16th European Signal Processing Conference (EUSIPCO 2008), Lausanne, Switzerland, August 25-29, 2008, copyright by EURASIP

ON THE CONVERGENCE SPEED OF ADAPTIVE IIR FILTERS WITH RATIONAL SPECTRUM INPUT SIGNALS

Thomas E. Filgueiras, F. and Phillip M. S. Burt

Dept. of Telecommunications and Control Engineering
Escola Politécnica, Universidade de São Paulo
Emails: tefil@lcs.poli.usp.br, phillip@lcs.poli.usp.br

ABSTRACT

A previously presented framework for the convergence speed analysis of adaptive IIR filters considered white input signals and an identification configuration. In this framework, a local model was built and the Hankel singular values of the system being identified had a fundamental role. It also led to the successive approximations (SA) algorithm, attaining a large speed gain. Here, an approach to extend this framework is presented, allowing more general rational spectrum input signals and unifying the analysis of the identification and inverse identification configurations. It leads to a model that has the same form and depends on the same number of Hankel singular values as in the white input case. The singular values, however, refer now to the result of a certain degree-preserving transformation of the system being identified. Conditions for faster convergence as well as the application of the SA algorithm follow naturally from the extended framework.

1. INTRODUCTION

A framework for the analysis of the convergence speed of adaptive IIR filters [1]-[3] with white input signals was presented in [4],[5]. It provides a better understanding of the effect of factors such as the order of the rational system being identified, the angular and radial distribution of its poles and zeros and the parameterization (direct form, lattice, etc.) that is employed for the adaptive filter. It also leads to an adaptive algorithm with faster convergence, the successive approximations algorithm (SA). In [6],[7] the same framework was extended to the case of inverse identification with white input and it was suggested that a similar approach could be applied to identification or inverse identification adaptive IIR filtering with non-white input signals. This idea is pursued in this paper, which is organized as follows.

In Section 2, properties of adaptive IIR filtering and of the framework for analysis of its convergence speed are reviewed. In Section 3, an approach to extend this framework to more general rational spectrum input signals is presented. Based on this, a condition for faster convergence is presented in Section 4 and the application of the SA algorithm is discussed in Section 5. In Section 6, perspectives for future work are discussed.

2. REVIEW OF ADAPTIVE IIR FILTERING PROPERTIES

In identification configuration adaptive IIR filtering, a rational function $\widehat{H}(z) = B(z)/A(z) = \sum_{k=0}^{M} b_k z^k / (1 + \sum_{k=1}^{M} a_k z^k)$ is adapted so as to minimize the mean square error $E[e^2(n)]$, where the error is $e(n) = y(n) - \widehat{y}(n)$, with $y(n) = H(z)u(n)$ being the system output for a known input $u(n)$ and $\widehat{y}(n) = \widehat{H}(z)u(n)$ being the adaptive filter output for the same input. In this mixed notation, z is the unit delay operator and $zu(n) = u(n-1)$. Using deg[.] to denote the minimum number of unit delays needed to implement a system, we assume throughout the paper that $\deg[H(z)] = N$.

For $X(z) = \sum_{n=-\infty}^{\infty} x_n z^n$ and $Y(z) = \sum_{n=-\infty}^{\infty} y_n z^n$, the standard inner product is $\langle X(z), Y(z) \rangle = \sum_{n=-\infty}^{\infty} x_n y_n$. The mean square error is given then by

$$E[e^2(n)] = \left\langle H(z) - \widehat{H}(z), S_u(z)[H(z) - \widehat{H}(z)] \right\rangle \quad (1)$$

$$\triangleq \left\| H(z) - \widehat{H}(z) \right\|_{S_u}^2, \quad (2)$$

where $S_u(z) = \sum_{k=-\infty}^{\infty} E[u(n)u(n-k)]z^k$.

2.1 Conditions for mean square error minimization

Given the poles of $\widehat{H}(z)$, the mean square error is minimized in relation to the zeros of $\widehat{H}(z)$ if and only if [3]

$$\left[S_u(z)[H(z) - \widehat{H}(z)] \right]_{\oplus} = zg(z)V(z), \quad (3)$$

where $[.]_{\oplus}$ is the causal projection operator (which eliminates the negative powers of z), $V(z) = z^M A(z^{-1})/A(z)$ is all-pass (that is, $|V(e^{j\omega})| \equiv 1$) and $g(z)$ belongs to the Hardy space $\mathscr{H}_2$ of the stable and causal functions of the complex variable z.

