ON ROBUSTNESS OF COUPLED ADAPTIVE FILTERS

Robert Dallinger and Markus Rupp

Vienna University of Technology
Institute of Communications and Radio-Frequency Engineering
Gusshausstrasse 25/389, A-1040 Vienna, Austria

ABSTRACT

We provide a time domain analysis of the robustness and stability performance for coupled adaptive algorithms of gradient type. The considered coupling may occur inherently as well as by desire of the designer. The presented analyses focus on system identification. Examples are presented to investigate convergence and steady-state behaviour by simulations which are compared to theory. In particular, the presented approach allows for a deeper understanding of cascaded adaptive filters in terms of robustness and l_2-stability.

Index Terms— Adaptive filters, system identification, robustness, l_2-stability, error bounds.

1. INTRODUCTION

In this paper, we provide a time-domain robustness and l_2-stability analysis for a certain class of adaptive filters following the lines of [1–5]. The class of adaptive filters we are investigating is characterised by coupling of their update errors. That is, at least two adaptive filters are running simultaneously, each impacting the error term of the other.

Note that we consider a type of coupling different to the one described in [6]. There, the output error of the adaptive filter is fed back to its own input. Here, in contrast, we study two more or less independent adaptive filters which are coupled via their error terms but not via their input sequences.

In [7] and [8] cascaded structures of adaptive filters were proposed. While in [7] the purpose was linear prediction, in [8] system identification in the context of echo cancellation was the main focus. We will show that cascaded structures can be approached by the proposed coupled filter algorithm and thus can be treated by our theory.

The paper is organized as follows: we present the basic problem of coupled filters and their robustness analysis in Sec. 2. How coupling affects the steady state behaviour is treated in Sec. 3. Sec. 4 investigates the coupled combination of a slow and a fast adaptive filter. Sec. 5 shows how the proposed theory can be applied to cascaded adaptive filters to analyse their stability. The paper closes with some concluding remarks.

2. COUPLED GRADIENT TYPE ALGORITHM

Consider two transversal filters, $\mathbf{g}$ of length M_g and $\mathbf{h}$ of length M_h, with in general differing input sequences x_k and u_k. System uncertainties are modelled at the output of $\mathbf{g}$ respectively $\mathbf{h}$ by the additive noise sequences $v_{g,k}$ and $v_{h,k}$. Thus, the overall outputs become (cmp. Fig. 1)

$$y_k = \mathbf{g}^{\mathsf{T}}\mathbf{x}_k + v_{g,k} \tag{1}$$

$$z_k = \mathbf{h}^{\mathsf{T}}\mathbf{u}_k + v_{h,k} \tag{2}$$

with the input vectors $\mathbf{x}_k = \left[x_k, x_{k-1}, \ldots, x_{k-M_g+1}\right]^{\mathsf{T}}$ and $\mathbf{u}_k = \left[u_k, u_{k-1}, \ldots, u_{k-M_h+1}\right]^{\mathsf{T}}$.

In a conventional situation, both filters are being identified by two adaptive Finite Impulse Response (FIR) filters, $\hat{\mathbf{g}}_k$ of length $M_{\hat{g}}$ and $\hat{\mathbf{h}}_k$ of length $M_{\hat{h}}$, separately minimizing the individual output errors $\tilde{e}_k$ and $\tilde{f}_k$. However, as a special case, we assume that the errors are linearly interfering with each other in the following manner:

$$\tilde{e}_k = y_k - \hat{\mathbf{g}}_k^{\mathsf{T}}\mathbf{x}_k + \nu_g(z_k - \hat{\mathbf{h}}_k^{\mathsf{T}}\mathbf{u}_k) \tag{3}$$

$$\tilde{f}_k = z_k - \hat{\mathbf{h}}_k^{\mathsf{T}}\mathbf{u}_k + \nu_h(y_k - \hat{\mathbf{g}}_k^{\mathsf{T}}\mathbf{x}_k), \tag{4}$$

where ν_g and ν_h are real valued coupling factors.

