An l_2–Stable Feedback Structure for Nonlinear Adaptive Filtering and Identification

Ali H. Sayed and Markus Rupp

Abstract—**This paper proposes a feedback structure for the design of l_2-stable algorithms for nonlinear adaptive filtering and identification, and establishes explicit connections between classical schemes in IIR modeling and more recent results in H^∞-theory. In particular, two algorithms due to Feintuch (1976) and Landau (1979), as well as the so-called pseudo-linear regression and Gauss-Newton algorithms, are discussed within the framework proposed herein. Additional examples and simulation results are included to illustrate the applicability of the approach to several nonlinear scenarios.**

Keywords— **Small-gain theorem, mean-value theorem, l_2-stability, positive-real function, averaging analysis, loop transformation, feedback structure, H^∞-filters, bootstrap method, IIR modeling, stochastic gradient filters, pseudo linear regression filter, Gauss-Newton filters.**

I. 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, an account of which is given by [14], the ODE approach of Ljung [17], [18], and the related class of averaging methods for trajectory approximation, as described in the books by [1] and [24].

Correspondingly, in the last decade, there has been an explosion of research in the areas of robust (H^∞) filtering and control, as indicated by several of the references at the end of this paper – e.g., [5], [13], [2], [26], [16], [9]. A major concern in the H^∞ setup has been 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 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.

Motivated by these results, we take here an alternative look at the analysis and design of adaptive and identification schemes. One of the objectives of this work is to show how to reconcile, within a nonlinear estimation context, earlier developments in stable adaptive schemes with more recent developments in H^∞ design. For this purpose, the discussion in this paper exploits a useful tool in system theory that is widely known as the small gain theorem. While this theorem can be reformulated in terms of hyperstability or passivity results [14], [1], [16], the analysis provided herein has several distinctive features that will become clear as the discussion progresses.

At this stage, however, we only wish to highlight the fact that by relying on the small gain theorem, we can now advantageously exploit the wealth of results that are already available in the H^∞−setting. This is especially helpful in the design (i.e., synthesis) phase. In particular, it will allow us to propose an adaptive structure that will be shown to include, as special cases, several algorithms that have been derived earlier in the literature (even prior to the emergence of the H^∞ point of view). The feedback formulation will also enable us to establish that these earlier schemes are special instances of the more recent class of H^∞−filters!

Moreover, although the feedback nature of adaptive schemes has been advantageously exploited in earlier places in the literature – e.g., [17], [14], 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 (as explained in [15]), but is instead motivated by our concern with the overall *robustness* performance of the algorithm. For this reason, the feedback configuration is defined here in such a way so as to *explicitly* consider the effect of *both* the measurement noise and the uncertainty in the initial weight-vector guess on the overall algorithm performance.

Notation. We use small boldface letters to denote vectors (e.g., $\mathbf{h}$) and capital boldface letters to denote matrices (e.g., $\mathbf{P}$). The symbol "$*$" denotes Hermitian conjugation (complex conjugation for scalars), and the notation $\|\mathbf{x}\|_2^2$ denotes the squared Euclidean norm of a vector. Also, $\mathbf{A}^{1/2}$ denotes a square-root factor of a matrix $\mathbf{A}$, viz., any matrix satisfying $\mathbf{A}^{1/2}\mathbf{A}^{*/2} = \mathbf{A}$.

We also use subscripts for time-indexing of vector quantities (e.g., $\mathbf{h}_i$) and parenthesis for time-indexing of scalar quantities (e.g., $d(i)$). We further employ the shift operator notation $q^{-1}u(k) = u(k-1)$. Hence, applying an operator $W(q^{-1}) = \sum_{k=1}^{M} w_k\, q^{-k}$ to a sequence $d(i)$ means $W(q^{-1})d(i) = \sum_{k=1}^{M} w_k\, d(i-k)$.

This material was based on work supported by the National Science Foundation under Award No. MIP-9409319. The work of M. Rupp was also supported in part by a scholarship from DAAD (German Academic Exchange Service) as well as the scientific division of NATO (Wissenschaftsausschuß der NATO). A preliminary version of this work was presented at the IFAC Symposium on Nonlinear Control System Design, Lake Tahoe, June 1995.

Department of Electrical Engineering, University of California, Los Angeles, CA 90095. E-mail: sayed@ee.ucla.edu.

This work was performed while M. Rupp was a postdoctoral fellow in the Department of Electrical and Computer Engineering, University of California, Santa Barbara, CA 93106. He is currently with the Wireless Technology Research Dept., Lucent Technologies, 791 Holmdel-Keyport Rd., Holmdel NJ 07733–0400.

II. An Example and Motivation

Consider a linear time-invariant autoregressive model that is described by the difference equation

$$d(i) = \sum_{k=1}^{M} a_k\, d(i-k) + u(i). \tag{2.1}$$

Here, M is the order of the filter (assumed known), $u(i)$ is the value of the input sequence at time i (also known), and the $\{a_k\}$ are unknown filter coefficients that we wish to estimate from noisy measurements of the output signal $d(i)$, say from $m(i) = d(i) + v(i)$ for $0 \le i \le N$. Here $m(i)$ and $v(i)$ denote the noisy measurement and the additive noise at time i, respectively.

If we collect the past M values of $d(i)$ into a row vector $\mathbf{d}_{i-1}$,

$$\mathbf{d}_{i-1} = \left[\begin{array}{ccc} d(i-1) & \ldots & d(i-M) \end{array}\right], \tag{2.2}$$

and the M unknown coefficients a_k into a column vector $\mathbf{w}$,

$$\mathbf{w} = \mathrm{col}\{a_1, a_2, \ldots, a_M\}, \tag{2.3}$$

then (2.1) can be rewritten in a compact vector form as $d(i) = \mathbf{d}_{i-1}\mathbf{w} + u(i)$. Consequently, the noisy measurements $m(i)$ satisfy

$$m(i) = \mathbf{d}_{i-1}\mathbf{w} + u(i) + v(i). \tag{2.4}$$

In expression (2.4), the quantities $m(i)$ and $u(i)$ are known and we can therefore introduce the known quantity $y(i) = m(i) - u(i)$ and write instead the equivalent expression:

$$y(i) = \mathbf{d}_{i-1}\mathbf{w} + v(i). \tag{2.5}$$

The problem can then be interpreted as follows: we are given noisy quantities (or measurements) $y(i)$ that are related to the unknown vector $\mathbf{w}$ via the term $\mathbf{d}_{i-1}\mathbf{w}$ as in (2.5). This term is not only time-variant, but also *nonlinear* in $\mathbf{w}$ because each entry of $\mathbf{d}_{i-1}$ is itself a function of $\mathbf{w}$, as is evident from (2.1) and (2.2).

We can indicate more explicitly the nonlinear dependency of the measurements $\{y(i)\}$ on the unknown vector $\mathbf{w}$ by rewriting equation (2.5) in the more generic form

$$y(i) = \mathbf{h}_i(\mathbf{w})\mathbf{w} + v(i), \tag{2.6}$$

where $\mathbf{h}_i(\mathbf{w})$ denotes a time-variant nonlinear (row vector) function of $\mathbf{w}$. In the special case of the above autoregressive example we have $\mathbf{h}_i(\mathbf{w}) = \mathbf{d}_{i-1}$. More generally, however, we may have situations with alternative forms for the nonlinear term $\mathbf{h}_i(\mathbf{w})$. Examples of such cases are provided later in Sec. 5.

Now given the noisy measurements $y(i)$ of (2.6), we may distinguish between two problems:

(i) The first problem is to use the given measurements $y(i)$ in order to estimate the unknown vector of parameters $\mathbf{w}$. This formulation has been extensively studied in the literature and several algorithms have been proposed (see, e.g., [14], [17], [18]). We shall return to these classical solutions in later sections of this work (Secs. 7 – 9).

(ii) The second problem is to use the given measurements $y(i)$ in order to estimate the uncorrupted terms $\mathbf{h}_i(\mathbf{w})\mathbf{w}$ in (2.6). While a distinction between cases (i) and (ii) may not seem necessary in least-squares formulations, it is nevertheless crucial in H^∞-based designs. It turns out that if one poses an H^∞ problem for estimating $\mathbf{w}$ and an H^∞ problem for estimating $\mathbf{h}_i(\mathbf{w})\mathbf{w}$, the solutions will in general be distinct. More interesting perhaps, it is the H^∞ estimation of $\mathbf{h}_i(\mathbf{w})\mathbf{w}$, rather than $\mathbf{w}$, that will allow us to establish connections with the classical schemes mentioned in (i).

To accommodate for case (ii) above, we shall define an auxiliary variable

$$z(i) = \mathbf{h}_i(\mathbf{w})\mathbf{w} \;\;, \tag{2.7}$$

and pose the problem of estimating $z(i)$ from the $y(i)$ (according to a certain H^∞ criterion to be defined further ahead). In fact, we can allow for more general cases and define auxiliary variables of the form

$$z(i) = \mathbf{g}_i(\mathbf{w})\mathbf{w}, \tag{2.8}$$

for some given (row vector) function $\mathbf{g}_i(\mathbf{w})$ that may also be time-variant and a function of the unknown $\mathbf{w}$. It may also be a constant vector that is independent of both time and $\mathbf{w}$, say $z(i) = \mathbf{g}\mathbf{w} = z$. This level of generality allows us to handle other situations of interest. For example, assume that we are only interested in estimating the third filter coefficient of the autoregressive model, then we would choose $\mathbf{g} = [0\;0\;1\;0\;0\ldots0]$ in $z = \mathbf{g}\mathbf{w}$.

We are now in a position to state the nonlinear estimation setting of this paper. For convenience of presentation, the problem will be formulated in a state-space context.

III. The Nonlinear H^∞ Problem

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

$$y(i) = \mathbf{h}_i(\mathbf{w})\mathbf{w} + v(i). \tag{3.1}$$

Here $v(i)$ stands for the noise component (or modeling uncertainties) at the discrete time instant i, and $\mathbf{h}_i(\mathbf{w})$ denotes a known time-variant row vector whose entries are themselves functions of the unknown entries of $\mathbf{w}$.

A special example of an autoregressive model that leads to an equation of the form (3.1) was discussed in the previous section. However, other nonlinear problems also lead to measurement structures that are similar to (3.1) and, hence, the discussion in this section applies to these problems as well. Examples to this effect will be postponed to Secs. 5, 6.2, and 6.3 (see though expressions (5.8), (5.12), and (5.17)).

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

$$\begin{aligned} \mathbf{x}_{i+1} &= \mathbf{x}_i, \quad \mathbf{x}_0 = \mathbf{w}, & (3.2)\\ y(i) &= \mathbf{h}_i(\mathbf{x}_i)\mathbf{x}_i + v(i). & (3.3) \end{aligned}$$

The state equation in (3.2) is trivial: the state vector does not change with time and it therefore remains equal to the initial state vector, which is taken to be $\mathbf{w}$, i.e.,

$$\mathbf{x}_{i+1} = \mathbf{x}_i = \ldots = \mathbf{x}_0 = \mathbf{w}.$$

[We may add though that the analysis of this paper extends to more general state-space models, e.g., with driving inputs in the state equation and nonunity transition matrices].

Let $z(i)$ denote a desired combination of the unknown vector $\mathbf{w}$, say

$$z(i) = \mathbf{g}_i(\mathbf{w})\mathbf{w} = \mathbf{g}_i(\mathbf{x}_i)\mathbf{x}_i \ , \tag{3.4}$$

for some known function form $\mathbf{g}_i(\cdot)$. The objective is to employ the available measurements $y(i)$ in order to estimate $z(i)$. The estimate of $z(i)$ is to be computed in a causal manner, i.e., it can only depend on the observations that are available up to time i, $\{y(j),\ 0 \le j \le i\}$. A motivation for this problem in the context of autoregressive modeling was provided in Sec. 2. Similar problem formulations also arise in other contexts and will be illustrated in Sec. 5.

So let $\hat{z}(i|i)$ denote an estimate for $z(i)$ that is dependent on the available observation data $\{y(\cdot)\}$ up to time i, and which is defined according to the following H^∞ criterion.

