# STABILITY OF ADAPTIVE FILTERS WITH LINEARLY INTERFERING UPDATE ERRORS

*Robert Dallinger and Markus Rupp*

Institute of Telecommunications, Vienna University of Technology, Austria
e-mail: {rdalling, mrupp}@nt.tuwien.ac.at

## ABSTRACT

We provide a time-domain analysis of the stability for two adaptive algorithms of gradient type that interfere with each other via their update errors. Such coupling can occur naturally as well as by desire of the designer. Especially, system identification algorithms that combine two adaptive schemes can often be described by such a structure. We derive precise statements on local contracting/expanding behaviour that in turn allow to deduce bounds ensuring Lyapunov stability. The application of our findings to a specific example shows how these bounds are obtained and how they outperform our previous results that were based on the small gain theorem.

***Index Terms—*** gradient type algorithms, system identification, Lyapunov stability, convergence, contraction mapping

## 1. INTRODUCTION

In this paper, we consider a structure of two gradient type adaptive algorithms that interfere with each other due to a linear memoryless coupling among their individual update errors. Based on our original idea [1], we investigate the contracting behaviour of the varying coefficient matrix for the underlying homogeneous system of difference equations. This leads to the main result of this paper, Theorem 1, which provides precise quantitative statements regarding local increase or decrease of the parameter error vectors (PEVs). These statements in turn, open up access to conditions ensuring Lyapunov stability [2]. In our previous work [3], the obtained $l_2$-stability bounds were rather conservative as they relied on the small gain theorem [4]. By the here presented novel results much tighter stability bounds can be deduced. We illustrate this claim by an example in Sec. 4, for which the obtained stability criteria are indeed found to be tight.

The here analysed structure may unintentionally occur due to some cross-talk or interference, and of course, due to the designers will. However, its relevance goes far beyond that, as especially many adaptive system identification schemes can be mapped to it and are thus inherently covered by this work. The class of such schemes includes for example algorithms that combine different adaptive filters to fuse their beneficial properties, e.g., fast convergence (short filter) and high accuracy (long filter) [5], cascaded structures like the adaptive Wiener and Hammerstein models [6–8], as well as, neural networks that are trained by the backpropagation algorithm [9]. Hence, for the latter, the robustness result in [10] can be extended to more than one single neuron by our theory, as we will publish elsewhere. Finally, least-mean-squares (LMS) algorithms with matrix step-size, e.g., the proportionate normalised LMS [11], are also found to be tightly related [12].

***Notation:*** Throughout this paper, we assume that the initial iteration occurs at $k = 0$. For some vectors $\mathbf{a}$ and $\mathbf{b}$, the column operator is meant in the sense $\mathrm{col}(\mathbf{a}, \mathbf{b}) = [\mathbf{a}^\mathsf{T}, \mathbf{b}^\mathsf{T}]^\mathsf{T}$. The angle brackets $\langle\cdot\rangle$ indicate a sequence of their argument. The superscripts $^*$, $^\mathsf{T}$, and $^\mathsf{H}$, stand for the complex conjugate, the transpose, and the conjugate transpose, respectively. The parameter vectors of the unknown reference systems are denoted by $\mathbf{g}, \mathbf{h}$, the ones of the estimated systems by $\hat{\mathbf{g}}_k, \hat{\mathbf{h}}_k$. Moreover, we assume that modelling uncertainties are covered by additive noise which allows to assume that the lengths of $\mathbf{g}$ and $\mathbf{h}$ coincide with the lengths of $\hat{\mathbf{g}}_k$ and $\hat{\mathbf{h}}_k$, respectively [13]. Thus, we can introduce the PEVs, $\tilde{\mathbf{g}}_k = \mathbf{g} - \hat{\mathbf{g}}_k$ and $\tilde{\mathbf{h}}_k = \mathbf{h} - \hat{\mathbf{h}}_k$. Except for the PEVs, a tilde on top of a scalar indicates that it is the noisy version of some noiseless entity. Finally, $\mathcal{N}_\mathbb{C}(\eta, \sigma^2)$ is the complex-valued normal distribution with mean $\eta$ and variance $\sigma^2$.

***Outline:*** The considered adaptive structure is introduced in Sec. 2. Sec. 3 first establishes the connection to our results in [3], then it presents our novel analysis. Sec. 4 treats a special configuration with strong symmetries and illustrates the impact of our novel approach by numerical experiments. Conclusions are drawn in Sec. 5. The proof of Theorem 1 is sketched in the Appendix.

