WA08 10:00

Proceedings of the 34th
Conference on Decision & Control
New Orleans, LA - December 1995

A ROBUSTNESS ANALYSIS OF GAUSS-NEWTON RECURSIVE METHODS

MARKUS RUPP AND ALI H. SAYED

Department of Electrical and Computer Engineering
University of California
Santa Barbara, CA 93106–9560

Abstract—We provide a time-domain robustness analysis of Gauss-Newton recursive methods that are often employed in identification and control. Several free parameters are included in the filter description while combining the covariance update and the weight-vector update, with the exponentially weighted recursive-least-squares (RLS) algorithm being an important special case. One of the contributions of this work is to show that by properly selecting the free parameters, the resulting filter can be made to impose certain bounds on the error quantities, thus resulting in desirable robustness properties (*cf.* H^∞-theory). We also show that an intrinsic feedback structure, mapping the noise sequence and the initial weight error to the apriori estimation errors and the final weight error, can be associated with such recursive schemes.

I. INTRODUCTION

This paper provides a time-domain feedback analysis of the class of Gauss-Newton (GN) recursive schemes, which have been employed in several areas of identification, control, signal processing, and communications (e.g., [1]-[4]). These are recursive estimators that are based on gradient-descent ideas and which involve two update relations: one updates the weight estimate while the other updates the inverse of the sample covariance matrix. Several free parameters are also included in the filter description, which allows for a reasonable degree of freedom in setting up a filter configuration. One of the contributions of this work is to show that by properly selecting the free parameters, the resulting filter can be made to impose certain bounds on the error quantities. These bounds are further shown to result in desirable robustness properties, along the lines of H^∞-filters [5, 6, 7].

We also establish that an intrinsic feedback structure, mapping the noise sequence and the initial weight error to the apriori estimation errors and the final weight error, 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.

This material was based on work supported by the National Science Foundation under Award No. MIP-9409319. A. H. Sayed is an Assistant Professor in the Department of Electrical and Computer Engineering, UCSB, CA 93106. M. Rupp is a post-doctoral fellow. His work was also supported by a scholarship from DAAD (German Academic Exchange Service) as well as the scientific division of NATO (Wissenschaftsausschuß der NATO).

It is then shown that the feedback configuration lends itself rather immediately to stability analysis via a so-called small gain theorem, which is a standard tool in system theory (e.g., [8, 9]). It provides contractivity conditions that are shown to guarantee the l_2-stability of the algorithm, with further implications on the convergence behaviour of the estimator. This is demonstrated by studying the energy flow through the feedback configuration and by exploiting the lossless nature of the feedforward path.

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

II. THE GAUSS-NEWTON RECURSIVE METHOD

There is an abundant literature on the analysis and design of GN methods, especially in the area of parametric system identification (see, e.g., [2, 3]).

We consider a collection of noisy measurements $\{d(i)\}_{i=0}^N$ that are assumed to arise from a linear model of the form

$$d(i) = \mathbf{u}_i\mathbf{w} + v(i), \tag{1}$$

where $v(i)$ denotes the measurement noise or disturbance and $\mathbf{u}_i$ denotes a row input vector. The column vector $\mathbf{w}$ consists of unknown parameters that we wish to estimate. In this paper we focus on the following so-called GN recursive method.

Algorithm 1 (Gauss-Newton Procedure) *Given measurements $\{d(i)\}_{i=0}^N$, an initial guess $\mathbf{w}_{-1}$, and a positive-definite matrix Π_0, recursive estimates of the weight vector $\mathbf{w}$ are obtained as follows:*

$$\mathbf{w}_i = \mathbf{w}_{i-1} + \mu(i)\,\mathbf{P}_i\mathbf{u}_i^*\,(d(i) - \mathbf{u}_i\mathbf{w}_{i-1}), \tag{2}$$

where $\mathbf{P}_i$ satisfies the Riccati equation update

$$\mathbf{P}_i = \frac{1}{\lambda(i)}\left(\mathbf{P}_{i-1} - \frac{\mathbf{P}_{i-1}\mathbf{u}_i^*\mathbf{u}_i\mathbf{P}_{i-1}}{\frac{\lambda(i)}{\beta(i)} + \mathbf{u}_i\mathbf{P}_{i-1}\mathbf{u}_i^*}\right),\ \mathbf{P}_{-1} = \Pi_0, \tag{3}$$

and $\{\lambda(i), \mu(i), \beta(i)\}$ are given positive scalar time-variant coefficients, with $\lambda(i) \leq 1$.

