Proc. SPIE Conf. on Advanced Signal Processing, vol. 2563, pp. 458–469, San Diego, July 1995.

A time-domain feedback analysis of adaptive gradient algorithms via the Small Gain Theorem*

ALI H. SAYED AND MARKUS RUPP

Department of Electrical and Computer Engineering
University of California
Santa Barbara, CA 93106.

ABSTRACT

This paper provides a time-domain feedback analysis of gradient-based 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.

Keywords: Adaptive gradient filters, filtered-error gradient algorithms, feedback connection, l_2-stability, the small gain theorem, contraction mapping, error bounds.

1 INTRODUCTION

Gradient-based identification schemes have become a standard tool in a wide range of applications in signal processing and control. Their simplicity and robustness have led to an increasing interest in the analysis of their stability and convergence properties. This paper suggests a time-domain approach that proves to be useful in the analysis and design of gradient-based estimators. It highlights and exploits an intrinsic feedback structure that can be associated with such schemes, mapping the noise sequence and the initial weight guess to the apriori estimation errors and the final weight estimate.

Although the feedback nature of these, and related recursive schemes, has been pointed out and advantageously exploited in earlier places in the the literature,[1–3] 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[4]), but is instead motivated by 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 the derivation here is not only interested in

*This material is based on work supported by the National Science Foundation under Award No. MIP-9409319. Any opinions, findings, and conclusions or recommendations expressed in this publication are those of the author(s) and do not necessarily reflect the views of the National Science Foundation. The work of M. Rupp is also supported by a scholarship from DAAD (German Academic Exchange Service) as well as the scientifical division of NATO (Wissenschaftsausschuß der NATO).

the analysis of the stability properties of the gradient-based schemes but also in a formal study of their robustness properties in the presence of uncertain disturbances. This is especially useful, for example, when the statistical properties of the disturbances are unknown.

In this regard, the feedback interconnection studied 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. Also, the blocks in *both* the feedforward and the feedback paths are allowed to be, and in fact are, *time-variant*. It is then shown that the feedback configuration lends to stability analysis via a so-called small gain theorem in system theory (e.g.,[5,6]).

An interesting fallout of the time-domain approach of this paper is that it can also be regarded as an extension of the so-called transfer function approach that is often used in the analysis of gradient recursions. The time-domain analysis is shown to avoid the restrictions and limitations that are usually imposed in the transfer-function description.

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). The symbol $\mathbf{I}$ will denote the identity matrix of appropriate dimensions, and the boldface letter $\mathbf{0}$ will either denote a zero vector or a zero matrix. Finally, the notation $\|\mathbf{x}\|_2^2$ will denote the squared Euclidean norm of a vector.

2 THE LEAST-MEAN-SQUARES ALGORITHM

One of the most widely used adaptive algorithms is the least-mean-squares (LMS) algorithm.[7] It starts with an initial guess $\mathbf{w}_{-1}$, for an unknown $M \times 1$ weight vector $\mathbf{w}$, and updates it via an update equation of the form

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

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

The difference $[d(i) - \mathbf{u}_i\mathbf{w}_{i-1}]$ will be denoted by $\tilde{e}_a(i)$ and will be referred to as the *output estimation error*. The following error measures will also be useful for our later analysis: $\tilde{\mathbf{w}}_i$ will denote the difference between the true weight $\mathbf{w}$ and its estimate $\mathbf{w}_i$, $\tilde{\mathbf{w}}_i \triangleq \mathbf{w} - \mathbf{w}_i$, and $e_a(i)$ will denote the *apriori estimation error*, $e_a(i) \triangleq \mathbf{u}_i\tilde{\mathbf{w}}_{i-1}$. It follows from the update equation (1) that the weight-error vector $\tilde{\mathbf{w}}_{i-1}$ satisfies the recursive equation:

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

It is also straightforward to verify that the apriori estimation error, $e_a(i)$, and the output estimation error, $\tilde{e}_a(i)$, differ by the disturbance $v(i)$, i.e., $\tilde{e}_a(i) = e_a(i) + v(i)$. We further define the *aposteriori estimation error*, $e_p(i) \triangleq \mathbf{u}_i\tilde{\mathbf{w}}_i$, and note that if we multiply (2) by $\mathbf{u}_i$ from the left we obtain the following relation (used later in (9)) between $e_p(i)$, $e_a(i)$, and $v(i)$,

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

2.1 A Transfer-Function Description of the LMS Algorithm

Before proceeding to the time-domain analysis of this paper, we shall first review a well-known approach to the analysis of LMS-type recursions that employs the concept of transfer functions.[8,9] In this method, the input vector $\mathbf{u}_i$ is assumed to have a shift structure, $\mathbf{u}_i = [u(i), u(i-1), \ldots, u(i-M+1)]$, and the individual entries are further assumed to arise from a sinusoidal excitation, say $u(i) = C\cos(\Omega\, i)$. Assuming a constant step-size μ and neglecting the initial condition $\tilde{\mathbf{w}}_{-1}$, the transfer function from the disturbance $v(i)$ to the apriori estimation error $e_a(i)$ can be shown to be approximately given by the following expression (in terms of z-transforms):

$$\frac{E_a(z)}{V(z)} \approx \frac{\frac{\mu C^2 M}{2}\left[1 - z\cos(\Omega)\right]}{z^2 - 2z\cos(\Omega)\left(1 - \frac{\mu C^2 M}{4}\right) + \left(1 - \frac{\mu C^2 M}{2}\right)} \; . \tag{4}$$

