EXISTENCE OF STATIONARY POINTS FOR REDUCED-ORDER HYPERSTABLE ADAPTIVE IIR FILTERS

Phillip A. Regalia

Département Signal & Image
Institut National des Télécommunications
9, rue Charles Fourier
91011 Evry cedex France

Mamadou Mboup and Mehdi Ashari

UFR Mathématiques et Informatique
Université René Descartes
45, rue des Saints-Pères
75270 Paris cedex 06 France

ABSTRACT

We establish existence of asymptotic stationary points for a class of adaptive IIR filtering algorithms, including (S)HARF, the Feintuch algorithm, and Landau's algorithm, for reduced-order cases. We show first that the nonlinear equations characterizing a stationary point admit a solution giving rise to a stable transfer function, when the input is white noise. We then show that an analytic procedure to construct the solution may be reduced to the Nevanlinna-Pick interpolation problem. The white noise assumption on the input simplifies the mathematics of an already difficult problem, although the existence proof appears extendable to correlated inputs as well.

1. INTRODUCTION

Most convergence results for adaptive IIR filters assume a sufficient order setting: the degree of the identifier is at least as large as that of the unknown system. Comparatively few convergence results are available for more realistic reduced-order settings; the equations characterizing stationary points are often nonlinear in the filter coefficients. A first step in studying convergence is to isolate the set of stationary points—if any exist—and we establish here existence of a stable transfer function satisfying such equations for a class of algorithms derived from hyperstability theory.

The existence proof assumes only that the unknown system is causal and L_2 stable. When in addition the unknown transfer function is rational, we outline a procedure for analytically constructing the set of transfer functions obtained at the stationary points. This proves quite useful for algorithm testing and validation, as well as for further analytic study of convergence properties.

2. ALGORITHM

Consider a system identification setup where $\{u(\cdot)\}$ is the input sequence and the output sequence $\{y(\cdot)\}$ is generated as

$$y(n) = \sum_{k=0}^{\infty} h_k\, u(n–k) + \zeta(n).$$

Here $\{\zeta(\cdot)\}$ is a stationary second-order process assumed independent of the input $\{u(\cdot)\}$. Let

$$\widehat{H}(z) = \frac{B(z)}{A(z)} = \frac{b_0 + b_1 z + \cdots + b_M z^M}{1 + a_1 z + \cdots + a_M z^M} \tag{1}$$

be a candidate M^{th}-order rational function, where z denotes the backward delay operator: $z\,u(n) = u(n–1)$. The coefficients $\{a_k\}$ and $\{b_k\}$ may be adjusted using the SHARF algorithm [1]:

$$\begin{aligned}
\hat{y}(n) &= -\sum_{k=1}^{M} a_k(n)\,\hat{y}(n–k) + \sum_{k=0}^{M} b_k(n)\,u(n–k) \\
\varepsilon(n) &= [y(n)–\hat{y}(n)] + \sum_{k=1}^{M} c_k\,[y(n–k)–\hat{y}(n–k)] \\
a_k(n+1) &= a_k(n) – \mu\,\varepsilon(n)\,\hat{y}(n–k), \quad k = 1,2,\ldots,M; \\
b_k(n+1) &= b_k(n) + \mu\,\varepsilon(n)\,u(n–k), \quad k = 0,1,\ldots,M.
\end{aligned}$$

Here $\{c_k\}$ are user-chosen constants. The Feintuch algorithm [2] results upon setting $c_k = 0$ for all k; the HARF algorithm [3] is a more sophisticated version using *a posteriori* quantities in the update equations; Landau's algorithm [4] replaces the stepsize μ with a matrix gain sequence, of which a QR variant may be found in [5].

