INVERSE IDENTIFICATION ADAPTIVE IIR FILTERING: CONVERGENCE SPEED ANALYSIS AND SUCCESSIVE APPROXIMATIONS ALGORITHM

Phillip M. S. Burt

Dept. of Telecommunications and Control Engineering
Escola Politécnica, University of São Paulo
E-mail: phillip@lcs.poli.usp.br

ABSTRACT

The approach based on balanced realization theory, previously used to analyze the convergence speed of adaptive IIR filters in the identification configuration and to propose a faster algorithm (successive approximations algorithm), is now used in the inverse identification configuration. This case is of interest for applications such as channel equalization and system identification itself. We show that, while in an identification configuration the Hankel singular values of the system being identified have an important effect on convergence speed, in the inverse identification case it is the Hankel singular values of a certain system related to the system being identified that have this role. From this result, a condition for faster convergence speed is obtained as well as an inverse identification version of the successive approximations algorithm. As in the identification case, it can lead to much faster convergence with a relatively small increase in computational complexity.

Index Terms— Adaptive IIR filtering, convergence speed.

1. INTRODUCTION

Adaptive IIR filters [1]-[3] frequently converge very slowly, which tends to inhibit their use in real applications. In [4]-[6], we presented an analysis of the local convergence speed of adaptive IIR filters in an identification configuration, based on balanced realization theory. From this, followed a new adaptive IIR algorithm with faster convergence, called the successive approximations (SA) algorithm. In [7] the same approach was used to analyze the convergence speed of adaptive IIR filters in the inverse identification configuration. This case is more commonly associated with channel equalization but can also be of interest to system identification itself, since, for instance, convergence speed properties are not necessarily equal to those of the identification configuration.

This work is organized as follows. In Section 2 we state the identification and inverse identification adaptive IIR filtering problems under consideration. Also, to motivate the analysis of the inverse identification problems we present a numerical example of slow convergence. In Section 3, aspects of the analysis of the identification case presented in [4]-[6] are reviewed. In Sections 4 and 5 this approach is extended to the inverse identification case. Based on this, in Section 6 we obtain a condition for faster convergence in the inverse identification case and in Section 7 we obtain an inverse identification version of the SA algorithm. Finally, in Section 8 we discuss perspectives for the continuation of this work.

This work was supported by FAPESP (Proc. 06/01113-0).

2. PROBLEM STATEMENT

2.1. Identification and Inverse Identification Configurations

Identification and inverse identification configurations for adaptive filtering [8] are represented in Fig. 1. In identification configuration adaptive IIR filtering, a rational function $\widehat{H}(z)$ is adapted so as to minimize the mean-square error $E[e^2(n)]$, where $e(n) = y(n) - \widehat{y}(n)$ is the error, $y(n) = H(z)u(n)$ is the system output (the "desired signal") for a known input $u(n)$ and $\widehat{y}(n) = \widehat{H}(z)u(n)$ is the filter output for the same input. In this mixed notation, z is the unit-delay operator, with $zu(n) = u(n-1)$. If the input $u(n)$ is white, the problem is equivalent to the minimization of the squared L_2 norm $||H(z) - \widehat{H}(z)||^2$. In the inverse identification case, a rational function $\widehat{H}_{\text{I}}(z)$ is adapted so as to minimize $E[e^2(n)]$, where now $e(n) = u(n) - \widehat{u}(n)$, with $\widehat{u}(n) = \widehat{H}_{\text{I}}(z)y(n) = \widehat{H}_{\text{I}}(z)H(z)u(n)$. In this case, for white $u(n)$ the problem is equivalent to the minimization of the squared L_2 norm $||1 - \widehat{H}_{\text{I}}(z)H(z)||^2$.