## 2. THE CONSIDERED COUPLED STRUCTURE

Consider two transversal filters $\mathbf{g}$ ($M_g$ taps) and $\mathbf{h}$ ($M_h$ taps), each driven by the in general different scalar input sequences $\langle x(k)\rangle$ and $\langle u(k)\rangle$, respectively. System uncertainties are modelled at the output of $\mathbf{g}$, respectively $\mathbf{h}$, by the additive noise sequences $\langle v_g(k)\rangle$ and $\langle v_h(k)\rangle$. Thus, at iteration $k$, the overall outputs of the filters $\mathbf{g}$ and $\mathbf{h}$ become (cmp. Fig. 1)

$$y_g(k) = \mathbf{g}^\mathsf{T}\mathbf{x}_k + v_g(k), \quad y_h(k) = \mathbf{h}^\mathsf{T}\mathbf{u}_k + v_h(k),$$

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

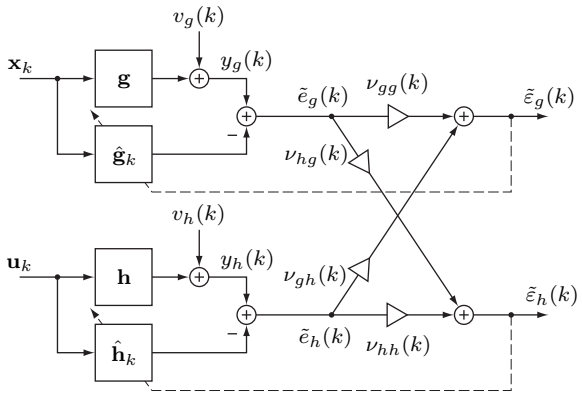



**Fig. 1**. Two adaptive filters with output errors coupled by the multiplicative cross-coupling factors $\nu_{gh}(k)$, $\nu_{hg}(k)$. The auto-coupling factors $\nu_{gg}(k)$ and $\nu_{hh}(k)$ are similar to commonly used step-sizes.

respectively, with the input delay vectors

$$\begin{aligned}\mathbf{x}_k &= [x(k), x(k-1), \ldots, x(k-M_g+1)]^\mathsf{T},\\ \mathbf{u}_k &= [u(k), u(k-1), \ldots, u(k-M_h+1)]^\mathsf{T}.\end{aligned} \tag{1}$$

In a standard system identification problem, both filters would be identified independently by two adaptive finite impulse response filters, $\hat{\mathbf{g}}_k$ and $\hat{\mathbf{h}}_k$. In this case, the LMS [13] algorithm could be employed for adaptation, i.e., in terms of the PEVs,

$$\begin{aligned}\tilde{\mathbf{g}}_{k+1} &= \tilde{\mathbf{g}}_k - \mu_g(k)\mathbf{x}_k^*\tilde{e}_g(k),\\ \tilde{\mathbf{h}}_{k+1} &= \tilde{\mathbf{h}}_k - \mu_h(k)\mathbf{u}_k^*\tilde{e}_h(k),\end{aligned} \tag{2}$$

with the noisy errors,

$$\tilde{e}_g(k) = e_g(k) + v_g(k), \quad \tilde{e}_h(k) = e_h(k) + v_h(k),$$

and their noiseless counterparts

$$\begin{aligned}e_g(k) &= \mathbf{g}^\mathsf{T}\mathbf{x}_k - \hat{\mathbf{g}}_k^\mathsf{T}\mathbf{x}_k = \tilde{\mathbf{g}}_k^\mathsf{T}\mathbf{x}_k,\\ e_h(k) &= \mathbf{h}^\mathsf{T}\mathbf{u}_k - \hat{\mathbf{h}}_k^\mathsf{T}\mathbf{u}_k = \tilde{\mathbf{h}}_k^\mathsf{T}\mathbf{u}_k.\end{aligned}$$

However, in this paper, we assume that this independence is violated due to a linear memoryless coupling among the output errors $\tilde{e}_g(k)$ and $\tilde{e}_h(k)$ as depicted in Fig. 1. Thus, the pair of update errors becomes

$$\begin{aligned}\tilde{\varepsilon}_g(k) &= \nu_{gg}(k)\tilde{e}_g(k) + \nu_{gh}(k)\tilde{e}_h(k),\\ \tilde{\varepsilon}_h(k) &= \nu_{hh}(k)\tilde{e}_h(k) + \nu_{hg}(k)\tilde{e}_g(k),\end{aligned}$$

with the cross-coupling factors $\nu_{gh}(k), \nu_{hg}(k)$, and the auto-coupling factors $\nu_{gg}(k), \nu_{hh}(k)$. All four coupling factors are assumed to be real-valued, whereupon the latter two are typically but not necessarily non-negative. Additionally, for reasons of structural clarity, the step-sizes in (2) are incorporated into the coupling factors, leading to the update equations

$$\tilde{\mathbf{g}}_{k+1} = \tilde{\mathbf{g}}_k - \mathbf{x}_k^*\tilde{\varepsilon}_g(k), \quad \tilde{\mathbf{h}}_{k+1} = \tilde{\mathbf{h}}_k - \mathbf{u}_k^*\tilde{\varepsilon}_h(k). \tag{3}$$

We emphasise that the recursions in (3) still aim to minimise $|\tilde{e}_g(k)|^2$ and $|\tilde{e}_h(k)|^2$, separately. However, in contrast to (2), due to the coupling, additional interference is introduced.

