ON GRADIENT TYPE ADAPTIVE FILTERS WITH NON-SYMMETRIC MATRIX STEP-SIZES

Markus Rupp, *Senior Member IEEE*
Vienna University of Technology, Institute of Telecommunications
Gusshausstr. 25/389, 1040 Vienna, Austria

***Abstract*—In this contribution we provide a thorough stability analysis of gradient type algorithms with non-symmetric matrix step-sizes. We hereby extend existing analyses for symmetric matrix step-sizes and present several methods to derive step-size bounds. Although we can guarantee the l_2-stability for such algorithms only under very restrictive conditions, we are able to proof convergence in the mean square sense under much more general conditions. Some of the derived step-size bounds turn out to very tight and allow for accurate algorithmic design.**

I. Introduction

Due to their low complexity and numerical robustness gradient type adaptive filter algorithms play the most important role when it comes to implementations. Their relatively low convergence rate is often overcome by clever step-size mechanisms. In literature matrix step-sizes have been proposed for speeding up convergence in echo compensation [1], [7] where convergence in the mean square sense was shown for diagonal matrices. The choice of a matrix as the inverse of the autocorrelation matrix of the driving process x_k is known as the Newton-LMS [2], [3] allowing to decorrelate the input process and thus to speed up convergence. Theoretical investigations of such algorithms treat positive definite matrices in the context of l_2-stability for time-invariant [4] and a class of time-variant matrices [5]. We extend such analysis here to the case of arbitrary non-symmetric but time-invariant matrices, e.g. for decorrelating an input process with low complexity or in the context of adaptive equalizer design [6]. The notion of time-variance may only be incorporated by a time-variant scalar step-size μ_k.

Consider the classical reference system for noise system identification with input $\mathbf{x}$ and plant $\mathbf{w}_o \in \mathbb{C}^{M\times 1}$ with the desired outcome

$$d_k = \mathbf{x}_k^T \mathbf{w}_o + v_k$$

additively disturbed by noise v_k of variance N_o. A gradient type algorithm for estimating $\mathbf{w}_o$ is given by

$$\hat{\mathbf{w}}_k = \hat{\mathbf{w}}_{k-1} + 2\mu_k \mathbf{G}\mathbf{x}_k^*[d_k - \mathbf{x}_k^T\hat{\mathbf{w}}_{k-1}]; k = 1, 2, ... \quad (1)$$

for which we included an arbitrary time-invariant matrix $\mathbf{G}$ together with a time-variant scalar step-size μ_k. We subtract $\mathbf{w}_o$ from its estimate and use only the parameter error vector $\tilde{\mathbf{w}}_k = \mathbf{w}_o - \hat{\mathbf{w}}_k$:

$$\begin{aligned} \tilde{\mathbf{w}}_k &= \tilde{\mathbf{w}}_{k-1} - 2\mu_k \mathbf{G}\mathbf{x}_k(d_k - \mathbf{x}_k^H\hat{\mathbf{w}}_{k-1}) & (2)\\ &= \tilde{\mathbf{w}}_{k-1} - 2\mu_k \mathbf{G}\mathbf{x}_k \underbrace{(\mathbf{x}_k^H\tilde{\mathbf{w}}_{k-1} + v_k)}_{\tilde{e}_{a,k}=e_{a,k}+v_k}. & (3)\end{aligned}$$

II. Stability Analysis

A. Analysis Method A

The central question of this paper is under which conditions such an adaptive gradient type algorithm will converge. We introduce an additional square matrix $\mathbf{F} \in \mathbb{C}^{M\times M}$ that we multiply from the left to obtain a modified update equation

$$\mathbf{F}\tilde{\mathbf{w}}_k = \mathbf{F}\tilde{\mathbf{w}}_{k-1} - 2\mu_k\mathbf{F}\mathbf{G}\mathbf{x}_k\tilde{e}_{a,k}; k = 1, 2, ... \quad (4)$$

A straightforward idea is now to compute the parameter vector error $\|\tilde{\mathbf{w}}_k\|_2^2$ in the light of such matrix $\mathbf{F}$ that is $\|\tilde{\mathbf{w}}_k\|_{\mathbf{F}}^2 = \tilde{\mathbf{w}}_k^H\mathbf{F}^H\mathbf{F}\tilde{\mathbf{w}}_k$. This weighted square Euclidean norm requires that $\mathbf{F}^H\mathbf{F} > 0$ is being positive definite or in this case equivalently $\mathbf{F}^H\mathbf{F}$ is being of full rank, which is a first condition and restriction on $\mathbf{F}$, obtaining:

$$\begin{aligned} \|\tilde{\mathbf{w}}_k\|_{\mathbf{F}}^2 &= \|\tilde{\mathbf{w}}_{k-1}\|_{\mathbf{F}}^2 - 2\mu_k\bar{e}_{A,k}^*\tilde{e}_{a,k} - 2\mu_k\bar{e}_{A,k}\tilde{e}_{a,k}^* \\ &\quad + 4\frac{\mu_k^2}{\bar{\mu}_{A,k}}|\tilde{e}_{a,k}|^2, & (5)\end{aligned}$$

where we employed the following notation

$$\begin{aligned} \bar{e}_{A,k} &= \mathbf{x}_k^H\mathbf{G}^H\mathbf{F}^H\mathbf{F}\tilde{\mathbf{w}}_{k-1}, & (6)\\ \bar{\mu}_{A,k} &= \frac{1}{\mathbf{x}_k^H\mathbf{G}^H\mathbf{F}^H\mathbf{F}\mathbf{G}\mathbf{x}_k}. & (7)\end{aligned}$$

