Proc. IFAC Symp. on Nonlinear Control System Design, (1)455–460, Lake Tahoe, June 95

A CLASS OF ADAPTIVE NONLINEAR H^{∞}–FILTERS WITH GUARANTEED l_2-STABILITY

A. H. SAYED and M. RUPP

Department of Electrical and Computer Engineering
University of California, Santa Barbara, CA.

Abstract. We pose an identification problem that involves a nonlinear output equation and proceed to suggest an approximate linear solution. The approximation is obtained in two stages. We first replace the nonlinear functional by a linear relation, thus reducing the problem to a standard linear H^{∞}–setting. We then suggest constructing an approximation that results in an overall feedback structure in order to meet desired robustness and stability properties. By combining the linear H^{∞} solution with a widely-used small gain theorem we show, under suitable conditions, that the approximate solution still leads to a filter with guaranteed l_2-stability. An example in the context of pole-zero (or IIR) system identification is discussed in details. The proposed structure is also shown to include, as special cases, several adaptive filters that have been employed earlier in the literature in the context of IIR modeling. In particular, two algorithms due to Feintuch and Landau, as well as the so-called pseudo-linear regression algorithm, are discussed within the framework proposed herein.

Key Words. Nonlinear output equation, small-gain theorem, l_2-stability, feedback structure, H^{∞}-filter, stochastic-gradient filters, IIR modeling.

INTRODUCTION

Considerable research activity has been devoted over the last two decades to the analysis and design of adaptive algorithms in both signal processing and control applications. In particular, several ingenious methods have been proposed for the performance and stability analysis of the varied adaptive schemes. Among these, the most notable are the hyperstability results of Popov, a nice account of which is given by (Landau, 1979), the ODE approach of Ljung (Ljung and Söderström, 1983), and the related class of averaging methods for trajectory approximation, as described in the recent book by (Solo and Kong, 1995).

Correspondingly, in the last decade, there has been an explosion of research in the areas of robust filtering and control, as indicated by some of the references at the end of this paper. A major concern here is the design of filters and controllers that are robust to parameter variations and to exogenous signals. In the filtering context, for instance, it is currently known how to design estimators with bounded H^{∞} (or 2–induced) norms, and the available results provide us with both (i) solvability and existence conditions, as well as (ii) recursive methods for the construction of a solution.

[0] * This material is based on work supported by the National Science Foundation under Award No. MIP-9409319. The work of M. Rupp is also supported in part by a scholarship from DAAD (German Academic Exchange Service) as well as the scientific division of NATO (Wissenschaftsausschuß der NATO).

Motivated by these results, we take here an alternative look at the performance analysis of adaptive schemes. The discussion in this paper is fundamentally based on a useful tool in system analysis widely known as the small gain theorem. In loose terms, the theorem states that a feedback interconnection of two systems is stable if the product of their gains (or induced norms) is less than unity. While this statement can be reformulated in terms of a hyperstability (or passivity) result, the analysis provided in this paper has several additional features.

First, by relying on the small gain theorem, we can advantageously exploit the wealth of results available in the H^{∞}–setting. In other words, since we essentially know how to design a robust filter, i.e., a system with a bounded gain, we can then guarantee an overall stable interconnection by imposing a condition on the gain of the feedback system. This is especially helpful in the design (i.e., synthesis) phase. In later sections, for example, we shall show that the adaptive scheme proposed herein includes, as special cases, several earlier algorithms, which, interestingly, also establishes that these earlier schemes can be regarded as special H^{∞}–filters.

Secondly, although the feedback nature of most adaptive schemes has been advantageously exploited in earlier places in the literature (see, e.g., (Landau, 1979; Ljung and Söderström, 1983)), the feedback configuration in this paper is of a different nature. It does not only refer to the fact that the update equations of an adaptive scheme can be put into a feedback form, but is instead motivated by our concern with the overall robustness performance of the algorithm. By this we mean that the feedback configuration is defined in such a way so as to explicitly consider the effect of *both* the measurement noise and the uncertainty in the initial guess on the algorithm performance.

Thirdly, the approach taken herein also addresses some new issues. In particular, we shall provide in a later section a stability analysis of two IIR adaptive schemes that are often attributed to (Landau, 1979) and (Feintuch, 1976). While a sufficient stability condition is available for Landau's scheme in terms of a positive-realness constraint (e.g., (Solo and Kong, 1995; pp.146–150)), there does not seem to exist a similar analysis for the closely related, yet different, Feintuch's algorithm. An explanation is provided here by showing that Feintuch's recursion requires an additional condition on the data. This is obtained by establishing the following interesting fact: Landau's scheme is shown to be a special case of a so-called *aposteriori* H^∞-filter while Feintuch's algorithm is shown to be a special case of a so-called *apriori* H^∞-filter. It is known in $H^\infty-$theory that the solvability and existence conditions for both filters are different. Here we show that in Landau's case, the condition trivializes and is therefore unnecessary, but it remains in Feintuch's case and is therefore required, along with a positive-realness condition.

Finally, and although not treated in this paper, we may remark that the feedback analysis suggested herein can be further shown to provide an interpretation of most adaptive schemes in terms of an interconnection that consists of a *lossless* (i.e., energy preserving) feedforward mapping and of either a memoryless or a dynamic feedback mapping, both of which are allowed to be time-variant. In this context, some interesting energy arguments can be invoked to further analyze the performance of the algorithms.

PROBLEM STATEMENT

Consider a collection of noisy measurements $\{y(i)\}_{i=0}^N$ that are related to a (column) vector of unknown parameters $\mathbf{w}$ via the nonlinear relation $y(i) = \mathbf{h}_i(\mathbf{w})\mathbf{w} + v(i)$. Here $v(i)$ stands for the noisy component at the discrete time instant i, and $\mathbf{h}_i(\mathbf{w})$ denotes a time-variant row vector whose entries are themselves functions of the unknown entries of $\mathbf{w}$. For notational convenience, we shall use boldface letters to denote vectors.

