ON THE ROBUSTNESS OF LMS ALGORITHMS WITH TIME-VARIANT DIAGONAL MATRIX STEP-SIZE

Robert Dallinger and Markus Rupp

Vienna University of Technology, Institute of Telecommunications, Austria; {rdalling, mrupp}@nt.tuwien.ac.at

ABSTRACT

The Proportionate Normalized Least Mean Squares (PNLMS) algorithm has been quite successful in combining higher convergence rates with low to moderate complexity that at the same time avoids numerical difficulties in fixed-point implementations. While the algorithm is stable in the mean square and l_2-sense for time-invariant matrices, the treatment of time-variant matrices requires additional approximations. These approximations are discarded in this paper which allows us to analyse the robustness in terms of l_2-stability for actually time-variant matrix step-sizes. This provides important results, as the algorithm in its variants also occurs in other fields of adaptive filtering such as cascaded filter structures. By simulations as well as by theoretical analysis, we demonstrate that in general, even small variations of the matrix step-size are sufficient for the algorithm to loose its robustness. Only in special cases, where specific constraints are imposed additionally, robustness can be guaranteed.

***Index Terms*—** PNLMS, matrix step-size, stability, convergence, robustness

1. INTRODUCTION

Adaptive gradient algorithms like the Least Mean Squares (LMS) algorithm are widely used for parameter estimation due to their readily achievable stability and their low computational complexity. The downside of such algorithms is their comparably slow convergence rate which reduces with increasing number of parameters to estimate [1, 2]. For example, in acoustic echo cancellation, this may become critical since the length of the expected room impulse responses (RIRs) can be considerably high, while a fast echo compensation is desired. Nevertheless, some a priori knowledge is available about typical RIRs which can be leveraged by tailored algorithms to speed up adaptation [3]. One class of such algorithms extends the LMS by introducing a (positive) definite diagonal matrix step-size. Algorithms of this kind are not restricted to acoustic echo cancellation, e.g., they are also useful for polynomial identification of static nonlinearities [4] and they occur in the analysis of cascaded adaptive filters [5]. The corresponding update equation reads

$$\hat{\mathbf{w}}_k = \hat{\mathbf{w}}_{k-1} + \mu_k \mathbf{L}_k \mathbf{u}_k \tilde{e}_{\mathrm{a},k}, \tag{1}$$

where $\mathbf{u}_k = [u_k, \ldots, u_{k-M+1}]$ is the input delay vector containing M consecutive samples of the input sequence u_k, $\hat{\mathbf{w}}_k$ is the adaptively estimated vector of echo weights, $\mu_k > 0$ is a scalar step-size, and $\tilde{e}_{\mathrm{a},k}$ is the error between the actual echo signal and the output generated by the current estimate $\hat{\mathbf{w}}_{k-1}$, distorted by an additive noise sequence v_k. Moreover, (1) contains the positive definite diagonal matrix step-size $\mathbf{L}_k = \mathrm{diag}_{i=1}^{M}\{l_{i,k}\}$ that is of special interest in this paper. It will be shown that the choice of $\mathbf{L}_k$ can cause divergence even under conformance with commonly known stability criteria. Based on the results of [6], which introduces the terms of symmetric and asymmetric gradient type algorithms, we will clarify why time-variant matrix step-sizes destroy the robustness of such algorithms. In the deterministic analysis of adaptive filters, a typical [2, 7–9] meaning of the term robustness refers to the ability of an adaptive algorithm to keep the energy ratio of estimation errors and unknown disturbances bounded from above. Applied to (1), this would mean that

$$\sup \left\{ \frac{\|\tilde{\mathbf{w}}_N\|_2^2 + \sum_{i=0}^{N} |\tilde{e}_{\mathrm{a},i}|^2}{\|\tilde{\mathbf{w}}_{-1}\|_2^2 + \sum_{i=0}^{N} |v_i|^2} \right\} \leq A, \tag{2}$$

with some constant $A < \infty$, N being any arbitrary positive integer, and $k = -1$ denoting the initial state of the algorithm. Additionally, we introduced the parameter error vector $\tilde{\mathbf{w}}_k = \mathbf{w} - \hat{\mathbf{w}}_k$ that represents the error between the unknown vector of echo weights $\mathbf{w}$ and the estimated parameters $\hat{\mathbf{w}}_k$. In this paper, we will use a related notion of robustness that can roughly be considered as unconditional convergence of $\tilde{\mathbf{w}}_k$. More details can be found in Sec. 3.