We introduce a proportionality factor $\lambda_{A,k}$ such that $\bar{e}_{A,k} = \lambda_{A,k}^* e_{a,k}$, which allows the simplification of (5) into

$$\begin{aligned} \|\tilde{\mathbf{w}}_k\|_{\mathbf{F}}^2 &= \|\tilde{\mathbf{w}}_{k-1}\|_{\mathbf{F}}^2 + 4\frac{\mu_k^2}{\bar{\mu}_{A,k}}|\tilde{e}_{a,k}|^2 & (8)\\ &\quad -2\mu_k\left[\lambda_{A,k}e_{a,k}^*\tilde{e}_{a,k} + \lambda_{A,k}^*e_{a,k}\tilde{e}_{a,k}^*\right]. \end{aligned}$$

Next to the additive noise v_k we can now form a new variable u_k:

$$\begin{aligned} |v_k|^2 &= |\tilde{e}_{a,k} - e_{a,k}|^2 & (9)\\ &= |\tilde{e}_{a,k}|^2 + |e_{a,k}|^2 - \tilde{e}_{a,k}e_{a,k}^* - \tilde{e}_{a,k}^*e_{a,k} \\ |u_k|^2 &= |\lambda_{A,k}\tilde{e}_{a,k} - e_{a,k}|^2 & (10)\\ &= |\lambda_{A,k}\tilde{e}_{a,k}|^2 + |e_{a,k}|^2 - \lambda_{A,k}\tilde{e}_{a,k}e_{a,k}^* - \lambda_{A,k}^*\tilde{e}_{a,k}^*e_{a,k}, \end{aligned}$$

which allows to reformulate (8)

$$\begin{aligned} \|\tilde{\mathbf{w}}_k\|_{\mathbf{F}}^2 &= \|\tilde{\mathbf{w}}_{k-1}\|_{\mathbf{F}}^2 + 4\frac{\mu_k^2}{\bar{\mu}_{A,k}}|\tilde{e}_{a,k}|^2 & (11)\\ &\quad -2\mu_k[|u_k|^2 - |\lambda_{A,k}|^2|\tilde{e}_{a,k}|^2 - |e_{a,k}|^2]. \end{aligned}$$

We can further bound $|u_k|^2$ by:

$$\begin{aligned} |u_k|^2 &= |\lambda_{A,k}\tilde{e}_{a,k} - e_{a,k}|^2 & (12)\\ &= |\lambda_{A,k}\tilde{e}_{a,k} - \tilde{e}_{a,k} + v_k|^2 & (13)\\ &\le (1+\gamma)|v_k|^2 + \frac{1+\gamma}{\gamma}|1-\lambda_{A,k}|^2|\tilde{e}_{a,k}|^2 \end{aligned}$$

for some positive value $\gamma > 0$ which in turn allows now to write

$$\begin{aligned} \|\tilde{\mathbf{w}}_k\|_{\mathbf{F}}^2 &+ 2\mu_k|e_{a,k}|^2 \le \|\tilde{\mathbf{w}}_{k-1}\|_{\mathbf{F}}^2 + 4\frac{\mu_k^2}{\bar{\mu}_{A,k}}|\tilde{e}_{a,k}|^2 \\ &+ 2\mu_k(1+\gamma)|v_k|^2 & (14)\\ &+ 2\mu_k\left[\frac{1+\gamma}{\gamma}|1-\lambda_{A,k}|^2 - |\lambda_{A,k}|^2\right]|\tilde{e}_{a,k}|^2. \\ &= \|\tilde{\mathbf{w}}_{k-1}\|_{\mathbf{F}}^2 + 2\mu_k(1+\gamma)|v_k|^2 & (15)\\ &\quad +2\mu_k\delta_{A,k}|\tilde{e}_{a,k}|^2 \end{aligned}$$

This work has been funded by the NFN SISE project (National Research Network Signal and Information Processing in Science and Engineering).

If the term

$$\delta_{A,k} = \frac{2\mu_k}{\bar{\mu}_{A,k}} + 1 - 2\Re\{\lambda_{A,k}\} + \frac{1}{\gamma}|1-\lambda_{A,k}|^2.$$

is negative or $0 < \frac{\mu_k}{\bar{\mu}_{A,k}} = \alpha < \Re\{\lambda_{A,k}\} - \frac{1}{2} - \frac{1}{2\gamma}|1-\lambda_{A,k}|^2$, the last term in (15) can simply be dropped and we obtain a first local stability condition relating the update from time instant $k-1$ to k:

Lemma 2.1: The adaptive gradient type algorithm with Update (3) exhibits the following local robustness properties from its inputs $\{\tilde{\mathbf{w}}_{k-1}, \sqrt{2\mu_k(1+\gamma)}v_k\}$ to its outputs $\{\tilde{\mathbf{w}}_k, \sqrt{2\mu_k}e_{a,k}\}$:

$$\frac{\|\tilde{\mathbf{w}}_k\|^2_{\mathbf{F}} + 2\mu_k|e_{a,k}|^2}{\|\tilde{\mathbf{w}}_{k-1}\|^2_{\mathbf{F}} + 2(1+\gamma)\mu_k|v_k|^2} \leq 1 \tag{16}$$

as long as μ_k can be selected so that $0 < \frac{\mu_k}{\bar{\mu}_{A,k}} = \alpha < \Re\{\lambda_{A,k}\} - \frac{1}{2} - \frac{1}{2\gamma}|1-\lambda_{A,k}|^2$ for some $\gamma > 0$, and $\mathbf{F}^H\mathbf{F} > 0$.

