ELSEVIER

Signal Processing 80 (2000) 1787–1794

SIGNAL PROCESSING

www.elsevier.nl/locate/sigpro

Robustness conditions of the LMS algorithm with time-variant matrix step-size

Markus Rupp*, Jürgen Cezanne

Bell-Labs, Lucent Technologies, Wireless Research Laboratory & Digital Technique Department, 791 Holmdel-Keyport Road, Holmdel, NJ 07733-0400, USA

Received 3 November 1999; received in revised form 9 February 2000

Abstract

Gradient-type algorithms commonly employ a scalar step-size, i.e., each entry of the regression vector is multiplied by the same value before updating the coefficients. More flexibility, however, is obtained when this step-size is of matrix size. It allows not only to individually scaling the entries of the regression vector but rotations and decorrelations are possible as well due to the choice of the matrix. A well-known example for the use of a fixed step-size matrix is the Newton–LMS algorithm. For such a fixed step-size matrix, conditions are well known under which a gradient-type algorithm converges. This article, however, presents robustness and convergence conditions for a least-mean-square (LMS) algorithm with time-variant matrix step-size. On the example of a channel estimator used in a cellular hand-phone, it is shown that the choice of a particular step-size matrix leads to considerable improvement over the fixed step-size case. © 2000 Elsevier Science B.V. All rights reserved.

Zusammenfassung

Gradientenalgorithmen nutzen typischerweise skalare Schrittweiten, d.h., vor der Koeffizientenerneuerung wird jeder Eintrag des Regressionsvektors mit dem selben Wert multipliziert. Eine grössere Flexibilität wird jedoch erreicht, wenn diese Schrittweite eine Matrix darstellt. Dies erlaubt nicht nur individuelles Skalieren des Regressionsvektors, sondern darüber hinaus, je nach Wahl der Matrix, auch Rotationen und Dekorrelationen. Ein sehr bekanntes Beispiel für den Gebrauch einer festen Schrittweitenmatrix ist der Newton–LMS Algorithmus. Für solche konstanten Schrittweitenmatrizen sind Bedingungen zur Konvergenz von Gradientenverfahren wohlbekannt. Der vorliegende Artikel jedoch präsentiert Bedingungen für Robustheit und Konvergenz des Least-Mean-Squares (LMS) Algorithmus bei zeit-varianter Schrittweitenmatrix. Am Beispiel eines Kanalschätzers für ein portables Telefon wird gezeigt, wie die Wahl einer speziellen Schrittweitenmatrix zu bedeutender Verbesserung gegenüber einer festen Schrittweite führt. © 2000 Elsevier Science B.V. All rights reserved.

Résumé

Les algorithmes de type gradient emploient communément un pas scalaire, à savoir, chaque entrée du vecteur de régression est multiplié par la même valeur avant la mise à jour des coefficients. Tourefois, l'utilisation d'un pas matriciel offre plus de flexibilité. Ceci permet non seulement une mise à l'échelle individuelle des entrées du vecteur de régression mais des rotations et des décorrélations sont également possibles selon le choix de la matrice. Un exemple bien connu de l'utilisation d'une matrice de pas fixe est l'algorithme Newton–LMS. Pour une telle matrice de pas fixe, les conditions

* Corresponding author. Tel.: + 1-732-888-7104; fax: + 1-732-888-7074.
E-mail address: rupp@lucent.com (M. Rupp).

PII: S0165-1684(00)00088-8

selon lesquelles un algorithme de type gradient converge sont bien connues. Cet article, de son côté, présente les conditions de robustesse et de convergence pour un algorithme à erreur quadratique moyenne minimale (LMS) avec une matrice de pas variant dans le temps. Il est montré sur l'exemple de l'estimateur de canal utilisé dans un téléphone cellulaire que le choix d'une matrice de pas particulière conduit à une amélioration considérable vis-à-vis de l'algorithme à pas fixe.

Keywords: Time-variant step-size matrix; Least-mean-squares algorithm; Channel estimation; Cellular phone

1. Introduction

For many applications, the convergence speed of the least-mean-squares (LMS) algorithm is rather limited but more sophisticated algorithms, like the recursive least-squares (RLS) algorithm are often too complex to implement. RLS algorithms also suffer from other problems like persistent excitation requirements and poor fixed-point properties. In a simple gradient-type algorithm, a matrix step-size instead of a scalar one can be of advantage, in particular when the matrix is of diagonal form. The algorithm then raises from $2N$ complexity to $3N$; this is in most cases affordable. In acoustic-echo cancellation such a successful application of a fixed diagonal matrix step-size has been reported by Makino [4]. There it has been shown that if the matrix entries are chosen proportional to the weights (magnitude) of the unknown system, the convergence speed can be increased considerably. In the case of acoustic echo cancellation, these weights are roughly known since the main echo usually occurs in the first few hundred taps whereas the weights that correspond to longer echo paths decrease approximately exponentially. Makino also showed some statistical convergence conditions. The robustness and in this context a more general statement of convergence of such a gradient-type algorithm with fixed matrix step-size was shown in [8].