Let Π_0 be a given positive-definite matrix and choose any initial guess for $\mathbf{w}$, which we shall denote by $\bar{\mathbf{x}}_0$. Define the weighted initial weight error $\tilde{\mathbf{x}}_0 = \Pi_0^{-1/2}(\mathbf{x}_0 - \bar{\mathbf{x}}_0)$, as well as the estimation error

$$e_p(i) = z(i) - \hat{z}(i|i). \tag{3.5}$$

For every time instant i, define the ratio:

$$r(i) \;=\; \frac{\sum_{j=0}^{i} |e_p(j)|^2}{\|\tilde{\mathbf{x}}_0\|_2^2 + \sum_{j=0}^{i} |v(j)|^2}. \tag{3.6}$$

This ratio provides a relative measure of the energies due to the estimation errors, the error in the initial guess $\bar{\mathbf{x}}_0$, and the disturbances $v(\cdot)$.

PROBLEM 1. [NONLINEAR H^∞ ESTIMATION] *Given (3.2)–(3.4), determine, if possible, causal estimates $\hat{z}(j|j)$, for $j = 0, 1, \ldots, N$, so as to guarantee that, for any $\tilde{\mathbf{x}}_0$ and $v(\cdot)$, the ratios $r(i)$ will be bounded by a given positive constant γ^2, say*

$$r(i) < \gamma^2 \quad \textit{for} \quad 0 \le i \le N. \tag{3.7}$$

Assume we collect the estimation errors $e_p(i)$ into a column vector, say

$$\mathbf{e}_p = \mathrm{col}\{e_p(0), e_p(1), \ldots, e_p(N)\}, \tag{3.8}$$

and the noise sequence and the initial weight-error into another column vector, say

$$\mathbf{n} = \mathrm{col}\{\tilde{\mathbf{x}}_0, \mathbf{v}\} = \text{disturbance vector}, \tag{3.9}$$

where $\mathbf{v}$ contains the additive noise sequence

$$\mathbf{v} = \mathrm{col}\{v(0), v(1), \ldots, v(N)\}. \tag{3.10}$$

That is, $\mathbf{n}$ contains the disturbance signals (these are signals that we have no control over) while $\mathbf{e}_p$ contains the resulting estimation errors (these are the errors that result from the solution). Now if a solution to the estimation problem exists, it should induce a mapping, say $\mathcal{T}_N$, from the quantities in (3.9) to the quantities in (3.8) satisfying $r(N) < \gamma^2$. In this case, we say that the level of robustness is γ.

As mentioned earlier, while the state equation (3.2) is trivial, the nonlinear H^∞ problem can be stated in full generality (i.e., for general state-space equations). In this paper, however, we focus on the special state equation form (3.2). Such forms arise also in adaptive least-squares problems [21] and, for this reason, we may refer to the special Problem 1 as a nonlinear $\mathrm{H}^\infty-$*adaptive* (or identification) problem.

IV. AN APPROXIMATE H^∞-LINEAR SOLUTION

The presence of the $\mathbf{w}-$dependent (nonlinear) functions $\mathbf{h}_i(\mathbf{w})$ and $\mathbf{g}_i(\mathbf{w})$, in both the numerator and the denominator of the cost ratio $r(i)$ in (3.6), complicates the solution of Problem 1. For this reason, we proceed here in two steps.
(i) We first assume that the nonlinear terms $\mathbf{h}_i(\mathbf{w})$ and $\mathbf{g}_i(\mathbf{w})$ (which are dependent on $\mathbf{w}$) are replaced by estimates $\hat{\mathbf{h}}_i$ and $\hat{\mathbf{g}}_i$ that do not depend on $\mathbf{w}$ but only *causally* on the given measurements. This allows us to *approximate* the nonlinear problem by a standard linear $\mathrm{H}^\infty-$setting, and therefore proceed with a linear filter design. Linearizations of this kind are common in the literature and have often been invoked in many different contexts in order to handle nonlinear situations. However, a linearized design need not (and it often does not) guarantee that performance specifications for the original nonlinear problem will necessarily be met by the linearized solution. For this reason, our design procedure proposes a second step, the purpose of which is to clarify under what conditions, and subject to what modifications, the linearized design can still guarantee the desired performance for the nonlinear setting.
(ii) More specifically, by using the estimates $\hat{\mathbf{h}}_i$ and $\hat{\mathbf{g}}_i$ to design a linear H^∞ filter with a desired level of robustness, we end up with a *modified* mapping, say $\mathcal{T}_N'$, from a *modified* disturbance vector $\mathbf{n}'$ to a *modified* estimation error vector, $\mathbf{e}_p' = \mathrm{col}\{e_p'(i)\}$. But since our objective is to induce a robust mapping relative to the original disturbance vector $\mathbf{n}$, rather than a modified version of it, we proceed to embed the linear H^∞-design into a feedback structure. The purpose of the feedback configuration is to guarantee that the resulting induced mapping from the *original* disturbance vector $\mathbf{n}$ (rather than $\mathbf{n}'$) to the modified estimation errors $\mathbf{e}_p'$ will satisfy a desired level of robustness (see Theorem 1 further ahead). We shall also argue later that, in several cases, the use of the modified estimation errors $e_p'(\cdot)$, instead of the original $e_p(\cdot)$, does not affect the overall desired performance (see, e.g., Sections 5 and 6.1).

A. Vector Estimates

We first assume that we have available at each time instant i, 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 and should only depend on known quantities (or measurements) up to time $(i-1)$.

One possibility is the so-called bootstrap technique [18]; it assumes that we have access to recursive estimates of the parameter vector $\mathbf{w}$, which are then employed in approximating $\mathbf{h}_i(\mathbf{w})$ and $\mathbf{g}_i(\mathbf{w})$; if we let $\mathbf{w}_{i-1}$ denote the estimate of $\mathbf{w}$ at time $i-1$ that is based on the measuremets available up to that time instant, then the bootstrap method uses

$$\hat{\mathbf{h}}_i = \mathbf{h}_i(\mathbf{w}_{i-1}) \quad \text{and} \quad \hat{\mathbf{g}}_i = \mathbf{g}_i(\mathbf{w}_{i-1}). \tag{4.1}$$

That is, it evaluates the nonlinear functions $\mathbf{h}_i(\mathbf{w})$ and $\mathbf{g}_i(\mathbf{w})$ at the weight estimate $\mathbf{w}_{i-1}$. A construction along these lines for the class of so-called modifiable nonlinear functions is discussed in [10] in the context of continuous-time filtering. In general, however, the bootstrap construction (4.1) may not be sufficient to guarantee an overall l_2-stable filter. Examples are provided in Secs. 5.1 and 7.

We may remark here that the time variations in $\hat{\mathbf{h}}_i$ and $\hat{\mathbf{g}}_i$ will be due to both known time dependencies in the original functions $\mathbf{h}_i(\mathbf{w})$ and $\mathbf{g}_i(\mathbf{w})$, as well as nonlinear dependencies on the weight-vector estimates themselves.

B. Step 1: A Linear Design

Assume that, in some way, at each time instant i we have available estimates $\hat{\mathbf{h}}_i$ and $\hat{\mathbf{g}}_i$ that may have been computed either according to (4.1) or according to some other construction.

Using the $\{\hat{\mathbf{h}}_i, \hat{\mathbf{g}}_i\}$ we can rewrite the earlier nonlinear state-space model (3.2)–(3.3) in terms of the "linearized" version

$$\mathbf{x}_{i+1} = \mathbf{x}_i, \quad \mathbf{x}_0 = \mathbf{w}, \tag{4.2}$$
$$y(i) = \hat{\mathbf{h}}_i \mathbf{x}_i + v'(i), \tag{4.3}$$

where $\hat{\mathbf{h}}_i$ is now independent of the state vector $\mathbf{x}_i$, and $v'(i)$ denotes the difference between $y(i)$ and the approximate term $\hat{\mathbf{h}}_i\mathbf{x}_i$. The $v'(i)$ can be interpreted as a modified disturbace and it can also be related to the original disturbance $v(i)$ since

$$\begin{aligned} v'(i) &= y(i) - \hat{\mathbf{h}}_i\mathbf{x}_i \ , \\ &= \mathbf{h}_i(\mathbf{x}_i)\mathbf{x}_i + v(i) - \hat{\mathbf{h}}_i\mathbf{x}_i \ , \\ &= \left[\mathbf{h}_i(\mathbf{w}) - \hat{\mathbf{h}}_i\right]\mathbf{w} + v(i). \end{aligned} \tag{4.4}$$

That is, $v'(i)$ and $v(i)$ differ by $[\mathbf{h}_i(\mathbf{w}) - \hat{\mathbf{h}}_i]\mathbf{w}$, which essentially measures how far is the approximation $\hat{\mathbf{h}}_i$ from $\mathbf{h}_i$.

Let also $z'(i)$ denote the approximation of $z(i)$ in (3.4), viz.,

$$z'(i) = \hat{\mathbf{g}}_i\mathbf{w} = \hat{\mathbf{g}}_i\mathbf{x}_i. \tag{4.5}$$

Since $\hat{\mathbf{g}}_i$ is a known vector, $z'(i)$ simply corresponds to a linear combination of the entries of $\mathbf{w}$.

We can now pose the problem of *linearly* and *causally* estimating the $z'(i)$ from the available measurements $y(i)$ and using the *linear* state-space model (4.2)-(4.3).

So let $\hat{z}'(i|i)$ denote an estimate for $z'(i)$ that is dependent on the available observation data $\{y(\cdot)\}$ up to time i, and which is defined according to the following H^∞ criterion. Define

$$e'_p(i) = z'(i) - \hat{z}'(i|i) \ , \tag{4.6}$$

as well as the ratios

$$r'(i) = \frac{\sum_{j=0}^{i} |e'_p(j)|^2}{\|\bar{\mathbf{x}}_0\|_2^2 + \sum_{j=0}^{i} |v'(j)|^2}. \tag{4.7}$$

<u>PROBLEM 2.</u> [LINEARIZED H^∞ ESTIMATION] *Given (4.2)–(4.5), determine, if possible, causal 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 bounded by a given positive constant ξ^2, say*

$$r'(i) < \xi^2 \quad \textit{for} \quad 0 \le i \le N. \tag{4.8}$$

Assume again that we collect the estimation errors $e'_p(i)$ into a column vector $\mathbf{e}'_p$, the modified noise sequence $v'(i)$ into $\mathbf{v}'$, and define the modified disturbance vector $\mathbf{n}' = \text{col}\{\bar{\mathbf{x}}_0, \mathbf{v}'\}$. If a solution to the linearized estimation problem exists, then it would induce a (block) lower triangular mapping from $\mathbf{n}'$ to $\mathbf{e}'_p$, say $\mathcal{T}'_N$, whose 2−induced norm will be bounded by ξ.

Problem 2 is a special case of a standard finite-horizon linear H^∞−filter design. One possible solution is the following H^∞− filter – e.g., [26], [9].

<u>ALGORITHM 1.</u> [A-POSTERIORI FILTER] *Estimates $\hat{z}'(j|j)$ that meet the requirements (4.8) exist if, and only if, the matrices $\mathbf{P}_{j+1}$ given below are positive definite for $j = 0, 1, \ldots, N$. In this case, we can take $\hat{z}'(j|j) = \hat{\mathbf{g}}_j\hat{\mathbf{x}}_{j|j}$, where the state estimates $\hat{\mathbf{x}}_{j|j}$ (also denoted by $\mathbf{w}_j$) can be evaluated recursively as follows:*

$$\hat{\mathbf{x}}_{j|j} = \hat{\mathbf{x}}_{j-1|j-1} + \tag{4.9}$$

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

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

$$\mathbf{P}_{j+1} = \mathbf{P}_j - \mathbf{P}_j \begin{bmatrix} \hat{\mathbf{g}}_j^* & \hat{\mathbf{h}}_j^* \end{bmatrix} \mathbf{R}_{e,j}^{-1} \begin{bmatrix} \hat{\mathbf{g}}_j \\ \hat{\mathbf{h}}_j \end{bmatrix} \mathbf{P}_j, \tag{4.10}$$

with

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

The Riccati difference equation (4.10) can be rewritten in an alternative form that will be more convenient for our later analysis. By employing the matrix-inversion formula (see, e.g., [11]), we obtain the following update for $\mathbf{P}_j^{-1}$,

$$\mathbf{P}_{j+1}^{-1} = \mathbf{P}_j^{-1} - \xi^{-2}\hat{\mathbf{g}}_j^*\hat{\mathbf{g}}_j + \hat{\mathbf{h}}_j^*\hat{\mathbf{h}}_j \ . \tag{4.11}$$

This form shows how the positivity of $\mathbf{P}_j$ is affected by the vectors $\{\hat{\mathbf{g}}_j, \hat{\mathbf{h}}_j\}$ and will be useful in later sections.