The measurements $\{y(i)\}$ can be alternatively interpreted as the noisy outputs of a simple state-space model of the form

$$\mathbf{x}_{i+1} = \mathbf{x}_i, \quad y(i) = \mathbf{h}_i(\mathbf{x}_i)\mathbf{x}_i + v(i), \quad \mathbf{x}_0 = \mathbf{w}. \qquad (1)$$

Let $z(i)$ denote a desired combination of the unknown vector $\mathbf{w}$, say $z(i) = \mathbf{g}_i(\mathbf{w})\mathbf{w} = \mathbf{g}_i(\mathbf{x}_i)\mathbf{x}_i$, and let $\hat{z}(i|i)$ denote an estimate for $z(i)$ that is dependent on the observation data $\{y(\cdot)\}$ up to time i, according to the following criterion. Let Π_0 be a positive-definite matrix and choose any initial guess for $\mathbf{w}$, which we shall denote by $\bar{\mathbf{x}}_0$ or $\mathbf{w}_{-1}$. For every time instant i, define the ratio: $r(i) \triangleq$

$$\frac{\sum_{j=0}^{i} |z(j) - \hat{z}(j|j)|^2}{(\mathbf{x}_0 - \bar{\mathbf{x}}_0)^* \Pi_0^{-1} (\mathbf{x}_0 - \bar{\mathbf{x}}_0) + \sum_{j=0}^{i} |y(j) - \mathbf{h}_j(\mathbf{x}_j)\mathbf{x}_j|^2},$$

which provides a relative measure of the energies due to the estimation error in $z(\cdot)$, the initial guess $\bar{\mathbf{x}}_0$, and the disturbance $v(\cdot)$. The objective is to determine, if possible, estimates $\hat{z}(j|j)$, for $j = 0, 1, \ldots, N$, so as to guarantee that, *for all* $\bar{\mathbf{x}}_0$ *and* $v(\cdot)$, the ratios $r(i)$ will be all bounded by a given positive constant, say $r(i) < \gamma^2$ for $0 \le i \le N$. These conditions guarantee that the transfer operator that maps the disturbances, $\{\Pi_0^{-1/2}(\mathbf{x}_0 - \bar{\mathbf{x}}_0),\ v(i)\}_{i=0}^N$ to the resulting estimation errors, $\{\tilde{z}(i|i) = z(i) - \hat{z}(i|i)\}_{i=0}^N$, will have a 2−induced (or $H^\infty-$)norm bounded by γ (i.e., is l_2 stable).

AN APPROXIMATE LINEAR MODEL

The presence of the $\mathbf{w}-$dependent (nonlinear) functions $\mathbf{h}_i(\cdot)$ and $\mathbf{g}_i(\cdot)$, in both the numerator and the denominator of the cost ratio $r(i)$, complicates the problem at hand. For this reason, we shall proceed here in two steps.

We shall first invoke an approximation that will replace the nonlinear inner products by linear relations. This will allow us to reduce the problem formulation to a standard linear $H^\infty-$setting, which has been widely studied in the literature. But a difficulty here is to exhibit an approximation that will still provide, under suitable conditions, an approximate linear filter with a guaranteed l_2-stable mapping from the <u>*original*</u> disturbances $\{\Pi_0^{-1/2}(\mathbf{x}_0 - \bar{\mathbf{x}}_0),\ v(i)\}_{i=0}^N$ to the resulting (yet modified) estimation errors, say $\{\tilde{z}'(i|i)\}_{i=0}^N$ (see Theorem 1 further ahead); we shall also argue later that, in important applications, the use of the modified estimation errors $\tilde{z}'(\cdot)$ does not affect the overall desired performance (see, e.g., Algorithm 2).

To begin with, assume we have available at each time instant i *row* vector estimates $\hat{\mathbf{h}}_i$ and $\hat{\mathbf{g}}_i$ for $\mathbf{h}_i(\mathbf{w})$ and $\mathbf{g}_i(\mathbf{w})$, respectively. These estimates may be computed in different forms. One possibility is the so-called bootstrap technique that is often used in system identification (e.g., (Ljung and Söderström, 1983)). This technique assumes that we have access to recursive estimates of the parameter vector $\mathbf{w}$, which are then employed in approximating $\mathbf{h}_i(\cdot)$ and $\mathbf{g}_i(\cdot)$; if we let $\mathbf{w}_{i-1}$ denote the estimate of $\mathbf{w}$ that is based on the data up to and including time $i-1$ (we also denote this by $\hat{\mathbf{x}}_{i-1|i-1}$), then the bootstrap method computes $\hat{\mathbf{h}}_i$ and $\hat{\mathbf{g}}_i$ via $\hat{\mathbf{h}}_i = \mathbf{h}_i(\mathbf{w}_{i-1})$ and $\hat{\mathbf{g}}_i = \mathbf{g}_i(\mathbf{w}_{i-1})$.

But many other possibilities for choosing the estimates $\hat{\mathbf{h}}_i$ and $\hat{\mathbf{g}}_i$ clearly exist. The bootstrap method need not guarantee an overall l_2-stable filter and a challenging task is to provide alternative constructions that guarantee a desired performance. We shall exhibit in the sequel one such example that guarantees the overall l_2-stability of the resulting filter despite the linear approximation.