The effect of the coefficients $\{\lambda(i), \mu(i), \beta(i)\}$ on the performance of the algorithm will be studied in the sequel. Here we note that, by applying the matrix inversion formula (e.g., [10]) to (3), the inverse of $\mathbf{P}_i$ satisfies the simple time-update

$$\mathbf{P}_i^{-1} \;=\; \lambda(i)\,\mathbf{P}_{i-1}^{-1} + \beta(i)\,\mathbf{u}_i^*\mathbf{u}_i, \qquad (4)$$

which establishes that $\mathbf{P}_i$ is guaranteed to be positive-definite for $\lambda(i), \beta(i) > 0$ since $\Pi_0 > 0$.

An important special case of (2) is the so-called Recursive-Least-Squares (RLS) algorithm (see, e.g., [10, 11]), which corresponds to the choices $\beta(i) = \mu(i) = 1$ and $\lambda(i) = \lambda = cte$. In this case, the Riccati recursion (3) reduces to

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

which leads to the update equation

$$\mathbf{w}_i \;=\; \mathbf{w}_{i-1} + \frac{\mathbf{P}_{i-1}\mathbf{u}_i^*}{\lambda + \mathbf{u}_i\mathbf{P}_{i-1}\mathbf{u}_i^*}\,(d(i) - \mathbf{u}_i\mathbf{w}_{i-1}),$$

the standard form of the RLS algorithm.

The difference $[d(i) - \mathbf{u}_i\mathbf{w}_{i-1}]$ in (2) will be denoted by $\tilde{e}_a(i)$ and will be referred to as the *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 $\mathbf{w}_i$, $\tilde{\mathbf{w}}_i = \mathbf{w} - \mathbf{w}_i$, and $e_a(i)$ will denote the *apriori estimation error*, $e_a(i) = \mathbf{u}_i\tilde{\mathbf{w}}_{i-1}$. It then follows from the update equation (2) that the weight-error vector $\tilde{\mathbf{w}}_{i-1}$ satisfies the recursive equation:

$$\tilde{\mathbf{w}}_i = \tilde{\mathbf{w}}_{i-1} - \mu(i)\mathbf{P}_i\mathbf{u}_i^*\tilde{e}_a(i)\,. \qquad (5)$$

It is also straightforward to verify that the apriori estimation error, $e_a(i)$, and the estimation error, $\tilde{e}_a(i)$, differ by the noise $v(i)$, i.e., $\tilde{e}_a(i) = e_a(i) + v(i)$. We further define the *aposteriori estimation* error, $e_p(i) = \mathbf{u}_i\tilde{\mathbf{w}}_i$, and note that if we multiply (5) by $\mathbf{u}_i$ from the left we get the following relation (used later in (13)) between $e_p(i)$, $e_a(i)$, and $v(i)$,

$$e_p(i) = [\,1 - \mu(i)\mathbf{u}_i\mathbf{P}_i\mathbf{u}_i^*\,]\;e_a(i) - \mu(i)\mathbf{u}_i\mathbf{P}_i\mathbf{u}_i^*\;v(i). \qquad (6)$$

III. A TIME-DOMAIN ANALYSIS

We now pursue a closer analysis of the GN recursion (2). For this purpose, we invoke the time-domain update recursion (5), multiply by $\mathbf{P}_i^{-\frac{1}{2}}$ from the left, and compute the squared norm (i.e., energies) of both sides of the resulting expression, i.e.,

$$\tilde{\mathbf{w}}_i^*\mathbf{P}_i^{-1}\tilde{\mathbf{w}}_i = \|\mathbf{P}_i^{-\frac{1}{2}}\tilde{\mathbf{w}}_{i-1} - \mu(i)\mathbf{P}_i^{\frac{*}{2}}\mathbf{u}_i^*\tilde{e}_a(i)\|_2^2\,. \qquad (7)$$

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

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

we conclude that the following equality always holds,

$$\tilde{\mathbf{w}}_i^*\mathbf{P}_i^{-1}\tilde{\mathbf{w}}_i + \mu(i)\,|e_a(i)|^2 + \mu(i)\,(\,1 - \mu(i)\mathbf{u}_i\mathbf{P}_i\mathbf{u}_i^*)\,|\tilde{e}_a(i)|^2$$
$$= \tilde{\mathbf{w}}_{i-1}^*\mathbf{P}_i^{-1}\tilde{\mathbf{w}}_{i-1} + \mu(i)\,|v(i)|^2\ . \qquad (8)$$

Substituting (4) for $\mathbf{P}_i^{-1}$ in the right-hand side, the last equality can be rewritten as

$$\begin{aligned}(\mu(i) - \beta(i))\,|e_a(i)|^2 + \mu(i)\,(\,1 - \mu(i)\mathbf{u}_i\mathbf{P}_i\mathbf{u}_i^*))|\tilde{e}_a(i)|^2 \\ = \lambda(i)\tilde{\mathbf{w}}_{i-1}^*\mathbf{P}_{i-1}^{-1}\tilde{\mathbf{w}}_{i-1} - \tilde{\mathbf{w}}_i^*\mathbf{P}_i^{-1}\tilde{\mathbf{w}}_i + \mu(i)\,|v(i)|^2\ . \quad (9)\end{aligned}$$

Equality (9) involves "energy" terms and allows us to establish that the following error bounds are always satisfied for the GN recursion (2) – in the statement of the Lemma, we employ the quantity $\bar{\mu}(i) = (\mathbf{u}_i\mathbf{P}_i\mathbf{u}_i^*)^{-1}$.