Several limitations and approximations are involved while establishing (4). These are avoided in the time-domain analysis of this paper. But for now, we wish to highlight the fact that an interesting feedback structure is implied by (4). To clarify this, we introduce the parameter $\bar{\mu} = \frac{2}{C^2 M}$, and define the normalized noise signal

$$\overline{V}(z) \triangleq \frac{\mu}{\bar{\mu}} V(z) - \left(1 - \frac{\mu}{\bar{\mu}}\right) E_a(z) \; . \tag{5}$$

It then follows that the transfer function relating $\overline{V}(z)$ and $E_a(z)$ is equal to

$$\frac{E_a(z)}{\overline{V}(z)} = \frac{z^{-1} - \cos(\Omega)}{z - \cos(\Omega)} \; , \tag{6}$$

which is an *all-pass* filter. Therefore, the transfer function (4) from $v(\cdot)$ to $e_a(\cdot)$ can be expressed as a feedback structure with an all-pass filter in the forward path and a constant gain in the feedback loop. This is depicted in Figure 1. The feedback gain is $(1 - \mu/\bar{\mu})$, which is thus equal to zero if we choose $\mu = \bar{\mu}$. This is known to be the choice that results in the highest convergence speed.

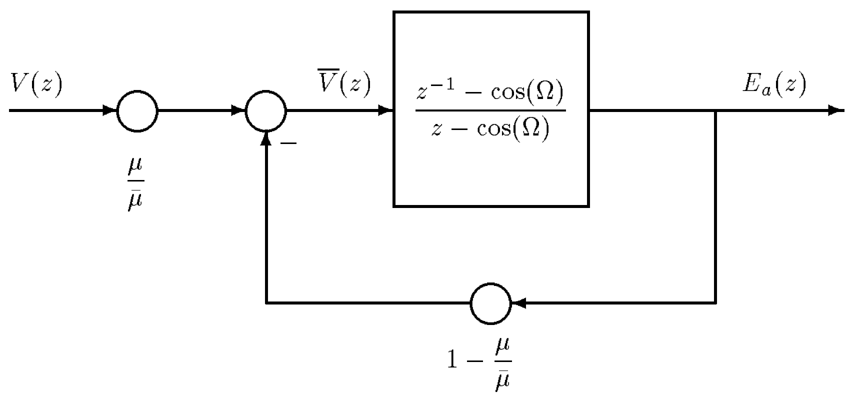

Figure 1: *A transfer function description for LMS: an all-pass mapping with gain feedback.*

The transfer function description suffers from limitations that hinders its applicability to more general scenarios. In particular, the argument explicitly assumes that the input vectors exhibit shift structure, which restricts the analysis to transversal filter structures. Also, the input sequence is further assumed to be sinusoidal, and the effect of initial conditions is ignored. Moreover, a constant, rather than a time-variant, step-size is assumed, and some nonlinear mixing terms need to be neglected. One of the contributions of this paper is to remove these limitations by employing an exact time-domain argument.

3 A TIME-DOMAIN ANALYSIS

The analysis that follows highlights an important feedback structure that is implied by gradient-type recursions of the form (1). For this purpose, we invoke (2) and compute the squared norm (i.e., energies) of both of its sides,

$$\|\tilde{\mathbf{w}}_i\|_2^2 \;=\; \|\tilde{\mathbf{w}}_{i-1}\|^2 - \mu(i)e_a(i)\tilde{e}_a^*(i) - \mu(i)e_a^*(i)\tilde{e}_a(i) + \mu^2(i)\|\mathbf{u}_i\|_2^2\,|\tilde{e}_a(i)|^2 \; .$$

If we replace $\tilde{e}_a(i)$ by $\tilde{e}_a(i) = e_a(i) + v(i)$ and use the fact that

$$|\tilde{e}_a(i)|^2 = |e_a(i) + v(i)|^2 = e_a(i)v^*(i) + v(i)e_a^*(i) + |e_a(i)|^2 + |v(i)|^2 \ ,$$

we conclude that the following equality always holds,

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

The following bounds are then always satisfied, where we have defined $\bar{\mu}^{-1}(i) = ||\mathbf{u}_i||_2^2$.

Lemma 3.1. *Consider the gradient recursion (1). It always holds that*

$$\frac{||\tilde{\mathbf{w}}_i||_2^2 + \mu(i)\,|e_a(i)|^2}{||\tilde{\mathbf{w}}_{i-1}||_2^2 + \mu(i)\,|v(i)|^2} \quad \begin{cases} \leq 1 & \text{for} \quad 0 < \mu(i) < \bar{\mu}(i), \\ = 1 & \text{for} \quad \mu(i) = \bar{\mu}(i), \\ \geq 1 & \text{for} \quad \mu(i) > \bar{\mu}(i). \end{cases} \tag{7}$$

In particular, the first two cases have an interesting interpretation that was exploited in[10] in order to provide a convergence and minimax analysis of gradient recursions of the form (1). They establish a local error-energy bound (or passivity relation) that also explains the *robustness nature* of such recursions: 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., $||\mathbf{w} - \mathbf{w}_i||_2^2 + \mu(i)|e_a(i)|^2$, will always be smaller than or equal to the sum of the energies of the starting errors (or disturbances), $||\mathbf{w} - \mathbf{w}_{i-1}||_2^2 + \mu(i)|v(i)|^2$. 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 the desired passivity relation over the interval $0 \leq i \leq N$. We may also add that other similar local, and global, passivity relations can be established by using aposteriori (rather than apriori) estimation errors.[10] But we shall forgo the details here and focus instead on the time-domain and feedback analysis.

