PART III

10. An Unbiased Equation-Error-Based Adaptive IIR Filtering Algorithm

The adaptive IIR filtering algorithms may be broadly classified into two classes namely output-error-based methods and equation-error-based methods [12]. The cost function of output-error-based methods is usually a non-quadratic function in the coefficients of the rational transfer function and usually has multiple minima. The non-convexity of the error surface makes it difficult to locate the global minimum of the error surface. The equation error surface, on the other hand, is quadratic in the coefficients, and hence, every minimum in the error surface is a global minimum of the error surface. Hence, simple gradient-based optimization procedures can be used to minimize the equation-error cost function. This renders the equation-error-based method more attractive for adaptive IIR filtering. A serious drawback of the equation-error-based adaptive filter is that the minimum point of the equation error surface provides a biased estimate of the true coefficients of the transfer function. Different techniques to mitigate/eliminate this problem have been studied. The Steiglitz-McBride method mitigates the bias by iteratively minimizing the equation error corresponding to prefiltered input and output signals [42]. The prefilter is modified at the end of each iteration based on the results of that iteration. Prefiltering, effectively, attempts to convert the equation error to output error. The instrumental variable method is another way of mitigating the bias [43]. Here, the information vector is modified so that the modified information vector is correlated to the original information vector, but uncorrelated to the measurement noise. In this chapter, we describe a method that, under certain conditions, completely eliminates the bias problem.

We first show that the minimum of the equation-error surface subject to the constraint that the autoregressive (AR) coefficient vector has unit norm provides an unbiased estimate of the true parameters. A system identification method based on this idea has been proposed [44]. This method uses an off-line procedure, where a batch of data is collected from the system and the collected data is used to construct a model with a separate (off-line) procedure. Off-line procedures are unsuitable for adaptive filtering, where a model of the plant, possibly time varying, is needed during the real time operation of the system. Here, a procedure is desired that updates the model after the arrival of each new data point. Such recursive procedures use less memory than off-line procedures, since there is no need to store all past data, and they may be used as computationally robust alternatives for the off-line identification methods. The existing adaptive (on-line) techniques attempt to solve this constrained optimization problem using a Lagrange multiplier method [45] or a generalized Rayleigh quotient [46]. We propose a hyper-spherical parameterization that converts this unit-norm-constrained optimization into an unconstrained optimization.

10.1 Unbiased Equation Error

The equation error [47] e_n, as shown in Figure 10.1, is characterized by the difference equation:

$$e_n = \sum_{k=0}^{M} b_k x_{n-k} - \sum_{k=0}^{M} a_k w_{n-k} \ , \tag{10.1}$$

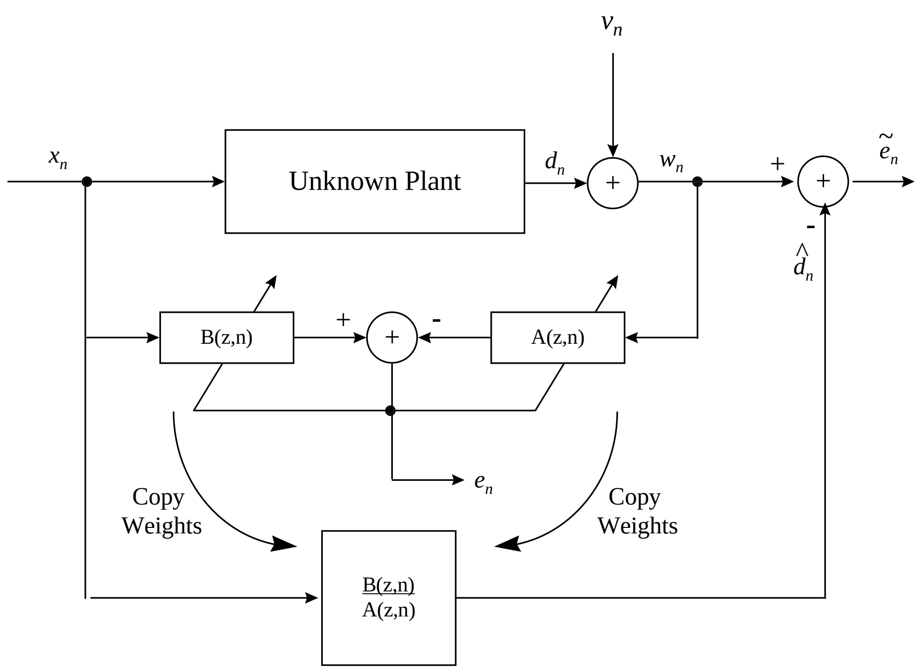

Figure 10.1 Equation-Error-Based Identifier used for Adaptive Filtering.

