Proc. SPIE Conf. on Advanced Signal Processing, San Diego, vol. 2563, pp. 218–229, July 1995.

Local and global passivity relations for Gauss-Newton methods in adaptive filtering *

MARKUS RUPP AND ALI H. SAYED

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

ABSTRACT

We provide a time-domain analysis of the robustness and stability performance of Gauss-Newton recursive methods that are often used in identification and control. Several free parameters are included in the filter description while combining the covariance update with the weight-vector update; 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 shown to impose certain bounds on the error quantities, thus resulting in desirable robustness and stability properties. We also show 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.

Keywords: Adaptive Gauss-Newton filters, feedback connection, l_2-stability, small gain theorem, contraction mapping.

1 INTRODUCTION

This paper provides a time-domain feedback analysis of the class of Gauss-Newton 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 involve two update relations: one updates the weight estimate, while the other updates the inverse of the sample covariance matrix. In this paper, we include several free parameters into the filter description while combining the covariance update with the weight-vector update. The parameters allow for a reasonable degree of freedom in setting up a filter configuration, and one of the contributions of this work is to show that by properly selecting the free parameters, the resulting filter can be shown to impose certain bounds on the error quantities. These bounds are further shown to result in desirable robustness and stability properties. In particular, we derive several new local and global passivity relations that are shown to explain the robust behaviour of this class algorithms.

We also establish that an intrinsic feedback structure, mapping the noise sequence and the initial weight guess

*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).

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 here to consist of two major blocks: a time-variant lossless (i.e., energy preserving) feedforward path and a time-variant feedback path.

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.,[5,6]). 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 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 denote either a zero vector or a zero matrix. The notation $||\mathbf{x}||_2^2$ will denote the squared Euclidean norm of a column (or row) vector $\mathbf{x}$, e.g., $||\mathbf{x}||_2^2 = \mathbf{x}^*\mathbf{x}$. Also, given a positive definite matrix $\mathbf{A}$, $\mathbf{A} > \mathbf{0}$, a square-root factor will be defined as any matrix, say $\mathbf{A}^{1/2}$, such that $\mathbf{A} = (\mathbf{A}^{1/2})(\mathbf{A}^{1/2})^*$. Such square-root factors are not unique. They can be made unique, e.g., by insisting that the factors be Hermitian or that they be triangular (with positive diagonal elements). In most applications, the triangular form is preferred. For convenience, we shall also write $(\mathbf{A}^{1/2})^* = \mathbf{A}^{*/2}$, $(\mathbf{A}^{1/2})^{-1} = \mathbf{A}^{-1/2}$, and $(\mathbf{A}^{-1/2})^* = \mathbf{A}^{-*/2}$.

2 THE GAUSS-NEWTON RECURSIVE METHOD

There is an abundant literature on the analysis and design of Gauss-Newton methods, especially in the area of parametric system identification (see, e.g.,[2,3,7]). Here, we only wish to briefly review this class of algorithms before proceeding into a closer analysis of their behaviour.

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

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

Here $v(i)$ 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 shall focus on the following so-called Gauss-Newton 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:*

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

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

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

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

Note that, for the sake of generality and for later purposes, we have included in the statement of the algorithm three coefficients $\{\lambda(i), \mu(i), \beta(i)\}$. The effect of these coefficients on the performance of the algorithm will be studied in later sections. Note also that by applying the matrix inversion formula (e.g.,[8]) to (3) we obtain that the inverse of $\mathbf{P}_i$ satisfies the simple time-update

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

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

2.1 The RLS Algorithm

An important special case of (2) is the so-called Recursive-Least-Squares (RLS) algorithm (see, e.g.,[8,9]), which corresponds to the choices $\beta(i) = \mu(i) = 1$ and $\lambda(i) = \lambda = cte$. In this case, the Riccati recursion (3) reduces to

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

which also implies that

$$\mathbf{P}_i\mathbf{u}_i^* = \frac{\mathbf{P}_{i-1}\mathbf{u}_i^*}{\lambda + \mathbf{u}_i\mathbf{P}_{i-1}\mathbf{u}_i^*}.$$

Using this last equality in (2) leads to the update equation

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

which is the standard form of the RLS algorithm.