Lemma 1 (A Local Passivity Relation) *Consider recursion (2). It always holds that*

$$\frac{\tilde{\mathbf{w}}_i^*\mathbf{P}_i^{-1}\tilde{\mathbf{w}}_i + (\mu(i) - \beta(i))\,|e_a(i)|^2}{\lambda(i)\tilde{\mathbf{w}}_{i-1}^*\mathbf{P}_{i-1}^{-1}\tilde{\mathbf{w}}_{i-1} + \mu(i)\,|v(i)|^2}\ \left\{\begin{array}{ll}\le 1; & \mu(i) < \bar{\mu}(i),\\ = 1; & \mu(i) = \bar{\mu}(i),\\ \ge 1; & \mu(i) > \bar{\mu}(i).\end{array}\right. \qquad (10)$$

Such relations also arise in the case of instantaneous-gradient-based algorithms, as detailed in [12].

The first two bounds in the above lemma admit an interesting interpretation that highlights a robustness property of the GN recursion (2). To clarify this, we assume that $\beta(i) \le \mu(i)$ in order to guarantee $(\mu(i) - \beta(i)) \ge 0$ and, hence, the factor $(\mu(i) - \beta(i))\,|e_a(i)|^2$ can be regarded as an energy term. In this case, we can interpret the first two bounds in the lemma as stating 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 weighted energies of the resulting errors, viz.,

$$\tilde{\mathbf{w}}_i^*\mathbf{P}_i^{-1}\tilde{\mathbf{w}}_i + (\mu(i) - \beta(i))\,|e_a(i)|^2\,,$$

will always be smaller than or equal to the sum of the weighted energies of the starting errors (or disturbances),

$$\lambda(i)\tilde{\mathbf{w}}_{i-1}^*\mathbf{P}_{i-1}^{-1}\tilde{\mathbf{w}}_{i-1} + \mu(i)\,|v(i)|^2\,.$$

The relations of Lemma 1 are local conclusions but similar results also hold over intervals of time. Indeed, if we assume $\mu(i) \le \bar{\mu}(i)$ for all i in the interval $0 \le i \le N$, then the following inequality holds for every time instant in the interval,

$$\begin{aligned}&(\mu(i) - \beta(i))\,|e_a(i)|^2 \\ &\le\ \lambda(i)\tilde{\mathbf{w}}_{i-1}^*\mathbf{P}_{i-1}^{-1}\tilde{\mathbf{w}}_{i-1} - \tilde{\mathbf{w}}_i^*\mathbf{P}_i^{-1}\tilde{\mathbf{w}}_i + \mu(i)\,|\bar{v}(i)|^2\ .\end{aligned}$$

Summing over i we conclude that

$$\begin{aligned}&\tilde{\mathbf{w}}_N^*\mathbf{P}_N^{-1}\tilde{\mathbf{w}}_N\ +\ \sum_{i=0}^{N}(\mu(i) - \beta(i))\lambda^{[i+1,N]}\,|e_a(i)|^2 \\ &\le \lambda^{[0,N]}\tilde{\mathbf{w}}_{-1}^*\mathbf{P}_{-1}^{-1}\tilde{\mathbf{w}}_{-1}\ +\ \sum_{i=0}^{N}\mu(i)\lambda^{[i+1,N]}|v(i)|^2\,, \qquad (11)\end{aligned}$$

which establishes a passivity relation over the interval $0 \le i \le N$. Here, we have used the notation $\lambda^{[i,j]} = \prod_{k=i}^{j}\lambda(k)$.

As a special case, assume $\lambda(i) = \mu(i) = \beta(i) = 1$ (which corresponds to an RLS problem in the absence of exponential weighting). Then the above conclusion implies that the mapping from $\left\{v(\cdot), \Pi_0^{-1/2}\tilde{\mathbf{w}}_{-1}\right\}$ to $\left\{\mathbf{P}_N^{-1/2}\tilde{\mathbf{w}}_N\right\}$ is always a contraction. That is,

$$\tilde{\mathbf{w}}_N^* \mathbf{P}_N^{-1} \tilde{\mathbf{w}}_N \leq \tilde{\mathbf{w}}_{-1}^* \Pi_0^{-1} \tilde{\mathbf{w}}_{-1} + \sum_{i=0}^{N} |v(i)|^2 ,$$

which is a known result.

IV. THE FEEDBACK STRUCTURE

Before proceeding, we first establish the following fact.

