PROC. 3RD IEEE MEDITERRANEAN SYMPOSIUM ON NEW DIRECTIONS IN CONTROL AND AUTOMATION, VOL. 1, PP. 249–256, CYPRUS, JULY 1995.

A TIME-DOMAIN FEEDBACK-ANALYSIS OF RECURSIVE IDENTIFICATION SCHEMES

ALI H. SAYED AND MARKUS RUPP

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

Abstract—This paper provides a time-domain feedback-analysis of several adaptive schemes with emphasis on stability and robustness issues. It is shown that an intrinsic feedback structure, mapping the noise sequence and the initial weight guess to the apriori estimation errors and the final weight estimate, can be associated with such schemes. The feedback configuration is motivated via energy arguments and is shown to consist of two major blocks: a time-variant lossless (i.e., energy preserving) feedforward path and a time-variant feedback path. The configuration is further shown to lend itself rather immediately to analysis via a so-called small gain theorem; thus leading to stability conditions that require the contractivity of certain operators.

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, a nice account of which is given by Landau [1], the ODE approach of Ljung [2], and the related class of averaging methods for trajectory approximation, as described in Anderson et al. [3] and in Solo and Kong [4].

I.1 Stability Analysis of Adaptive Schemes

The hyperstability approach is based on a stability theorem of Popov [5], which extends a result from linear feedback theory to a class of nonlinear feedback systems. The theorem is applicable to a feedback interconnection with a linear time-invariant system, say $H(z)$, in the feedforward path and a possibly nonlinear time-variant system in the feedback loop. Such interconnections often arise in IIR (i.e., infinite-impulse-response) modeling and have received considerable attention in the literature. In simple terms, the hyperstability theorem states that if the nonlinear system obeys a so-called *Popov's inequality* then the overall connection is stable provided the feedforward transfer function $H(z)$ is *strictly positive real.* The application of this condition to the analysis and development of stable adaptive algorithms was pioneered by Landau [1] and further refined by Johnson [6] , thus leading to algorithms that go by the names of HARF, SHARF and PLR.

The stability analysis in the hyperstability framework is usually carried out in the noise-free case, i.e., in the absence of measurement disturbances. The effect of the disturbances can be examined by resorting to another widely used method for the convergence analysis of recursive schemes, known as the ODE approach. Its application to the adaptive context was pioneered by Ljung [3], and the basic idea is to associate a differential equation with a discrete-time recursive scheme. Under some technical assumptions, including slow adaptation, a stable point of the differential equation can be shown to be a convergent point in the mean for the adaptive algorithm.

ODE-based analyses have been carried out for both cases when the adaptation gain is a constant, say μ [7], or vanishing, say $\mu(i) \rightarrow 0$ [2]. The latter case leads to algorithms that essentially turn off as time progresses, thus debilitating the tracking capabilities of the algorithm in potentially nonstationary (or time-variant) environments. In the former case of constant μ, the ODE analysis requires slow adaptation and the convergence conclusions only hold for sufficiently small μ. This is an issue that usually hinders obtaining more powerful conclusions from an ODE analysis.

This situation is in fact characteristic of the class of averaging methods, of which the ODE approach is a prominent member. The averaging methods are tools designed for the approximation of the trajectories of either difference or differential equations, accounts of which can be found in [3,4]. In the discrete-time case, for instance, the approximation is achieved by associating with a general, possibly nonlinear and time-variant, difference equation a so-called averaged system that is described by a time-invariant equation. The analysis of the time-invariant model is often more amenable to standard stability tools, such as those from Lyapunov analysis, than the original system. Here also the convergence properties of the original and the averaged systems can be tied together as long as a certain parameter,

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 by a scholarship from DAAD (German Academic Exchange Service) as well as the scientific division of NATO (Wissenschaftsausschuß der NATO). Email-contact: sayed@ece.ucsb.edu.

say the adaptation gain μ in our case, is sufficiently small (e.g., Ch. 7 of [4]).

I.2 A Time-Domain Feedback Analysis

The above results have motivated us to take an alternative look at two issues that are relevant to any performance analysis of recursive identification schemes. We focus not only on the potential *stability* of an adaptive scheme but also on its *robustness* performance in the presence of disturbances. Here, the term *disturbance* is used to refer to both *measurement noise* and *uncertainties* in the initial condition.

In this respect, this paper suggests a time-domain feedback framework that is useful for both the analysis and design of recursive estimators. A major motivation throughout our derivation is to see how far we can progress in the analysis without imposing restrictive assumptions on the data (noisy or not) and on the adaptation gain (small or not) right from the early stages of the analysis.

Moreover, the notion of robustness that we employ here is in perfect agreement with the notion of robust filters in studies on H^∞-filtering (see, e.g., [8] and the references therein). By explicitly bringing up the connection with H^∞-theory, we are able to exploit the wealth of results that are already available in this area.

Indeed, and in order to further highlight these connections, the derivation in future sections will be fundamentally based on a useful tool in system theory that is widely known as the small gain theorem. In loose terms, the theorem states that the 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 passivity result, the analysis provided in this paper has three distinctive features.

First, by relying on the small gain theorem, we can advantageously exploit many results that are 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 fact, it can be shown that several of the algorithms that have been derived earlier in the literature, with the objective of meeting the requirements of the hyperstability approach, are different forms of H^∞−filters, as will become apparent throughout our analysis.

We have in fact pursued this point of view in [9] where we have provided a stability analysis, along the lines of this paper, of two IIR adaptive schemes that are often attributed to Landau [1] and Feintuch [10]. While a sufficient stability condition is available for Landau's scheme in terms of a positive-realness constraint [4, pp.146–150], there did not seem to exist a similar analysis for the closely related, yet different, Feintuch's algorithm. An explanation was provided in [9] by showing that Feintuch's recursion requires 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 *aposteriori* H^∞-filter while Feintuch's algorithm was shown to be a special case of a so-called *apriori* H^∞-filter. Moreover, it is known in H^∞−theory that the solvability and existence conditions for both filters are different. We showed in [9] 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.

Similar conclusions hold for a variety of recursive schemes, which include the class of instantaneous-gradient-based estimators (e.g., LMS, normalized LMS, projection LMS, general time-variant step-sizes), the class of Gauss-Newton estimators, and the class of filtered-error variants in both the linear and nonlinear settings (e.g., filtered-error LMS, filtered-error Gauss-Newton recursions). In particular, it will follow from the arguments in this paper that these different variants are also robust filters.