where x_n and w_n are the input and output (corrupted by measurement noise) of the unknown plant, M is the order of the model, and $\{b_k, a_k\}$ are the coefficients of the IIR model $\hat{H}(z)$, that produces the estimate of the plant output, and is defined as

$$\hat{H}(z) = \frac{B(z)}{A(z)} = \frac{\sum_{k=0}^{M} b_k z^{-k}}{\sum_{k=0}^{M} a_k z^{-k}}. \tag{10.2}$$

The coefficients of the model are chosen so that the mean square equation error

$$\begin{aligned} E\left[e_n^2\right] &= \begin{bmatrix}\mathbf{b}^t & -\mathbf{a}^t\end{bmatrix} \begin{bmatrix} E\left(\mathbf{x}_n\mathbf{x}_n^t\right) & E\left(\mathbf{x}_n\mathbf{w}_n^t\right) \\ E\left(\mathbf{w}_n\mathbf{x}_n^t\right) & E\left(\mathbf{w}_n\mathbf{w}_n^t\right) \end{bmatrix} \begin{bmatrix}\mathbf{b} \\ -\mathbf{a}\end{bmatrix} \\ &= \begin{bmatrix}\mathbf{b}^t & -\mathbf{a}^t\end{bmatrix} \begin{bmatrix} \mathbf{R}_{xx,n} & \mathbf{R}_{xw,n} \\ \mathbf{R}_{xw,n}^t & \mathbf{R}_{ww,n} \end{bmatrix} \begin{bmatrix}\mathbf{b} \\ -\mathbf{a}\end{bmatrix}, \end{aligned} \tag{10.3}$$

where

$$\begin{aligned} \mathbf{b} &= \begin{bmatrix} b_0 & b_1 & \cdots & b_M \end{bmatrix}^t, \\ \mathbf{a} &= \begin{bmatrix} a_0 & a_1 & \cdots & a_M \end{bmatrix}^t, \\ \mathbf{x}_n &= \begin{bmatrix} x_n & x_{n-1} & \cdots & x_{n-M} \end{bmatrix}^t, \text{ and} \\ \mathbf{w}_n &= \begin{bmatrix} w_n & w_{n-1} & \cdots & w_{n-M} \end{bmatrix}^t, \end{aligned}$$

is minimized. If we assume that the signals are stationary and that the measurement noise v_n (shown in Figure 10.1) is a stationary white noise, with variance σ_v^2, which is independent of the input, then (10.3) may be rewritten as

$$E\left[e_n^2\right] = \begin{bmatrix}\mathbf{b}^t & -\mathbf{a}^t\end{bmatrix} \begin{bmatrix} \mathbf{R}_{xx} & \mathbf{R}_{xd} \\ \mathbf{R}_{xd}^t & \mathbf{R}_{dd} \end{bmatrix} \begin{bmatrix}\mathbf{b} \\ -\mathbf{a}\end{bmatrix} + \sigma_v^2 \mathbf{a}^t \mathbf{a}, \tag{10.4}$$

where d_n is the true output of the plant. A minimizing solution of (10.4) is $\mathbf{a} = \mathbf{b} = \mathbf{0}$. Some constraint is needed to avoid this trivial solution. Typically, a_0 is set to 1. This results in a monic characteristic polynomial $A(z)$. With the monic constraint, the mean square equation error is

$$E\left[e_n^2\right] = \begin{bmatrix}\mathbf{b}^t & -\mathbf{a}^t\end{bmatrix}\begin{bmatrix}\mathbf{R}_{xx} & \mathbf{R}_{xd} \\ \mathbf{R}_{xd}^t & \mathbf{R}_{dd}\end{bmatrix}\begin{bmatrix}\mathbf{b} \\ -\mathbf{a}\end{bmatrix} + \sigma_v^2\left(1+\sum_{k=1}^{M} a_k^2\right) \tag{10.5}$$

Clearly, the "unwanted" second term on the right hand side of (10.5) adds a penalty function proportional to the norm of $\mathbf{a}$. This introduces an undesirable bias to the minimizing solution, which in addition depends on the variance of the measurement noise.

