ELSEVIER

Signal Processing 50 (1996) 165–187

SIGNAL PROCESSING

Robustness of Gauss–Newton recursive methods: A deterministic feedback analysis [1]

Markus Rupp, Ali H. Sayed*

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

Received 2 March 1995; revised 20 October 1995

Abstract

This paper provides a time-domain feedback analysis of the robustness performance of Gauss–Newton recursive methods that are often used in identification and control. These are recursive estimators that also involve updates of sample covariance matrices. Several free parameters are included in the filter descriptions while combining the covariance updates with the weight-vector updates. One of the contributions of this work is to show that by properly selecting the free parameters, the resulting filters can be made to impose certain bounds on the error quantities, thus resulting in desirable robustness properties (along the lines of H_∞ filter designs). It is also shown that an intrinsic feedback structure, mapping the noise sequence and the initial weight error to the a priori 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. Emphasis is further given to filtered-error variants that give rise to dynamic time-variant feedback loops rather than memoryless loops. Such variants arise in IIR modeling.

Zusammenfassung

Dieser Artikel präsentiert eine Robustheitsanalyse im Zeitbereich von rekursiven Gauss–Newton Verfahren wie sie oft bei der Systemidentifikation und im regelungstechnischen Bereich auftreten. Es handelt sich hierbei um rekursive Schätzer, die implizit die Zeitkovarianzmatrix des Anregungssignals invertieren. Die Filterbeschreibung beinhaltet verschiedene freie Parameter sowohl für die Bestimmung der Parameterschätzwerte als auch der Zeitkovarianzmatrizen. Ein Beitrag dieser Arbeit besteht darin zu zeigen, daß bei sorgfältig gewählten freien Parametern die Fehlergrößen des resultierenden Filters beschränkt bleiben, wodurch die gewünschte Eigenschaft der Robustheit (gemäß der H_∞ Terminologie) erzielt wird. Außerdem wird gezeigt, daß für diese Schätzverfahren stets eine rückgekoppelte Struktur existiert, welche die Störsequenz und den Initialschätzfehler auf die a priori Fehler abbildet. Die Rückkopplungsstruktur wird durch Energiegrößen motiviert und besteht aus zwei Blöcken: einem zeitvarianten, verlustfreien (also energieerhaltenden) Vorwärtspfad und einem zeitvarianten Rückwärtspfad. Ferner werden Algorithmusvarianten mit gefiltertem Fehler beschrieben, die statt zu einer gedächtnisfreien zu einer dynamischen Rückkopplung führen. Solche Varianten treten bei der Modellierung von IIR Filtern auf.

* Corresponding author. Fax: (805)893-3262; e-mail: sayed@ece.ucsb.edu.

[1] This material was 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 was also supported by a scholarship from DAAD (German Academic Exchange Service) as well as the scientifical division of NATO (Wissenschaftsausschuß der NATO).

PII: S0165-1684(96)00022-9

Résumé

Cet article fournit une analyse par rétroaction temporelle des performances en termes de robustesse des méthodes récursives de type Gauss–Newton qui sont souvent utilisées en identification et contrôle. Ce sont des estimateurs récursifs qui impliquent également des mises à jour des matrices de covariance obtenues à partir des échantillons. Plusieurs paramètres libres sont inclus dans les descriptions des filtres lorsque les mises à jour de la covariance sont combinées avec les mises à jour du vecteur de coefficients. Une des contributions de ce travail est de montrer que par une sélection appropriée des paramètres libres, les filtres résultants peuvent être capables d'imposer certaines limites sur les erreurs, ce qui donne comme résultat des propriétés souhaitables de robustesse (selon la terminologie des filtres H_∞). Il est également montré qu'une structure de rétroaction intrinsèque, faisant correspondre à la séquence de bruit et à l'erreur initial sur les coefficients de pondération les erreurs d'estimation a priori et l'erreur finale sur les coefficients de pondération, peut être associées à de tels algorithmes. La configuration de rétroaction est justifiée par des arguments sur l'énergie et est montrée consister en deux blocs principaux: un chemin direct variant dans le temps et sans perte (conservant l'énergie) et un chemin de rétroaction variant dans le temps. L'accent est également mis sur les variantes à erreur filtrée qui donnent lieu à des boucles de rétroaction variant dans le temps dynamiques plutôt qu'à des boucles sans mémoire. De telles variantes apparaissent en modélisation IIR.

Keywords: Adaptive Gauss–Newton filters; Filtered-error algorithms; Feedback connection; l_2-stability; The small gain theorem; Contraction mapping; Passivity relations; Error bounds

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., [6, 13, 16, 17, 19, 31]). These are recursive methods that are based on gradient-descent ideas and employ sample covariance matrices to control the update directions. Their descriptions involve two update relations: one for the update of the weight estimate and the other for the update of the inverse of the sample covariance matrix. An important special case is the exponentially weighted recursive-least-squares (RLS) algorithm (e.g., [9, 24]), which employs a decaying exponential weighting in the covariance update. But other possibilities exist and in this paper several free parameters are included in the filter descriptions while combining the covariance updates with the weight-vector updates. These parameters allow for a reasonable degree of freedom in setting up filter configurations, and one of the contributions of this work is to show that by properly selecting the free parameters, the resulting filters 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_∞ filtering [7, 8, 23, 29].

Intuitively, a robust filter is one for which the estimation errors are consistent with the disturbances in the sense that 'small' disturbances would lead to 'small' estimation errors, no matter what the disturbances are. This is not generally true for any adaptive filter: the estimation errors can still be 'large' even in the presence of small disturbances (e.g., RLS algorithms).

The robustness issue is addressed here in a purely deterministic framework and without assuming prior knowledge of noise statistics. This is especially useful in situations where prior statistical information is missing since a robust design would guarantee a desired level of robustness independent of the noise statistics. In loose terms, by robustness we mean that the ratio of the estimation error energy to the noise or disturbance energy will be guaranteed to be upper bounded by a positive constant, say the constant one,

$$\frac{\text{estimation error energy}}{\text{disturbance energy}} \leqslant 1. \tag{1}$$

The disturbance energy may include the energies of the measurement noise, modeling uncertainties, errors in initial weight guess, etc. From a practical point of view, a relation of the form (1) is desirable since it guarantees that the resulting estimation error energy will be upper bounded by the disturbance energy, no matter what the nature and the statistics of the disturbances are. In this case, the algorithm will not unnecessarily magnify the disturbance energy and, consequently, small estimation errors will result when small disturbances occur. One of the contributions of this work is to show how to select the parameters that define a Gauss–Newton update in order to guarantee such robust behavior. This can be contrasted with results in stochastic settings where stability and convergence statements are often given in the mean and mean-square sense. In such settings, even for the simple LMS algorithm, a constant step-size μ that is bounded by twice the inverse of the maximal eigenvalue of the autocorrelation matrix can still lead to blow up in a practical experiment (see, e.g., [21]).

To address the above issues, the paper suggests a time-domain approach that proves to be useful in both the analysis and design of robust estimators. It highlights and exploits an intrinsic feedback structure that can be associated with Gauss–Newton schemes, mapping the noise sequence and the initial weight error to the a priori estimation errors and the final weight error. 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.

Although the feedback nature of these, and related recursive schemes, has been pointed out and advantageously exploited in earlier places in the literature (see, e.g., the discussions in [13, 17]), 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, 14, 18]), but is instead motivated via energy arguments that also *explicitly* take into consideration *both* the effect of the measurement *noise* and the effect of the uncertainty in the *initial guess* for the weight vector. These extensions are incorporated into the feedback arguments of this paper because the derivation here is interested in a formal study of the robustness properties of Gauss–Newton schemes in the presence of uncertain disturbances. This is especially useful when the statistical properties of the disturbances are unknown.

In this regard, the feedback connection provided herein is shown to exhibit three main features that distinguish it from earlier studies in the literature: the feedforward path in the connection consists of a *lossless* (i.e., energy preserving) mapping while the feedback path consists either of a *memoryless* interconnection or, in the case of filtered-error variants, of a *dynamic* system that is dependent on the error filter. 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 itself rather immediately to stability analysis via a so-called small gain theorem, which is a standard tool in system theory (e.g., [11, 32]); it provides contractivity conditions that are shown to guarantee the l_2-stability of the algorithms, with further implications on the convergence behavior 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.

Emphasis is also given to filtered-error variants that are useful in IIR modeling and in noise control applications (e.g., [2, 5, 12, 15]). Such variants give rise to dynamic time-variant feedback loops and a procedure for computing optimal step-sizes in order to guarantee robustness and improved speeds of convergence is suggested. Simulation results are included to support the theoretical conclusions.

1.1. Notation

We use lower case boldface letters to denote vectors and capital boldface letters to denote matrices. Also, the symbol '$*$' denotes Hermitian conjugation (complex conjugation for scalars). The symbol $\boldsymbol{I}$ denotes the identity matrix of appropriate dimensions, and the boldface letter $\mathbf{0}$ denotes either a zero vector or a zero matrix. The notation $\|\boldsymbol{x}\|^2$ denotes the squared Euclidean norm of a vector, and $\boldsymbol{A}^{1/2}$ denotes a square-root factor of $\boldsymbol{A}$, viz., any matrix satisfying $\boldsymbol{A}^{1/2}\boldsymbol{A}^{*/2} = \boldsymbol{A}$. For convenience, we shall also write $(\boldsymbol{A}^{1/2})^* = \boldsymbol{A}^{*/2}$, $(\boldsymbol{A}^{1/2})^{-1} = \boldsymbol{A}^{-1/2}$, $(\boldsymbol{A}^{-1/2})^* = \boldsymbol{A}^{-*/2}$. Thus note the expressions $\boldsymbol{A} = \boldsymbol{A}^{1/2}\boldsymbol{A}^{*/2}$ and $\boldsymbol{A}^{-1} = \boldsymbol{A}^{-*/2}\boldsymbol{A}^{-1/2}$. Also, all vectors are column vectors except for the so-called input data vector denoted by $\boldsymbol{u}_i$, which is taken to be a row vector.

2. Gauss–Newton recursive methods

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., [10, 16, 17, 31]). Here we only wish to briefly review this class of algorithms before proceeding into a closer analysis of their behavior.

We consider a collection of noisy measurements $\{d(i)\}_{i=0}^{N}$ that arise from a linear model of the form (we use subscripts for indexing of vector quantities and parenthesis for indexing of scalar quantities)

$$d(i) = \boldsymbol{u}_i \boldsymbol{w} + v(i), \tag{2}$$

where $v(i)$ denotes measurement noise or disturbance (or modeling errors/uncertainties) and $\boldsymbol{u}_i$ denotes a row input vector. The column vector $\boldsymbol{w}$ consists of unknown parameters that we wish to estimate. In this paper we focus on the following so-called Gauss–Newton recursive estimator.

Algorithm 2.1 (*Gauss–Newton procedure*). *Given measurements* $\{d(i)\}_{i=0}^{N}$, *an initial guess* $\boldsymbol{w}_{-1}$, *and a positive-definite matrix* Π_0, *recursive estimates of the weight vector* $\boldsymbol{w}$ *are obtained as follows*:

$$\boldsymbol{w}_i = \boldsymbol{w}_{i-1} + \mu(i)\, \boldsymbol{P}_i \boldsymbol{u}_i^* (d(i) - \boldsymbol{u}_i \boldsymbol{w}_{i-1}), \tag{3}$$

where $\boldsymbol{P}_i$ *satisfies the* (*Riccati*) *update*

$$\boldsymbol{P}_i = \frac{1}{\lambda(i)} \left(\boldsymbol{P}_{i-1} - \frac{\boldsymbol{P}_{i-1}\boldsymbol{u}_i^* \boldsymbol{u}_i \boldsymbol{P}_{i-1}}{\frac{\lambda(i)}{\beta(i)} + \boldsymbol{u}_i \boldsymbol{P}_{i-1} \boldsymbol{u}_i^*} \right), \quad \boldsymbol{P}_{-1} = \Pi_0, \tag{4}$$

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

The effect of the coefficients $\{\lambda(i), \mu(i), \beta(i)\}$ on the performance of the algorithm (3)–(4) will be studied in later sections. In particular, conditions will be derived for choosing these parameters in order to guarantee that small disturbance energies will necessarily result in small estimation error energies (and, hence, in robust filters).