The second distinctive feature of the approach suggested herein is that although the feedback nature of most of the above recursive schemes has been pointed out and advantageously exploited in earlier places in the literature (e.g., [1,2]), the feedback configuration in this paper is of a different nature. It does not only refer to the fact that the update equations can be put into a feedback form (as explained in [11]),) but is instead motivated via energy arguments that also *explicitly* take into consideration *both* the effect of the measurement *noise* and the effect of the uncertainty in the *initial guess* for the weight vector. These extensions are incorporated into the feedback arguments of this paper because our derivation is also concerned with the robustness properties of the algorithms in the presence of uncertain disturbances. This is especially useful, for example, when the statistical properties of the disturbances are unknown.

Furthermore, the feedback connection provided herein is shown to exhibit three main features that distinguish it from earlier studies in the literature: the feedforward path in the connection consists of a *lossless* (i.e., energy preserving) mapping while the feedback path consists either of a *memoryless* interconnection or, in the case of filtered-error variants, of a *dynamic* system that is dependent on the error filter. The blocks in *both* the feedforward and the feedback paths are allowed to be, and in fact are, *time-variant.* This is a distinctive feature, especially when compared with the hyperstability analysis which requires that one of the paths be time-invariant.

I.3 Notation

In the sequel, we shall use small boldface letters to denote vectors and capital boldface letters to denote matrices. Also, the symbol "$*$" will denote Hermitian conjugation (complex conjugation for scalars), and the notation $\|\mathbf{x}\|_2^2$ will denote the squared Euclidean norm of a vector.

II. STOCHASTIC GRADIENT METHODS

In order to highlight the major features of the framework proposed herein, we shall first focus on the important subclass of instantaneous-gradient-based schemes, which includes as a special case the famed least-mean-squares algorithm (LMS). Once the basic ideas are introduced, we shall then proceed to more involved situations, which include the study of filtered-error variants and Gauss-Newton methods.

II.1 The Least-Mean-Squares Algorithm

One of the most widely used adaptive algorithms is the least-mean-squares (LMS) algorithm. It starts with an initial guess $\mathbf{w}_{-1}$, for an unknown $M \times 1$ weight vector $\mathbf{w}$, and updates it via the update equation

$$\begin{aligned} \mathbf{w}_i &= \mathbf{w}_{i-1} + \mu(i)\,\mathbf{u}_i^*\,[d(i) - \mathbf{u}_i\mathbf{w}_{i-1}] \\ &= \mathbf{w}_{i-1} + \mu(i)\mathbf{u}_i^*\tilde{e}_a(i) \end{aligned} \tag{1}$$

where the $\{\mathbf{u}_i\}$ are given row vectors and the $\{d(i)\}$ are noisy measurements of the terms $\{\mathbf{u}_i\mathbf{w}\}$, viz., $d(i) = \mathbf{u}_i\mathbf{w} + v(i)$. The nonnegative factor $\mu(i)$ is the adaptation gain (step-size) parameter.

The difference $[d(i) - \mathbf{u}_i\mathbf{w}_{i-1}]$ will be denoted by $\tilde{e}_a(i)$, $\tilde{\mathbf{w}}_i$ will denote the difference between the true weight vector $\mathbf{w}$ and its estimate $\mathbf{w}_i$, $\tilde{\mathbf{w}}_i = (\mathbf{w} - \mathbf{w}_i)$, $e_a(i)$ will denote the *apriori estimation error*, $e_a(i) = \mathbf{u}_i\tilde{\mathbf{w}}_{i-1}$, and $e_p(i)$ will denote the *aposteriori estimation error*, $e_p(i) = \mathbf{u}_i\tilde{\mathbf{w}}_i$. It follows from the update equation (1) that the following relations hold:

$$\tilde{\mathbf{w}}_i = \tilde{\mathbf{w}}_{i-1} - \mu(i)\mathbf{u}_i^*\tilde{e}_a(i) \tag{2}$$

$$e_p(i) = \left[\, 1 - \mu(i)\|\mathbf{u}_i\|_2^2 \,\right]\, e_a(i) - \mu(i)\|\mathbf{u}_i\|_2^2\, v(i) \tag{3}$$

II.2 Local Error-Energy Bounds

We first establish several local error-energy bounds that can in effect be used to account for the robust behaviour of the LMS recursion in practice. For this purpose, we invoke the time-domain update recursion (2) and compute the squared norm (i.e., energies) of both of its sides. This leads to the equality:

$$\|\tilde{\mathbf{w}}_{i-1}\|_2^2 + \mu(i)\,|v(i)|^2 =$$
$$\|\tilde{\mathbf{w}}_i\|_2^2 + \mu(i)\,|e_a(i)|^2 + \mu(i)\left(1 - \mu(i)\bar{\mu}^{-1}(i)\right))|\tilde{e}_a(i)|^2$$

where we have defined $\bar{\mu}^{-1}(i) = \|\mathbf{u}_i\|_2^2$. Consequently, the following *local-bounds* always hold:

$$\frac{\|\tilde{\mathbf{w}}_i\|_2^2 + \mu(i)\,|e_a(i)|^2}{\|\tilde{\mathbf{w}}_{i-1}\|_2^2 + \mu(i)\,|v(i)|^2} \quad \begin{cases} \leq 1 & \text{for } \mu(i) < \bar{\mu}(i) \\ = 1 & \text{for } \mu(i) = \bar{\mu}(i) \\ \geq 1 & \text{for } \mu(i) > \bar{\mu}(i) \end{cases}$$

In particular, the first two bounds have an interesting robustness interpretation: they state that no matter what the value of the noise component $v(i)$ is, and no matter how far the estimate $\mathbf{w}_{i-1}$ is from the true vector $\mathbf{w}$, the sum of the energies of the resulting errors, viz., $\|\tilde{\mathbf{w}}_i\|_2^2 + \mu(i)|e_a(i)|^2$, will always be smaller than or equal to the sum of the energies of the starting errors (or disturbances), $\|\tilde{\mathbf{w}}_{i-1}\|_2^2 + \mu(i)|v(i)|^2$. This is a local conclusion but a similar result also holds over intervals of time. Indeed, note that if we assume $\mu(i) \leq \bar{\mu}(i)$ for all i in the interval $0 \leq i \leq N$, then the following inequality holds for every time instant in the interval,

$$\mu(i)\,|e_a(i)|^2 \;\leq\; \|\tilde{\mathbf{w}}_{i-1}\|_2^2 - \|\tilde{\mathbf{w}}_i\|_2^2 + \mu(i)\,|\bar{v}(i)|^2\,.$$

Summing over i we conclude that

$$\|\tilde{\mathbf{w}}_N\|_2^2 + \sum_{i=0}^{N} \mu(i)\,|e_a(i)|^2 \;\leq\; \|\tilde{\mathbf{w}}_{-1}\|_2^2 + \sum_{i=0}^{N} \mu(i)|v(i)|^2$$

which establishes a fundamental contractivity relation over the interval $0 \leq i \leq N$.