The filter of Algorithm 1 is a so-called *filtered* or *a-posteriori* version since each estimate $\hat{z}'(i|i)$ also depends on the most recent measurement $y(i)$.

There is a related *estimation* or *a-priori* version, which estimates the $z'(i)$ by using only the data $\{y(j)\}$ up to time $(i-1)$ rather than time i. If we denote these estimates by $\hat{z}'(i|i-1)$ and the corresponding estimation errors by

$$e_a'(i) = z'(i) - \hat{z}'(i|i-1), \tag{4.12}$$

then the a-priori filter considers instead the ratios

$$r''(i) = \frac{\sum_{j=0}^{i} |e_a'(i)|^2}{\|\tilde{\mathbf{x}}_0\|_2^2 + \sum_{j=0}^{i-1} |v'(j)|^2} , \tag{4.13}$$

and tries to bound them, say by

$$r''(i) < \xi^2 \quad \text{for} \quad 0 \le i \le N. \tag{4.14}$$

<u>Algorithm 1a.</u> [A-priori Filter] *Estimates $\hat{z}'(j|j-1)$ that meet the requirements (4.14) exist if, and only if, the matrices $\tilde{\mathbf{P}}_j$ given below are positive definite for $j = 0, 1, \ldots, N$. In this case, we can take $\hat{z}'(j|j-1) = \hat{\mathbf{g}}_j\hat{\mathbf{x}}_{j|j-1}$, where the state estimates $\hat{\mathbf{x}}_{j|j-1}$ (also denoted by $\mathbf{w}_{j-1}$) 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

$$\hat{\mathbf{x}}_{j+1|j} = \hat{\mathbf{x}}_{j|j-1} + \tag{4.15}$$

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

with initial condition $\hat{\mathbf{x}}_{0|-1} = \bar{\mathbf{x}}_0$ and $\mathbf{P}_j$ is as in Algorithm 1.

We focus in the next few sections on the a-posteriori filter (Algorithm 1) and return in later sections to the a-priori version (Algorithm 1A).

C. Step 2: A Feedback Structure

As mentioned earlier, the solution given by Algorithm 1 induces a mapping $\mathcal{T}_N'$ from the modified disturbance vector $\mathbf{n}'$ to the modified estimation errors $\mathbf{e}_p'$ (see definitions after (4.8)). Its 2−induced norm is guaranteed to be bounded by ξ in view of (4.8). This is schematically indicated in Fig. 1.

In view of (4.4), the modified disturbance $v'(i)$ is related to the original disturbance $v(i)$ and to the difference $[\mathbf{h}_i(\mathbf{w}) - \hat{\mathbf{h}}_i]\mathbf{w}$. However, it may often happen (see examples in next section) that the difference $[\mathbf{h}_i(\mathbf{w}) - \hat{\mathbf{h}}_i]\mathbf{w}$ can be related to $e_p'(i)$, and in these cases we would be able to re-express the modified disturbance $v'(i)$ in terms of $\{v(j), e_p'(j)\}$.

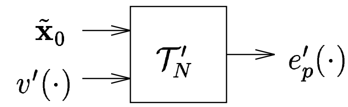

Fig. 1. *Induced map.*

This possibility is dependent on how the estimates $\{\hat{\mathbf{h}}_i, \hat{\mathbf{g}}_i\}$ are constructed and several examples are provided in the next section. For the time being, let us simply assume that such a construction has been determined. It would then induce a relation between the original noise vector $\mathbf{v}$, the modified estimation errors $\mathbf{e}_p'$, and the modified noise vector $\mathbf{v}'$, say of the general form

$$\mathbf{v}' = \mathcal{V}_N\mathbf{v} + \mathcal{F}_N\mathbf{e}_p' \ , \tag{4.16}$$

where $\mathcal{F}_N$ and $\mathcal{V}_N$ denote causal linear operators (or filters). Incorporating these operators into Fig. 1 would lead to the feedback structure of Fig. 2.

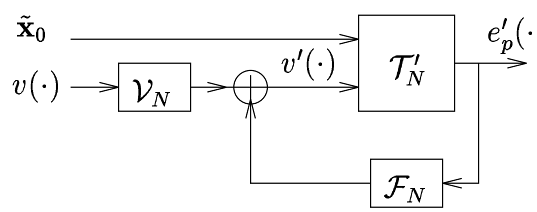

Fig. 2. *Feedback structure of the linearized solution.*

V. Several Illustrative Examples

Before discussing the implications of the feedback scheme of Fig. 2 on the overall desired robustness performance, we first exhibit several examples that illustrate how such feedback structures can be induced by proper constructions of the estimates $\{\hat{\mathbf{h}}_i, \hat{\mathbf{g}}_i\}$. We start with the autoregressive model of Section 2.

A. Back to the Autoregressive Example

In the autoregressive problem of Section 2, we were given noisy measurements $y(i)$ satisfying (2.6) with the row vector $\mathbf{h}_i(\mathbf{w})$ being equal to $\mathbf{d}_{i-1}$ in (2.2). That is, the entries of $\mathbf{h}_i(\mathbf{w})$ were time-delayed versions of each other,

$$\mathbf{h}_i(\mathbf{w}) = \begin{bmatrix} d(i-1) & d(i-2) & \ldots & d(i-M) \end{bmatrix}, \tag{5.1}$$

where each $d(i)$ in turn satisfied

$$d(i) = \mathbf{d}_{i-1}\mathbf{w} + u(i) = \mathbf{h}_i(\mathbf{w})\mathbf{w} + u(i). \tag{5.2}$$

Moreover, we were interested in estimating the uncorrupted term $\mathbf{d}_{i-1}\mathbf{w}$, which we defined as $z(i) = \mathbf{h}_i(\mathbf{w})\mathbf{w}$ in (2.7). Therefore, for this example, we have $\mathbf{g}_i(\mathbf{w}) = \mathbf{h}_i(\mathbf{w})$.

A construction for the estimate $\hat{\mathbf{h}}_i$ is suggested by the above expressions. Indeed, assume we incorporate a similar time-shift structure into the entries of $\hat{\mathbf{h}}_i$ (or $\hat{\mathbf{d}}_{i-1}$), say (compare with (5.1))

$$\hat{\mathbf{h}}_i = \left[\ \hat{d}(i-1) \quad \hat{d}(i-2) \quad \ldots \quad \hat{d}(i-M)\ \right], \qquad (5.3)$$

where the estimates $\hat{d}(i)$ are further evaluated as suggested by the defining relation (5.2), viz.,

$$\hat{d}(i) = \hat{\mathbf{d}}_{i-1}\mathbf{w}_i + u(i) = \hat{\mathbf{h}}_i\mathbf{w}_i + u(i). \qquad (5.4)$$

Here, as mentioned eariler, $\mathbf{w}_i$ denotes an estimate for $\mathbf{w}$ that is based on the data up to time i. We may further note that the estimate $\hat{\mathbf{h}}_i$ in (5.3) is *not* of the bootstrap type (4.1). The reason being that each entry $\hat{d}(j)$ in (5.3) is dependent on $\mathbf{w}_j$ (in view of (5.4)) and, consequently, $\hat{\mathbf{h}}_i$ is not only a function of the most recent estimate $\mathbf{w}_{i-1}$.

The constructions (5.3) and (5.4) lead to a relation of the form (4.16) between $v'(i)$ and $\{v(i), e'_p(i)\}$, where now (since $\hat{\mathbf{g}}_i = \hat{\mathbf{h}}_i$)

$$e'_p(i) = \hat{\mathbf{h}}_i\mathbf{w} - \hat{\mathbf{h}}_i\mathbf{w}_i. \qquad (5.5)$$

To verify this claim, let us denote the difference $[d(i) - \hat{d}(i)]$ by $\tilde{d}(i)$ (which is in fact equal to $e_p(i)$). This measures the error in estimating $d(i)$ by employing $\hat{d}(i)$. Let us also associate with the unknown vector $\mathbf{w}$ the polynomial $A(q^{-1}) = \sum_{k=1}^{M} a_k q^{-k}$, where the $\{a_k\}$ denote the entries of $\mathbf{w}$.

It is then easy to see that expression (4.4) for $v'(i)$ becomes

$$\begin{aligned} v'(i) &= \left[\mathbf{h}_i(\mathbf{w}) - \hat{\mathbf{h}}_i\right]\mathbf{w} + v(i), \\ &= A(q^{-1})\tilde{d}(i) \ + \ v(i). \end{aligned} \qquad (5.6)$$

We now relate $\tilde{d}(i)$ to $e'_p(i)$. Indeed,

$$\begin{aligned} \tilde{d}(i) &= \mathbf{h}_i(\mathbf{w})\mathbf{w} - \hat{\mathbf{h}}_i\mathbf{w}_i, \\ &= \left\{\mathbf{h}_i(\mathbf{w}) - \hat{\mathbf{h}}_i\right\}\mathbf{w} \ + \ e'_p(i), \\ &= A(q^{-1})\tilde{d}(i) \ + \ e'_p(i), \\ &= \frac{1}{1 - A(q^{-1})}\, e'_p(i). \end{aligned} \qquad (5.7)$$

Combining with (5.6) we see that

$$v'(i) = v(i) \ + \ \frac{A(q^{-1})}{1 - A(q^{-1})}\, e'_p(i),$$

which provides an explicit relation among the variables $\{v'(i), v(i), e'_p(i)\}$, as desired.

In terms of the structure of FIG. 2 we have $\mathcal{V}_N = I$ (the identity operator), and $\mathcal{F}_N$ equal to the $(N+1) \times (N+1)$ leading triangular Toeplitz operator that describes the action of $A/(1-A)$ over the first $(N+1)$ samples of $\{e'_p(\cdot)\}$ (in the absence of initial conditions). The entries of $\mathcal{F}_N$ are the first $(N+1)$ coefficients of the expansion of $A/1 - A$ in terms of powers of q^{-1}. This is depicted schematically in FIG. 3.

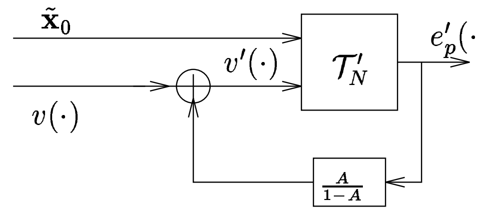

Fig. 3. *The autoregressive model.*

B. An Example of a Sinusoidal Nonlinearity

Assume w is a scalar parameter that we wish to identify. For this purpose, noisy measurements $\{y(i)\}$ are available that are related to w as follows:

$$y(i) = (c + a(i)\cos[b(i)w])\ w \ + \ v(i), \qquad (5.8)$$

where $a(i)$ and $b(i)$ are known (scalar-valued) sequences, and c is a known positive constant. In the language of the model (3.1), this corresponds to the choice

$$h_i(w) = c + a(i)\cos[b(i)w].$$

[$h_i(w)$ is now a scalar-valued function rather than a vector function. For this reason, we are not employing boldface notation to refer to $h_i(w)$ in order to be consistent with our earlier convention].

The quantity $z(i)$ that is of interest in this problem is $z(i) = w$ and therefore (cf. (3.4)), $g_i(w) = 1$. That is, $g_i(w)$ is simply the unity constant and, consequently, we can set $\hat{g}_i = 1 = g_i(w)$. In this example, the functions $h_i(w)$ and $g_i(w)$ are different and, furthermore, $e'_p(i) = e_p(i)$ since no approximations are needed for $g_i(w)$.

We proceed to replace $h_i(w)$ by the estimate

$$\hat{h}_i = c + a(i)\cos[b(i)w_{i-1}] = h_i(w_{i-1}), \qquad (5.9)$$

where w_{i-1} denotes an estimate for the coefficient w that is based on the observations up to time $(i-1)$. This is now an estimate of the bootstrap form (4.1). Based on this construction we have

$$e_p(i) = e'_p(i) = \hat{g}_i(w - w_i) = w - w_i, \qquad (5.10)$$

and (cf. (4.4))

$$v'(i) = v(i) + [h_i(w) - h_i(w_{i-1})]w. \qquad (5.11)$$

Assuming the function $h_i(w)$ is real-valued and sufficiently smooth (or continuous), we now invoke a useful result from mathematical analysis, viz., the mean-value theorem. It allows us to replace the difference $[h_i(w) - h_i(w_{i-1})]$ by a scalar multiple of $(w - w_{i-1})$. More precisely, the mean-value theorem guarantees the existence of a point $\bar{w}_{i-1}$ (lying along the segment connecting w and w_{i-1}) such that the following exact equality holds:

$$h_i(w) - h_i(w_{i-1}) \ = \ \dot{h}_i(\bar{w}_{i-1})\,[w - w_{i-1}]\,,$$

where $\dot{h}_i$ denotes the derivative of h_i and is equal to $\dot{h}_i(w) = -a(i)b(i)\sin[b(i)w]$. This allows us to rewrite (5.11) as

$$v'(i) = v(i) - a(i)b(i)we_p(i-1)\sin[b(i)\bar{w}_{i-1}] \; .$$

This again establishes a relation among the quantities $\{v'(\cdot), v(\cdot), e'_p(\cdot)\}$, as shown in Fig. 4. In this case, the feedback loop is a unit delay with coefficient $-a(i)b(i)w\sin[b(i)\bar{w}_{i-1}]$.