In the special case where the unknown transfer function

$$H(z) = \sum_{k=0}^{\infty} h_k z^k, \qquad |z| < 1, \tag{2}$$

is a rational function of degree M (our chosen filter order), we have $H(z) = \mathcal{B}(z)/\mathcal{A}(z)$ for polynomials $\mathcal{A}(z)$ and $\mathcal{B}(z)$. The (S)HARF algorithms are (weakly) convergent for slow adaptation provided we choose the compensation coefficients in $C(z) = 1 + c_1 z + \cdots + c_M z^M$ to give $C(z)/\mathcal{A}(z)$ as

a strictly positive real transfer function [6]; for the Feintuch algorithm the same applies to $1/\mathcal{A}(z)$, whereas Landau's algorithm is strongly convergent provided $[C(z)/\mathcal{A}(z)]-\frac{1}{2}$ is strictly positive real [7].

We consider here the more realistic reduced-order case, which assumes that, no matter what finite order M we choose for our "identifer" in (1), the degree of $H(z)$ in (2) is greater than M. Identification in the literal sense is then unattainable, and one is led to inquire whether the SHARF algorithm (or its variants) might converge to a useful approximation to $H(z)$. Anderson and Johnson [8] show the nice result that this algorithm class does not diverge, but this alone need not imply convergence to some useful approximation.

In order to study convergence, it is useful to first isolate the set of asymptotic stationary points of the algorithm; these correspond to those values of the filter coefficients $\{a_k\}$ and $\{b_k\}$ which, if held fixed, would yield vanishing mean update terms for the coefficients. For all the algorithms considered, this reads as (E = expectation)

$$\begin{aligned} E[\varepsilon(n)\hat{y}(n-k)] &= 0, \quad k=1,2,\ldots,M; \\ E[\varepsilon(n)u(n-k)] &= 0, \quad k=0,1,\ldots,M. \end{aligned} \tag{3}$$

Although this system is nonlinear in the filter coefficients $\{a_k\}$, solutions may be expressed in the transfer function space. To this end, let

$$V(z)=\frac{a_M+a_{M-1}z+\cdots+a_1 z^{M-1}+z^M}{1+a_1 z+\cdots+a_{M-1}z^{M-1}+a_M z^M} \tag{4}$$

be an M^{th}-order all-pass function having the same poles as $\widehat{H}(z)$. If the input sequence $\{u(\cdot)\}$ is white noise, then from [9, p. 534] $\widehat{H}(z)$ will be a stable transfer function from $\{a_k\}$ and $\{b_k\}$, obtained at a solution to (3), if and only if

$$H(z)-\widehat{H}(z)=z^{M+1}V(z)R(z) \tag{5}$$

for some stable and causal function $R(z)$. The factor z^{M+1} means that the impulse responses of $H(z)$ and $\widehat{H}(z)$ coincide in the first $M+1$ terms, while the factor $V(z)$ implies that $H(z)$ and $\widehat{H}(z)$ take the same values at the reciprocals of the poles of $\widehat{H}(z)$ [as these points are zeros of $V(z)$].

If $V(z)$ is expanded in terms of its impulse response as

$$V(z)=\sum_{k=0}^{\infty} v_k z^k,$$

then one may show [9] that (5) may be rearranged as

$$\begin{bmatrix} h_1 & h_2 & h_3 & \cdots \\ h_2 & h_3 & h_4 & \cdots \\ \vdots & \vdots & \vdots & \ddots \\ h_M & h_{M+1} & h_{M+2} & \cdots \end{bmatrix} \begin{bmatrix} v_0 \\ v_1 \\ v_2 \\ \vdots \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ \vdots \\ 0 \end{bmatrix} \tag{6}$$

or, in inner product form,

$$\langle H(z), z^k V(z)\rangle = 0, \quad k=1,2,\ldots,M.$$

If we can find an M^{th}-order all-pass function fulfilling these constraints, then we shall have the poles of $\widehat{H}(z)=B(z)/A(z)$. The numerator $B(z)$ can then be found by solving

$$\frac{B(1/z_i)}{A(1/z_i)}=H(1/z_i)$$

where $\{z_i\}$ are the poles of $\widehat{H}(z)$; this yields a linear system in the coefficients $\{b_k\}$, which is always solvable. The resulting $\widehat{H}(z)$ then fulfills (5), and accordingly we concentrate on finding an M^{th}-order all-pass function $V(z)$ fulfilling the orthogonality constraints (6). The next section establishes existence of such an all-pass function whenever $H(z)$ has a square-summable impulse response. Section 4 then shows how to construct such a $V(z)$ in the special case where $H(z)$ is rational.