This bias problem can be eliminated by using a unit-norm constraint on $\mathbf{a}$ instead of the monic constraint [44]. With this unit-norm constraint, the equation error in (10.4) becomes

$$E\left[e_n^2\right] = \begin{bmatrix}\mathbf{b}^t & -\mathbf{a}^t\end{bmatrix}\begin{bmatrix}\mathbf{R}_{xx} & \mathbf{R}_{xd} \\ \mathbf{R}_{xd}^t & \mathbf{R}_{dd}\end{bmatrix}\begin{bmatrix}\mathbf{b} \\ -\mathbf{a}\end{bmatrix} + \sigma_v^2 . \tag{10.6}$$

Since σ_v^2 is a constant, minimizing (10.6) is equivalent to minimizing the equation error under the noise-free condition, i.e., noise does not influence the solution. The Hessian matrix of the mean square equation error, the matrix in the quadratic expression in (10.6), is positive semidefinite. Hence, the error surface is convex [16]. Furthermore, if the input signal is persistently exciting of degree $2M+1$ and the model order M equals the true order of the unknown plant, the minimizing solution of (10.6) gives the true parameters of the plant. If M is greater than the true order, the minimizing solution of (10.6) gives the true parameters after common poles and zeros in the model are canceled [44]. These properties of the unbiased equation error criterion in (10.6) make it attractive for adaptive filtering using gradient based algorithms.

10.2 Recursive Adaptive Filtering Algorithm

The objective is to recursively minimize the mean squared value of the equation error shown in (10.1), under the unit-norm constraint $\mathbf{a}^t\mathbf{a}=1$. The gradient descent algorithm [47], with suitable step-size, can be used to minimize the mean square equation error. However, true gradient computation requires estimation of second order statistics of the signals. This can be avoided by using the stochastic gradient algorithm [47]. Here, the instantaneous squared value of the equation error, rather than its mean, is minimized. We enforce the unit-norm constraint by using a hyper-spherical parameterization of $\mathbf{a}$. That is, rather than directly adapting the MA coefficients $\mathbf{b} = (b_0, b_1, \ldots, b_M)^T$, and the AR coefficients $\mathbf{a} = (a_0, a_1, \ldots, a_M)^T$, where $(\cdot)^T$ is the vector transpose operator, we propose adapting $\overline{\mathbf{b}} = (\overline{b}_0, \overline{b}_1, \ldots, \overline{b}_M)^T$, r, and $\boldsymbol{\theta} = (\theta_1, \theta_2, \ldots, \theta_M)^T$, where

$$\begin{aligned}
b_j &= r\overline{b}_j, & \forall j \in \{0,1,\ldots,M\} \\
a_0 &= r\cos\theta_1, & \\
a_k &= r\left(\prod_{i=1}^{k} \sin\theta_i\right)\cos\theta_{k+1}, & \forall k \in \{1,2,\ldots,M-1\} \\
a_M &= r\prod_{i=1}^{M} \sin\theta_i. &
\end{aligned} \tag{10.7}$$

It can be shown that $\|\mathbf{a}\| = r$ and that $\mathbf{a} = \mathbf{a}(r, \boldsymbol{\theta}) = r\mathbf{a}(1, \boldsymbol{\theta})$. To insure a one-to-one correspondence between $\boldsymbol{\alpha} = [\mathbf{b}^T \quad \mathbf{a}^T]^T$ and $\boldsymbol{\beta} = [\overline{\mathbf{b}}^T \quad r \quad \boldsymbol{\theta}^T]^T$, we restrict the values of θ_i such that $\theta_i \in [0, 2\pi)\ \forall i \in \{1,2,\ldots,M\}$ and r such that $r > 0$.