This paper presents in Sec. 2 two gradient algorithms with matrix step-size in order to demonstrate the occurring robustness issues. This will give motivation for the analysis presented in Sec. 3, which investigates the reason why a matrix step-size introduces the risk of divergence. Moreover, Sec. 3 presents a systematic method to reveal such diverging behaviour and applies this method to a very basic algorithm with matrix step-size. Eventually, Sec. 4 provides a detailed investigation why only a time-variant matrix step-size can actually lead to divergence. The paper closes with conclusions in Sec. 5.

2. MOTIVATION

For a constant matrix step-size, the l_2-stability of such algorithms has been studied in [8] and a stability analysis in the mean square sense was provided in [10]. However, for a time-varying matrix step-size, to the authors knowledge, equivalent results do not exist. One of the few works on this topic even showed that convergence of the parameter error vector requires to impose impractically strong conditions on the matrix step-size [11].

2.1. Exponentially Weighted & Proportionate Normalised LMS

The Exponentially Weighted Step-Size Normalised LMS (EWNLMS) proposed by Makino et al. [12] is an algorithm with a time-invariant diagonal matrix step-size. It is based on the observation that RIRs typically show an exponential decay, which motivated the authors to introduce a constant diagonal matrix step-size $\mathbf{L}_{\mathrm{M}} = \mathrm{diag}_{i=0}^{M-1}\{\delta^i\}$, $0 < \delta < 1$, with exponentially decaying entries. A second member

This work has been funded by the Austrian Science Fund (FWF) under Awards S10611-N13 in the framework of NFN SISE.

out of the class of matrix step-size LMS algorithms is the Proportionate Normalised LMS (PNLMS) introduced by Duttweiler [10]. This algorithm leverages the sparsity of the RIR in order to speed up the learning rate. This is achieved by introducing a time-varying diagonal matrix step-size $\mathbf{L}_{\mathrm{D},k}$ that basically weights each dimension i of the update of $\hat{\mathbf{w}}_k$ by the magnitude of its corresponding weight $\hat{w}_{i,k-1}$ (refer to [10] for further details). Both algorithms also normalise the scalar step-size by the power of the input delay vector, i.e.,

$$\mu_k = \bar{\mu}/\|\mathbf{u}_k\|_2^2, \tag{3}$$

with some step-size factor $\bar{\mu} > 0$, leading to a normalised LMS (NLMS) with matrix step-size. The PNLMS has been continuously investigated and extended since its introduction in [10], and many recent papers are based on it directly or on variants of the LMS that use a diagonal matrix step-size [3, 13–16].

2.2. Hidden divergence for time-variant matrices

In [5,6], the robustness, in the sense of unconditional convergence of the parameter error vector, was considered for so called asymmetric gradient type algorithms. Such algorithms are obtained as soon as the direction of descent does not coincide with the direction of the regression vector. Clearly, (1), and thus, the two above mentioned echo cancellation schemes, belong to this class, since in general, $\mathbf{L}_k\mathbf{u}_k$ will not be in parallel (or anti-parallel) with $\mathbf{u}_k$. It can be observed that such algorithms have an inherent potential to divergence, as soon as the step-size is larger than zero. Such tendency to diverge can be revealed by simulations that excite the adaptive system with a worst-case sequence, which aims to maximise the energy of the parameter error vector in each adaptation step (see Sec. 3.1 for details). Fig. 1 presents the results of a Monte Carlo simulation for the EWNLMS and for the PNLMS. The latter is implemented exactly as in [10] (with $\delta_{[10]} = \rho_{[10]} = 10^{-4}$, where we augmented the symbols by a subscript to indicate that they have a different meaning than the ones used in this paper). For both algorithms, all signals and systems were real-valued. For each Monte Carlo run, the $M = 500$ filter weights of the reference system $\mathbf{w}$ were randomly generated from a zero-mean Gaussian distribution, then it was scaled to have norm $\|\mathbf{w}\| = 1$, and finally each tap $i = 1, \ldots, M$ of $\mathbf{w}$ was weighted by a factor δ^{i-1}, with $\delta = 0.95$, leading to a random but (in the mean) exponentially decaying reference impulse response. The update error in (1) is given by $\tilde{e}_{\mathrm{a},k} = \tilde{\mathbf{w}}_{k-1}^{\mathsf{T}}\mathbf{u}_k + v_k$, with the additive statistically independent zero-mean white Gaussian noise v_k of variance $\sigma_v^2 = 10^{-6}$. Two simulation sets were performed. In the first set, the input samples u_k were randomly drawn from a uniform distribution over the interval $[-1, 1]$. The corresponding behaviour of the relative misadjustment