If the input signal has a rational spectrum, then $S_u(z) = G(z)G(z^{-1})$, where $G(z)$ is a rational function in $\mathscr{H}_2$ and can also be constrained to have minimum phase. In particular, when the input is white and has unitary variance then $S_u(z) \equiv 1$. In this case, from (3) and from $V(z)$ being all-pass, the mean square error is given by $E[e^2(n)] = \|H(z) - \widehat{H}(z)\|^2 = \|g(z)\|^2$, where $||g(z)||^2$ as a function of the poles of $\widehat{H}(z)$ is called the "reduced error surface".

Considering now the optimization of the zeros and the poles of $\widehat{H}(z)$, it can be shown [3] that $\widehat{H}(z)$ is at a stationary point (minimum or saddle point) of the mean square error if and only if $[S_u(z)[H(z) - \widehat{H}(z)]]_{\oplus} = zQ(z)V^2(z)$, for some $Q(z) \in \mathscr{H}_2$. From this and $S_u(z) = G(z)G(z^{-1})$, with $\deg[G(z)] = L$, it can be shown (see also [8]) that the condition $M \geq N + L$ ensures there are no local minima of the mean square error. For a white input signal ($L = 0$) this requires at least sufficient modelling ($M = N$), while for non-white inputs ($L > 0$) it requires over-modelling ($M > N$).

2.2 Mean square error orthonormal decomposition

As follows from [3, p.122], with $\widehat{H}(z) = B(z)/A(z) = \sum_{k=0}^{M} b_k z^k / (1 + \sum_{k=1}^{M} a_k z^k)$, we can always obtain a set of functions

$$\left\{ \frac{P_k(z)}{A(z)} \right\}_{k=0}^{M} \cup \left\{ z^k V(z) \right\}_{k=1}^{\infty}, \quad (4)$$

where $\deg[P_k(z)] \leq M$, which is a complete orthonormal basis for $\mathscr{H}_2$. In this basis, $\widehat{H}(z) = \sum_{k=0}^{M} \nu_k P_k(z)/A(z)$ and $H(z) = \sum_{k=0}^{M} \rho_k P_k(z)/A(z) + \sum_{k=1}^{\infty} \tau_k z^k V(z)$, which, with $\tau_k = \langle H(z), z^k V(z) \rangle$, leads to

$$\left\| H(z) - \widehat{H}(z) \right\|^2 = \sum_{k=0}^{M} (\rho_k - \nu_k)^2 + \sum_{k=1}^{\infty} \left\langle H(z), z^k V(z) \right\rangle^2. \quad (5)$$

T. E. Filgueiras, F. and P. M. S. Burt were supported by CNPq and FAPESP (Proc. 06/01113-0), respectively.

Optimizing the zeros of $\widehat{H}(z)$ given its poles makes $\sum_{k=0}^{M}(\rho_k - \nu_k)^2 = 0$. With $V(z) = \sum_{k=0}^{\infty} v_n z^n$ and $\mathbf{v}^t = [\ v_0 \quad v_1 \quad \ldots]$, the reduced error surface is given then by [3]

$$\left\| H(z) - \widehat{H}(z) \right\|^2 = \sum_{k=1}^{\infty} \left\langle H(z), z^k V(z) \right\rangle^2 = \mathbf{v}^t \mathbf{\Gamma}_H^2 \mathbf{v}, \qquad (6)$$

where $\mathbf{\Gamma}_H$ is the Hankel form of $H(z)$.

2.3 Balanced form factorization

The Hankel form $\mathbf{\Gamma}_H$ can be factored into the controllability and observability forms of a balanced realization of $H(z)$, which, with $\deg[H(z)] = N$, allows us to rewrite (6) so that the reduced error surface is given by

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

where $\mathbf{\Sigma}$ is diagonal and composed of the N Hankel singular values σ_k of $H(z)$ and the elements of $\boldsymbol{\alpha}$ are $\alpha_k = \langle \zeta_k(z), V(z) \rangle$, $k = 1, \ldots, N$, with $\zeta_k(z)$ being the normalized controllability transfer functions of a balanced realization of $H(z)$[4],[5].

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

$$\boldsymbol{\alpha}(n+1) \approx \left[\mathbf{I} - \mu_a \mathbf{J}\mathbf{J}^t \mathbf{\Sigma}^2 \right] \boldsymbol{\alpha}(n), \qquad (7)$$

where μ_a is the adaptation stepsize of the pole paramenters and $\mathbf{J}$ is a Jacobian sensitivy matrix, with elements $[\mathbf{J}]_{i,j} = \partial \alpha_i / \partial a_j$. It varies with pole parameters a_j, but for greater simplicity we do not explicity denote this. 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 true gradient adaptation algorithm is usually a good approximation of the stochastic gradient algorithm.