In other scenarios, the weights and thus the choice of the matrix step-size might not be known a priori as in the acoustic scenario. In these cases it can be of advantage to couple the diagonal elements of the matrix to the estimated weights (see [9] and references therein). If the matrix is of diagonal form, the algorithm is called PNLMS, for proportionate (weight) NLMS. Simulation results and variants with some improvements compared to the fixed-weight case are reported in [9].

In this paper, the robustness of the matrix step-size LMS algorithm is considered and sufficient conditions on the matrix to guarantee convergence independent of the input statistics are presented. We hereby follow the method in [6–8] and extend the results for the time-variant matrix case. A choice for the elements of a (not necessarily) diagonal matrix will be discussed with statistical approaches. We will further show with some channel estimation examples, as they are common in wireless communications that the algorithm leads to considerable improvements.

2. Robustness of the algorithm

The LMS algorithm with matrix step-size can be written in the following form:

$$\boldsymbol{w}_{k+1} = \boldsymbol{w}_k + \mu(k)\frac{\boldsymbol{A}_k\boldsymbol{x}_k^{\mathrm{H}}}{\boldsymbol{x}_k\boldsymbol{A}_k\boldsymbol{x}_k^{\mathrm{H}}}e(k) \tag{1}$$

together with the error

$$e(k) = d(k) - \boldsymbol{x}_k\boldsymbol{w}_k \tag{2}$$

$$= v(k) + \boldsymbol{x}_k(\boldsymbol{w} - \boldsymbol{w}_k) \tag{3}$$

$$= v(k) + \boldsymbol{x}_k\tilde{\boldsymbol{w}}_k, \tag{4}$$

where the weight-error vector

$$\tilde{\boldsymbol{w}}_k \triangleq \boldsymbol{w} - \boldsymbol{w}_k \tag{5}$$

has been introduced, a performance measure that will be used throughout the paper. The so-called desired signal $d(k)$ consists of additive (observation) noise $v(k)$ and the linear combination of the input signal $x(k)$ and the unknown weights $\boldsymbol{w}$: $d(k) = v(k) + \boldsymbol{x}_k \boldsymbol{w}$. Note that this particular form (1) allows for a time-variant step-size $\mu(k)$ in order to control convergence rate, noise sensitivity, etc. The integration of the matrix $\boldsymbol{A}_k$ in the normalization term leads to a compact notation. We therefore decided to use this particular normalization throughout the paper and only alter the step-size $\mu(k)$. Note further that throughout the paper it is assumed that the matrix $\boldsymbol{A}_k$ is symmetric and positive definite, i.e., all eigenvalues of $\boldsymbol{A}_k$ are positive and real-valued.

Bold-faced small letters for vectors and capital letters for matrices will be used. The input vector $\boldsymbol{x}_k$ is assumed to be of dimension $1 \times M$,

$$\boldsymbol{x}_k = [x(k), x(k-1), \ldots, x(k-M+1)],$$

while the weight vector $\boldsymbol{w}_k$ is of dimension $M \times 1$,

$$\boldsymbol{w}_k^{\mathrm{T}} = [w_1(k), w_2(k), \ldots, w_M(k)].$$

Since the following analysis is purely deterministic, there is no restriction on the structure of $\boldsymbol{x}_k$, thus the series of vectors $\boldsymbol{x}_k$ can have arbitrary statistical dependencies between its elements. Furthermore, "H" will be used to denote conjugate transposition. The expression $\|\cdot\|^2$ will be used for the squared l_2-norm, while $\|\cdot\|_2$ simply denotes the l_2-norm and $\|\cdot\|$ denotes any arbitrary norm.

2.1. Time-invariant matrix $\boldsymbol{A}_k = \boldsymbol{A}$

For a time-invariant matrix $\boldsymbol{A}_k = \boldsymbol{A}$ it is straightforward to derive convergence conditions. By multiplying the update equation (1) with $\boldsymbol{A}^{-1/2}$ from[1] the left one can substitute $\boldsymbol{A}^{-1/2}\tilde{\boldsymbol{w}}_k = \bar{\boldsymbol{w}}_k$ and $\bar{\boldsymbol{x}}_k = \boldsymbol{x}_k \boldsymbol{A}^{1/2}$. The update in terms of the weight-error vector $\tilde{\boldsymbol{w}}_k$ can now be rewritten in terms of the new variables $\bar{\boldsymbol{w}}_k$ and $\bar{\boldsymbol{x}}_k$:

$$\bar{\boldsymbol{w}}_{k+1} = \bar{\boldsymbol{w}}_k - \mu(k) \frac{\bar{\boldsymbol{x}}_k^{\mathrm{H}}}{\|\bar{\boldsymbol{x}}_k\|^2} e(k). \tag{6}$$