Such a local robustness property however is only useful if it can be extended towards a global property. To this end we sum up the energy terms over a finite horizon from $k = 1, ..., N$ and compute norms:

$$\sqrt{\sum \mu_k|e_{a,k}|^2} \leq \sqrt{\frac{\|\tilde{\mathbf{w}}_0\|^2_{\mathbf{F}}}{2}} + \sqrt{\sum \mu_k(1+\gamma)|v_k|^2}. \tag{17}$$

The expression makes sense as long as $\delta_{A,k} < 0$. However we can extend the result even for $\delta_{A,k} < 1$. To show this property we start with summing up (15) under the condition that $0 < \delta_{A,k} < 1$, remembering that $\tilde{e}_{a,k} = e_{a,k} + v_k$ and obtain

$$\begin{aligned}\sqrt{\sum \mu_k|e_{a,k}|^2} &\leq \sqrt{\frac{\|\tilde{\mathbf{w}}_0\|^2_{\mathbf{F}}}{2}} + \sqrt{\sum \mu_k(1+\gamma)|v_k|^2} \\ &+ \sqrt{\sum \mu_k\delta_{A,k}|e_{a,k}|^2} + \sqrt{\sum \mu_k\delta_{A,k}|v_k|^2}\end{aligned} \tag{18}$$

$$\begin{aligned}\min(1-\delta_{A,k})\sqrt{\sum \mu_k|e_{a,k}|^2} &\leq \sqrt{\frac{\|\tilde{\mathbf{w}}_0\|^2_{\mathbf{F}}}{2}} \\ &+ \max(1-\delta_{A,k})\sqrt{\sum \mu_k(1+\gamma)|v_k|^2},\end{aligned} \tag{19}$$

for which both terms $\min(1-\delta_{A,k})$ and $\max(1-\delta_{A,k})$ remain positive and bounded. We thus can conclude on global robustness:

Lemma 2.2: The adaptive gradient type algorithm with Update (3) exhibits a global robustness from initial uncertainties $\|\tilde{\mathbf{w}}_0\|_{\mathbf{F}}$ and the additive noise energy sequence $\{\sqrt{\mu_k(1+\gamma)}v_k\}_{k=1,2,...,N}$ to its a-priori error sequence $\{\sqrt{\mu_k}e_{a,k}\}_{k=1,2,...,N}$ if the normalized step-size $0 < \alpha = \mu_k/\bar{\mu}_{A,k} < \Re\{\lambda_{A,k}\} - \frac{1}{2\gamma}|1-\lambda_{A,k}|^2$ for some $\gamma > 0$, and $\mathbf{F}^H\mathbf{F} > 0$.

While such statement ensures the LMS algorithm with non-symmetric matrix step-size $\mathbf{G}$ to be l_2-stable, it actually is based on the condition that $\Re\{\lambda_{A,k}\} > 0$. This brings us back to the choice of $\lambda_{A,k}$ which we will have to analyze further. Recall that we defined $\bar{e}_{A,k} = \lambda^*_{A,k}e_{a,k}$ that is we relate $e_{a,k} = \mathbf{x}_k^H\tilde{\mathbf{w}}_{k-1}$ and $\bar{e}_{A,k} = \mathbf{x}_k^H\mathbf{G}^H\mathbf{F}^H\mathbf{F}\tilde{\mathbf{w}}_{k-1}$. As these inner vector products, defining $e_{a,k}$ as well as $\bar{e}_{A,k}$, can take on every arbitrary value, independent of each other, there is no relation in form of a bound from one to the other and as a consequence a strict l_2 stability analysis must end here. Note however, if the relations of the previous lemma hold for any signal they also hold for random processes following some statistics. Thus, placing the expectation operation over all energy terms results in correct statements even though somewhat restricted now by the imposed statistics. Note further that even if $e_{a,k}$ and $\bar{e}_{A,k}$ is hard to be related for general signals, from a statistical point of view the two signals are related. This can be seen when we compute their average energy, that is $\mathbb{E}[e^*_{a,k}\bar{e}_{A,k}] = \mathbb{E}[\tilde{\mathbf{w}}^H_{k-1}\mathbf{x}_k\mathbf{x}_k^H\mathbf{G}^H\mathbf{F}^H\mathbf{F}\tilde{\mathbf{w}}_{k-1}] = \mathbb{E}[\tilde{\mathbf{w}}^H_{k-1}\mathbf{R}_{\text{xx}}\mathbf{G}^H\mathbf{F}^H\mathbf{F}\tilde{\mathbf{w}}_{k-1}]$. Starting with (5), taking expectations on both side and solving for steady-state, that is $\mathbb{E}[\|\tilde{\mathbf{w}}_k\|^2_{\mathbf{F}}] = \mathbb{E}[\|\tilde{\mathbf{w}}_{k-1}\|^2_{\mathbf{F}}] = \mathbb{E}[\|\tilde{\mathbf{w}}_\infty\|^2_{\mathbf{F}}]$ we find

$$\mathbb{E}[|e_{a,\infty}|^2] = \frac{\mathbb{E}\left[\frac{\mu_k}{\bar{\mu}_{A,k}}\right]\sigma_v^2}{\bar{\lambda}_A - \mathbb{E}\left[\frac{\mu_k}{\bar{\mu}_{A,k}}\right]} \tag{20}$$

$$\bar{\lambda}_A = \frac{\Re\left\{\mathbb{E}[\tilde{\mathbf{w}}^H_\infty\mathbf{R}_{\text{xx}}\mathbf{G}^H\mathbf{F}^H\mathbf{F}\tilde{\mathbf{w}}_\infty]\right\}}{\mathbb{E}[\tilde{\mathbf{w}}^H_\infty\mathbf{R}_{\text{xx}}\tilde{\mathbf{w}}_\infty]} \tag{21}$$

