \mu e(i)\left(\mathbf{d}_{i-1} + \frac{e(i)(\mathbf{R}_{vv}\mathbf{a}_i - \mathbf{p}_{vv})}{1 - 2\mathbf{p}_{vv}^T\mathbf{a}_i + \mathbf{a}_i^T\mathbf{R}_{vv}\mathbf{a}_i}\right) & (23)\\ \mathbf{b}_{i+1} &= \mathbf{b}_i + \mu e(i)\mathbf{x}_i, & (24)\end{aligned}$$

where $\alpha_0^2(i)$ is absorbed into μ as both are always positive.

The algorithm in (23)–(24) is a modified equation error scheme in which a correction term is added to the update of $\mathbf{a}_i$. If $v(i)$ is uncorrelated, this update is

$$\mathbf{a}_{i+1} = \mathbf{a}_i + \mu e(i)\left(\mathbf{d}_{i-1} + \frac{e(i)}{1 + \mathbf{a}_i^T\mathbf{a}_i}\mathbf{a}_i\right). \tag{25}$$

Note that the quantity $E[e^2(i)/(1 + \mathbf{a}_i^T\mathbf{a}_i)]$ is equal to the noise variance $\sigma_v^2(i)$ as $\mathbf{a}_i \to \mathbf{a}$ and $\mathbf{b}_i \to \mathbf{b}$ for a system that is not undermodeled. Thus, (25) is analogous to (1) where $\widehat{\sigma_v^2}(i) = e^2(i)/(1 + \mathbf{a}_i^T\mathbf{a}_i)$ is the estimate of $\sigma_v^2(i)$.

2.2. Algorithm #2

We can derive a second gradient descent algorithm by differentiating $J(i)$ with respect to $\underline{\alpha}_i$ and $\underline{\beta}_i$, where $\alpha_0(i)$ is expressed in terms of $\underline{\alpha}_i$ and $\underline{\beta}_i$. Then, using (7)–(8), we express the resulting updates in terms of $\mathbf{a}_i$ and $\mathbf{b}_i$.

Differentiating both sides of (14) with respect to $\underline{\alpha}_i$ gives

$$\frac{\partial \alpha_0^2(i)}{\partial \underline{\alpha}_i} + 2\frac{\partial \alpha_0(i)}{\partial \underline{\alpha}_i}\mathbf{p}_{vv}^T\underline{\alpha}_i + 2\alpha_0(i)\mathbf{p}_{vv} + 2\mathbf{R}_{vv}\underline{\alpha}_i = 0 \tag{26}$$

or, equivalently,

$$\frac{\partial\alpha_0(i)}{\partial\underline{\alpha}_i} = -\frac{\alpha_0(i)\mathbf{p}_{vv} + \mathbf{R}_{vv}\underline{\alpha}_i}{\alpha_0(i) + \mathbf{p}_{vv}^T\underline{\alpha}_i}. \tag{27}$$

By substituting $\underline{\alpha}_i = -\alpha_0(i)\mathbf{a}_i$ into (27), we arrive at

$$\frac{\partial\alpha_0(i)}{\partial\underline{\alpha}_i} = \frac{\mathbf{R}_{vv}\mathbf{a}_i - \mathbf{p}_{vv}}{1 - \mathbf{p}_{vv}^T\mathbf{a}_i}. \tag{28}$$

Taking the derivative of $J(i)$ with respect to $\underline{\beta}_i$ gives

$$\frac{\partial J(i)}{\partial\underline{\beta}_i} = E\left[\alpha_0(i)e(i)\mathbf{x}_i\right]. \tag{29}$$

Differentiating $J(i)$ with respect to $\underline{\alpha}_i$ gives

$$\frac{\partial J(i)}{\partial\underline{\alpha}_i} = E\left[\alpha_0(i)e(i)\left(\mathbf{d}_{i-1} + d(i)\frac{\partial\alpha_0(i)}{\partial\underline{\alpha}_i}\right)\right]. \tag{30}$$