$$m_{\mathbf{w}}(k) = \mathrm{E}\left\{\frac{\|\tilde{\mathbf{w}}_{k-1}\|_2^2}{\|\tilde{\mathbf{w}}_{-1}\|_2^2}\right\}, \tag{4}$$

with the initial value $\tilde{\mathbf{w}}_{-1}$ and the expectation $\mathrm{E}\{\cdot\}$, shows the expected converging behaviour. The second set of simulations uses for the input samples the same interval, i.e., $u_k \in [-1, 1]$, but it searches in each adaptation step for the above mentioned worst-case vector $\hat{\mathbf{u}}_k$ (cf. (7)) by generating $N_{\mathrm{wc}} = 500$ realisations and choosing that one which maximises $\|\tilde{\mathbf{w}}_k\|_2$. As can be observed in Fig. 1, for the EWNLMS, the learning rate reduces but it still tends to converge. What is striking, is that the PNLMS diverges although it looked stable in the first set of simulations with randomly chosen input. The reason for this difference between EWNLMS and PNLMS motivates the investigations presented in the rest of this paper.

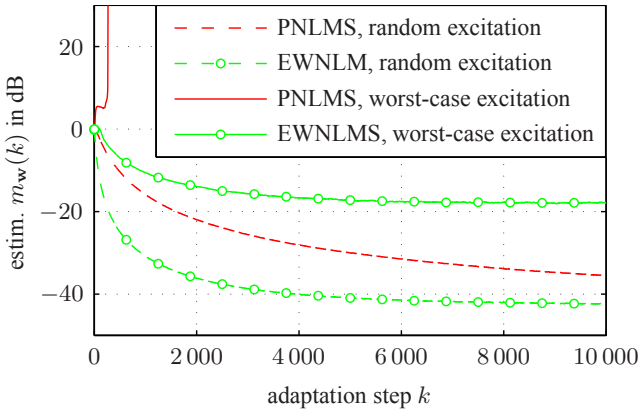


Fig. 1. EWNLMS and PNLMS with random excitation and with worst-case sequences.

3. DIVERGENCE OF THE MATRIX STEP-SIZE LMS

In [6], gradient type algorithms were separated into two classes, symmetric and asymmetric algorithms. For the primer, the direction of descent is parallel to the direction of the regression vector. For such algorithms, stability has been investigated in several senses and corresponding bounds for the step-size are available, e.g., in [1, 2, 6, 7]. The asymmetric algorithms are characterised by a direction of descent that differs from the direction of the regression vector. Clearly, for $\mathbf{L}_k \neq \mathbf{I}$, the algorithm in (1), belongs to this class since in general $\mathbf{u}_k \nparallel \mathbf{L}_k\mathbf{u}_k$. In the noiseless case, (1) represents a recursion of the form

$$\tilde{\mathbf{w}}_k = \mathbf{B}_k\tilde{\mathbf{w}}_{k-1} \qquad \text{with} \qquad \mathbf{B}_k = \mathbf{I} - \mu_k\mathbf{L}_k\mathbf{u}_k\mathbf{u}_k^{\mathsf{T}}. \tag{5}$$

At this point, we are now able to clarify the here used notion of robustness, as also done in [5, 6]. It focuses on the convergence of the sequence formed by the energy $\|\tilde{\mathbf{w}}_k\|_2^2$ of the parameter error vector. If this sequence does not diverge under any circumstances, the algorithm is said to be robust. As can be seen from (5), the matrix sequence of time-variant linear mappings $\mathbf{B}_k$ will decide on this, if a bounded initial parameter error is assumed. Clearly, when considering the adaptation up to $k \to \infty$, to ensure a bounded parameter error in the limit, the product $\prod_{k=0}^{\infty}\mathbf{B}_k$ has to have a finite induced 2-norm, or equivalently, its maximum singular value is required to be finite. In the sequel, we will focus on the local behaviour of $\mathbf{B}_k$ and start with a brief summary of the results from [6], which presents a singular value analysis of the matrix $\mathbf{B}_k$. Accordingly, for $\mu_k \neq 0$, the matrix $\mathbf{B}_k$ can be shown to have $M - 2$ unit singular values and at least one singular value that is larger than one. In the sequel, we assume without loss of generality (WOLOG) that the latter is $\sigma_{1,k}$ and the second non-unit singular value is $\sigma_{2,k}$. Consequently, such algorithms tend to diverge, provided that the mode of $\sigma_{1,k}$ is sufficiently excited. It is also found that for a constant matrix step-size, i.e., $\mathbf{L}_k = \mathbf{L}$, the Cholesky factorisation of $\mathbf{L}$ always allows to map (1) to an equivalent symmetric algorithm. Consequently, for constant matrix step-sizes the above mentioned diverging behaviour cannot be observed.