## 3. STABILITY ANALYSIS

In [3], we provided criteria for $l_2$-stability of an equivalent coupled structure. There, the step-sizes were assumed fixed. As it is rather straightforward to modify the results in [3] to varying step-sizes, we skip the derivation and refer the reader to [14] for further details. Applying the so obtained expressions to the structure in Fig. 1, yields the $l_2$-stability condition

$$|\nu_{gh}(k)\nu_{hg}(k)| < |\nu_{gg}(k)\nu_{hh}(k)|\,, \tag{4}$$

if for all $k$, the auto-coupling factors fulfil

$$\nu_{gg}(k) < \|\mathbf{x}_k\|_2^{-2}\,, \qquad \nu_{hh}(k) < \|\mathbf{u}_k\|_2^{-2}\,.$$

This result is rather conservative as it is based on the small gain theorem. In the next section, we go down a different path which enables us to state tighter stability conditions.

### 3.1. Novel stability approach

In a first step, we discard all noise terms in (3) leading to

$$\begin{bmatrix}\tilde{\mathbf{g}}_{k+1}\\ \tilde{\mathbf{h}}_{k+1}\end{bmatrix} = \underbrace{\left(\mathbf{I} - \begin{bmatrix}\nu_{gg}(k)\mathbf{x}_k^*\mathbf{x}_k^\mathsf{T} & \nu_{gh}(k)\mathbf{x}_k^*\mathbf{u}_k^\mathsf{T}\\ \nu_{hg}(k)\mathbf{u}_k^*\mathbf{x}_k^\mathsf{T} & \nu_{hh}(k)\mathbf{u}_k^*\mathbf{u}_k^\mathsf{T}\end{bmatrix}\right)}_{\mathbf{B}_k \in \mathbb{C}^{(M_g+M_h)\times(M_g+M_h)}} \begin{bmatrix}\tilde{\mathbf{g}}_k\\ \tilde{\mathbf{h}}_k\end{bmatrix}. \tag{5}$$

In Sec. 3.2 we will justify this assumption. For the homogeneous recursion in (5), the following theorem holds for which the proof is sketched in the Appendix.

**Theorem 1** *Consider recursion (5), or equivalently, (3) with* $v_g(k) = v_h(k) \equiv 0$. *Furthermore, let* $\mathbf{x}_k \neq \mathbf{0}$, $\mathbf{u}_k \neq \mathbf{0}$, *and*

$$\alpha(k) = 2\nu_{gg}(k) - \nu_{gg}^2(k)\|\mathbf{x}_k\|^2 - \nu_{hg}^2(k)\|\mathbf{u}_k\|^2, \tag{6}$$

$$\beta(k) = \nu_{gh}(k) - \nu_{gg}(k)\nu_{gh}(k)\|\mathbf{x}_k\|^2 + \nu_{hg}(k) - \nu_{hh}(k)\nu_{hg}(k)\|\mathbf{u}_k\|^2, \tag{7}$$

$$\gamma(k) = 2\nu_{hh}(k) - \nu_{hh}^2(k)\|\mathbf{u}_k\|^2 - \nu_{gh}^2(k)\|\mathbf{x}_k\|^2. \tag{8}$$

*With the three conditions*

A. $\alpha(k)\gamma(k) \geq \beta^2(k)$,
B. $\alpha(k)\|\mathbf{x}_k\|^2 + \gamma(k)\|\mathbf{u}_k\|^2 > 0$,
C. $\alpha(k)\|\mathbf{x}_k\|^2 + \gamma(k)\|\mathbf{u}_k\|^2 < 0$,

*for increasing* $k$, *the (Euclidean) distance* $\|\operatorname{col}(\tilde{\mathbf{g}}_k, \tilde{\mathbf{h}}_k)\|$ *between the estimated parameter vector* $\operatorname{col}(\hat{\mathbf{g}}_k, \hat{\mathbf{h}}_k)$ *and the vector* $\operatorname{col}(\mathbf{g}, \mathbf{h})$ *of reference parameters*

1. *can never increase, if and only if (iff), A and B are true,*
2. *can never decrease, iff A and C are fulfilled,*
3. *can do both, increase or decrease, depending on* $\mathbf{x}_k$ *and* $\mathbf{u}_k$, *iff A is violated, and B or C is satisfied,*
4. *is ensured to remain constant, iff B and C are violated.*