Now, we can substitute (28) for $\partial\alpha_0(i)/\partial\underline{\alpha}_i$ on the RHS of (30), giving

$$\frac{\partial J(i)}{\partial\underline{\alpha}_i} = E\left[\alpha_0(i)e(i)\left(\mathbf{d}_{i-1} + d(i)\frac{\mathbf{R}_{vv}\mathbf{a}_i - \mathbf{p}_{vv}}{1 - \mathbf{p}_{vv}^T\mathbf{a}_i}\right)\right]. \tag{31}$$

Equations (29) and (31) are the gradients of the cost function with respect to $\underline{\alpha}_i$ and $\underline{\beta}_i$. Since the gradients are already in terms of $\mathbf{a}_i$ and $\mathbf{b}_i$, we can immediately specify the coefficient updates using (7)–(8) as

$$\mathbf{a}_{i+1} = \mathbf{a}_i + \mu e(i)\left(\mathbf{d}_{i-1} + \frac{d(i)(\mathbf{R}_{vv}\mathbf{a}_i - \mathbf{p}_{vv})}{1 - \mathbf{p}_{vv}^T\mathbf{a}_i}\right) \tag{32}$$

$$\mathbf{b}_{i+1} = \mathbf{b}_i + \mu e(i)\mathbf{x}_i. \tag{33}$$

Similar to Algorithm #1, (32)–(33) is a modified equation error scheme with a correction term in the update for $\mathbf{a}_i$. If $v(i)$ is uncorrelated, this update is

$$\mathbf{a}_{i+1} = \mathbf{a}_i + \mu e(i)(\mathbf{d}_{i-1} + d(i)\mathbf{a}_i). \tag{34}$$

For $\mathbf{a}_i = \mathbf{a}$ and $\mathbf{b}_i = \mathbf{b}$, the quantity $E[d(i)e(i)]$ is equal to the noise variance $\sigma_v^2(i)$; thus, $\widehat{\sigma_v^2}(i) = d(i)e(i)$ is an estimate of the noise variance. The algorithm in (33)–(34) is different from that in [10] as our algorithm uses the standard equation-error update for $\mathbf{b}_i$.

3. IMPLEMENTATION ISSUES

The two algorithms we have derived require knowledge of the correlation statistics of the output noise $v(i)$ corrupting the desired response signal $d(i)$. If this noise is uncorrelated, $\mathbf{R}_{vv} = \mathbf{I}$ and $\mathbf{p}_{vv} = \mathbf{0}$, and thus (25) or (34) can be used to update $\mathbf{a}_i$ in this case. Algorithm #2 requires only $3N+2M+1$ multiplies per iteration to implement in this situation. For Algorithm #1, the quantity $\mathbf{a}_i^T\mathbf{a}_i = ||\mathbf{a}_i||_2^2$ can be computed from its previous value and other quantities that can be recursively updated, such that its complexity is also $O(3N + 2M)$.

In cases where $v(i)$ is correlated, we can employ estimators for $\mathbf{R}_{vv}$ and $\mathbf{p}_{vv}$ within the algorithms. From the nature of the bias removal term in Algorithm #2, we have several possibilities. It can be shown that

$$\widehat{\mathbf{R}}_{vv}^{(1)} = E[\mathbf{d}_{i-1}\mathbf{e}_{i-1}^T]/\sigma_v^2(i) \tag{35}$$

$$\widehat{\mathbf{R}}_{vv}^{(2)} = E[\mathbf{e}_{i-1}\mathbf{d}_{i-1}^T]/\sigma_v^2(i) \tag{36}$$

$$\widehat{\mathbf{R}}_{vv} = E[\mathbf{d}_{i-1}\mathbf{e}_{i-1}^T + \mathbf{e}_{i-1}\mathbf{d}_{i-1}^T]/(2\sigma_v^2(i)) \tag{37}$$