But for now, let us simply assume that, in some way, we have available estimates $\hat{\mathbf{h}}_i$ and $\hat{\mathbf{g}}_i$. This then allows us to replace the earlier ratio $r(i)$ in (2) by a linearized version given by $r'(i) =$

$$\frac{\sum_{j=0}^{i} |\hat{\mathbf{g}}_j \mathbf{x}_j - \hat{\mathbf{g}}_j \hat{\mathbf{x}}_{j|j}|^2}{(\mathbf{x}_0 - \bar{\mathbf{x}}_0)^* \Pi_0^{-1} (\mathbf{x}_0 - \bar{\mathbf{x}}_0) + \sum_{j=0}^{i} |y(j) - \hat{\mathbf{h}}_j \mathbf{x}_j|^2}.$$

We now proceed to determine state-estimates $\hat{\mathbf{x}}_{j|j}$ so as to guarantee that the above ratio will be bounded by a given constant, say ξ^2, $r'(i) < \xi^2$ for $0 \leq i \leq N$. The approximate problem is now formulated in a form that is standard in the $H^\infty-$ literature. One possible solution is the following $H^\infty-$adaptive filter, e.g., (Yaesh and Shaked, 1991; Hassibi, Sayed, and Kailath, 1993*b*).

<u>ALGORITHM 1.</u> [APOSTERIORI FILTER] *The estimates* $\hat{\mathbf{x}}_{j|j}$ *(also equal to* $\mathbf{w}_j$*) can be evaluated recursively as follows:* $\hat{\mathbf{x}}_{j|j} =$

$$\hat{\mathbf{x}}_{j-1|j-1} + \mathbf{P}_j \hat{\mathbf{h}}_j^* \left[1 + \hat{\mathbf{h}}_j \mathbf{P}_j \hat{\mathbf{h}}_j^*\right]^{-1} \left[y(j) - \hat{\mathbf{h}}_j \hat{\mathbf{x}}_{j-1|j-1}\right],$$

with $\hat{\mathbf{x}}_{-1|-1} = \bar{\mathbf{x}}_0$ *and where* $\mathbf{P}_j$ *satisfies the Riccati difference equation:* $\mathbf{P}_0 = \Pi_0$,

$$\mathbf{P}_{j+1} = \mathbf{P}_j - \mathbf{P}_j \begin{bmatrix} \hat{\mathbf{g}}_j^* & \hat{\mathbf{h}}_j^* \end{bmatrix} \mathbf{R}_{e,j}^{-1} \begin{bmatrix} \hat{\mathbf{g}}_j \\ \hat{\mathbf{h}}_j \end{bmatrix} \mathbf{P}_j,$$

$$\mathbf{R}_{e,j} = \left\{ \begin{bmatrix} -\xi^2 & \mathbf{0} \\ \mathbf{0} & 1 \end{bmatrix} + \begin{bmatrix} \hat{\mathbf{g}}_j \\ \hat{\mathbf{h}}_j \end{bmatrix} \mathbf{P}_j \begin{bmatrix} \hat{\mathbf{g}}_j^* & \hat{\mathbf{h}}_j^* \end{bmatrix} \right\}.$$

Moreover, this solution guarantees $\{r'(i) < \xi^2\}_{i=0}^{N}$ *if, and only if,* $\mathbf{P}_{j+1} > \mathbf{0}$ *for* $j = 0, 1, \ldots, N$.

The above filter is a so-called *filtered* or *aposteriori* version. There is a related *estimation* or *apriori* version, which estimates $\hat{\mathbf{g}}_i \mathbf{x}_i$ by using only the data $\{y(j)\}$ up to $(i-1)$ rather than i, i.e., it tries to bound the ratios $r''(i) =$

$$\frac{\sum_{j=0}^{i} |\hat{\mathbf{g}}_j \mathbf{x}_j - \hat{\mathbf{g}}_j \hat{\mathbf{x}}_{j|j-1}|^2}{(\mathbf{x}_0 - \bar{\mathbf{x}}_0)^* \Pi_0^{-1} (\mathbf{x}_0 - \bar{\mathbf{x}}_0) + \sum_{j=0}^{i-1} |y(j) - \hat{\mathbf{h}}_j \mathbf{x}_j|^2},$$

where $\hat{\mathbf{x}}_{j|j-1}$ denotes an estimate for $\mathbf{x}_j$ that is based on the data up to time $j-1$. It is often denoted by the shorthand notation $\hat{\mathbf{x}}_j$ and can also be interpreted as equal to $\mathbf{w}_{j-1}$.

<u>ALGORITHM 1A.</u> [APRIORI FILTER] *The estimates* $\hat{\mathbf{x}}_{j|j-1}$ *can be evaluated recursively as follows. Let* $\tilde{\mathbf{P}}_j = \left[\mathbf{P}_j^{-1} - \xi^{-2} \hat{\mathbf{g}}_j^* \hat{\mathbf{g}}_j\right]^{-1}$. *Then* $\hat{\mathbf{x}}_{j+1|j} =$

$$\hat{\mathbf{x}}_{j|j-1} + \tilde{\mathbf{P}}_j \hat{\mathbf{h}}_j^* \left[1 + \hat{\mathbf{h}}_j \tilde{\mathbf{P}}_j \hat{\mathbf{h}}_j^*\right]^{-1} \left[y(j) - \hat{\mathbf{h}}_j \hat{\mathbf{x}}_{j|j-1}\right],$$

where $\hat{\mathbf{x}}_{0|-1} = \bar{\mathbf{x}}_0$ *and* $\mathbf{P}_j$ *is as above. This solution guarantees* $\{r''(i) < \xi^2\}_{i=0}^{N}$ *if, and only if,* $\tilde{\mathbf{P}}_j > \mathbf{0}$ *for* $j = 0, 1, \ldots, N$.

In the following discussion, we shall mainly focus on the aposteriori version, viz., Algorithm 1. We shall return however to the second version (Algorithm 1A) in the last section when we discuss Feintuch's algorithm.