Note that by applying the matrix inversion formula (e.g., [9]) to (4) we obtain that the inverse of $\boldsymbol{P}_i$ satisfies the simple time-update

$$\boldsymbol{P}_i^{-1} = \lambda(i)\, \boldsymbol{P}_{i-1}^{-1} + \beta(i)\, \boldsymbol{u}_i^* \boldsymbol{u}_i. \tag{5}$$

This establishes that $\boldsymbol{P}_i$ 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 (3) is the so-called recursive-least-squares (RLS) algorithm (see, e.g., [9, 24]), which corresponds to the choices $\beta(i) = \mu(i) = 1$ and $\lambda(i) = \lambda = cte$. In this case, the Riccati recursion (4) reduces to

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

which also implies that

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

Using this last equality in (3) leads to the update equation

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

which is the standard form of the RLS algorithm.

2.2. Error signals

The difference $[d(i) - \boldsymbol{u}_i \boldsymbol{w}_{i-1}]$ in (3) will be denoted by $\tilde{e}_{\mathrm{a}}(i)$ and will be referred to as the *output estimation error*. The following error measures will be useful for our later analysis: $\tilde{\boldsymbol{w}}_i$ denotes the difference between the true weight $\boldsymbol{w}$ and its estimate $\boldsymbol{w}_i$, $\tilde{\boldsymbol{w}}_i = \boldsymbol{w} - \boldsymbol{w}_i$, $e_{\mathrm{a}}(i)$ denotes the *a priori estimation error*, $e_{\mathrm{a}}(i) = \boldsymbol{u}_i \tilde{\boldsymbol{w}}_{i-1}$, and $e_{\mathrm{p}}(i)$ denotes the *a posteriori estimation error*, $e_{\mathrm{p}}(i) = \boldsymbol{u}_i \tilde{\boldsymbol{w}}_i$.

It follows from the update equation (3) that the weight-error vector $\tilde{\boldsymbol{w}}_{i-1}$ satisfies the recursive equation:

$$\tilde{\boldsymbol{w}}_i = \tilde{\boldsymbol{w}}_{i-1} - \mu(i) \boldsymbol{P}_i \boldsymbol{u}_i^* \tilde{e}_{\mathrm{a}}(i). \tag{6}$$

It is also straightforward to verify that the a priori estimation error, $e_{\mathrm{a}}(i)$, and the output estimation error, $\tilde{e}_{\mathrm{a}}(i)$, differ by the disturbance $v(i)$, i.e., $\tilde{e}_{\mathrm{a}}(i) = e_{\mathrm{a}}(i) + v(i)$. Hence, the update relation (3) is also equivalent to

$$\boldsymbol{w}_i = \boldsymbol{w}_{i-1} + \mu(i)\, \boldsymbol{P}_i \boldsymbol{u}_i^* \, [e_{\mathrm{a}}(i) + v(i)]. \tag{7}$$

If we further multiply (6) by $\boldsymbol{u}_i$ from the left we obtain the following relation (used later in (21)) between $e_{\mathrm{p}}(i)$, $e_{\mathrm{a}}(i)$ and $v(i)$,

$$e_{\mathrm{p}}(i) = \left[1 - \mu(i) \boldsymbol{u}_i \boldsymbol{P}_i \boldsymbol{u}_i^*\right] e_{\mathrm{a}}(i) - \mu(i) \boldsymbol{u}_i \boldsymbol{P}_i \boldsymbol{u}_i^* \, v(i). \tag{8}$$

3. A time-domain analysis

We now pursue a closer analysis of the Gauss–Newton recursion (3) in order to highlight an important feedback structure that is implied by the algorithm. This structure plays an important role in our discussions and serves as a basis for the robustness analysis provided herein.

Our purpose in this and the following two sections is threefold:
1. to show that certain local and global energy (passivity) relations can be associated with the Gauss–Newton recursion (3);
2. to employ these relations in order to highlight a feedback structure;
3. to derive conditions on the parameters $\{\lambda(i), \mu(i), \beta(i)\}$ in order to guarantee that the feedback structure will behave as a robust filter.

3.1. Local passivity relations

To begin with, the following result is immediate from the time-domain update recursion (6) – see Appendix A for a proof.

Lemma 1 (Energy Relation). *The following equality holds for all i:*

$$\tilde{\boldsymbol{w}}_i^* \boldsymbol{P}_i^{-1} \tilde{\boldsymbol{w}}_i + [\mu(i) - \beta(i)]\, |e_{\mathrm{a}}(i)|^2$$
$$+\mu(i)\left[\,1 - \mu(i)\boldsymbol{u}_i \boldsymbol{P}_i \boldsymbol{u}_i^*\right]\, |\tilde{e}_a(i)|^2$$
$$= \lambda(i)\tilde{\boldsymbol{w}}_{i-1}^* \boldsymbol{P}_{i-1}^{-1} \tilde{\boldsymbol{w}}_{i-1} + \mu(i)\, |v(i)|^2 . \tag{9}$$

Expression (9) relates five 'energy' terms:
1. The weighted energy of the current weight error $\tilde{\boldsymbol{w}}_i$, namely $\tilde{\boldsymbol{w}}_i^* \boldsymbol{P}_i^{-1} \tilde{\boldsymbol{w}}_i$, where the weighting is with respect to the inverse of $\boldsymbol{P}_i$.
2. The weighted energy of the previous weight error $\tilde{\boldsymbol{w}}_{i-1}$, namely $\lambda(i)\tilde{\boldsymbol{w}}_{i-1}^* \boldsymbol{P}_{i-1}^{-1} \tilde{\boldsymbol{w}}_{i-1}$, where the weighting is with respect to the inverse of $\lambda^{-1}(i)\boldsymbol{P}_i$.
3. The weighted disturbance energy, $\mu(i)\,|v(i)|^2$.
4. The weighted a priori error energy $[\mu(i) - \beta(i)]\,|e_{\mathrm{a}}(i)|^2$ (assuming $\mu(i) \geqslant \beta(i)$; more on this condition later).
5. The weighted output-error energy, $\mu(i)[1 - \mu(i)\boldsymbol{u}_i \boldsymbol{P}_i \boldsymbol{u}_i^*]|\tilde{e}_{\mathrm{a}}(i)|^2$.

Eq. (9) therefore allows us to study how the energies of the error terms propagate as the algorithm progresses. In particular, it follows from (9) that the following energy sum (which appears on the left-hand side),

$$\tilde{\boldsymbol{w}}_i^* \boldsymbol{P}_i^{-1} \tilde{\boldsymbol{w}}_i + [\mu(i) - \beta(i)]\, |e_{\mathrm{a}}(i)|^2 ,$$

can be larger, smaller, or equal to the energy sum on the right-hand side,

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

depending on whether the term $\mu(i)\left[\,1 - \mu(i)\boldsymbol{u}_i \boldsymbol{P}_i \boldsymbol{u}_i^*\right] |\tilde{e}_a(i)|^2$ is negative, positive or zero, respectively. This is summarized in the following statement where we have introduced the notation $\bar{\mu}(i) = (\boldsymbol{u}_i \boldsymbol{P}_i \boldsymbol{u}_i^*)^{-1}$ (the inverse of the weighted energy of the input data).

Lemma 2 (A local passivity relation). *Consider the Gauss–Newton recursion* (3). *It always holds that*

$$\frac{\tilde{\boldsymbol{w}}_i^* \boldsymbol{P}_i^{-1} \tilde{\boldsymbol{w}}_i + (\mu(i) - \beta(i))\, |e_{\mathrm{a}}(i)|^2}{\lambda(i)\tilde{\boldsymbol{w}}_{i-1}^* \boldsymbol{P}_{i-1}^{-1} \tilde{\boldsymbol{w}}_{i-1} + \mu(i)\, |v(i)|^2}
\begin{cases} \leqslant 1 & \textit{for } 0 < \mu(i) < \bar{\mu}(i), \\ = 1 & \textit{for } \mu(i) = \bar{\mu}(i), \\ \geqslant 1 & \textit{for } \mu(i) > \bar{\mu}(i). \end{cases} \tag{10}$$

(Such relations also arise in the case of instantaneous-gradient-based algorithms (i.e., algorithms that avoid the propagation of Riccati variables $\boldsymbol{P}_i$, as detailed in [25, 28].)

Interpretation. The first two bounds in the above lemma admit an interesting interpretation that highlights a robustness property of the Gauss–Newton recursion (3). To clarify this, assume that $\beta(i) \leqslant \mu(i)$ in order to guarantee that the factor $(\mu(i) - \beta(i))\, |e_{\mathrm{a}}(i)|^2$ can be interpreted as an energy term.

In this case, we can interpret the first two bounds in the lemma to state the following: if the adaptation gain $\mu(i)$ is chosen according to $\mu(i) \leqslant \bar{\mu}(i)$, then no matter what the value of the denominator in (10) is, the numerator will always be smaller than the denominator. Intuitively, the denominator consists of the sums of the 'disturbance' energies, namely noise and previous weight error, while the numerator consists of the sums of the energies of the current weight error and

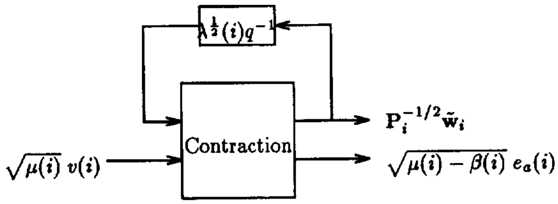

Fig. 1. A local contraction mapping.

the a priori error,

$$\tilde{\boldsymbol{w}}_i^* \boldsymbol{P}_i^{-1} \tilde{\boldsymbol{w}}_i + (\mu(i) - \beta(i))\, |e_a(i)|^2 \leqslant \lambda(i) \tilde{\boldsymbol{w}}_{i-1}^* \boldsymbol{P}_{i-1}^{-1} \tilde{\boldsymbol{w}}_{i-1} + \mu(i)\, |v(i)|^2 \quad \text{if } \mu(i) \leqslant \bar{\mu}(i). \tag{11}$$

This is a desirable robustness property in the sense that it guarantees that if the 'disturbance' energy is small then the resulting estimation error energies will accordingly be small (more on this in the next section). It establishes, as shown in Fig. 1, a contractive map from

$$\{\sqrt{\lambda(i)}\boldsymbol{P}_{i-1}^{-1/2} \tilde{\boldsymbol{w}}_{i-1}, \sqrt{\mu(i)} v(i)\} \tag{12}$$

to

$$\{\boldsymbol{P}_i^{-1/2} \tilde{\boldsymbol{w}}_i, \sqrt{\mu(i) - \beta(i)} e_a(i)\}. \tag{13}$$

(In simple terms, a map that takes x to y, say $y = T[x]$, is said to be contractive if for all x we have $\|T[x]\|^2 \leqslant \|x\|^2$. That is, the output energy does not exceed the input energy). The symbol q^{-1} in Fig. 1 denotes a unique delay.

3.2. A global contraction mapping

Since the contractivity relation (11) holds for each time instant i, it should also hold globally over an interval of time. Indeed, assuming $\mu(i) \leqslant \bar{\mu}(i)$ over $0 \leqslant i \leqslant N$, it follows from (11) that

$$\tilde{\boldsymbol{w}}_N^* \boldsymbol{P}_N^{-1} \tilde{\boldsymbol{w}}_N + \sum_{i=0}^{N} (\mu(i) - \beta(i)) \lambda^{[i+1,N]}\, |e_a(i)|^2 \leqslant \lambda^{[0,N]} \tilde{\boldsymbol{w}}_{-1}^* \boldsymbol{P}_{-1}^{-1} \tilde{\boldsymbol{w}}_{-1} + \sum_{i=0}^{N} \mu(i) \lambda^{[i+1,N]} |v(i)|^2, \tag{14}$$

where we have employed the notation $\lambda^{[i,j]} = \Pi_{k=i}^{j} \lambda(k)$, with $\lambda^{[N+1,N]} = 1$.