Equation	*Mults.*
Initialization: $\mathbf{a}_0 = \mathbf{p}_0 = \mathbf{0}$, $\mathbf{b}_0 = \mathbf{0}$, $\mathbf{R}_0 = \delta\mathbf{I}$, $\sigma_v^2(0) = \delta$.	
$e(i) = d(i) - \mathbf{a}_i^T\mathbf{d}_{i-1} - \mathbf{b}_i^T\mathbf{x}_i$	$N + M$
$\overline{e}(i) = \mu e(i)$	1
$g(i) = \overline{e}(i)d(i)$	1
$\mathbf{g}_i = \overline{e}(i)\mathbf{d}_{i-1}$	N
$\mathbf{p}_i = \lambda\mathbf{p}_{i-1} + \mathbf{g}_i + d(i)\overline{\mathbf{e}}_{i-1}$	$2N$
$\sigma_v^2(i) = \lambda\sigma_v^2(i-1) + 2g(i)$	1
$\mathbf{r}_i = \mathbf{R}_{i-1}\mathbf{a}_i$	N^2
$f(i) = \begin{cases} e(i)\overline{e}(i)/(\sigma_v^2(i) + (\mathbf{r}_i - 2\mathbf{p}_i)^T\mathbf{a}_i) & \text{Alg. \#1} \\ g(i)/(\sigma_v^2(i) - \mathbf{p}_i^T\mathbf{a}_i) & \text{Alg. \#2} \end{cases}$	$N + 1$ $(1\div)$
$\mathbf{a}_{i+1} = \mathbf{a}_i + \mathbf{g}_i + f(i)(\mathbf{r}_i - \mathbf{p}_i)$	N
$\mathbf{b}_{i+1} = \mathbf{b}_i + \overline{e}(i)\mathbf{x}_i$	M
$\begin{bmatrix} \mathbf{R}_i & \vdots \\ \cdots & \cdot \end{bmatrix} = \begin{bmatrix} \sigma_v^2(i) & \mathbf{p}_i^T \\ \mathbf{p}_i & \mathbf{R}_{i-1} \end{bmatrix}$	0
Total Computation: $N^2 + 6N + 2M + 4$ (one$\div$)	

Table 1: Two unbiased adaptive IIR filtering algorithms for correlated observation noise.

are all unbiased estimators for the normalized covariance matrix $\mathbf{R}_{vv}$ at parameter convergence. The unbiased estimators for $\mathbf{p}_{vv}$ at convergence are similar. By replacing expectations in (35)–(37) with filtered sample averages, several different algorithms are possible. The complexities of the algorithms are $N^2 + O(N) + O(M)$ with one divide.

To maintain symmetry of the estimates, we employ exponentially-windowed time averages of the quantities $d(i)e(i)$, $(d(i)\mathbf{e}_{i-1} + e(i)\mathbf{d}_{i-1})$ and $(\mathbf{d}_{i-1}\mathbf{e}_{i-1}^T + \mathbf{e}_{i-1}\mathbf{d}_{i-1}^T)$. To further reduce the complexity, we calculate scaled versions of $\widehat{\sigma_v^2}(i)$, $\widehat{\mathbf{p}}_{vv,i}$, and $\widehat{\mathbf{R}}_{vv,i}$. The resulting algorithms are given in Table 1. The complexities of the algorithms are quite manageable for typical choices of M and N.

4. STOCHASTIC ANALYSIS

In this section, we provide stochastic analyses of algorithms #1 and #2 assuming a system identification model of the form in (4)–(5). For both algorithms, we derive recursive equations characterizing the mean behaviors of the coefficients. From these equations, we derive necessary conditions on the step size for the stability of the algorithms. Moreover, we show that one of the stability points of the recursions yields the optimum parameter values for both algorithms. Note that it has been proven that the constrained cost function in (11) is convex with a single minimum [7]. Finally, we provide a mean-square analysis of Algorithm #2 as given in (33)–(34) for white observation noises assuming Gaussian input and desired response signals.

For these analyses, we define the vectors $\mathbf{w}$, $\mathbf{w}_i$, $\widetilde{\mathbf{w}}_i$, $\mathbf{u}_i$, $\overline{\mathbf{u}}_i$, and $\mathbf{u}_i^{(j)}$, as

$$\mathbf{w} = \begin{bmatrix} \mathbf{a} \\ \mathbf{b} \end{bmatrix}, \quad \mathbf{w}_i = \begin{bmatrix} \mathbf{a}_i \\ \mathbf{b}_i \end{bmatrix}, \quad \widetilde{\mathbf{w}}_i = \mathbf{w}_i - \mathbf{w}, \tag{38}$$