To analyse the algorithm in (1), we follow the idea from [6] and try to identify the conditions which lead in the noiseless case to

$$D_{\mathbf{w},k}^{(p)} = \|\tilde{\mathbf{w}}_k\|_p - \|\tilde{\mathbf{w}}_{k-1}\|_p \geq 0 \quad \text{or} \quad Q_{\mathbf{w},k}^{(p)} = \frac{\|\tilde{\mathbf{w}}_k\|_p}{\|\tilde{\mathbf{w}}_{k-1}\|_p} \geq 1, \tag{6}$$

where we can use any p-norm. This is possible due to the equivalence of these vector norms [17, p.272]. For the 2-norm and for given μ_k and $\tilde{\mathbf{w}}_{k-1}$, this obviously leads to the question, which $\mathbf{u}_k$ results in a $\mathbf{B}_k$ such that its right singular vector $\mathbf{v}_{1,k}$, corresponding to the largest singular value $\sigma_{1,k}$, is in parallel to $\tilde{\mathbf{w}}_{k-1}$. This would lead to the one specific vector $\hat{\mathbf{u}}_k$ which actually maximises $D^{(2)}_{\mathbf{w},k}$ and $Q^{(2)}_{\mathbf{w},k}$, i.e.,

$$\hat{\mathbf{u}}_k = \arg\max_{\mathbf{u}_k} \frac{\left\|[\mathbf{I} - \mu_k \mathbf{L}_k \mathbf{u}_k \mathbf{u}_k^{\mathsf{T}}]\tilde{\mathbf{w}}_{k-1}\right\|_2}{\|\tilde{\mathbf{w}}_{k-1}\|_2}. \quad (7)$$

A not as strict alternative would be to require the right singular vector $\mathbf{v}_{2,k}$ that belongs to the second non-unit singular value $\sigma_{2,k} \leq 1 \leq \sigma_{1,k}$ to be orthogonal to $\tilde{\mathbf{w}}_{k-1}$. This would still be sufficient to satisfy (6), since we know from the analysis in [6] that all other singular values are greater or equal to one. Note that $\sigma_{2,k}$ can also become larger than one, however then, the algorithm will inevitably never converge. Ensuring $\sigma_{2,k} \leq 1$ is always possible by choosing a sufficiently small step-size.

3.1. Construction of worst-case sequences

We are thus interested in constructing such worst-case sequences systematically. We first notice that the (left and right) singular vectors that correspond to $\sigma_{1,k}$ and $\sigma_{2,k}$ are contained in the hyperplane $\mathcal{S}$ that is spanned by $\mathbf{u}_k$ and $\mathbf{L}_k\mathbf{u}_k$. Accordingly, the singular vectors belonging to the other $M-2$ unit singular values, are orthogonal to $\mathcal{S}$. Consequently, the worst-case excitation vector $\hat{\mathbf{u}}_k$ and the a priori parameter error vector $\tilde{\mathbf{w}}_{k-1}$ are related by

$$\tilde{\mathbf{w}}_{k-1} = a_k[\alpha_k\mathbf{I} + \mathbf{L}_k]\hat{\mathbf{u}}_k \;\Leftrightarrow\; \hat{\mathbf{u}}_k = \frac{1}{a_k}[\alpha_k\mathbf{I} + \mathbf{L}_k]^{-1}\tilde{\mathbf{w}}_{k-1}, \quad (8)$$

with some adequately chosen factors $a_k, \alpha_k \in \mathbb{R}$. Inserting (8) and the update error $\tilde{e}_{\mathrm{a},k}$ for $v_k \equiv 0$, leads to the following update equation for worst-case excitation by $\hat{\mathbf{u}}_k$

$$\tilde{\mathbf{w}}_k = \tilde{\mathbf{w}}_{k-1} - \frac{\mu_k}{a_k^2}\tilde{\mathbf{w}}_{k-1}^{\mathsf{H}}[\alpha_k\mathbf{I} + \mathbf{L}_k]^{-1}\tilde{\mathbf{w}}_{k-1} \cdot \mathbf{L}_k[\alpha_k\mathbf{I} + \mathbf{L}_k]^{-1}\tilde{\mathbf{w}}_{k-1}. \quad (9)$$