Lemma 2 (A Lower Bound on $\bar{\mu}(i)$) *Consider the GN algorithm (2) with the positive parameters $\{\lambda(i), \mu(i), \beta(i)\}$. Define $\bar{\mu}(i)$ as before, $\bar{\mu}(i) = (\mathbf{u}_i \mathbf{P}_i \mathbf{u}_i^*)^{-1}$. It always holds, for nonzero vectors $\mathbf{u}_i$, that $\bar{\mu}(i) > \beta(i)$.*

Proof: Introduce the notation $\bar{\mu}(i|i-1) \triangleq (\mathbf{u}_i \mathbf{P}_{i-1} \mathbf{u}_i^*)^{-1}$. Then we can write

$$\bar{\mu}^{-1}(i) = \mathbf{u}_i \mathbf{P}_i \mathbf{u}_i^* = \frac{1}{\beta(i) + \lambda(i)\bar{\mu}(i|i-1)}.$$

In other words, $\bar{\mu}(i) = \beta(i) + \lambda(i)\bar{\mu}(i|i-1)$, where the term $\lambda(i)\bar{\mu}(i|i-1)$ is strictly positive since $\mathbf{P}_{i-1} > 0$. ∎

We now show that the bounds in Lemma 1 can be described via an alternative form that leads to an interesting feedback structure. To clarify this, we rewrite (6) as

$$e_p(i) = \left(1 - \frac{\mu(i)}{\bar{\mu}(i)}\right) e_a(i) - \frac{\mu(i)}{\bar{\mu}(i)} v(i), \qquad (12)$$

and use it to re-express the update equation (2) in the following form:

$$\begin{aligned} \mathbf{w}_i &= \mathbf{w}_{i-1} + \mu(i)\mathbf{P}_i\mathbf{u}_i^* e_a(i) + \mu(i)\mathbf{P}_i\mathbf{u}_i^* v(i) \\ &= \mathbf{w}_{i-1} + \bar{\mu}(i)\mathbf{P}_i\mathbf{u}_i^* e_a(i) \\ &\quad + \mathbf{P}_i\mathbf{u}_i^* \underbrace{[\mu(i)v(i) - (\bar{\mu}(i) - \mu(i))e_a(i)]}_{-\bar{\mu}(i)e_p(i) = \bar{\mu}(i)\bar{v}(i)}. \end{aligned} \qquad (13)$$

This shows that the weight-update equation (2) can be rewritten in terms of a new step-size parameter $\bar{\mu}(i)$ and a modified "noise" term $\bar{v}(i) = -e_p(i)$. This should be compared with (2), which corresponds to

$$\mathbf{w}_i = \mathbf{w}_{i-1} + \mu(i)\mathbf{P}_i\mathbf{u}_i^*[e_a(i) + v(i)].$$

If we now apply arguments similar to those prior to (10) to (13), we readily conclude that the following equality holds for all $\mu(i)$ and $v(i)$,

$$\frac{\tilde{\mathbf{w}}_i^* \mathbf{P}_i^{-1} \tilde{\mathbf{w}}_i + (\bar{\mu}(i) - \beta(i))\, |e_a(i)|^2}{\lambda(i)\tilde{\mathbf{w}}_{i-1}^* \mathbf{P}_{i-1}^{-1} \tilde{\mathbf{w}}_{i-1} + \bar{\mu}(i)\, |e_p(i)|^2} = 1 . \qquad (14)$$

Recall that we have shown earlier that $\bar{\mu}(i) > \beta(i)$. Hence, the above relation establishes that 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 $\sqrt{\bar{\mu}(\cdot) - \beta(\cdot)}e_a(\cdot)$ can then be expressed in terms of a feedback structure as shown in Figure 1. The feedback loop consists of a gain factor that is equal to $(1 - \mu(i)/\bar{\mu}(i))/\sqrt{1 - \beta(i)/\bar{\mu}(i)}$. Also,

$$\bar{\mu}^{\frac{1}{2}}(i)\bar{v}(i) = \frac{\mu(i)}{\bar{\mu}^{\frac{1}{2}}(i)} v(i) - \left(1 - \frac{\mu(i)}{\bar{\mu}(i)}\right) \bar{\mu}^{\frac{1}{2}}(i) e_a(i) . \qquad (15)$$

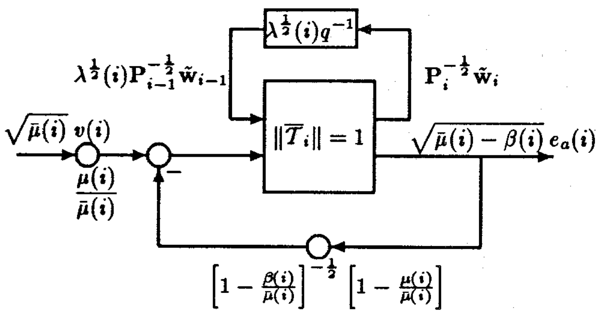

Figure 1: *A time-variant lossless mapping with feedback.*

Lemma 3 (Feedback Representation) *Consider the GN recursion (2) and (4). It always holds, for any $\mu(i)$, that*

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

where $\bar{\mu}^{-1}(i) = \mathbf{u}_i \mathbf{P}_i \mathbf{u}_i^*$. That is, the map $\overline{\mathcal{T}}_i$ is always lossless. Moreover, this map leads to the feedback structure with a lossless forward path and a gain feedback loop as shown in Figure 1.