Although Theorem 1 only provides local statements, Cases 1, 2, and 4 directly translate to global statements. Only Case 3 does not directly provide insight to global behaviour. Here, divergence and convergence of $\|\operatorname{col}(\tilde{\mathbf{g}}_k, \tilde{\mathbf{h}}_k)\|$ can occur, depending on the excitation sequence $\langle \operatorname{col}(\mathbf{x}_k, \mathbf{u}_k)\rangle$. There always exists at least one worst-case (w.c.) sequence that maximises $\|\operatorname{col}(\tilde{\mathbf{g}}_{k+1}, \tilde{\mathbf{h}}_{k+1})\|$ for each iteration $k$. However, finding such an excitation sequence is not straightforward. Often this can only be approximated by a random search. Only additional restrictions on the excitation vectors, as explained in the scenario of Sec. 4 further ahead, may make a full search feasible that ensures a conclusive answer regarding convergence or divergence. Such restricting conditions in turn, may cause an algorithm that is covered by Case 3 to behave stable [1]. At this point, we emphasise that convergence means boundedness of the PEVs, not necessarily a zero limit.

### 3.2. Presence of noise with finite energy

Now, we will justify the assumption made at the beginning of Sec. 3.1 by considering the presence of noise. To do so, we start with the noisy version of (5),

$$\begin{bmatrix}\tilde{\mathbf{g}}_{k+1}\\ \tilde{\mathbf{h}}_{k+1}\end{bmatrix} = \mathbf{B}_k \begin{bmatrix}\tilde{\mathbf{g}}_{k}\\ \tilde{\mathbf{h}}_{k}\end{bmatrix} - \begin{bmatrix}\overbrace{(\nu_{gg}(k)v_g(k) + \nu_{gh}(k)v_h(k))}^{\tilde{v}_g(k)}\,\mathbf{x}_k\\ \underbrace{(\nu_{hg}(k)v_g(k) + \nu_{hh}(k)v_h(k))}_{\tilde{v}_h(k)}\,\mathbf{u}_k\end{bmatrix}.$$

If we take the Euclidean norm and apply the triangular inequality, we obtain with the maximum singular value (SV) of the matrix $\mathbf{B}_k$ denoted by $\sigma_{\max}(k)$,

$$\left\|\begin{bmatrix}\tilde{\mathbf{g}}_{k+1}\\ \tilde{\mathbf{h}}_{k+1}\end{bmatrix}\right\| \le \sigma_{\max}(k)\left\|\begin{bmatrix}\tilde{\mathbf{g}}_{k}\\ \tilde{\mathbf{h}}_{k}\end{bmatrix}\right\| + \left\|\begin{bmatrix}\tilde{v}_g(k)\mathbf{x}_k\\ \tilde{v}_h(k)\mathbf{u}_k\end{bmatrix}\right\|.$$

If Case 1 (or 4) in Theorem 1 is satisfied, $\sigma_{\max}(k) = 1$, and we find by substituting iteratively from $k = 0$ up to infinity

$$\left\|\begin{bmatrix}\tilde{\mathbf{g}}_{\infty}\\ \tilde{\mathbf{h}}_{\infty}\end{bmatrix}\right\| \le \left\|\begin{bmatrix}\tilde{\mathbf{g}}_{0}\\ \tilde{\mathbf{h}}_{0}\end{bmatrix}\right\| + \sum_{k=0}^{\infty}\left\|\begin{bmatrix}\tilde{v}_g(k)\mathbf{x}_k\\ \tilde{v}_h(k)\mathbf{u}_k\end{bmatrix}\right\|,$$

where the sum on the right hand side exists and is bounded, as long as the noise sequences $\langle\tilde{v}_g(k)\rangle$ and $\langle\tilde{v}_h(k)\rangle$ have finite energy. Then, also the norm of the combined PEV is bounded.

## 4. COUPLING WITH STRONG SYMMETRIES

In order to demonstrate the tighter nature of Theorem 1 in contrast to the bound in (4), we consider the specific scenario

$$\nu_{gg}(k) = \nu_{hh}(k) = \mu, \quad \nu_{gh}(k) = \nu_{hg}(k) = \mu\nu, \tag{9}$$

$$\|\mathbf{x}_k\|_2^2 = \|\mathbf{u}_k\|_2^2 = M. \tag{10}$$

Note that (10) is not as restricting as it may appear, since it is even exactly fulfilled in many communication systems using constant modulus transmission. Applying Theorem 1 with

$$\alpha(k) = \gamma(k) = \mu\left[2 - \mu M(1+\nu^2)\right],$$
$$\beta(k) = 2\mu\nu(1-\mu M).$$

allows to identify the four areas in the plane spanned by $\mu M$ and $\nu$ that correspond to Cases 1–4 in Theorem 1. Fig. 2 visualises the result, which we summarise in the following corollary (the proof is skipped for reasons of brevity).