CONVERGENCE AND l_2-STABILITY OF THE APPROXIMATE FILTER

The numerator in the modified ratio $r'(i)$ includes the *aposteriori*-error term $\hat{e}_p(j) = \hat{\mathbf{g}}_j \mathbf{w} - \hat{\mathbf{g}}_j \mathbf{w}_j$, while the denominator includes energy terms of the form $|y(j) - \hat{\mathbf{h}}_j \mathbf{w}|^2$, with the estimate $\hat{\mathbf{h}}_j$ replacing the true function $\mathbf{h}_j(\mathbf{w})$. To clarify the implications of this approximation, note that we can write

$$y(j) - \hat{\mathbf{h}}_j \mathbf{w} = \left[\mathbf{h}_j(\mathbf{w}) - \hat{\mathbf{h}}_j\right] \mathbf{w} + v(j) \ \stackrel{\Delta}{=} \ \hat{v}(j), \qquad (2)$$

which shows that the extra (approximation error) term $\left[\mathbf{h}_j(\mathbf{w}) - \hat{\mathbf{h}}_j\right] \mathbf{w}$ is added to the noise component $v(j)$ when compared to the denominator of $r(i)$.

Now assume that the estimates $\hat{\mathbf{h}}_j$ and $\hat{\mathbf{g}}_j$ are chosen in such a way (see example in next section) so as to result in an explicit relation between $\hat{v}(j)$ and $\{v(j), \hat{e}_p(j)\}$. This is shown in the feedback structure in Figure 1 where we have denoted the difference $(\mathbf{w} - \mathbf{w}_{-1})$ by $\tilde{\mathbf{w}}_{-1}$. The symbol $\mathcal{T}_N$ denotes the (causal) operator that maps the (modified) disturbances $\{\Pi_0^{-1/2}(\mathbf{w} - \mathbf{w}_{-1}),\ \hat{v}(j)\}_{j=0}^{N}$ to the estimation errors $\{\hat{e}_p(j)\}_{j=0}^{N}$. In view of Algorithm 1, this operator is constructed so as to have a $2-$induced (or $H^\infty-$) norm bounded by ξ. The symbols $\mathcal{F}_N$ and $\mathcal{V}_N$ denote causal (linear) operators relating the sequence $\{\hat{v}(j)\}_{j=0}^{N}$ to the sequences $\{v(j), \hat{e}_p(j)\}_{j=0}^{N}$ (assuming zero initial conditions).

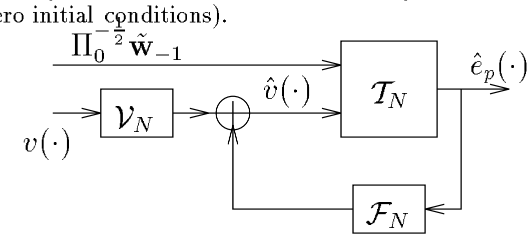

Fig. 1. *Feedback structure of the linearized solution.*

Let $\|\cdot\|_\infty$ denote the $2-$induced norm of a (linear) operator. Using the triangle inequality of norms, the fact that $r'(N) \leq \xi^2$, along with (3) below, we conclude the following.

<u>THEOREM 1.</u> *Consider the recursive solution of Algorithm 1 and assume that the estimates* $\hat{\mathbf{h}}_j$ *and* $\hat{\mathbf{g}}_j$ *result in a feedback structure of the form indicated in Figure 1. If the following condition is satisfied,*

$$\|\mathcal{T}_N\|_\infty \|\mathcal{F}_N\|_\infty < 1, \qquad (3)$$

then the mapping from $\{\Pi_0^{-1/2} \tilde{\mathbf{w}}_{-1}, v(\cdot)\}$ *to* $\{\hat{e}_p(\cdot)\}$ *is* l_2-*stable with finite gain in the following sense,*

$$\sqrt{\sum_{j=0}^{N} |\hat{e}_p(j)|^2} \ \leq \qquad (4)$$

$$k_N \left[\|\Pi_0^{-1/2} \tilde{\mathbf{w}}_{-1}\|_2 + \|\mathcal{V}_N\|_\infty \sqrt{\sum_{j=0}^{N} |v(j)|^2} \ + \ \beta'_N \right],$$

where $k_N \stackrel{\Delta}{=} \|\mathcal{T}_N\|_\infty / (1 - \|\mathcal{T}_N\|_\infty \|\mathcal{F}_N\|_\infty)$, *and* β'_N *is a nonnegative finite constant that accounts for possibly nonzero initial conditions.*

Expression (4) establishes that the map from $\{\Pi_0^{-1/2}\tilde{\mathbf{w}}_{-1}, \|\mathcal{V}_N\|_\infty v(\cdot)\}$ to $\{\hat{e}_p(\cdot)\}$ is l_2-stable (with finite gain – see, e.g, (Vidyasagar, 1993; p.365) and (Khalil, 1992; 213–215)). Note that this map involves the original disturbances $\{v(\cdot)\}$ rather than $\{\hat{v}(\cdot)\}$. Note also that since we already know that $\|\mathcal{T}_N\|_\infty < \xi$, then a sufficient condition for (3) to hold is to require $\|\mathcal{F}_N\|_\infty \le 1/\xi$. These results can be regarded as an immediate application of the small gain theorem (see, e.g., (Khalil, 1992; p.214) and (Vidyasagar, 1993; p.337)) to the feedback connection of Figure 1.

In the limit, if the noise sequence $\{v(\cdot)\}$ has finite energy, i.e., $\sum_{j=0}^{\infty} |v(j)|^2 < \infty$, and if we further let $\mathcal{T}$, $\mathcal{F}$ and $\mathcal{V}$ denote the (corresponding) semi-infinite operators satisfying (3) with $N \to \infty$ and $\|\mathcal{V}\|_\infty < \infty$, then we also get that (assuming $\beta'_\infty < \infty$ or that the effect of initial conditions dies out with time) $\sum_{j=0}^{\infty} |\hat{e}_p(j)|^2 < \infty$, which implies that error convergence is guaranteed, i.e., $\lim_{j\to\infty} \hat{e}_p(j) = 0$.