3. EXISTENCE IN THE GENERAL CASE

Rather than parametrize $V(z)$ in direct form, as in (4), we find it more convenient to use its reflection coefficients. Let $W_M(z)=V(z)$ [where $M=\deg V(z)$], and determine lower order all-pass functions $W_k(z)$ from the Schur recursion

$$W_{k-1}(z)=z^{-1}\frac{W_k(z)-s_k}{1-s_k W_k(z)}, \quad s_k=W_k(0), \tag{7}$$

for $k=M,\ \ldots,\ 1$. The all-pass function $W_M(z)=V(z)$ is stable with degree M if and only if $-1<s_k<1$ for $k=M,\ldots,1$. We shall write $V(z,s_M,\ldots,s_1)$ when necessary to indicate the dependence of $V(z)$ on $\{s_k\}$.

The following properties are obtained readily from (7):

1. If $s_k=\pm 1$, but $|s_i|<1$ for $i>k$, then $V(z)$ degenerates to a stable all-pass function of degree $M-k$, parametrized by $s_{k+1},\ \ldots,\ s_M$.
2. If we negate s_k at its boundary, then for all z

$$V(z,s_M,\ldots,s_{k+1},s_k)\Big|_{s_k=+1} = -V(z,-s_M,\ldots,-s_{k+1},s_k)\Big|_{s_k=-1}$$

Introduce now the vector-valued function

$$\mathbf{F}(s_M,\ldots,s_1)=\begin{bmatrix} \langle H(z),zV(z)\rangle \\ \vdots \\ \langle H(z),z^M V(z)\rangle \end{bmatrix},$$

which is continuous in the reflection coefficients $\{s_k\}$ which parametrize $V(z)$. We shall vary the reflection coefficients over the *closed* hypercube

$$|s_k|\le 1, \quad k=1,2,\ldots,M,$$

and deduce that $\mathbf{F}$ must vanish somewhere in this domain. For any choice of $\{s_k\}$, $\mathbf{F}$ is bounded in Euclidean norm as

$$\|\mathbf{F}\|^2 \le \sum_{k=1}^{\infty} h_k^2. \tag{8}$$

This follows because the functions $\{z^k V(z)\}$ are orthonormal; the elements of $\mathbf{F}$ are the coefficients of projection of $H(z)$ on these functions, whose sum of squares can never exceed the L_2-norm squared of the function so projected.

From property 1 above, we have

$$\lim_{s_M \to +1} V(z) = +1, \quad \lim_{s_M \to -1} V(z) = -1,$$

irrespective of the remaining reflection coefficients. As such,

$$\lim_{s_M \to +1} \mathbf{F} = \begin{bmatrix} h_1 \\ \vdots \\ h_M \end{bmatrix}, \quad \lim_{s_M \to -1} \mathbf{F} = - \begin{bmatrix} h_1 \\ \vdots \\ h_M \end{bmatrix}.$$

Denote the first point by $\mathcal{B}_M$; the second becomes $-\mathcal{B}_M$. Choose now any set of values for $s_{M-1}, \ldots, s_1$, and let s_M vary from +1 to −1. We generate a curve whose path depends on the chosen values of $s_{M-1}, \ldots, s_1$, but each curve begins at the point $\mathcal{B}_M$ and ends at $-\mathcal{B}_M$.