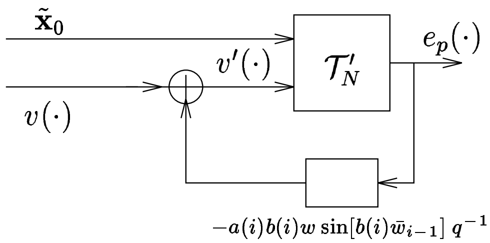

Fig. 4. *A sinusoidal nonlinearity.*

We shall see later in our analysis of the l_2-stability of the approximate solution that what matters is the norm of the feedback loop. In other words, the fact that $\bar{w}_{i-1}$ is an unknown in Fig. 4 does not pose a significant problem since its presence will be overcome by noting that a sinusoidal function is always bounded by one no matter what the value of its argument is (see the discussion in Sec. 6.2).

A related example that fits into this remark arises in the context of Perceptron training in neural networks. In this case, a feedback structure of the form shown in Fig. 4 also arises, with a feedback loop that depends on the derivative of the activation function. A discussion along these lines can be found in [23], where it is further shown how to improve the convergence speed of the training phase by studying the energy flow through the feedback interconnection.

C. A Third Example: A Squaring System

Assume again that w is a scalar parameter that we wish to identify and that

$$y(i) = w^2 \; + \; v(i). \tag{5.12}$$

This corresponds to $h_i(w) = w$ and $g_i(w) = 1$. We set $\hat{h}_i(w) = w_{i-1}$ and $\hat{g}_i = g_i(w) = 1$. Consequently,

$$e_p(i) = e'_p(i) = w - w_i, \tag{5.13}$$

and (cf. (4.4))

$$v'(i) = v(i) + (w - w_{i-1})w = v(i) + e'_p(i-1)w.$$

In this case, the feedback loop is also a unit delay with gain equal to w itself.

D. A Fourth Example: The Vector Case

We may as well mention here that we can replace the *row* vectors $\mathbf{h}_i(\mathbf{w})$ and $\mathbf{g}_i(\mathbf{w})$ in (3.3) and (3.4) by *matrices*, say $\mathbf{H}_i(\mathbf{w})$ and $\mathbf{G}_i(\mathbf{w})$, respectively. Accordingly, the scalar quantities $\{y(i), v(i), z(i), e'_p(i), e'_a(i)\}$ would become column vectors, say $\{\mathbf{y}_i, \mathbf{v}_i, \mathbf{z}_i, \mathbf{e}'_{p,i}, \mathbf{e}'_{a,i}\}$ [recall that we use subscripts, rather than parenthesis, to time-index vector and matrix quantities].

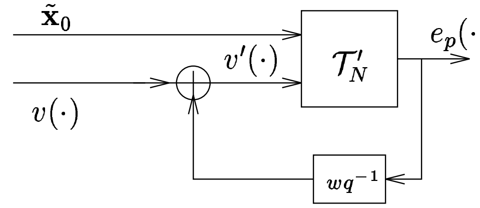

Fig. 5. *A squaring system.*

The statements and solutions of Problems 1 and 2 remain unchanged except for the notational change of replacing $\hat{\mathbf{h}}_i$ and $\hat{\mathbf{g}}_i$ by $\hat{\mathbf{H}}_i$ and $\hat{\mathbf{G}}_i$, resp. We also replace the scalar quantities by the respective vector quantities.

For example, consider again expression (3.3),

$$y(i) = \mathbf{h}_i(\mathbf{w})\mathbf{w} \; + \; v(i), \tag{5.14}$$

and assume we are interested in estimating the column vector $\mathbf{w}$. In this case, $\mathbf{h}_i(\mathbf{w})$ is still a row vector but $\mathbf{G}_i(\mathbf{w}) = \mathbf{I}$ is now the identity matrix.

If we replace $\mathbf{h}_i(\mathbf{w})$ by $\mathbf{h}_i(\mathbf{w}_{i-1})$ and use

$$\mathbf{e}_{p,i} = \mathbf{e}'_{p,i} = \mathbf{w} - \mathbf{w}_i. \tag{5.15}$$

Then, according to (4.4),

$$v'(i) = v(i) + [\mathbf{h}_i(\mathbf{w}) - \mathbf{h}_i(\mathbf{w}_{i-1})]\mathbf{w}. \tag{5.16}$$

As an illustration, consider the following contrived example:

$$y(i) = v(i) + \tag{5.17}$$

$$\begin{bmatrix} e^{\alpha(i)w_1} & \ln[\beta(i)w_2] & \gamma(i)w_3^2 \end{bmatrix} \begin{bmatrix} w_1 \\ w_2 \\ w_3 \end{bmatrix},$$

where $\{w_1, w_2, w_3\}$ are the entries of the unknown vector $\mathbf{w}$, and $\{\alpha(i), \beta(i), \gamma(i)\}$ are known sequences. Let $\{w_{1,i}, w_{2,i}, w_{3,i}\}$ denote estimates of the weight-vector entries. Then, using the mean-value theorem, we can write

$$v'(i) = v(i) +$$

$$\mathbf{e}^{\mathbf{T}}_{p,i-1} \begin{bmatrix} \alpha(i)e^{\alpha(i)\bar{w}_{1,i-1}} & 0 & 0 \\ 0 & \frac{1}{\bar{w}_{2,i-1}} & 0 \\ 0 & 0 & 2\gamma(i)\bar{w}_{3,i-1} \end{bmatrix} \mathbf{w},$$

where

$$\mathbf{e}^{\mathbf{T}}_{p,i} = \begin{bmatrix} w_1 - w_{1,i} & w_2 - w_{2,i} & w_3 - w_{3,i} \end{bmatrix}.$$

VI. l_2−STABILITY OF THE FEEDBACK STRUCTURE

We now return to the general setting of the feedback structure shown in Fig. 2 and assume that the mappings $\mathcal{T}'_N$, $\mathcal{V}_N$, and $\mathcal{F}_N$ have been identified (as demonstrated in the above examples).

The question of interest in this section is to verify under what conditions on the norms of $\{\mathcal{T}'_N, \mathcal{V}_N, \mathcal{F}_N\}$ the overall mapping from the original disturbance vector $\mathbf{n}$ (which includes $\tilde{\mathbf{x}}_0$ and $v(\cdot)$, as defined in (3.9)) to the modified estimation errors $\mathbf{e}'_p$ (defined after (4.8)) is l_2−stable.

For this purpose, we let $\|\cdot\|_{2,ind}$ denote the 2−induced norm of a linear operator, e.g.,

$$\|\mathcal{T}'_N\|_{2,ind} \;=\; \sup_{\mathbf{x}\neq 0} \frac{\|\mathcal{T}'_N\mathbf{x}\|_2}{\|\mathbf{x}\|_2} \;.$$

<u>THEOREM 1.</u> *Consider the recursive solution of Algorithm 1 and assume that the estimates $\hat{\mathbf{h}}_i$ and $\hat{\mathbf{g}}_i$ result in a feedback structure of the form indicated in FIG. 2. If the following condition is satisfied,*

$$\|\mathcal{T}'_N\|_{2,ind}\|\mathcal{F}_N\|_{2,ind} < 1 \;, \tag{6.1}$$

then the mapping from the original disturbances $\mathbf{n}$ to the modified errors $\mathbf{e}'_p$ is l_2−stable in the following sense,

$$\|\mathbf{e}'_p\|_2 \;\leq\; k_N\,\{\,\|\tilde{\mathbf{x}}_0\|_2 \;+\; \|\mathcal{V}_N\|_{2,ind}\|\mathbf{v}\|_2\,\}\,, \tag{6.2}$$

where

$$k_N = \frac{\|\mathcal{T}'_N\|_{2,ind}}{1-\|\mathcal{T}'_N\|_{2,ind}\|\mathcal{F}_N\|_{2,ind}}.$$

<u>PROOF</u>: The filter $\mathcal{F}_N$ maps the vector $\mathbf{e}'_p$ to another vector, say $\mathbf{n}_f$. Likewise, the filter $\mathcal{V}_N$ maps the vector $\mathbf{v}$ to another vector, say $\mathbf{n}_v$. It follows from the definition of the 2−induced norm of an operator that

$$\begin{aligned}\|\mathbf{n}_v\|_2 &\leq \|\mathcal{V}_N\|_{2,ind}\,\|\mathbf{v}\|_2,\\ \|\mathbf{n}_f\|_2 &\leq \|\mathcal{F}_N\|_{2,ind}\,\|\mathbf{e}'_p\|_2.\end{aligned}$$

But $\mathbf{v}' = \mathbf{n}_v + \mathbf{n}_f$ and it follows from the triangle inequality of norms that

$$\|\mathbf{v}'\|_2 \;\leq\; \|\mathbf{n}_v\|_2 + \|\mathbf{n}_f\|_2.$$

Therefore,

$$\|\mathbf{v}'\|_2 \leq \|\mathcal{V}_N\|_{2,ind}\|\mathbf{v}\|_2 \;+\; \|\mathcal{F}_N\|_{2,ind}\|\mathbf{e}'_p\|_2. \tag{6.3}$$

Now $\mathcal{T}'_N$ maps $\{\tilde{\mathbf{x}}_0, \mathbf{v}'\}$ into $\mathbf{e}'_p$ and, hence,

$$\|\mathbf{e}'_p\|_2 \;\leq\; \|\mathcal{T}_N\|_{2,ind}\;\{\,\|\tilde{\mathbf{x}}_0\|_2 \;+\; \|\mathbf{v}'\|_2\}\,. \tag{6.4}$$

Using (6.1) and (6.3) in (6.4) we obtain (6.2). ■

The above statement can be regarded as a special manifestation of the small gain theorem – e.g., [12, p.214] and [25, p.337] – when applied to the feedback connection of FIG. 2.

Note also that since we already know that $\|\mathcal{T}'_N\|_{2,ind} < \xi$ (when a solution to the design Problem 2 exists), then a sufficient condition for (6.1) to hold is to require

$$\|\mathcal{F}_N\|_{2,ind} \leq 1/\xi. \tag{6.5}$$

A. A Remark on the Limit Behaviour

Assume a limiting (infinite-horizon) solution $\mathcal{T}'_N$ exists to Problem 2 as $N\to\infty$ (examples are discussed further ahead in Sections 7 and 8, where it is shown that in some cases of interest the solvability condition of Algorithm 1 becomes trivial). Accordingly, let $\mathcal{T}'$, $\mathcal{F}$, and $\mathcal{V}$ denote the (corresponding) semi-infinite operators satisfying (6.1) and (6.2) with $N\to\infty$ and $\|\mathcal{V}\|_{2,ind} < \infty$.

In this situation, with the noise sequence $\{v(\cdot)\}$ having finite energy, i.e.,

$$\sum_{j=0}^{\infty} |v(j)|^2 < \infty,$$

the estimation errors will also have finite energy (cf. (6.2)),

$$\sum_{j=0}^{\infty} |e'_p(j)|^2 < \infty.$$

This implies that error convergence is guaranteed, i.e.,

$$\lim_{j\to\infty} e'_p(j) = 0.$$

If we return to the example of sub-section 5.1 we see from expression (5.7) that a convergent $e'_p(i)$ would imply a convergent $\tilde{d}(i)$, which would in turn imply that $z(i)$ is recovered since $\hat{d}(i)\to d(i)$ and $z(i) = d(i) - u(i)$.

B. Back to the Sinusoidal Example

In the sinusoidal example of Section 5.2, the feedback loop is a unit delay that is given by $-a(i)b(i)w\sin[b(i)\bar{w}_{i-1}]q^{-1}$. That is, at each time instant i, the error $e_p(i-1)$ is simply scaled by the scalar quantity $-a(i)b(i)w\sin[b(i)\bar{w}_{i-1}]$.