Table 10.1 shows the resulting stochastic-gradient-based adaptation algorithm.

Table 10.1 Hyper-spherical LMS Equation-Error Algorithm.

Initialize $\overline{\mathbf{b}}$ and $\boldsymbol{\theta}$ (Note: $r \equiv 1$, so that $\mathbf{b} \equiv \overline{\mathbf{b}}$)
For $n = 0,1,2,\ldots$, repeat the following:
(1) $a_{n,0} = \cos\theta_{n,1}$
(2) $a_{n,k} = \left(\prod_{i=1}^{k} \sin\theta_{n,i}\right)\cos\theta_{n,k+1}, \forall k \in \{1,2,\ldots,M-1\}$
(3) $a_{n,M} = \prod_{i=1}^{M} \sin\theta_{n,i}$
(4) $e_n = \sum_{k=0}^{M} \overline{b}_{n,k} x_{n-k} - \sum_{k=0}^{M} a_{n,k} w_{n-k}$
(5) $\overline{b}_{n+1,k} = \overline{b}_{n,k} - \mu e_n x_{n-k}$, for $k = 0,1,\ldots,M$
(6) $\theta_{n+1,k} = \theta_{n,k} - \mu e_n \mathbf{w}_n^t \left(\frac{\partial \mathbf{a}}{\partial \theta_k}\right)_{\theta=\theta_n}$, for $k = 1,2,\ldots,M$

10.3 (Non-) Unimodality Issues

For small enough step-sizes, the above algorithm is guaranteed to converge to a point where the gradient is zero. Hence it is desirable to have the gradient being zero at a point as a sufficient condition for that point to be a global optimum of the function being optimized. In this section, we analyze whether the equation-error surface, after hyper-spherical transformation, satisfies the global optimality property.

Lemma: If $a_M \neq 0$, the Jacobian $\mathbf{J}_M$ of the hyper-spherical transformation is non-singular.

Proof: The Jacobian of the hyper-spherical transformation can be written as

$$\mathbf{J}_M = \begin{pmatrix} \mathbf{J}_M^{11} & \mathbf{J}_M^{12} \\ \mathbf{0} & \mathbf{J}_M^{22} \end{pmatrix} \tag{10.8}$$

$$\mathbf{J}_M^{11} = r\mathbf{I}_{(M+1)\times(M+1)} \tag{10.9}$$

$$\mathbf{J}_M^{12} = \begin{bmatrix} \overline{\mathbf{b}} & \mathbf{0}_{(M+1)\times M} \end{bmatrix} \tag{10.10}$$

$$\mathbf{J}_M^{22} = \begin{pmatrix} c_1 & -rs_1 & 0 & \cdots & 0 \\ s_1c_2 & rc_1c_2 & -rs_1s_2 & \ddots & \vdots \\ \vdots & \vdots & \vdots & & \\ & & & \ddots & 0 \\ s_1\ldots s_{M-1}c_M & rc_1s_2\ldots s_{M-1}c_M & rs_1c_2s_3\ldots s_{M-1}c_M & \cdots & -rs_1\ldots s_M \\ s_1\ldots s_M & rc_1s_2\ldots s_M & rs_1c_2s_3\ldots s_M & \cdots & rs_1\ldots s_{M-1}c_M \end{pmatrix}_{(M+1)\times(M+1)} \tag{10.11}$$

where c_m and s_n denote $\cos\theta_m$ and $\sin\theta_n$, respectively. Using the partitioned-matrix determinant identity [48], $\det(\mathbf{J}_M) = \det(\mathbf{J}_M^{11})\det(\mathbf{J}_M^{22})$. From (10.9) follows that $\det(\mathbf{J}_M^{11}) = r^{M+1}$. We claim that

$$\det(\mathbf{J}_M^{22}) = r^M \sin^{M-1}(\theta_1)\sin^{M-2}(\theta_2)\cdots\sin(\theta_{M-1}). \tag{10.12}$$