2.4 Faster Convergence and SA Algorithm

Still assuming a white input signal, expression (7) suggests a condition for faster convergence. Since the direct term of $H(z)$ does not affect the associated Hankel form, if $H(z)$ has the form $s[z^N D(z^{-1})/D(z) + c]$ (where we note that $z^N D(z^{-1})/D(z)$ is all-pass), then all Hankel singular values of $H(z)$ are equal to s, which tends to contribute to a smaller eigenvalue spread of $\mathbf{J}\mathbf{J}^t\mathbf{\Sigma}^2$, and thus to a faster convergence of the adaptation. In the successive approximations (SA) algorithm proposed in [4], an auxiliary block is used in an iterative manner to turn $H(z)$ into an all-pass function and thus speed up convergence.

3. CONVERGENCE SPEED WITH RATIONAL SPECTRUM INPUT SIGNALS

Adaptive filtering in the identification configuration and with a rational spectrum input signal is represented in Fig. 1, where $w(n)$ is a unitary variance white signal. It can be seen that the particular case in which, for some minimum-phase $H_I(z) \in \mathscr{H}_2$, we have $H(z) = 1/H_I(z)$ and $G(z) = H_I(z)$ is equivalent to the inverse identification configuration with white input, also represented in Fig. 1 and which was analyzed in [6],[7]. Moreover, inverse identification with non-white input can be obtained with $G(z) = G'(z)H_I(z)$. Thus, both configurations can be analyzed in an unified manner.

3.1 Example of the effect of a non-white input

In order to motivate the following analysis, we consider now a numerical example. A standard direct-form stochastic gradient algorithm [3, p.271] is applied to the identification of an all-pass system for which $H(z^{-1})$ has poles at $0.8\angle \pm 30^\circ$ and $0.9\angle \pm 100^\circ$

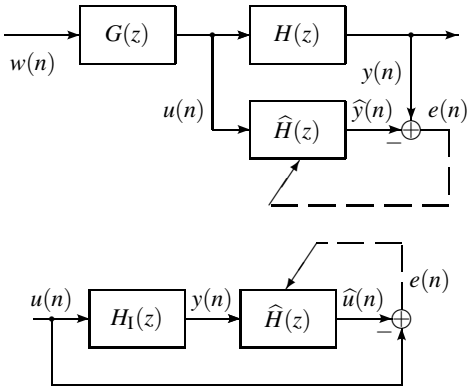

Figure 1: Identification (top) and inverse identification (bottom) adaptive filter configurations.

and zeros at $1/0.8\angle \pm 30^\circ$ and $1/0.9\angle \pm 100^\circ$. The input $u(n)$ is white Gaussian noise with unitary variance. With $M = N = 4$ and $H(z) = \sum_{k=0}^{N} c_k z^k / (1 + \sum_{k=1}^{N} d_k z^k)$, the parameters were initialized at $b_k = 0.4 \times c_k$ and $a_k = 0.4 \times d_k$. Adaptation steps $\mu_a = 0.02$ and $\mu_b = 0.04$ were used for the pole and zero adapted parameters, respectively. They are within a factor of 2 of the values that lead the adaptation to diverge (considering a larger set of initializations), which also holds for the other simulations presented here. The resulting squared error $e^2(n)$ is shown in Fig. 2(A), in logarithmic scale and smoothed for better visualization. As could be expected from the discussion in Section 2, convergence is relatively fast: an error of -60 dB, for instance, is reached in less than 2×10^3 iterations.

We consider now that the input is given by $u(n) = G(z)w(n)$, where $w(n)$ is white Gaussian noise with unitary variance and $G(z^{-1})$ has poles at $0.7\angle \pm 110^\circ$, zeros at $0.75\angle \pm 30^\circ$and unitary L_2 norm. The frequency response $|G(e^{j\omega})|$ is high-pass. With $\mu_a = 0.02$, $\mu_b = 0.02$, $b_k = 0.9 \times c_k$ and $a_k = 0.9 \times d_k$, the resulting squared error $e^2(n)$ is shown in Fig. 2(B). It can be seen that despite starting closer to the optimum, convergence is much slower than for the white input case. It can be also shown that parameters b_k remain in this case very close to their optimum values given parameters a_k.