$$\mathbf{u}_i = \begin{bmatrix} \mathbf{y}_{i-1} \\ \mathbf{x}_i \end{bmatrix}, \quad \overline{\mathbf{u}}_i = \mathbf{u}_i + \mathbf{z}_i, \quad \mathbf{u}_i^{(j)} = \overline{\mathbf{u}}_i - \mathbf{z}_i^{(j)}, \tag{39}$$

respectively, where $\mathbf{z}_i^{(j)}$ for $j = \{1, 2\}$ and $\mathbf{z}_i$ are given by

$$\mathbf{z}_i^{(1)} = \begin{bmatrix} \dfrac{\mathbf{p}_{vv} - \mathbf{R}_{vv}\mathbf{a}_i}{1 - 2\mathbf{p}_{vv}^T\mathbf{a}_i + \mathbf{a}_i^T\mathbf{R}_{vv}\mathbf{a}_i}e(i) \\ \mathbf{0} \end{bmatrix} \tag{40}$$

$$\mathbf{z}_i^{(2)} = \begin{bmatrix} \dfrac{\mathbf{p}_{vv} - \mathbf{R}_{vv}\mathbf{a}_i}{1 - \mathbf{p}_{vv}^T\mathbf{a}_i}d(i) \\ \mathbf{0} \end{bmatrix}, \tag{41}$$

$$\mathbf{z}_i = \begin{bmatrix} \mathbf{v}_{i-1} \\ \mathbf{0} \end{bmatrix}, \tag{42}$$

and $\mathbf{v}_i = [v(i) \cdots v(i-N+1)]^T$, respectively.

4.1. Mean Behaviors

Using the notation defined in (38)–(42), we can compactly express the updates in both (23)–(24) and (32)–(33) as

$$\mathbf{w}_{i+1} = \mathbf{w}_i + \mu e(i)\mathbf{u}_i^{(j)} \quad (43)$$

where $e(i)$ can be expressed as

$$e(i) = d(i) - \mathbf{w}_i^T\overline{\mathbf{u}}_i = v_a(i) - \mathbf{u}_i^T\widetilde{\mathbf{w}}_i \quad (44)$$

and $v_a(i)$ is defined as

$$v_a(i) = v(i) - \mathbf{a}_i^T\mathbf{v}_{i-1}. \quad (45)$$

We subtract $\mathbf{w}$ from both sides of (43), giving

$$\begin{aligned}\widetilde{\mathbf{w}}_{i+1} &= \widetilde{\mathbf{w}}_i - \mu\left(\mathbf{u}_i\mathbf{u}_i^T + \mathbf{z}_i\mathbf{u}_i^T - \mathbf{z}_i^{(j)}\mathbf{u}_i^T\right)\widetilde{\mathbf{w}}_i \\ &\quad + \mu v_a(i)\left(\mathbf{u}_i + \mathbf{z}_i - \mathbf{z}_i^{(j)}\right). \end{aligned} \quad (46)$$

For both algorithms, we find an approximate evolution equation for the means of the filter coefficients by taking expectations of both sides of (46) assuming that i) the commonly-used independence assumptions hold [11] and ii) the fluctuations in the coefficient errors are small such that the elements of $\mathbf{w}_i$ can be replaced by their expected values within the nonlinear terms of (46). This operation gives

$$\begin{aligned}E[\widetilde{\mathbf{w}}_{i+1}] &= (\mathbf{I} - \mu(\mathbf{R}_{uu} - E[\mathbf{z}_i^{(j)}\mathbf{u}_i^T]))E[\widetilde{\mathbf{w}}_i] \\ &\quad + \mu\left(E[v_a(i)\mathbf{z}_i] - E[v_a(i)\mathbf{z}_i^{(j)}]\right). \end{aligned} \quad (47)$$