Calculating $Q^{(1)}_{\mathbf{w},k}$ from (6) leads to

$$Q^{(1)}_{\mathbf{w},k} = \frac{\sum_{i=1}^{M} q_{i,k}|\tilde{w}_{i,k-1}|}{\sum_{i=1}^{M}|\tilde{w}_{i,k-1}|}, \quad (10)$$

with

$$q_{i,k} = \frac{|\tilde{w}_{i,k}|}{|\tilde{w}_{i,k-1}|} = \left|1 - \frac{\mu_k}{a_k^2}\underbrace{\frac{l_{i,k}}{\alpha_k + l_{i,k}}\sum_{j=1}^{M}\frac{|\tilde{w}_{k-1,j}|^2}{\alpha_k + l_{j,k}}}_{g_{i,k}(\alpha_k)}\right|. \quad (11)$$

Determining from this expression the a_k and α_k that lead to the worst-case is even for very low dimensions M a challenging if not hopeless task. Although not presented here, note that also the 2-norm leads to expressions that have an infeasible complexity. Consequently, we need to simplify (11) to make it more accessible. This can be achieved by the fact that

$$\min_{i=1}^{M}\{q_{i,k}\} \leq Q^{(1)}_{\mathbf{w},k} \leq \max_{i=1}^{M}\{q_{i,k}\}. \quad (12)$$

Therefore, the algorithm diverges at least for one sequence, if for almost all k, a_k and α_k can be found such that $\min_{i=1}^{M}\{q_{i,k}\} > 1$. Similarly, if $\max_{i=1}^{M}\{q_{i,k}\} < 1$ is fulfilled for almost all k and all pairs of a_k and α_k, convergence is ensured.

3.2. The Bipartite matrix LMS

We will now briefly present one of the simplest LMS algorithms with diagonal time-variant matrix step-size to illustrate the use of the above introduced method. We consider the algorithm in (1) with

$$\mathbf{L}_k = \begin{bmatrix}\mathbf{I}_{M/2} & \mathbf{0}\\ \mathbf{0} & \eta_k\mathbf{I}_{M/2}\end{bmatrix} = \begin{bmatrix}1 & 0\\ 0 & \eta_k\end{bmatrix} \otimes \mathbf{I}_{M/2}, \quad (13)$$

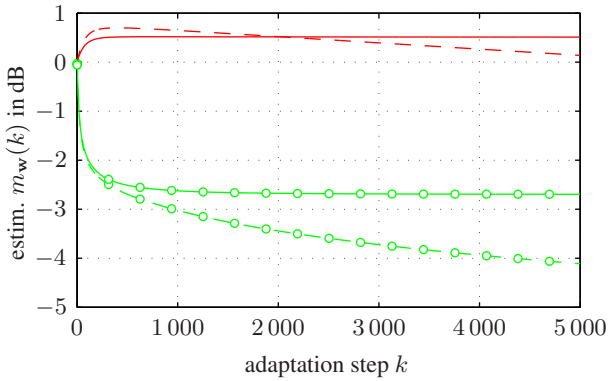
where we assumed that M is even and where η_k is some positive scalar. The symbol $\otimes$ denotes the Kronecker product. Clearly, the $l_{i,k}$ in (11) can either be one or η_k and also the sum in (11) reduces to an addition of two fractions. Such algorithms are typically encountered when two adaptive filters are concatenated. Also adaptive Wiener and Hammerstein models can be mapped to such an algorithm [5, 18]. Even in the here considered simple case, an explicit solution for a_k and α_k is hard to find. However, the above presented insight can be used to reduce the space in which the worst-case vectors $\hat{\mathbf{u}}_k$ have to be searched for. Introducing the vectors $\mathbf{u}_{1,k}$ and $\mathbf{u}_{2,k}$, both of length $M/2$, for the upper and the lower half of $\mathbf{u}_k$, respectively, we applied a small normalized step-size of $\mu_k = 0.1/[\|\mathbf{u}_{1,k}\|_2^2 + \eta_k\|\mathbf{u}_{2,k}\|_2^2]$ for which stability in the mean square sense, as well as, in the l_2-sense is guaranteed. We performed two sets of Monte Carlo simulations, a first one with constant $\eta_k = \eta = 2$, and a second one, with η_k being time-variant and randomly selected from a uniform distribution between zero and two. Figures 2(a) and 2(b) depict the corresponding simulation results, respectively. We considered two cases of excitation sequences, one generates the samples u_k from a uniform distribution on the interval $[-\frac{\sqrt{6}}{2}, \frac{\sqrt{6}}{2}]$. The other one generates a sequence of $u_k \in \{-1, 1\}$ with identical probability. The latter is termed 'bipolar' in Fig. 2. Additionally, we considered the linear combiner (LC) case, i.e., no shift dependency among the elements of $\mathbf{u}_k$, and the finite impulse response (FIR) filter case.