2.2 Error Signals

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 *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 then follows from the update equation (2) that the weight-error vector $\tilde{\mathbf{w}}_{i-1}$ satisfies the recursive equation:

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

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 (5) by $\mathbf{u}_i$ from the left we obtain the following relation (used later in (19)) between $e_p(i)$, $e_a(i)$, and $v(i)$,

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

3 A TIME-DOMAIN ANALYSIS

We now pursue a closer analysis of the Gauss-Newton recursion (2) in order to highlight an important feedback structure that is implied by the algorithm. This structure will play an important role in our discussions and will serve as a basis for the robustness analysis provided herein. 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.,

$$\begin{aligned}
\tilde{\mathbf{w}}_i^*\mathbf{P}_i^{-1}\tilde{\mathbf{w}}_i \quad &= \quad ||\mathbf{P}_i^{-\frac{1}{2}}\tilde{\mathbf{w}}_{i-1} - \mu(i)\mathbf{P}_i^{\frac{*}{2}}\mathbf{u}_i^*\tilde{e}_a(i)||_2^2\,, \\
&= \quad \tilde{\mathbf{w}}_{i-1}^*\mathbf{P}_i^{-1}\tilde{\mathbf{w}}_{i-1} - \mu(i)e_a(i)\tilde{e}_a^*(i) - \mu(i)e_a^*(i)\tilde{e}_a(i) + \mu^2(i)\mathbf{u}_i\mathbf{P}_i\mathbf{u}_i^*\,|\tilde{e}_a(i)|^2\,.
\end{aligned}$$

If we now 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,

$$\tilde{\mathbf{w}}_i^* \mathbf{P}_i^{-1} \tilde{\mathbf{w}}_i + \mu(i) \, |e_a(i)|^2 + \mu(i) \, (\, 1 - \mu(i) \mathbf{u}_i \mathbf{P}_i \mathbf{u}_i^*) \, |\tilde{e}_a(i)|^2 = \tilde{\mathbf{w}}_{i-1}^* \mathbf{P}_i^{-1} \tilde{\mathbf{w}}_{i-1} + \mu(i) \, |v(i)|^2 \ . \tag{7}$$

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

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

This is an important equality that involves "energy" terms. In loose terms, it tells us how the weighted "energies" of the error signals $\tilde{\mathbf{w}}_i$ and $e_a(i)$ relate to the weighted energies of the noise $v(i)$ and the weight error $\tilde{\mathbf{w}}_{i-1}$. The implications of this observation will be detailed throughout our discussions, starting with the next section.

3.1 A Local Apriori-Based Passivity Relation

Expression (8) allows us to establish that the following error bounds are always satisfied for the Gauss-Newton recursion (2) – in the statement of the Lemma, we employ the quantity $\bar{\mu}(i) \triangleq (\mathbf{u}_i \mathbf{P}_i \mathbf{u}_i^*)^{-1}$.

LEMMA 3.1 (A LOCAL PASSIVITY RELATION). *Consider the Gauss-Newton recursion (2). It always holds that*

$$\frac{\tilde{\mathbf{w}}_i^* \mathbf{P}_i^{-1} \tilde{\mathbf{w}}_i + (\mu(i) - \beta(i)) \, |e_a(i)|^2}{\lambda(i) \tilde{\mathbf{w}}_{i-1}^* \mathbf{P}_{i-1}^{-1} \tilde{\mathbf{w}}_{i-1} + \mu(i) \, |v(i)|^2} \begin{cases} \leq 1 & \text{for} \quad 0 < \mu(i) < \bar{\mu}(i), \\ = 1 & \text{for} \quad \mu(i) = \bar{\mu}(i), \\ \geq 1 & \text{for} \quad \mu(i) > \bar{\mu}(i). \end{cases} \tag{9}$$

Such relations also arise in the case of instantaneous-gradient-based algorithms (i.e., algorithms that avoid the propagation of Riccati variables $\mathbf{P}_i$), as detailed in.[10]

The first two bounds in the above lemma admit an interesting interpretation that highlights a robustness property of the Gauss-Newton recursion (2). To clarify this, we assume that $\beta(i) \leq \mu(i)$ in order to guarantee $(\mu(i) - \beta(i)) \geq 0$ 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),