where we applied the independence assumption [3][Chapter 9] on the regression vectors $\mathbf{x}_k$ with autocorrelation matrix $\mathbf{R}_{\text{xx}} = \mathbb{E}[\mathbf{x}_k\mathbf{x}_k^H]$ of the driving process x_k and the corresponding parameter error vectors $\tilde{\mathbf{w}}_{k-1}$. The so defined $\bar{\lambda}_A$ can be interpreted as the mean of $\lambda_{A,k}$. The term $\mathbb{E}\left[\frac{\mu_k}{\bar{\mu}_{A,k}}\right]$ takes on a particular simple form ($= \alpha$) when a normalized step-size is applied: $\mu_k = \alpha\bar{\mu}_{A,k}$. The steady-state solution can be a means for defining a step-size bound: $\alpha < \bar{\lambda}_A$. As $\tilde{\mathbf{w}}_\infty$ is typically unknown, it would be difficult to evaluate $\bar{\lambda}_A$. A conservative bound is easy to derive by the Rayleigh factor:

$$\bar{\lambda}_{A,\min} \leq \frac{\mathbf{w}^H\mathbf{R}_{\text{xx}}^{\frac{1}{2}}\left[\mathbf{G}^H\mathbf{F}^H\mathbf{F} + \mathbf{F}^H\mathbf{F}\mathbf{G}\right]\mathbf{R}_{\text{xx}}^{\frac{H}{2}}\mathbf{w}}{2\mathbf{w}^H\mathbf{w}} \leq \bar{\lambda}_{A,\max}.$$

Let us summarize the previous considerations in the following theorem.

Theorem 2.1: The adaptive filter with Update (3) with non-symmetric step-size matrix $\mathbf{G}$, some square matrix $\mathbf{F}$ that satisfies the condition $\mathbf{F}^H\mathbf{F} > 0$, and normalized step-size $\alpha = \frac{\mu_k}{\bar{\mu}_{A,k}}$ guarantees convergence in the mean square sense of its parameter error vector $\tilde{\mathbf{w}}_k$ if the step-size

$$0 < \alpha < \bar{\lambda}_{A,\min} \leq \bar{\lambda}_A \tag{22}$$

under the independence assumption of the regression vectors $\mathbf{x}_k$ with $\mathbf{R}_{\text{xx}} = \mathbb{E}[\mathbf{x}_k\mathbf{x}_k^H]$.

If the minimum Rayleigh factor $\bar{\lambda}_{A,\min}$ is negative we cannot conclude convergence. If $\mu > \bar{\lambda}_{A,\max}$ we expect divergence.

Example A: Let us use $\mathbf{F} = \mathbf{I}$ and $\mathbf{R}_{\text{xx}} = \mathbf{I}$. In this case we find

$$\bar{\mu}_{A,k} = \frac{1}{\mathbf{x}_k^H\mathbf{G}^H\mathbf{G}\mathbf{x}_k} \tag{23}$$

and convergence in the mean square sense for $0 < \alpha < \frac{1}{2}\min\{\text{eig}(\mathbf{G} + \mathbf{G}^H)\}$.

B. Analysis Method B

We now modify the previous method by the following idea. Let us assume again an additional matrix $\mathbf{F}$ that is multiplied from the left. However, now we will not compute the norm in $\mathbf{F}^H\mathbf{F}$ but the inner vector product including $\mathbf{F}$ only. We repeat the process with $\mathbf{F}^H$ and obtain so the conjugate complex of the first part. We add the terms:

$$\begin{aligned}\|\tilde{\mathbf{w}}_k\|^2_{\mathbf{F}+} &= \|\tilde{\mathbf{w}}_{k-1}\|^2_{\mathbf{F}+} \\ &+ 4\frac{\mu_k^2}{\bar{\mu}_{B,k}}|\tilde{e}_{a,k}|^2 - 2\mu_k\bar{e}^*_{B,k}\tilde{e}_{a,k} - 2\mu_k\bar{e}_{B,k}\tilde{e}^*_{a,k}\end{aligned} \tag{24}$$

with the new abbreviations

$$\|\tilde{\mathbf{w}}_k\|^2_{\mathbf{F}+} = \tilde{\mathbf{w}}_k^H(\mathbf{F} + \mathbf{F}^H)\tilde{\mathbf{w}}_k \tag{25}$$

$$\bar{e}_{B,k} = \mathbf{x}_k^H\mathbf{G}^H[\mathbf{F}^H + \mathbf{F}]\tilde{\mathbf{w}}_{k-1}, \tag{26}$$

$$\bar{\mu}_{B,k} = \frac{1}{\mathbf{x}_k^H\mathbf{G}^H[\mathbf{F}^H + \mathbf{F}]\mathbf{G}\mathbf{x}_k} = \frac{1}{\|\mathbf{G}\mathbf{x}_k\|^2_{\mathbf{F}+}}. \tag{27}$$

From here the derivation follows the same path as before, we thus will present the important highlights so that the reader can follow easily. Note that the norm in which we require convergence of the parameter error vector is in $\sqrt{\|\cdot\|^2_{\mathbf{F}+}}$ which makes Method B distinctively different to the previous one. We follow the previous method

$$\begin{aligned}\|\tilde{\mathbf{w}}_k\|^2_{\mathbf{F}+} + 2\mu_k|e_{a,k}|^2 \leq \|\tilde{\mathbf{w}}_{k-1}\|^2_{\mathbf{F}+} \\ + \quad 2\mu_k(1+\gamma)|v_k|^2 + 2\mu_k\delta_{B,k}|\tilde{e}_{a,k}|^2 \\ \lambda_{B,k} \quad = \quad 2\frac{\mu_k}{\bar{\mu}_{B,k}} + 1 - 2\Re\{\lambda_{B,k}\} + \frac{1}{\gamma}|1-\lambda_{B,k}|^2.\end{aligned} \quad (28)$$