In other words, the properties of the LMS algorithm remain unaltered for this case, however, with

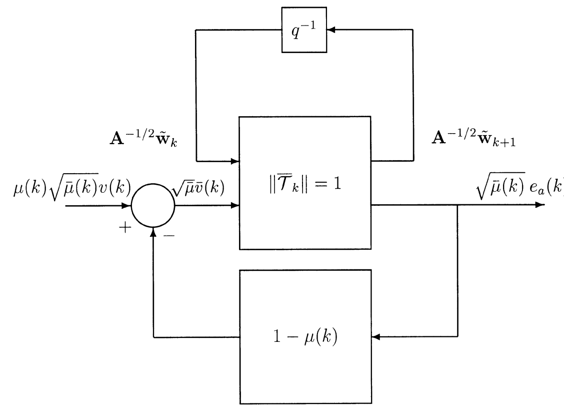

Fig. 1. A time-variant loss-less mapping with gain feedback for the time-constant matrix step-wise algorithm with time-variant step-size $\mu(k)$.

transformed variables. For the update in (6), robustness and convergence is well known. Due to the chosen normalization $1/[\boldsymbol{x}_k \boldsymbol{A} \boldsymbol{x}_k^{\mathrm{H}}]$ convergence is guaranteed in $\boldsymbol{A}^{-1/2}\tilde{\boldsymbol{w}}_k$ rather than in $\tilde{\boldsymbol{w}}_k$. The norm $\|\tilde{\boldsymbol{w}}_k\|_2$ has just been transformed into $\tilde{\boldsymbol{w}}_k^{\mathrm{H}} \boldsymbol{A}^{-1} \tilde{\boldsymbol{w}}_k = \|\tilde{\boldsymbol{w}}\|_{A^{-1}}$. Due to relations between norms, the convergence in l_2-norm can be concluded from the matrix $\boldsymbol{A}$-norm as long as the matrix $\boldsymbol{A}$ is positive definite, i.e., all eigenvalues of $\boldsymbol{A}$ are positive. In [6–8] a deterministic analysis method is described that allows to show how such a mapping (6) is robust even in the presence of observation noise $v(k)$. The main idea is to rewrite the update equation into two energy terms as reflected in Fig. 1. The figure depicts the loss-less instantaneous mapping $\overline{\mathscr{T}}_k$ of the uncertainties $\boldsymbol{A}^{-1/2}\boldsymbol{w}_k$ and $\bar{v}(k)/\sqrt{\boldsymbol{x}_k \boldsymbol{A} \boldsymbol{x}_k^{\mathrm{H}}}$ to the a priori error[2] $e_{\mathrm{a}}(k) \triangleq \boldsymbol{x}_k \tilde{\boldsymbol{w}}_k$. This mapping is defined by a loss-less, energy-preserving relation (see [7,8] for more details)

$$\|\bar{\boldsymbol{w}}_{k+1}\|^2 + \bar{\mu}(k)|e_{\mathrm{a}}(k)|^2 = \|\bar{\boldsymbol{w}}_k\|^2 + \bar{\mu}(k)|\bar{v}(k)|^2, \tag{7}$$

where the short form $\bar{\mu}(k) = 1/[\boldsymbol{x}_k \boldsymbol{A} \boldsymbol{x}_k^{\mathrm{H}}]$ has been used. The unit time delay is denoted as the operator

[1] Any regular matrix $\boldsymbol{A} = \boldsymbol{V}\Lambda\boldsymbol{V}^{-1}$, Λ being a diagonal matrix with the eigenvalues of the matrix $\boldsymbol{A}$. The notation $\boldsymbol{A}^{1/2} = \boldsymbol{V}\Lambda^{1/2}\boldsymbol{V}^{-1}$.

[2] Note that in some adaptive filter literature it is also common to denote the a priori error as the disturbed version, i.e., $e_{\mathrm{a}}(k) + v(k)$.

q^{-1} in the figure. Note that this mapping does not use the original noise term $v(k)$ but a new noise term composed of the original one and a feedback part from the a prior error:

$$\bar{v}(k) \triangleq \mu(k)v(k) - (1-\mu(k))e_a(k). \quad (8)$$

Summing up the terms from $k=0$ to N leads to the global property

$$\|\bar{\boldsymbol{w}}_{N+1}\|^2 + \sum_{k=0}^{N} \bar{\mu}(k)|e_a(k)|^2 = \|\bar{\boldsymbol{w}}_0\|^2 + \sum_{k=0}^{N} \bar{\mu}(k)|\bar{v}(k)|^2. \quad (9)$$