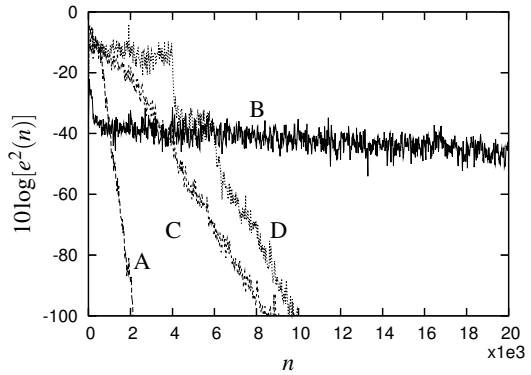

Figure 2: Squared error, $H(z)$ with white (A) and non-white (B) inputs, $H_s(z)$ with non-white input (C) and $H(z)$ with non-white input and SA algorithm (D) .

3.2 Error decomposition

Assuming the minimum-phase function $G(z) \in \mathscr{H}_2$ associated with the input has the form $G(z) = E(z)/F(z)$ and $\deg[G(z)] = L$, a nat-

ural extension to (4) is a complete orthonormal basis for $\mathscr{H}_2$ with the form

$$\left\{\frac{P_k(z)}{A(z)F(z)}\right\}_{k=0}^{M+L} \cup \left\{z^k V(z)\frac{\overline{F}(z)}{F(z)}\right\}_{k=1}^{\infty}, \tag{8}$$

where $\deg[P_k(z)] \leq M+L$ and $\overline{F}(z) = z^L F(z^{-1})$. Now, since $G(z)H(z)$ and $G(z)\widehat{H}(z)$ are both in $\mathscr{H}_2$, we can decompose them using this basis as $G(z)\widehat{H}(z) = \sum_{k=0}^{M+L} \nu_k P_k(z)/[A(z)F(z)]$ and

$$\begin{aligned} G(z)H(z) &= \sum_{k=0}^{M+L} \rho_k \frac{P_k(z)}{A(z)F(z)} + \\ &\quad + \sum_{k=1}^{\infty} \tau_k z^k V(z)\frac{\overline{F}(z)}{F(z)}. \end{aligned} \tag{9}$$

The mean square error $||H(z)-\widehat{H}(z)||_{S_u}^2 = ||G(z)[H(z)-\widehat{H}(z)]||^2$ can then be decomposed as

$$\left\|H(z)-\widehat{H}(z)\right\|_{S_u}^2 = \sum_{k=0}^{M+L}(\rho_k-\nu_k)^2 + \sum_{k=1}^{\infty}\tau_k^2. \tag{10}$$

The particular case in which $H(z) = 1/H_I(z)$ and $G(z)H(z) = 1$ results in $\tau_k = \langle 1, z^k V(z)\overline{F}(z)/F(z)\rangle \equiv 0$ and, as seen, is equivalent to the inverse identification of $H_I(z)$, assuming $H_I(z)$ is minimum-phase. As discussed in more detail in [6], the decomposition of the mean square error obtained in this case is not very useful. The case $G(z)H(z) \neq 1$ would deserve further attention, but we postpone it at this moment. Instead, generalizing the procedure in [6],[7] we consider the decomposition of the maximum-phase component $\overline{E}(z)/F(z)$ of $G(z)$, where $\overline{E}(z) = z^L E(z^{-1})$, multiplied by $H(z)$ or $\widehat{H}(z)$, products which are also in $\mathscr{H}_2$:

$$\frac{\overline{E}(z)}{F(z)}\widehat{H}(z) = \sum_{k=0}^{M+L}\overline{\nu}_k\frac{P_k(z)}{A(z)F(z)}, \tag{11}$$

$$\begin{aligned} \frac{\overline{E}(z)}{F(z)}H(z) &= \sum_{k=0}^{M+L}\overline{\rho}_k\frac{P_k(z)}{A(z)F(z)} + \\ &= + \sum_{k=1}^{\infty}\overline{\tau}_k z^k V(z)\frac{\overline{F}(z)}{F(z)}. \end{aligned} \tag{12}$$

Now, since $\overline{E}(z)/E(z)$ is all-pass it does not affect the mean square error and we can write then

$$\left\|H(z)-\widehat{H}(z)\right\|_{S_u}^2 = \sum_{k=0}^{M+L}(\overline{\rho}_k-\overline{\nu}_k)^2 + \sum_{k=1}^{\infty}\overline{\tau}_k, \tag{13}$$

where the coefficients $\overline{\tau}_k$ are given by

$$\begin{aligned} \overline{\tau}_k &= \left\langle \frac{\overline{E}(z)}{F(z)}H(z), z^k V(z)\frac{\overline{F}(z)}{F(z)}\right\rangle \\ &= \left\langle \frac{E(z^{-1})}{F(z^{-1})}H(z), z^k V(z)\right\rangle. \end{aligned}$$

Moreover, since $z^k V(z)$ is (strictly) causal we can write $\overline{\tau}_k = \langle H_p(z), z^k V(z)\rangle$, where

$$H_p(z) \triangleq \left[G\left(z^{-1}\right)H(z)\right]_{\oplus}. \tag{14}$$

Synthesizing, then:

Property 1 *With a rational spectrum input signal, the mean square error when identifying $H(z)$ can be written as*

$$\left\|H(z)-\widehat{H}(z)\right\|_{S_u}^2 = \sum_{k=0}^{M+L}(\overline{\rho}_k-\overline{\nu}_k)^2 + \sum_{k=1}^{\infty}\left\langle H_p(z), z^k V(z)\right\rangle^2, \tag{15}$$

where $H_p(z)$ is given by (14).