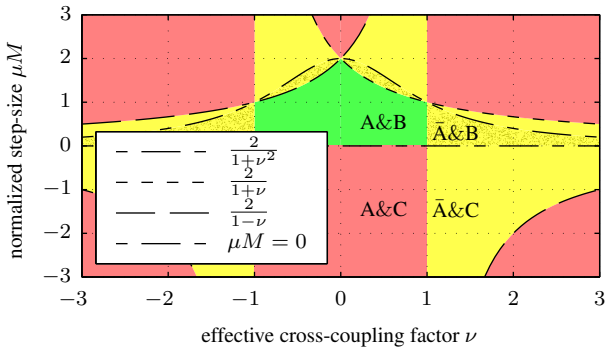



**Fig. 2**. Results of the convergence analysis for the scenario given by (9) and (10). The four areas correspond to the Cases 1–4 in Theorem 1. Contracting behaviour is only ensured for A&B (green). For the (red) areas with A&C, the parameter error will never decrease. Case 4 from Theorem 1 coincides with the two dash-dotted lines.

**Corollary 1** *Consider the homogeneous recursion (5), or equivalently, (3) with $v_g(k) = v_h(k) \equiv 0$, and the coupling scenario specified by (9) and (10). Then, the largest SV of the matrix $\mathbf{B}_k$ is ensured to be less or equal to one (entailing non-diverging behaviour) iff*

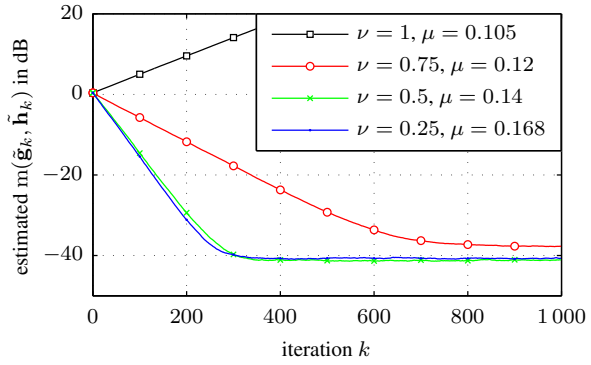
$$-1 \le \nu \le 1, \qquad \textit{and} \qquad 0 \le \mu M \le \frac{2}{1+|\nu|}. \tag{11}$$

Clearly, the stability bound becomes larger with looser coupling. If we would have consulted (4) instead of Theorem 1, we would have found $\mu M \le 1$, for all $|\nu| < 1$; for any other $\nu$ no decision would have been possible. This well demonstrates that Theorem 1 leads to tighter bounds than those presented in [3]. Moreover, for $|\nu| \ge 1$, the latter cannot be applied at all, while Theorem 1 and Corollary 1 still hold.

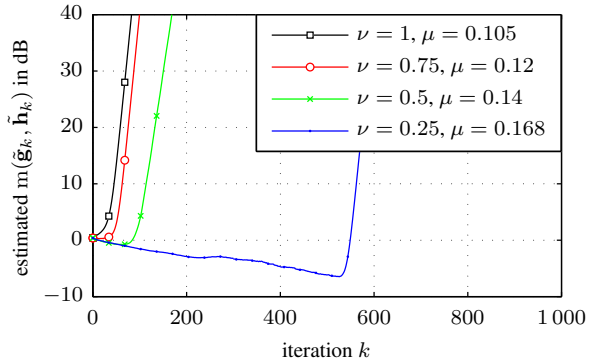
As short excursus, consider the slight modification of (9) to

$$\nu_{gg}(k) = \nu_{hg}(k) = \mu, \quad \nu_{hh}(k) = \nu_{gh}(k) = \mu\nu, \tag{12}$$

which changes the situation drastically. Then, (4) is found to be violated, which means that based on [3], no conclusive statement regarding $l_2$-stability can be obtained. In contrast, (for $\mu \neq 0$ and $|\nu| \neq 1$) Theorem 1 reveals that only Case 3 can occur, which obviously provides much more insight. Before we end our excursus and return to (9), we mention that (12) is obtained if the update errors of both branches in Fig. 1 are identical, i.e., $\tilde{\varepsilon}_g(k) = \tilde{\varepsilon}_h(k)$. This is exactly the situation that occurs if two LMS algorithms are cascaded [3, 15].




(a) Results obtained for simple random QPSK input sequences. All cases except $\nu = 1$ show converging behaviour.




(b) Results for the simulations which include a w.c. search. For all cases divergence is revealed.

**Fig. 3**. Numeric evaluation of the coupling with strong symmetries. The step-size $\mu$ is chosen such that it exceeds the bound in (11) by $5\,\%$. The combined parameter mismatch $\mathrm{m}(\tilde{\mathbf{g}}_k, \tilde{\mathbf{h}}_k)$ (cf. (13)) is estimated by averaging over 200 simulation runs.