Assume, for now, that a feedforward robust filter $\mathcal{T}'_N$ of level ξ exists [that is, that the solvability conditions of Algorithm 1 are satisfied]. The l_2−stability of the overall system of Fig. 4 can then be guaranteed if we require (cf. (6.5))

$$|a(i)b(i)w\sin[b(i)\bar{w}_{i-1}]| \;\leq\; \frac{1}{\xi} \;\text{ for all i.} \tag{6.6}$$

This condition is in terms of $\bar{w}_{i-1}$, which is unknown. To overcome this difficulty, we may simply invoke the fact that $-1\leq \sin[b(i)\bar{w}_{i-1}]\leq 1$ no matter what $\bar{w}_{i-1}$ is. In this case, a sufficient condition for (6.6) to hold is to require

$$\sup_i\; |a(i)b(i)w| \leq \frac{1}{\xi}. \tag{6.7}$$

This in effect specifies a region of the real axis for which l_2−stability can be guaranteed: if w lies in this region then a robust nonlinear estimator can be designed according to the explanation in the earlier sections. We now clarify this statement.

Recall from Algorithm 1 that the existence of a robust feedforward filter $\mathcal{T}'_N$ requires the positivity of the Riccati variables $\mathbf{P}_j$. In the current context, using (5.9), expression (4.11) becomes ($\mathbf{P}_j$ is now a scalar quantity written as $p(j)$)

$$p^{-1}(j+1) = p^{-1}(j) - \xi^{-2} + |c + a(j)\cos[b(j)w_{j-1}]|^2.$$

Since the initial value $p(0)$ is assumed positive (due to the choice of Π_0), the above update shows that as long as $|c+a(j)\cos[b(j)w_{j-1}]|^2$ is not smaller than ξ^{-2}, the successive $p(j)$ will be guaranteed to remain positive. Now the function $\cos(\cdot)$ is always bounded by 1. Therefore, a sufficient condition for the existence of the filter $\mathcal{T}'_N$ is for the $\{a(j)\}$ to satisfy

$$(c - |a(j)|)^2 \geq \xi^{-2}.$$

Assume $c > \sup_j |a(j)|$. Then the above result may also be interpreted as follows: it suggests a choice for ξ. In other words, if one chooses ξ such that

$$\xi^{-1} \leq c - \sup_j |a(j)|\ , \qquad (6.8)$$

then a filter $\mathcal{T}'_N$ will be guaranteed to exist.

Assume, for example, that $c = 6$ and

$$|b(i)| \leq 0.2\ , \quad 2.0 \leq |a(i)| \leq 4.0 \qquad (6.9)$$

for all i. According to (6.8), one can choose any ξ such that $\xi^{-1} \leq 2.0$ and the feedforward filter $\mathcal{T}'_N$ of Algorithm 1 will exist.

The particular choice of ξ will end up restricting the interval over which w can lie for a guaranteed overall l_2-stable design (because of (6.7)). In particular, the choice $\xi = 1$ requires (cf. (6.7)) $|0.8w| \leq 1$ or, equivalently,

$$w \in [-1.25, 1.25].$$

To confirm the above results, noisy measurements $y(i)$ were generated via

$$y(i) = [6 + a(i)\cos(b(i)w)]\ w + v(i),$$

with $w = 0.54$ and where the sequences $a(i)$ and $b(i)$ were generated randomly with values bounded as in (6.9). Also, the noise sequence was randomly chosen in l_2, and the level of robustness ξ was chosen to be $\xi = 1$. Algorithm 1 was then used with initial conditions $\bar{x}_0 = 0$ and $p(0) = 1$. Figure 6 shows the signal $y(i)$ and the resulting error signal $(w - w_i)$. It is clear from the figure that the error $e_p(i)$ of the robust filter approaches zero rather fast.

C. Back to the Square System

A similar discussion holds for the squaring system of Sec. 5.3, where the feedback loop is determined by w itself. Assume again that a feedforward robust filter $\mathcal{T}'_N$ with level ξ exists. The l_2-stability of the overall system of Fig. 5 would then require (cf. (6.5)) that

$$|w| \leq \frac{1}{\xi} \iff -\frac{1}{\xi} \leq w \leq \frac{1}{\xi}.$$

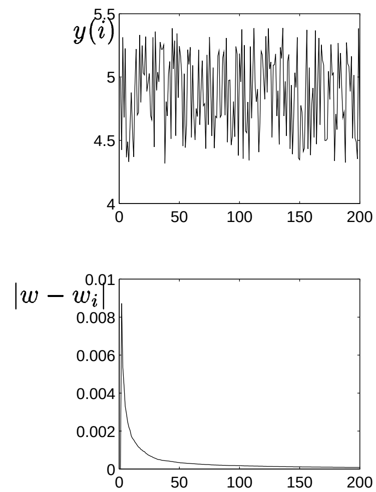

Fig. 6. *Simulation for a sinusoidal nonlinearity.*

In other words, if the unknown w lies in the interval $[-1/\xi, 1/\xi]$ then a robust nonlinear estimator can be designed by following the discussion of the earlier sections.

Moreover, in this example, $\hat{h}_i = w_{i-1}$ and the solvability conditions would require the positivity of the Riccati variables $p(j)$ in (cf. (4.11))

$$p^{-1}(j+1) = p^{-1}(j) - \xi^{-2} + |w_{j-1}|^2.$$

This suggests that if ξ^{-2} is chosen small enough (smaller than the energies of the successive weight estimates) then $|w_{j-1}|^2 - \xi^{-2} \geq 0$ for all j and the successive $p(j)$ will be positive.

Figure 7 shows the results of a simulation for a squaring system $y(i) = w^2 + v(i)$ with $w = 0.54$. The noise sequence was randomly chosen in l_2, and the initial guess was $\bar{x}_0 = 0.5$ with $p(0) = 1$ and $\xi = 3.0$. It was observed for this example that the difference $|w_i|^2 - \xi^{-2}$ remained positive so that the solvability conditions were satisfied during the simulation.

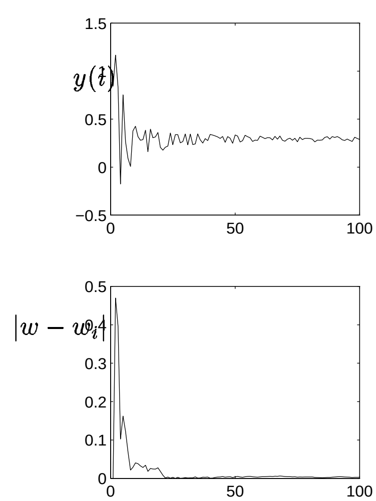

Fig. 7. *Simulation for a squaring system.*

VII. 7. The Case of Systems with Shift Structure

The case of systems where the $\mathbf{h}_i(\mathbf{w})$ vector has time-shifted entries often arises in applications. We examined an example to this effect in the context of autoregressive modeling in Sec. 2. A similar situation also arises in ARMA modeling as discussed later. For this reason, we shall now study in more details this subclass of systems and clarify the connections with several earlier results in the literature.

We still consider the measurement model $y(i) = \mathbf{h}_i(\mathbf{w})\mathbf{w} + v(i)$ with the desired signal taken as $z(i) = \mathbf{h}_i(\mathbf{w})\mathbf{w}$, and assume the row vector $\mathbf{h}_i(\mathbf{w})$ has *shift* structure ,

$$\mathbf{h}_i(\mathbf{w}) = \left[\begin{array}{cccc} d(i-1) & d(i-2) & \ldots & d(i-M) \end{array}\right]. \quad (7.1)$$

For generality, we further assume that each entry $d(i)$ is generated via a relation of the form

$$d(i) = S(q^{-1})[\mathbf{h}_i(\mathbf{w})\mathbf{w}], \quad (7.2)$$

where $S(q^{-1})$ denotes a known linear time-invariant filter. This means that $d(i)$ is obtained by filtering the signal $\mathbf{h}_i(\mathbf{w})\mathbf{w}$ through S. The special case $S(q^{-1}) = 1$ arises in autoregressive modeling and was considered in Section 2.

Following the discussion of sub-section 5.1, we again define

$$\hat{\mathbf{h}}_i = \left[\begin{array}{cccc} \hat{d}(i-1) & \hat{d}(i-2) & \ldots & \hat{d}(i-M) \end{array}\right], \quad (7.3)$$

where the $\{\hat{d}(\cdot)\}$ are evaluated via (as suggested by (7.2))

$$\hat{d}(i) = S(q^{-1})[\hat{\mathbf{h}}_i\mathbf{w}_i]. \quad (7.4)$$

In this case, we have

$$e_p'(i) = \hat{\mathbf{h}}_i\mathbf{w} - \hat{\mathbf{h}}_i\mathbf{w}_i,$$

and we get

$$v'(i) \;=\; W(q^{-1})[\tilde{d}(i)] \;+\; v(i), \quad (7.5)$$

with

$$\tilde{d}(i) \;=\; \frac{S(q^{-1})}{1 - S(q^{-1})W(q^{-1})}[e_p'(i)], \quad (7.6)$$

where $W(q^{-1})$ is the polynomial associated with the entries of $\mathbf{w}$. Consequently,

$$v'(i) = v(i) \;+\; \frac{S(q^{-1})W(q^{-1})}{1 - S(q^{-1})W(q^{-1})}[e_p'(i)],$$

which again provides an explicit relation between the variables $\{v'(\cdot), v(\cdot), e_p'(\cdot)\}$. This is depicted in Fig. 8.

A sufficient condition for (6.1) to hold is to require $\|\mathcal{F}_N\|_{2,ind} < 1/\xi$. This is satisfied if

$$\frac{SW}{1-SW} \text{ is stable}, \quad (7.7)$$

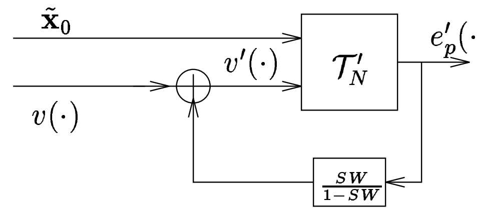

Fig. 8. *The case of systems with shift structure.*

and

$$\max_{\omega} \left| \frac{\xi S(e^{j\omega})W(e^{j\omega})}{1 - S(e^{j\omega})W(e^{j\omega})} \right| < 1, \quad (7.8)$$

for $0 \leq \omega < 2\pi$.

Writing down Algorithm 1 for this case (with $\hat{\mathbf{h}}_i = \hat{\mathbf{g}}_i$) we obtain the following filter equations for $0 \leq j \leq N$ [the equations are now written, for convenience of exposition and for ease of comparison with results from the literature, in terms of $\mathbf{w}_j$ rather than $\hat{\mathbf{x}}_{j|j}$ – the initial condition is now also denoted by $\mathbf{w}_{-1}$ rather than $\bar{\mathbf{x}}_0$]:

$$\mathbf{w}_j = \mathbf{w}_{j-1} + \quad (7.9)$$

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

$$\mathbf{P}_{j+1} = \quad (7.10)$$

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

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

The positivity of the matrices $\mathbf{P}_{j+1}$ over $0 \leq j \leq N$, as well as condition (7.8), guarantee that the above filter attains a level of robustness that is upper-bounded by

$$k_N \;=\; \frac{\|\mathcal{T}_N'\|_{2,ind}}{1 - \|\mathcal{T}_N'\|_{2,ind}\|\mathcal{F}_N\|_{2,ind}}.$$

That is, in view of (6.2),

$$\|\mathbf{e}_p'\|_2 \;\leq\; k_N\left\{\,\|\tilde{\mathbf{x}}_0\|_2 \;+\; \|\mathbf{v}\|_2\,\right\}. \quad (7.11)$$

The positivity condition is in fact always met if ξ is chosen to be a positive real number not smaller than one. This is clarified in the next section, where we highlight a connection to the so-called PLR algorithm.