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

3.2 A Global Contraction Mapping

The relations of Lemma 3.1 are local conclusions but similar results also hold 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,

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

Summing over i we conclude that

$$\tilde{\mathbf{w}}_N^* \mathbf{P}_N^{-1} \tilde{\mathbf{w}}_N \ + \ \sum_{i=0}^{N} (\mu(i) - \beta(i)) \lambda^{[i+1,N]} \, |e_a(i)|^2 \ \leq \ \lambda^{[0,N]} \tilde{\mathbf{w}}_{-1}^* \mathbf{P}_{-1}^{-1} \tilde{\mathbf{w}}_{-1} \ + \ \sum_{i=0}^{N} \mu(i) \lambda^{[i+1,N]} |v(i)|^2, \tag{10}$$

which establishes a passivity relation over the interval $0 \le i \le N$. Here, we have used the notation $\lambda^{[i,j]} = \prod_{k=i}^{j} \lambda(k)$. Alternatively, if we denote by $\epsilon_N(\mathbf{w}_{-1}, v(\cdot))$ the difference between the left- and the right-hand sides of (10),

$$\epsilon_N(\mathbf{w}_{-1}, v(\cdot)) \triangleq \tag{11}$$

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

then we also conclude from the argument prior to (10) that we always have, for any $\mathbf{w}_{-1}$ and $v(\cdot)$,

$$\epsilon_N(\mathbf{w}_{-1}, v(\cdot)) \le 0, \tag{12}$$

as long as $\mu(i) \le \bar{\mu}(i)$. If we further have $\beta(i) \le \mu(i)$ then we can interpret the above result as establishing the existence of a contractive mapping from the signals

$$\left\{ \sqrt{\mu(\cdot)} \sqrt{\lambda^{[\cdot+1,N]}} v(\cdot), \; \sqrt{\lambda^{[0,N]}} \Pi_0^{-1/2} \tilde{\mathbf{w}}_{-1} \right\} \tag{13}$$

to the signals

$$\left\{ \sqrt{(\mu(\cdot) - \beta(\cdot))} \sqrt{\lambda^{[\cdot+1,N]}} e_a(\cdot), \mathbf{P}_N^{-1/2} \tilde{\mathbf{w}}_N \right\}. \tag{14}$$

The quantities in (13) involve the disturbances, i.e., noise and initial uncertainty in the guess for $\mathbf{w}$. The quantities in (14), on the other hand, involve the resulting estimation errors $e_a(\cdot)$ and the final weight-error $\tilde{\mathbf{w}}_N$. Consequently, the above statement establishes the following interesting fact: the Gauss-Newton algorithm (2), under the assumption $\beta(i) \le \mu(i) \le \bar{\mu}(i)$, always guarantees that the (weighted) error energy due the initial disturbances will *not* be magnified.

LEMMA 3.2 (A GLOBAL RELATION). *If $\mu(i) \le \bar{\mu}(i)$ over $0 \le i \le N$, then the Gauss-Newton algorithm (2) 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 (13) to (14).*

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

3.3 A Local Aposteriori-Based Passivity Relation

Note that the error bounds derived in Lemma 3.1 are in terms of the apriori estimation error $e_a(i)$. But other bounds can be derived as well in terms of the aposteriori error $e_p(i)$ (and, in fact, combinations thereof). It follows from (6) and (7) that the apriori and aposteriori error-energies are related via the expression

$$\bar{\mu}^{-1}(i) \left(\tilde{\mathbf{w}}_i^* \mathbf{P}_i^{-1} \tilde{\mathbf{w}}_i - \tilde{\mathbf{w}}_{i-1}^* \mathbf{P}_i^{-1} \tilde{\mathbf{w}}_{i-1} \right) = |e_p(i)|^2 - |e_a(i)|^2 .$$