In the sequel of this section, we validate Corollary 1 by two Monte-Carlo simulations. Both of them are based on Fig. 1 where all filters have a transversal structure of length $M_g = M_h = M = 10$. At each iteration $k$, the two scalar inputs $x(k)$ and $u(k)$ in (1) are taken from a quadrature phase shift keying (QPSK) alphabet with unit modulus. This ensures that (10) is satisfied and allows us to introduce the normalised step-size $\mu M$. The noise samples $v_g(k)$ and $v_h(k)$ are drawn from $\mathcal{N}_{\mathbb{C}}(0, 10^{-4})$. All results are obtained by averaging over 200 simulation runs, where for each run, the elements of the reference vectors $\mathbf{g}$ and $\mathbf{h}$ are randomly generated from $\mathcal{N}_{\mathbb{C}}(0, \frac{1}{M})$. For each of the considered coupling values $\nu = \{0.25, 0.5, 0.75, 1\}$, the step-size is chosen $5\,\%$ above the bound given in (11).

In the first experiment, $x(k)$ and $u(k)$ are randomly drawn from the QPSK alphabet. Fig. 3(a) shows an estimate of the combined parameter mismatch, i.e.,

$$\mathrm{m}(\tilde{\mathbf{g}}_k, \tilde{\mathbf{h}}_k) = \frac{\mathrm{E}\{\|\tilde{\mathbf{g}}_k\|^2 + \|\tilde{\mathbf{h}}_k\|^2\}}{\mathrm{E}\{\|\tilde{\mathbf{g}}_0\|^2 + \|\tilde{\mathbf{h}}_0\|^2\}}, \tag{13}$$

which converges for all cases except for $\nu = 1$. The second experiment only differs in the generation of the excitation samples, as they are chosen such that $\mathrm{m}(\tilde{\mathbf{g}}_k, \tilde{\mathbf{h}}_k)$ is maximised during each iteration. This is achieved by a complete search over all 16 possible combined excitation vectors. There are only 16 possibilities, as except from $x(k)$ and $u(k)$, the elements in (1) are all fixed from previous iterations, and due to the QPSK alphabet the pair $\{x(k), u(k)\}$ can only assume one out of $2^4 = 16$ combinations. Fig. 3(b) depicts the averaged results obtained with the w.c. input sequences. Clearly, the w.c. search reveals divergence for all considered cases, even for those which lead to convergence in the first experiment.

We repeated the above experiments with the step-size $\mu$ chosen $5\,\%$ below the bound given in (11). Due to space restrictions, the results are not included. However, as expected from Corollary 1, under no circumstance, divergence occurs. Of course, the results obtained with w.c. excitation lead to slower convergence and degraded steady-state behaviour.

## 5. CONCLUSION

The here presented stability analysis is based on the norm of the combined PEV. As this norm fulfils the requirements of a Lyapunov function, Theorem 1 and Corollary 1 establish conditions for Lyapunov stability of the considered coupled structure. Although the energy of the update errors is not further investigated, it ensures robustness of the parameter error mismatch with respect to additive noise of finite energy. Similar to the results in [3], for the here presented stability analysis, persistence of excitation has an impact to the convergence of the PEVs, as it also restricts the space of w.c. excitation sequences. However, these effects are beyond the scope of this paper and will be treated elsewhere.

## APPENDIX: SKETCHED PROOF OF THEOREM 1

For the combined PEV $\mathrm{col}(\tilde{\mathbf{g}}_k, \tilde{\mathbf{h}}_k)$ at iteration $k$, in the noiseless case, we find from (5)

$$\begin{bmatrix} \tilde{\mathbf{g}}_k \\ \tilde{\mathbf{h}}_k \end{bmatrix} = \underbrace{\prod_{m=0}^{k-1} \mathbf{B}_m}_{\mathbf{C}_k} \begin{bmatrix} \tilde{\mathbf{g}}_0 \\ \tilde{\mathbf{h}}_0 \end{bmatrix}. \tag{14}$$

Clearly, with $\sigma_{\mathbf{C},\max}(k)$ denoting the maximum SV of $\mathbf{C}_k$, $\|\mathrm{col}(\tilde{\mathbf{g}}_k, \tilde{\mathbf{h}}_k)\| \leq \|\mathrm{col}(\tilde{\mathbf{g}}_0, \tilde{\mathbf{h}}_0)\|$, if $\sigma_{\mathbf{C},\max}(k) \leq 1$. The SV $\sigma_{\mathbf{C},\max}(k)$ is bounded from above by

$$\sigma_{\mathbf{C},\max}(k) \leq \prod_{m=0}^{k-1} \sigma_{\max}(m). \tag{15}$$

Thus, $\sigma_{\max}(m) \leq 1$ for all $m \in \{0, \ldots, k-1\}$ is sufficient for $\sigma_{\mathbf{C},\max}(k) \leq 1$.