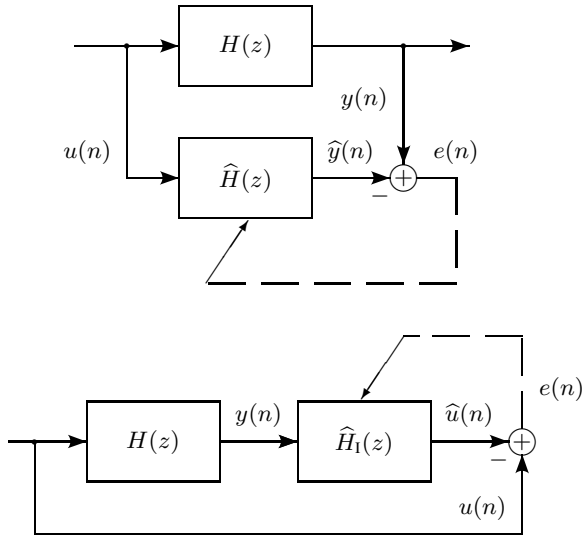

Fig. 1. Identification (top) and inverse identification (bottom) adaptive filter configurations.

2.2. Example of Slow Convergence

To motivate the analysis that follows, we consider, in the inverse identification configuration, the application of a standard direct-form stochastic gradient algorithm [3, p.271] to the case in which $H(z^{-1})$ has zeros at $0.6\angle \pm 95^\circ$ and $0.8\angle \pm 110^\circ$, poles at $0.7\angle \pm 30^\circ$ and $0.7\angle \pm 90^\circ$ and unit L_2 norm, the signal $u(n)$ is white Gaussian noise with unit variance and we have sufficient modelling. Initial parameter values equal to 0.5 times the true parameter values were chosen. Adaptation steps $\mu_a = 0.02$ and $\mu_b = 0.2$, which are within a factor of 2 of the values that lead the adaptation to diverge, were used for, respectively, the pole and zero parameters of the adaptive filter. The squared error $e^2(n)$ is shown in Fig. 2, in logarithmic scale. It can be seen that to reach an error of -60 dB, for instance, it takes over 2×10^5 iterations.

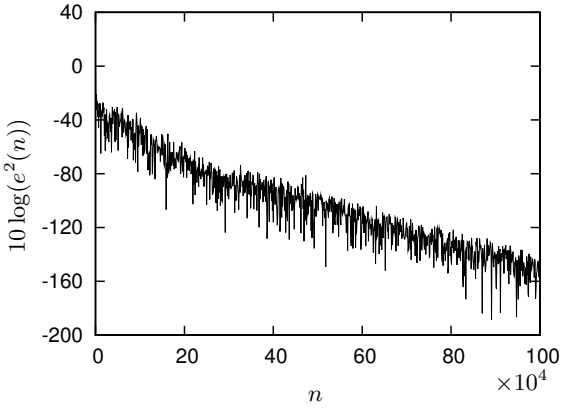

Fig. 2. Example of slow convergence in inverse identification.

3. REVIEW OF IDENTIFICATION CONFIGURATION

We now review some aspects of the analysis of convergence speed in the identification configuration, presented in [4], [6]. The ideas involved will then be applied in the next section to the inverse identification configuration.

3.1. Reduced Error Surface

We assume a white input, so $E[e^2(n)] = ||H(z) - \widehat{H}(z)||^2$. With $\widehat{H}(z) = (b + b_1 z + \ldots + b_M z^M)/(1 + a_1 z + \ldots + a_M z^M) = B(z)/A(z)$ and $\overline{A}(z) = z^M A(z^{-1})$, we can always obtain a set of functions

$$\left\{ \frac{P_k(z)}{A(z)} \right\}_{k=0}^{M} \cup \left\{ z^k \frac{\overline{A}(z)}{A(z)} \right\}_{k=1}^{\infty},$$

with $P_M(z) = \overline{A}(z)$, such that $\widehat{H}(z) = \sum_{k=0}^{M} \nu_k P_k(z)/A(z)$, which is a complete orthonormal basis for the Hardy space $\mathcal{H}_2$ of causal and stable functions of the complex variable z (as follows from [3, p.122]). We note that the basis functions $z^k \overline{A}(z)/A(z)$ are allpass, that is, $|e^{j\omega k}\overline{A}(e^{j\omega})/A(e^{j\omega})| = 1$.