3.1 The Feedback Structure

The bounds in (7) can be described via an alternative form that will lead us to an interesting feedback structure. The structure will be shown to constitute the proper extension of the transfer function description of Figure 1 to the general time-variant scenario. To clarify this, we first show how to rewrite the gradient recursion (1) in an alternative convenient form. We rewrite (3) as

$$e_p(i) = \left(1 - \frac{\mu(i)}{\bar{\mu}(i)}\right) e_a(i) - \frac{\mu(i)}{\bar{\mu}(i)} v(i), \tag{8}$$

and use it to re-express the update equation (1) in the following form:

$$\begin{aligned} \mathbf{w}_i &= \mathbf{w}_{i-1} + \mu(i)\mathbf{u}_i^* e_a(i) + \mu(i)\mathbf{u}_i^* v(i) \\ &= \mathbf{w}_{i-1} + \bar{\mu}(i)\mathbf{u}_i^* e_a(i) + \mathbf{u}_i^*[\mu(i)v(i) - (\bar{\mu}(i) - \mu(i))e_a(i)] \\ &= \mathbf{w}_{i-1} + \bar{\mu}(i)\mathbf{u}_i^*[e_a(i) - e_p(i)] = \mathbf{w}_{i-1} + \bar{\mu}(i)\mathbf{u}_i^*[e_a(i) + \bar{v}(i)]. \end{aligned} \tag{9}$$

This shows that the weight-update equation (1) 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 prior to (7), we readily conclude that the following equality holds <u>for all</u> $\{\mu(i), v(i)\}$,

$$\frac{||\tilde{\mathbf{w}}_i||_2^2 + \bar{\mu}(i)\,|e_a(i)|^2}{||\tilde{\mathbf{w}}_{i-1}||_2^2 + \bar{\mu}(i)\,|e_p(i)|^2} = 1\,. \tag{10}$$

This relation 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 2. The feedback loop consists of a gain factor that is equal to $(1-\mu(i)/\bar{\mu}(i))$. Also,

$$\bar{\mu}^{\frac{1}{2}}(i)\bar{v}(i) = \frac{\mu(i)}{\bar{\mu}^{\frac{1}{2}}(i)}v(i) - \left(1 - \frac{\mu(i)}{\bar{\mu}(i)}\right)\bar{\mu}^{\frac{1}{2}}(i)e_a(i)\,. \tag{11}$$

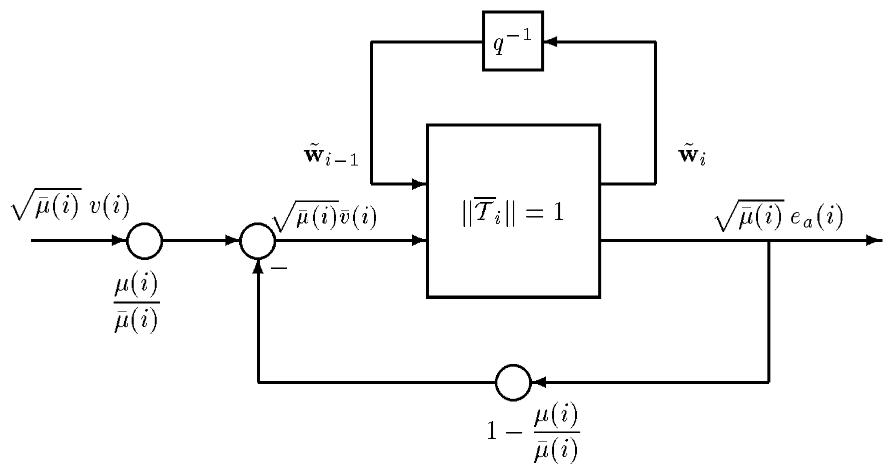

Figure 2: *A time-variant lossless mapping with gain feedback for gradient algorithms.*

It is easily seen that the lossless mapping $\overline{\mathcal{T}}_i$ is given by the following relation:

$$\begin{bmatrix} \sqrt{\bar{\mu}(i)}\ e_a(i) \\ \tilde{\mathbf{w}}_i \end{bmatrix} = \begin{bmatrix} \sqrt{\bar{\mu}(i)}\ \mathbf{u}_i & 0 \\ \mathbf{I} - \bar{\mu}(i)\mathbf{u}_i^*\mathbf{u}_i & -\sqrt{\bar{\mu}(i)}\mathbf{u}_i^* \end{bmatrix} \begin{bmatrix} \tilde{\mathbf{w}}_{i-1} \\ \sqrt{\bar{\mu}(i)}\ \bar{v}(i) \end{bmatrix}.$$

The similarity between Figure 1 and Figure 2 is obvious. However, it should be stressed that the analysis that led to Figure 2 is *exact* and valid in a general time-variant setting. No approximations or assumptions were made on the data.