Contrarily, for $\sigma_{\max}(k) \geq (1 + \delta)$ with $\delta > 0$, the contained equality in (15) entails that $\sigma_{\mathbf{C},\max}(k)$ keeps growing

with increasing $k$. Hence, $\|\mathrm{col}(\tilde{\mathbf{g}}_k, \tilde{\mathbf{h}}_k)\|$ increases as well. Whether then, in the w.c., $\|\mathrm{col}(\tilde{\mathbf{g}}_k, \tilde{\mathbf{h}}_k)\|$ remains bounded or tends towards infinity depends on the domain of $\mathrm{col}\,(\mathbf{x}_k, \mathbf{u}_k)$.

The above considerations establish the basis for Theorem 1. Accordingly, we will now consider the SVs $\sigma_i(k)$ of $\mathbf{B}_k$ in more detail. *In the sequel, we omit indices or arguments referring to the iteration index $k$.* Based on (5), the squared SVs $\sigma_i$ are identical to the eigenvalues of the matrix

$$\mathbf{B}^{\mathsf{H}}\mathbf{B} = \mathbf{I} - \begin{bmatrix} \alpha\mathbf{x}^*\mathbf{x}^{\mathsf{T}} & \beta\mathbf{x}^*\mathbf{u}^{\mathsf{T}} \\ \beta\mathbf{u}^*\mathbf{x}^{\mathsf{T}} & \gamma\mathbf{u}^*\mathbf{u}^{\mathsf{T}} \end{bmatrix},$$

where we used (6)–(8). We skip the derivation and just summarise that $\mathbf{B}$ has $M_g + M_h - 2$ (mutually orthogonal right-sided) singular vectors (SVCs) $\mathbf{v}_i$, without loss of generality (WOLOG) $i \in \{3, \ldots, M_g + M_h\}$, that coincide with the unit SV. For the remaining two SVs, WOLOG, we assume $\sigma_1 \geq \sigma_2$ and obtain their squares as

$$\begin{aligned} \sigma_{1,2}^2 &= 1 - \frac{1}{2}(\alpha\|\mathbf{x}\|^2 + \gamma\|\mathbf{u}\|^2) \\ &\quad \pm \frac{1}{2}\sqrt{(\alpha\|\mathbf{x}\|^2 - \gamma\|\mathbf{u}\|^2)^2 + 4\beta^2\|\mathbf{x}\|^2\|\mathbf{u}\|^2} \quad (16) \\ &= 1 - \frac{1}{2}(\alpha\|\mathbf{x}\|^2 + \gamma\|\mathbf{u}\|^2) \\ &\quad \times \left(1 \pm \sqrt{1 + \frac{4(\beta^2 - \alpha\gamma)\|\mathbf{x}\|^2\|\mathbf{u}\|^2}{(\alpha\|\mathbf{x}\|^2 + \gamma\|\mathbf{u}\|^2)^2}}\right). \quad (17) \end{aligned}$$

Additionally, let $\mathcal{S}_{1,2} = \mathrm{span}\{\mathbf{v}_1, \mathbf{v}_2\}$ be the hyperplane spanned by the corresponding right-sided SVCs.

In order to proof the four cases in Theorem 1, first recognise that Cases 1 and 2 are equivalent to $\sigma_1 \leq 1$ and $\sigma_2 \geq 1$, respectively. Case 3 covers the situation $\sigma_1 > 1 > \sigma_2$, and Case 4 is obtained when $\sigma_1 = \sigma_2 = 1$. Moreover, note that Theorem 1 ensures $\|\mathbf{x}\| > 0$ and $\|\mathbf{u}\| > 0$.

***Case 1, i.e.,*** $1 \geq \sigma_1 > \sigma_2$***:*** It can directly be seen from (16) that for Case 1, we have to require

$$\sqrt{(\alpha\|\mathbf{x}\|^2 - \gamma\|\mathbf{u}\|^2)^2 + 4\beta^2\|\mathbf{x}\|^2\|\mathbf{u}\|^2} < \alpha\|\mathbf{x}\|^2 + \gamma\|\mathbf{u}\|^2,$$

which it satisfied iff Conditions A and B in Theorem 1 are fulfilled. For any vector $\mathrm{col}(\tilde{\mathbf{g}}, \tilde{\mathbf{h}}) \not\perp \mathcal{S}_{1,2}$, the norm of the combined PEV decreases, otherwise, it remains unchanged.

***Case 2, i.e.,*** $\sigma_1 > \sigma_2 \geq 1$***:*** Similar to Case 1, we see from (16) that Case 2 requires

$$\sqrt{(\alpha\|\mathbf{x}\|^2 - \gamma\|\mathbf{u}\|^2)^2 + 4\beta^2\|\mathbf{x}\|^2\|\mathbf{u}\|^2} < -(\alpha\|\mathbf{x}\|^2 + \gamma\|\mathbf{u}\|^2),$$

which is true iff Conditions A and C are fulfilled. Then, the norm of the combined PEV increases for any vector $\mathrm{col}(\tilde{\mathbf{g}}, \tilde{\mathbf{h}}) \not\perp \mathcal{S}_{1,2}$, otherwise it remains constant.