The feedback configuration of Figure 1 lends itself rather immediately to stability analysis, as we now explain. It follows from the equality in Lemma 3 that for every time instant i, and for any $\mu(i)$, we have

$$\begin{aligned} &(\bar{\mu}(i) - \beta(i))\, |e_a(i)|^2 = \\ &\lambda(i)\tilde{\mathbf{w}}_{i-1}^* \mathbf{P}_{i-1}^{-1} \tilde{\mathbf{w}}_{i-1} - \tilde{\mathbf{w}}_i^* \mathbf{P}_i^{-1} \tilde{\mathbf{w}}_i + \bar{\mu}(i)\, |\bar{v}(i)|^2 . \end{aligned} \qquad (16)$$

This allows us to conclude that the system in Figure 1 is l_2−stable, i.e., it maps a bounded energy sequence $\{\sqrt{\bar{\mu}(\cdot)}\ v(\cdot)\}$ to a bounded energy sequence $\{\sqrt{\bar{\mu}(\cdot) - \beta(\cdot)}\ e_a(\cdot)\}$ in a sense precised in (20) below. In fact, we shall also conclude that the same result holds even if we replace $\bar{\mu}(\cdot)$ by $\mu(\cdot)$. Such a result is desirable because it allows us to conclude the convergence of $e_a(\cdot)$ to zero, provided the noise sequence satisfies an additional condition.

For this purpose, assume we run the GN recursion (2) from time $i = 0$ up to time N. If we compute the sum of both sides of the equality (16) we obtain,

$$\begin{aligned} \sum_{i=0}^{N} \lambda^{[i+1,N]}(\bar{\mu}(i) - \beta(i))\, |e_a(i)|^2 &= \lambda^{[0,N]}\tilde{\mathbf{w}}_{-1}^* \mathbf{P}_{-1}^{-1} \tilde{\mathbf{w}}_{-1} \\ &- \tilde{\mathbf{w}}_N^* \mathbf{P}_N^{-1} \tilde{\mathbf{w}}_N + \sum_{i=0}^{N} \lambda^{[i+1,N]}\bar{\mu}(i)\, |\bar{v}(i)|^2 , \end{aligned}$$

which also implies that (by ignoring $\tilde{\mathbf{w}}_N^* \mathbf{P}_N^{-1} \tilde{\mathbf{w}}_N$)

$$\sum_{i=0}^{N} \lambda^{[i+1,N]}(\bar{\mu}(i) - \beta(i))\, |e_a(i)|^2$$

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

Consequently,

$$\sqrt{\sum_{i=0}^{N} \lambda^{[i+1,N]}(\bar{\mu}(i) - \beta(i))\,|e_a(i)|^2} \tag{17}$$

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

But it follows from (15), and from the triangular inequality for norms, that

$$\sqrt{\sum_{i=0}^{N} \lambda^{[i+1,N]}\bar{\mu}(i)\,|\bar{v}(i)|^2} \leq \sqrt{\sum_{i=0}^{N} \lambda^{[i+1,N]}\frac{\mu^2(i)}{\bar{\mu}(i)}|v(i)|^2}$$

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

We thus conclude that

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

$$+ \sqrt{\sum_{i=0}^{N} \lambda^{[i+1,N]}\frac{\mu^2(i)}{\bar{\mu}(i)}|v(i)|^2}$$

$$+ \sqrt{\sum_{i=0}^{N} \lambda^{[i+1,N]}\left|1 - \frac{\mu(i)}{\bar{\mu}(i)}\right|^2 \bar{\mu}(i)|e_a(i)|^2} . \tag{18}$$

Define

$$\Delta(N) = \max_{0\leq i\leq N} \left|\frac{1 - \frac{\mu(i)}{\bar{\mu}(i)}}{\sqrt{1 - \frac{\beta(i)}{\bar{\mu}(i)}}}\right| \quad \text{and} \quad \gamma(N) = \max_{0\leq i\leq N} \frac{\mu(i)}{\bar{\mu}(i)} . \tag{19}$$