3.2 l_2-Stability and the Small Gain Theorem

The feedback configuration of Figure 2 lends itself rather immediately to stability analysis, as we now explain. It follows from the equality (10) that for every time instant i, and for any $\mu(i)$, we have

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

This will allow us to conclude that the system in Figure 2 is l_2-stable, i.e., it maps a bounded energy sequence $\{\sqrt{\bar{\mu}(\cdot)}\ v(\cdot)\}$ to a bounded energy sequence $\{\sqrt{\bar{\mu}(\cdot)}\ e_a(\cdot)\}$ in a sense precised in (16) below. In fact, we shall also conclude that the same result holds even if we replace $\bar{\mu}(\cdot)$ by $\mu(\cdot)$.

For this purpose, assume we run the gradient recursion (1) from time $i=0$ up to time N. If we compute the sum of both sides of the above equality (12) we obtain,

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

which also implies that (by ignoring the term $\|\tilde{\mathbf{w}}_N\|_2^2$)

$$\sqrt{\sum_{i=0}^{N} \bar{\mu}(i)\,|e_a(i)|^2} \;\le\; \|\tilde{\mathbf{w}}_{-1}\|_2 + \sqrt{\sum_{i=0}^{N} \bar{\mu}(i)\,|\bar{v}(i)|^2}. \tag{13}$$

But it follows from (11), and from the triangular inequality for norms, that

$$\sqrt{\sum_{i=0}^{N} \bar{\mu}(i)\,|\bar{v}(i)|^2} \;\le\; \sqrt{\sum_{i=0}^{N} \frac{\mu^2(i)}{\bar{\mu}(i)}|v(i)|^2} \;+\; \sqrt{\sum_{i=0}^{N} \left|1-\frac{\mu(i)}{\bar{\mu}(i)}\right|^2 \bar{\mu}(i)|e_a(i)|^2}\;.$$

We thus conclude that

$$\sqrt{\sum_{i=0}^{N} \bar{\mu}(i)\,|e_a(i)|^2} \;\le\; \|\tilde{\mathbf{w}}_{-1}\|_2 + \sqrt{\sum_{i=0}^{N} \frac{\mu^2(i)}{\bar{\mu}(i)}|v(i)|^2} + \sqrt{\sum_{i=0}^{N} \left|1-\frac{\mu(i)}{\bar{\mu}(i)}\right|^2 \bar{\mu}(i)|e_a(i)|^2}\;. \tag{14}$$

Define

$$\Delta(N) = \max_{0\le i\le N} \left|1-\frac{\mu(i)}{\bar{\mu}(i)}\right| \quad \text{and} \quad \gamma(N) = \max_{0\le i\le N} \frac{\mu(i)}{\bar{\mu}(i)}\;. \tag{15}$$

It then follows that

$$\sqrt{\sum_{i=0}^{N} \bar{\mu}(i)\,|e_a(i)|^2} \;\le\; \|\tilde{\mathbf{w}}_{-1}\|_2 + \gamma(N)\sqrt{\sum_{i=0}^{N} \bar{\mu}(i)|v(i)|^2} + \Delta(N)\sqrt{\sum_{i=0}^{N} \bar{\mu}(i)|e_a(i)|^2}\;.$$

If $(1-\Delta(N)) > 0$ we conclude from the last inequality that

$$\sqrt{\sum_{i=0}^{N} \bar{\mu}(i)\,|e_a(i)|^2} \;\le\; \frac{1}{1-\Delta(N)}\left[\;\|\tilde{\mathbf{w}}_{-1}\|_2 + \gamma(N)\sqrt{\sum_{i=0}^{N} \bar{\mu}(i)|v(i)|^2}\;\right], \tag{16}$$

which establishes the l_2-stability of the system.

The condition $(1-\Delta(N)) > 0$ is equivalent to requiring $\Delta(N) < 1$. This can be viewed as a manifestation of the so-called small gain theorem in system analysis.[5,6] In simple terms, the theorem states that the l_2-stability of a feedback configuration (that includes Figure 2 as a special case) requires that the product of the norms of the feedforward and the feedback operators be strictly bounded by one. Here, the feedforward map has (2−induced) norm equal to one (due to its losslessness) while the 2−induced norm of the feedback map is $\Delta(N)$. Note also that for $\Delta(N) < 1$ we clearly need that $0 < \mu(i) < 2\bar{\mu}(i)$. This leads to the following conclusion.

THEOREM 3.2. *Consider the gradient-recursion (1) and define* $\Delta(N)$ *and* $\gamma(N)$ *as in (15). 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 sense of (16). 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:*

$$\sqrt{\sum_{i=0}^{N} \mu(i)\,|e_a(i)|^2} \;\le\; \frac{\gamma^{1/2}(N)}{1-\Delta(N)}\left[\;\|\tilde{\mathbf{w}}_{-1}\|_2 + \gamma^{1/2}(N)\sqrt{\sum_{i=0}^{N} \mu(i)|v(i)|^2}\;\right]. \tag{17}$$

Proof. The proof of the first bound (16) was provided prior to the statement of the theorem. As for the second bound, we first note that (14) implies

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

Now (17) follows by noting that

$$\sum_{i=0}^{N} \mu(i)|e_a(i)|^2 = \sum_{i=0}^{N} \frac{\mu(i)}{\bar{\mu}(i)}\; \bar{\mu}(i)|e_a(i)|^2 \leq \gamma(N) \sum_{i=0}^{N} \bar{\mu}(i)|e_a(i)|^2.$$

∎

In fact, a stronger upper bound than (17) 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 Lemma 3.1:

$$\sqrt{\sum_{i=0}^{N} \mu(i)\, |e_a(i)|^2} \;\leq\; \left[\; ||\tilde{\mathbf{w}}_{-1}||_2 + \sqrt{\sum_{i=0}^{N} \mu(i)|v(i)|^2} \;\right]. \tag{18}$$

The fact that the bound in (17) 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 (7), should also exist for this interval. In fact, this is also the case, as stated below. The proof is omitted for brevity.