Comparing with (5), we note that the first part of this decomposition has $M+L$ terms instead of M terms and the second part involves $H_p(z)$ instead of $H(z)$. We note also that $H_p(z)$ is equivalent to the function associated to the weighted Hankel form considered in [3, p. 623], where it is shown that from $G(z) \in \mathscr{H}_2$ being minimum-phase it follows that

$$\deg[H_p(z)] = \deg[H(z)].$$

That is, the transformation in (14) is degree-preserving. Finally, the function $H_p(z) = [H_1(z^{-1})/H_1(z)]_{\oplus}$ considered in [6],[7] (this particular form suggested in [9]) is a special case of (14) with, as already discussed, $H(z) = 1/H_1(z)$ and $G(z) = H_1(z)$.

3.3 The restricted error surface

In the white input case, the assumption that the zeros of $\widehat{H}(z)$ are optimized leads to (6) which, in turn, was used to analyze the convergence speed properties discussed in [4],[5]. We will consider now a similar procedure, starting from (15).

We note, initially, that if the points for which, for all adapted parameters p_j,

$$\frac{\partial}{\partial p_j}\sum_{k=0}^{M+L}(\overline{\nu}_k-\overline{\rho}_k)^2 = 0 \tag{16}$$

form a manifold in the parameter space, and if the adaptation takes place on this manifold, then, from (15), all the convergence speed properties of a true gradient algorithm already obtained for the white input case would apply with $H_p(z)$ being used in place of $H(z)$. At this moment we don't know a necessary and sufficient condition for (16), so we consider in the following a condition which is sufficient but maybe not necessary.

Clearly, if

$$\overline{\nu}_k = \overline{\rho}_k,\ k = 0,1,\ldots,M+L, \tag{17}$$

then (16) is satisfied. In this case, we will use the term "restricted error surface" to refer to the mean square error $\|H(z)-\widehat{H}(z)\|_{S_u}^2 = \sum_{k=1}^{\infty}\langle H_p(z), z^k V(z)\rangle^2$ as a function of the pole parameters. We note that the restricted error surface is contained in the reduced error surface.

Now, from (11) and (12), if (17) is true then

$$\frac{\overline{E}(z)}{F(z)}\left[H(z)-\widehat{H}(z)\right] = zQ(z)V(z)\frac{\overline{F}(z)}{F(z)}, \tag{18}$$

for some $Q(z) \in \mathscr{H}_2$. With $H(z) = C(z)/D(z)$, $\widehat{H}(z) = B(z)/A(z)$ and $V(z) = \overline{A}(z)/A(z)$ it follows that $Q(z)$ has the form $R(z)/D(z)$, for some $R(z)$. If we also rewrite $\overline{E}(z)$ as

$$\overline{E}(z) = z^l\overline{E'}(z),\ \text{where } l = \begin{cases} 0, & \text{if } \deg[E(z)] = L \\ 1, & \text{if } \deg[E(z)] < L \end{cases},$$

then we can rewrite (18) as

$$z^l\overline{E'}(z)C(z)A(z) - z^l\overline{E'}(z)B(z)D(z) = zR(z)\overline{F}(z)\overline{A}(z). \tag{19}$$

Assuming now $\overline{A}(z)$ and $\overline{E'}(z)$ have exactly K roots in common, we can group these roots in $\overline{P}(z)$ and write $\overline{A}(z) = \overline{P}(z)\overline{A'}(z)$, where $\deg[\overline{A'}(z)] = M-K$ and $A(z) = P(z)A'(z)$, where $\deg[A'(z)] \leq M-$

K. Proceeding similary, we can also write $\overline{E}'(z) = \overline{P}(z)\overline{E''}(z)$ and $E'(z) = E''(z)P(z)$. Using these expressions in (19) simplifies it to

$$z^l\overline{E''}(z)C(z)P(z)A'(z) - z^l\overline{E''}(z)B(z)D(z) =$$

$$= zR(z)\overline{F}(z)\overline{A'}(z). \qquad (20)$$

Since none of the roots of $\overline{E''}(z)$ can be shared by $\overline{A'}(z)$ (by construction) or $\overline{F}(z)$ (since $\deg[G(z)] = L$ is assumed), it follows that $R(z)$ has the form $T(z)\overline{E''}(z)$ and (20) simplifies further to

$$C(z)P(z)A'(z) - B(z)D(z) = z^{1-l}T(z)\overline{F}(z)\overline{A'}(z). \qquad (21)$$