II.3 The Feedback Structure

The previous local bounds can be described via an alternative form that will lead us to an interesting feedback structure that characterizes update relations of the form (1). To clarify this, we re-express the update equation (1) in the alternative form:

$$\begin{aligned} \mathbf{w}_i &= \mathbf{w}_{i-1} + \bar{\mu}(i)\mathbf{u}_i^* e_a(i) + \\ &\quad \mathbf{u}_i^*\underbrace{[\mu(i)v(i) - (\bar{\mu}(i) - \mu(i))e_a(i)]}_{-\bar{\mu}(i)e_p(i)} \end{aligned}$$

That is,

$$\begin{aligned} \mathbf{w}_i &= \mathbf{w}_{i-1} + \bar{\mu}(i)\mathbf{u}_i^*[e_a(i) - e_p(i)] \\ &= \mathbf{w}_{i-1} + \bar{\mu}(i)\mathbf{u}_i^*[e_a(i) + \bar{v}(i)] \end{aligned} \tag{4}$$

This shows that the weight-update equation can be rewritten in terms of a new step-size parameter $\bar{\mu}(i)$ and a modified "noise" term $\bar{v}(i) = -e_p(i)$ (compare with (1)). If we now follow arguments similar to those in Sec. II.2, we readily conclude that the following equality holds <u>for all</u> $\{\mu(i), v(i)\}$,

$$\frac{\|\tilde{\mathbf{w}}_i\|_2^2 + \bar{\mu}(i)\,|e_a(i)|^2}{\|\tilde{\mathbf{w}}_{i-1}\|_2^2 + \bar{\mu}(i)\,|e_p(i)|^2} = 1 \tag{5}$$

This establishes that the map from $\{\tilde{\mathbf{w}}_{i-1}, \sqrt{\bar{\mu}(i)}\bar{v}(i)\}$ to $\{\tilde{\mathbf{w}}_i, \sqrt{\bar{\mu}(i)}e_a(i)\}$, denoted by $\overline{\mathcal{T}}_i$, is always *lossless*, i.e., it preserves energy. The overall mapping from the *original* disturbance $\sqrt{\bar{\mu}(\cdot)}v(\cdot)$ to the resulting apriori estimation error $\sqrt{\bar{\mu}(\cdot)}e_a(\cdot)$ can then be expressed in terms of a feedback structure, as in Figure 1. The feedback loop consists of a gain factor that is equal to $(1 - \mu(i)/\bar{\mu}(i))$. Moreover, using (3), the terms $\bar{v}(i)$ and $v(i)$ are related via

$$\bar{\mu}^{\frac{1}{2}}(i)\bar{v}(i) = \frac{\mu(i)}{\bar{\mu}^{\frac{1}{2}}(i)}v(i) - \left(1 - \frac{\mu(i)}{\bar{\mu}(i)}\right)\bar{\mu}^{\frac{1}{2}}(i)e_a(i) \tag{6}$$

II.4 The Small Gain Theorem: l_2-Stability

The feedback configuration of Figure 1 lends itself rather immediately to stability analysis via the small-gain theorem [12,13]. Indeed, by invoking the fact that $\overline{\mathcal{T}}_i$ is a lossless system, the overall interconnection will then be guaranteed to result in an l_2-stable mapping from $\{\sqrt{\bar{\mu}(\cdot)}\; v(\cdot), \tilde{\mathbf{w}}_{-1}\}$ to $\{\sqrt{\bar{\mu}(\cdot)}\; e_a(\cdot)\}$ provided we impose, for all i,

$$\left|1 - \frac{\mu(i)}{\bar{\mu}(i)}\right| < 1 \iff 0 < \mu(i) < 2\bar{\mu}(i).$$

More specifically, if we define

$$\Delta(N) = \max_{0 \leq i \leq N} \left|1 - \frac{\mu(i)}{\bar{\mu}(i)}\right|, \quad \gamma(N) = \max_{0 \leq i \leq N} \frac{\mu(i)}{\bar{\mu}(i)}$$

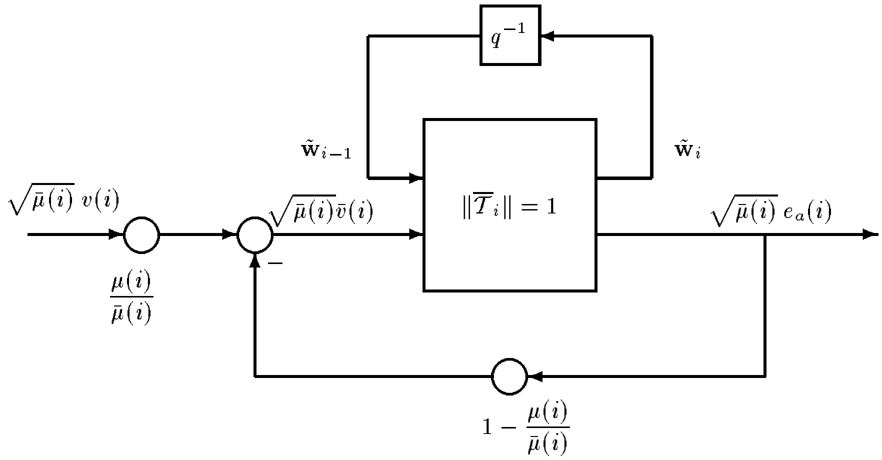

Figure 1: *A time-variant lossless mapping with time-variant gain feedback for stochastic gradient algorithms.*

Then it can be shown that the following results hold.