In Fig. 2(a), depicting the fixed matrix step-size case, using LCs, the relative misadjustment shows some potential to grow initially, but then, it levels out at some steady-state or even tends to converge. If the LCs are replaced by FIR filters, independently of the excitation sequence, the relative misadjustment monotonously decreases. In contrast, for the time-variant matrix step-size, in Fig. 2(b), all but the FIR filter case with bipolar excitation lead to diverging behaviour. The reason for the convergence of the FIR filter case with bipolar input is that the available space for the construction of $\hat{\mathbf{u}}_k$ is dramatically small. Actually, in each adaptation step, there are only two possible values to choose from.

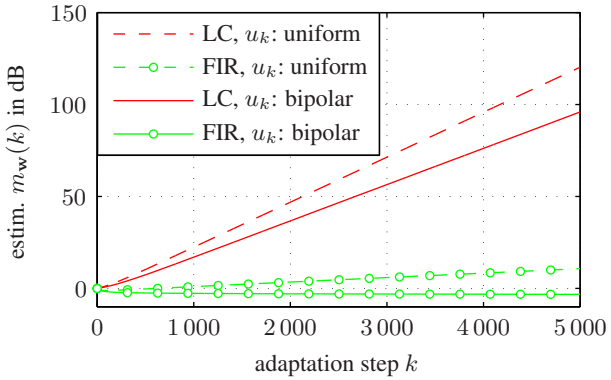
Note that for the bipolar excitation sequence, it was possible to generate the worst-case sequences by a full search, since even in the LC case, the search depth was only 1 024 for the chosen dimension $M = 10$. For uniform excitation, a worst-case search over 1 024 random realisations was performed. Additionally, for the simulations using LCs, the search space was restricted to $\mathcal{S}$. All results were obtained by averaging over 500 Monte Carlo runs. Although not presented here, further simulations showed that unstable behaviour can also be observed if the entries $l_{i,k}$ are changing only slightly.

4. ANALYSIS OF THE DIFFERENCE BETWEEN CONSTANT AND TIME-VARIANT MATRIX STEP-SIZES

What still remains open, is the discussion why a time-variant matrix step-size makes such a big difference. To do so, we consider a simple example with constant $\mathbf{L}_k = \mathbf{L} = \mathrm{diag}\{1, 2, 3\}$. Then clearly, the three poles p_i of the functions $g_{i,k}(\alpha_k)$ in (11) are at the



(a) Fixed matrix step-size (same legend as below).



(b) Randomly varying matrix step-size.

Fig. 2. The worst-case learning behaviour of the Bipartite PNLMS algorithm under various conditions: FIR filter vs. linear combiner (LC), bipolar vs. uniform excitation.

positions $\mathbf{p} = [-1, -2, -3]^\mathsf{T}$, where we collected them in a vector $\mathbf{p}$ of length $M = 3$ with descending order. The left-hand side of Fig. 3 depicts the $g_{i,0}(\alpha_0)$, under the assumption that the initial parameter error vector $\tilde{\mathbf{w}}_{-1} = [1, 1, 1]^\mathsf{T}$. For divergence, we require $\max_{i=1}^{M}\{q_{i,k}\} < 1$. As (11) shows, this is fulfilled if the corresponding $g_{i,k}(\alpha_k) < 0$ or $g_{i,k}(\alpha_k) > 1$. It can be verified that the latter case can always be prevented by normalising the step-size as $\mu_k = \bar{\mu}(\mathbf{u}_k^\mathsf{T}\mathbf{L}_k\mathbf{u}_k)^{-1}$ with an appropriate positive step-size factor $\bar{\mu}$. For $\bar{\mu} = 2$, this normalisation actually coincides with the bound for mean square stability [10]. Consequently, we only need to consider the case $g_{i,k}(\alpha_k) < 0$. Since we require $\hat{\mathbf{u}}_k$ to be finite, α_k must not become equal to one of the poles of $g_{i,k}(\alpha_k)$ (to see this, cf. (8)). From Fig. 3, we find that choosing $\alpha_k > p_1$ or $\alpha_k < p_M$ leads to $g_{i,k}(\alpha_k) > 0$ for all i. Therefore, the α_k which leads to the worst-case fulfils $p_M < \alpha_k < p_1$. Obviously, it is never possible to achieve $g_{i,k}(\alpha_k) < 0$ for all i. A worst-case search would then lead to $p_2 \ll \alpha_k < p_1$. If we assume WOLOG that the diagonal of $\mathbf{L}$ is also sorted, however, in ascending order, this will lead to a growth of the parameter error at positions $i = 2, \ldots, M$ but $\tilde{w}_{1,k}$ will contract. Since the $g_{i,k}(\alpha_k)$ depend on $\tilde{\mathbf{w}}_{k-1}$, the change of parameter errors also changes the $g_{i,k}(\alpha_k)$ with progressing adaptation. This process will go on until $\tilde{w}_{1,k}$ reaches zero, if we exclude the worst-case condition that the algorithmic update freezes. Then, the pole p_1 is cancelled and only the poles p_2 to p_M remain. This situation is depicted on the right-hand side of Fig. 3. However, in