***Case 3, i.e.,*** $\sigma_1 > 1 > \sigma_2$***:*** If Condition A is violated, the argument of the square-root in (17) is larger than one. This in turn entails that the bracket expression in the second line of (17) is larger than two for the positive square-root, and less than zero for the negative square-root. Consequently, as long as $\alpha\|\mathbf{x}\|^2 + \gamma\|\mathbf{u}\|^2 \neq 0$, $\sigma_1 > 1$ and $\sigma_2 < 1$. Thus, choosing $\mathrm{col}(\tilde{\mathbf{g}}, \tilde{\mathbf{h}}) \parallel \mathbf{v}_1$ leads to an increasing, $\mathrm{col}(\tilde{\mathbf{g}}, \tilde{\mathbf{h}}) \parallel \mathbf{v}_2$ to a decreasing norm of the combined PEV.

***Case 4, i.e.,*** $\sigma_1 = \sigma_2 = 1$***:*** Since $\alpha\|\mathbf{x}\|^2 + \gamma\|\mathbf{u}\|^2 = 0$, it follows from (17) that $\sigma_1 = \sigma_2 = 1$. Hence, in (14), $\mathbf{C} \equiv \mathbf{I}$ which entails that the combined PEV remains constant.

## REFERENCES


[1] R. Dallinger and M. Rupp, "A strict stability limit for adaptive gradient type algorithms," in *Rec. 43rd Asilomar Conf. on Signals, Systems and Computers*, Pacific Grove, CA, USA, Nov. 2009, pp. 1370–1374.

[2] K. J. Åström and B. Wittenmark, *Adaptive control*, Addison-Wesley Company, Reading, MA, USA, 2nd edition, 1995.

[3] R. Dallinger and M. Rupp, "On robustness of coupled adaptive filters," in *Proc. IEEE Int. Conf. on Acoustics, Speech, and Signal Process.*, Taipei, Taiwan, Apr. 2009, pp. 3085–3088.

[4] M. Vidyasagar, *Nonlinear systems analysis*, SIAM, Philadelphia, PA, USA, 2nd edition, 2002.

[5] E. L. O. Batista, O. J. Tobias, and R. Seara, "New insights in adaptive cascaded FIR structure: Application to fully adaptive interpolated FIR structures," in *Proc. 15th European Signal Process. Conf.*, Poznan, Poland, Sept. 2007.

[6] D. Schreurs, M. O'Droma, A. Goacher, and M. Gadringer, Eds., *RF power amplifier behavioural modeling*, Cambridge University Press, Cambridge, UK, 2009.

[7] E. Aschbacher and M. Rupp, "Modelling and identification of a nonlinear power-amplifier with memory for nonlinear digital adaptive pre-distortion," in *Proc. IEEE Int. Worksh. on Signal Process. Adv. in Wireless Commun.*, Rome, Italy, June 2003, pp. 658–662.

[8] P. Celka, N. J. Bershad, and J.-M. Vesin, "Stochastic gradient identification of polynomial Wiener systems: Analysis and application," *IEEE Trans. Signal Process.*, vol. 49, no. 2, pp. 301–313, Feb. 2001.

[9] S. Haykin, *Neural networks*, Macmillan College Publishing Company, New York, NY, USA, 1994.

[10] B. Hassibi, A. H. Sayed, and T. Kailath, "LMS and backpropagation are minimax filters," in *Theoretical advances in neural computation and learning*, pp. 425–447. Kluwer Academic Publishers, Norwell, MA, USA, Nov. 1994.

[11] D. L. Duttweiler, "Proportionate normalized least-mean-squares adaptation in echo cancelers," *IEEE Trans. Speech Audio Process.*, vol. 8, no. 5, pp. 508–518, Sept. 2000.

[12] R. Dallinger and M. Rupp, "On the robustness of LMS algorithms with time-variant diagonal matrix step-size," in *Proc. IEEE Int. Conf. on Acoustics, Speech, and Signal Process.*, Vancouver, BC, Canada, May 2013, pp. 5691–5695.

[13] S. Haykin, *Adaptive filter theory*, Prentice-Hall, Upper Saddle River, NJ, USA, 4th edition, 2002.

[14] A. H. Sayed, *Fundamentals of adaptive filtering*, John Wiley & Sons, Hoboken, NJ, USA, 2003.

[15] R. Dallinger and M. Rupp, "Stability analysis of an adaptive Wiener structure," in *Proc. IEEE Int. Conf. on Acoustics, Speech, and Signal Process.*, Dallas, TX, USA, Mar. 2010, pp. 3718–3721.