If we substitute the recursion (4) for $\mathbf{P}_i$ we obtain

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

This relation suggests the following bound in terms of the aposteriori-estimation error:

$$\frac{\nu(i) \tilde{\mathbf{w}}_i^* \mathbf{P}_i^{-1} \tilde{\mathbf{w}}_i + \frac{\mu(i) - \beta(i)}{1 - \beta(i) \bar{\mu}^{-1}(i)} \, |e_p(i)|^2}{\nu(i) \lambda(i) \; \tilde{\mathbf{w}}_{i-1}^* \mathbf{P}_{i-1}^{-1} \tilde{\mathbf{w}}_{i-1} + \mu(i) \, |v(i)|^2} \le 1, \tag{15}$$

for $\mu(i) \leq \bar{\mu}(i)$ and where we have defined

$$\nu(i) \triangleq \frac{1-\mu(i)\bar{\mu}^{-1}(i)}{1-\beta(i)\bar{\mu}^{-1}(i)}\,. \tag{16}$$

This local relation can also be used to derive a global bound that is valid over an interval of time. Following the same argument prior to (10), we can sum over N terms and obtain

$$\frac{\tilde{\mathbf{w}}_N^* \mathbf{P}_N^{-1} \tilde{\mathbf{w}}_N + \sum_{i=0}^{N} \frac{\mu(i)-\beta(i)}{1-\mu(i)\bar{\mu}^{-1}(i)} \lambda^{[i+1,N]} \left|e_p(i)\right|^2}{\lambda^{[0,N]}\, \tilde{\mathbf{w}}_{-1}^* \Pi_0^{-1} \tilde{\mathbf{w}}_{-1} + \sum_{i=0}^{N} \mu(i)\nu^{-1}(i)\lambda^{[i+1,N]} \left|v(i)\right|^2} \leq 1. \tag{17}$$

4 THE FEEDBACK STRUCTURE

Before proceeding to a discussion of the feedback structure alluded to earlier, we shall first establish the following useful fact.

LEMMA 4.1 (A LOWER BOUND ON $\bar{\mu}(i)$). *Consider the Gauss-Newton algorithm (2) with the free 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

$$\bar{\mu}^{-1}(i) = \mathbf{u}_i \mathbf{P}_i \mathbf{u}_i^* \;=\; \frac{1}{\lambda(i)} \left(\mathbf{u}_i \mathbf{P}_{i-1} \mathbf{u}_i^* - \frac{(\mathbf{u}_i \mathbf{P}_{i-1} \mathbf{u}_i^*)^2}{\frac{\lambda(i)}{\beta(i)} + \mathbf{u}_i \mathbf{P}_{i-1} \mathbf{u}_i^*} \right) = \frac{1}{\beta(i) + \lambda(i)\bar{\mu}(i|i-1)}.$$

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 3.1 can be described via an alternative form that will lead us to an interesting feedback structure. To clarify this, we first show how to rewrite the Gauss-Newton recursion (2) in an alternative convenient form. We rewrite (6) as

$$e_p(i) = \left(1 - \frac{\mu(i)}{\bar{\mu}(i)}\right) e_a(i) - \frac{\mu(i)}{\bar{\mu}(i)} v(i), \tag{18}$$

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

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

That is,

$$\mathbf{w}_i = \mathbf{w}_{i-1} + \bar{\mu}(i)\mathbf{P}_i \mathbf{u}_i^* [e_a(i) - e_p(i)] \triangleq \mathbf{w}_{i-1} + \bar{\mu}(i)\mathbf{P}_i \mathbf{u}_i^* [e_a(i) + \bar{v}(i)]. \tag{19}$$

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 (9) to (19), we readily conclude that the following equality holds <u>for all</u> $\mu(i)$ and $v(i)$,

$$\frac{\tilde{\mathbf{w}}_i^* \mathbf{P}_i^{-1} \tilde{\mathbf{w}}_i + (\bar{\mu}(i) - \beta(i))\, |e_a(i)|^2}{\lambda(i) \tilde{\mathbf{w}}_{i-1}^* \mathbf{P}_{i-1}^{-1} \tilde{\mathbf{w}}_{i-1} + \bar{\mu}(i)\, |e_p(i)|^2} = 1\,. \tag{20}$$

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,

$$\bar{\mu}^{\frac{1}{2}}(i) \bar{v}(i) = \frac{\mu(i)}{\bar{\mu}^{\frac{1}{2}}(i)} v(i) - \left(1 - \frac{\mu(i)}{\bar{\mu}(i)}\right) \bar{\mu}^{\frac{1}{2}}(i) e_a(i)\,. \tag{21}$$