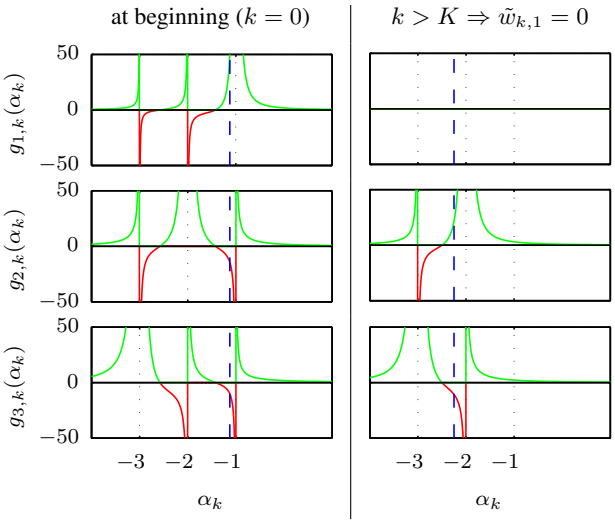


Fig. 3. Behaviour of the functions $g_{i,k}(\alpha_k)$ in (11). The left-hand side depicts the initial state, i.e., $k = 0$ and under the assumption that $\tilde{\mathbf{w}}_{k-1} = [1, 1, 1]^\mathsf{T}$. The right-hand side shows the situation after convergence of the first dimension, i.e., $k > K$. The functions $g_{2,k}(\alpha_k)$ and $g_{3,k}(\alpha_k)$ are assumed to be adequately scaled. On both sides, the blue dashed line indicates a possible value $\hat{\alpha}$ for α_k, as it my be provided by a worst-case search (left-hand side: $\hat{\alpha}_0 = -1.125$, right-hand side: $\hat{\alpha}_K = -2.5$).

contrast to the left-hand side, where $\tilde{\mathbf{w}}_{-1} = [1, 1, 1]^\mathsf{T}$, here, the depicted curves represent adequately scaled versions of $g_{2,k}(\alpha_k)$ and $g_{3,k}(\alpha_k)$. Now, the worst-case search will find $p_3 \ll \alpha_k < p_2$, which leads to an analogous progress of the parameter error vector. This cancellation of poles will go on until one but the last pole is compensated. Then, $g_{i,k}(\alpha_k) = 0$ for $i = 1, \ldots, M-1$ and $g_{M,k}(\alpha_k) > 0$, which finally also lets $\tilde{w}_{M,k}$ decrease. Eventually, this shows that $\|\tilde{\mathbf{w}}_k\|_2$ will remain bounded.

If we now assume that at least one of the poles keeps changing its position, then a permanent cancellation of the corresponding dimension will not be achievable. Therefore, for the worst-case search, it will always be possible to find an α_k which leads to an increase of the $\|\tilde{\mathbf{w}}_k\|_2$. By this, we recognize that the use of a time-variant matrix step-size gives reason for diverging behaviour of the algorithm (1). Of course, as the simulation examples in Sec. 3.2 have shown, if the space from which $\hat{\mathbf{u}}_k$ can be chosen is restricted, divergence might be prevented.