The global property still holds from the original observation noise $v(k)$ to the a priori error $e_a(k)$ in inequality for step-sizes $0<\mu(k)<1$ as is demonstrated in [6,7]. There it is also shown that similar relations are valid from the original observation noise $v(k)$ to the a priori error $e_a(k)$, the most important one shown here:

$$\|\bar{\boldsymbol{w}}_{N+1}\|^2 + \sum_{k=0}^{N} \mu(k)\bar{\mu}(k)|e_a(k)|^2 \leqslant \|\bar{\boldsymbol{w}}_0\|^2 + \sum_{k=0}^{N} \mu(k)\bar{\mu}(k)|v(k)|^2. \quad (10)$$

2.2. Contracting matrix A_k

If the matrix $\boldsymbol{A}_k$ is of contracting nature, i.e., $\|\boldsymbol{A}_k\| \leqslant \|\boldsymbol{A}_{k-1}\|$, the global property (9) holds in inequality since $\tilde{\boldsymbol{w}}_k \boldsymbol{A}_k^{-1}\tilde{\boldsymbol{w}}_k^{\mathrm{H}} \geqslant \tilde{\boldsymbol{w}}_k \boldsymbol{A}_{k-1}^{-1}\tilde{\boldsymbol{w}}_k^{\mathrm{H}}$ and

$$\|\bar{\boldsymbol{w}}_{N+1}\|^2 + \sum_{k=0}^{N} \bar{\mu}(k)|e_a(k)|^2 \leqslant \|\bar{\boldsymbol{w}}_0\|^2 + \sum_{k=0}^{N} \bar{\mu}(k)|\bar{v}(k)|^2, \quad (11)$$

now with redefined value $\bar{\mu}(k) = 1/[\boldsymbol{x}_k\boldsymbol{A}_k\boldsymbol{x}_k^{\mathrm{H}}]$, is obtained. Again, applying the techniques from [6,7], similar contracting properties hold for $0<\mu(k)<1$ as can be readily shown. Does this exclude expanding matrices $\boldsymbol{A}_k$? The answer is no as the following theorem will show.

2.3. General time-variant matrix A_k

In the case of a time-variant matrix, the properties of the update equation are of a different nature. Since the matrix $\boldsymbol{A}_k$ changes with every time instant, a fixed matrix norm does not exist and therefore a conclusion to the l_2-norm is not possible. Other properties can be derived instead. As a first property, the following theorem on the boundedness of the parameter error vector will be proven.

Theorem 1 (Boundedness). *The parameter error vectors $\tilde{\boldsymbol{w}}_k = \boldsymbol{w} - \boldsymbol{w}_k$ of the matrix step-size LMS algorithm given in* (1) *with fixed step-size μ are bounded if the matrix $\boldsymbol{A}_k \in \mathbb{R}^{M\times M}$ and the additive noise satisfy the following requirements*

(1) *$\boldsymbol{A}_k$ is positive definite[3] with $\boldsymbol{0} < \sigma\boldsymbol{I} < \boldsymbol{A}_k < \tau\boldsymbol{I} < \infty$*
   (a) *$\boldsymbol{A}_k$ converges such that $\sum_{k=0}^{\infty}\|\boldsymbol{A}_k^{1/2} - \boldsymbol{A}_{k+1}^{1/2}\| < M_1 < \infty$*
   (b) *or, in the case that $\boldsymbol{A}_k$ is of diagonal form, $\boldsymbol{A}_k$ converges such that $\sum_{k=0}^{\infty}\|\boldsymbol{A}_k^L - \boldsymbol{A}_{k+1}^L\| < M_2 < \infty$ for arbitrary L.*

(2) *The l_1-norm of the noise is bounded such that $\mu\sum_{k=0}^{\infty}|v(k)(|/\sqrt{\boldsymbol{x}_k\boldsymbol{A}_k\boldsymbol{x}_k^{\mathrm{H}}}) < M_3 < \infty$.*

In order to prove the theorem, we start with (6) here rewritten in terms of $\tilde{\boldsymbol{w}}_k$ and $\boldsymbol{x}_k$ for convenience:

$$\boldsymbol{A}_k^{-1/2}\tilde{\boldsymbol{w}}_{k+1} = \boldsymbol{A}_k^{-1/2}\tilde{\boldsymbol{w}}_k - \mu\frac{\boldsymbol{A}_k^{1/2}\boldsymbol{x}_k^{\mathrm{H}}}{\boldsymbol{x}_k\boldsymbol{A}_k\boldsymbol{x}_k^{\mathrm{H}}}e(k). \quad (12)$$