LEMMA 3.3. *Consider the gradient recursion (1). Then the following bounds always hold:*

$$0 \;\leq\; \frac{||\tilde{\mathbf{w}}_i||_2^2 + \mu(i)\,|e_a(i)|^2}{||\tilde{\mathbf{w}}_{i-1}||_2^2 + \mu(i)\,|v(i)|^2} \leq 1 \qquad \textit{for } \; 0 < \mu(i) \leq \bar{\mu}(i)\;,$$

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

Before proceeding further, it will be convenient here to introduce a matrix notation that will be helpful in the sequel. Define the diagonal matrices

$$\mathbf{M}_N = \text{diag}\;\{\mu(0), \mu(1), \ldots, \mu(N)\}, \quad \overline{\mathbf{M}}_N = \text{diag}\;\{\bar{\mu}(0), \bar{\mu}(1), \ldots, \bar{\mu}(N)\}, \tag{19}$$

and the vectors

$$\mathbf{e}_{a,N}^* = [e_a^*(0), e_a^*(1), ..., e_a^*(N)]\;, \quad \mathbf{v}_N^* = [v^*(0), v^*(1), ..., v^*(N)]. \tag{20}$$

It is easy to see that due to the diagonal structure of $\mathbf{M}_N$ and $\overline{\mathbf{M}}_N$, the 2−induced norms of the matrices $(\mathbf{I} - \mathbf{M}_N \overline{\mathbf{M}}_N^{-1})$ and $\mathbf{M}_N \overline{\mathbf{M}}_N^{-1}$ are equal to $\Delta(N)$ and $\gamma(N)$, respectively. We can then rewrite (16) in matrix form,

$$||\overline{\mathbf{M}}_N^{1/2} \mathbf{e}_{a,N}||_2 \;\leq\; \frac{1}{1 - ||\mathbf{I} - \mathbf{M}_N \overline{\mathbf{M}}_N^{-1}||_{2,ind}} \left[||\tilde{\mathbf{w}}_{-1}||_2 \;+\; ||\mathbf{M}_N \overline{\mathbf{M}}_N^{-1}||_{2,ind} ||\overline{\mathbf{M}}_N^{1/2} \mathbf{v}_N||_2 \right].$$

The l_2-stability condition then amounts to guaranteeing a contractive feedback map, $(\mathbf{I} - \mathbf{M}_N \overline{\mathbf{M}}_N^{-1})$. The map is rather trivial (i.e., memoryless) in this case: it consists of a diagonal matrix that simply scales the input sequence $\{\sqrt{\bar{\mu}(\cdot)}\; e_a(\cdot)\}$. We shall see later, especially in the context of filtered-error variants, that a more involved *dynamic* feedback map arises.

3.3 On Convergence and Energy Propagation

In order to further appreciate the significance of the bounds of Theorem 3.2, we now exhibit a convergence result that follows as a consequence of the l_2-stability property. Indeed, assume that the normalized noise sequence $\{\sqrt{\mu(\cdot)}\, v(\cdot)\}$ has finite energy, i.e., $\sum_{i=0}^{\infty} \mu(i)|v(i)|^2 < \infty$. It then follows from (17) that $\sum_{i=0}^{\infty} \mu(i)|e_a(i)|^2 < \infty$. This is true since, for any N, we always have $0 < \gamma(N) < 2$ and $0 < 1 - \Delta(N) < 1$. We therefore conclude that $\{\sqrt{\mu(\cdot)}\, e_a(\cdot)\}$ is a Cauchy sequence and, hence, $\lim_{i\to\infty} \sqrt{\mu(i)}\, e_a(i) = 0$.

If a persistence of excitation condition is further imposed on the the input vectors $\mathbf{u}_i$, then we can also conclude $\lim_{i\to\infty} \mathbf{w}_i = \mathbf{w}$. We omit the details here and instead stress that 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 2.

Indeed, 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 in the stochastic setting that for Gaussian processes,[11] as well as for spherically invariant random processes,[12] 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 2.

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

$$E_w(i) = E_w(i-1) - E_e(i), \tag{21}$$

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 (21) 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. Strict inequality is guaranteed if $E_e(i) \neq 0$. This is in general the case especially when the input vectors $\mathbf{u}_i$ are assumed persistently exciting.

But what if $\mu(i) \neq \bar{\mu}(i)$? In this case the feedback path is active and we now verify that the convergence speed is affected (in fact, it becomes slower) 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

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

where we have defined the coefficient $\tau(i)$ (compare with (21)). It is easy to verify that as long as $\mu(i) \neq \bar{\mu}(i)$ we always have $0 < \tau(i) < 1$. That is, $\tau(i)$ is strictly less than one and the rate of decrease in the energy of $\tilde{\mathbf{w}}_i$ is lowered, thus confirming our earlier remark.