Consider the specific curve

$$\mathcal{B}_{M-1} \stackrel{\Delta}{=} \{\mathbf{F} : -1 \le s_M \le 1, s_{M-1} = +1\}.$$

By property 1 above, this curve does not depend on $s_{M-2}, \ldots, s_1$. And by property 2, defining a similar curve with $s_{M-1} = -1$ must generate $-\mathcal{B}_{M-1}$:

$$-\mathcal{B}_{M-1} = \{\mathbf{F} : -1 \le s_M \le 1, s_{M-1} = -1\}.$$

Now set $s_{M-2} = +1$, and let s_M and s_{M-1} vary between +1 and −1:

$$\mathcal{B}_{M-2} = \{\mathbf{F} : -1 \le s_M, s_{M-1} \le +1, s_{M-2} = +1\}.$$

This surface must contain $\mathcal{B}_{M-1}$ and $-\mathcal{B}_{M-1}$ as edges (obtained with $s_{M-1} = \pm 1$) and $\pm\mathcal{B}_M$ as endpoints (with $s_M = \pm 1$). Again from property 2 above, a similar surface with $s_{M-2} = -1$ must generate $-\mathcal{B}_{M-2}$:

$$-\mathcal{B}_{M-2} = \{\mathbf{F} : -1 \le s_M, s_{M-1} \le +1, s_{M-2} = -1\}.$$

Continuing this procedure for s_{M-3} down to s_1, we generate sucessively higher dimensional surfaces $\pm\mathcal{B}_{M-3}$ down to $\pm\mathcal{B}_1$; the surfaces $\pm\mathcal{B}_k$ will always have $\mathcal{B}_{k+1}$ and $-\mathcal{B}_{k+1}$ as edges, and ultimately $\pm\mathcal{B}_M$ as endpoints.

Now, the M-dimensional surfaces $\mathcal{B}_1$ and $-\mathcal{B}_1$ will lie on exactly opposite sides of the origin, and since they share common M–1-dimensional edges $\mathcal{B}_2$ and $-\mathcal{B}_2$, their union $\mathcal{B}_1 \cup -\mathcal{B}_1$ must completely surround the origin.

Now vary s_1 from +1 to −1 and, for each value of s_1, let

$$\mathcal{B}_0(s_1) = \{\mathbf{F} : -1 \le s_k \le +1, \quad k = 2, \ldots, M\}$$

denote the surface obtained as the remaining reflection coefficients exhaust their hypercube. This gives $\mathcal{B}_0(+1) = \mathcal{B}_1$ and $\mathcal{B}_0(-1) = -\mathcal{B}_1$. The surfaces $\pm\mathcal{B}_2$ are still edges of $\mathcal{B}_0(s_1)$ for all s_1, and varying s_1 from +1 to −1 causes $\mathcal{B}_0(s_1)$ to continuously deform from $\mathcal{B}_1$ to $-\mathcal{B}_1$; an intermediate value of s_1 must yield an intermediate surface passing through the origin. This shows that the range space of $\mathbf{F}$ includes the origin, the desired goal.

We remark that $\mathbf{F}$ may vanish for some boundary point $|s_k| = 1$; in this case, we will have found an all-pass function $V(z)$ of degree less than M fulfilling the orthogonality constraints (6). In this case, the transfer function $\widehat{H}(z)$ reconstruced from $V(z)$ will have pole-zero cancellations, since $\deg \widehat{H}(z) = \deg V(z)$ may be shown.

4. ANALYTIC CONSTRUCTION

With existence established whenever $H(z)$ is L_2 stable, we pursue construction of $V(z)$ in the special case where $H(z)$ is a rational function, of degree p, say. To this end, let $(\mathbf{A}, \mathbf{b}, \mathbf{c})$ be a minimal realization of the strictly causal part of $H(z)$:

$$\sum_{k=1}^{\infty} h_k z^k = z\,\mathbf{c}(\mathbf{I} - z\mathbf{A})^{-1}\mathbf{b} \quad \Rightarrow \quad h_k = \mathbf{c}\mathbf{A}^{k-1}\mathbf{b}.$$