A. *The Pseudo-Linear Regression Algorithm*

An interesting point to note here is that the $\mathrm{H}^\infty-$based algorithm (7.9) can in fact be related to so-called *pseudo-linear regression* (PLR) algorithms in IIR modeling – e.g., [14][p. 167]. To clarify this, we first note, as in (4.11), that the Riccati recursion for $\mathbf{P}_j$ in the above algorithm is equivalent to

$$\mathbf{P}_{j+1}^{-1} = \mathbf{P}_j^{-1} + (1 - \xi^{-2})\hat{\mathbf{h}}_j^*\hat{\mathbf{h}}_j. \quad (7.12)$$

This means that the positivity condition on the $\{\mathbf{P}_j\}$ is *always satisfied* for any choice $\xi \geq 1$ and $\Pi_0 > 0$. Moreover, if we again invert (7.12) we obtain that the Riccati recursion (7.10) can be rewritten in the equivalent (and more recognizable) form

$$\mathbf{P}_{j+1} = \mathbf{P}_j - \frac{\mathbf{P}_j \hat{\mathbf{h}}_j^* \hat{\mathbf{h}}_j \mathbf{P}_j}{(1-\xi^{-2})^{-1} + \hat{\mathbf{h}}_j \mathbf{P}_j \hat{\mathbf{h}}_j^*} . \tag{7.13}$$

Expressions (7.9) and (7.13) constitute the a-posteriori form of the so-called PLR algorithm. A related a-priori version is derived further ahead at the end of Section 9.2.

Expression (7.9) also includes as special cases other forms of recursive identification schemes, other than the PLR algorithm discussed above, such as the important class of instantaneous-gradient-based filters. This is detailed in the next section.

VIII. Instantaneous-Gradient-Based IIR Adaptive Filters

Note that the Riccati recursion (7.10) (or (7.12)) trivializes in an important special case given below. This fact was first noted in [8] in the linear context of FIR (or MA) identification [see also [4] and [22] for a related discussion in the continuous-time case]. We now extend the result to the nonlinear scenario of the previous sections.

If ξ is chosen to be one, $\xi = 1$, then recursion (7.12) 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 $\Pi_0 > 0$, which is always satisfied since Π_0 is assumed to be positive-definite. In particular, this holds for a special choice of the form $\Pi_0 = \alpha \mathbf{I}$, a (positive) constant multiple of the identity. For this choice, the update of the weight estimate (7.9) reduces to

$$\mathbf{w}_j = \mathbf{w}_{j-1} + \frac{\alpha}{1+\alpha\|\hat{\mathbf{h}}_j\|_2^2} \hat{\mathbf{h}}_j^* \left[y(j) - \hat{\mathbf{h}}_j \mathbf{w}_{j-1} \right], \tag{8.1}$$

which is an instantaneous-gradient-based recursion (also known as NLMS [7], [1]) with a step-size of the form

$$\frac{\alpha}{1+\alpha\|\hat{\mathbf{h}}_j\|_2^2}.$$

Exponential stability of the NLMS algorithm (8.1) in the *noise-free* case ($v'(i) = 0$) has been extensively discussed in the literature (e.g., [1][Sec. 2.6]). It has been shown to require, along with a persistence of excitation condition on $\{\hat{\mathbf{h}}_j\}$, the strict positive-realness of the transfer function $1/(1-SW)$ [1][Thm. 2.9].

In the next subsection we connect the framework of this paper with these earlier studies by showing that the positive-realness condition alone also arises in the *noisy case*, as a result of (7.8), and it serves to guarantee an overall l_2-stable (or robust) algorithm.

A. Stability Analysis and Loop Transformations

The l_2-stability condition (7.8) for Figure 8 requires $SW/(1-SW)$ to be strictly contractive (since ξ is now taken to be one). Noting that we can write

$$\frac{SW}{1-SW} = \frac{1}{1-SW} - 1, \tag{8.2}$$

and using the fact that for any complex number z, the following conditions are equivalent,

$$|z-1| < 1 \iff \text{Re}\,(1/z) > 1/2,$$

we conclude that the contractivity requirement on $SW/(1-SW)$ is equivalent to the positive-real part of $(1-SW)$ being larger than $1/2$,

$$\text{Re}\left[1 - S(e^{j\omega})W(e^{j\omega})\right] > 1/2. \tag{8.3}$$

This guarantees an l_2-stable system

$$\text{from} \quad \left\{\alpha^{-1/2}\tilde{\mathbf{w}}_{-1}, v(\cdot)\right\} \quad \text{to} \quad \{e_p'(\cdot)\} \tag{8.4}$$

where $\tilde{\mathbf{w}}_{-1} = \mathbf{w} - w_{-1}$. The condition (8.3) can be relaxed by applying a scaling tool that can be related to so-called loop transformations in passivity analysis. For this purpose, we first establish two preliminary results. The first result rewrites (8.1) in an alternative form, which has already been noted earlier in the literature in the *noise free* case ($v'(i) = 0$) – e.g., [1][Sec. 2.6.1].

<u>Lemma 1.</u> [Alternative Update] *The update equation (8.1) can be rewritten in the equivalent form*

$$\mathbf{w}_j = \mathbf{w}_{j-1} + \alpha \hat{\mathbf{h}}_j^* \left[e_p'(j) + v'(j) \right], \tag{8.5}$$

where, as defined earlier, $e_p'(j) = \hat{\mathbf{h}}_j \mathbf{w} - \hat{\mathbf{h}}_j \mathbf{w}_j$, *and* $v'(j) = y(j) - \hat{\mathbf{h}}_j \mathbf{w}$.

<u>Proof</u>: Note that $e_p'(j) + v'(j) = y(j) - \hat{\mathbf{h}}_j \mathbf{w}_j$. Therefore, all we need to establish is the identity

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

But it follows from the update equation (8.1) that

$$\hat{\mathbf{h}}_j \mathbf{w}_j = \frac{\hat{\mathbf{h}}_j \mathbf{w}_{j-1}}{1+\alpha\|\hat{\mathbf{h}}_j\|_2^2} + \frac{\alpha\|\hat{\mathbf{h}}_j\|_2^2\, y(j)}{1+\alpha\|\hat{\mathbf{h}}_j\|_2^2}.$$

Subtracting $y(j)$ from both sides and multiplying by α leads to the desired equality.

■

The map from $\{\alpha^{-1/2}\tilde{\mathbf{w}}_{-1}, v'(\cdot)\}$ to $\{e_p'(\cdot)\}$ is a strict contraction since, as argued above, the recursion (8.5), which is equivalent to (8.1), is an a-posteriori $H^\infty-$filter and the positivity condition is always satisfied due to $\mathbf{P}_j = \Pi_0 = \alpha \mathbf{I} > 0$, and $\xi = 1$.

It is also clear that, in fact, this result holds for any update filter of the form (8.5) and for any noise sequence

$v'(\cdot)$. In other words, it holds whenever we have a recursive equation of the form (we are deliberately changing the notation here to $\{\beta, \bar{e}_p, \bar{v}\}$ for generality):

$$\mathbf{w}_j = \mathbf{w}_{j-1} + \beta\hat{\mathbf{h}}_j^* \left[\bar{e}_p(j) \ + \ \bar{v}(j)\right], \tag{8.6}$$

for an arbitrary noise sequence $\{\bar{v}(\cdot)\}$, an arbitrary positive number β, and for $\bar{e}_p(j) = \hat{\mathbf{h}}_j\mathbf{w} - \hat{\mathbf{h}}_j\mathbf{w}_j$. We summarize this in the form of a lemma, the proof of which follows immediately from the original H^∞ motivation.

<u>LEMMA 2.</u> [A CONTRACTIVE MAPPING] *Consider an update relation of the form (8.6). It always holds that*

$$\frac{\sum_{j=0}^{N} |\bar{e}_p(j)|^2}{\beta^{-1}\|\tilde{\mathbf{w}}_{-1}\|_2^2 \ + \ \sum_{j=0}^{N} |\bar{v}(j)|^2} < 1. \tag{8.7}$$

<u>PROOF</u>: Given an arbitrary noise sequence $\{\bar{v}(\cdot)\}$, we can define the sequence $\{\bar{y}(\cdot)\}$

$$\bar{y}(j) = \hat{\mathbf{h}}_j\mathbf{w} + \bar{v}(j).$$

Now, recursion (8.6) can be equivalently rewritten in the form (in view of Lemma 1)

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

which can be readily seen to be a special case of the $\mathrm{H}^\infty-$filter of Algorithm 1 with $\Pi_0 = \beta\mathbf{I} > 0$ and $\xi = 1$. This establishes that (8.7) holds. ∎

Using the relation established after (7.6), and expression (8.2), we have

$$v'(j) = -e_p'(j) + \frac{1}{1-SW}[e_p'(j)] + v(j),$$

which allows us to rewrite (8.5) in the form

$$\mathbf{w}_j \ = \ \mathbf{w}_{j-1} + \alpha\hat{\mathbf{h}}_j^* \left[\frac{1}{1-SW}[e_p'(j)] + v(j)\right].$$

These observations motivate us to reexpress the update recursion (8.5) in the equivalent form:

$$\begin{aligned}\mathbf{w}_j \ &= \ \mathbf{w}_{j-1} + \beta\hat{\mathbf{h}}_j^* \left[\frac{\alpha/\beta}{1-SW}[e_p'(j)] + \frac{\alpha}{\beta}v(j)\right], \\ &= \ \mathbf{w}_{j-1} + \beta\hat{\mathbf{h}}_j^* \left[e_p'(j) \ + \ v''(j)\right], \end{aligned} \tag{8.8}$$

where we have now defined

$$v''(j) = -e_p'(j) + \frac{\alpha/\beta}{1-SW}[e_p'(j)] + \frac{\alpha}{\beta}v(j), \tag{8.9}$$

and where β is *any* positive real number. The recursion (8.8) guarantees, in view of Lemma 2 above, a strictly contractive map from $\{\beta^{-1/2}\tilde{\mathbf{w}}_{-1}, v''(\cdot)\}$ to $\{e_p'(\cdot)\}$. Accordingly, an overall l_2-stable system (compare with (8.4))

$$\text{from} \quad \left\{\beta^{-1/2}\tilde{\mathbf{w}}_{-1}, \frac{\alpha}{\beta}v(\cdot)\right\} \quad \text{to} \quad \{e_p'(\cdot)\} \tag{8.10}$$

will be guaranteed if we impose

$$\max_{\omega}\left|\frac{\alpha/\beta}{1-S(e^{j\omega})W(e^{j\omega})} \ - 1\right| \ < \ 1, \tag{8.11}$$

which requires (compare with (8.3))

$$\mathrm{Re}\ \left[1 - S(e^{j\omega})W(e^{j\omega})\right] > \frac{\alpha}{2\beta}. \tag{8.12}$$

Since this should be true for any choice of β, we therefore conclude, by choosing β large enough (but finite), that a sufficient condition for the l_2-stability of (8.1), in the sense of (8.10), is the strict positive-realness of the function $(1 - SW)$.

Now recall that if the real part of a complex number z is positive then the same is true about the real part of $1/z$. This allows us to conclude that a sufficient condition for the l_2-stability of (8.1) is the strict positive-realness of the function $1/(1 - SW)$. Hence, if $\{v(\cdot)\}$ is a finite-noise energy sequence, then so is $\{\frac{\alpha}{\beta}v(\cdot)\}$, and we can still conclude from the l_2-stability condition that $e_p'(i) \to 0$.

In the next section we consider an important special case that arises in ARMA modeling (see also [1][Sec. 5.1.3]).

B. 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

$$\begin{aligned} d(j) \ &= \ \sum_{k=1}^{M_a} a_k\, d(j-k) + \sum_{k=0}^{M_b-1} b_k\, u(j-k), \\ &= \ \begin{bmatrix} \mathbf{d}_{j-1} & \mathbf{u}_j \end{bmatrix} \begin{bmatrix} \mathbf{a} \\ \mathbf{b} \end{bmatrix} \\ &= \ \mathbf{h}_j(\mathbf{w})\mathbf{w}, \end{aligned} \tag{8.13}$$

where

$$\begin{aligned} \mathbf{d}_{j-1} \ &= \ \begin{bmatrix} d(j-1) & \ldots & d(j-M_a) \end{bmatrix}, \\ \mathbf{u}_j \ &= \ \begin{bmatrix} u(j) & \ldots & u(j-M_b+1) \end{bmatrix}, \\ \mathbf{a} \ &= \ \mathrm{col}\{a_1, a_2, \ldots, a_{M_a}\}, \\ \mathbf{b} \ &= \ \mathrm{col}\{b_0, b_1, \ldots, b_{M_b-1}\}, \end{aligned}$$

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 [14], is one that updates the weight estimate according to expression (8.1) and where $\hat{\mathbf{h}}_j$ is computed as

$$\hat{\mathbf{h}}_j = \left[\begin{array}{cc} \hat{\mathbf{d}}_{j-1} & \mathbf{u}_j \end{array} \right].$$

Note that $\mathbf{u}_j$ is known, while the entries of

$$\hat{\mathbf{d}}_{j-1} = \left[\begin{array}{ccc} \hat{d}(j-1) & \ldots & \hat{d}(j-M_a) \end{array} \right],$$

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:

$$\hat{d}(j) \;=\; \hat{\mathbf{d}}_{j-1}\mathbf{a}_j + \mathbf{u}_j\mathbf{b}_j, \tag{8.14}$$

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 (7.4) (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 $A(q^{-1})$, where $A(q^{-1})$ is the shift polynomial that is associated with the coefficients in $\mathbf{a}$. This is because the difference $(\mathbf{h}_j - \hat{\mathbf{h}}_j)$ now has the form $\left[\begin{array}{cc} (\mathbf{d}_{j-1} - \hat{\mathbf{d}}_{j-1}) & \mathbf{0} \end{array} \right]$. 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 the l_2-stability of Landau's scheme is to require the strict positive-realness of $1/(1-A)$.