5. CONCLUSION

We presented a convergence analysis of the (N)LMS with diagonal matrix step-size and demonstrated that the use of a time-variant matrix introduces the risk of diverging behaviour. Actually, only strong restrictions on the allowed excitation sequences can ensure convergence and thus stability in the l_2-sense. Based on a 1-norm analysis, the subtle but important difference between constant and varying matrix step-sizes was investigated. Simulations for the very basic bipartite matrix step-size LMS also showed the numerical relevance of the presented approach. Although closed form results can barely be found for the construction of worst-case sequences, the numerical evaluation still suffices to reveal divergent behaviour for broadly used algorithms such as the PNLMS.

6. REFERENCES

[1] Simon Haykin, *Adaptive Filter Theory*, Prentice-Hall, Upper Saddle River, NJ, USA, 4th edition, 2002.

[2] Ali H. Sayed, *Fundamentals of Adaptive Filtering*, John Wiley & Sons, Inc., Hoboken, NJ, USA, 2003.

[3] Steven L. Gay, "An efficient, fast converging adaptive filter for network echo cancellation," in *Record of the Thirty-Second Asilomar Conference on Signals, Systems and Computers*, Pacific Grove, CA, USA, Nov. 1998, pp. 394–398.

[4] Robert Dallinger, Henri Ruotsalainen, Risto Wichman, and Markus Rupp, "Adaptive pre-distortion techniques based on orthogonal polynomials," in *Record of the Forty-Fourth Asilomar Conference on Signals, Systems and Computers*, Pacific Grove, CA, USA, Nov. 2010, pp. 1945–1950.

[5] Robert Dallinger and Markus Rupp, "Stability analysis of an adaptive Wiener structure," in *Proceedings of the IEEE International Conference on Acoustics, Speech, and Signal Processing*, Dallas, TX, USA, Mar. 2010, pp. 3718–3721.

[6] Robert Dallinger and Markus Rupp, "A strict stability limit for adaptive gradient type algorithms," in *Record of the Forty-Third Asilomar Conference on Signals, Systems and Computers*, Pacific Grove, CA, USA, Nov. 2009, pp. 1370–1374.

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

[8] Ali H. Sayed and Markus Rupp, "Error-energy bounds for adaptive gradient algorithms," *IEEE Transactions on Signal Processing*, vol. 44, no. 8, pp. 1982–1989, Aug. 1996.

[9] Ali H. Sayed and Markus Rupp, "Robustness issues in adaptive filtering," in *The Digital Signal Processing Handbook*, chapter 20. CRC Press, Boca Raton, FL, USA, Jan. 1998.

[10] Donald L. Duttweiler, "Proportionate normalized least-mean-squares adaptation in echo cancelers," *IEEE Transactions on Speech and Audio Processing*, vol. 8, no. 5, pp. 508–518, Sept. 2000.

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

[12] Shoji Makino, Yutaka Kaneda, and Nobuo Koizumi, "Exponentially weighted stepsize NLMS adaptive filter based on the statistics of a room impulse response," *IEEE Transactions on Speech and Audio Processing*, vol. 1, no. 1, pp. 101–108, Jan. 1993.

[13] Jacob Benesty and Steven L. Gay, "An improved PNLMS algorithm," in *Proceedings of the IEEE International Conference on Acoustics, Speech, and Signal Processing*, Orlando, FL, USA, May 2002, vol. 2, pp. 1881–1884.

[14] Hongyang Deng and Miloš Doroslovački, "Improving convergence of the PNLMS algorithm for sparse impulse response identification," *IEEE Signal Processing Letters*, vol. 12, no. 3, pp. 181–184, Mar. 2005.

[15] Hongyang Deng and Roman A. Dyba, "Partial update PNLMS algorithm for network echo cancellation," in *Proceedings of the IEEE International Conference on Acoustics, Speech, and Signal Processing*, Taipei, Taiwan, Apr. 2009, pp. 1329–1332.

[16] Francisco das Chagas de Souza, Orlando José Tobias, Rui Seara, and Dennis R. Morgan, "A PNLMS algorithm with individual activation factors," *IEEE Transactions on Signal Processing*, vol. 58, no. 4, pp. 2036–2047, Apr. 2010.

[17] Roger A. Horn and Charles R. Johnson, *Matrix Analysis*, Cambridge University Press, Cambridge, UK, 1999.

[18] Robert Dallinger and Markus Rupp, "On robustness of coupled adaptive filters," in *Proceedings of the IEEE International Conference on Acoustics, Speech, and Signal Processing*, Taipei, Taiwan, Apr. 2009, pp. 3085–3088.