<u>THEOREM 1.</u> *Consider the gradient-recursion (1) and define $\Delta(N)$ and $\gamma(N)$ as above. If $0 < \mu(i) < 2\bar{\mu}(i)$, then the map from $\{\sqrt{\bar{\mu}(\cdot)}v(\cdot), \tilde{\mathbf{w}}_{-1}\}$ to $\{\sqrt{\bar{\mu}(\cdot)}e_a(\cdot)\}$ is l_2-stable in the following sense,*

$$\sqrt{\sum_{i=0}^{N} \bar{\mu}(i)\,|e_a(i)|^2} \leq \tag{7}$$

$$\frac{1}{1-\Delta(N)}\left[\|\tilde{\mathbf{w}}_{-1}\|_2 + \gamma(N)\sqrt{\sum_{i=0}^{N}\bar{\mu}(i)|v(i)|^2}\right]$$

Moreover, the map from $\{\sqrt{\mu(\cdot)}v(\cdot), \tilde{\mathbf{w}}_{-1}\}$ to $\{\sqrt{\mu(\cdot)}e_a(\cdot)\}$ (i.e., with $\bar{\mu}(\cdot)$ replaced by $\mu(\cdot)$) is also l_2-stable in the following sense:

$$\sqrt{\sum_{i=0}^{N} \mu(i)\,|e_a(i)|^2} \leq \tag{8}$$

$$\frac{\gamma^{1/2}(N)}{1-\Delta(N)}\left[\|\tilde{\mathbf{w}}_{-1}\|_2 + \gamma^{1/2}(N)\sqrt{\sum_{i=0}^{N}\mu(i)|v(i)|^2}\right]$$

In fact, a stronger upper bound than (8) can be given when $\mu(i)$ is further restricted to the interval $0 < \mu(i) \leq \bar{\mu}(i)$. This follows from the arguments after the local error-bounds of Sec. II.2, namely, if $0 < \mu(i) \leq \bar{\mu}(i)$ then

$$\sqrt{\sum_{i=0}^{N} \mu(i)\,|e_a(i)|^2} \leq \left[\ \|\tilde{\mathbf{w}}_{-1}\|_2 + \sqrt{\sum_{i=0}^{N}\mu(i)|v(i)|^2}\ \right]$$

The fact that the bound in (8) is valid even for $\mu(i)$ in the interval $\bar{\mu}(i) \leq \mu(i) < 2\bar{\mu}(i)$ suggests that a local bound, along the lines of those in Sec. II.2, should also exist for this interval. In fact, this is the case and it can be established that

$$1 \quad \leq \quad \frac{\|\tilde{\mathbf{w}}_i\|_2^2 + \mu(i)\,|e_a(i)|^2}{\|\tilde{\mathbf{w}}_{i-1}\|_2^2 + \mu(i)\,|v(i)|^2} \quad \leq \frac{\mu(i)}{2\bar{\mu}(i) - \mu(i)}$$

for $\bar{\mu}(i) \leq \mu(i) < 2\bar{\mu}(i)$.

II.5 A Convergence Result

In order to appreciate the significance of the global bounds of Theorem 1, assume that the normalized noise sequence $\{\sqrt{\mu(\cdot)}\ v(\cdot)\}$ has finite energy. It then follows from (8) that $\sum_{i=0}^{\infty}\mu(i)|e_a(i)|^2 < \infty$. We therefore conclude that $\{\sqrt{\mu(\cdot)}\ e_a(\cdot)\}$ is a Cauchy sequence and, hence, $\sqrt{\mu(i)}\ e_a(i)$ converges to zero.

For the special case of a constant step-size in (1), the condition $0 < \mu(i) < 2\bar{\mu}(i)$ that appears in the statement of Theorem 1 collapses to requiring $0 < \mu < 2\,[\inf_i\ \bar{\mu}(i)]$. This latter condition coincides with the standard one required in [4, Ch. 6] in the *noise free* case ($v(i) = 0$) in order to conclude the stability of the LMS recursion (1). Here, however, we have argued that the condition can still be employed in the *noisy* and also *time-variant* case, by showing how to explicitly incorporate the noise signal into the feedback structure and by studying the l_2-stability of the overall interconnection.

In fact, more physical insights into the convergence behaviour of the gradient recursion (1) can be obtained by studying the energy flow through the feedback configuration of Figure 1, as shown in the next section.

II.6 Energy Flow in the Feedback Structure

The feedback structure, and the associated lossless block in the direct path, provide a helpful physical picture for the energy flow through the system. To clarify this, let us ignore the measurement noise $v(i)$ and assume that we have noiseless measurements $d(i) = \mathbf{u}_i\mathbf{w}$. It is known that in a stochastic Gaussian setting, the maximal speed of convergence is obtained for $\mu(i) = \bar{\mu}(i)$, i.e., for the so-called projection LMS algorithm. We shall now argue that this conclusion is consistent with the feedback configuration of Figure 1.

Indeed, for $\mu(i) = \bar{\mu}(i)$, the feedback loop is disconnected. This means that there is no energy flowing back into the lower input of the lossless section from its lower output $e_a(\cdot)$. The losslessness of the feedforward path then implies that

$$E_w(i) = E_w(i-1) - E_e(i) \tag{9}$$

where we are denoting by $E_e(i)$ the energy of $\sqrt{\bar{\mu}(i)}\, e_a(i)$ and by $E_w(i)$ the energy of $\tilde{\mathbf{w}}_i$. Expression (9) implies that the weight-error energy is a non-increasing function of time, i.e., $E_w(i) \leq E_w(i-1)$ for all i.

But what if $\mu(i) \neq \bar{\mu}(i)$? In this case the feedback path is active and the convergence speed is affected since the rate of decrease in the energy of the weight-error vector is now lowered. Indeed, for $\mu(i) \neq \bar{\mu}(i)$ we obtain (compare with (9))

$$E_w(i) = E_w(i-1) - \underbrace{\left(1 - \left|1 - \frac{\mu(i)}{\bar{\mu}(i)}\right|^2\right)}_{\tau(i)} E_e(i).$$

It is easy to verify that as long as $\mu(i) \neq \bar{\mu}(i)$ we always have $0 < \tau(i) < 1$, which shows that the rate of decrease in the energy of $\tilde{\mathbf{w}}_i$ is lowered.