This relation has essentially the same structure as the one in (11), except for the additional scaling of the terms $|e_a(i)|^2$ and $|v(i)|^2$ by $\lambda^{[i+1,N]} = \lambda(i+1)\lambda(i+2)\cdots\lambda(N)$. It can as well be interpreted as establishing a contractive mapping from the signals

$$\left\{\sqrt{\lambda^{[0,N]}} \Pi_0^{-1/2} \tilde{\boldsymbol{w}}_{-1}, \sqrt{\mu(0)}\sqrt{\lambda^{[1,N]}} v(0), \ldots, \sqrt{\mu(N)} v(N)\right\} \tag{15}$$

to the signals

$$\left\{\sqrt{(\mu(0) - \beta(0))}\sqrt{\lambda^{[1,N]}} e_a(0), \ldots, \sqrt{(\mu(N) - \beta(N))} e_a(N), \boldsymbol{P}_N^{-1/2} \tilde{\boldsymbol{w}}_N\right\}. \tag{16}$$

In other words, assume we stack the entries of (15) into a column vector and the entries of (16) into a second column vector, and let $\mathcal{T}_N$ denote the mapping that maps the first vector (15) to the second one (16). The entries of this mapping can be determined from the update relation (6) and from the definitions of the a priori estimation errors $e_a(\cdot)$. The specific values of these entries are not of immediate interest here except to say that it can be verified that $\mathcal{T}_N$ turns out to be a block lower triangular operator of the form

$$\begin{bmatrix} \sqrt{(\mu(0) - \beta(0))}\sqrt{\lambda^{[1,N]}} e_a(0) \\ \sqrt{(\mu(1) - \beta(1))}\sqrt{\lambda^{[2,N]}} e_a(1) \\ \vdots \\ \boldsymbol{P}_N^{-1/2} \tilde{\boldsymbol{w}}_N \end{bmatrix} = \underbrace{\begin{bmatrix} x & & & \\ x & x & & \\ \vdots & & \ddots & \\ x & x & x & x \end{bmatrix}}_{\mathcal{T}_N} \begin{bmatrix} \sqrt{\lambda^{[0,N]}} \Pi_0^{-1/2} \tilde{\boldsymbol{w}}_{-1} \\ \sqrt{\mu(0)}\sqrt{\lambda^{[1,N]}} v(0) \\ \vdots \\ \sqrt{\mu(N)} v(N) \end{bmatrix}. \tag{17}$$

(For example, the first block entry of $\mathcal{T}_N$, which relates $e_a(0)$ to $\tilde{\boldsymbol{w}}_{-1}$, can be easily seen to be $\sqrt{\frac{\mu(0)-\beta(0)}{\lambda(0)}} \boldsymbol{u}_0 \Pi_0^{1/2}$.) The contractivity of $\mathcal{T}_N$ means that its maximum singular value is always guaranteed to be upper bounded by one,

$$\bar{\sigma}(\mathcal{T}_N) \leqslant 1. \tag{18}$$

The quantities in (15) involve the disturbances, i.e., noise and initial uncertainty in the guess for $\boldsymbol{w}$. The quantities in (16), on the other hand, involve the resulting estimation errors $e_a(\cdot)$ and $\tilde{w}_N$. In loose terms, the statement (18) establishes the following interesting fact: the Gauss–Newton algorithm (3), under the assumption $\beta(i) \leqslant \mu(i) \leqslant \bar{\mu}(i)$, always guarantees that the weighted error energy due to the initial disturbances will *not* be magnified.

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 $\{v(\cdot), \Pi_0^{-1/2}\tilde{w}_{-1}\}$ to $\{P_N^{-1/2}\tilde{w}_N\}$ is always a contraction. That is,

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

which is a well-known result in least-squares theory [17].

4. The feedback structure

We now employ the above local and global passivity relations in order to highlight a feedback structure that can be associated with Gauss–Newton recursions. This structure will be helpful in determining more relaxed conditions on the parameters $\{\mu(i), \lambda(i), \beta(i)\}$ for robustness and also for suggesting choices for the adaptation gains that would result in faster speeds of convergence.

We have argued above that if the adaptation gains are chosen such that $\beta(i) \leqslant \mu(i) \leqslant \bar{\mu}(i)$ then robustness (or contractivity) is guaranteed in the sense that the 'resulting weighted energies' will never exceed the 'weighted disturbance energies' (cf. (11) and (18)). The term 'robustness' is used in this paper to indicate that a ratio of suitably defined weighted energies is bounded by some positive number. The analysis in the earlier sections guarantees that for the choice $\beta(i) \leqslant \mu(i) \leqslant \bar{\mu}(i)$, the ratio of the energies of the signals in (15) and (16) will be bounded by one. However, the condition on $\mu(i)$ can be further relaxed at the expense of guaranteeing energy ratios that will be bounded by some other positive number (possibly larger than one), say

$$\frac{\text{weighted error energy}}{\text{disturbance energy}} \leqslant K, \tag{19}$$

for some constant K (to be determined). This is also a desirable property since it means that the disturbance energy will be at most magnified (or de-magnified) by a factor of K. Hence, from a practical point of view, it will mean that small disturbance energies cannot result in relatively substantial error energies.

These issues will be clarified in this section, along with procedures for choosing the adaptation gains $\mu(i)$ in order to result in faster speeds of convergence. The latter point will be addressed by studying the energy flow through the feedback structure.

But before proceeding to a discussion of the feedback structure, we first state the following two useful facts (proofs of which can be found in Appendix B). The first lemma provides a lower bound for $\bar{\mu}(i)$.

Lemma 3 (A lower bound on $\bar{\mu}(i)$). *Consider the Gauss–Newton algorithm* (3) *with the free positive parameters* $\{\lambda(i), \mu(i), \beta(i)\}$. *Define* $\bar{\mu}(i)$ *as before, viz.,* $\bar{\mu}(i) = (u_i P_i u_i^*)^{-1}$. *It always holds, for nonzero vectors* u_i, *that* $\bar{\mu}(i) > \beta(i)$.

The second lemma rewrites the Gauss–Newton update equation (3) in an alternative convenient form by using (8), which is rewritten here as

$$e_p(i) = \left(1 - \frac{\mu(i)}{\bar{\mu}(i)}\right) e_a(i) - \frac{\mu(i)}{\bar{\mu}(i)} v(i). \tag{20}$$

Lemma 4 (Alternative update form). *The Gauss–Newton algorithm* (3) *can be rewritten as follows:*

$$w_i = w_{i-1} + \bar{\mu}(i) P_i u_i^* [e_a(i) - e_p(i)]. \tag{21}$$

The alternative form (21) shows that the original weight-update equation (3) can be rewritten in terms of a new step-size parameter $\bar{\mu}(i)$ and a modified 'noise' term that we denote by $\bar{v}(i)$, and which is equal to $-e_p(i)$ (compare with (7)).

If we now follow arguments similar to those prior to (10), we readily conclude that the following equality holds for all $\mu(i)$ and $v(i)$ since the adaptation gain in (21) is now equal to $\bar{\mu}(i)$ itself, as required by the second condition in (10),

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

The quantities $\{\tilde{w}_i, \tilde{w}_{i-1}, e_a(i), P_i, P_{i-1}\}$ in the above ratio are exactly the same as those in the ratio (10). However, the normalizing factor $\mu(i)$ is now replaced by $\bar{\mu}(i)$ and the noise term $v(i)$ is now replaced by the a posteriori estimation error $e_p(i)$. Moreover, in view of the auxiliary result of Lemma 3, the difference $[\bar{\mu}(i) - \beta(i)]$ is now always nonnegative. Hence, the term $(\bar{\mu}(i) - \beta(i))\,|e_a(i)|^2$ can be interpreted as an energy term regardless of any condition on the original $\{\mu(i), \beta(i)\}$.

Relation (22) therefore establishes that the local map, denoted by $\overline{\mathcal{T}}_i$, from

$$\{\sqrt{\lambda(i)}P_{i-1}^{-1/2}\tilde{w}_{i-1}, \sqrt{\bar{\mu}(i)}e_p(i)\} \tag{23}$$

to

$$\{P_i^{-1/2}\tilde{w}_i, \sqrt{\bar{\mu}(i)-\beta(i)}e_a(i)\}, \tag{24}$$

is always *lossless*, i.e., it preserves energy (compare with the signals in (12) and (13) where the mapping was contractive and not lossless). The lossless map can be explicitly seen to be given by (in terms of $\bar{\mu}(i)$):

$$\begin{bmatrix} [\bar{\mu}(i)-\beta(i)]^{1/2}\, e_a(i) \\ P_i^{-1/2}\tilde{w}_i \end{bmatrix} =$$

$$\underbrace{\begin{bmatrix} \left[\dfrac{\bar{\mu}(i)-\beta(i)}{\lambda(i)}\right]^{1/2} u_i P_{i-1}^{1/2} & 0 \\ P_i^{-1/2}\left[I - \bar{\mu}(i)P_i u_i^* u_i\right] P_{i-1}^{1/2}\lambda^{-1/2}(i) & \bar{\mu}^{1/2}(i)P_i^{1/2}u_i^* \end{bmatrix}}_{\overline{\mathcal{T}}_i}$$

$$\times \begin{bmatrix} \lambda^{1/2}(i)P_{i-1}^{-1/2}\tilde{w}_{i-1} \\ \bar{\mu}^{1/2}(i)\, e_p(i) \end{bmatrix}.$$

This map involves the a posteriori estimation error $e_p(i)$ or, equivalently, the modified noise disturbance $-\bar{v}(i)$. But since $e_p(i)$ (or $\bar{v}(i)$) can be expressed in terms of the original disturbance $v(i)$ and the a priori estimation error (cf. (20)),

$$\bar{\mu}^{1/2}(i)\bar{v}(i) = \frac{\mu(i)}{\bar{\mu}^{1/2}(i)}v(i)\left(1 - \frac{\mu(i)}{\bar{\mu}(i)}\right)\bar{\mu}^{1/2}(i)e_a(i), \tag{25}$$

then the lossless map from (23) to (24) can be expressed as a feedback map that involves the original weighted disturbance $\sqrt{\bar{\mu}(\cdot)}v(\cdot)$. This is shown in Fig. 2. The feedback loop consists of a memoryless (or static) system that performs the scaling by $(1 - \mu(i)/\bar{\mu}(i))/\sqrt{1 - \beta(i)/\bar{\mu}(i)}$.

Comparing Fig. 2 with the earlier Fig. 1 we see that the main differences are (i) in representing the contractive map of the earlier Fig. 1 as a feedback interconnection of a lossless system with a memoryless system and (ii) in scaling the input and output signals $v(i)$ and $e_a(i)$ differently, where $\mu(i)$ is now replaced by $\bar{\mu}(i)$. Intuitively, the configuration of Fig. 2 shows how the different error and disturbance signals $\{\tilde{w}_i, \tilde{w}_{i-1}, e_a(i), v(i)\}$ in a Gauss–Newton recursion are connected. It highlights an implicit feedback structure that consists of two simple blocks: a lossless block that preserves energy and a static block with no memory. The feedback structure also explicitly incorporates the noise $v(i)$.

The point now is that such feedback configurations lend themselves rather immediately to stability and robustness analysis via tools that are standard in system theory, such as the small gain theorem [11, 32], as we now verify.