It then follows that

$$\sqrt{\sum_{i=0}^{N} \lambda^{[i+1,N]}(\bar{\mu}(i) - \beta(i))\,|e_a(i)|^2} \leq \lambda^{[0,N]}\tilde{\mathbf{w}}_{-1}\mathbf{P}_{-1}^{-1}\tilde{\mathbf{w}}_{-1}$$

$$+ \gamma(N)\sqrt{\sum_{i=0}^{N} \lambda^{[i+1,N]}\bar{\mu}(i)|v(i)|^2}$$

$$+ \Delta(N)\sqrt{\sum_{i=0}^{N} \lambda^{[i+1,N]}(\bar{\mu}(i) - \beta(i))|e_a(i)|^2} .$$

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

$$\sqrt{\sum_{i=0}^{N} \lambda^{[i+1,N]}(\bar{\mu}(i) - \beta(i))\,|e_a(i)|^2} \leq \frac{1}{1-\Delta(N)} \tag{20}$$

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

which establishes the desired 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 [8, 9]. In simple terms, the theorem states that the l_2-stability of a feedback configuration 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 while the 2-induced norm of the feedback map is $\Delta(N)$. Note also that for $\Delta(N) < 1$ we clearly need that, for all i,

$$0 < \mu(i) < \bar{\mu}(i)\left(1 + \sqrt{1 - \frac{\beta(i)}{\bar{\mu}(i)}}\right) . \tag{21}$$

Theorem 1 (l_2-Stability) *Consider the GN recursion (2) and define $\Delta(N)$ and $\gamma(N)$ as in (19) and also*

$$\tilde{\gamma}(N) \triangleq \max_{0\leq i\leq N} \frac{\mu(i) - \beta(i)}{\bar{\mu}(i) - \beta(i)}.$$

If (21) holds then 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 in the sense of (20). Moreover, if $\beta(i) \leq \mu(i)$ 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 also l_2-stable in the following sense:

$$\sqrt{\sum_{i=0}^{N} \lambda^{[i+1,N]}(\mu(i) - \beta(i))\,|e_a(i)|^2} \leq \frac{\tilde{\gamma}^{1/2}(N)}{1-\Delta(N)}$$

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

Proof: (of second bound). Note that (18) implies

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

$$+ \gamma^{1/2}(N)\sqrt{\sum_{i=0}^{N} \lambda^{[i+1,N]}\mu(i)|v(i)|^2}$$

$$+ \Delta(N)\sqrt{\sum_{i=0}^{N} \lambda^{[i+1,N]}(\bar{\mu}(i) - \beta(i))|e_a(i)|^2} .$$

Now (22) follows by noting that

$$\sum_{i=0}^{N} \lambda^{[i+1,N]}(\mu(i) - \beta(i))|e_a(i)|^2$$

$$\leq \tilde{\gamma}(N)\sum_{i=0}^{N} \lambda^{[i+1,N]}(\bar{\mu}(i) - \beta(i))|e_a(i)|^2.$$

■

In fact, a stronger upper bound than (22) can be given when $\mu(i)$ is further restricted to the interval $0 < \beta(i) \leq \mu(i) \leq \bar{\mu}(i)$. This follows from the arguments in Section III.

Lemma 4 (A Tighter Bound) *Consider the GN recursion (2). If $0 < \beta(i) \leq \mu(i) \leq \bar{\mu}(i)$ then a tighter bound is the following:*

$$\sqrt{\sum_{i=0}^{N} \lambda^{[i+1,N]}(\mu(i) - \beta(i))\,|e_a(i)|^2} \leq$$

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

V. ON CONVERGENCE AND ENERGY PROPAGATION

We now exhibit a convergence result that follows as a consequence of the l_2-stability property. Assume that the normalized noise sequence $\{\sqrt{\lambda^{[\cdot+1,N]}\mu(\cdot)}\; v(\cdot)\}$ has finite energy, i.e., $\sum_{i=0}^{\infty} \lambda^{[i+1,N]}\mu(i)|v(i)|^2 < \infty$. It then follows from (22) that $\sum_{i=0}^{\infty} \lambda^{[i+1,N]}(\mu(i) - \beta(i))|e_a(i)|^2 < \infty$ (for $\beta(i) \leq \mu(i)$). This is true since, for any N, we always have $0 < \gamma(N) < 2$, $0 < \bar{\gamma}(N) < 2$ and $0 < 1 - \Delta(N) < 1$. We therefore conclude that $\{\sqrt{\lambda^{[\cdot+1,N]}(\mu(\cdot) - \beta(\cdot))}\; e_a(\cdot)\}$ is a Cauchy sequence and, hence,

$$\lim_{i\to\infty} \sqrt{\lambda^{[i+1,N]}(\mu(i) - \beta(i))}\; e_a(i) = 0. \tag{22}$$

A similar analysis can also be carried out for finite-power noise but we shall omit the details here for brevity. Instead we stress that more physical insights into the convergence behaviour of the GN recursion (2) can be obtained by studying the energy flow through the feedback configuration of Figure 1.