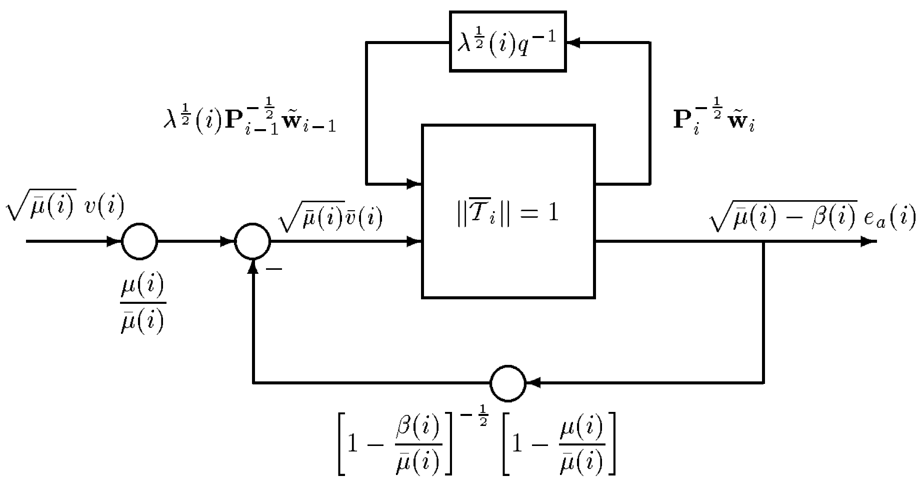

Figure 1: *A time-variant lossless mapping with gain feedback.*

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

$$\begin{bmatrix} [\bar{\mu}(i) - \beta(i)]^{\frac{1}{2}}\ e_a(i) \\ \mathbf{P}_i^{-\frac{1}{2}} \tilde{\mathbf{w}}_i \end{bmatrix} = \underbrace{\begin{bmatrix} \left[\frac{\bar{\mu}(i) - \beta(i)}{\lambda(i)}\right]^{\frac{1}{2}} \mathbf{u}_i \mathbf{P}_{i-1}^{\frac{1}{2}} & 0 \\ \mathbf{P}_i^{-\frac{1}{2}} \left[\mathbf{I} - \bar{\mu}(i) \mathbf{P}_i \mathbf{u}_i^* \mathbf{u}_i\right] \mathbf{P}_{i-1}^{\frac{1}{2}} \lambda^{-\frac{1}{2}}(i) & -\bar{\mu}^{\frac{1}{2}}(i) \mathbf{P}_i^{\frac{1}{2}} \mathbf{u}_i^* \end{bmatrix}}_{\overline{\mathcal{T}}_i} \begin{bmatrix} \lambda^{\frac{1}{2}}(i) \mathbf{P}_{i-1}^{-\frac{1}{2}} \tilde{\mathbf{w}}_{i-1} \\ \bar{\mu}^{\frac{1}{2}}(i)\ \bar{v}(i) \end{bmatrix}.$$

We collect the results of this section into a lemma.

LEMMA 4.2 (FEEDBACK REPRESENTATION). *Consider the Gauss-Newton 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 $\bar{\mathcal{T}}_i$ 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.

5 l_2−STABILITY AND THE SMALL GAIN THEOREM

The feedback configuration of Figure 1 lends itself rather immediately to stability analysis, as we now explain. It follows from the equality in Lemma 4.2 that for every time instant i, and for any $\mu(i)$, we have

$$(\bar{\mu}(i) - \beta(i))\,|e_a(i)|^2 = \lambda(i)\tilde{\mathbf{w}}_{i-1}^* \mathbf{P}_{i-1}^{-1} \tilde{\mathbf{w}}_{i-1} - \tilde{\mathbf{w}}_i^* \mathbf{P}_i^{-1} \tilde{\mathbf{w}}_i + \bar{\mu}(i)\,|\bar{v}(i)|^2\,. \tag{22}$$

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 (26) 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 would be desirable because it will then allow us to conclude the convergence of $e_a(\cdot)$ to zero.