5. l_2-stability and the small gain theorem

The purpose of this section is to derive conditions on the adaptation gain $\mu(i)$ in order to guarantee that the overall system of Fig. 2 is l_2-stable. By this we mean that it maps a finite energy sequence $\{\sqrt{\bar{\mu}(\cdot)}\ v(\cdot)\}$ to a finite energy sequence $\{\sqrt{\bar{\mu}(\cdot) - \beta(\cdot)}\ e_a(\cdot)\}$ in a sense precised in (27) below. In other words, while the choice $\beta(i) \leqslant \mu(i) \leqslant \bar{\mu}(i)$ guarantees the contractivity property (18), with the signals defined in (17), the more relaxed condition on $\mu(i)$ that is derived further ahead in (30),

(i) will guarantee the l_2-stability of the mapping in Fig. 2, which involves the alternative signals $\{\sqrt{\bar{\mu}(\cdot)}v(\cdot)\}$ and $\{\sqrt{\bar{\mu}(\cdot) - \beta(\cdot)}e_a(\cdot)\}$ with $\bar{\mu}(i)$ replacing $\mu(i)$ (see (28) and (29) below), and

(ii) will also allow us to conclude, under an additional condition on $\beta(i)$, that the original mapping in (17) has a 2-induced norm that is bounded by another positive number (generally different from one).

In either case, the point is that we can relax the condition on $\mu(i)$ and still conclude that a relation of the form (19) can be established.

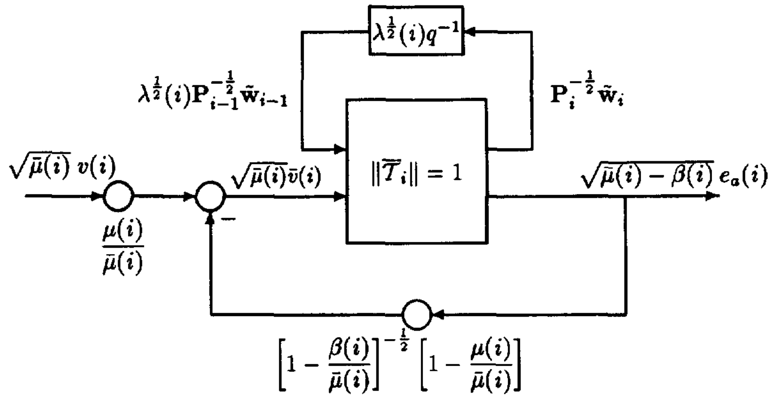

Fig. 2. A time-variant lossless mapping with gain feedback.

The main result is stated below (a proof of which can be found in Appendix C – the proof is based on the triangle inequality of norms).

Define

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

$$\text{and}\quad \gamma(N) = \max_{0\leqslant i\leqslant N}\frac{\mu(i)}{\bar{\mu}(i)}. \tag{26}$$

That is, $\Delta(N)$ is the maximum absolute value of the gain of the feedback loop over the interval of time $0 \leqslant i \leqslant N$. Likewise, $\gamma(N)$ is the maximum value of the scaling factor $\mu(i)/\bar{\mu}(i)$ at the input of the feedback interconnection.

Theorem 1 (l_2-stability). *Consider the Gauss–Newton recursion* (3). *If* $\Delta(N) < 1$ *then*

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

Interpretation. The terms $\{(\bar{\mu}(i)-\beta(i))\lambda^{[i+1,N]}|e_a(i)|^2,\ \bar{\mu}(i)\lambda^{[i+1,N]}|v(i)|^2\}$ that appear in expression (27) are essentially the same as the ones that appear in the earlier expression (14), except that $\mu(i)$ is now replaced by $\bar{\mu}(i)$. Therefore, expression (27) is still a relation that compares the energies of two weighted sequences. Moreover, the energy of the sequence $\{\bar{\mu}(i)\lambda^{[i+1,N]}|v(i)|^2\}$ is further scaled by $\gamma(N)$. Relation (27) then shows that the sum of the energies due to the weighted disturbances (noise and initial weight error) can at most be magnified by $1/[1-\Delta(N)]$. In particular, if the right-hand-side energy in (27) is bounded (finite) and if the gain $1/[1-\Delta(N)]$ is finite, then relation (27) also means that the map from

$$\{\sqrt{\lambda^{[\cdot+1,N]}\bar{\mu}(\cdot)}\,v(\cdot),\ \sqrt{\lambda^{[0,N]}}\boldsymbol{P}_{-1}^{-1/2}\tilde{\boldsymbol{w}}_{-1}\} \tag{28}$$

to

$$\{\sqrt{\lambda^{[\cdot+1,N]}(\bar{\mu}(\cdot)-\beta(i))}\,e_a(\cdot)\} \tag{29}$$

is l_2-stable (it maps a finite energy sequence to another finite energy sequence).

The difference $[1-\Delta(N)]$ is positive and smaller than one in view of the restriction on $\Delta(N)$, which is necessarily less than one. Therefore, the gain $1/[1-\Delta(N)]$ can become large only when $\Delta(N)$ approaches 1, but is otherwise reasonable for other values of $\Delta(N)$. For example, $\Delta(N) = 0.9$ leads to a magnification factor of 10.

In fact, the condition $\Delta(N) < 1$ can also be interpreted as a manifestation of the so-called small gain theorem in system analysis [11, 32]. In simple terms, the theorem states that the l_2-stability of a feedback configuration (that includes Fig. 2 as a special case) requires that the product of the 2-induced norms of the feedforward and the feedback operators be strictly bounded by one. Here, the feedforward map has the 2-induced norm equal to one (due to its losslessness) while the 2-induced norm of the feedback map is $\Delta(N)$, which therefore needs to be less than one.

It is also straightforward to verify that $\Delta(N) < 1$ is equivalent to requiring that $\mu(i)$ be chosen in the interval

$$0 < \mu(i) < \bar{\mu}(i)\left(1 + \sqrt{1 - \frac{\beta(i)}{\bar{\mu}(i)}}\right). \tag{30}$$

If $\beta(i)$ is further assumed to satisfy $\beta(i) \leqslant \mu(i)$ then it also holds that the map from

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

to

$$\left\{\sqrt{\lambda^{[\cdot+1,N]}\mu(\cdot) - \beta(\cdot)}\, e_{\mathrm{a}}(\cdot)\right\}$$

(i.e., with $\bar{\mu}(\cdot)$ replaced by $\mu(\cdot)$, as in the original relations (14) and (17)) is l_2-stable in the following sense. Define

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

This measures how far are the differences $[\mu(i) - \beta(i)]$ and $[\bar{\mu}(i) - \beta(i)]$ apart from each other (see Appendix C for a proof of the following result).

Theorem 2 (l_2-stability). *Consider again the Gauss–Newton recursion* (3). *If*

$$\beta(i) \leqslant \mu(i) < \bar{\mu}(i)\left(1 + \sqrt{1 - \frac{\beta(i)}{\bar{\mu}(i)}}\right) \tag{31}$$

holds, then

$$\sqrt{\sum_{i=0}^{N}(\mu(i) - \beta(i))\lambda^{[i+1,N]}\,|e_{\mathrm{a}}(i)|^2}$$
$$\leqslant \frac{\tilde{\gamma}^{1/2}(N)}{1 - \Delta(N)}\left[\sqrt{\lambda^{[0,N]}\tilde{\boldsymbol{w}}_{-1}^{*}\boldsymbol{P}_{-1}^{-1}\tilde{\boldsymbol{w}}_{-1}}\right.$$
$$\left. + \gamma^{1/2}(N)\sqrt{\sum_{i=0}^{N}\mu(i)\lambda^{[i+1,N]}|v(i)|^2}\right]. \tag{32}$$

Interpretation. We see that by allowing $\mu(i)$ to assume larger values than $\bar{\mu}(i)$, as suggested by (31), the gain factor that was one in (14) and (18) is now changed to $\tilde{\gamma}^{1/2}(N)/[1 - \Delta(N)]$. How big or how large this gain is depends on how close or how far is $\Delta(N)$ from one and also on the value of $\tilde{\gamma}(N)$. Relation (32) tells us how much the (weighted) estimation error energies in a Gauss–Newton recursion are magnified relative to the (weighted) disturbance energies.

Therefore, we conclude from the last two theorems that a sufficient condition for the energy amplification to remain bounded is to choose the adaptation gains as indicated in the statements of the theorems (cf. (30) or (31)).

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

$$\begin{aligned} \boldsymbol{M}_N &= \mathrm{diag}\{\mu(0), \mu(1), \ldots, \mu(N)\}, \\ \overline{\boldsymbol{M}}_N &= \mathrm{diag}\{\bar{\mu}(0), \bar{\mu}(1), \ldots, \bar{\mu}(N)\}, \end{aligned} \tag{33}$$

$$\begin{aligned} \boldsymbol{B}_N &= \mathrm{diag}\{\beta(0), \beta(1), \ldots, \beta(N)\}, \\ \Lambda_N &= \mathrm{diag}\{\lambda^{[0,N]}, \lambda^{[1,N]}, \ldots, \lambda(N)\}, \end{aligned} \tag{34}$$

and the vectors

$$\begin{aligned} \boldsymbol{e}_{\mathrm{a},N}^{*} &= [e_{\mathrm{a}}^{*}(0), e_{\mathrm{a}}^{*}(1), \ldots, e_{\mathrm{a}}^{*}(N)], \\ \boldsymbol{v}_N^{*} &= [v^{*}(0), v^{*}(1), \ldots, v^{*}(N)]. \end{aligned} \tag{35}$$

That is, $\boldsymbol{M}_N$ is the diagonal matrix with the adaptation gains along its diagonal. Likewise, $\overline{\boldsymbol{M}}_N$ has the (inverse of the) input energies along its diagonal. The $\boldsymbol{e}_{a,N}$ and $\boldsymbol{v}_N$ are column vectors with the a priori estimation errors and the noise sequence, respectively.

Correspondingly, the matrix $[\boldsymbol{I} - \boldsymbol{B}_N\overline{\boldsymbol{M}}_N^{-1}]^{-1/2}[\boldsymbol{I} - \boldsymbol{M}_N\overline{\boldsymbol{M}}_N^{-1}]$ is diagonal with the gains of the feedback loop along its diagonal. Likewise, the matrix $\boldsymbol{M}_N\overline{\boldsymbol{M}}_N^{-1}$ is diagonal with the input gains $\mu(i)/\bar{\mu}(i)$ along its diagonal.

If we denote the 2-induced norms of the matrices $[\boldsymbol{I} - \boldsymbol{B}_N\overline{\boldsymbol{M}}_N^{-1}]^{-1/2}[\boldsymbol{I} - \boldsymbol{M}_N\overline{\boldsymbol{M}}_N^{-1}]$ and $\boldsymbol{M}_N\overline{\boldsymbol{M}}_N^{-1}$ by

$$\left\|[\boldsymbol{I} - \boldsymbol{B}_N\overline{\boldsymbol{M}}_N^{-1}]^{-1/2}[\boldsymbol{I} - \boldsymbol{M}_N\overline{\boldsymbol{M}}_N^{-1}]\right\|_{2,\mathrm{ind}}$$

and $\left\|\boldsymbol{M}_N\overline{\boldsymbol{M}}_N^{-1}\right\|_{2,\mathrm{ind}}$, respectively, it is then easy to see, due to the diagonal structure of M_N and $\overline{M}_N$, that these norms are given by $\Delta(N)$ and $\gamma(N)$, respectively,

$$\Delta(N) = \left\| [I - B_N \overline{M}_N^{-1}]^{-1/2} [I - M_N \overline{M}_N^{-1}] \right\|_{2,\text{ind}}$$

and $\gamma(N) = \left\| M_N \overline{M}_N^{-1} \right\|_{2,\text{ind}}$.

The l_2-stability condition then amounts to guaranteeing a contractive feedback map,

$$[I - B_N \overline{M}_N^{-1}]^{-1/2} [I - M_N \overline{M}_N^{-1}].$$

The map is memoryless in this case: it consists of a diagonal matrix that simply scales the input sequence $\{\sqrt{\lambda^{[\cdot+1,N]}(\bar{\mu}(\cdot) - \beta(\cdot))}\, e_a(\cdot)\}$. We shall see later, especially in the context of filtered-error variants, that a more involved *dynamic* feedback map arises.