Writing now $H(z)$ in terms of this basis, we have $H(z) = \sum_{k=0}^{M} \rho_k P_k(z)/A(z) + \sum_{k=1}^{\infty} \tau_k z^k \overline{A}(z)/A(z)$, which leads to

$$||H(z) - \widehat{H}(z)||^2 = \sum_{k=0}^{M} (\nu_k - \rho_k)^2 + \sum_{k=1}^{\infty} \tau_k^2. \qquad (1)$$

With $\langle ., . \rangle$ denoting the standard inner product, since $\tau_k = \langle H(z), z^k \overline{A}(z)/A(z) \rangle$, we have then

$$||H(z) - \widehat{H}(z)||^2 = \sum_{k=0}^{M} (\nu_k - \rho_k)^2 + \sum_{k=1}^{\infty} \langle H(z), z^k \frac{\overline{A}(z)}{A(z)} \rangle^2. \quad (2)$$

Optimizing the zeros of $\widehat{H}(z)$ given its poles (that is, working on the "reduced error surface") makes $\sum_{k=0}^{M} (\nu_k - \rho_k)^2 = 0$ [3]. The remaining term in (2) can then be written in matrix form, so that

$$||H(z) - \widehat{H}(z)||^2 = \sum_{k=1}^{\infty} \langle H(z), z^k \frac{\overline{A}(z)}{A(z)} \rangle^2 = \mathbf{v}^t \mathbf{\Gamma}_H^2 \mathbf{v}, \quad (3)$$

where $\mathbf{v}$ contains the coefficients of the expansion of $\overline{A}(z)/A(z)$, superscript t denotes transpose, and $\mathbf{\Gamma}_H$ is the Hankel form of $H(z)$.

3.2. Balanced Form Factorization

We now assume that $\deg H(z) = N$. The Hankel form $\mathbf{\Gamma}_H$ can be factored into the controlability and observability forms of a balanced realization of $H(z)$, which allows us to rewrite (3) so that on the reduced error surface

$$||H(z) - \widehat{H}(z)||^2 = \boldsymbol{\alpha}^t \mathbf{\Sigma}^2 \boldsymbol{\alpha},$$

where the elements of $\boldsymbol{\alpha}$ are $\alpha_k = \langle \zeta_k(z), \overline{A}(z)/A(z) \rangle, k = 1, .., N$, with $\zeta_k(z)$ being the normalized controllability transfer functions of a balanced realization of $H(z)$, and where $\mathbf{\Sigma}$ is diagonal and composed of the Hankel singular values σ_k of $H(z)$.

If we consider a true gradient adaptation algorithm working on the reduced error surface, a local description of the adaptation process is then obtained as

$$\boldsymbol{\alpha}(n+1) \approx \left\{ \mathbf{I} - \mu_a \mathbf{J}[\boldsymbol{\alpha}(n)] \mathbf{J}^t[\boldsymbol{\alpha}(n)] \mathbf{\Sigma}^2 \right\} \boldsymbol{\alpha}(n), \qquad (4)$$

where μ_a is the adaptation stepsize of the pole paramenters and $\mathbf{J}$ is a Jacobian sensitivy matrix, with elements $[\mathbf{J}]_{i,j} = \partial \alpha_i / \partial a_j$. If the eigenvalue spread of $\mathbf{J}\mathbf{J}^t\mathbf{\Sigma}^2$ is large, then convergence is slow at points where $\boldsymbol{\alpha}(n)$ is mainly in the directions of eigenvectors associated with small eigenvalues of $\mathbf{J}\mathbf{J}^t\mathbf{\Sigma}^2$. In these slowly converging situations, the behaviour of the true gradient adaptation algorithm is usually a good approximation of the behaviour of the stochastic gradient algorithm.

3.3. Faster Convergence Condition and SA Algorithm

Now, if $H(z) = \overline{D}(z)/D(z)$, with $\overline{D}(z) = z^N D(z^{-1})$, then $H(z)$ is allpass and its Hankel singular values are all equal to 1. This tends to contribute to a smaller eigenvalue spread of $\mathbf{J}\mathbf{J}^t\Sigma^2$ and, thus, to faster convergence of the adaptation. In fact, since the direct term of the transfer function does not affect the associated Hankel form, any

$$H(z) = s \left[\frac{\overline{D}(z)}{D(z)} + c \right] = s \frac{\overline{D}(z) + cD(z)}{D(z)} \qquad (5)$$

will also have equal Hankel singular values (in this case, equal to s).