For this purpose, assume we run the Gauss-Newton recursion (2) from time $i = 0$ up to time N. If we compute the sum of both sides of the above equality (22) we obtain,

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

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

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

Consequently,

$$\sqrt{\sum_{i=0}^{N} \lambda^{[i+1,N]}(\bar{\mu}(i) - \beta(i))\,|e_a(i)|^2} \ \leq\ \sqrt{\lambda^{[0,N]}\tilde{\mathbf{w}}_{-1}^* \mathbf{P}_{-1}^{-1} \tilde{\mathbf{w}}_{-1}} + \sqrt{\sum_{i=0}^{N} \lambda^{[i+1,N]}\bar{\mu}(i)\,|\bar{v}(i)|^2}. \tag{23}$$

But it follows from (21), 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

$$\begin{aligned}\sqrt{\sum_{i=0}^{N} \lambda^{[i+1,N]}(\bar{\mu}(i) - \beta(i))\,|e_a(i)|^2} \quad &\leq \quad \sqrt{\lambda^{[0,N]}\tilde{\mathbf{w}}_{-1}^* \mathbf{P}_{-1}^{-1} \tilde{\mathbf{w}}_{-1}} + \sqrt{\sum_{i=0}^{N} \lambda^{[i+1,N]}\frac{\mu^2(i)}{\bar{\mu}(i)}|v(i)|^2} \\ &\quad + \sqrt{\sum_{i=0}^{N} \lambda^{[i+1,N]}\left|1 - \frac{\mu(i)}{\bar{\mu}(i)}\right|^2 \bar{\mu}(i)|e_a(i)|^2}\ .\end{aligned} \tag{24}$$

Define

$$\Delta(N) = \max_{0\leq i\leq N} \left| \frac{1 - \frac{\mu(i)}{\bar{\mu}(i)}}{\sqrt{1 - \frac{\beta(i)}{\bar{\mu}(i)}}} \right| \quad \text{and} \quad \gamma(N) = \max_{0\leq i\leq N}\ \frac{\mu(i)}{\bar{\mu}(i)}\ . \tag{25}$$

It then follows that

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

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

$$\sqrt{\sum_{i=0}^{N} \lambda^{[i+1,N]}(\bar{\mu}(i)-\beta(i))\,|e_a(i)|^2} \le \frac{1}{1-\Delta(N)} \left[\sqrt{\lambda^{[0,N]}\tilde{\mathbf{w}}_{-1}\mathbf{P}_{-1}^{-1}\tilde{\mathbf{w}}_{-1}} + \gamma(N)\sqrt{\sum_{i=0}^{N} \lambda^{[i+1,N]}\bar{\mu}(i)|v(i)|^2} \right], \tag{26}$$

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.[5,6] In simple terms, the theorem states that the l_2-stability of a feedback configuration (that includes Figure 1 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, for all i,

$$0 < \mu(i) < \bar{\mu}(i)\left(1+\sqrt{1-\frac{\beta(i)}{\bar{\mu}(i)}}\right). \tag{27}$$

THEOREM 5.1 (l_2-STABILITY). *Consider the Gauss-Newton recursion (2) and define $\Delta(N)$ and $\gamma(N)$ as in (25) and also*

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

If (27) holds then the map from

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

is l_2-stable in the sense of (26). Moreover, if $\beta(i) \le \mu(i)$ then it also holds that the map from

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

(i.e., with $\bar{\mu}(\cdot)$ replaced by $\mu(\cdot)$) is also l_2-stable in the following sense:

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

Proof: The proof of the first bound was provided prior to the statement of the theorem. As for the second bound, we first note that (24) implies

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

Now (28) follows by noting that

$$\begin{aligned}\sum_{i=0}^{N} \lambda^{[i+1,N]}(\mu(i)-\beta(i))|e_a(i)|^2 \quad &= \quad \sum_{i=0}^{N} \lambda^{[i+1,N]}\frac{\mu(i)-\beta(i)}{\bar{\mu}(i)-\beta(i)}\;(\bar{\mu}(i)-\beta(i))|e_a(i)|^2 \\ &\le \quad \tilde{\gamma}(N)\sum_{i=0}^{N} \lambda^{[i+1,N]}(\bar{\mu}(i)-\beta(i))|e_a(i)|^2.\end{aligned}$$

■

In fact, a stronger upper bound than (28) 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 3.2.