4 FILTERED-ERROR GRADIENT ALGORITHMS

The feedback loop concept of the former sections applies equally well to gradient algorithms that employ filtered versions of the output estimation error, $\tilde{e}_a(i) = d(i) - \mathbf{u}_i\mathbf{w}_{i-1}$. 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 3. 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,

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

It may also be a time-variant filter, in which case the coefficients f_j will vary with time, say $f_j(i)$. A typical application where the need for such algorithms arises is in the active control of noise.[8] 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).

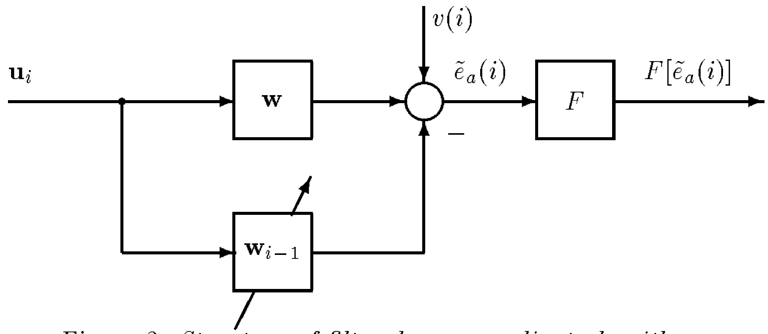

Figure 3: *Structure of filtered-error gradient algorithms.*

4.1 The Modified Filtered-x LMS Algorithm

The filtered-x LMS algorithm employs a recursive update of the form

$$\mathbf{w}_i = \mathbf{w}_{i-1} + \mu(i) F[\mathbf{u}_i]^* \, F[\tilde{e}_a(i)] \; , \tag{22}$$

where the input data $\mathbf{u}_i$ is also processed by the filter F. The linearity of F implies that $F[\tilde{e}_a(i)] = F[\mathbf{u}_i \tilde{\mathbf{w}}_{i-1}] + F[v(i)]$. If slow adaptation is assumed,[8,13] i.e., if the weight estimates do not vary considerably over the length of F, $\mathbf{w}_i \approx \mathbf{w}_{i-1} \approx \ldots \approx \mathbf{w}_{i-M_F}$, 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 standard 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.[14,15] This is achieved by modifying the update expression as follows:

$$\mathbf{w}_i = \mathbf{w}_{i-1} + \mu(i) F[\mathbf{u}_i]^* \left(F[\tilde{e}_a(i)] + F[\mathbf{u}_i \mathbf{w}_{i-1}] - F[\mathbf{u}_i]\mathbf{w}_{i-1} \right) \; . \tag{23}$$

The extra terms that are added to the update recursion have the following effect:

$$F[\tilde{e}_a(i)] + F[\mathbf{u}_i \mathbf{w}_{i-1}] - F[\mathbf{u}_i]\mathbf{w}_{i-1} \quad = \quad F[v(i)] + F[\mathbf{u}_i] \left(\mathbf{w} - \mathbf{w}_{i-1} \right) \; ,$$

where we have invoked the fact that $F[\mathbf{u}_i \mathbf{w}] = F[\mathbf{u}_i]\mathbf{w}$ for a time-invariant plant $\mathbf{w}$. This again reduces the update equation to

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

which 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 $F[v(i)]$ – but this time no approximation is employed. The results of the previous sections will then be immediately applicable with the proper change of variables: $\mathbf{u}_i \longleftrightarrow F[\mathbf{u}_i]$, $v(i) \longleftrightarrow F[v(i)]$.

4.2 The Filtered-Error LMS Algorithm

The so-called filtered-error LMS update[16] retains the input vector unchanged, viz.,

$$\mathbf{w}_i = \mathbf{w}_{i-1} + \mu(i)\mathbf{u}_i^* \, F[\tilde{e}_a(i)] \,. \tag{24}$$

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. Following the discussion that led to (9), we get

$$\mathbf{w}_i \quad = \quad \mathbf{w}_{i-1} + \bar{\mu}(i)\mathbf{u}_i^* \left[e_a(i) + \bar{v}(i) \right] \,, \tag{25}$$

where the modified noise sequence $\{\bar{v}(\cdot)\}$ is now given by,

$$\bar{\mu}(i)\bar{v}(i) = \mu(i)F[v(i)] - \bar{\mu}(i)e_a(i) + \mu(i)F[e_a(i)],$$

and $\bar{\mu}(i) = 1/\|\mathbf{u}_i\|_2^2$. This is of the same form as the one discussed in Section 3.1, which readily implies that the following relation also holds:

$$\frac{\|\tilde{\mathbf{w}}_i\|_2^2 + \bar{\mu}(i)\,|e_a(i)|^2}{\|\tilde{\mathbf{w}}_{i-1}\|_2^2 + \bar{\mu}(i)\,|\bar{v}(i)|^2} = 1 \,. \tag{26}$$

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 4. We remark that the notation,

$$1 - \frac{\mu(i)}{\sqrt{\bar{\mu}(i)}}\, F[\cdot]\, \frac{1}{\sqrt{\bar{\mu}(i)}},$$

that appears in the feedback loop should be interpreted as follows: we first divide $\sqrt{\bar{\mu}(i)}\, e_a(i)$ by $\sqrt{\bar{\mu}(i)}$, followed by the filter F and then by a subsequent scaling by $\frac{\mu(i)}{\sqrt{\bar{\mu}(i)}}$. Likewise, the term $\sqrt{\bar{\mu}(i)}\, v(i)$ is first divided by $\sqrt{\bar{\mu}(i)}$, then filtered by F and finally scaled by $\frac{\mu(i)}{\sqrt{\bar{\mu}(i)}}$. The feedback loop now consists of a dynamic system. But we can