For this purpose, assume we have noiseless measurements $d(i) = \mathbf{u}_i\mathbf{w}$. 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 $e_a(\cdot)$. At time $i = -1$, the initial energy fed into the system is due to the initial guess $\tilde{\mathbf{w}}_{-1}$ and is equal to $\tilde{\mathbf{w}}_{-1}^{*}\mathbf{P}_{-1}^{-1}\tilde{\mathbf{w}}_{-1}$. We shall denote this energy by $E_w(-1)$. Now, at any subsequent time instant i, the total energy entering the lossless system should be equal to the total energy exiting the system, viz., $\lambda(i)E_w(i-1) = E_w(i)+E_e(i)$, or, equivalently,

$$E_w(i) = \lambda(i)E_w(i-1) - E_e(i), \tag{23}$$

where we are denoting by

$$E_e(i) \triangleq (\bar{\mu}(i) - \beta(i))|\mathbf{u}_i\tilde{\mathbf{w}}_{i-1}|^2, \quad E_w(i) \triangleq \tilde{\mathbf{w}}_i^{*}\mathbf{P}_i^{-1}\tilde{\mathbf{w}}_i.$$

Expression (23) implies that, for $\lambda(i) \leq 1$, 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 $E_e(i) \neq 0$. Note also that the so-called forgetting factor $\lambda(i)$ plays an important role.

But what if $\mu(i) \neq \bar{\mu}(i)$? In this case, the feedback path is active and the convergence speed is now affected (in fact, it becomes slower) since the rate of decrease in the energy of the weight-error vector is lowered. Indeed, for $\mu(i) \neq \bar{\mu}(i)$, we always have part of the output energy at $e_a(\cdot)$ fed-back into the input of the lossless system. More precisely, if we let $E_{\bar{v}}(i)$ denote the energy term $\bar{\mu}(i)|\bar{v}(i)|^2$, then the following equality must hold:

$$\lambda(i)E_w(i-1) + E_{\bar{v}}(i) = E_w(i) + E_e(i)$$

at any time instant i. Also, the feedback loop implies that

$$E_{\bar{v}}(i) = \left| \frac{1 - \frac{\mu(i)}{\bar{\mu}(i)}}{\sqrt{1 - \frac{\beta(i)}{\bar{\mu}(i)}}} \right|^2 E_e(i) < E_e(i),$$

since we are assuming a contractive feedback connection. Therefore,

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

where we have defined the coefficient $\tau(i)$ (compare with (23)). 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.

VI. FILTERED-ERROR ALGORITHMS WITH GAUSS-NEWTON UPDATES

The feedback loop concept of the former sections applies equally well to Gauss-Newton algorithms that employ filtered versions of the output estimation error, $\tilde{e}_a(i)$. Such algorithms are useful when the error $\tilde{e}_a(i)$ can not be observed directly, but rather a filtered version of it, say $F[\tilde{e}_a(i)]$ with a finite-impulse response filter of order M_F and coefficients $\{f_j\}$. A typical application arises in the active control of noise. The filtered-error Gauss-Newton algorithm (FEGN) uses an update equation of the form $\mathbf{w}_i = \mathbf{w}_{i-1} + \mu(i)\mathbf{P}_i\mathbf{u}_i^{*}\, F[\tilde{e}_a(i)]$. In this case, it can be verified that

$$\frac{\tilde{\mathbf{w}}_i^{*}\mathbf{P}_i^{-1}\tilde{\mathbf{w}}_i + (\bar{\mu}(i) - \beta(i))\,|e_a(i)|^2}{\lambda(i)\tilde{\mathbf{w}}_{i-1}^{*}\mathbf{P}_{i-1}^{-1}\tilde{\mathbf{w}}_{i-1} + \bar{\mu}(i)\,|\bar{v}(i)|^2} = 1\,, \tag{24}$$

where we have introduced the modified noise sequence $\{\bar{v}(\cdot)\}$,

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

This again establishes a feedback interconnection but with a *dynamic* feedback loop that is given by

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

It can be verified that the l_2-stability of the feedback structure now requires the contractivity of the matrix