III. FILTERED-ERROR GRADIENT METHODS

We now move a step further and consider the class of filtered-error gradient algorithms. We shall show that the feedback loop concept of the former sections applies equally well to these variants, which employ filtered versions of the output estimation error, $\tilde{e}_a(i) = d(i) - \mathbf{u}_i\mathbf{w}_{i-1}$.

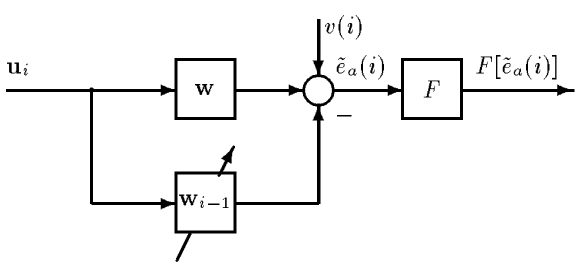

Figure 2: *Structure of filtered-error algorithms.*

Such algorithms are useful when the error $\tilde{e}_a(i)$ can not be observed directly, but rather a filtered version of it, as indicated in Figure 2. The operator F denotes the filter that operates on $\tilde{e}_a(i)$. It may be assumed to be a finite-impulse response filter of order M_F, say

$$F(q^{-1})[x(i)] = F[x(i)] = \sum_{j=0}^{M_F-1} f_j x(i-j)$$

It may also be a time-variant filter. A typical application where the need for such algorithms arises is in the active control of noise. In the sequel we shall discuss two important classes of algorithms that employ filtered error measurements; the so-called Modified filtered-x LMS (MFxLMS) and filtered error LMS (FELMS) [14,15,16].

III.1 The MFxLMS Algorithm

A common algorithm that is used to handle the filtered error case is the filtered-x LMS algorithm. It employs a recursive update of the form

$$\mathbf{w}_i = \mathbf{w}_{i-1} + \mu(i)F[\mathbf{u}_i]^* F[\tilde{e}_a(i)] \tag{10}$$

where the input data $\mathbf{u}_i$ is also processed by the filter F. If slow adaptation is assumed, i.e., if the variation in the weight estimates do not vary considerably over the length of the filter F, then we can approximate $F[\mathbf{u}_i\mathbf{w}_{i-1}]$ by $F[\mathbf{u}_i]\mathbf{w}_{i-1}$, which leads to the approximate update

$$\mathbf{w}_i = \mathbf{w}_{i-1} + \mu(i)F[\mathbf{u}_i]^* \left(F[d(i)] - F[\mathbf{u}_i]\mathbf{w}_{i-1}\right)$$

with $F[d(i)] = F[\mathbf{u}_i]\mathbf{w} + F[v(i)]$. This is of the same form as the LMS update (1) with the quantities $\{\mathbf{u}_i, d(i), v(i)\}$ replaced by their filtered versions $\{F[\mathbf{u}_i], F[d(i)], F[v(i)]\}$. In this case, the conclusions of the previous sections hold. For example, the stability condition now becomes approximately, $0 < \mu(i) < 2/\|F[\mathbf{u}_i]\|_2^2$.

Recently, an improvement has been proposed that avoids the slow adaptation assumption [15]. This is achieved by modifying the update expression as follows: $\mathbf{w}_i =$

$$\mathbf{w}_{i-1} + \mu(i)F[\mathbf{u}_i]^* \left(F[\tilde{e}_a(i)] + F[\mathbf{u}_i\mathbf{w}_{i-1}] - F[\mathbf{u}_i]\mathbf{w}_{i-1}\right)$$

which is equivalent to the update equation

$$\mathbf{w}_i = \mathbf{w}_{i-1} + \mu(i)F[\mathbf{u}_i]^* \left(\ F[v(i)] + F[\mathbf{u}_i]\tilde{\mathbf{w}}_{i-1}\ \right)$$

This is of the same form as the LMS update (1) but with the filtered input sequence $F[\mathbf{u}_i]$ and the filtered noise sequence $F[v(i)]$. This time, however, no approximation is employed. The results of the previous sections will then be immediately applicable with the proper change of variables.

III.2 The FELMS Algorithm

The so-called filtered-error LMS algorithm retains the input vector unchanged and uses

$$\mathbf{w}_i = \mathbf{w}_{i-1} + \mu(i)\mathbf{u}_i^* F[\tilde{e}_a(i)] \tag{11}$$

In contrast to the FxLMS algorithm, and its modified form, the error-path filter F does not need to be known explicitly, and the algorithm also requires less computation. Similar update forms also arise in the context of IIR modeling, such as Feintuch's algorithm [9,17] and the SHARF algorithm [6]. Following the discussion that led to (4), we get

$$\begin{aligned} \mathbf{w}_i &= \mathbf{w}_{i-1} + \bar{\mu}(i)\mathbf{u}_i^*[e_a(i) + \bar{v}(i)] \\ \bar{\mu}(i)\bar{v}(i) &= \mu(i)F[v(i)] - \bar{\mu}(i)e_a(i) + \mu(i)F[e_a(i)] \end{aligned}$$

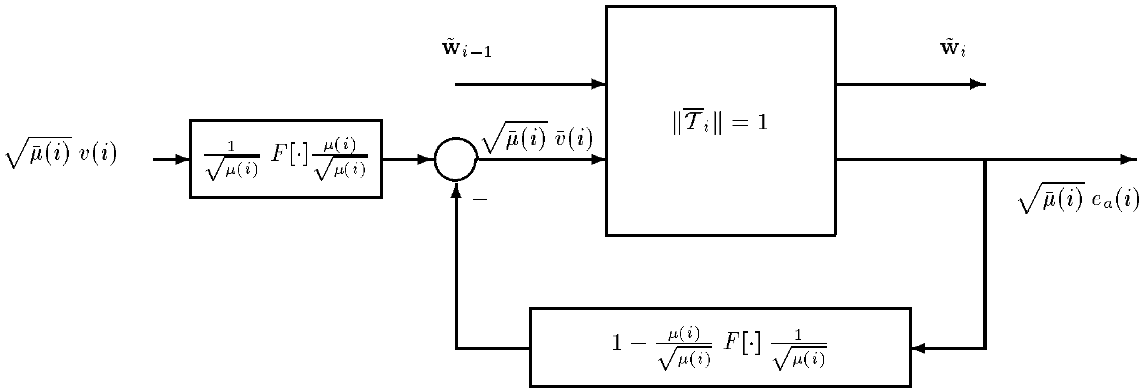

Figure 3: *Filtered-error LMS algorithm as a time-variant lossless mapping with dynamic feedback.*

and $\bar{\mu}(i) = 1/\|\mathbf{u}_i\|_2^2$. This is of the same form as the update discussed in Section II.3, which readily implies that the following relation always holds:

$$\frac{\|\tilde{\mathbf{w}}_i\|_2^2 + \bar{\mu}(i)\,|e_a(i)|^2}{\|\tilde{\mathbf{w}}_{i-1}\|_2^2 + \bar{\mu}(i)\,|\bar{v}(i)|^2} = 1 \tag{12}$$