While this is a known result for Landau's scheme – e.g., [24, 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 *a-posteriori* H^∞-filter of Algorithm 2, and that the corresponding solvability condition has been trivialized by choosing $\Pi_0 = \alpha\mathbf{I}$.

IX. l_2−Stability of the Approximate A-Priori Filter

The analysis of Sections 4 through 8 are equally applicable to the a-priori filter of Algorithm 1A. For this reason, we shall be brief in this section and highlight only the points that are distinctive of the a-priori case. As it turns out, some subtle points persist that turn out to mark the difference between the a-posteriori and a-priori filters.

In fact, most of the analyses in the literature address stability issues of a-posteriori versions only, such as the NLMS algorithm, the PLR algorithm, and Landau's scheme that were discussed in the earlier sections – e.g., [14][Sec. 5.3], [1][Secs. 2.6 and 5.1], and [24][Sec. 6.2].

While an averaging analysis, along the lines of [3], [24], can be pursued for a-priori adaptive schemes, the conclusions would generally hold only for very small adaptation gains. In Sections 9.4–9.5 further ahead, we study the a-priori versions without requiring beforehand that the adaptation gains be very small. Instead, the solvability conditions of the a-priori H^∞ formulation will be shown to indicate how large the adaptation gains can be for guaranteed l_2−stability (see, e.g., (8.32) and (8.33)).

In other words, the point of view taken in this paper helps clarify some subtle differences that exist between the a-priori and a-posteriori versions. This is achieved by raising and exploiting connections with the design of a-priori and a-posteriori H^∞−filters and by highlighting the differences in their solvability (or existence) conditions.

A. The Approximate A-priori Solution

To begin with, note that the numerator in the ratio $r''(i)$ in (4.13) includes the *a-priori* error term

$$e'_a(j) = \hat{\mathbf{g}}_j\mathbf{w} - \hat{\mathbf{g}}_j\mathbf{w}_{j-1},$$

while the denominator includes the modified noise sequence $v'(j) = y(j) - \hat{\mathbf{h}}_j\mathbf{w}$.

Assume again that we collect the a-priori errors into a column vector,

$$\mathbf{e}'_a = \mathrm{col}\{e_a(0), e_a(1), \ldots, e_a(N)\},$$

the original and modified noise sequences into two vectors,

$$\mathbf{v} = \mathrm{col}\{v(0), v(1), \ldots, v(N-1)\},$$

$$\mathbf{v}' = \mathrm{col}\{v'(0), v'(1), \ldots, v'(N-1)\},$$

and the initial weight error $\tilde{\mathbf{x}}_0$ along with $\mathbf{v}'$ into a modified disturbance vector $\mathbf{n}' = \mathrm{col}\{\tilde{\mathbf{x}}_0, \mathbf{v}'\}$. Define also $\mathbf{n} = \mathrm{col}\{\tilde{\mathbf{x}}_0, \mathbf{v}\}$. Let $\mathcal{T}'_N$ denote the (causal) operator that maps the modified disturbances $\mathbf{n}'$ to the estimation errors $\mathbf{e}'_a$. In view of Algorithm 1A, this operator is constructed so as to have a 2−induced norm that is bounded by ξ.

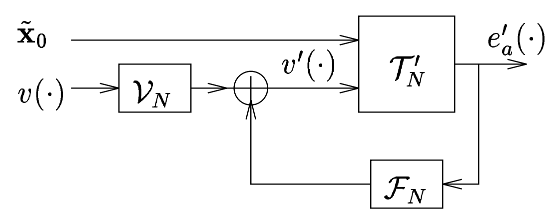

Fig. 9. *Feedback structure of the a-priori solution.*

If we further assume that estimates $\hat{\mathbf{h}}_j$ and $\hat{\mathbf{g}}_j$ are constructed in such a way so as to result in an explicit relation between $v'(j)$ and $\{v(j), e'_a(j)\}$ (FIG. 9), then the following statement is immediate in much the same way as in the a-posteriori version of Theorem 1.

<u>Theorem 1A.</u> *Consider the recursive solution of Algorithm 1A and assume that the estimates $\hat{\mathbf{h}}_j$ and $\hat{\mathbf{g}}_j$ result in a feedback structure of the form indicated in FIG. 9. If the following condition is satisfied,*

$$\|\mathcal{T}'_N\|_{2,ind}\|\mathcal{F}_N\|_{2,ind} < 1\,, \tag{8.15}$$

then the mapping from the original disturbances $\mathbf{n}$ to the modified errors $\mathbf{e}'_a$ is l_2−stable with finite gain in the following sense,

$$\|\mathbf{e}'_a\|_2 \;\leq\; k_N\,\{\; \|\tilde{\mathbf{x}}_0\|_2 \;+\; \|\mathcal{V}_N\|_{2,ind}\|\mathbf{v}\|_2 \;\}\,, \tag{8.16}$$

where

$$k_N = \frac{\|\mathcal{T}'_N\|_{2,ind}}{1 - \|\mathcal{T}'_N\|_{2,ind}\|\mathcal{F}_N\|_{2,ind}}.$$

B. The Case of Systems With Shift Structure

We reconsider the example of Section 7 and also assume that $\mathbf{h}_i(\mathbf{w}) = \mathbf{g}_i(\mathbf{w})$ exhibits *shift* structure as in (7.1), with the estimate $\hat{\mathbf{h}}_i$ given by (7.3) except that now the $\{\hat{d}(\cdot)\}$ are evaluated via

$$\hat{d}(i) = S(q^{-1})[\hat{\mathbf{h}}_i \mathbf{w}_{i-1}]. \tag{8.17}$$

Comparing with (7.4) we see that we now have $\mathbf{w}_{i-1}$ instead of $\mathbf{w}_i$. This allows us to relate $v'(i)$ and $\{v(i), e'_a(i)\}$ as follows (FIG. 10):

$$v'(i) = v(i) + \frac{S(q^{-1})W(q^{-1})}{1 - S(q^{-1})W(q^{-1})}\, e'_a(i).$$

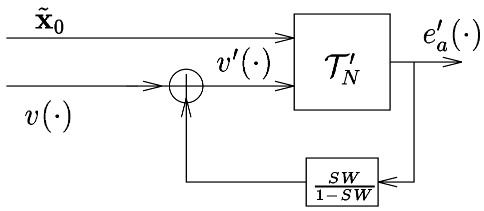

Fig. 10. *Systems with shift structure: A-priori case.*

Again, a sufficient condition for (8.15) to hold is to require $\|\mathcal{F}_N\|_{2,ind} < 1/\xi$. This is satisfied if

$$\frac{SW}{1 - SW} \text{ is stable,} \tag{8.18}$$

and

$$\max_{\omega} \left| \frac{\xi S(e^{j\omega})W(e^{j\omega})}{1 - S(e^{j\omega})W(e^{j\omega})} \right| < 1, \tag{8.19}$$

for $0 \leq \omega < 2\pi$.

Writing down Algorithm 1A for this case we obtain [we use $\mathbf{w}_j$ instead of $\hat{\mathbf{x}}_{j+1|j}$ and also denote the initial condition by $\mathbf{w}_{-1}$]:

$$\mathbf{w}_j = \mathbf{w}_{j-1} + \tag{8.20}$$

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

$$\mathbf{P}_{j+1} = \tag{8.21}$$

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

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

$$\tilde{\mathbf{P}}_j = \left[\mathbf{P}_j^{-1} - \xi^{-2}\hat{\mathbf{h}}_j^* \hat{\mathbf{h}}_j\right]^{-1}.$$

The positivity of the matrices $\tilde{\mathbf{P}}_j$ over $0 \leq j \leq N$, as well as condition (8.19), guarantee that the above filter attains a level of robustness that is upper-bounded by

$$k_N = \frac{\|\mathcal{T}'_N\|_{2,ind}}{1 - \|\mathcal{T}'_N\|_{2,ind}\|\mathcal{F}_N\|_{2,ind}}.$$

That is, in view of (8.16),

$$\|\mathbf{e}'_a\|_2 \leq k_N \{ \|\tilde{\mathbf{x}}_0\|_2 + \|\mathbf{v}\|_2 \}, \tag{8.22}$$

C. Gauss Newton Updates and PLR Algorithm

Recursion (8.21) can also be rewritten in a more familiar form. As argued before, the inverse of the Riccati variable $\mathbf{P}_{j+1}$ in (8.21) can be updated as in (7.12). Therefore,

$$\tilde{\mathbf{P}}_j^{-1} + \hat{\mathbf{h}}_j^* \hat{\mathbf{h}}_j = \mathbf{P}_{j+1}^{-1}.$$

Consequently, using the matrix inversion formula [11], we obtain

$$\begin{aligned}(1 + \hat{\mathbf{h}}_j \tilde{\mathbf{P}}_j \hat{\mathbf{h}}_j^*)^{-1} &= 1 - \hat{\mathbf{h}}_j(\hat{\mathbf{h}}_j^* \hat{\mathbf{h}}_j + \tilde{\mathbf{P}}_j^{-1})^{-1}\hat{\mathbf{h}}_j^*, \\ &= 1 - \hat{\mathbf{h}}_j \mathbf{P}_{j+1}^{-1} \hat{\mathbf{h}}_j^*.\end{aligned}$$

Likewise,

$$\begin{aligned}\tilde{\mathbf{P}}_j \hat{\mathbf{h}}_j^* &= \left[\mathbf{P}_j^{-1} - \xi^{-2}\hat{\mathbf{h}}_j^* \hat{\mathbf{h}}_j\right]^{-1} \hat{\mathbf{h}}_j^*, \\ &= \left[\mathbf{P}_{j+1}^{-1} - \hat{\mathbf{h}}_j^* \hat{\mathbf{h}}_j\right]^{-1} \hat{\mathbf{h}}_j^*, \\ &= \mathbf{P}_{j+1}\hat{\mathbf{h}}_j^* \left[1 - \hat{\mathbf{h}}_j \mathbf{P}_{j+1}^{-1} \hat{\mathbf{h}}_j^*\right]^{-1}.\end{aligned}$$

Therefore,

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

This allows us to rewrite the recursions (8.20)–(8.21) in the compact (and more recognizable a-priori PLR) form:

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

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

In the special case of FIR (or MA) filters, the above form is often known as the Gauss-Newton update. More details on the robustness and stability of such updates can be found in [19].

D. Instantaneous-Gradient-Based Algorithms

The recursions (8.20)–(8.21) also collapse to an instantaneous-gradient-based filter. Indeed, if ξ is chosen to be one, $\xi = 1$, then we obtain

$$\mathbf{P}_{j+1}^{-1} = \mathbf{P}_j^{-1} = \Pi_0^{-1},$$

where Π_0 is the initial condition. In particular, for the special choice $\Pi_0 = \mu\mathbf{I}$, a (positive) constant multiple of the identity, we obtain

$$\mathbf{w}_j = \mathbf{w}_{j-1} + \mu\hat{\mathbf{h}}_j^* \left[y(j) - \hat{\mathbf{h}}_j \mathbf{w}_{j-1}\right], \tag{8.23}$$

which is an instantaneous-gradient-based recursion with a constant step-size parameter (also known as the LMS algorithm [8]).

Now note that the solvability requirement $\tilde{\mathbf{P}}_j > 0$ is not automatically satisfied. It requires

$$\mu^{-1}\mathbf{I} - \hat{\mathbf{h}}_j^* \hat{\mathbf{h}}_j > \mathbf{0}, \quad 0 \leq i \leq N.$$

The matrix $\mu^{-1}\mathbf{I} - \hat{\mathbf{h}}_j^* \hat{\mathbf{h}}_j$ is a rank-one modification of the identity and hence, we may equivalently require

$$\mu < \min_{0 \leq i \leq N} \frac{1}{\|\hat{\mathbf{h}}_j\|_2^2}. \tag{8.24}$$

This is in clear contrast to the a-posteriori case in (8.1), where no condition is imposed on α. Moreover, in order to guarantee a non-zero upper bound on μ (as N becomes larger) we now require that the sequence $\{\hat{\mathbf{h}}_j\}$ to be bounded.