By straightforward evaluation of the expectations on the RHS of (47), we find that

$$E[\mathbf{z}_i^{(j)}\mathbf{u}_i^T]E[\widetilde{\mathbf{w}}_i] = c_i^{(j)}\begin{bmatrix}\mathbf{p}_{vv} - \mathbf{R}_{vv}E[\mathbf{a}_i] \\ \mathbf{0}\end{bmatrix} \quad (48)$$

$$E[v_a(i)\mathbf{z}_i] = \sigma_v^2(i)\begin{bmatrix}\mathbf{p}_{vv} - \mathbf{R}_{vv}E[\mathbf{a}_i] \\ \mathbf{0}\end{bmatrix} \quad (49)$$

$$E[v_a(i)\mathbf{z}_i^{(j)}] = \sigma_v^2(i)\begin{bmatrix}\mathbf{p}_{vv} - \mathbf{R}_{vv}E[\mathbf{a}_i] \\ \mathbf{0}\end{bmatrix}, \quad (50)$$

where $c_i^{(j)}$ for $j = \{1, 2\}$ is

$$c_i^{(1)} = -\frac{E[\widetilde{\mathbf{w}}_i^T]\mathbf{R}_{uu}E[\widetilde{\mathbf{w}}_i]}{1 - 2\mathbf{p}_{vv}^T E[\mathbf{a}_i] + E[\mathbf{a}_i^T]\mathbf{R}_{vv}E[\mathbf{a}_i]} \quad (51)$$

$$c_i^{(2)} = \frac{\mathbf{w}^T\mathbf{R}_{uu}E[\widetilde{\mathbf{w}}_i]}{1 - \mathbf{p}_{vv}^T E[\mathbf{a}_i]}. \quad (52)$$

Substituting the relations in (48)–(50) into (47), we find

$$E[\widetilde{\mathbf{w}}_{i+1}] = (\mathbf{I} - \mu(\mathbf{R}_{uu} - E[\mathbf{z}_i^{(j)}\mathbf{u}_i^T]))E[\widetilde{\mathbf{w}}_i]. \quad (53)$$

This nonlinear equation is difficult to analyze without knowledge of the input, desired response, and output noise statistics. However, note that $E[\widetilde{\mathbf{w}}_i] = \mathbf{0}$ is a stationary point of (53), indicating that one of the solutions obtainable by the algorithm is the correct one. In addition, by premultiplying (53) by $E[\mathbf{w}_i]$, it can be shown for Algorithm #1 that $c_i^{(1)} = 0$ at the only stationary point of the analysis equation, yielding $E[\widetilde{\mathbf{w}}_i] = \mathbf{0}$ at convergence. For Algorithm #2, it can be shown that for uncorrelated noises $v(i)$, the correct minima is the only achievable one within the set of stable IIR models for $N \leq 2$. Although such statements are not proofs of convergence, extensive simulations of the two algorithms have shown that they have well-behaved adaptation characteristics.

We can use (53) to develop necessary conditions on the step size μ for stability of the algorithms. Noting that $E[\widetilde{\mathbf{w}}_i] \approx \mathbf{0}$ about the optimum solution, the conditions

$$0 < \mu < 2/\lambda_{max} \quad (54)$$

are necessary for convergence of (53), where λ_{max} is the maximum eigenvalue of $\mathbf{R}_{uu}$. Typically, μ is chosen in the range $0 < \mu \ll 2/(NE[d^2(i)] + ME[x^2(i)])$.

4.2. Mean Square Behavior

We now analyze the mean-square behavior of algorithm #2 in (33)–(34) for jointly-Gaussian input and desired response signals and uncorrelated output noises. Since this algorithm is a modified two-channel version of the anti-Hebbian adaptive FIR filter analyzed in [12], the analytical results from this work can be applied. The equations describing the approximate mean-square evolution of the coefficients are

$$\mathbf{m}_{i+1} = (\mathbf{I} - \mu\overline{\mathbf{R}})\mathbf{m}_i + \mu(\mathbf{I} - \mathbf{I}_a\overline{\mathbf{K}}_i)\overline{\mathbf{p}} \quad (55)$$

$$\overline{\mathbf{K}}_{i+1} = \mathbf{F}_i + \mathbf{G}_i + \mathbf{H}_i + \mathbf{H}_i^T \quad (56)$$

$$\begin{aligned}\mathbf{F}_i &= \overline{\mathbf{K}}_i - \mu(\overline{\mathbf{R}}\overline{\mathbf{K}}_i + \overline{\mathbf{K}}_i\overline{\mathbf{R}}) + \mu^2(2\mathbf{R}_{\overline{uu}}\overline{\mathbf{K}}_i\mathbf{R}_{\overline{uu}} \\ &\quad + \mathbf{R}_{\overline{uu}}\mathrm{tr}[\mathbf{R}_{\overline{uu}}\overline{\mathbf{K}}_i] + 3\sigma_d^4\mathbf{I}_a\overline{\mathbf{K}}_i\mathbf{I}_a \\ &\quad - \mathbf{C}\overline{\mathbf{K}}_i\mathbf{I}_a - \mathbf{I}_a\overline{\mathbf{K}}_i\mathbf{C}) \end{aligned} \quad (57)$$