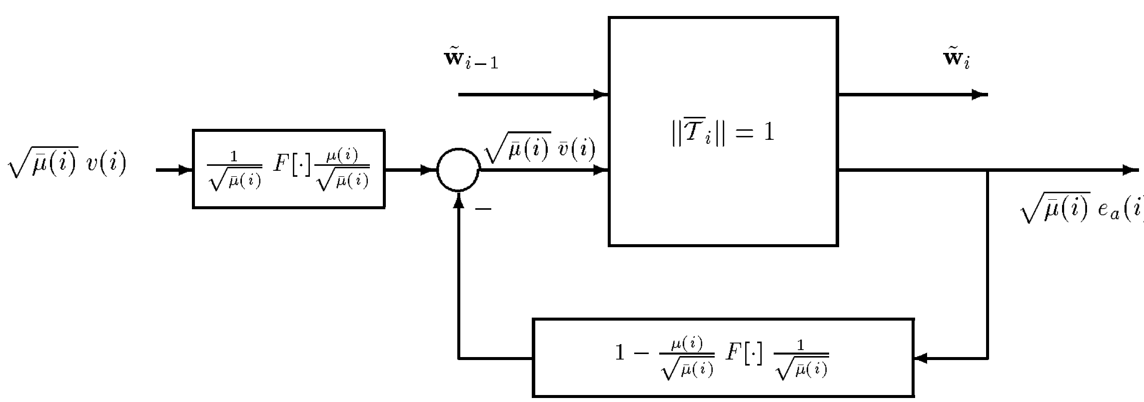

Figure 4: *Filtered-error LMS algorithm as a time-variant lossless mapping with dynamic feedback.*

still proceed to study the l_2-stability of the overall configuration in much the same way as we did in Section 3.1. For this purpose, we use the vector and matrix quantities as in (19)–(20) and define a vector $\bar{\mathbf{v}}_N$ similar to $\mathbf{v}_N$ but with the entries $\bar{v}(\cdot)$ instead of $v(\cdot)$. Also define the lower triangular matrix $\mathbf{F}_N$ that describes the action of the filter F on a sequence at its input. For a time-invariant filter F, this is a Toeplitz lower-triangular matrix with band with equal to $M_F \ll N$. It follows that we can write

$$\overline{\mathbf{M}}_N^{\frac{1}{2}}\bar{\mathbf{v}}_N \quad = \quad \left(\mathbf{I} - \overline{\mathbf{M}}_N^{-\frac{1}{2}}\mathbf{M}_N\mathbf{F}_N\overline{\mathbf{M}}_N^{-\frac{1}{2}}\right)\overline{\mathbf{M}}_N^{\frac{1}{2}}\mathbf{e}_{a,N} + \left(\overline{\mathbf{M}}_N^{-\frac{1}{2}}\mathbf{M}_N\mathbf{F}_N\overline{\mathbf{M}}_N^{-\frac{1}{2}}\right)\overline{\mathbf{M}}_N^{\frac{1}{2}}\mathbf{v}_N \,.$$

Define

$$\Delta(N) \triangleq \|\mathbf{I} - \overline{\mathbf{M}}_N^{-\frac{1}{2}} \mathbf{M}_N \mathbf{F}_N \overline{\mathbf{M}}_N^{-\frac{1}{2}}\|_{2,ind} \text{ and } \gamma(N) \triangleq \|\overline{\mathbf{M}}_N^{-\frac{1}{2}} \mathbf{M}_N \mathbf{F}_N \overline{\mathbf{M}}_N^{-\frac{1}{2}}\|_{2,ind}.$$

If we now follow the arguments of Section 3.1 we easily obtain the following result, which extends Theorem 3.2.

THEOREM 4.1. *Consider the filtered-error LMS recursion (24) and define* $\Delta(N)$ *and* $\gamma(N)$ *as above. If* $\Delta(N) < 1$ *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,*

$$\sqrt{\sum_{i=0}^{N} \bar{\mu}(i)\,|e_a(i)|^2} \leq \frac{1}{1-\Delta(N)} \left[\|\tilde{\mathbf{w}}_{-1}\|_2 + \gamma(N) \sqrt{\sum_{i=0}^{N} \bar{\mu}(i)|v(i)|^2} \right]. \tag{27}$$

Moreover, the map from $\{\sqrt{\mu(\cdot)}\, v(\cdot),\ \tilde{\mathbf{w}}_{-1}\}$ *to* $\{\sqrt{\mu(\cdot)}\, e_a(\cdot)\}$ *will also be* l_2-*stable with*

$$\sqrt{\sum_{i=0}^{N} \mu(i)\,|e_a(i)|^2} \leq \frac{\gamma^{1/2}(N)}{1-\Delta(N)} \left[\|\tilde{\mathbf{w}}_{-1}\|_2 + \gamma^{1/2}(N) \sqrt{\sum_{i=0}^{N} \mu(i)|v(i)|^2} \right]. \tag{28}$$