Since the degrees of the left and right handsides of this expression don't exceed $M+N$ and $M-K+L+1-l+\deg[T(z)]$, respectively, it follows that

$$\deg[T(z)] \leq N+K-L-1+l. \qquad (22)$$

Writing (21) in matrix form we have

$$\mathbf{\Theta}_{CP}\begin{bmatrix} 1 \\ \mathbf{a}' \end{bmatrix} - \mathbf{\Theta}_D\mathbf{b} = \mathbf{\Theta}_{T\overline{F}}\mathbf{\Pi}\begin{bmatrix} 1 \\ \mathbf{a}' \end{bmatrix},$$

where $\mathbf{\Theta}_{CP}$ and $\mathbf{\Theta}_D$ are the $(M+N+1)\times(M-K+1)$ convolution matrices formed from the coefficients of $C(z)P(z)$ and $D(z)$ respectively, $\mathbf{\Theta}_{T\overline{F}}$ is the $(M+N+1)\times(M-K+1)$ convolution matrix formed from the coefficients of $z^{1-l}T(z)\overline{F}(z)$, $\mathbf{\Pi}$ is the antidiagonal permutation matrix, and $\mathbf{a}'$ and $\mathbf{b}$ are the vectors formed from the coefficients of $A'(z)$ and $B(z)$ respectively. Separating their first columns we can write $\mathbf{\Theta}_{CP} = [\boldsymbol{\theta}_{CP}\ \mathbf{\Phi}_{CP}]$ and $\mathbf{\Theta}_{T\overline{F}}\mathbf{\Pi} = [\boldsymbol{\theta}_{T\overline{F}}\ \mathbf{\Phi}_{T\overline{F}}]$. We arrive then at the following system of equations:

$$[\ -\mathbf{\Theta}_D \quad \mathbf{\Phi}_{CP} - \mathbf{\Phi}_{T\overline{F}}]\begin{bmatrix} \mathbf{b} \\ \mathbf{a}' \end{bmatrix} = \boldsymbol{\theta}_{T\overline{F}} - \boldsymbol{\theta}_{CP}. \qquad (23)$$

For each $T(z)$, (23) is a system of $M+N+1$ equations and $2M-K+1$ unknowns. Therefore, as long as

$$M \geq N+K,$$

the solutions to this system are a linear space of $M-N-K$ dimensions. Assuming, additionally, from (22), that $T(z)$ has

$$N+K-L+l > 0$$

degrees of freedom, the solution of the original problem (17), or in other terms, the domain of the restricted error surface, is a set of manifolds, each one with

$$(M-N-K)+(N+K-L+l) = M-L+l \qquad (24)$$

dimensions. Each manifold is associated to a particular set of roots shared by $\overline{A}(z)$ and $\overline{E'}(z)$.

As discussed in Section 2.1, adopting $M = N+L$ has the advantage of ruling out local minima. In this case, with (24), the dimension of the manifolds above is $M-L+l = N+l$.

3.4 Numerical example

Let us now apply the ideas discussed in this section to the non-white input case of the example presented is Subsection 3.1. Initially, the zero/pole diagrams of $H(z)$ and $H_p(z)$ are shown in Fig. 3. It can be seen that the effect of $G(z)$ in this case is to make one pair of zeros of $H_p(z)$ almost cancel one pair of its poles.

In Fig. 4 we show the mean square error $E[e^2(n)]$ that results from the true gradient algorithm, employing the same adaptation steps as the stochastic gradient algorithm. As seen, it is a good estimate of the squared error $e^2(n)$ resulting from the stochastic gradient algorithm. Moreover, as also shown in the figure, the part

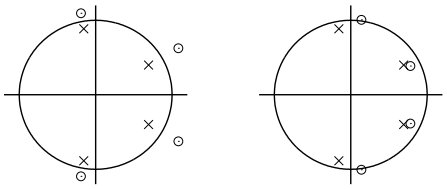

Figure 3: Zeros ("$\circ$") and poles ("$\times$") of $H(z^{-1})$ (left) and $H_p(z^{-1})$(right).