$$\begin{aligned}\mathbf{G}_i &= \mu^2(\mathbf{C} - \mathbf{C}(\overline{\mathbf{K}}_i - \mathbf{m}_i\mathbf{m}_i^T)\mathbf{I}_a - \mathbf{I}_a(\overline{\mathbf{K}}_i - \mathbf{m}_i\mathbf{m}_i^T)\mathbf{C} \\ &\quad + 2\mathbf{I}_a\overline{\mathbf{K}}_i\mathbf{C}\overline{\mathbf{K}}_i\mathbf{I}_a + \mathbf{I}_a\overline{\mathbf{K}}_i\mathbf{I}_a\mathrm{tr}[\overline{\mathbf{K}}_i\mathbf{C}] \\ &\quad - 2\mathbf{I}_a\mathbf{m}_i\mathbf{m}_i^T(\mathbf{m}_i^T\mathbf{C}\mathbf{m}_i)\mathbf{I}_a) \end{aligned} \quad (58)$$

$$\begin{aligned}\mathbf{H}_i &= \mu(\mathbf{I} + 3\mu\sigma_d^2\mathbf{I}_a)(\mathbf{m}_i\overline{\mathbf{p}}^T + \overline{\mathbf{p}}^T\mathbf{m}_i(2\mathbf{m}_i\mathbf{m}_i^T - \overline{\mathbf{K}}_i)\mathbf{I}_a \\ &\quad - (\overline{\mathbf{K}}_i\overline{\mathbf{p}}\mathbf{m}_i^T + \mathbf{m}_i\overline{\mathbf{p}}^T\overline{\mathbf{K}}_i)\mathbf{I}_a) + \mu^2((2\mathbf{R}_{\overline{uu}}\overline{\mathbf{K}}_i \\ &\quad + \mathrm{tr}[\mathbf{R}_{\overline{uu}}\overline{\mathbf{K}}_i]\mathbf{I})\overline{\mathbf{p}}\mathbf{m}_i^T\mathbf{I}_a + (\mathbf{R}_{\overline{uu}}\mathbf{m}_i\overline{\mathbf{p}}^T + \overline{\mathbf{p}}\mathbf{m}_i^T\mathbf{R}_{\overline{uu}} \\ &\quad + \overline{\mathbf{p}}^T\mathbf{m}_i\mathbf{R}_{\overline{uu}})(2\overline{\mathbf{K}}_i\mathbf{I}_a - 2\mathbf{m}_i\mathbf{m}_i^T\mathbf{I}_a - \mathbf{I})) \end{aligned} \quad (59)$$

where $\mathbf{m}_i = E[\mathbf{w}_i]$, $\overline{\mathbf{K}}_i = E[\mathbf{w}_i\mathbf{w}_i^T]$, $\mathbf{R}_{\overline{uu}} = E[\overline{\mathbf{u}}_i\overline{\mathbf{u}}_i^T]$, $\overline{\mathbf{p}} = E[d(i)\overline{\mathbf{u}}_i]$, $\overline{\mathbf{R}} = \mathbf{R}_{\overline{uu}} - \sigma_d^2\mathbf{I}_a$, $\mathbf{C} = 2\overline{\mathbf{p}}\overline{\mathbf{p}}^T + \sigma_d^2\mathbf{R}_{\overline{uu}}$, and $[\mathbf{I}_a]_{i,j} = 1$ for $1 \leq i = j \leq N$ and is zero otherwise.

The analysis in (55)–(59) can be iterated via computer given the initial coefficient values and the signal statistics, and the effects of step size on the algorithm's behavior can be easily studied.

5. SIMULATIONS

We now verify and explore the performance of the new algorithms via simulations. Averaged results have been obtained from 100 independent simulation runs in each case.