E. A Stability Analysis

The stability condition of Algorithm 1A requires $SW/(1-SW)$ to be strictly contractive, which is equivalent to requiring the positive-real part of $(1-SW)$ to be larger than 1/2,

$$\text{Re}\left[1 - S(e^{j\omega})W(e^{j\omega})\right] > \frac{1}{2}.$$

This condition, along with (8.24), guarantees an l_2-stable map from $\{\tilde{\mathbf{x}}_0, \mathbf{v}\}$ to $\mathbf{e}'_a$, where $\tilde{\mathbf{x}}_0 = \mu^{-1/2}\tilde{\mathbf{w}}_{-1}$.

Now note that an l_2-stable map allows us to conclude convergence of $e'_a(\cdot)$ to zero for a finite-energy noise sequence $\{v(\cdot)\}$. But if $\{v(\cdot)\}$ has finite-energy then the same holds for any noise sequence that is a constant multiple of $v(\cdot)$, say $\{\lambda v(\cdot)\}$, for a finite constant λ. This suggests that we may replace (8.24) by another condition, with the intent of guaranteeing an l_2-stable map from $\{\lambda v(\cdot)\}$ to $\{e'_a(\cdot)\}$ rather than from $\{v(\cdot)\}$ to $\{e'_a(\cdot)\}$, for some constant λ. This is clarified in the sequel.

First, note that

$$\begin{aligned} y(j) - \hat{\mathbf{h}}_j \mathbf{w}_{j-1} &= \hat{\mathbf{h}}_j \mathbf{w} + v'(j) - \hat{\mathbf{h}}_j \mathbf{w}_{j-1}, \\ &= e'_a(j) + v'(j). \end{aligned}$$

Hence, the update equation (8.23) can be rewritten in the form

$$\mathbf{w}_j = \mathbf{w}_{j-1} + \mu \hat{\mathbf{h}}_j^* \left[e'_a(j) + v'(j)\right], \tag{8.25}$$

where, as established earlier,

$$v'(j) = -e'_a(j) + \frac{1}{1-SW}[e'_a(j)] + v(j). \tag{8.26}$$

We can also use (8.26) to write (8.25) in the equivalent form

$$\mathbf{w}_j = \mathbf{w}_{j-1} + \mu \hat{\mathbf{h}}_j^* \left[\frac{1}{1-SW}[e'_a(j)] + v(j)\right].$$

Let us for the moment ignore any restriction on μ, such as (8.24), and simply require that it be a positive constant. Now choose any constant β and assume that it satisfies

$$\beta < \inf_j \frac{1}{\|\hat{\mathbf{h}}_j\|_2^2}. \tag{8.27}$$

Using β, we can rewrite the above update recursion for $\mathbf{w}_j$ in the equivalent form

$$\begin{aligned} \mathbf{w}_j &= \mathbf{w}_{j-1} + \beta \hat{\mathbf{h}}_j^* \left[\frac{\mu/\beta}{1-SW}[e'_a(j)] + \frac{\mu}{\beta} v(j)\right], \\ &= \mathbf{w}_{j-1} + \beta \hat{\mathbf{h}}_j^* \left[e'_a(j) + \hat{v}'(j)\right], \end{aligned} \tag{8.28}$$

where we have now defined

$$\hat{v}'(j) = -e'_a(j) + \frac{\mu/\beta}{1-SW}[e'_a(j)] + \frac{\mu}{\beta} v(j). \tag{8.29}$$

In view of (8.27), the recursion (8.28) also guarantees a strict contraction map from $\{\beta^{-1/2}\tilde{\mathbf{w}}_{-1}, \hat{v}'(\cdot)\}$ to $\{e'_a(\cdot)\}$. Accordingly, an overall l_2-stable system

$$\text{from } \left\{\beta^{-1/2}\tilde{\mathbf{w}}_{-1}, \frac{\mu}{\beta} v(\cdot)\right\} \text{ to } \{e'_a(\cdot)\}$$

will be guaranteed if we impose

$$\max_\omega \left| \frac{\mu/\beta}{1 - S(e^{j\omega})W(e^{j\omega})} - 1 \right| < 1, \tag{8.30}$$

which requires

$$\text{Re}\left[1 - S(e^{j\omega})W(e^{j\omega})\right] > \frac{\mu}{2\beta} > 0. \tag{8.31}$$

Assume $[1 - S(e^{j\omega})W(e^{j\omega})]$ has a strictly positive real part and let

$$\kappa = \min_\omega \text{Re}\left[1 - S(e^{j\omega})W(e^{j\omega})\right]. \tag{8.32}$$

Then, according to (8.27) and (8.31), the step size μ has to be chosen so as to satisfy

$$\mu < 2\beta\kappa < 2\kappa \inf_j \frac{1}{\|\hat{\mathbf{h}}_j\|_2^2}. \tag{8.33}$$

If this restriction is satisfied, then an l_2-stable map from $\{\beta^{-1/2}\tilde{\mathbf{w}}_{-1}, \frac{\mu}{\beta} v(\cdot)\}$ to $\{e'_a(\cdot)\}$ will result, as desired.

We thus see that the following sufficient conditions will guarantee $e'_a(i) \to 0$ for a finite-energy noise $\{v(\cdot)\}$:

(i) $[1 - S(e^{j\omega})W(e^{j\omega})]$ is strictly positive-real. Let κ be as in (8.32).

(ii) Choose μ as in (8.33).

In the next section we consider a special case that arises in IIR modeling.

F. Feintuch's Scheme for IIR Modeling

Recursion (8.23) was suggested by [6] in the context of IIR modeling, though from a very different point of view. Here we have established that it is in fact a special case of the nonlinear $H^\infty-$structure studied in this paper.

Moreover, and considering the same setting of Section 6.2, the estimate $\hat{\mathbf{h}}_j$ is computed with the $\{\mathbf{a}_j, \mathbf{b}_j\}$ in (8.14) replaced by $\{\mathbf{a}_{j-1}, \mathbf{b}_{j-1}\}$, viz., we now use

$$\hat{d}(j) \quad = \quad \hat{\mathbf{d}}_{j-1}\mathbf{a}_{j-1} + \mathbf{u}_j\mathbf{b}_{j-1}. \qquad (8.34)$$

Hence, the sufficient conditions derived in the previous section are applicable here as well. We nevertheless see that, in addition to a strict positive-realness condition, we also require that the step-size parameter be properly chosen as in (8.33).

X. Concluding Remarks

We posed a nonlinear identification problem and proposed a solution in terms of a feedback structure that consists of two steps. First, a linear approximation was employed and a standard H$^\infty-$filter design was carried out. This provided a filter with a 2−induced norm that was guaranteed to be bounded by a given constant. Then, a feedback interconnection was introduced in order to guarantee an overall l_2-stable filter, under suitable conditions on the data and system parameters.

An interesting fall out of the discussion was that it explicitly clarified the connections among several earlier IIR modeling schemes with more recent results in H$^\infty-$theory. In particular, we have addressed the so-called Landau's scheme, Feintuch's scheme, PLR algorithm, Gauss-Newton updates, and instantaneous-gradient schemes, as special cases of the general algorithms (Algorithms 1 and 1A) derived herein.

Moreover, the approach of this paper further clarified the connections between Landau's and Feintuch's schemes in IIR modeling. While a sufficient stability condition has been available for Landau's scheme in terms of a positive-realness constraint – e.g., [24, pp.146–150], a more restrictive condition is required for the closely related, yet different, Feintuch's algorithm. An explanation was provided here by showing that Feintuch's recursion required an additional condition on the data. This was obtained by establishing the following interesting fact: Landau's scheme was shown to be a special case of a so-called *a-posteriori* H$^\infty$-filter while Feintuch's algorithm was shown to be a special case of a so-called *a-priori* H$^\infty$-filter. It is known in H$^\infty-$theory that the solvability and existence conditions for both filters are different. Here we showed that in Landau's case, the condition trivialized and was therefore unnecessary, but it remained in Feintuch's case and was 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 further be shown to provide an interpretation of most adaptive schemes in terms of a feedback interconnection that consists of two major blocks: i) a *lossless* (i.e., energy preserving) feedforward mapping and ii) either a memoryless or a dynamic feedback mapping. In contrast to some earlier analyses via hyperstability results that require one of the paths to be time-invariant [14][p. 381], both mappings in the feedback composition of this work are allowed to be time-variant [20]. Examples to this effect were in fact provided in the paper (e.g., Fig. 4 and also Sec. 6.2). Moreover, the losslessness of the feedforward path can be shown to allow for interesting energy arguments that help analyze the performance of the algorithms as well as design variants with improved convergence speed (e.g., [23]). These details will be discussed elsewhere.

Acknowledgment

The authors would like to thank the anonymous reviewers for their constructive comments and suggestions that have considerably improved the presentation of the results.

References

[1] Anderson, B. D. O., R. R. Bitmead, Jr. C. R. Johnson, P. V. Kokotovic, R. L. Kosut, I. M. Y. Mareels, L. Praly, and B. D. Riedle (1986). *Stability of Adaptive Systems: Passivity and Averaging Analysis.* MIT Press, Cambridge, MA.

[2] Basar, T. and P. Bernhard (1991). *H^∞-Optimal Control and Related Minimax Design Problems - A Dynamic Game Approach.* Birkhauser, Boston, Berlin, Basel.

[3] Benveniste, A., M. Métivier and P. Priouret (1990). *Adaptive Algorithms and Stochastic Approximations.* Springer-Verlag, Berlin Heidelberg,

[4] Didinsky, G., Z. Pan and T. Basar (1995). Parameter identification for uncertain plants using $H^\infty-$methods. *Automatica,* **31**(9), 1227–1250.

[5] 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.

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

[7] Goodwin, G. C. and K. S. Sin (1984). *Adaptive Filtering, Prediction and Control*, Prentice-Hall, NJ.

[8] Hassibi, B., A. H. Sayed and T. Kailath (1993). LMS is H^∞ optimal. In: *Proc. CDC.* San Antonio, Texas, vol.1, pp. 74–79. See also *IEEE Transactions on Signal Processing*, **44**(2), 267–280, 1996.

[9] Hassibi, B., A. H. Sayed and T. Kailath (1996). Linear estimation in Krein spaces - Part I: Theory and Part II: Applications, *IEEE Trans. Automatic Control* **4**(1), 18–33 and 34-49.

[10] Jang, J. and J. L. Speyer, (1994). A class of nonlinear game-theoretic filters and controllers using a dissipative approach. In: *Proc. CDC*, Lake Buena Vista, FL, vol.1, pp. 799–804.

[11] Kailath, T. (1980). *Linear Systems.* Prentice Hall.

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

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

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

[15] Landau, I. D. (1984). A feedback system approach to adaptive filtering. *IEEE Transactions on Information Theory,* **IT-30**, no. 2, pp. 251–262.

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

[17] Ljung, L. (1977) "On positive real transfer functions and the convergence of some recursive schemes. *IEEE Transactions on Automatic Control* **22**, 539–551.

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

[19] Rupp, M. and A. H. Sayed (1996*a*). Robustness of Gauss-Newton recursive methods: A deterministic feedback analysis. *Signal Processing*, vol. 50, no.3, pp. 165–188.

[20] Rupp, M. and A. H. Sayed (1996*b*). A time-domain feedback analysis of filtered-error adaptive gradient algorithms. *IEEE Transactions on Signal Processing,* **44**(6), 1428–1439. See also *Proc. SPIE Conference*, San Diego, CA, vol. 2563, pp. 458–469, 1995.

[21] Sayed, A. H. and T. Kailath (1994). A state-space approach to adaptive RLS filtering. *IEEE Signal Processing Magazine*, vol. 11, no. 3, pp. 18–60.

[22] Sayed, A. H. and P. V. Kokotovic (1995). A note on a Lyapunov argument for stochastic gradient methods in the presence of noise. *Proc. ACC*, Seattle, WA, vol. 1, pp. 587–588.

[23] Sayed, A. H. and M. Rupp (1995). A feedback analysis of perceptron learning for neural networks. *Proc. Asilomar Conference on Signals, Systems, and Computers*, vol. 2, pp. 894–898.

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

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

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