By expanding the error signal $e(k)$, this expression can be reformulated as

$$\boldsymbol{A}_k^{-1/2}\tilde{\boldsymbol{w}}_{k+1} = \left[\mathrm{I} - \mu\frac{\boldsymbol{A}_k^{1/2}\boldsymbol{x}_k^{\mathrm{H}}\boldsymbol{x}_k\boldsymbol{A}_k^{\mathrm{H}/2}}{\boldsymbol{x}_k\boldsymbol{A}_k\boldsymbol{x}_k^{\mathrm{H}}}\right]\boldsymbol{A}_k^{-1/2}\tilde{\boldsymbol{w}}_k - \mu\frac{\boldsymbol{A}_k^{1/2}\boldsymbol{x}_k^{\mathrm{H}}}{\boldsymbol{x}_k\boldsymbol{A}_k\boldsymbol{x}_k^{\mathrm{H}}}v(k). \quad (13)$$

The term $[\boldsymbol{I} - \mu\boldsymbol{A}_k^{1/2}\boldsymbol{x}_k^{\mathrm{H}}\boldsymbol{x}_k\boldsymbol{A}_k^{\mathrm{H}/2}/\boldsymbol{x}_k\boldsymbol{A}_k\boldsymbol{x}_k^{\mathrm{H}}]$ is contracting as long as $0<\mu<2$. Thus, computing the l_2-norm of (12) leads to the following inequality:

$$\|\boldsymbol{A}_k^{-1/2}\tilde{\boldsymbol{w}}_{k+1}\|_2 \leqslant \|\boldsymbol{A}_k^{-1/2}\tilde{\boldsymbol{w}}_k\|_2 + \mu\frac{|v(k)|}{\sqrt{\boldsymbol{x}_k\boldsymbol{A}_k\boldsymbol{x}_k^{\mathrm{H}}}}. \quad (14)$$

[3] The notation $\boldsymbol{A} < \boldsymbol{B}$ means $\mathrm{eig}(\boldsymbol{A}-\boldsymbol{B}) < 0$, see for example [3].

Also utilizing a well-known norm property

$$\|A_{k+1}^{-1/2}\tilde{w}_{k+1}\|_2 \leqslant \|A_k^{1/2}A_{k+1}^{-1/2}\|_2 \|A_k^{-1/2}\tilde{w}_{k+1}\|_2 \quad (15)$$

leaves us with

$$\|A_{k+1}^{-1/2}\tilde{w}_{k+1}\|_2 \leqslant \|A_k^{1/2}A_{k+1}^{-1/2}\|_2 \left(\|A_k^{-1/2}\tilde{w}_k\|_2 + \mu \frac{|v(k)|}{\sqrt{x_k A_k x_k^{\mathrm{H}}}} \right). \quad (16)$$

If the iteration is followed backwards from time instant N to its beginning at time instant $k = 0$, one obtains

$$\|A_{N+1}^{-1/2}\tilde{w}_{N+1}\|_2 \leqslant \prod_{k=0}^{N} \|A_k^{1/2}A_{k+1}^{-1/2}\|_2 \|A_0^{-1/2}\tilde{w}_0\|_2 + \mu \sum_{l=0}^{N} \prod_{k=l}^{N} \|A_k^{1/2}A_{k+1}^{-1/2}\|_2 \frac{|v(l)|}{\sqrt{x_k A_k x_k^{\mathrm{H}}}}. \quad (17)$$

The product term can be rewritten as

$$\prod_{k=l}^{N} \|A_k^{1/2}A_{k+1}^{-1/2}\|_2$$

$$= \prod_{k=l}^{N} \|I - [I - A_k^{1/2}A_{k+1}^{-1/2}]\|_2 \quad (18)$$

$$\leqslant \prod_{k=l}^{N} (1 + \|I - A_k^{1/2}A_{k+1}^{-1/2}\|_2) \quad (19)$$

$$\leqslant \mathrm{e}^{\sum_{k=l}^{N} \|I - A_k^{1/2}A_{k+1}^{-1/2}\|_2} \quad (20)$$

$$\leqslant \mathrm{e}^{\sum_{k=l}^{N} \|A_{k+1}^{-1/2}\|_2 \|A_k^{1/2} - A_{k+1}^{1/2}\|_2} \quad (21)$$

$$\leqslant \mathrm{e}^{1/\sqrt{\sigma} \sum_{k=l}^{N} \|A_k^{1/2} - A_{k+1}^{1/2}\|_2}. \quad (22)$$

From here it can be concluded that the products are bounded if the sum

$$\sum_{k=l}^{N} \|A_k^{1/2} - A_{k+1}^{1/2}\|_2 < \infty$$

for every N. Due to the norm relations, for diagonal matrices A_k, it is equivalent to require that

$$\prod_{k=l}^{N} \|A_k^L A_{k+1}^{-L}\| < M_2^L \quad (23)$$

which leads to the condition that $\sum_{k=l}^{N} \|A_k^L - A_{k+1}^L\|_2 < \infty$.