AN EXAMPLE: THE CASE OF SYSTEMS WITH SHIFT STRUCTURE

In order to illustrate the above discussion, we now consider in details an important situation that often arises in system identification. We shall employ here the shift operator notation $q^{-1}[u(k)] = u(k-1)$. Thus, applying an operator $W(q^{-1}) = \sum_{k=1}^{M} w_k\, q^{-k}$ to a sequence $d(j)$ means

$$W(q^{-1})[d(j)] \quad = \quad \sum_{k=1}^{M} w_k\, d(j-k)\,.$$

The case we consider assumes that the nonlinear vectors $\mathbf{h}_j(\mathbf{w})$ and $\mathbf{g}_j(\mathbf{w})$ are equal and exhibit *shift* structure (a motivation for this situation is provided in the next section). By this we mean that their entries are of the form $\mathbf{h}_j(\mathbf{w}) = \mathbf{g}_j(\mathbf{w}) \triangleq$

$$\begin{bmatrix} d(j-1) & d(j-2) & \ldots & d(j-M) \end{bmatrix}, \qquad (5)$$

where M denotes the size of $\mathbf{h}_j(\mathbf{w})$, and $d(\cdot)$ is a scalar entry that is still assumed to depend nonlinearly on $\mathbf{w}$. Setting $\mathbf{g}_j(\mathbf{w})$ equal to $\mathbf{h}_j(\mathbf{w})$ implies that $z(j) = \mathbf{h}_j(\mathbf{w})\mathbf{w}$ and, consequently, $z(j)$ is equal to the uncorrupted component in the noisy measurement $y(j) = \mathbf{h}_j(\mathbf{w})\mathbf{w} + v(j)$.

It is further assumed that each entry $d(j)$ is recursively generated via

$$d(j) = S(q^{-1})[\mathbf{h}_j(\mathbf{w})\mathbf{w}], \qquad (6)$$

where $S(q^{-1})$ denotes an autoregressive filter. This means that $d(j)$ is obtained by filtering $\mathbf{h}_j(\mathbf{w})\mathbf{w}$ through an autoregressive filter. The special case $S(q^{-1}) = 1$ often arises in autoregressive modeling, and is discussed in the next section. We also assume that a similar shift structure is incorporated into the construction of the estimate $\hat{\mathbf{h}}_j$, say

$$\hat{\mathbf{h}}_j = \begin{bmatrix} \hat{d}(j-1) & \hat{d}(j-2) & \ldots & \hat{d}(j-M) \end{bmatrix}, \quad (7)$$

where the choice of the $\hat{d}(\cdot)$ is as explained below.

The difference $[d(j) - \hat{d}(j)]$ is denoted by $\tilde{d}(j)$, and we further associate with $\mathbf{w}$ the polynomial $W(q^{-1})$, where the $\{w_i\}$ are the entries of $\mathbf{w}$. It is then easy to see that the expression (2) for $\hat{v}(j)$ becomes

$$\begin{aligned} \hat{v}(j) &= \left[\mathbf{h}_j(\mathbf{w}) - \hat{\mathbf{h}}_j\right]\mathbf{w} + v(j), \\ &= W(q^{-1})[\tilde{d}(j)] \; + \; v(j). \end{aligned} \qquad (8)$$

Our purpose is to relate $\tilde{d}(j)$ to $\hat{e}_p(j)$, which will then lead to a desired relation between $\hat{v}(j)$ and $\{\hat{e}_p(j), v(j)\}$. For this purpose, we shall now exhibit a possibility for choosing the $\hat{d}(\cdot)$ and, consequently, for completely defining the estimate $\hat{\mathbf{h}}_j$, in order to achieve such relation. The $\hat{d}(j)$ will be computed as follows:

$$\hat{d}(j) = S(q^{-1})[\hat{\mathbf{h}}_j \mathbf{w}_j]. \qquad (9)$$

Note that this estimate is not of the bootstrap type that we referred to earlier since the computation of $\hat{d}(j)$ is highly dependent not only on the $\mathbf{w}_{j-1}$ but also on all previous estimates of $\mathbf{w}$.

The point now is that construction (9) allows us to relate $\tilde{d}(\cdot)$ to $\hat{e}_p(\cdot)$ via a filtering operation as follows:

$$\begin{aligned} \tilde{d}(j) &= d(j) - \hat{d}(j), \\ &= S(q^{-1})\left[\mathbf{h}_j(\mathbf{w})\mathbf{w} - \hat{\mathbf{h}}_j\mathbf{w}_j\right], \\ &= S(q^{-1})\left[\left\{\mathbf{h}_j(\mathbf{w}) - \hat{\mathbf{h}}_j\right\}\mathbf{w} \; + \; \hat{e}_p(j)\right], \\ &= S(q^{-1})W(q^{-1})[\tilde{d}(j)] \; + \; S(q^{-1})\left[\hat{e}_p(j)\right], \\ &= \frac{S(q^{-1})}{1 - S(q^{-1})W(q^{-1})}[\hat{e}_p(j)]. \end{aligned} \qquad (10)$$

Combining with (8) we see that

$$\hat{v}(j) = v(j) \; + \; \frac{S(q^{-1})W(q^{-1})}{1 - S(q^{-1})W(q^{-1})}[\hat{e}_p(j)].$$

In terms of the structure of Figure 1 we have $\mathcal{V}_N = I$ (the identity operator) and $\mathcal{F}_N$ equal to the $(N+1) \times (N+1)$ leading triangular operator that describes the action of $SW/(1 - SW)$ over the first $(N+1)$ samples of $\{\hat{e}_p(\cdot)\}$ (in the absence of initial conditions).