6. 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 ignore the measurement noise $v(i)$ and assume that we have noiseless measurements $d(i) = u_i w$. The weight-error update equation (6) can then be easily seen to collapse to

$$\tilde{w}_i = [I - \mu(i) P_i u_i^* u_i]\, \tilde{w}_{i-1}, \tag{36}$$

where the $M \times M$ coefficient matrix $[I - \mu(i) P_i u_i^* u_i]$ is simply a rank one modification of the identity matrix. It is known in the stochastic setting that for Gaussian processes [1], as well as for spherically invariant random processes [20], 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 (with P_i replaced by the identity matrix) [33]. We now argue that this conclusion is consistent with the feedback configuration of Fig. 2 in the case of Gauss–Newton methods.

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 $\tilde{w}_{-1}$ and is equal to $\tilde{w}_{-1}^* P_{-1}^{-1} \tilde{w}_{-1}$. 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,

$$E_w(i) = \lambda(i) E_w(i-1) - E_e(i), \tag{37}$$

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 $P_i^{-1/2} \tilde{w}_i$,

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

Expression (37) implies that, for $\lambda(i) \leqslant 1$, the weight-error energy is a non-increasing function of time, i.e., $E_w(i) \leqslant E_w(i-1)$ for all i. Strict inequality is guaranteed if $E_e(i) \neq 0$. This is in general the case especially when the input vectors u_i satisfy some persistence of excitation conditions. Note also that the so-called forgetting factor $\lambda(i)$ plays an important role. The smaller the $\lambda(i)$, the smaller the $E_w(i)$.

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 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 (37)). 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 $\tilde{w}_i$ is lowered, thus confirming our earlier remark. Simulation examples are included in a later section.

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 (6) collapses to $\tilde{w}_i = \tilde{w}_{i-1}$ and, hence, $w_i = w_{-1}$ for all time instants i. This means that no improvements over the initial guess are obtained and, consequently, convergence will never be attained if $w_{-1} \neq w$. 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, from 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 Fig. 2 where it is clear that the fine structures 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 u_i.

7. Filtered-error algorithms with Gauss–Newton update

The feedback loop concept of the former sections applies equally well to Gauss–Newton algorithms that

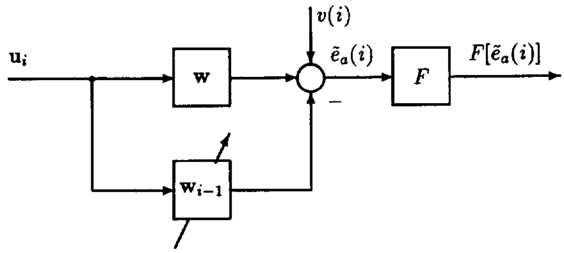

Fig. 3. Structure of filtered-error adaptive algorithms.

employ filtered versions of the output estimation error, $\tilde{e}_a(i) = d(i) - u_i w_{i-1}$. Such algorithms are useful when the error $\tilde{e}_a(i)$ cannot be observed directly, but rather a filtered version of it, as indicated in Fig. 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, say

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

It may also be a time-variant filter, 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 (see, e.g., [5, 12, 15]). In the sequel we shall discuss two important classes of algorithms that employ filtered error measurements: the filtered error Gauss–Newton algorithm for transversal filters and its counterpart for adaptive IIR filters, also called the pseudo linear regression (PLR) algorithm.

7.1. The filtered-error Gauss–Newton algorithm

The so-called filtered-error Gauss–Newton algorithm (FEGN) retains the input vector unchanged, viz., it uses an update equation of the form

$$w_i = w_{i-1} + \mu(i)P_i u_i^* F[d(i) - u_i w_{i-1}]$$

$$= w_{i-1} + \mu(i)P_i u_i^* F[\tilde{e}_a(i)], \tag{38}$$

where the difference $[d(i) - u_i w_{i-1}]$ is further filtered by F. We can repeat here the same feedback analysis as in the previous sections. This will establish a similar feedback structure to the one in Fig. 2, except that the feedback loop will now be dynamic; the impulse response of the filter F will also enter into the composition of the feedback loop.

Since the discussion in this section follows rather closely the development in the earlier sections, we shall be brief and will only emphasize the differences that arise in the current context.

7.1.1. The feedback sructure for the filtered-error Gauss–Newton algorithms

Following the discussion of Appendix B that led to (21), we get the following sequence of equalities:

$$\begin{aligned}
\boldsymbol{w}_i &= \boldsymbol{w}_{i-1} + \mu(i)\boldsymbol{P}_i\boldsymbol{u}_i^* F[e_a(i)] + \mu(i)\boldsymbol{P}_i\boldsymbol{u}_i^* F[v(i)] \\
&= \boldsymbol{w}_{i-1} + \mu(i)\boldsymbol{P}_i\boldsymbol{u}_i^* F[e_a(i)] + \mu(i)\boldsymbol{P}_i\boldsymbol{u}_i^* F[v(i)] \\
&\quad + \bar{\mu}(i)\boldsymbol{P}_i\boldsymbol{u}_i^* e_a(i) - \bar{\mu}(i)\boldsymbol{P}_i\boldsymbol{u}_i^* e_a(i) \\
&= \boldsymbol{w}_{i-1} + \bar{\mu}(i)\boldsymbol{P}_i\boldsymbol{u}_i^* e_a(i) \\
&\quad + \boldsymbol{P}_i\boldsymbol{u}_i^*[\mu(i)F[v(i)] - \bar{\mu}(i)e_a(i) \\
&\quad + \mu(i)F[e_a(i)]] \\
&= \boldsymbol{w}_{i-1} + \bar{\mu}(i)\boldsymbol{P}_i\boldsymbol{u}_i^*[e_a(i) + \bar{v}(i)],
\end{aligned} \tag{39}$$

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

$$\bar{\mu}(i)\bar{v}(i) = \mu(i)F[v(i)] - \bar{\mu}(i)e_a(i) + \mu(i)F[e_a(i)], \tag{40}$$

and $\bar{\mu}^{-1}(i) = \boldsymbol{u}_i\boldsymbol{P}_i\boldsymbol{u}_i^*$.

Expression (39) rewrites the filtered-error update relation (38) in a form similar to the original Gauss–Newton update (3) (i.e., without an explicit filtering operation by F). This is achieved by redefining a modified noise sequence $\bar{v}(i)$ that incorporates the action of the filter F, as well as by employing an adaptation gain that is equal to $\bar{\mu}(i)$.

Based on this rewriting, and using the same arguments following Lemma 4, we readily obtain that the following equality should hold for all i:

$$\frac{\tilde{\boldsymbol{w}}_i^*\boldsymbol{P}_i^{-1}\tilde{\boldsymbol{w}}_i + (\bar{\mu}(i) - \beta(i))\,|e_a(i)|^2}{\lambda(i)\tilde{\boldsymbol{w}}_{i-1}^*\boldsymbol{P}_{i-1}^{-1}\tilde{\boldsymbol{w}}_{i-1} + \bar{\mu}(i)\,|\bar{v}(i)|^2} = 1. \tag{41}$$

This again establishes that the map from $\{\sqrt{\lambda(i)}\,\boldsymbol{P}_{i-1}^{-1/2}\tilde{\boldsymbol{w}}_{i-1}, \sqrt{\bar{\mu}(i)}\bar{v}(i)\}$ to $\{\boldsymbol{P}_i^{-1/2}\tilde{\boldsymbol{w}}_i, \sqrt{\bar{\mu}(i) - \beta(i)}\,e_a(i)\}$, 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 a priori estimation error $\sqrt{\bar{\mu}(\cdot) - \beta(i)}e_a(\cdot)$ can be expressed in terms of a feedback structure, as shown in Fig. 4, which now explicitly includes the action of the filter F in the feedback loop.

We further remark that the notation,

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

that appears in the feedback loop should be interpreted as follows: we first divide $\sqrt{\bar{\mu}(i) - \beta(i)}\,e_a(i)$ by $\sqrt{\bar{\mu}(i) - \beta(i)}$, followed by the filter F and then by a subsequent scaling by $\mu(i)/\sqrt{\bar{\mu}(i)}$. 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 $\mu(i)/\sqrt{\bar{\mu}(i)}$.

The feedback loop now consists of a dynamic system. But we can still proceed to study the l_2-stability of the overall configuration in much the same way as we did in Section 4. Similar arguments will lead to a sufficient condition for stability that we now exhibit. For this purpose, we use the convenient vector and matrix quantities as in (33)–(35) and also define a vector $\bar{\boldsymbol{v}}_N$ similar to $\boldsymbol{v}_N$ but with the entries $\bar{v}(\cdot)$ instead of $v(\cdot)$.

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

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

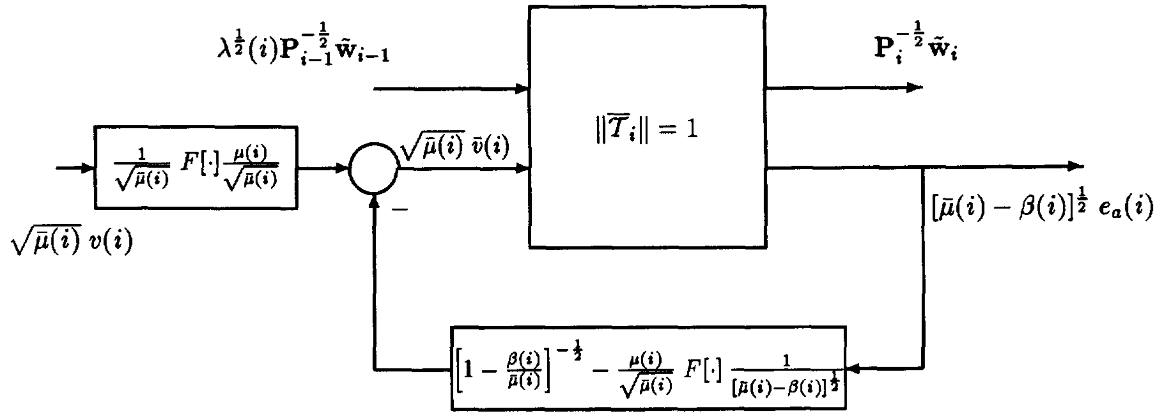

Fig. 4. Filtered-error Gauss–Newton algorithm as a time-variant lossless mapping with dynamic feedback.

It is thus immediate to verify that the successive outputs of $F[e_a(\cdot)]$ can be obtained by simply computing the matrix–vector product $\boldsymbol{F}_N \boldsymbol{e}_{a,N}$. Also, if the filter F were time-variant all that changes is that the matrix $\boldsymbol{F}_N$ will not be Toeplitz anymore. Instead, its first diagonal will consist of the values of the first coefficient $f_0(\cdot)$ at the successive time-instants, and so on.

The following result follows from precisely similar arguments to those in the proof of Theorems 1 and 2 in Appendix C. Define

$$\Delta(N) = \left\| \overline{\boldsymbol{M}}_N^{1/2} (\boldsymbol{I} - \boldsymbol{M}_N \overline{\boldsymbol{M}}_N^{-1} \boldsymbol{F}_N)[\overline{\boldsymbol{M}}_N - \boldsymbol{B}_N]^{-1/2} \right\|_{2,\text{ind}}$$

and $\gamma(N) = \left\| \overline{\boldsymbol{M}}_N^{-1/2} \boldsymbol{M}_N \boldsymbol{F}_N \overline{\boldsymbol{M}}_N^{-1/2} \right\|_{2,\text{ind}}$.

Here, $\Delta(N)$ is the 2-induced norm of the feedback loop whose action is described by the matrix $\overline{\boldsymbol{M}}_N^{1/2} (\boldsymbol{I} - \boldsymbol{M}_N \overline{\boldsymbol{M}}_N^{-1} \boldsymbol{F}_N)[\overline{\boldsymbol{M}}_N - \boldsymbol{B}_N]^{-1/2}$. Likewise, $\gamma(N)$ is the 2-induced norm of the filter at the input of the feedback structure and whose action is described by $\overline{\boldsymbol{M}}_N^{-1/2} \boldsymbol{M}_N \boldsymbol{F}_N \overline{\boldsymbol{M}}_N^{-1/2}$.