The orthogonality constraint (6) then reads as

$$\begin{bmatrix} 0 \\ 0 \\ \vdots \\ 0 \end{bmatrix} = \underbrace{\begin{bmatrix} \mathbf{c} \\ \mathbf{cA} \\ \vdots \\ \mathbf{cA}^{M-1} \end{bmatrix}}_{\mathcal{O}_{M\times p}} \begin{bmatrix} \mathbf{b} & \mathbf{Ab} & \mathbf{A}^2\mathbf{b} & \cdots \end{bmatrix} \begin{bmatrix} v_0 \\ v_1 \\ v_2 \\ \vdots \end{bmatrix}. \qquad (9)$$

in which the $M \times p$ observability matrix $\mathcal{O}_{M\times p}$ is distinguished. To simplify the procedure, suppose the eigenvalues of $\mathbf{A}$ are distinct, so that $\mathbf{A}$ is diagonalizable.[1] By a similarity transformation if necessary, we may consider a particular realization for which

$$\mathbf{A} = \begin{bmatrix} \lambda_1 & & & \\ & \lambda_2 & & \\ & & \ddots & \\ & & & \lambda_p \end{bmatrix}, \quad \mathbf{b} = \begin{bmatrix} 1 \\ 1 \\ \vdots \\ 1 \end{bmatrix}, \qquad (10)$$

and $\mathbf{c}$ contains the pole residues. This then gives in (9)

$$M \text{ terms} \left\{ \begin{bmatrix} 0 \\ 0 \\ \vdots \\ 0 \end{bmatrix} = \underbrace{\begin{bmatrix} \mathbf{c} \\ \mathbf{cA} \\ \vdots \\ \mathbf{cA}^{M-1} \end{bmatrix}}_{\mathcal{O}_{M\times p}} \begin{bmatrix} V(\lambda_1) \\ V(\lambda_2) \\ \vdots \\ V(\lambda_p) \end{bmatrix}. \right.$$

Seen from this angle, we now seek an M^{th}-order all-pass function which "interpolates" the nullspace of the observability matrix.

[1] When $\mathbf{A}$ has repeated eigenvalues, one may replace (10) with a controllable Jordan form, and the rest carries through with straightforward modifications.

To facilitate the procedure to follow, consider first the special case in which $p = M+1$. The matrix $\mathcal{O}_{M\times(M+1)}$ then has a nullspace of dimension one and, since $\mathcal{O}$ has a Vandermonde structure, a particular vector $\mathbf{x} = [x_1, \ldots, x_p]^t$ in its nullspace may be written as

$$x_i = 1 \Bigg/ \left(c_i \prod_{\substack{k=1 \\ k\neq i}}^{p} (\lambda_i - \lambda_k) \right). \qquad (11)$$

All null vectors may thus be written as $\alpha\mathbf{x}$, where α is a free scale factor.

Consider now the problem of finding a Schur function $S(z)$ (i.e., analytic in $|z| < 1$ and bounded by unit modulus) which fulfills the interpolation conditions

$$S(\lambda_i) = \alpha x_i, \qquad i = 1, 2, \ldots, p.$$

This is the classic Nevanlinna-Pick interpolation problem, and a solution exists if and only if a certain $p \times p$ Pick matrix $\mathbf{P}$, with (i, j)-element

$$\mathbf{P}_{ij} = \frac{1 - |\alpha|^2 x_i x_j^*}{1 - \lambda_i \lambda_j^*}, \qquad i, j = 1, 2, \ldots, p,$$

is positive (semi-) definite [10]. (Here "*" denotes Hermitian transposition). The solution set degenerates to an all-pass function $V(z)$ of degree M (or less) if and only if $\mathbf{P}$ becomes positive semi-definite and singular [10]; in that case, $\deg V(z) = \operatorname{rank} \mathbf{P}$. Now, one can readily check that $\mathbf{P}$ can be decomposed as $\mathbf{P} = \mathbf{Q} - |\alpha|^2\mathbf{R}$, where

$$\mathbf{Q}_{ij} = \frac{1}{1 - \lambda_i \lambda_j^*}, \quad \mathbf{R}_{ij} = \frac{x_i x_j^*}{1 - \lambda_i \lambda_j^*}.$$