Figure 1 shows the convergence of the average total squared parameter error, given by $||\widetilde{\mathbf{w}}_i||_2^2$, for several algorithms operating on signals generated from the system

$$\mathbf{a} = \begin{bmatrix} 1.1314 \\ -0.25 \end{bmatrix}, \qquad \mathbf{b} = \begin{bmatrix} 0 \\ 0.05 \\ -0.4 \end{bmatrix}, \quad (60)$$

where both the input and observation noise signals are uncorrelated zero-mean Gaussian with variances $\sigma_x^2(i) = 1$ and $\sigma_v^2(i) = 0.01$, respectively. Here, we have chosen step

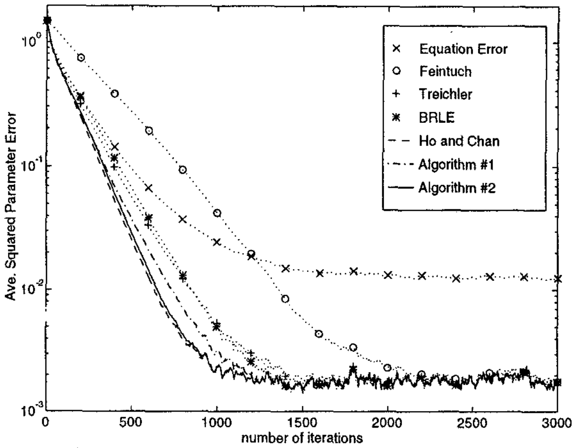

Fig. 1: Convergence of average squared parameter errors for the competing algorithms, uncorrelated output noise.

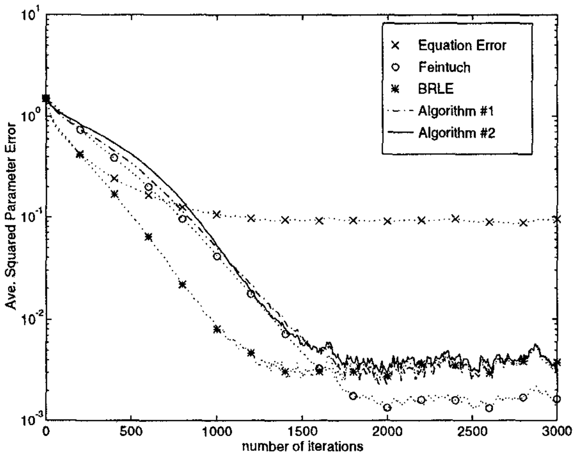

Fig. 2: Convergence of average squared parameter errors for the competing algorithms, correlated output noise.

size parameters of 0.03, 0.031, 0.033, 0.034, 0.031, 0.033, and 0.03, for the equation error, Feintuch' [1], Treichler's [5], BRLE [4], Ho and Chan's [8], and Algorithms #1 and #2, respectively, to obtain approximately the same average parameter error powers in steady-state, if possible. In this case, we have set $\tau(i) = 1 - \lambda^i$, $\lambda = 0.99$ for the BRLE algorithm. As can be seen, the two new algorithms outperform all but the algorithm in [8]. Note that the behavior of Algorithm #2 closely follows that of the algorithm in [8] without need for either costly divides or an additional adaptive parameter. Because of their computational simplicity, the new algorithms are to be preferred in this situation.

Figure 2 shows the average total squared parameter error for the various algorithms for a system identification task identical to that above, except that $v(i)$ is correlated Gaussian with $E[v(i)v(i-m)] = 0.01(-0.9)^{|m|}$. In this case, Algorithms #1 and #2 are implemented as in Table 1 with $\lambda = 0.99$, and $\mu = 0.03$ for all algorithms. While all algorithms except the equation error scheme produce unbiased estimates, the new algorithms offer robust adaptation without careful choice of the convergence parameters.

Figure 3 shows the mean behavior of the filter coefficients for a two-parameter system identification task in which $a_1 = 0.9$, $b_0 = \sqrt{0.19}$, $\mu = 0.003$, and the input and observation noise signals are uncorrelated zero-mean Gaussian with variances $\sigma_x^2(i) = 1$ and $\sigma_v^2(i) = 0.1$, respectively.