Theorem 3. *Consider the filtered-error Gauss–Newton recursion* (38). *If $\Delta(N) < 1$ then*

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

$$\leqslant \frac{1}{1 - \Delta(N)} \left[\sqrt{\lambda^{[0,N]} \tilde{\boldsymbol{w}}_{-1}^* \boldsymbol{P}_{-1}^{-1} \tilde{\boldsymbol{w}}_{-1}} + \gamma(N) \sqrt{\sum_{i=0}^{N} \bar{\mu}(i) \lambda^{[i+1,N]} |v(i)|^2} \right]. \tag{42}$$

Moreover, if we further have $\beta(i) \leqslant \mu(i)$ then

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

$$\leqslant \frac{\bar{\gamma}^{1/2}(N)}{1 - \Delta(N)} \left[\sqrt{\lambda^{[0,N]} \tilde{\boldsymbol{w}}_{-1}^* \boldsymbol{P}_{-1}^{-1} \tilde{\boldsymbol{w}}_{-1}} + \gamma^{1/2}(N) \sqrt{\sum_{i=0}^{N} \mu(i) \lambda^{[i+1,N]} |v(i)|^2} \right]. \tag{43}$$

We thus see that the major requirement is for the feedback matrix

$$\overline{\boldsymbol{M}}_N^{1/2} (\boldsymbol{I} - \boldsymbol{M}_N \overline{\boldsymbol{M}}_N^{-1} \boldsymbol{F}_N)[\overline{\boldsymbol{M}}_N - \boldsymbol{B}_N]^{-1/2}$$

to be contractive. We shall denote this matrix by $\boldsymbol{G}_N$. It can be easily seen to have the following triangular form (it also has a band of width M_F):

$$\boldsymbol{G}_N = \begin{pmatrix} \dfrac{1-\frac{\mu(0)}{\bar{\mu}(0)}f_0}{\sqrt{1-\frac{\beta(0)}{\bar{\mu}(0)}}} & & 0 \\ -\dfrac{\mu(1)f_1}{\sqrt{\bar{\mu}(1)}\sqrt{\bar{\mu}(0)-\beta(0)}} & \dfrac{1-\frac{\mu(1)}{\bar{\mu}(1)}f_0}{\sqrt{1-\frac{\beta(1)}{\bar{\mu}(1)}}} & \\ -\dfrac{\mu(2)f_2}{\sqrt{\bar{\mu}(2)}\sqrt{\bar{\mu}(0)-\beta(0)}} & -\dfrac{\mu(2)f_1}{\sqrt{\bar{\mu}(2)}\sqrt{\bar{\mu}(1)-\beta(1)}} & \dfrac{1-\frac{\mu(2)}{\bar{\mu}(2)}f_0}{\sqrt{1-\frac{\beta(2)}{\bar{\mu}(2)}}} \\ \vdots & & & \ddots \end{pmatrix}. \tag{44}$$

The entries of $\boldsymbol{G}_N$ depend on four parameters: the step-sizes $\mu(i)$, the $\beta(i)$, the energies of the input sequence $\bar{\mu}(i)$, and the error filter F. Several special cases may be of interest. For example, the special case $F = 1$ (i.e., no filter) immediately leads to

$$\boldsymbol{G}_N = \operatorname{diag}\left\{ \frac{1-\frac{\mu(0)}{\bar{\mu}(0)}}{\sqrt{1-\frac{\beta(0)}{\bar{\mu}(0)}}}, \ldots, \frac{1-\frac{\mu(N)}{\bar{\mu}(N)}}{\sqrt{1-\frac{\beta(N)}{\bar{\mu}(N)}}} \right\},$$

which is precisely the case we encountered earlier in Section 4. A sufficient condition for the contractivity of $\boldsymbol{G}_N$ was seen to choose $\mu(i)$ such that $0 < \mu(i) < \bar{\mu}(i)(1 + \sqrt{1-\beta(i)/\bar{\mu}(i)})$. Another special case is $F = q^{-1}$ (i.e., a simple delay). The filtered-error Gauss–Newton recursion (38) then collapses to

$$\boldsymbol{w}_i = \boldsymbol{w}_{i-1} + \mu(i)\boldsymbol{P}_i\boldsymbol{u}_i^*\, \tilde{e}_a(i-1). \tag{45}$$

The corresponding $\boldsymbol{G}_N$ is given by (since $f_0 = 0$ and $f_1 = 1$)

$$\boldsymbol{G}_N = \begin{pmatrix} \dfrac{1}{\sqrt{1-\frac{\beta(0)}{\bar{\mu}(0)}}} & & 0 \\ -\dfrac{\frac{\mu(1)}{\sqrt{\bar{\mu}(1)}}}{\sqrt{\bar{\mu}(0)-\beta(0)}} & \dfrac{1}{\sqrt{1-\frac{\beta(1)}{\bar{\mu}(1)}}} & \\ & -\dfrac{\frac{\mu(2)}{\sqrt{\bar{\mu}(2)}}}{\sqrt{\bar{\mu}(2)-\beta(1)}} & \dfrac{1}{\sqrt{1-\frac{\beta(2)}{\bar{\mu}(2)}}} \\ 0 & & \ddots & \ddots \end{pmatrix}. \tag{46}$$

We see that $\boldsymbol{G}_N$ cannot be a contractive matrix since, for example, its leading $(0,0)$ entry is not less than one, because of $\beta(i) < \bar{\mu}(i)$ (if a matrix is contractive, then all its principal leading submatrices have also to be contractive). This is consistent with results in the literature where it has been observed that the delayed-error algorithm usually leads to unstable behavior.

We also see from the general expression for $\boldsymbol{G}_N$ in (44) that a simple gain filter $F = f_0$ with a *negative* f_0 leads to a noncontractive $\boldsymbol{G}_N$. We consider in the next section another important special case.

7.2. The projection filtered-error algorithm

The projection filtered-error algorithm employs a special choice for the step-size parameter, viz., a scaled multiple of the reciprocal input energy,

$$\mu(i) = \alpha\, \bar{\mu}(i), \quad \alpha > 0. \tag{47}$$

This leads to an update recursion that is often referred to as a projection update,

$$\boldsymbol{w}_i = \boldsymbol{w}_{i-1} + \alpha \frac{\boldsymbol{P}_i\boldsymbol{u}_i^*}{\boldsymbol{u}_i\boldsymbol{P}_i\boldsymbol{u}_i^*} F[\tilde{e}_a(i)].$$

In this case, it can be seen that the contractivity requirement now collapses to requiring the contractivity of $\boldsymbol{G}_N$, where $\boldsymbol{G}_N$ is now given by

$$\begin{pmatrix} \frac{1-\alpha f_0}{\sqrt{1-\frac{\beta(0)}{\bar{\mu}(0)}}} & & 0 & \\ -\frac{\alpha\sqrt{\bar{\mu}(1)}}{\sqrt{\bar{\mu}(0)-\beta(0)}} & \frac{1-\alpha f_0}{\sqrt{1-\frac{\beta(1)}{\bar{\mu}(1)}}} & & \\ -\frac{\alpha\sqrt{\bar{\mu}(2)}}{\sqrt{\bar{\mu}(0)-\beta(0)}} & -\frac{\alpha\sqrt{\bar{\mu}(2)}}{\sqrt{\bar{\mu}(1)-\beta(1)}} & \frac{1-\alpha f_0}{\sqrt{1-\frac{\beta(2)}{\bar{\mu}(2)}}} & \\ \vdots & & & \ddots \end{pmatrix}. \tag{48}$$

We further assume that the (weighted) energy of the input sequence $\boldsymbol{u}_i$ does not change very rapidly over the filter length M_F, i.e., $\bar{\mu}(i) \approx \bar{\mu}(i-1) \approx \cdots \approx \bar{\mu}(i-M_F)$. This is a reasonable assumption since, as mentioned earlier, we often have $M_F \ll M$. We also assume that the $\beta(i)$ satisfy $\beta(i) = \bar{\beta}\bar{\mu}(i)$ with $\bar{\beta} \ll 1$. (A particular choice with $\alpha = 1$ and $\bar{\beta} = 0.5$ will be discussed later in the context of the PLR algorithm.) In this case, $\boldsymbol{G}_N$ collapses to

$$\boldsymbol{G}_N \approx \boldsymbol{I} - \alpha \boldsymbol{F}_N, \tag{49}$$

and the contractivity of $[\boldsymbol{I} - \alpha\boldsymbol{F}_N]$ can be guaranteed if we choose the α so as to satisfy

$$\max_{\Omega} \left|1 - \alpha F(\mathrm{e}^{\mathrm{j}\Omega})\right| < 1. \tag{50}$$

This is equivalent to requiring that $F(\mathrm{e}^{\mathrm{j}\Omega})$ should lie inside a circle of center $(1,0)$ and radius $1/\alpha$. Thus, the smaller the step-size α, the wider the circle, and, consequently, the easier for the condition to be met. This also suggests that, for faster convergence (i.e., for smallest feedback gain), we should choose α optimally by solving the min-max problem:

$$\min_{\alpha} \max_{\Omega} \left|1 - \alpha F(\mathrm{e}^{\mathrm{j}\Omega})\right|. \tag{51}$$

If the resulting minimum is less than 1 then the corresponding optimum α will result in faster convergence and also in an overall robust system (cf. (43)).

7.3. A note on the pseudo-linear regression algorithm

Another case of filtered error variants that arises in IIR modeling is the so-called pseudo-linear regression algorithm [3, 30]. In this case, the filter F in the feedback path takes the special form (see also the discussion in [22])

$$F(q^{-1}) = \frac{1}{1 - A(q^{-1})},$$

for some unknown recursive filter A. The results of this paper are also applicable to this case. In particular, l_2-stability would require the following contractivity condition:

$$\left\| \overline{\boldsymbol{M}}_N^{1/2} (\boldsymbol{I} - \boldsymbol{M}_N \overline{\boldsymbol{M}}_N^{-1} [\boldsymbol{I} - \boldsymbol{A}_N]^{-1}) \times [\overline{\boldsymbol{M}}_N - \boldsymbol{B}_N]^{-1/2} \right\|_{2,\mathrm{ind}} < 1,$$

where the matrix $\boldsymbol{A}_N$ is lower triangular (as in the definition of $\boldsymbol{F}_N$) and describes the action of the unknown filter $A(q^{-1})$.

Assume, for example, that $\beta(i) = \gamma_1 \bar{\mu}(i)$ and $\mu(i) = \gamma_2 \bar{\mu}(i)$, with $\gamma_1 \leqslant \gamma_2 < 1$. The condition then simplifies to choosing the two parameters γ_1 and γ_2 so as to satisfy

$$\|\overline{\boldsymbol{M}}_N^{1/2} [\boldsymbol{I}_N - \gamma_2 [\boldsymbol{I} - \boldsymbol{A}_N]^{-1}] \overline{\boldsymbol{M}}_N^{-1/2}\|_{2,\mathrm{ind}} < \sqrt{\gamma_1}.$$

A sufficient but conservative condition is to require

$$\|\boldsymbol{I}_N - \gamma_2 [\boldsymbol{I} - \boldsymbol{A}_N]^{-1}\|_{2,\mathrm{ind}} < \sqrt{\gamma_1} \frac{\min_{0 \leqslant i \leqslant N} \sqrt{\bar{\mu}(i)}}{\max_{0 \leqslant i \leqslant N} \sqrt{\bar{\mu}(i)}},$$

which establishes a dependency on the time-variations in the input vector energy.