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

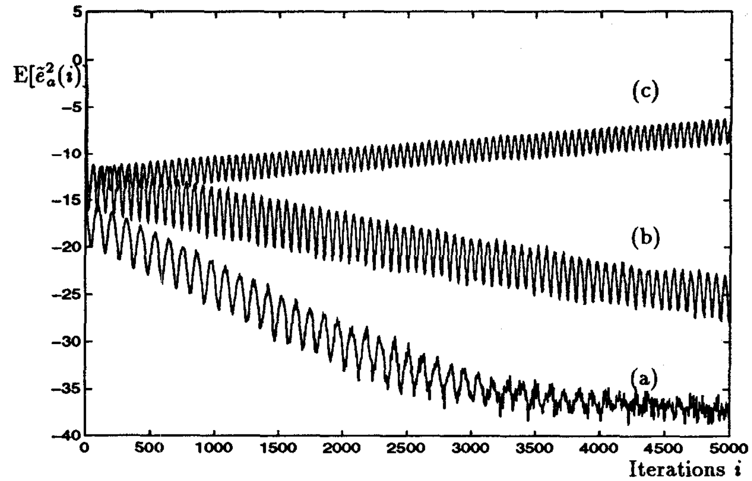

Figure 2: *Convergence behaviour for FEGN algorithm with sinusoidal input sequence and various step-sizes* $\alpha = 0.085(a), 0.15(b), 0.18(c)$.

An important special choice for the step-size parameter corresponds to $\mu(i) = \alpha\, \bar{\mu}(i)$, $\alpha > 0$. If we assume that the (weighted) energy of the input sequence $\mathbf{u}_i$ does not change very rapidly over the filter length M_F, i.e., $\bar{\mu}(i) \approx \bar{\mu}(i-1) \approx \ldots \approx \bar{\mu}(i - M_F)$, and that the $\beta(i)$ satisfy $\beta(i) = \bar{\beta}\bar{\mu}(i)$ with $\bar{\beta} \ll 1$, then the contractivity condition, and faster convergence, can be satisfied by choosing α as

$$\min_{\alpha} \max_{\Omega} \left|1 - \alpha F(e^{j\Omega})\right|. \tag{25}$$

If the resulting minimum is less than 1 then the corresponding optimum α will result in l_2-stability.

A demonstration to this effect is shown in the following simulation for the filter $F(q) = 1 - 1.2\, q^{-1} + 0.72\, q^{-2}$ with $\beta(i) = 0.05\bar{\mu}(i), \lambda = 0.99$, and $\mu(i) = \alpha\bar{\mu}(i)$. The input sequence $u(i)$ (assuming a $\mathbf{u}_i$ with shift structure) is chosen as sinusoidal with frequency Ω_0, so that the apriori error signal can be assumed to be dominated by this frequency. In this case, the optimum α can be found via the simpler expression $\min_\alpha \left|1 - \alpha F\left(e^{j\Omega_o}\right)\right|$, which can be solved explicitly and we get $\alpha_{opt} = \text{Real}\left\{\frac{1}{F(e^{-j\Omega_0})}\right\}$. Following the same procedure, the step-size α_{lim} for which the stability limit is achieved can be calculated to be $\alpha_{lim} = 2\alpha_{opt}$. To verify these statements, we created an input sequence of the form $u(i) = \sin[1.2i + \phi]$, where 50 different values for ϕ were uniformly chosen from the interval $[-\pi, \pi]$. The optimal step-size α_{opt} can thus be calculated to be $\alpha_{opt} = 0.085$ and the stability bound is obtained for $\alpha_{lim} = 2\alpha_{opt} = 0.17$. Figure 2 shows three runs of the FEGN for the choices $\alpha = 0.085, \alpha = 0.15$ and $\alpha = 0.18$. As expected, the first value of α leads to the fastest convergence speed. In every simulation we averaged over 50 trials. The additive noise $v(i)$ was assumed to be -40dB below the input power during the experiments and the order of the adaptive filter was set to $M = 10$. The algorithm was run for $N = 5000$ iterations.

Further connections of the methods of this paper to results in H$^\infty$-filtering can be found in [13].

References

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

[2] H. W. Sorenson, *Parameter Estimation*, Marcel Dekker, New York, 1980.

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

[4] G. C. Goodwin and K. S. Sin, *Adaptive Filtering, Prediction and Control*, Prentice-Hall, Englewood Cliffs, NJ, 1984.

[5] P. P. Khargonekar and K. M. Nagpal, "Filtering and smoothing in an H^∞- setting," *IEEE Trans. on Automatic Control*, AC-36, pp. 151–166, 1991.

[6] U. Shaked and Y. Theodor, "H^∞-optimal estimation: A tutorial," *Proc. IEEE Conference on Decision and Control*, pp. 2278–2286, Tucson, AZ, Dec. 1992.

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

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

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

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

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

[12] A. H. Sayed and M. Rupp, "A time-domain feedback analysis of adaptive gradient algorithms via the small gain theorem," *Proc. SPIE Conference on Advanced Signal Processing: Algorithms, Architectures, and Implementations*, vol. 2563, pp. 458–469, San Diego, July 1995.

[13] A. H. Sayed and M. Rupp, A class of adaptive nonlinear H$^\infty$-filters with guaranteed l_2-stability, *Proc. IFAC Symposium on Nonlinear Control System Design*, vol. 1, pp. 455-460, Tahoe City, CA, June 1995.