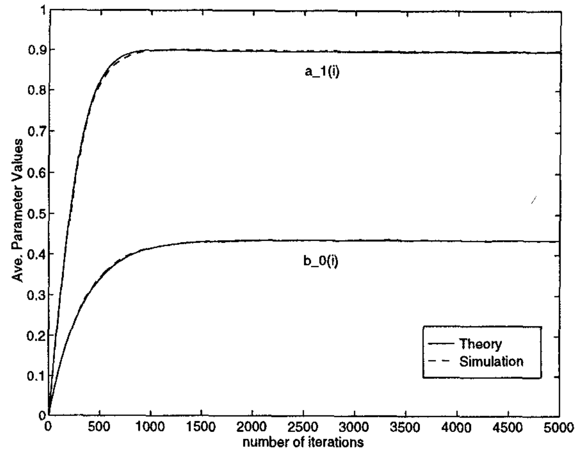

Fig. 3: Convergence of parameters as predicted from (55)–(59) and from simulation.

Also shown for comparison is the predicted performance of the system using (55)–(59). The steady-state values of the coefficients are close to their optimum values, and the mean-square analysis accurately predicts the behavior of the system in this situation.

6. CONCLUSIONS

In this paper, we have derived two new algorithms for unbiased adaptive IIR filtering in the presence of potentially-correlated output noise. Analyses and simulations indicate that the algorithms are well-behaved and provide unbiased estimates at convergence. An important issue for these algorithms is the existence and nature of any strictly positive real (SPR) conditions that the unknown system must satisfy for algorithm convergence. These issues and the robustness of the algorithms are explored in a companion paper [13].

REFERENCES

[1] P.L. Feintuch, "An adaptive recursive LMS filter," *Proc. IEEE*, vol. 64, no. 11, pp. 1622-1624, Nov. 1976.

[2] J.J. Shynk, "Adaptive IIR filtering," *IEEE Signal Processing Mag.*, vol. 6, no. 2, pp. 4-21, April 1989.

[3] P.A. Regalia, *Adaptive IIR Filtering in Signal Processing and Control* (New York: Marcel Dekker, 1995).

[4] J. Lin and R. Unbehauen, "Bias-remedy least-mean-square equation error algorithm for IIR system parameter recursive estimation," *IEEE Trans. Signal Processing*, vol. 40, no. 1, pp. 62–69, Jan. 1992.

[5] J.R. Treichler, "γ-LMS and its use in a noise-compensating adaptive spectral analysis technique," *Proc. IEEE Int. Conf. Acoust., Speech, Signal Processing*, pp. 933-936, April 1979.

[6] L. Ljung and T. Söderström, *Theory and Practice of Recursive Identification* (Cambridge, MA: MIT Press, 1983).

[7] P.A. Regalia, "An unbiased equation error identifier and reduced- order approximations," *IEEE Trans. Signal Processing*, vol. 42, no. 6, pp. 1297-1412, June 1994.

[8] K.C. Ho and Y.T. Chan, "Bias removal in equation error adaptive IIR filters," *IEEE Trans. Signal Processing*, vol. 43, no. 1, pp. 51-62, Jan. 1995.

[9] C.E. Davila, "An algorithm for efficient, unbiased, equation-error infinite impulse response adaptive filtering," *IEEE Trans. Signal Processing*, vol. 42, no. 5, pp. 1221-1226, May 1994.

[10] K. Gao, M.O. Ahmad, and M.N.S. Swamy, "A constrained anti-Hebbian learning algorithm for total least-squares estimation with applications to adaptive FIR and IIR filtering," *IEEE Trans. Circ. Syst. II*, vol. 41, no. 11, pp. 718-729, Nov. 1994.

[11] J.E. Mazo, "On the independence theory of equalizer convergence," *Bell Sys. Tech. J.* , vol. 58, no. 5, pp. 963-993, May-June 1979.

[12] S.C. Douglas, "Analysis of an anti-Hebbian adaptive FIR filtering algorithm," *IEEE Trans. Circ. Syst. II*, vol. 43, no. 11, Nov. 1996 (to appear).

[13] M. Rupp and S.C. Douglas, "Deterministic stability analyses of unit-norm constrained algorithms for unbiased adaptive IIR filtering," to be presented at *IEEE Int. Conf. Acoust., Speech, Signal Processing*, Munich, Germany, April 1997.