Other choices of the adaptation gains have been shown to lead to results independent of these variations (see [26]). The above condition is still dependent on the (usually unknown) filter $A(q^{-1})$. However, if we could choose $\mu(i) = \beta(i) = \bar{\mu}(i)$, then the scaling factor $\sqrt{\bar{\mu}(i) - \beta(i)}$ of the a priori error sequence would become zero and the feedback-path will have no effect on the overall stability. But, unfortunately, $\bar{\mu}(i)$ is a function of $\beta(i)$ and they can only become equal for $\lambda(i) = 0$. The choice of the projection stepsize helps here, if $\mu(i) = \alpha\bar{\mu}(i)$ and $\beta(i) = \bar{\beta}\bar{\mu}(i)$. In a recent paper [3], the following choices were proposed:

$$\bar{\beta} = 0.5, \quad \alpha = 1, \quad \lambda(i) = 0.5\bar{\mu},$$

with the intent of minimizing the a posteriori error. In this case, $\bar{\mu}(i) - \beta(i) = 1/(1 + \bar{\mu}^{-1}(i|i-1))$ does not become zero, but $\mu(i) = \bar{\mu}(i)$ and the feedback loop gain becomes zero in the unfiltered case. The forgetting factor λ has been chosen to be $1 - \beta(i)$, since this choice is advantageous for fixed-point realizations. The elements of the Riccati matrix $\boldsymbol{P}_i$ can then easily be bounded.

8. Simulation results

In the following simulations we demonstrate several of the points raised in our previous discussions. We considered the following two error-path filter functions:

$$F_1(q) = 1 - 1.2\,q^{-1} + 0.72\,q^{-2}$$

$$\text{and} \quad F_2(q) = \sum_{k=0}^{19} q^{-k} = \frac{1 - q^{-20}}{1 - q^{-1}}.$$

The real parts of F_1 and F_2 are shown in Fig. 5, which shows that F_1 has a strictly positive real part while F_2 does not.

We employed the projection normalization (47) 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 $\boldsymbol{u}_i$ with shift structure) was chosen sinusoidal with frequency Ω_0 so that the a priori error signal can be assumed to be dominated by this frequency. In this case, the optimum α of (50) can be approximated by the simpler expression

$$\min_{\alpha} \left|1 - \alpha F\left(e^{j\Omega_0}\right)\right|,$$

which can be solved explicitly and we get

$$\alpha_{\text{opt}} = \text{Real}\left\{\frac{1}{F_1(e^{-j\Omega_0})}\right\}$$

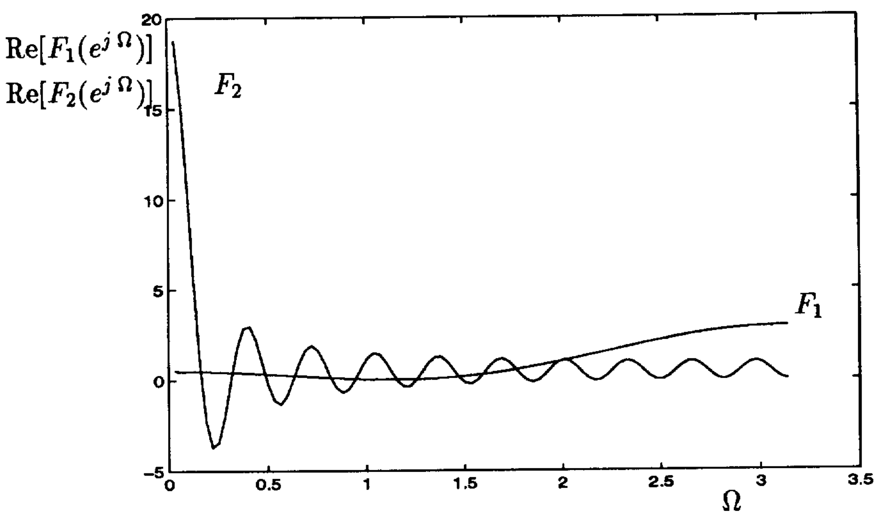

Fig. 5. Real parts of $F_1(e^{j\Omega})$ and $F_2(e^{j\Omega})$.

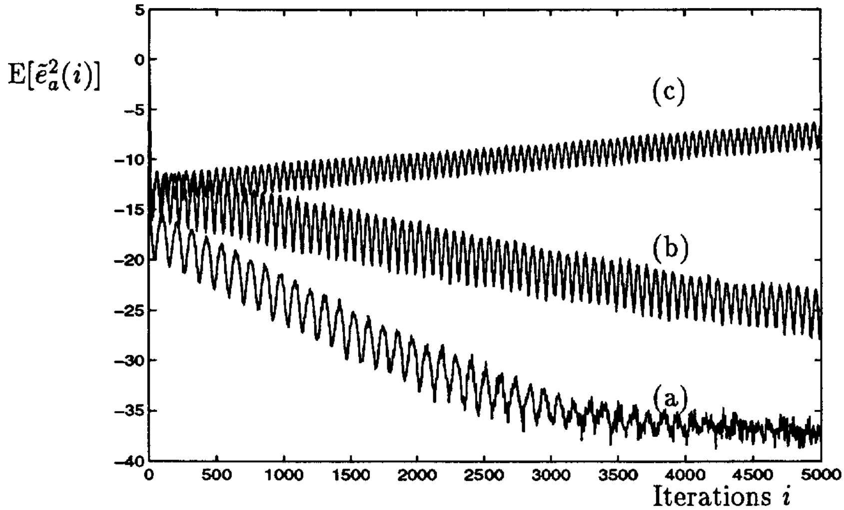

Fig. 6. Convergence behavior for the FEGN algorithm with sinusoidal input sequence and various step-sizes (a) $\alpha = 0.085$, (b) $\alpha = 0.15$, (c) $\alpha = 0.18$.

to be the step-size that causes faster convergence speed.

Following the same procedure, the step-size α_{lim} for which the stability limit (49) is achieved can be calculated to be $\alpha_{\text{lim}} = 2\alpha_{\text{opt}}$.

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 reason for adding a random phase ϕ is to obtain smoother learning curves after averaging. The optimal step-size α_{opt} can thus be calculated to be $\alpha_{\text{opt}} = 0.085$ and the stability bound is obtained for $\alpha_{\text{lim}} = 2\alpha_{\text{opt}} = 0.17$. Fig. 6 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 -40 dB 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. We see that for the first two values of α, the sample average of $|\tilde{e}_a(i)|^2$ decreases with time, while for the last value it increases.

To derive conditions (50) and (51) we used the approximation (49) with the requirement that the energy of the input vector changes slowly over the filter length M. To measure the impact of this variation, the change in the input vector energy from one time-instant to the following can be used, viz.,

$$\left(\bar{\mu}^{-1}(i) - \bar{\mu}^{-1}(i-1)\right)^2.$$

This difference collapses to the simple expression $\left(|u(i)|^2 - |u(i-M)|^2\right)^2$ when a time-shift structure for $\boldsymbol{u}_i$ is assumed. In a stochastic setting, the above measure of the input energy variation can be explicitly evaluated via

$$E\left(|u(i)|^2 - |u(i-M)|^2\right)^2 = 2(E|u(i)|^4 - E|u(i)|^2|u(i-M)|^2),$$

which can further be reduced to

$$2(\kappa - 1)E[|u(i)|^2]^2,$$

if we assume that $|u(i)|^2$ is uncorrelated with $|u(i-M)|^2$, which is a reasonable assumption for M relatively large. Hence, the kurtosis parameter κ gives a measure of the input energy variations. To describe the impact of these variations on the FEGN algorithm, we applied various random input sequences with different kurtosis as listed in Table 1. A bipolar random sequence with amplitude A avoids, of course, any variation in the input energy since we have now a constant $\bar{\mu}^{-1}(i) = MA$ and the kurtosis becomes one. For Gaussian sequences the kurtosis becomes three and for a

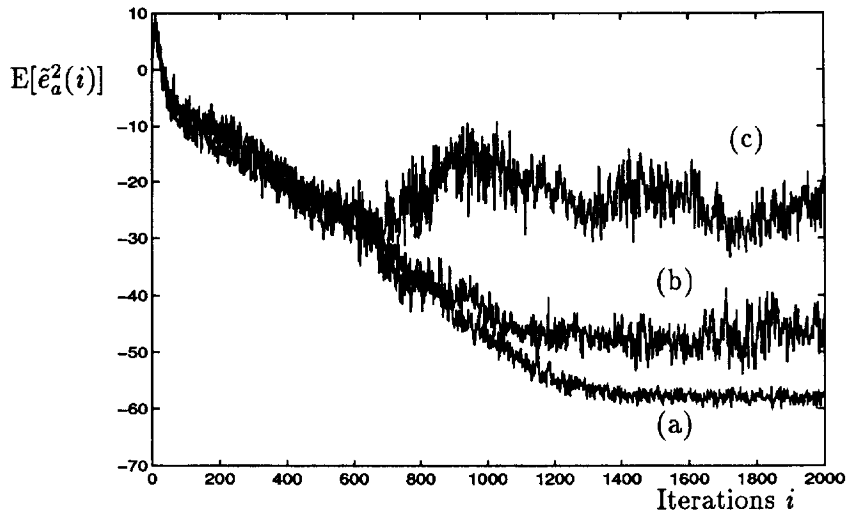

Fig. 7. Learning curves for the FEGN algorithm with F_1 and various random input processes with different pdf ($\alpha = \alpha_{opt} = 0.3$). (a) bipolar, uniform, gaussian; (b) K_0; (c) gamma.

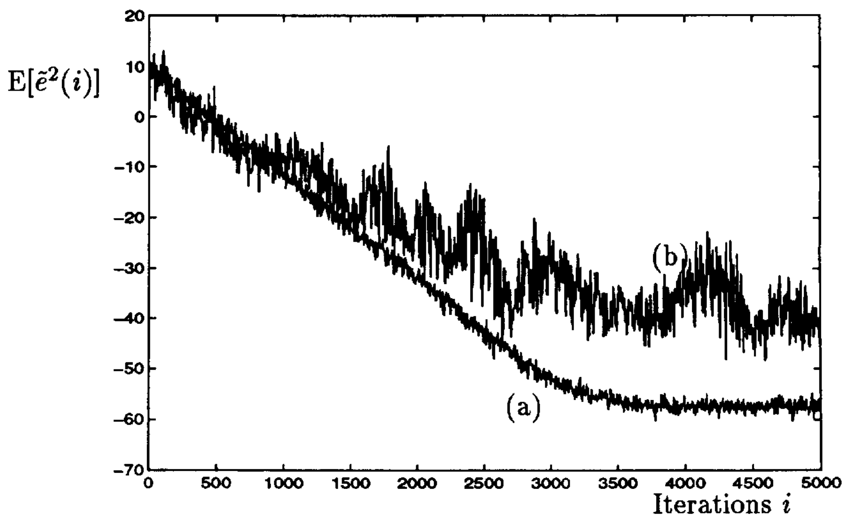

Fig. 8. Learning curves for the FEGN algorithm with F_2 and random input process: (a) Gaussian, $\alpha = 0.05$; (b) K_0, $\alpha = 0.05$.

K_0 density, a modified Bessel function or McDonald function of the second kind for order zero, which is used to describe the density of speech samples, the kurtosis is equal to nine. All random sequences were white processes with variance one. Fig. 7 depicts the simulation results, where we again averaged over 50 trials. Unlike the previous simulations, the power of the additive noise $v(i)$ was set at -60 dB relative to the input sequence since then the effect of the various kurtosis can be observed better. As can be seen from the figure, the higher the kurtosis, the more the variation in the steady-state value. The convergence speed and the stability bounds, however, remain unchanged. We repeated the above experiment with white random processes for the filter F_2. The stability bound was calculated at $\alpha = 0.097$ and the fastest convergence at $\alpha = 0.05$. The simulation resulted in a stability bound at 0.09 and fastest convergence at 0.05 for the first

Table 1
Kurtosis of various density functions.

Distribution	Kurtosis
Bipolar	1
Uniform	1.8
Gaussian	3
K_0	9
Gamma	11.66

four random processes of the table. For the gamma density, however, the values were smaller (stability bound at 0.05 and fastest convergence at 0.025). This is due to the strongly variant input sequence in this case. Since the filter F_2 is much longer than F_1, the effect of approximation (49) becomes visible. Fig. 8 depicts some learning curves for F_2. The noise was again set to -60 dB. One learning curve was obtained for a Gaussian sequence and the other for a K_0 density.