For $M = 1$, $\det(\mathbf{J}_M^{22}) = r$, which satisfies (10.12). Assume that (10.12) is true for $\mathbf{J}_{M-1}^{22}$. Using (10.11)

$$\begin{aligned}\det(\mathbf{J}_M^{22}) &= rs_1\ldots s_{M-1}c_M^2 \det(\mathbf{J}_{M-1}^{22}) + rs_1\ldots s_{M-1}s_M^2 \det(\mathbf{J}_{M-1}^{22}) \\ &= rs_1\ldots s_{M-1} \det(\mathbf{J}_{M-1}^{22})\end{aligned} \tag{10.13}$$

Hence, $\det\left(\mathbf{J}_M^{22}\right) = r^M \sin^{M-1}(\theta_1)\sin^{M-2}(\theta_2)\cdots\sin(\theta_{M-1})$ and, by induction, (10.12) holds for all positive integer values of M . Hence,

$$\det(\mathbf{J}_M) = r^{2M+1}\sin^{M-1}(\theta_1)\sin^{M-2}(\theta_2)\cdots\sin(\theta_{M-1}). \tag{10.14}$$

Then $a_M = r\sin(\theta_1)\sin(\theta_2)...\sin(\theta_M) \neq 0$ implies none of $\sin(\theta_i)$ or r in (2) equals 0. Hence $\mathbf{J}_M$ is non-singular.

Minimizing the equation error (in terms of direct coefficient parameterization) under the unit-norm constraint is equivalent to minimizing the generalized Rayleigh quotient [46] given by

$$E_1(\boldsymbol{\alpha}) \overset{\Delta}{=} \frac{\begin{bmatrix}\mathbf{b}^T & \mathbf{a}^T\end{bmatrix}\mathbf{R}\begin{bmatrix}\mathbf{b}\\ \mathbf{a}\end{bmatrix}}{\mathbf{a}^T\mathbf{a}} \tag{10.15}$$

where $\mathbf{R}$ is a positive-definite auto-correlation matrix. Since there is a one-to-one correspondence between $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$, (3) can be rewritten as follows.

$$\begin{aligned} E_1(\boldsymbol{\alpha}) &= \frac{\begin{bmatrix} r\overline{\mathbf{b}}^T & r\mathbf{a}(1,\boldsymbol{\theta})^T\end{bmatrix}\mathbf{R}\begin{bmatrix} r\overline{\mathbf{b}}\\ r\mathbf{a}(1,\boldsymbol{\theta})\end{bmatrix}}{r^2} \\ &= \begin{bmatrix}\overline{\mathbf{b}}^T & \mathbf{a}(1,\boldsymbol{\theta})^T\end{bmatrix}\mathbf{R}\begin{bmatrix}\overline{\mathbf{b}}\\ \mathbf{a}(1,\boldsymbol{\theta})\end{bmatrix} \overset{\Delta}{=} E_2(\boldsymbol{\beta}) \end{aligned} \tag{10.16}$$

Thus $E_2(\boldsymbol{\beta})$ is the unit-norm-constrained equation-error cost function after the hyper-spherical transformation.

Theorem: Every stationary point of $E_1(\boldsymbol{\alpha})$ is a stationary point of $E_2(\boldsymbol{\beta})$. Furthermore, any newly formed stationary point of $E_2(\boldsymbol{\beta})$ is a saddle point and the AR polynomial corresponding to these saddle points has degree less than *M*.

Proof: The gradients of $E_1(\boldsymbol{\alpha})$ and $E_2(\boldsymbol{\beta})$ are related as follows:

$$\nabla_{\boldsymbol{\beta}} E_2(\boldsymbol{\beta}) = \mathbf{J}_M^T \nabla_{\boldsymbol{\alpha}} E_1(\boldsymbol{\alpha}), \tag{10.17}$$

where $\boldsymbol{\alpha} = \begin{bmatrix}\mathbf{b}^T & \mathbf{a}^T(r,\boldsymbol{\theta})\end{bmatrix}^T$. From (10.17), it follows that if $\nabla_{\boldsymbol{\alpha}} E_1(\boldsymbol{\alpha}^*) = \mathbf{0}$, then the corresponding point in hyper-spherical coordinates $\boldsymbol{\beta}^* = \begin{bmatrix}\overline{\mathbf{b}}^{*T} & r^* & \boldsymbol{\theta}^{*T}\end{bmatrix}^T$, where $\mathbf{b}^* = r^*\overline{\mathbf{b}}^*$ and