LEMMA 5.2 (A TIGHTER BOUND). *Consider the Gauss-Newton recursion (2). If $0 < \beta(i) \leq \mu(i) \leq \bar{\mu}(i)$ then a tighter bound is the following:*

$$\sqrt{\sum_{i=0}^{N} \lambda^{[i+1,N]}(\mu(i) - \beta(i))\, |e_a(i)|^2} \leq \left[\sqrt{\lambda^{[0,N]}\tilde{\mathbf{w}}_{-1}^{*}\mathbf{P}_{-1}^{-1}\tilde{\mathbf{w}}_{-1}} + \sqrt{\sum_{i=0}^{N} \lambda^{[i+1,N]}\mu(i)|v(i)|^2} \right]. \tag{29}$$

The fact that the bound in (28) is valid even for $\mu(i)$ in the interval $\bar{\mu}(i) \leq \mu(i) < \bar{\mu}(i)\left(1 + (1 - \beta(i)/\bar{\mu}(i))^{\frac{1}{2}}\right) < 2\bar{\mu}(i)$ suggests that a local bound, along the lines of (9), should also exist for this interval. In fact, this is also the case as stated below – the proof is omitted for brevity.

LEMMA 5.3 (A FINER LOCAL BOUND). *Consider the Gauss-Newton recursion (2). Then the following bounds always hold:*

$$\begin{array}{lll} 0 \;\leq\; \dfrac{\tilde{\mathbf{w}}_i^{*}\mathbf{P}_i^{-1}\tilde{\mathbf{w}}_i + \mu(i)\,|e_a(i)|^2}{\tilde{\mathbf{w}}_{i-1}^{*}\mathbf{P}_i^{-1}\tilde{\mathbf{w}}_{i-1} + \mu(i)\,|v(i)|^2} & \leq 1 & \textit{for } 0 < \mu(i) \leq \bar{\mu}(i)\,, \\ 1 \;\leq\; \dfrac{\tilde{\mathbf{w}}_i^{*}\mathbf{P}_i^{-1}\tilde{\mathbf{w}}_i + \mu(i)\,|e_a(i)|^2}{\tilde{\mathbf{w}}_{i-1}^{*}\mathbf{P}_i^{-1}\tilde{\mathbf{w}}_{i-1} + \mu(i)\,|v(i)|^2} & \leq \dfrac{\mu(i)}{2\bar{\mu}(i) - \mu(i)} & \textit{for } \bar{\mu}(i) \leq \mu(i) < 2\bar{\mu}(i)\,. \end{array}$$

6 ON CONVERGENCE AND ENERGY PROPAGATION

In order to further appreciate the significance of the bounds of Theorem 5.1, we now exhibit a convergence result that follows as a consequence of the l_2−stability property. Indeed, assume that the normalized noise sequence $\{\sqrt{[\lambda^{[\cdot+1,N]}\mu(\cdot)}\; v(\cdot)\}$ has finite energy, i.e., $\sum_{i=0}^{\infty} \lambda^{[i+1,N]}\mu(i)|v(i)|^2 < \infty$. It then follows from (28) 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 $0 < \gamma(N) < 2$, $0 < \tilde{\gamma}(N) < 2$ 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,

$$\lim_{i\to\infty} \sqrt{\lambda^{[i+1,N]}(\mu(i) - \beta(i))}\; e_a(i) = 0. \tag{30}$$

If a persistence of excitation condition is further imposed on the input vectors $\mathbf{u}_i$, then we can also conclude convergence of the weight vector, i.e., $\lim_{i\to\infty} \mathbf{w}_i = \mathbf{w}$. Details will be provided elsewhere. But here, we would like to note that more physical insights into the convergence behaviour of the Gauss-Newton recursion (2) can be obtained by studying the energy flow through the feedback configuration of Figure 1, as shown in the next section.