Lemma 2.3: The adaptive gradient type algorithm with Update (3) exhibits the following local robustness properties from its inputs $\{\tilde{\mathbf{w}}_{k-1}, \sqrt{\mu_k(1+\gamma)}v_k\}$ to its outputs $\{\tilde{\mathbf{w}}_k, \sqrt{\mu_k}e_{a,k}\}$:

$$\frac{\|\tilde{\mathbf{w}}_k\|^2_{\mathbf{F}+} + 2\mu_k|e_{a,k}|^2}{\|\tilde{\mathbf{w}}_{k-1}\|^2_{\mathbf{F}+} + 2(1+\gamma)\mu_k|v_k|^2} \leq 1 \quad (29)$$

as long as μ_k can be selected so that $0 < \frac{\mu_k}{\bar{\mu}_{B,k}} = \alpha < \Re\{\lambda_{B,k}\} - \frac{1}{2} - \frac{1}{2\gamma}|1-\lambda_{B,k}|^2$ for $\gamma > 0$, and $\mathbf{F}+\mathbf{F}^H > 0$.

Following the same method as before, we find the global statement:

Lemma 2.4: The adaptive gradient type algorithm with Update (3) exhibits a global robustness from initial uncertainty $\tilde{\mathbf{w}}_0$ and the additive noise sequence $\{\sqrt{2\mu_k(1+\gamma)}v_k\}_{k=1,2,\ldots,N}$ to its a-priori error sequence $\{\sqrt{2\mu_k}e_{a,k}\}_{k=1,2,\ldots,N}$ if the normalized step-size $0 < \alpha = \mu_k/\bar{\mu}_{B,k} < \Re\{\lambda_{B,k}\} - \frac{1}{2\gamma}|1-\lambda_{B,k}|^2$ for $\gamma > 0$ and $\mathbf{F}^H + \mathbf{F} > 0$.

This lemma offers similar properties than Lemma 2.2 of Method A and thus the problem of the in general unknown $\lambda_{B,k}$. We thus also follow the steady-state computation as in the previous A and find

$$\bar{\lambda}_{B,\min} \quad \leq \quad \frac{\mathbf{w}^H\mathbf{R}^{\frac{1}{2}}_{\text{xx}}\left[\mathbf{G}^H[\mathbf{F}^H+\mathbf{F}]+[\mathbf{F}^H+\mathbf{F}]\mathbf{G}\right]\mathbf{R}^{\frac{H}{2}}_{\text{xx}}\mathbf{w}}{2\mathbf{w}^H\mathbf{w}} \leq \bar{\lambda}_{B,\max}.$$

Theorem 2.2: The adaptive filter with Update (3) with non-symmetric step-size matrix $\mathbf{G}$, some square matrix $\mathbf{F}$ that satisfies the condition $\mathbf{F}^H + \mathbf{F} > 0$, and normalized step-size $\alpha = \frac{\mu_k}{\bar{\mu}_{B,k}}$ guarantees convergence in the mean square sense of its parameter error vector $\tilde{\mathbf{w}}_k$ if the step-size

$$0 < \alpha < \bar{\lambda}_{B,\min} \leq \bar{\lambda}_B \quad (30)$$

under the independence assumption of the regression vectors $\mathbf{x}_k$ with $\mathbf{R}_{\text{xx}} = \mathbb{E}[\mathbf{x}_k\mathbf{x}_k^H]$.

If the minimum Rayleigh factor $\bar{\lambda}_{B,\min}$ is negative we cannot conclude convergence. If $\mu > \bar{\lambda}_{B,\max}$ we expect divergence.

Example B: Let us use $\mathbf{F} = \frac{1}{2}\mathbf{I}$ and $\mathbf{R}_{\text{xx}} = \mathbf{I}$. In this case we find

$$\bar{\mu}_{B,k} = \frac{1}{\mathbf{x}_k^H\mathbf{G}^H\mathbf{G}\mathbf{x}_k} \quad (31)$$

and convergence for $0 < \alpha < \frac{1}{2}\min\{\text{eig}(\mathbf{G}+\mathbf{G}^H)\}$. Thus, for this choice methods A and B coincide (compare to Example A).

C. Analysis Method C

We now continue in a similar way as in the previous Method B but assume that $\mathbf{F} = \mathbf{G}^{-1}$ exists. We find the inner vector product:

$$\begin{aligned}\tilde{\mathbf{w}}_k^H\mathbf{G}^{-1}\tilde{\mathbf{w}}_k \;=\; & \tilde{\mathbf{w}}_{k-1}^H\mathbf{G}^{-1}\tilde{\mathbf{w}}_{k-1} + 4\mu_k^2\mathbf{x}_k^H\mathbf{G}\mathbf{x}_k|\tilde{e}_{a,k}|^2 \\ - \;& 2\mu_k[\tilde{\mathbf{w}}_{k-1}^H\mathbf{x}_k\tilde{e}_{a,k}+\mathbf{x}_k^H\mathbf{G}^H\mathbf{G}^{-1}\tilde{\mathbf{w}}_{k-1}\tilde{e}^*_{a,k}],\end{aligned}$$

that we complement by its conjugate complex part just as in the previous Method B. However now some terms compensate as $\mathbf{G}\mathbf{G}^{-1} = \mathbf{I}$.