A sufficient condition for (3) to hold is to require $\|\mathcal{F}_N\|_\infty < 1/\xi$. This is satisfied if $SW/1 - SW$ is stable and

$$\max_{\omega} \left| \frac{\xi S(e^{j\omega})W(e^{j\omega})}{1 - S(e^{j\omega})W(e^{j\omega})} \right| < 1, \;\; 0 \le \omega < 2\pi. \quad (11)$$

The stability of $SW/(1 - SW)$ also implies that the effect of initial conditions will die out as time progresses. If the noise sequence $\{v(\cdot)\}$ further has finite energy, then we also conclude that $\lim_{j\to\infty} \hat{e}_p(j) = 0$. It also follows from the stability of $S/(1 - SW)$, from the convergence of $\hat{e}_p(j)$, and from (10) that $\lim_{j\to\infty} \tilde{d}(j) = 0$. We summarize the example of this section in the form of an algorithm, which can be seen as special case of Algorithm 1 (it is also written, for convenience, in terms of $\mathbf{w}_j$ rather than $\hat{\mathbf{x}}_{j|j}$).
<u>ALGORITHM 2.</u> [SYSTEMS WITH SHIFT STRUCTURE] *Consider* $\{y(j) = \mathbf{h}_j(\mathbf{w})\mathbf{w} + v(j)\}$, *where* $\mathbf{h}_j(\mathbf{w})$ *is assumed to exhibit a shift structure as in (5) with its en-*

tries $\{d(\cdot)\}$ computed via (6), for a given filter $S(q^{-1})$. Assume further that $\hat{\mathbf{h}}_j$ is also constructed so as to exhibit a shift structure as in (7) and that its entries are computed via (9).

If the quantity $\mathbf{P}_{j+1}$ computed below is positive-definite for $0 \leq j \leq N$, and if (11) holds, then the following filter, with initial guess $\mathbf{w}_{-1}$ and $\mathbf{P}_0 = \Pi_0$,

$$\mathbf{w}_j = \mathbf{w}_{j-1} + \mathbf{P}_j \hat{\mathbf{h}}_j^* \left[1 + \hat{\mathbf{h}}_j \mathbf{P}_j \hat{\mathbf{h}}_j^*\right]^{-1} \left[y(j) - \hat{\mathbf{h}}_j \mathbf{w}_{j-1}\right],$$

$$\mathbf{P}_{j+1} = \mathbf{P}_j - \mathbf{P}_j \left[\begin{array}{cc} \hat{\mathbf{h}}_j^* & \hat{\mathbf{h}}_j^* \end{array}\right] \mathbf{R}_{e,j}^{-1} \left[\begin{array}{c} \hat{\mathbf{h}}_j \\ \hat{\mathbf{h}}_j \end{array}\right] \mathbf{P}_j,$$

$$\mathbf{R}_{e,j} = \left\{ \left[\begin{array}{cc} -\xi^2 & \mathbf{0} \\ \mathbf{0} & 1 \end{array}\right] + \left[\begin{array}{c} \hat{\mathbf{h}}_j \\ \hat{\mathbf{h}}_j \end{array}\right] \mathbf{P}_j \left[\begin{array}{cc} \hat{\mathbf{h}}_j^* & \hat{\mathbf{h}}_j^* \end{array}\right] \right\},$$

guarantees (4) with $\|\mathcal{V}_N\|_\infty = 1$. If $\{v(\cdot)\}$ has finite energy, then $\lim_{j\to\infty} \hat{e}_p(j) = 0$.

The above algorithm can in fact be related to so-called pseudo-linear regression (PLR) algorithms in IIR modeling (e.g., (Landau, 1979)[p. 167]). To clarify this, we first note, via the so-called matrix inversion formula, that the Riccati recursion in Algorithm 2 is equivalent to

$$\mathbf{P}_{j+1}^{-1} = \mathbf{P}_j^{-1} + (1 - \xi^{-2}) \hat{\mathbf{h}}_j^* \hat{\mathbf{h}}_j.$$

This means that the positivity condition on the $\{\mathbf{P}_j\}$ is always satisfied for any choice $\xi \geq 1$. Moreover, if we again invert the above expression we obtain that the Riccati recursion can be rewritten in the equivalent (and more recognizable) form

$$\mathbf{P}_{j+1} = \mathbf{P}_j - \frac{\mathbf{P}_j \hat{\mathbf{h}}_j^* \hat{\mathbf{h}}_j \mathbf{P}_j}{(1 - \xi^{-2})^{-1} + \hat{\mathbf{h}}_j \mathbf{P}_j \hat{\mathbf{h}}_j^*}.$$

INSTANTANEOUS-GRADIENT-BASED IIR ADAPTIVE FILTERS

The Riccati recursion of Algorithm 2 can also be shown to trivialize in an important special case. This fact was noted in (Hassibi, Sayed, and Kailath, 1993*a*) in the linear context of FIR (or MA) identification and will now be extended to the nonlinear scenario of the previous section.

If ξ is chosen to be one, $\xi = 1$, then the recursion trivializes to $\mathbf{P}_{j+1}^{-1} = \mathbf{P}_j^{-1} = \Pi_0^{-1}$, where Π_0 is the initial condition. The solvability condition then becomes $\Pi_0 > 0$. In particular, this is always satisfied if we choose $\Pi_0 = \alpha \mathbf{I}$, a (positive) constant multiple of the identity. Under these conditions, the update of the weight estimate in Algorithm 2 reduces to

$$\mathbf{w}_j = \mathbf{w}_{j-1} + \frac{\alpha}{1 + \alpha \|\hat{\mathbf{h}}_j\|_2^2} \hat{\mathbf{h}}_j^* \left[y(j) - \hat{\mathbf{h}}_j \mathbf{w}_{j-1}\right], \quad (12)$$

which is an instantaneous-gradient-based recursion with a step-size of the form $\alpha/(1 + \alpha\|\hat{\mathbf{h}}_j\|_2^2)$.