We thus see that the major requirement is for the feedback matrix $\left(\mathbf{I} - \overline{\mathbf{M}}_N^{-\frac{1}{2}} \mathbf{M}_N \mathbf{F}_N \overline{\mathbf{M}}_N^{-\frac{1}{2}}\right)$ to be contractive. We shall denote it by $\mathbf{P}_N$, which can be easily seen to be

$$\mathbf{P}_N = \begin{pmatrix} 1 - \frac{\mu(0)}{\bar{\mu}(0)} f_0 & & & \mathbf{O} \\ -\frac{\mu(1)}{\sqrt{\bar{\mu}(0)\bar{\mu}(1)}} f_1 & 1 - \frac{\mu(1)}{\bar{\mu}(1)} f_0 & & \\ -\frac{\mu(2)}{\sqrt{\bar{\mu}(0)\bar{\mu}(2)}} f_2 & -\frac{\mu(2)}{\sqrt{\bar{\mu}(1)\bar{\mu}(2)}} f_1 & 1 - \frac{\mu(2)}{\bar{\mu}(2)} f_0 & \\ \vdots & & & \ddots \end{pmatrix}. \tag{29}$$

4.3 The Projection FELMS Algorithm

We focus now on an important choice for the step-size parameter, viz., $\mu(i) = \alpha\, \bar{\mu}(i)$, a scaled multiple of the reciprocal input energy. This leads to an update recursion that is often referred to as a projection update (or Projection FELMS). In this case, it can be seen that the contractivity requirement now collapses to requiring the contractivity of

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

We 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., $\bar{\mu}(i) \approx \bar{\mu}(i-1) \approx \ldots \approx \bar{\mu}(i - M_F)$, which is a reasonable assumption for $M_F \ll M$. In this case, the contractivity of $\mathbf{P}_N$ can be guaranteed if we choose the α so as to satisfy

$$\max_{\Omega} \left|1 - \alpha F(e^{j\Omega})\right| < 1\,. \tag{30}$$

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

$$\min_{\alpha} \max_{\Omega} \left|1 - \alpha F(e^{j\Omega})\right|. \tag{31}$$

Simulation results have confirmed these conclusions. But these are excluded from this paper for reasons of brevity.

5 CONCLUDING REMARKS

The analysis provided herein extends to other forms of adaptive algorithms, and can also be shown to be related to developments in $H^\infty-$theory. Results in this direction are reported in the companion papers.[17–19]

6 REFERENCES

[1] B. Widrow. "Adaptive sampled-data systems." Proc. 1st IFAC Congress, Moscow, 1960.

[2] I. D. Landau. Adaptive Control – The Model Reference Approach. New York: Dekker, 1979.

[3] L. Ljung and T. Söderström. Theory and Practice of Recursive Identification. MIT Press, 1983.

[4] I. D. Landau. "A feedback system approach to adaptive filtering." IEEE Transactions on Information Theory, vol. IT-30, no. 2, March 1984.

[5] H. K. Khalil. Nonlinear Systems. Mac Millan, 1992.

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

[7] B. Widrow and S.D. Stearns. Adaptive Signal Processing. Englewood Cliffs, NJ: Prentice–Hall, 1985.

[8] J. R. Glover, Jr. "Adaptive noise canceling applied to sinusoidal interferences." IEEE Trans. Acoust., Speech, Signal Processing, vol. 25, no. 6., pp. 484–491, Dec. 1977.

[9] P. M. Clarkson and P. R. White. "Simplified analysis of the LMS adaptive filter using a transfer function approximation." IEEE Trans. Acoust., Speech, Signal Processing, vol. 35, no. 7, pp. 987–993, July 1987.

[10] A. H. Sayed and M. Rupp. "On the robustness, convergence, and minimax performance of instantaneous-gradient adaptive filters." Proc. of Asilomar Conference on Signals, Systems, and Computers, pp. 592–596, Oct. 1994.Oct. 1994.

[11] N. J. Bershad. "Analysis of the normalized LMS algorithm with Gaussian inputs." IEEE Trans. Acoust., Speech, Signal Processing, vol. 34, no. 4, pp. 793–806, Aug. 1986.

[12] M. Rupp. "The behavior of LMS and NLMS algorithms in the presence of spherically invariant processes." IEEE Trans. Signal Processing, vol. 41, no. 3, pp. 1149-1160, Mar. 1993.

[13] D. R. Morgan. "An analysis of multiple correlation cancellation loops with a filter in the auxiliary path." IEEE Trans. Acoust., Speech, Signal Processing, vol. 28, no. 4, pp. 454–467, Aug. 1980.

[14] E. Bjarnason. "Noise cancellation using a modified form of the filtered–XLMS algorithm." Proc. Eusipco SignalProcessing V, Brüssel, 1992.

[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., vol. 95, no. 6, pp. 3397-3389, June 1994.

[16] S. Dasgupta, J. S. Garnett and C. R. Johnson, Jr., "Convergence of an adaptive filter with signed filtered error." IEEE Trans. Signal Processing, vol. 42, no. 4, pp. 946-950, April. 1994.

[17] 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, Lake Tahoe, CA, June 1995.

[18] M. Rupp and A. H. Sayed. "Local and global passivity relations for Gauss-Newton methods in adaptive filtering." Proc. (this) SPIE Conference on Advanced Signal Processing: Algorithms, Architectures, and Implementations, San Diego, CA, July 1995.

[19] M. Rupp and A. H. Sayed. "On the stability and convergence of Feintuch's algorithm for adaptive IIR filtering." *Proc. IEEE ICASSP*, Detroit, MI, May 1995.