In the successive approximations (SA) algorithm proposed in [4], an auxiliary block is used in an iterative manner to turn $H(z)$ into an allpass function and thus speed up convergence.

4. EXTENSION TO INVERSE IDENTIFICATION

With $H(z) = C(z)/D(z) \in \mathcal{H}_2$ and assuming $\deg[H(z)] = M$ (sufficient modelling), we initially build a complete orthonormal basis for $\mathcal{H}_2$,

$$\left\{ \frac{P_k(z)}{A(z)D(z)} \right\}_{k=0}^{2M} \cup \left\{ z^k \frac{\overline{A}(z)\overline{D}(z)}{A(z)D(z)} \right\}_{k=1}^{\infty}, \quad (6)$$

where $\overline{D}(z) = z^M D(z^{-1})$. We now modify this basis to

$$\left\{ \frac{P_k(z)}{A(z)D(z)} \cdot \frac{C(z)}{\overline{C}(z)} \right\}_{k=0}^{2M} \cup \left\{ z^k \frac{\overline{A}(z)\overline{D}(z)}{A(z)D(z)} \cdot \frac{C(z)}{\overline{C}(z)} \right\}_{k=1}^{\infty}, \quad (7)$$

where $\overline{C}(z) = z^M C(z^{-1})$.

Although this new set does not necessarily span Hardy space $\mathcal{H}_2$, it is still orthonormal, since for any $X(z)$ and $Y(z)$, $\langle X(z)C(z)/\overline{C}(z), Y(z)C(z)/\overline{C}(z)\rangle = \langle X(z), Y(z)\rangle$. Moreover, restricting ourselves to the case in which $H(z)$ is minimum phase, it is easy to show [7] that $\widehat{H}_{\rm I}(z)H(z)$ and the constant function $F(z) = 1$ are in the space spanned by this basis, with $\widehat{H}_{\rm I}(z)H(z) = \sum_{k=0}^{2M} \overline{\nu}_k P_k(z) C(z)/G(z)$ and $1 = \sum_{k=0}^{2M} \overline{\rho}_k P_k(z)C(z)/G(z) + \sum_{k=1}^{\infty} \overline{\tau}_k z^k \overline{A}(z)\overline{D}(z)C(z)/G(z)$, where $G(z) = A(z)D(z)\overline{C}(z)$. This result allows us to write the mean- square error as

$$||1 - \widehat{H}_{\rm I}(z)H(z)||^2 = \sum_{k=0}^{2M} (\overline{\nu}_k - \overline{\rho}_k)^2 + \sum_{k=1}^{\infty} \overline{\tau}_k^2,$$

which reminds us of expression (1) for the identification configuration.

This correspondence can be driven further if we note that coefficients $\overline{\tau}_k$ are given by

$$\overline{\tau}_k = \left\langle 1, z^k \frac{\overline{A}(z)\overline{D}}{A(z)D(z)} \cdot \frac{C(z)}{\overline{C}(z)} \right\rangle = \left\langle \frac{\overline{C}(z)D(z)}{C(z)\overline{D}(z)}, z^k \frac{\overline{A}(z)}{A(z)} \right\rangle.$$

Moreover, since $z^k\overline{A}(z)/A(z)$ is strictly causal, we have, with $[.]_{\oplus}$ being the causal projection operator,

$$\overline{\tau}_k = \left\langle H_{\rm p}(z), z^k \frac{\overline{A}(z)}{A(z)} \right\rangle, \quad H_{\rm p}(z) \triangleq \left[\frac{\overline{C}(z)D(z)}{C(z)\overline{D}(z)} \right]_{\oplus}.$$

Thus, we arrive at:

Property 1 *In the inverse identification configuration, with minimum phase $H(z) = C(z)/D(z)$, $\widehat{H}_{\rm I}(z) = B(z)/A(z)$ and sufficient modelling, the mean-square error can be written as*

$$||1 - \widehat{H}_{\rm I}(z)H(z)||^2 = \sum_{k=0}^{2M} (\overline{\nu}_k - \overline{\rho}_k)^2 + \sum_{k=1}^{\infty} \left\langle H_{\rm p}(z), z^k \frac{\overline{A}(z)}{A(z)} \right\rangle^2, \quad (8)$$

where $\overline{\nu}_k$ and $\overline{\rho}_k$ are coefficients of the expansion of $\widehat{H}_{\rm I}(z)H(z)$ and 1, respectively, in the basis defined by (7) and $H_{\rm p}$ is the causal projection of $\overline{C}(z)D(z)/[C(z)\overline{D}(z)]$.

We can see that (8) has the same form as (2), with $H_{\rm p}(z)$ in place of $H(z)$. In the identification configuration, as seen, the assumption that the parameters are on the reduced error surface leads to (3) which, in turn, leads to the convergence properties discussed in Section 3 (and others, discussed in [4], [6]). In the next section we will see how a similar procedure can be followed starting from (8).

5. SLOW CONVERGENCE ON THE RESTRICTED ERROR SURFACE

From (8), it follows that at points where

$$\frac{\partial}{\partial p} \sum_{k=0}^{2M} (\overline{\nu}_k - \overline{\rho}_k)^2 = 0 \quad (9)$$

for all adapted parameters p, a local description of the adaptation such as (4) can be obtained, and all the resulting local convergence speed properties of the identification configuration, discussed in Section 3, apply to the inverse identification configuration if $H(z)$ is replaced by $H_{\rm p}(z)$. If there is a manifold of points that satisfy (9) rather than a discrete set of them, then it is more likely that they will be relevant to the global behaviour of the adaptation.

When $\overline{\nu}_k = \overline{\rho}_k$ for $k = 0, 1, \ldots 2M-1$ we say the parameters are on the "second restricted error surface", which is shown [7] to be a one-dimensional manifold. Specifically, solutions $\widehat{H}_{\rm I}(z) = B(z)/A(z)$ to these equations are given by

$$\begin{bmatrix} \mathbf{b} \\ \mathbf{a} \end{bmatrix} = [\ \boldsymbol{\Theta}_C \quad -(\boldsymbol{\Phi}_D + \gamma\boldsymbol{\Phi}_{\overline{D}})]^{-1}(\boldsymbol{\theta}_D + \gamma\boldsymbol{\theta}_{\overline{D}}),$$

where $\boldsymbol{\Theta}_C$, $\boldsymbol{\Theta}_D = [\boldsymbol{\theta}_D\ \boldsymbol{\Phi}_D]$ and $\boldsymbol{\Theta}_{\overline{D}} = [\boldsymbol{\theta}_{\overline{D}}\ \boldsymbol{\Phi}_{\overline{D}}]\mathbf{P}$, with $\mathbf{P}$ being the anti-diagonal permutation matrix, are $2M \times (M+1)$ convolution matrices formed from the coefficients of $C(z)$, $D(z)$, and $\overline{D}(z)$, respectively, and $\mathbf{b}$ and $[1\ \mathbf{a}^t]^t$ are $(M+1)\times 1$ vectors formed from the coefficients of $B(z)$ and $A(z)$, respectively. The only restriction on scalar γ is that $B(z)/A(z) \in \mathcal{H}_2$.

We conjecture that close to every point of the second restricted error surface there is a point where (9) is valid. This is plausible since, as we move away from the second restricted error surface, all terms $(\overline{\nu}_k - \overline{\rho}_k)^2$ for $k = 0, 1, \ldots, 2M-1$ grow and only $(\overline{\nu}_{2M} - \overline{\rho}_{2M})^2$ may become smaller.