We may finally remark that we have also realized experiments under the same conditions for the so-called filtered-error LMS (least-mean-squares) algorithm. This is an instantaneous-gradient-based estimator that avoids the propagation of a Riccati variable [28]. Not only the conditions for stability and faster convergence speed were the same, but also the learning curves were the same. Hence, although Gauss–Newton type algorithms exhibit a higher convergence speed in the unfiltered case, especially for strongly correlated signals, this behavior is apparently lost in the filtered variants, i.e., when a filter F exists in the error path. In other words, in the filtered case, the Gauss–Newton schemes do not seem to converge faster than the LMS schemes and, hence, the additional complexity required to run Gauss–Newton filters in this case may not be justified. One possible explanation for this is that the Riccati variable $\boldsymbol{P}_i$, which is regarded as an estimate of the autocorrelation matrix of the input vectors $\{\boldsymbol{u}_i\}$, loses this interpretation due to the filtering operation.

9. Concluding remarks

We have provided a time-domain analysis of Gauss–Newton-based adaptive schemes with emphasis on l_2-stability or robustness issues. This was achieved by highlighting an intrinsic feedback structure that arises in this context, and which maps the noise sequence and the initial weight error to the a priori estimation errors and the final weight error. The feedback configuration was motivated via energy arguments and was shown to consist of two major blocks: a time-variant lossless (i.e., energy preserving) feedforward path and a time-variant feedback path. Emphasis was further given to filtered-error variants that give rise to dynamic time-variant feedback loops rather than memoryless loops.

The analysis was carried out in a purely deterministic framework and conditions on the adaptation gains $\mu(i)$ and on the parameters $\beta(i)$ were derived so as to guarantee robustness in the presence of disturbances, in the sense that small energy disturbances would consistently lead to small estimation error energies. Such a property does not hold if the above parameters are not properly chosen.

The discussion was also extended to the filtered-error case, where a filter is included in the update equation. It was shown to lead to a dynamic feedback structure with implications on the convergence speed and robustness performance of the Gauss–Newton updates. Optimal choices for the step-size parameters in the projection filtered-error case were derived and simulation results confirmed the validity of the expressions.

The analysis presented here extends to other classes of algorithms as well, e.g., to block adaptive filters and to updates that involve nonlinear functionals (such as Perceptron training). The results can also be shown to be related to developments in H_∞-theory. Results and connections in this direction are reported in [22, 26, 27].

Appendix A (Proof of Lemma 1)

Multiply (6) by $\boldsymbol{P}_i^{-1/2}$ from the left, and compute the squared norm (i.e., energies) of both sides of the resulting expression, i.e.,

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

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

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

we conclude that the following equality always holds:

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

Substituting recursion (5) for $\boldsymbol{P}_i^{-1}$ in the right-hand side, we obtain (9). □

Appendix B (Proofs of Lemmas 3 and 4)

Proof of Lemma 3. Introduce the notation $\bar{\mu}(i|i-1) = (\boldsymbol{u}_i \boldsymbol{P}_{i-1} \boldsymbol{u}_i^*)^{-1}$. Then, we can write

$$\begin{aligned}\bar{\mu}^{-1}(i) &= \boldsymbol{u}_i \boldsymbol{P}_i \boldsymbol{u}_i^* \\ &= \frac{1}{\lambda(i)}\left(\boldsymbol{u}_i \boldsymbol{P}_{i-1}\boldsymbol{u}_i^* - \frac{(\boldsymbol{u}_i \boldsymbol{P}_{i-1}\boldsymbol{u}_i^*)^2}{\frac{\lambda(i)}{\beta(i)} + \boldsymbol{u}_i \boldsymbol{P}_{i-1}\boldsymbol{u}_i^*}\right) \\ &= \frac{1}{\lambda(i)}\left(\bar{\mu}^{-1}(i|i-1) - \frac{\bar{\mu}^{-2}(i|i-1)}{\frac{\lambda(i)}{\beta(i)} + \bar{\mu}^{-1}(i|i-1)}\right) \\ &= \frac{1}{\beta(i) + \lambda(i)\bar{\mu}(i|i-1)}.\end{aligned}$$

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

Proof of Lemma 4. We use (20) to re-express the update equation (3) as follows:

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

which leads to (21). □

Appendix C (Proofs of Theorems 1 and 2)

Proof of Theorem 1. It follows from (22) that, over $0 \leqslant i \leqslant N$,

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

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

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

Consequently,

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

But it follows from (25), and from the triangular inequality for norms, that

$$\begin{aligned}&\sqrt{\sum_{i=0}^{N} \lambda^{[i+1,N]}\bar{\mu}(i)\,|\bar{v}(i)|^2} \\ &\quad \leqslant \sqrt{\sum_{i=0}^{N} \lambda^{[i+1,N]}\frac{\mu^2(i)}{\bar{\mu}(i)}|v(i)|^2} \\ &\qquad + \sqrt{\sum_{i=0}^{N} \lambda^{[i+1,N]}\left|1 - \frac{\mu(i)}{\bar{\mu}(i)}\right|^2 \bar{\mu}(i)|e_a(i)|^2}.\end{aligned}$$

We thus conclude that

$$\sqrt{\sum_{i=0}^{N} \lambda^{[i+1,N]}(\bar{\mu}(i) - \beta(i))\,|e_a(i)|^2}$$
$$\leqslant \sqrt{\lambda^{[0,N]}\tilde{\boldsymbol{w}}_{-1}^* \boldsymbol{P}_{-1}^{-1}\tilde{\boldsymbol{w}}_{-1}}$$
$$+\sqrt{\sum_{i=0}^{N} \lambda^{[i+1,N]}\frac{\mu^2(i)}{\bar{\mu}(i)}|v(i)|^2}$$
$$+\sqrt{\sum_{i=0}^{N} \lambda^{[i+1,N]}\left|1-\frac{\mu(i)}{\bar{\mu}(i)}\right|^2 \bar{\mu}(i)|e_a(i)|^2}, \qquad (*)$$

and, consequently,

$$\sqrt{\sum_{i=0}^{N} \lambda^{[i+1,N]}(\bar{\mu}(i) - \beta(i))\,|e_a(i)|^2}$$
$$\leqslant \sqrt{\lambda^{[0,N]}\tilde{\boldsymbol{w}}_{-1}\boldsymbol{P}_{-1}^{-1}\tilde{\boldsymbol{w}}_{-1}}$$
$$+\gamma(N)\sqrt{\sum_{i=0}^{N} \lambda^{[i+1,N]}\bar{\mu}(i)|v(i)|^2}$$
$$+\Delta(N)\sqrt{\sum_{i=0}^{N} \lambda^{[i+1,N]}(\bar{\mu}(i) - \beta(i))|e_a(i)|^2}.$$

If $[1-\Delta(N)] > 0$ we conclude from the last inequality that (27) holds. □

Proof of Theorem 2. Expression $(*)$ in the above proof implies that

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

But note that

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

Hence (32) follows. □

References

[1] N.J. Bershad, "Analysis of the normalized LMS algorithm with Gaussian inputs", *IEEE Trans. Acoust. Speech Signal Process.*, Vol. 34, No. 4, August 1986, pp. 793–806.

[2] E. Bjarnason, "Noise cancellation using a modified form of the filtered–XLMS algorithm", *Proc. Eusipco Signal Processing V*, Brussel, 1992.

[3] J.E. Cousseau and P.S.R. Diniz, "On optimal convergence factor for IIR adaptive filters", *Proc. ISCAS*, Vol. 2, May 1994, pp. 137–140.

[4] P.J. Gawthrop, "On the stability and convergence of a self-tuning controller", *Internat. J. Control.*, Vol. 31, 1980, pp. 973–998.

[5] J.R. Glover, Jr., "Adaptive noise canceling applied to sinusoidal interferences", *IEEE Trans. Acoust. Speech Signal Process.*, Vol. 25, No. 6, December 1977, pp. 484–491.

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

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

[8] B. Hassibi, A.H. Sayed and T. Kailath, "Linear estimation in Krein spaces - Part I: Theory", *IEEE Trans. Automat. Control*, Vol. 41, No. 1, January 1996, pp. 18–33.

[9] S. Haykin, *Adaptive Filter Theory*, Prentice-Hall, Englewood Cliffs, NJ, 2nd Edition, 1991.

[10] R. Isermann, *Digital Control Systems*, Springer, Berlin, 1991.

[11] H.K. Khalil, *Nonlinear Systems*, MacMillan, New York, 1992.

[12] 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. Amer.*, Vol. 95, No. 6, June 1994, pp. 3397–3389.

[13] Y.D. Landau, *Adaptive Control: The Model Reference Approach*, Marcel Dekker, NY, 1979.

[14] I.D. Landau, "A feedback system approach to adaptive filtering", *IEEE Trans. Inform. Theory*, Vol. 30, No. 2, March 1984, pp. 251–262.

[15] S. Laugesen and S. Elliott, "Multichannel active control of random noise in a small reverberant room", *IEEE Trans. Speech Audio Process.*, Vol. 1, No. 2, April 1993.

[16] L. Ljung, *System Identification: Theory for the User*, Prentice-Hall, Englewood Cliffs, NJ, 1987.

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

[18] R. Ortega, L. Praly and I.D. Landanu, "Robustness of discrete-time direct adaptive controllers", *IEEE Trans. Automat. Control*, Vol. 30, No. 12, December 1985, pp. 1179–1187.

[19] T. Petillon, A. Gilloire and S. Theodoridis, "The fast Newton transversal filter: An efficient scheme for acoustic echo cancellation in mobile radio", *IEEE Trans. Signal Processing*, Vol. 42, No. 3, March 1994, pp. 509–517.

[20] M. Rupp, "The behavior of LMS and NLMS algorithms in the presence of spherically invariant processes", *IEEE Trans. Signal Processing*, Vol. 41, No. 3, March 1993, pp. 1149–1160.

[21] M. Rupp, "Bursting in the LMS algorithm", *IEEE Trans. Signal Processing*, Vol. 43, No. 10, October 1995, pp. 2414–2417.

[22] M. Rupp and A.H. Sayed, "On the stability and convergence of Feintuch's algorithm for adaptive IIR filtering", *Proc. IEEE Internat. Conf. Acoust. Speech Signal Process.*, Detroit, MI, Vol. 2, May 1995, pp. 1388–1391.

[23] A.H. Sayed, B. Hassibi and T. Kailath, "Inertia conditions for the minimization of quadratic forms in indefinite metric spaces", *Operator Theory: Advances and Applications*, 1996, To appear.

[24] A.H. Sayed and T. Kailath, "A state-space approach to adaptive RLS filtering", *IEEE Signal Process. Mag.*, Vol. 11, No. 3, July 1994, pp. 18–60.

[25] A.H. Sayed and M. Rupp, "Error energy bounds for adaptive gradient algorithms", *IEEE Trans. Signal Processing*, August 1996, to appear.

[26] A.H. Sayed and M. Rupp, "A class of adaptive nonlinear H^{∞}-filters with guaranteed l_2-stability", *Proc. IFAC Symp. on Nonlinear Control System Design*, Tahoe City, CA, Vol. 1, June 1995, pp. 455–460.

[27] A.H. Sayed and M. Rupp, "A feedback analysis of Perceptron learning for neural networks", *Proc. 29th Asilomar Conf. on Signals, Systems, and Computers*, Pacific Grove, CA, October 1995, To appear.

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

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

[30] J.J. Shynk, "Adaptive IIR filtering", *IEEE Acoust. Speech Signal Process. Mag.*, Vol. 6, No. 2, April 1989.

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

[32] M. Vidyasagar, *Nonlinear Systems Analysis*, Prentice-Hall, Englewood Cliffs, NJ, 2nd Edition, 1993.

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