If it is assumed furthermore that the noise $\mu \sum |v(k)| / \sqrt{x_k A_k x_k^{\mathrm{H}}} < M_3 < \infty$, it can be concluded that $\|A_{N+1}^{-1/2}\tilde{w}_{N+1}\|_2$ is bounded, and since A_k is positive definite the parameter error vectors $\tilde{w}_k$ are bounded as well. Note that the condition (1a) on the matrix is not very restrictive. Even a matrix of the form $A_k = B - 1/(K+1)C$ with $B - C > 0$, B, C diagonal, i.e., a matrix whose sum is not convergent or bounded, even such a matrix satisfies the requirements since

$$\sum_{k=0}^{\infty} \|A_k - A_{k+1}\|_2 = \|C\|_2 \sum_{k=0}^{\infty} \frac{1}{(k+1)(k+2)} < M < \infty.$$

This result now enables us to give convergence conditions of the algorithm.

Theorem 2 (Convergence). *The matrix step-size LMS algorithm given in* (1) *with constant step-size* μ *is robust in the* l_2 *sense and the a priori errors* $e_{\mathrm{a}}(k)$ *converge to zero if the matrix* A_k *and the additive noise satisfy the conditions from Theorem* 1 *and the step-size* μ *remains in the range* (0,1].

Computing the squared l_2 norm of Eq. (12) for $0 < \mu \leqslant 1$ leads to

$$\tilde{w}_{k+1}^{\mathrm{H}} A_k^{-1} \tilde{w}_{k+1} + \mu \frac{|x_k \tilde{w}_k|^2}{x_k A_k x_k^{\mathrm{H}}} \leqslant \tilde{w}_k^{\mathrm{H}} A_k^{-1} \tilde{w}_k + \mu \frac{|v(k)|^2}{x_k A_k x_k^{\mathrm{H}}}. \quad (24)$$

The first term in (24) can be rewritten

$$\tilde{w}_{k+1}^{\mathrm{H}} A_k^{-1} \tilde{w}_{k+1} = \tilde{w}_{k+1}^{\mathrm{H}} A_{k+1}^{-1} \tilde{w}_{k+1} + \tilde{w}_{k+1}^{\mathrm{H}} [A_k^{-1} - A_{k+1}^{-1}] \tilde{w}_{k+1}$$

$$= \tilde{w}_{k+1}^{\mathrm{H}} A_{k+1}^{-1} \tilde{w}_{k+1} + \Delta_k. \quad (25)$$

It can be shown that (see Gradshteyn 0.228 [2]) the Δ_k converges if $\sum_{k=0}^{\infty} [A_k^{-1} - A_{k+1}^{-1}]$ converges and $\|\tilde{w}_k\|_2 \leqslant M$. The first condition is readily shown by

$$\sum_{k=0}^{\infty} [A_k^{-1} - A_{k+1}^{-1}] = A_0^{-1} - A_{\infty}^{-1}. \quad (26)$$

Due to the requirements in Theorem 1, the matrix A_k is positive definite in the bounds (σ,τ), its inverse is also bounded and therefore $A_0^{-1} - A_{\infty}^{-1}$ must be finite.

With these conditions the remaining terms become easier to handle. Now (24) can be rewritten as

$$\tilde{\boldsymbol{w}}_{N+1}^{\mathrm{H}}\boldsymbol{A}_{N+1}^{-1}\tilde{\boldsymbol{w}}_{N+1} + \mu\sum_{k=0}^{N}\frac{|\boldsymbol{x}_k\tilde{\boldsymbol{w}}_k|^2}{\boldsymbol{x}_k\boldsymbol{A}_k\boldsymbol{x}_k^{\mathrm{H}}} \leqslant \tilde{\boldsymbol{w}}_0^{\mathrm{H}}\boldsymbol{A}_0^{-1}\tilde{\boldsymbol{w}}_0 + \mu\sum_{k=0}^{N}\frac{|v(k)|^2}{\boldsymbol{x}_k\boldsymbol{A}_k\boldsymbol{x}_k^{\mathrm{H}}}M_2, \tag{27}$$

where M_2 is the constant, which the term $\sum_{k=0}^{\infty}\Delta_k$ converges to. Note that the terms $\tilde{\boldsymbol{w}}_{k+1}^{\mathrm{H}}[\boldsymbol{A}_k^{-1} - \boldsymbol{A}_{k+1}^{-1}]\tilde{\boldsymbol{w}}_{k+1}$ disappear in the constant matrix case and case be handled easily for a monotonically decreasing matrix $\boldsymbol{A}_k$.