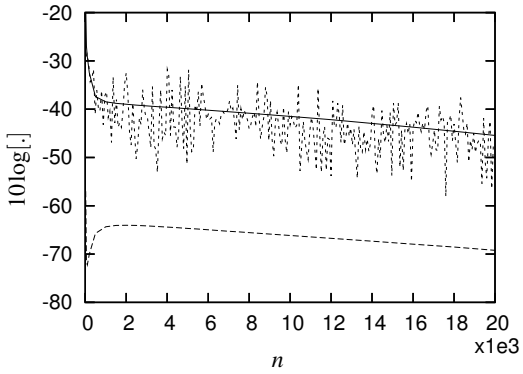

Figure 4: $E[e^2(n)]$ and $e^2(n)$ of true and stochastic gradient algorithms (top traces), respectively, and the part $\sum_{k=0}^{M+L}(\overline{\rho}_k - \overline{v}_k)^2$of $E[e^2(n)]$ (bottom trace).

$\sum_{k=0}^{M+L}(\overline{\rho}_k - \overline{v}_k)^2$ in (15) of $E[e^2(n)]$ is quite smaller than the total (more than 20 dB, that is, less than 1 %). In other terms, the adaptation occurs close to the restricted error surface. As discussed above, this suggest that in this case the model (7) is a good aproximation of the adaptation process, with $H_p(z)$ used in place of $H(z)$.

Indeed, the model (7) accounts for the slow convergence that is observed: the eigenvalue spread of $\mathbf{J}\mathbf{J}'\mathbf{\Sigma}^2$ is quite large (the eigenvalues are almost constant during the adaptation, with $\lambda_1 = 3.2$, $\lambda_2 = 2.2$, $\lambda_3 = 0.032$ and $\lambda_4 = 0.005$) and $\boldsymbol{\alpha}(n)$ is mainly in the directions of the eigenvectors associated to the smaller eigenvalues, as shown in Fig. 5. This suggests that model (7) can be use to establish conditions that contribute to a faster convergence, which is carried out in the next section.

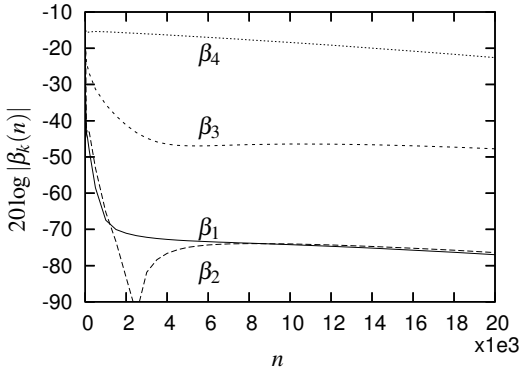

Figure 5: Components $\beta_k(n)$ of $\boldsymbol{\alpha}(n)$ in the directions of the eigenvectors associated to the eigenvalues $\lambda_k(n)$.

4. SMALL SINGULAR VALUE SPREAD CONDITION

4.1 Condition on the zeros of $H(z)$

In our model (7) of the adaptation, if the Hankel singular values of $H_p(z)$ are all equal this will tend to reduce the spread of the eigenvalues of $\mathbf{J}\mathbf{J}^t\mathbf{\Sigma}^2$. Initially, we note that with $H(z)=C(z)/D(z)$ and since $G(z^{-1})$ is anti-causal, $H_p(z)$ has the form

$$H_p(z) \triangleq \left[G(z^{-1})H(z)\right]_{\oplus} = \frac{S(z)}{D(z)},$$

where $\deg[S(z)] \leq N$. So, if

$$S(z) = s\left[\overline{D}(z) + cD(z)\right] \tag{25}$$

for any c, all the Hankel singular values of $H_p(z)$ will be equal to s. To find a $C_s(z)$ such that $H_s(z) = C_s(z)/D(z)$ leads to $S(z)$ with the form above, let us separate

$$\begin{aligned} G(z^{-1})H(z) &= \frac{E(z^{-1})}{F(z^{-1})}\frac{C_s(z)}{D(z)} = \frac{z^L E(z^{-1})}{z^L F(z^{-1})}\frac{C_s(z)}{D(z)} \\ &= \frac{\overline{E}(z)}{\overline{F}(z)}\frac{C_s(z)}{D(z)} \end{aligned}$$

in its causal and strictly anti-causal parts, writing

$$\frac{\overline{E}(z)}{\overline{F}(z)}\frac{C_s(z)}{D(z)} = \frac{S(z)}{D(z)} + \frac{\overline{Q}(z)}{\overline{F}(z)}, \tag{26}$$

for some $\overline{Q}(z) = z^L Q(z^{-1})$, where $\deg[Q(z)] < L$. Substituting (25) in (26), and with some algebraic manipulation we get