This establishes that the map from $\{\tilde{\mathbf{w}}_{i-1}, \sqrt{\bar{\mu}(i)}\bar{v}(i)\}$ to $\{\tilde{\mathbf{w}}_i, \sqrt{\bar{\mu}(i)}e_a(i)\}$, denoted by $\overline{\mathcal{T}}_i$, is also *lossless*, and that the overall mapping from the original disturbance $\sqrt{\bar{\mu}(\cdot)}v(\cdot)$ to the resulting apriori estimation error $\sqrt{\bar{\mu}(\cdot)}e_a(\cdot)$ can be expressed in terms of a feedback structure, as shown in Figure 3.

The feedback loop now consists of a dynamic system. But we can still proceed to study the l_2-stability of the overall configuration in much the same way as we did in Section II.3. Similar arguments will easily lead to a sufficient condition for stability that we now exhibit. But first we introduce a compact matrix notation and define

$$\begin{aligned} \mathbf{M}_N &= \text{diag}\,\{\mu(0), \mu(1), \ldots, \mu(N)\} \\ \overline{\mathbf{M}}_N &= \text{diag}\,\{\bar{\mu}(0), \bar{\mu}(1), \ldots, \bar{\mu}(N)\} \end{aligned}$$

We also define the lower triangular matrix $\mathbf{F}_N$ that describes the action of the filter F on a sequence at its input. This is generally a band matrix since $M_F \ll N$, as shown below for the special case $M_F = 3$,

$$\mathbf{F}_N = \begin{bmatrix} f_0 & & & & \\ f_1 & f_0 & & & \\ f_2 & f_1 & f_0 & & \\ & f_2 & f_1 & f_0 & \\ & & \ddots & \ddots & \ddots \end{bmatrix}$$

It is then immediate to verify that the interconnection is l_2-stable provided the matrix $\mathbf{G}_N$ defined by

$$\mathbf{G}_N \triangleq \left(\mathbf{I} - \overline{\mathbf{M}}_N^{-\frac{1}{2}}\mathbf{M}_N\mathbf{F}_N\overline{\mathbf{M}}_N^{-\frac{1}{2}}\right)$$

is strictly contractive. The matrix $\mathbf{G}_N$ can be easily seen to have the following triangular form (it also has a band of width M_F): $\mathbf{G}_N =$

$$\begin{pmatrix} 1 - \frac{\mu(0)}{\bar{\mu}(0)}f_0 & & \mathbf{O} \\ -\frac{\mu(1)}{\sqrt{\bar{\mu}(0)\bar{\mu}(1)}}f_1 & 1 - \frac{\mu(1)}{\bar{\mu}(1)}f_0 & \\ -\frac{\mu(2)}{\sqrt{\bar{\mu}(0)\bar{\mu}(2)}}f_2 & -\frac{\mu(2)}{\sqrt{\bar{\mu}(1)\bar{\mu}(2)}}f_1 & 1 - \frac{\mu(2)}{\bar{\mu}(2)}f_0 \\ \vdots & & & \ddots \end{pmatrix}$$

Several special cases may be of interest. For example, the special case $F = 1$ (i.e., no filter) immediately leads to the case we encountered earlier in Section II.3. Another special case is $F = q^{-1}$ (i.e., a simple delay). The filtered-error LMS recursion (11) then collapses to the so-called delayed-error LMS. The corresponding $\mathbf{G}_N$ matrix can not be a strict contraction since its $(1,1)$ entry will be equal to 1. This is consistent with results in the literature where it has been observed that the delayed-error LMS algorithm usually leads to unstable behaviour. We also see from the general expression for $\mathbf{G}_N$ that a simple gain filter $F = f_0$ with a *negative* f_0 leads to a non-contractive $\mathbf{G}_N$.

III.3 The Projection FELMS Algorithm

We focus now on an important choice for the step-size parameter, viz., $\mu(i) = \alpha\ \bar{\mu}(i)$ with $\alpha > 0$. In this case, it can be seen that the contractivity requirement now collapses to

$$\left\|\mathbf{I} - \alpha\overline{\mathbf{M}}_N^{\frac{1}{2}}\mathbf{F}_N\overline{\mathbf{M}}_N^{-\frac{1}{2}}\right\|_{2,ind} < 1$$

with

$$\mathbf{G}_N = \begin{pmatrix} 1 - \alpha f_0 & & \mathbf{O} \\ -\alpha\frac{\sqrt{\bar{\mu}(1)}}{\sqrt{\bar{\mu}(0)}}f_1 & 1 - \alpha f_0 & \\ -\alpha\frac{\sqrt{\bar{\mu}(2)}}{\sqrt{\bar{\mu}(0)}}f_2 & -\alpha\frac{\sqrt{\bar{\mu}(2)}}{\sqrt{\bar{\mu}(1)}}f_1 & 1 - \alpha f_0 \\ \vdots & & & \ddots \end{pmatrix}$$

We shall further assume that the energy of the input sequence $\mathbf{u}_i$ does not change very rapidly over the filter length

M_F, i.e., $\bar{\mu}(i) \approx \bar{\mu}(i-1) \approx \ldots \approx \bar{\mu}(i-M_F)$. This is a reasonable assumption since, as mentioned earlier, we often have $M_F \ll M$. In this case, $\mathbf{G}_N$ collapses to $\mathbf{G}_N \approx \mathbf{I} - \alpha\mathbf{F}_N$. It is now easy to see that the contractivity of $(\mathbf{I} - \alpha\mathbf{F}_N)$ can be guaranteed if we choose the α so as to satisfy

$$\max_{\Omega} \left|1 - \alpha F(e^{j\Omega})\right| < 1 \qquad (13)$$

This also suggests that, for faster convergence (i.e., for smaller feedback gain), we may choose α optimally by solving the min-max problem:

$$\min_{\alpha} \max_{\Omega} \left|1 - \alpha F(e^{j\Omega})\right| \qquad (14)$$

If the resulting minimum is less than 1 then the corresponding optimum α will result in faster convergence. Simulations results have confirmed this conclusion. But we omit the details here for brevity.