As long as the sequence of noise is bounded (which we already assumed above), the sequence of a priori errors is bounded as well, and they will form a Cauchy series. Thus,

$$\frac{|e_{\mathrm{a}}(k)|^2}{\boldsymbol{x}_k\boldsymbol{A}_k\boldsymbol{x}_k^{\mathrm{H}}} \to 0. \tag{28}$$

With the conditions on $\boldsymbol{A}_k$ it can be concluded that $e_{\mathrm{a}}(k)\to 0$.

3. Matrix selection

Although some important properties of the time-variant matrix step-size have been derived, we are not yet able to explain why a diagonal step-size proportional to the weight vector as proposed in the PNLMS algorithm is a good choice. Assuming that $x(k)$ is random white sequence with constant modulus property as in typical wireless scenarios, the variance $\boldsymbol{K}_k = E[\tilde{\boldsymbol{w}}_k\tilde{\boldsymbol{w}}_k^{\mathrm{H}}]$ of the parameter error vector can be written as

$$\boldsymbol{K}_{k+1} = \boldsymbol{K}_k - \frac{\boldsymbol{A}_k\boldsymbol{K}_k + \boldsymbol{K}_k\boldsymbol{A}_k}{\mathrm{trace}\boldsymbol{A}_k} + \frac{\boldsymbol{A}_k^2}{\mathrm{trace}^2\boldsymbol{A}_k}\sigma_v^2 + \frac{\boldsymbol{A}_k}{\mathrm{trace}\boldsymbol{A}_k}[\boldsymbol{I}\,\mathrm{trace}(\boldsymbol{K}_k) + \boldsymbol{K}_k - \mathrm{diag}(\boldsymbol{K}_k)] \times\frac{\boldsymbol{A}_k}{\mathrm{trace}\boldsymbol{A}_k}, \tag{29}$$

where diagonal matrices $\boldsymbol{A}_k$ were assumed. The derivation for this follows the presentation in [1,5] for Gaussian and spherically invariant processes. In the case of constant modulus sequences the fourth-order terms are different: $E[\boldsymbol{x}_k^{\mathrm{H}}\boldsymbol{x}_k\boldsymbol{K}_k\boldsymbol{x}_k^{\mathrm{H}}\boldsymbol{x}_k] = [\boldsymbol{I}\ \mathrm{trace}(\boldsymbol{K}_k) + \boldsymbol{K}_k - \mathrm{diag}(\boldsymbol{K}_k)]$.

Due to the constant modulus property the expression $E[1/[\boldsymbol{x}_k\boldsymbol{A}_k\boldsymbol{x}_k^{\mathrm{H}}]] = 1/\mathrm{trace}\boldsymbol{A}_k$ is independent of the statistics. If only the diagonal terms of $\boldsymbol{K}_k$ are written into a vector $\tilde{\boldsymbol{k}}_k$ as

$$\tilde{\boldsymbol{k}}_{k+1} = \left(\boldsymbol{I} - 2\frac{\boldsymbol{A}_k}{\mathrm{trace}\boldsymbol{A}_k} + \frac{\boldsymbol{A}_k^2}{\mathrm{trace}^2\boldsymbol{A}_k}\boldsymbol{1}\boldsymbol{1}^{\mathrm{T}}\right)\tilde{\boldsymbol{k}}_k + \frac{\boldsymbol{A}_k^2}{\mathrm{trace}^2\boldsymbol{A}_k}\boldsymbol{1}\sigma_v^2, \tag{30}$$

where $\boldsymbol{1}$ denotes a column vector with ones as elements. Now consider one element $\tilde{k}(k)$ of $\tilde{\boldsymbol{k}}_k$. It evolves over time by approximately $(1 - a_l(k)/\mathrm{trace}(\boldsymbol{A}_k))^2\tilde{k}^2(k)$, where $a_l(k)$ denote the diagonal elements of the matrix $\boldsymbol{A}_k$. The convergence rate is given by the terms $(1 - a_l(k)/\mathrm{trace}(\boldsymbol{A}_k))^2$. For a fixed step-size the parameter errors all converge with the same rate. If we like to speed up the convergence of the larger errors, the convergence speed should be proportional to the error magnitude itself. This can be achieved, if $a_l(k)$ is proportional to $|\tilde{k}(k)|$. A choice thus could be $a_l(k) = |\tilde{k}(k)|$. However, the errors are not known. Only in the beginning of the update with zero initialization of $\boldsymbol{w}_0$, the errors are identical to the weights of the unknown system. With continuing adaptation, the estimates will become closer and closer to the true values. Larger weights also cause larger parameter errors. Thus, a possible solution is to set $\boldsymbol{A}_k = \mathrm{diag}(|w_k|)$.

After the initialization, it can therefore be of advantage to switch back to a more standard-mode for the step-size, i.e., a diagonal matrix with identical entries. This can avoid slow convergence of the smaller parameter errors. In practice, the range for the elements of $\boldsymbol{A}_k$ will be lower bounded by some small constant (see [9] and the following example).