We now introduce

$$\|\mathbf{x}_k\|^2_{\mathbf{G}+} \;=\; \mathbf{x}_k^H[\mathbf{G}^H+\mathbf{G}]\mathbf{x}_k, \quad (32)$$

$$\bar{\mu}_{C,k} \;=\; \frac{1}{\|\mathbf{x}_k\|^2_{\mathbf{G}+}}, \quad (33)$$

$$\bar{e}_{C,k} \;=\; \mathbf{x}_k^H\mathbf{G}^H\mathbf{G}^{-1}\tilde{\mathbf{w}}_{k-1}, \quad (34)$$

$$\;=\; \lambda^*_{C,k}e_{a,k} \quad (35)$$

$$\delta_{C,k} \;=\; 2\frac{\mu_k}{\bar{\mu}_k} - |1+\lambda_{C,k}|^2 + \frac{1+\gamma}{\gamma}|\lambda_{C,k}|^2.$$

Note that $\delta_{C,k}$ now takes a slightly different form compared to the values in Methods A and B, leading to much tighter bounds.

Lemma 2.5: The adaptive gradient type algorithm with Update (3) exhibits the following local robustness properties from its input values $\left\{\sqrt{\|\tilde{\mathbf{w}}_{k-1}\|^2_{\mathbf{G^{-1}}+}}, \sqrt{2\mu_k(1+\gamma)}v_k\right\}$ to its output values $\left\{\sqrt{\|\tilde{\mathbf{w}}_k\|^2_{\mathbf{G^{-1}}+}}, 2\sqrt{\mu_k}e_{a,k}\right\}$

$$\frac{\|\tilde{\mathbf{w}}_k\|^2_{\mathbf{G^{-1}}+} + 4\mu_k|e_{a,k}|^2}{\|\tilde{\mathbf{w}}_{k-1}\|^2_{\mathbf{G^{-1}}+} + 2(1+\gamma)\mu_k|v_k|^2} \leq 1$$

as long as μ_k can be selected so that $0 < \frac{\mu_k}{\bar{\mu}_{C,k}} = \alpha < \Re\{\lambda_{C,k}\} - \frac{1}{2\lambda_{C,k}}|\lambda_{C,k}|^2$ for some $\gamma > 0$ and as long as the matrix $\mathbf{G}+\mathbf{G}^H$ is positive definite.

Lemma 2.6: The adaptive gradient type algorithm with Update (3) exhibits a global robustness from initial uncertainties $\sqrt{\tilde{\mathbf{w}}_0^H[\mathbf{G}^{-1}+\mathbf{G}^{-H}]\tilde{\mathbf{w}}_0}$ and the additive noise energy sequence $\{\sqrt{2\mu_k(1+\gamma)}v_k\}_{k=1,2,\ldots,N}$ to its a-priori error energy sequence $\{\sqrt{4\mu_k}e_{a,k}\}_{k=1,2,\ldots,N}$ if $0 < \alpha < \Re\{\lambda_{C,k}\} + \frac{1}{2} - \frac{1}{2\gamma}|\lambda_{C,k}|^2$ for some $\gamma > 0$ and $\mathbf{G}^{-1} + \mathbf{G}^{-H} > 0$

Note that this analysis method compared to the previous two methods delivers a stronger argument when compared to Methods A and B. Here the step-size bound could become positive and it might be even possible to guarantee l_2-stability in some scenarios.

Following the stochastic approach as before, the steady state is

$$\mathbb{E}[|e_{a,\infty}|^2] \;=\; \frac{2\mathbb{E}\left[\frac{\mu_k}{\bar{\mu}_{C,k}}\right]\sigma_v^2}{\bar{\lambda}_C + 1 - 2\mathbb{E}\left[\frac{\mu_k}{\bar{\mu}_{C,k}}\right]} \quad (36)$$

$$\bar{\lambda}_C \;=\; \frac{\Re\left\{\mathbb{E}[\tilde{\mathbf{w}}_\infty^H\mathbf{R}_{\text{xx}}\mathbf{G}^H\mathbf{G}^{-1}\tilde{\mathbf{w}}_\infty]\right\}}{\mathbb{E}[\tilde{\mathbf{w}}_\infty^H\mathbf{R}_{\text{xx}}\tilde{\mathbf{w}}_\infty]}$$

We find the mean $\bar{\lambda}_C$ of $\lambda_{C,k}$ to be bounded by

$$\bar{\lambda}_{C,\min} \leq \frac{\mathbf{w}^H\mathbf{R}^{\frac{1}{2}}_{\text{xx}}\left[\mathbf{G}^H\mathbf{G}^{-1}+\mathbf{G}^{-H}\mathbf{G}\right]\mathbf{R}^{\frac{H}{2}}_{\text{xx}}\mathbf{w}}{2\mathbf{w}^H\mathbf{w}} \leq \bar{\lambda}_{C,\max}$$

Theorem 2.3: The adaptive filter with Update (3) with non-symmetric step-size matrix $\mathbf{G}$, satisfying $\mathbf{G}+\mathbf{G}^H > 0$ and normalized step-size $\alpha = \frac{\mu_k}{\bar{\mu}_{C,k}}$ guarantees convergence in the mean square sense of its parameter error vector $\tilde{\mathbf{w}}_k$ if the step-size

$$0 < \alpha < \frac{1}{2} + \frac{1}{2}\bar{\lambda}_{C,\min} \leq \frac{1}{2} + \frac{1}{2}\bar{\lambda}_C \quad (37)$$

under the independence assumption of the regression vectors $\mathbf{x}_k$ with $\mathbf{R}_{\text{xx}} = \mathbb{E}[\mathbf{x}_k\mathbf{x}_k^H]$. Alternatively, the so normalized algorithm also converges if the matrix $\mathbf{G}+\mathbf{G}^H$ is negative definite.