IV. GAUSS-NEWTON METHODS

We now discuss the so-called Gauss-Newton recursive method, which updates the weight-estimate according to the following relation

$$\mathbf{w}_i = \mathbf{w}_{i-1} + \mu(i)\,\mathbf{P}_i\mathbf{u}_i^*\,\left(d(i) - \mathbf{u}_i\mathbf{w}_{i-1}\right) \qquad (15)$$

where $\mathbf{P}_i$ satisfies the Riccati equation update, with initial condition Π_0,

$$\mathbf{P}_i = \frac{1}{\lambda(i)}\left(\mathbf{P}_{i-1} - \frac{\mathbf{P}_{i-1}\mathbf{u}_i^*\mathbf{u}_i\mathbf{P}_{i-1}}{\frac{\lambda(i)}{\beta(i)} + \mathbf{u}_i\mathbf{P}_{i-1}\mathbf{u}_i^*}\right) \qquad (16)$$

and $\{\lambda(i), \mu(i), \beta(i)\}$ are given positive scalar time-variant coefficients, with $\lambda(i) \le 1$. An important special case of (15) is the so-called Recursive-Least-Squares (RLS) algorithm [18,19], which corresponds to the choices $\beta(i) = \mu(i) = 1$ and $\lambda(i) = \lambda = cte$.

It further follows from the update equation (15) that

$$\tilde{\mathbf{w}}_i = \tilde{\mathbf{w}}_{i-1} - \mu(i)\mathbf{P}_i\mathbf{u}_i^*\tilde{e}_a(i) \qquad (17)$$

If we multiply (17) by $\mathbf{u}_i$ from the left we also obtain the following relation between $e_p(i)$, $e_a(i)$, and $v(i)$,

$$e_p(i) = [\,1 - \mu(i)\mathbf{u}_i\mathbf{P}_i\mathbf{u}_i^*\,]\;e_a(i) - \mu(i)\mathbf{u}_i\mathbf{P}_i\mathbf{u}_i^*\;v(i)$$

The same line of reasoning that we followed in Sec.II.2 can be repeated here to show that

$$\frac{\tilde{\mathbf{w}}_i^*\mathbf{P}_i^{-1}\tilde{\mathbf{w}}_i + (\mu(i) - \beta(i))\,|e_a(i)|^2}{\lambda(i)\tilde{\mathbf{w}}_{i-1}^*\mathbf{P}_{i-1}^{-1}\tilde{\mathbf{w}}_{i-1} + \mu(i)\,|v(i)|^2}\left\{\begin{array}{ll} \le 1 & \mu(i) < \bar{\mu}(i) \\ = 1 & \mu(i) = \bar{\mu}(i) \\ \ge 1 & \mu(i) > \bar{\mu}(i) \end{array}\right.$$

Define $\epsilon_N(\mathbf{w}_{-1}, v(\cdot)) \triangleq$

$$\left\{\tilde{\mathbf{w}}_N^*\mathbf{P}_N^{-1}\tilde{\mathbf{w}}_N + \sum_{i=0}^{N}(\mu(i) - \beta(i))\lambda^{[i+1,N]}\,|e_a(i)|^2\right\} -$$

$$\left\{\lambda^{[0,N]}\tilde{\mathbf{w}}_{-1}^*\mathbf{P}_{-1}^{-1}\tilde{\mathbf{w}}_{-1} + \sum_{i=0}^{N}\mu(i)\lambda^{[i+1,N]}\,|v(i)|^2\right\}$$

where we have employed the notation $\lambda^{[i,j]} = \prod_{k=i}^{j}\lambda(k)$. If $\mu(i) \le \bar{\mu}(i)$ over $0 \le i \le N$, then the first two local bounds allow us to conclude that the Gauss-Newton algorithm (15) always guarantees $\epsilon_N(\mathbf{w}_{-1}, v(\cdot)) \le 0$ for any $\mathbf{w}_{-1}$ and $v(\cdot)$. If we further have $\beta(i) \le \mu(i) \le \bar{\mu}(i)$, then this also establishes the existence of a contraction mapping from

$$\left\{\sqrt{\mu(\cdot)}\sqrt{\lambda^{[\cdot+1,N]}}v(\cdot),\ \sqrt{\lambda^{[0,N]}}\Pi_0^{-1/2}\tilde{\mathbf{w}}_{-1}\right\}$$

to $\left\{\sqrt{(\mu(\cdot) - \beta(\cdot))}\sqrt{\lambda^{[\cdot+1,N]}}e_a(\cdot), \mathbf{P}_N^{-1/2}\tilde{\mathbf{w}}_N\right\}$.

It can also be verified that $\bar{\mu}(i) > \beta(i)$ for all i. Now, by following an argument similar to the one presented in Sec. II.3 we can verify that it always holds, <u>for all</u> $\mu(i)$ and $v(i)$, that

$$\frac{\tilde{\mathbf{w}}_i^*\mathbf{P}_i^{-1}\tilde{\mathbf{w}}_i + (\bar{\mu}(i) - \beta(i))\,|e_a(i)|^2}{\lambda(i)\tilde{\mathbf{w}}_{i-1}^*\mathbf{P}_{i-1}^{-1}\tilde{\mathbf{w}}_{i-1} + \bar{\mu}(i)\,|e_p(i)|^2} = 1 \qquad (18)$$

with $e_p(i) = -\bar{v}(i)$ and

$$\bar{\mu}^{\frac{1}{2}}(i)\bar{v}(i) = \frac{\mu(i)}{\bar{\mu}^{\frac{1}{2}}(i)}v(i) - \left(1 - \frac{\mu(i)}{\bar{\mu}(i)}\right)\bar{\mu}^{\frac{1}{2}}(i)e_a(i)$$

Hence, the map from

$$\{\sqrt{\lambda(i)}\mathbf{P}_{i-1}^{-\frac{1}{2}}\tilde{\mathbf{w}}_{i-1}, \sqrt{\bar{\mu}(i)}\bar{v}(i)\}$$

to

$$\{\mathbf{P}_i^{-\frac{1}{2}}\tilde{\mathbf{w}}_i, \sqrt{\bar{\mu}(i) - \beta(i)}e_a(i)\},$$

denoted by $\overline{\mathcal{T}}_i$, is always *lossless*, i.e., it preserves energy. The overall mapping from the *original* disturbance $\sqrt{\bar{\mu}(\cdot)}v(\cdot)$ to the resulting apriori estimation error $\sqrt{\bar{\mu}(\cdot)}e_a(\cdot)$ can then be expressed in terms of a feedback structure as shown in Figure 4.