4. Channel estimation example

The following example is in the context of cellular TDMA phones as described in the IS-54, IS-136, and IS-137 standards (see for example [10]). The standards require bit-error rates (BER) on the down-link, from base station to mobile receiver, of lower than 3% at 22 dB SNR (this assumes a noise figure of 8 dB for the receiver) at two path Rayleigh fading channels with $T/4$, $T/2$ and T delay spread at a Doppler frequency of 181 Hz. This Doppler frequency corresponds to a vehicle moving with 100 km/h when transmitting with 1880 MHz carrier frequency. The receiver of a cellular handphone was designed to run with a Viterbi decoder based on channel estimates of an LMS algorithm. In simulation runs a fixed step-size ($\mu = 0.24$) was optimized to minimize the BER for the three different delay spreads mentioned above. The BER result from averaging over 500 slots of data. Table 1 compares these results with those of the new step-side algorithm. A considerable improvement for all three cases is achieved. In particular, cases with higher delay spread show much better performance now. The values in the column entitled "matrix step-size" are obtained for the algorithm including divisions and square roots. A DSP version without complicated mathematical operations has been developed. As indicated in Table 1 by "approximation", it shows for most cases even better performance.

The algorithm has been rewritten so that it fits perfectly the needs for a fixed-point DSP implementation. An additional step-size of $\mu = 0.9$ has been used. Suppose the weights are given by w(1), w(2), w(3) and the diagonal entries of the matrix $\boldsymbol{A}_k$ are given by A(1), A(2), A(3). The final algorithm is given by the following Matlab code: Notice that the division of the term norm_A

```
norm_A = 0.1;
for i = 1:3
    A(i) = abs(real(w(i)))+abs(imag(w(i)));
    A(i) = 0.45+A(i)*0.5;
    A(i) = min(2,max(w(i),0.1));
    norm_A = norm_A+A(i);
    end;
if (norm_A < 1.5)
    norm_A = 2-norm_A;
  else
    norm_A = 0.75 - 0.125*norm_A;
    end;
for i = 1:3
    A(i) = A(i)*norm_A;
    end;
```

Table 1
Bit-error rates with fixed step-size and matrix step-size

Channel condition	Fixed step-size (%)	Matrix step-size (%)	Approximation (%)
$T/4$ delay spread	2.6	2.3	2.2
$T/2$ delay spread	3.6	2.5	2.7
T delay spread	2.9	1.9	2.1

is replaced by a one-step iteration. The operation A(i) = 0.45+A(i)*0.5; approximates the square-root-operation. As mentioned in [9] bounds of the norm (norm_A $\geqslant 0.1$) lead to more stable behaviour.

5. Conclusion

In this paper, conditions on the time variations of a matrix step-size in gradient-type stochastic algorithms were derived based on l_2-stability criterions. A motivation for the choice of diagonal matrices was discussed and on an example of a cellular hand set, the advantages of the method were presented.

References

[1] A. Feuer, E. Weinstein, Convergence analysis of LMS filters with uncorrelated Gaussian data, IEEE Trans. Acoust. Speech Signal Process. ASSP-33 (1) (February 1985) 222–230.

[2] I.S. Gradshteyn, I.M. Ryzhik, Table of Integrals, Series and Products, Academic Press, New York, 1980.

[3] S. Haykin, Adaptive Filter Theory, Prentice-Hall, Englewood Cliff, NJ, 1996.

[4] S. Makino, Y. Kaneda, N. Koizumi, Exponentially weighted step-size NLMS adaptive filter based on the statistics of a room impulse response, IEEE Trans. Speech Audio Process. 1 (1) (January 1993) 101–108.

[5] M. Rupp, The behavior of LMS and NLMS algorithms in the presence of spherically invariant processes, IEEE Trans. Signal Process. SP-41 (3) (March 1993) 1149–1160.

[6] M. Rupp, A.H. Sayed, A time-domain feedback analysis of filtered-error adaptive gradient algorithms, IEEE Trans. Signal Process. 44 (6) (June 1996) 1428–1440.

[7] M. Rupp, A.H. Sayed, V.K. Madisetti, D.B. Williams (Eds.), Robustness in Adaptive Filters, in: The DSP Handbook, CRC Press, Boca Raton, 1997.
[8] A.H. Sayed, M. Rupp, Error energy bounds for adaptive gradient algorithms, IEEE Trans. Signal Process. 44 (8) (August 1996) 1982–1989.
[9] Steven L. Gay, An efficient, fast converging adaptive filter for network echo cancellation, Proceedings of Asilomar Conference, November 1998, pp. 394–398.
[10] TIA/EIA/Interim Standard IS 137, Telecommunication Association, December 1994.