The matrix $\mathbf{R}$ is positive definite since all the x_i are nonzero [cf. (11)]. There thus exists an invertible matrix $\mathbf{T}$ for which $\mathbf{T}\mathbf{R}\mathbf{T}^* = \mathbf{I}_p$; a congruence transformation applied to $\mathbf{P}$ then gives

$$\mathbf{T}\mathbf{P}\mathbf{T}^* = \mathbf{T}\mathbf{Q}\mathbf{T}^* - |\alpha|^2\mathbf{I},$$

from which it is clear that choosing $|\alpha|^2 = \lambda_{min}[\mathbf{T}\mathbf{Q}\mathbf{T}^*]$ renders $\mathbf{P}$ positive semi-definite. We now have the interpolation points $\{\lambda_i\}$ and the interpolation values $\{\alpha x_i\}$ for an M^{th} (or lower) order all-pass function $V(z)$. The algorithm of [10] (among others) can then be used to construct $V(z)$.

When $p > M+1$, the dimension of the nullspace of $\mathcal{O}_{M\times p}$ increases to $p - M$. Let $\mathbf{x}_1, \ldots \mathbf{x}_{p-M}$ be any basis for this nullspace. We then seek an all-pass function $V(z)$ of degree M (or less) for which

$$\begin{bmatrix} V(\lambda_1) \\ \vdots \\ V(\lambda_p) \end{bmatrix} = \mathbf{y} = \alpha_1 \mathbf{x}_1 + \alpha_2 \mathbf{x}_2 + \cdots + \alpha_{p-M} \mathbf{x}_{p-M},$$

for some choice of the constants $\{\alpha_k\}$; these must be chosen in such a way that the Pick matrix

$$\mathbf{P}_{ij} = \frac{1 - y_i y_j^*}{1 - \lambda_i \lambda_j^*}$$

becomes positive semi-definite of rank M or less. Such a choice of $\{\alpha_k\}$ exists since the previous section showed that $V(z)$ exists. Uniqueness of the $\{\alpha_k\}$, however, is not in general ensured, which complicates a direct procedure for determining these constants. Once found, though, we again have the interpolation points $\{\lambda_i\}$ and the interpolation values $\{y_i\}$ of an M^{th} (or lower) order all-pass function $V(z)$, from which a solution may readily be constructed.

5. REFERENCES

[1] M. G. Larimore, J. R. Treichler, and C. R. Johnson, Jr., "SHARF: An algorithm for adapting IIR digital filters," *IEEE Trans. Acoustics, Speech, and Signal Processing*, vol. 28, pp. 428–440, August 1980.

[2] P. L. Feintuch, "An adaptive recursive LMS filter," *Proc. IEEE*, vol. 64, pp. 1622–1624, Nov. 1976.

[3] C. R. Johnson, Jr., "A convergence proof for a hyperstable adaptive recursive filter," *IEEE Trans. Information Theory*, vol. 25, pp. 745–749, Nov. 1979.

[4] I. D. Landau, "Unbiased recursive identification using model reference adaptive techniques," *IEEE Trans. Automatic Control*, vol. 21, pp. 194–202, April 1976.

[5] M. Harteneck, R. W. Stewart, J. G. McWhirter, and I. K. Proudler, "An effective approach to adaptive IIR filtering," *Proc. ICASSP*, Atlanta, GA, 1996; pp. 1634–1637.

[6] H. Fan, "Application of Benveniste's convergence results in the study of adaptive IIR filtering algorithms," *IEEE Trans. Inf. Th.*, vol. 34, pp. 692–709, July 1988.

[7] L. Ljung, "On positive real transfer functions and the convergence of some recursive schemes," *IEEE Trans. Automatic Control*, vol. 22, pp. 539–551, August 1977.

[8] B. D. O. Anderson and C. R. Johnson, Jr., "On reduced-order adaptive output error identification and adaptive IIR filtering," *IEEE Trans. Automatic Control*, vol. 27, pp. 927–933, Aug. 1982.

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

[10] Ph. Delsarte, Y. Genin, and Y. Kamp, "On the role of the Nevanlinna-Pick problem in circuit and system theory," *Int. J. Circuit Theory and Applications*, vol. 9, pp. 177–187, 1981.