The feedback configuration of Figure 4 also lends itself to analysis via the small gain theorem, along the same lines of Sec. II.4. If we define

$$\Delta(N) = \max_{0 \le i \le N}\left|\frac{1 - \frac{\mu(i)}{\bar{\mu}(i)}}{\sqrt{1 - \frac{\beta(i)}{\bar{\mu}(i)}}}\right|$$

then the small-gain condition would require that we impose $(1 - \Delta(N)) > 0$, which is equivalent to

$$0 < \mu(i) < \bar{\mu}(i)\left(1 + \sqrt{1 - \frac{\beta(i)}{\bar{\mu}(i)}}\right) \qquad (19)$$

Under this condition, we can verify that the map from

$$\{\sqrt{\lambda^{[\cdot+1,N]}\bar{\mu}(\cdot)}\;v(\cdot),\ \sqrt{\lambda^{[0,N]}}\mathbf{P}_{-1}^{-\frac{1}{2}}\tilde{\mathbf{w}}_{-1}\}$$

to

$$\{\sqrt{\lambda^{[\cdot+1,N]}(\bar{\mu}(\cdot) - \beta(i))}\;e_a(\cdot)\}$$

is l_2-stable. Moreover, if $\beta(i) \le \mu(i)$ then it also holds that the map from

$$\{\sqrt{\lambda^{[\cdot+1,N]}\mu(\cdot)}\;v(\cdot),\ \sqrt{\lambda^{[0,N]}}\mathbf{P}_{-1}^{-\frac{1}{2}}\tilde{\mathbf{w}}_{-1}\}$$

to

$$\{\sqrt{\lambda^{[\cdot+1,N]}\mu(\cdot) - \beta(\cdot)}\;e_a(\cdot)\}$$

(i.e., with $\bar{\mu}(\cdot)$ replaced by $\mu(\cdot)$) is l_2-stable.

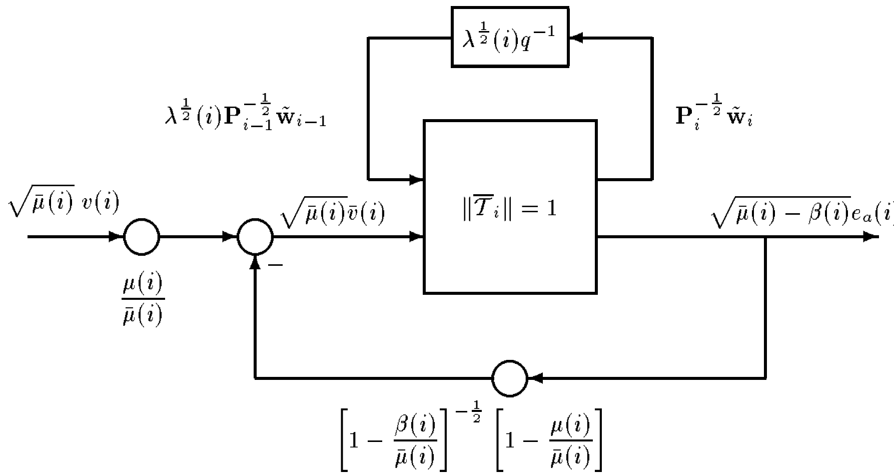

Figure 4: *A time-variant lossless mapping with time-variant gain feedback for Gauss-Newton methods.*

V. CONCLUDING REMARKS

More variants of the Gauss-Newton type update occur if the underlying model is not a transversal model but rather an IIR model. In this case, the update recursion is of filtered-error type, and leads to a dynamic feedback – see, e.g., [9,17] for more details and for connections with results in $H^\infty-$theory.

REFERENCES

[1] I. D. Landau. *Adaptive Control: The Model Reference Approach.* Control and Systems Theory. Marcel Dekker, Inc., NY, 1979.

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

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

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

[5] V. M. Popov. *Hyperstability of Control Systems.* Springer-Verlag, NY, 1973.

[6] C. R. Johnson. Adaptive IIR filtering: Current results and open issues. *IEEE Trans. Inform. Theory*, 30(2):237–250, March 1984.

[7] H. J. Kushner. *Weak Convergence Methods and Singularly Perturbed Stochastic Control and Filtering Problems.* Birkhäuser, Boston, MA, 1990.

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

[9] A. H. Sayed and M. Rupp. A class of nonlinear adaptive H^∞-filters with guaranteed l_2-stability. *Proc. IFAC Symposium on Nonlinear Control System Design*, June 1995.

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

[11] I. D. Landau. A feedback system approach to adaptive filtering. *IEEE Transactions on Information Theory*, 30(2), March 1984.

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

[13] M. Vidyasagar. *Nonlinear Systems Analysis.* Prentice Hall, New Jersey, second edition, 1993.

[14] Jr. J. R. Glover. Adaptive noise canceling applied to sinusoidal interferences. *IEEE Trans. Acoust., Speech, Signal Processing*, 25(6):484–491, December 1977.

[15] I. Kim, H. Na, K. Kim, and Y. Park. Constraint filtered-x and filtered-u algorithms for the active control of noise in a duct. *J. Acoust. Soc. Am.*, 95(6):3397–3389, June 1994.

[16] B. Widrow and S. D. Stearns. *Adaptive Signal Processing.* Prentice-Hall, Inc., Englewood Cliffs, NJ, 1985.

[17] M. Rupp and A. H. Sayed. On the stability and convergence of Feintuch's algorithm for adaptive IIR filtering. *Proc. IEEE ICASSP*, vol. 2, pp. 1388–1391, MI, May 1995.

[18] S. Haykin. *Adaptive Filter Theory.* Prentice Hall, Englewood Cliffs, NJ, second edition, 1991.

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