$$C_s(z)\overline{E}(z) - D(z)\overline{Q}(z) = s\left[\overline{D}(z) + cD(z)\right]\overline{F}(z),$$

which can be written in matrix form as

$$\begin{bmatrix} \mathbf{\Theta}_{\overline{E}} & -\mathbf{\Theta}_D \end{bmatrix}\begin{bmatrix} \mathbf{c} \\ \mathbf{q} \end{bmatrix} = \mathbf{b}, \tag{27}$$

where $\mathbf{\Theta}_{\overline{E}}$ is the $(N+L+1)\times(N+1)$ convolution matrix formed from the coefficients of $\overline{E}(z)$, $\mathbf{\Theta}_D$ is the $(N+L+1)\times L$ convolution matrix formed from the coefficients of $D(z)$, $\mathbf{c}$ contains the $N+1$ coefficients of $C_s(z)$, $\mathbf{q}$ contains the L coefficients of $\overline{Q}(z)$ and $\mathbf{b}$ contains the $N+L+1$ coefficients of $s\left[\overline{D}(z)+cD(z)\right]\overline{F}(z)$. Solving the system gives the $C_s(z)$ we seek.

4.2 Examples

Let us return to the non-white input case of the example presented is Subsection 3.1, for which we compute $C_s(z)$ according to (27), with $s=0.5$ and $c=0$. The result of the identification of $H_s(z)$ with the stochastic gradient algorithm with adaptation steps $\mu_a = 0.02$ and $\mu_b = 0.04$, and initial parameters t $b_k = 0.4\times c_{s,k}$ and $a_k = 0.4\times d_k$ is shown in Fig. 2(C). It can be seen that, indeed, the convergence is much faster for $H_s(z)$ than for $H(z)$, although not as fast as in the white input case.

5. SUCCESSIVE APPROXIMATIONS ALGORITHM

The condition for small Hankel singular value spread obtained in the previous section can be used now to extend for rational spectrum inputs the Successive Approximations (SA) algorithm, which was originally presented in [4] for the white input case. The resulting procedure is :

1) Starting from an initial estimate $H_*(z) = C_*(z)/D_*(z)$ of $H(z) = C(z)/D(z)$, $C_s(z)$ is calculated with (27) and an auxiliary block $H_a(z) = [C_*(z) - C_s(z)]/D_*(z)$ is used in parallel with the adaptive IIR filter, which is initialized at $C_s(z)/D_*(z)$.

(*Observation*: with such an auxiliary block, if $H_*(z) \approx H(z)$ then $H(z) - H_a(z) \approx C_s(z)/D_*(z)$, which has equal singular values, and the adaptation of $\widehat{H}(z)$ will tend to be fast, as discussed in the previous section. Differently from [4], a direct form adaptive filter is used for comparison with previous simulations.)

2) After a certain number n_a of samples, the transfer function $\widehat{H}(z)$ of the adaptive filter is used to obtain a new estimate $H_*(z)$ such that $||\widehat{H}(z) + H_a(z) - H_*(z)||^2$ is approximately minimized. This computation is carried out in a certain number n_x of samples, and with the new $H_*(z)$ we go back to Step 1.

The result of the SA algorithm with $n_a = 2000$, $n_x = 200$ and applied to the example as above is presented in Fig. 2(D). It can be seen that once in the neighbourhood of the optimum it converges with the same rate of the identification of $H_s(z)$, as would be expected.

6. CONCLUSION

The analytical framework presented in [4],[5] was extended to the more general case of rational spectrum inputs, which also allowed an unification with the inverse identification of a minimum-phase system, considered in [6],[7]. Desirable further developments would be a deeper understanding of the role of the restricted error surface, which was introduced in this work, and inclusion of the non-minimum phase case of inverse identification.

REFERENCES

[1] J. Treichler, C. R. Johnson, Jr., and M. G. Larimore, *Theory and design of adaptive filters*, Prentice-Hall, Upper Saddle River, 2001.

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

[3] P. A. Regalia, *Adaptive IIR filtering in signal processing and control*, Marcel Dekker, New York, 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, "A new framework for convergence analysis and algorithm development of adaptive IIR filters," *IEEE Transactions on Signal Processing*, vol. 53, no. 8, pp. 3129–3140, Aug. 2005.

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

[7] P. M. S. Burt, "Inverse identification adaptive IIR filtering: convergence speed analysis and successive approximations algorithm," in *Proc. IEEE Icassp*, Honolulu, 2007.

[8] T. Söderström and P. Stoica, "Some properties of the output error method," *Automatica*, vol. 18, pp. 93–99, 1982.

[9] M. Mboup, *(private communication)*, 2007.