Corroborating this conjecture, it is verified [7] that for the slowly converging example presented in Section 2, adaptation readily reaches a point close to the second restricted error surface and that from there on, convergence is very slow. From there on, also, refering to (4) with $\mathbf{J}\mathbf{J}^t\boldsymbol{\Sigma}^2$ associated now with $H_{\rm p}(z)$ instead of $H(z)$, $\boldsymbol{\alpha}(n)$ is mainly in the direction of the eigenvectors associated with the small eigenvalues of $\mathbf{J}\mathbf{J}^t\boldsymbol{\Sigma}^2$, which effectively explains why convergence is slow. Figure 3 shows the evolution of eigenvalues λ_k of $\mathbf{J}\mathbf{J}^t\boldsymbol{\Sigma}^2$.

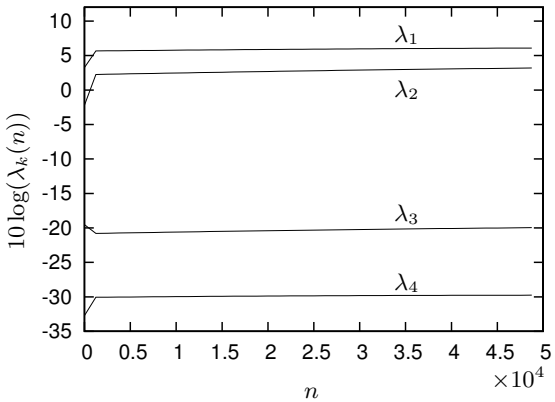

Fig. 3. True gradient adaptation: eigenvalues λ_k of $\mathbf{J}\mathbf{J}^t\boldsymbol{\Sigma}^2$.

6. CONDITION FOR SMALL SINGULAR VALUE SPREAD

We now obtain a condition relative to $H(z)$ in order that $H_{\rm p}(z) = \{\overline{C}(z)D(z)/[C(z)\overline{D}(z)]\}_{\oplus}$ has a small Hankel singular value spread, contributing to faster convergence. Initially, we note that $H_{\rm p}(z)$ has the form $S(z)/C(z)$ and that if

$$S(z) = s[\overline{C}(z) + cC(z)], \tag{10}$$

for any c, then $H_{\rm p}(z) = S(z)/C(z)$ will have all Hankel singular values equal to s. Now, separating the causal and strictly anticausal parts of $\overline{C}(z)D(z)/[C(z)\overline{D}(z)]$, we can write $S(z)\overline{D}(z) + T(z)C(z) = \overline{C}(z)D(z)$, which with (10) results in $s[\overline{C}(z)+cC(z)]\overline{D}(z) + T(z)C(z) = \overline{C}(z)D(z)$. Rearranging, we get

$$[T(z) + sc\overline{D}(z)]C(z) = [D(z) - s\overline{D}(z)]\overline{C}(z).$$

Since minimum phase $C(z)$ does not have roots in common with $\overline{C}(z)$, it follows that $C(z) = t[D(z) - s\overline{D}(z)]$ and

$$H(z) = t\frac{D(z) - s\overline{D}(z)}{D(z)}, \tag{11}$$

for any t, which should be compared with (5) for the identification configuration.

7. SUCCESSIVE APPROXIMATIONS ALGORITHM

With (11) we now obtain an inverse identification version of the successive approximations algorithm (SA) presented in [4]. Starting from an initial estimate $H_*(z)$ of $H(z)$, an auxiliary block $H_{\rm a}(z)$ is employed as depicted in Figure 4, such that the input to $\widehat{H}_{\rm I}(z)$ is now $[H(z)+H_{\rm a}(z)]u(n)$ and $H(z)+H_{\rm a}(z) \approx t[D(z)-s\overline{D}(z)]/D(z)$. A normalized lattice is adapted with the simplified partial stochastic gradient algorithm [3], with adaptation steps μ_a and μ_b. After a certain number $n_{\rm a}$ of samples, $n_{\rm x}$ samples are used to compute a new estimate $H_*(z)$ such that $H_*(z) \approx 1/\widehat{H}_{\rm I}(z) - H_{\rm a}(z)$. With it, a new $H_{\rm a}(z)$ is obtained and the process repeats.