Stability Analysis and Loop Transformations

The stability condition of Algorithm 2 requires $SW/(1-SW)$ to be strictly contractive, which in turn requires the positive-real part of $(1-SW)$ to be larger than $1/2$. This range can be extended by applying a useful scaling tool that is related to so-called loop transformations in passivity analysis. Indeed, recursion (12) can be rewritten in the equivalent form

$$\begin{aligned} \mathbf{w}_j &= \mathbf{w}_{j-1} + \alpha \hat{\mathbf{h}}_j^* \left[\frac{1}{1 - SW}[\hat{e}_p(j)] + v(j)\right] & (13) \\ &= \mathbf{w}_{j-1} + \alpha \hat{\mathbf{h}}_j^* \left[\hat{e}_p(j) + \hat{v}(j)\right], & (14) \end{aligned}$$

where we have defined

$$\hat{v}(j) \triangleq -\hat{e}_p(j) + \frac{1}{1 - SW}[\hat{e}_p(j)] + v(j). \quad (15)$$

The map from $\{\alpha^{-1/2}\tilde{\mathbf{w}}_{-1}, \hat{v}(\cdot)\}$ to $\{\hat{e}_p(\cdot)\}$ is a strict contraction since, as argued above, the recursion (14) is an aposteriori $H^\infty-$filter and the positivity condition is always satisfied due to $\mathbf{P}_j = \Pi_0 = \alpha \mathbf{I} > 0$. This result holds for any update filter of the form (14) and for any noise sequence $\hat{v}(\cdot)$. The special construction (15) was further exploited while studying the stability of the feedback interconnection. These observations motivate us to rewrite the update recursion (13) in the equivalent form

$$\begin{aligned} \mathbf{w}_j &= \mathbf{w}_{j-1} + \beta \hat{\mathbf{h}}_j^* \left[\frac{\alpha/\beta}{1 - SW}[\hat{e}_p(j)] + \frac{\alpha}{\beta} v(j)\right], \\ &= \mathbf{w}_{j-1} + \beta \hat{\mathbf{h}}_j^* \left[\hat{e}_p(j) + \hat{v}'(j)\right], & (16) \end{aligned}$$

where we have now defined

$$\hat{v}'(j) \triangleq -\hat{e}_p(j) + \frac{\alpha/\beta}{1 - SW}[\hat{e}_p(j)] + \frac{\alpha}{\beta} v(j), \quad (17)$$

and β is *any* positive real number. The recursion (16) guarantees a strict contraction map from $\{\beta^{-1/2}\tilde{\mathbf{w}}_{-1}, \hat{v}'(\cdot)\}$ to $\{\hat{e}_p(\cdot)\}$. Accordingly, an overall l_2-stable system from $\{\beta^{-1/2}\tilde{\mathbf{w}}_{-1}, \alpha/\beta v(\cdot)\}$ to $\{\hat{e}_p(\cdot)\}$ will be guaranteed if we impose

$$\max_\omega \left| \frac{\alpha/\beta}{1 - S(e^{j\omega})W(e^{j\omega})} - 1 \right| < 1, \quad (18)$$

which requires Real $[1 - S(e^{j\omega})W(e^{j\omega})] > \alpha/2\beta$. Since this should be true for any choice of β, we therefore conclude, by choosing β large enough, that a sufficient condition for the l_2-stability of (12) is the strict positive-realness of the function $1/(1-SW)$. This condition guarantees that the update solution (12) will always result in a convergent sequence $\{\hat{e}_p(\cdot)\}$, under a finite-energy assumption on $\{v(\cdot)\}$.

In the next section we consider two important special cases that arise in IIR modeling.

Landau's Scheme for IIR Modeling

Consider a linear time-invariant system that is described by a recursive (i.e., pole-zero or IIR) difference equation of the form

$$d(j) = \sum_{k=1}^{M_a} a_k\, d(j-k) + \sum_{k=0}^{M_b - 1} b_k\, u(j-k),$$

$$= \begin{bmatrix} \mathbf{d}_{j-1} & \mathbf{u}_j \end{bmatrix} \begin{bmatrix} \mathbf{a} \\ \mathbf{b} \end{bmatrix} \equiv \mathbf{h}_j(\mathbf{w})\mathbf{w}, \quad (19)$$

where $\mathbf{d}_{j-1} = \begin{bmatrix} d(j-1) & \ldots & d(j-M_a) \end{bmatrix}$ and $\mathbf{u}_j = \begin{bmatrix} u(j) & \ldots & u(j-M_b+1) \end{bmatrix}$, while $\mathbf{a} = \text{col}\{a_1, \ldots, a_{M_a}\}$, and $\mathbf{b} = \text{col}\{b_0, \ldots, b_{M_b-1}\}$. The row (data) vector $\mathbf{h}_j(\mathbf{w}) = \begin{bmatrix} \mathbf{d}_{j-1} & \mathbf{u}_j \end{bmatrix}$ is dependent on $\mathbf{w}$ since the entries of $\mathbf{d}_{j-1}$ depend on $\mathbf{w}$. Here $\mathbf{w}$ is a column vector that contains the parameters $\mathbf{a}$ and $\mathbf{b}$.

The problem of interest is the following: given noisy measurements $\{y(\cdot)\}$ of the output of the system, $\{d(\cdot)\}$, in response to a known input sequence $\{u(\cdot)\}$, say $y(j) = d(j) + v(j) = \mathbf{h}_j(\mathbf{w})\mathbf{w} + v(j)$, estimate the system parameters $\mathbf{a}$ and $\mathbf{b}$ (or $\mathbf{w}$).