Note that due to the normalization of the step-size by terms in $\mathbf{G}$, replacing $\mathbf{G}$ by $-\mathbf{G}$ causes a positive definite matrix $\mathbf{G}+\mathbf{G}^H$ to become negative definite so that the product $\bar{\mu}_{C,k}\mathbf{G}$ remains positive. Also due to the products $\mathbf{G}^H\mathbf{G}^{-1}$ effects compensate each other. The positive upper bound for the normalized step-size is thus not changed by this. The derivation simply requires in this case to define $\|\tilde{\mathbf{w}}_k\|_{\mathbf{G}^{-1}+} = -\tilde{\mathbf{w}}_k^H[\mathbf{G}^{-1}+\mathbf{G}^{-H}]\tilde{\mathbf{w}}_k$ to be a norm.

Name	$1/\bar{\mu}_k$	$\bar{\lambda}_{\min} = \frac{1}{2}\min\text{eig}\{\cdot\}$	comment
Alg. 1	$\mathbf{x}_k^H \mathbf{G}^H \mathbf{G} \mathbf{x}_k$	$\mathbf{G} + \mathbf{G}^H$	Example A+B
Alg. 2	$\mathbf{x}_k^H [\mathbf{G} + \mathbf{G}^H] \mathbf{x}_k$	$\mathbf{G}^H \mathbf{G}^{-1} + \mathbf{G}^{-H} \mathbf{G}$	Method C
Alg. 3	$\mathbf{x}_k^H \mathbf{x}_k$	$(\mathbf{G}^H \mathbf{G})^{-1} \mathbf{G}^H + \mathbf{G}(\mathbf{G}^H \mathbf{G})^{-1}$	Method A, $\mathbf{F} = (\mathbf{G}^H \mathbf{G})^{-1} \mathbf{G}^H$
Alg. 4	$\mathbf{x}_k^H (\mathbf{G}^H \mathbf{G})^2 \mathbf{x}_k$	$\mathbf{G}\mathbf{G}^H \mathbf{G} + \mathbf{G}^H \mathbf{G} \mathbf{G}^H$	Method A, $\mathbf{F} = \mathbf{G}^H$
Alg. 5	$\mathbf{x}_k^H [\mathbf{G} + \mathbf{G}^H] \mathbf{x}_k$	$\mathbf{G}^H (\mathbf{G}^H \mathbf{G})^{-1} \mathbf{G}^H + \mathbf{G}^H (\mathbf{G}^H \mathbf{G})^{-1} \mathbf{G} + 2\mathbf{I}$	Method B, $\mathbf{F} = (\mathbf{G}^H \mathbf{G})^{-1} \mathbf{G}^H$
Alg. 6	$\mathbf{x}_k^H [\mathbf{G}^H (\mathbf{G}^H \mathbf{G})^{-2} \mathbf{G}] \mathbf{x}_k$	$\mathbf{G}^H (\mathbf{G}^H \mathbf{G})^{-2} + (\mathbf{G}^H \mathbf{G})^{-2} \mathbf{G}$	Method A, $\mathbf{F}^H \mathbf{F} = (\mathbf{G}^H \mathbf{G})^{-2}$

TABLE I

Various algorithmic normalizations based on the proposed methods A,B,C with corresponding condition $\bar{\lambda}_{\min} > 0$ (Alg.2: $\bar{\lambda}_{\min} > -1$).

D. Consequences

Corollary 2.1: Consider the three update equations:

$$\hat{\mathbf{w}}_k = \hat{\mathbf{w}}_{k-1} + 2\alpha\bar{\mu}_k \mathbf{G}\mathbf{x}_k \tilde{e}_{a,k} \quad (38)$$

$$\hat{\mathbf{w}}_k = \hat{\mathbf{w}}_{k-1} + 2\alpha\bar{\mu}_k \mathbf{G}^H \mathbf{x}_k \tilde{e}_{a,k} \quad (39)$$

$$\hat{\mathbf{w}}_k = \hat{\mathbf{w}}_{k-1} + \alpha\bar{\mu}_k [\mathbf{G} + \mathbf{G}^H] \mathbf{x}_k \tilde{e}_{a,k} \quad (40)$$

with $\tilde{e}_{a,k} = \mathbf{x}_k^H \tilde{\mathbf{w}}_{k-1} + v_k$ and $\bar{\mu}_k = \mathbf{x}_k^H [\mathbf{G} + \mathbf{G}^H] \mathbf{x}_k$. All three algorithms converge in the mean square sense as long as $\alpha\bar{\mu}_k[\mathbf{G} + \mathbf{G}^H]$ is positive definite for sufficiently small step-size α. Note that this can even include that $[\mathbf{G} + \mathbf{G}^H]$ is negative definite.

The steady-state of such algorithms can also be computed. Starting from (5) we compute the expectation of the energy terms considering a fixed start value $\tilde{\mathbf{w}}_0$ as well as random excitation $\mathbf{x}_k$ and additive noise v_k. For steady-state we find that $\mathbb{E}[\|\tilde{\mathbf{w}}_k\|_2^2] = \mathbb{E}[\|\tilde{\mathbf{w}}_{k-1}\|_2^2] = \mathbb{E}[\|\tilde{\mathbf{w}}_\infty\|_2^2]$ and obtain for normalized step-sizes $\alpha = \mu_k/\bar{\mu}_k$:

$$S_{\text{rel}} = \frac{\mathbb{E}\|\tilde{\mathbf{w}}_\infty\|_2^2}{\|\mathbf{w}_o\|_2^2} = \frac{\alpha N_o}{\bar{\lambda} - \alpha}. \quad (41)$$

The only difference to other LMS algorithms shows in the value of $\bar{\lambda}$ which is two in an NLMS. However, the actual value of $\bar{\lambda}$ is difficult to compute. For white driving processes x_k its bounds are $\bar{\lambda}_{\min} = \frac{1}{2}\min\{\text{eig}(\mathbf{G} + \mathbf{G}^H)\} \le \bar{\lambda} \le \frac{1}{2}\max\{\text{eig}(\mathbf{G} + \mathbf{G}^H)\} = \bar{\lambda}_{\max}$.