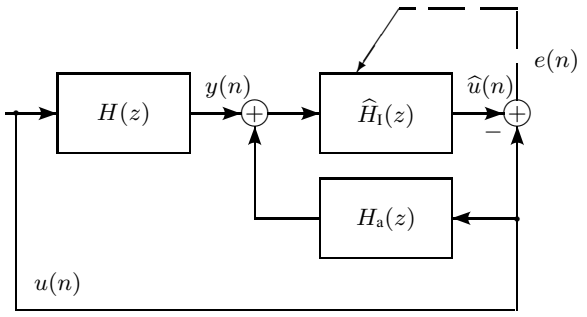

Fig. 4. Diagram of inverse identification version of SA algorithm

This inverse identification version of the SA algorithm was applied to the slow converging example of Section 2, with $t = 1$, $s = 0.6$, $\mu_a = \mu_b = 0.1$, $n_{\rm a} = 450$ and $n_{\rm x} = 200$. An initial estimate $H_*(z)$ obtained with the direct form parameters at 0.7 times their true value was assumed. The result is in Figure 5, along with the result for the normalized lattice alone with $\mu_a = \mu_b = 0.03$ (which are within a factor of 2 of their maximum value for stability) and the same initialization $H_*(z)$. As can be seen, the SA algorithm is much faster.

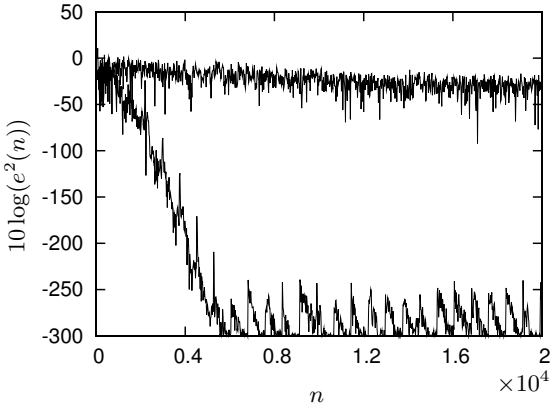

Fig. 5. Inverse identification SA algorithm (faster) and normalized lattice gradient algorithm (slower).

8. CONCLUSIONS AND PERSPECTIVES

The analytical approach in [4], [6], [7] was applied to inverse identification adaptive IIR filtering and an inverse identification version of the successive approximations algorithm was obtained. Other developments of this work that would be of interest are the analysis of the effect, in inverse identification, of the choice of parameterization (direct form, lattice, etc.) and the analysis of the non-minimum phase case. For both the identification and inverse identification configurations, the case of non-white input signals is of great interest. It is possible that the approach followed here can be applied to this problem.

9. REFERENCES

[1] J. Treichler, C. R. Johnson, Jr., and M. G. Larimore, *Theory and Design of Adaptive Filters*. Upper Saddle River, NJ: Prentice-Hall, 2001.

[2] J. J. Shynk, "Adaptive IIR filtering," *IEEE ASSP Mag.*, vol. 6, no. 2, pp. 4–21, Apr. 1989.

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

[4] P. M. S. Burt and P. A. Regalia, "A new framework for convergence analysis and algorithm development of adaptive IIR filters," in *Proc. IEEE ICASSP*, Montreal, 2004, vol. 2, pp. 441–444.

[5] P. M. S. Burt and P. A. Regalia, "IIR adaptive filtering: convergence speed properties in the undermodelled case," in *Proc. IEEE ICASSP*, Philadelphia, 2005, vol. 4, pp. 45–48.

[6] P. M. S. Burt and P. A. Regalia, "A new framework for convergence analysis and algorithm development of adaptive IIR filters," *IEEE Trans. on Signal Processing*, vol. 53, no. 8, pp. 3129–3140, Aug. 2005.

[7] P. M. S. Burt, "On the convergence speed of inverse identification adaptive IIR filters," in *Proc. SBrT/IEEE International Telecommunications Symposium*, 2006, pp. 51–56.

[8] S. Haykin, *Adaptive Filter Theory*, Upper Saddle River, NJ: Prentice-Hall, 2004.