An existing approximate solution, which is based on instantaneous-gradient ideas (Landau, 1979), is one that updates the weight estimate according to expression (12) and where $\hat{\mathbf{h}}_j$ is computed as $\hat{\mathbf{h}}_j = \begin{bmatrix} \hat{\mathbf{d}}_{j-1} & \mathbf{u}_j \end{bmatrix}$. Here, the $\mathbf{u}_j$ is known, while the entries of $\hat{\mathbf{d}}_{j-1} = \begin{bmatrix} \hat{d}(j-1) & \ldots & \hat{d}(j-M_a) \end{bmatrix}$, are estimated recursively: start with initial guesses $\{\hat{d}(-1), \hat{d}(-2), \ldots, \hat{d}(-M_a)\}$ and compute successive estimates $\hat{d}(j)$, for $j \geq 0$, via the recursion:

$$\hat{d}(j) = \hat{\mathbf{d}}_{j-1}\mathbf{a}_j + \mathbf{u}_j\mathbf{b}_j, \quad (20)$$

where $\{\mathbf{a}_j, \mathbf{b}_j\}$ denote estimates of $\{\mathbf{a}, \mathbf{b}\}$ at the j^{th} iteration. This is a special case of the construction (9) (with $S = 1$). We also see here that we only need to estimate the leading part of $\mathbf{h}_j$ (the part corresponding to $\mathbf{d}_{j-1}$) since the $\mathbf{u}_j$ part is given. Nevertheless, the same framework discussed so-far in the paper applies. All we have to do is employ the results of Algorithm 2 with $W(q^{-1})$ replaced by $A(q^{-1})$. This is because the difference $(\mathbf{h}_j - \hat{\mathbf{h}}_j)$ now has the form $\begin{bmatrix} (\mathbf{d}_{j-1} - \hat{\mathbf{d}}_{j-1}) & \mathbf{0} \end{bmatrix}$. That is, its second block entry is zero and, consequently, $(\mathbf{h}_j - \hat{\mathbf{h}}_j)\mathbf{w} = (\mathbf{d}_{j-1} - \hat{\mathbf{d}}_{j-1})\mathbf{a}$. We then conclude that a sufficient condition for l_2-stability is to require the strict positive-realness of $1/(1-A)$.

While this is a known result for Landau's scheme (e.g., (Solo and Kong, 1995; pp.146–150)), we have rederived it here within the general framework of this paper. In particular, we have established that Landau's scheme is in fact a special case of the *aposteriori* H^∞-filter of Algorithm 2, and that the corresponding solvability condition has been trivialized by choosing $\Pi_0 = \alpha\mathbf{I}$.

Feintuch's Scheme for IIR Modeling

A related discussion is provided in (Rupp and Sayed, 1995), where a variant of recursion (12) is used, viz., one that employs a constant step size, as suggested by (Feintuch, 1976),

$$\mathbf{w}_j = \mathbf{w}_{j-1} + \mu\hat{\mathbf{h}}_j^* \left[y(j) - \hat{\mathbf{h}}_j\mathbf{w}_{j-1}\right], \quad (21)$$

where $\{\mathbf{a}_j, \mathbf{b}_j\}$ in (20) are further replaced by $\{\mathbf{a}_{j-1}, \mathbf{b}_{j-1}\}$. This is in fact a special case of the *apriori* filter of Algorithm 1A, in exactly the same way as (12) is a special case of the *aposteriori* filter of Algorithm 1. By setting $\hat{\mathbf{g}}_j = \hat{\mathbf{h}}_j$, $\xi = 1$, and $\Pi_0 = \mu\mathbf{I}$ in Algorithm 1A we obtain (21). Now, however, the solvability condition requires $(\mu^{-1}\mathbf{I} - \hat{\mathbf{h}}_j^*\hat{\mathbf{h}}_j) > 0$ or, equivalently, $\mu\|\hat{\mathbf{h}}_j\|^2 < 1$. Under this additional assumption, a sufficient condition for l_2-stability is to require the contractiveness of $A/(1-A)$ – more details along the lines of this paper can be found in (Sayed and Rupp, 1994; Sayed and Kailath, 1994; Rupp and Sayed, 1995).

1. REFERENCES

Doyle, J. C., K. Glover, P. Khargonekar and B. Francis (1989). State-space solutions to standard H_2 and H_∞ control problems. *IEEE Transactions on Automatic Control* **34**(8), 831–847.

Feintuch, P. L. (1976). An adaptive recursive LMS filter. *Proc. IEEE* **64**(11), 1622–1624.

Hassibi, B., A. H. Sayed and T. Kailath (1993*a*). LMS is H^∞ optimal. In: *Proc. CDC*. San Antonio, Texas, vol.1 , pp. 74–79.

Hassibi, B., A. H. Sayed and T. Kailath (1993*b*). Recursive linear estimation in Krein spaces - Part II: Applications. In: *Proc. CDC*. San Antonio, Texas, vol. 4, pp. 3495–3500.

Khalil, H. K. (1992). *Nonlinear Systems*. MacMillan.

Khargonekar, P.P. and K. M. Nagpal (1991). Filtering and smoothing in an $H^\infty-$ setting. *IEEE Trans. on Automatic Control* **AC-36**, 151–166.

Landau, Y. D. (1979). *Adaptive Control: The Model Reference Approach*. Control and Systems Theory. Marcel Dekker. NY.

Green, M. and D. J. N. Limebeer (1995).*Linear Robust Control*. Prentice Hall. NJ.

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

Rupp, M. and A. H. Sayed (1995). On the stability and convergence of Feintuch's algorithm for adaptive IIR filtering. In *Proc. IEEE ICASSP*. Detroit, MI.

Sayed, A. H. and T. Kailath (1994). A state-space approach to adaptive RLS filtering. *IEEE Signal Processing Magazine* **11**(3), pp. 18–60.

Sayed, A. H. and M. Rupp (1994). On the robustness, convergence, and minimax performance of instantaneous-gradient adaptive filters. In: *Proc. Asilomar Conference on Signals, Systems, and Computers*, pp. 592–596.

Solo, V. and X. Kong (1995). *Adaptive Signal Processing Algorithms: Stability and Performance*, Prentice Hall, New Jersey.

Vidyasagar, M. (1993). *Nonlinear Systems Analysis*. Second ed. Prentice Hall. New Jersey.

Yaesh, I. and U. Shaked (1991). $H^\infty-$optimal estimation: The discrete time case. In: *Proc. Inter. Symp. on MTNS*. Kobe, Japan. pp. 261–267.