$\mathbf{a}^* = \mathbf{a}(r^*, \theta^*)$, satisfies $\nabla_\beta E_2(\beta^*) = \mathbf{0}$. This proves the first part of the theorem. Furthermore, since the Hessians of the error surfaces $E_1(\alpha)$ and $E_2(\beta)$ are related as follows

$$\nabla_\beta^2 E_2(\beta) = \mathbf{J}_M^T \nabla_\alpha^2 E_1(\alpha) \mathbf{J}_M \tag{10.18}$$

the nature of the stationary points of $E_1(\alpha)$ is preserved by the hyper-spherical transformation. That is, minima get mapped to minima and saddle points get mapped to saddle points.

If there exists a point β^* such that $\nabla_\beta E_2(\beta^*) = \mathbf{0}$ while $\nabla_\alpha E_1(\alpha^*) \neq \mathbf{0}$, where $\alpha^* = [r^* \overline{\mathbf{b}}^{*T} \quad \mathbf{a}^T(r^*, \theta^*)]^T$, it follows from (10.18) that $\mathbf{J}_M$ must be singular. From the Lemma above, $\mathbf{J}_M$ being singular implies $a_M = 0$, which in turn shows that the AR polynomial corresponding to any newly formed stationary point of $E_2(\beta)$ has degree less than *M*. Nayeri and Jenkins have proved that the newly formed stationary points due to any continuous transformation are saddle points [4].

From the above theorem, it is clear that the unimodal constrained equation-error surface [46] in ARMA parameterizations, upon hyper-spherical transformation, may have stationary points in addition to the global minimum of the error surface. Hence, the gradient being zero at a point is not a sufficient condition for that point to be a global minimum. However, the newly formed stationary points are saddle points. If parameter trajectories go through or near saddle points this can potentially slow down the convergence of the algorithm. However, the gradient algorithm will still converge, albeit perhaps more slowly, to the global minimum provided there is some noise present to perturb the adaptation algorithm away from saddle points and provided the saddle points are not dense. It is worth adding that there is always some noise due to measurement and quantization errors.

10.4 Simulation Results

The proposed algorithm is simulated in MATLAB. The plant to be modeled is a fifth order system with poles and zeros as shown in Figure 10.2. The model order M is chosen to be 5, equal to the true system order. The input x_n and the measurement noise v_n are assumed to be white. The signal-to-noise ratio (SNR) of the measured output is 10 dB. The steady-state pole and zero estimates from the proposed algorithm, along with the estimates from the traditional equation error based algorithm (with monic constraint), are shown in Figure 10.2.

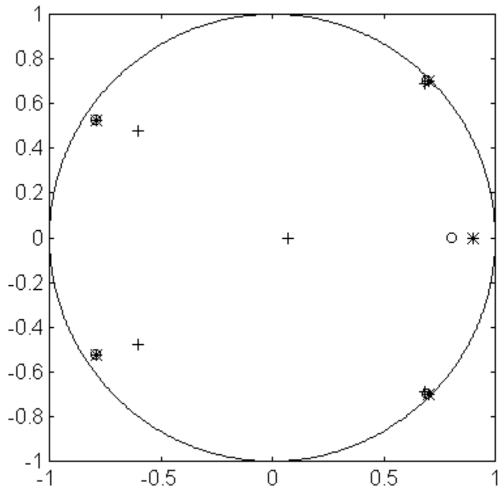

Figure 10.2(a) True Pole Locations (*), Estimate using Unbiased Equation Error (o), and Estimate using Traditional Equation Error with Monic Constraint(+).