6.1 Energy propagation and convergence speed

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 for now ignore the measurement noise $v(i)$ and assume that we have noiseless measurements $d(i) = \mathbf{u}_i\mathbf{w}$. The weight-error update equation (5) can then be easily seen to collapse to

$$\tilde{\mathbf{w}}_i = [\mathbf{I} - \mu(i)\mathbf{P}_i\mathbf{u}_i^{*}\mathbf{u}_i]\, \tilde{\mathbf{w}}_{i-1}, \tag{31}$$

where the $M \times M$ coefficient matrix $[\mathbf{I} - \mu(i)\mathbf{P}_i\mathbf{u}_i^*\mathbf{u}_i]$ is simply a rank one modification of the identity matrix. 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 for gradient-type algorithms is obtained for $\mu(i) = \bar{\mu}(i)$, i.e., for the so-called projection LMS algorithm. Gauss-Newton-type algorithms are known to uncorrelate the input process, so that the convergence speed no longer depends on the excitation process.[13] 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 $e_a(\cdot)$. To understand the implications of this fact, let us study the energy flow through the system as time progresses. 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,

$$E_w(i) = \lambda(i)E_w(i-1) - E_e(i), \tag{32}$$

where we are denoting by $E_e(i)$ the energy of $\sqrt{\bar{\mu}(i) - \beta(i)}\, e_a(i)$ and by $E_w(i)$ the energy of $\mathbf{P}_i^{-\frac{1}{2}}\tilde{\mathbf{w}}_i$,

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

Expression (32) 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 $E_e(i) \neq 0$. This is in general the case especially when the input vectors $\mathbf{u}_i$ are assumed persistently exciting. This means that they are rich enough so as to avoid situations of the form $\mathbf{u}_i\tilde{\mathbf{w}}_{i-1} = 0$ (viz., $\mathbf{u}_i$ orthogonal to $\tilde{\mathbf{w}}_{i-1}$). Under this condition, the weight-error energy is guaranteed to decrease with time, thus tending to zero and we obtain $\mathbf{w}_i \to \mathbf{w}$. Note also that the so-called forgetting factor $\lambda(i)$ plays an important role. The smaller the $\lambda(i)$ the faster is the convergence of the algorithm.

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 $\mu(i) \neq \bar{\mu}(i)$, the feedback path is connected and, therefore, 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 (due to energy conservation):

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

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,

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

where we have defined the coefficient $\tau(i)$ (compare with (32)). 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.

Finally, what if the measurement noise $v(\cdot)$ is nonzero? In a deterministic setting, the samples $v(\cdot)$ can assume any values. In particular, we can envision a noise sequence that happens to assume the special value

$v(i) = -e_a(i)$. In this case, the update relation (5) collapses to $\tilde{\mathbf{w}}_i = \tilde{\mathbf{w}}_{i-1}$ and, hence, $\mathbf{w}_i = \mathbf{w}_{-1}$ for all time instants i. This means that no improvements over the initial guess are obtained and, consequently, convergence will never be attained if $\mathbf{w}_{-1} \neq \mathbf{w}$. In other words, no sensible statements can be made about the convergence of the algorithm if no restrictions are imposed on the noise sequence $v(\cdot)$. However, it is known, form theoretical as well as experimental results for stochastic noise sequences, that the noise does not affect the rate of convergence but rather the steady-state value. This is consistent with the feedback configuration of Figure 1 where it is clear that the fine structure of the feedforward and the feedback paths are independent of the specific values of the noise sequence; it only depends on $\{\mu(i), \lambda(i), \beta(i)\}$ and $\mathbf{u}_i$.

7 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, since it involves a filtered version of the apriori error. It can be shown that such variants also admit a feedback structure of the form derived in this paper, except that the feedback loop is now dynamic (i.e., not memoryless). The analysis of this paper can also be shown to suggest a procedure for computing optimal step-sizes in order to guarantee stability and faster convergence. These facts will be detailed elsewhere – though see[14–16] for related discussion and for connections with results in $H^\infty-$theory.

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

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

[7] R. Isermann. Digital Control Systems. Springer, second edition, 1991.

[8] S. Haykin. Adaptive Filter Theory. Englewood Cliffs, NJ: Prentice–Hall, second edition, 1991.

[9] 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.

[10] A.H. Sayed and M. Rupp. "On the robustness, convergence, and minimax performance of instantaneous-gradient adaptive filters." Proc. Asilomar Conference on Signals, Systems, and Computers, pp. 592–596, 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] B. Widrow and S.D. Stearns. Adaptive Signal Processing. Englewood Cliffs, NJ: Prentice–Hall, 1985.

[14] 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.

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

[16] 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, (this conference), San Diego, CA, July 1995.