E. Validation

In an MC experiment we run simulations (20 runs for each parameter setup) for $M = 50$ with a noise variance of $N_o = 0.0001$. Excitation signals are white QPSK symbols. The experiment applies

$$\mathbf{G} = \begin{pmatrix} 1 & a & a^2 & \dots & a^{M-1} \\ 0 & 1 & a & \dots & a^{M-2} \\ 0 & 0 & 1 & \dots & a^{M-3} \\ \vdots & & & & \\ 0 & 0 & \dots & 1 & a \\ 0 & 0 & \dots & 0 & 1 \end{pmatrix}, \quad (42)$$

where we vary a from zero to one.[1] Independent of the value a the matrix is always regular. We are interested in correctness and the precision of our derived bounds. We thus use the normalized step-sizes and normalize them w.r.t. their bounds, that is $\mu_k = \alpha\bar{\mu}_k\bar{\lambda}_{\min}$. We thus expect to find converging algorithms for $\alpha < 1$. Table I depicts a list of choices. Figure 1 exhibits the observed bounds for $\max\alpha$ from Alg. 1 to 6 when ranging $0 \le a < 1$. Compare Alg. 2 and Alg. 5, being identical but with different bounds, the bound of Alg. 2 being about twice as large as that of Alg. 5. Alg. 1 and Alg. 3 as well as Alg. 4 and Alg. 6 show almost identical behavior, respectively. Above all, if $\mathbf{G}$ is unknown, only Alg. 3 is of practical interest.

[1]Note that the corresponding Matlab code is available under https://www.nt.tuwien.ac.at/downloads/featured-downloads .

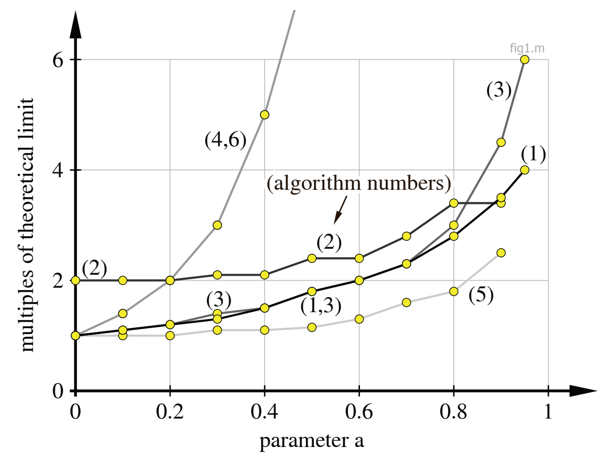

Fig. 1. *Convergence bound α_l over parameter a.*

III. Conclusion

We provided a multitude of potential analysis methods when including non-symmetric step-size matrices. Although we could only prove l_2 stability under very restrictive conditions, we were able to prove convergence in the mean square sense. Simulating the algorithms with worst case sequences (as in [11]) indicate that such algorithms behave indeed robustly, provided the step-sizes are sufficiently small.

References

[1] S. Makino, Y. Kaneda, and N. Koizumi, "Exponentially weighted step-size NLMS adaptive filter based on the statistics of a room impulse response," IEEE Trans. on Speech and Audio Proc., vol. 1, no. 1, pp. 101-108, Jan. 1993.

[2] S. Haykin, *Adaptive Filter Theory*, fourth edition, Prentice Hall, 2002.

[3] A.H. Sayed, "Fundamentals of Adaptive Filtering," Wiley 2003.

[4] A.H. Sayed and M. Rupp, "Error-energy bounds for adaptive gradient algorithms," IEEE Trans. on SP, vol. 44, pp. 1982-1989, Aug. 1996.

[5] M. Rupp and J. Cezanne, "Robustness conditions of the LMS algorithm with time-variant matrix step-size," Signal Processing, vol. 80, no. 9, pp. 1787-1794, Sept. 2000.

[6] M. Rupp, "Convergence Properties of Adaptive Equalizer Algorithms," accepted with minor revisions at IEEE Trans.on Signal Processing, 2011.

[7] D.L. Duttweiler, "Proportionate normalized least mean square adaptation in echo cancellers," IEEE Trans. Speech & Audio Proc., vol. 8, no. 5, pp. 508-518, Sept. 2000.

[8] M. Rupp, A.H. Sayed, "A time-domain feedback analysis of filtered-error adaptive gradient algorithms," IEEE Trans. on SP, vol. 44, no. 6, pp. 1428-1440, June 1996.

[9] A.H. Sayed and M. Rupp, *Chapter: Robustness Issues in Adaptive Filters* in The DSP Handbook, CRC Press, 1998.

[10] A. Feuer and E. Weinstein, "Convergence analysis of LMS filters with uncorrelated Gaussian data," IEEE Trans. ASSP-33, Feb. 1985.

[11] R. Dallinger and M. Rupp, "A Strict Stability Limit for Adaptive Gradient Type Algorithms," in Proc. of the 43rd Asilomar Conf. on Signals, Systems, and Computers, pp. 1370 - 1374, Nov. 2009.