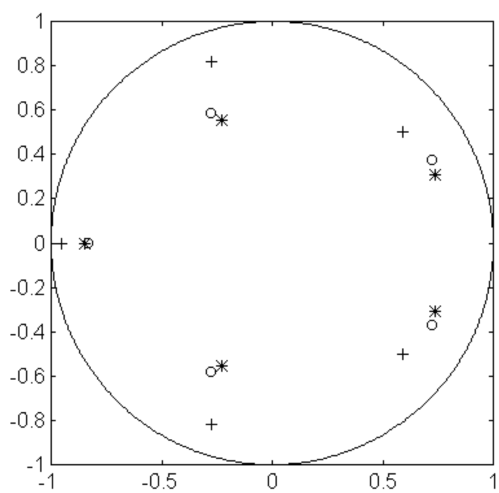

Figure 10.2(b) True Zero Locations (*), Estimate using Unbiased Equation Error (o), and Estimate using Traditional Equation Error with Monic Constraint(+).

The unbiased equation error approach produces estimated poles that nearly coincide with the original poles, and estimated zeros that are mostly very close to the original zeros. The corresponding results from the monic equation error approach exhibit a pronounced bias. Alternatively we can look at the corresponding magnitude responses shown in Figure 10.3. Clearly, the estimate from the traditional algorithm is biased, while the estimate from the proposed algorithm is unbiased.

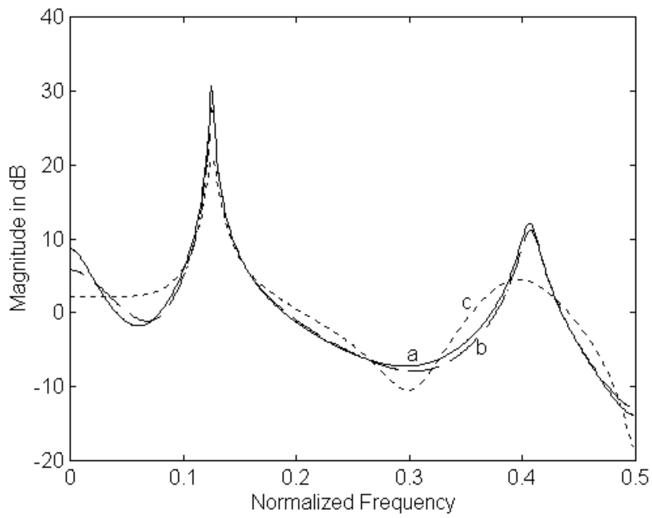

Figure 10.3 Magnitude Responses: (a) True Response, (b) Estimate from Equation Error with Unit-Norm Constraint, and (c) Estimate from Equation Error with Monic Constraint.

Figure 10.4 shows the convergence behavior of the estimate for $a_{n,0}$. The estimates for the other $a_{n,k}$'s behave similarly.

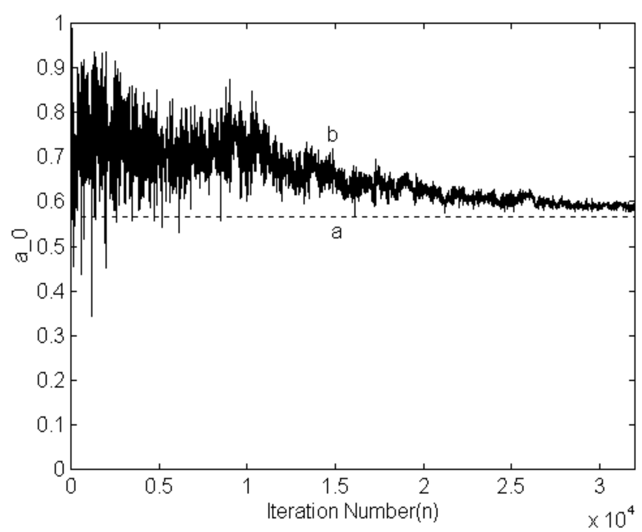

Figure 10.4 (a) True Value of a_0 and (b) Estimated a_0 from the Adaptive IIR Filtering Algorithm.

10.5 Conclusion

An algorithm for adaptive IIR filtering based on a parameterization that converts the unit-norm constrained adaptive filtering problem into an unconstrained adaptive filtering problem is presented. The parameter transformation may introduce additional stationary points in the error surface. However, additional stationary points in the transformed error surface, if any, are saddle points, so that the unique global minimum is preserved. The proposed algorithm does not have any bias problem, if the measurement noise is white, and it converges to the global minimum of the unbiased equation error surface.