2.1. Noisefree stability analysis

If the adaptation is performed by the Least-Mean-Squares (LMS) algorithm, $\hat{\mathbf{g}}_k$ and $\hat{\mathbf{h}}_k$ are updated following

$$\hat{\mathbf{g}}_{k+1} = \hat{\mathbf{g}}_k + \mu_g\mathbf{x}_k^*\tilde{e}_k \tag{5}$$

$$\hat{\mathbf{h}}_{k+1} = \hat{\mathbf{h}}_k + \mu_h\mathbf{u}_k^*\tilde{f}_k, \tag{6}$$

where for simplicity reasons the positive step-sizes μ_g and μ_h are assumed to be constant. If the filters $\mathbf{g}$ and $\hat{\mathbf{g}}_k$ respectively $\mathbf{h}$ and $\hat{\mathbf{h}}_k$ are same length, the tap error weights $\tilde{\mathbf{g}}_k = \mathbf{g} - \hat{\mathbf{g}}_k$ and $\tilde{\mathbf{h}}_k = \mathbf{h} - \hat{\mathbf{h}}_k$ can be introduced. Consequently, combining equations (5) and (6) leads in the noise free case (i.e., $v_{g,k} = v_{h,k} = 0$) to

$$\begin{bmatrix} \tilde{\mathbf{g}}_{k+1} \\ \tilde{\mathbf{h}}_{k+1} \end{bmatrix} = \begin{bmatrix} \mathbf{I} - \mu_g\mathbf{x}_k^*\mathbf{x}_k^{\mathsf{T}} & -\nu_g\mu_g\mathbf{x}_k^*\mathbf{u}_k^{\mathsf{T}} \\ -\nu_h\mu_h\mathbf{u}_k^*\mathbf{x}_k^{\mathsf{T}} & \mathbf{I} - \mu_h\mathbf{u}_k^*\mathbf{u}_k^{\mathsf{T}} \end{bmatrix} \begin{bmatrix} \tilde{\mathbf{g}}_k \\ \tilde{\mathbf{h}}_k \end{bmatrix}. \tag{7}$$

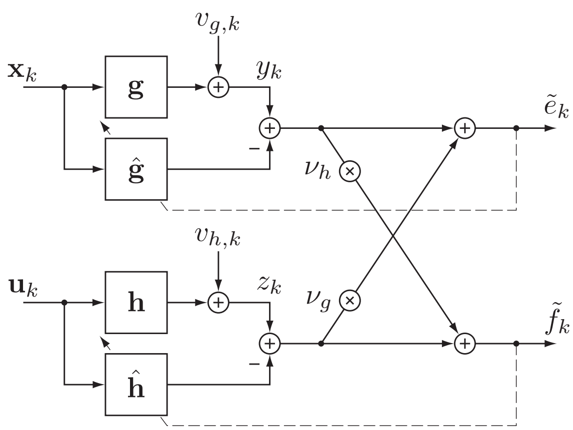

Fig. 1. *Coupled adaptive filters.*

It can be shown that only two singular values of the matrix in (7) can differ from one. Although, the matrix in general changes with each iteration, those two singular values decide on stability or instability. However, a further analysis will be presented elsewhere.

Incorporating also the noise terms in (7) allows for a classical second order moments analysis (see e.g. [5]). Such analyses provided the theoretical graphs which are presented in Fig. 2 and Fig. 3 together with the corresponding simulation results.

2.2. Robustness analysis

Including uncertainties like modelling errors and noise, this section will lead to a sufficient criterion for stability in the l_2-sense, following the robustness analysis from [1–5].

In the sequel, we will consider the adaptive filter $\hat{\mathbf{g}}_k$ in Fig. 1, due to symmetry the results analogously apply to $\hat{\mathbf{h}}_k$. In [5] (p.1042ff), it is shown that for any constant step-size $\mu_g \leq \max_{1\leq l\leq k} \|\mathbf{x}_l\|_2^{-2}$ the energy of the undisturbed error $e_k = \mathbf{g}_k^{\mathsf{T}}\mathbf{x}_k - \hat{\mathbf{g}}_k^{\mathsf{T}}\mathbf{x}_k$ at time instant k is bounded by

$$\sum_{l=1}^{k} |e_l|^2 \leq \frac{1}{\mu_g}\|\mathbf{g}-\hat{\mathbf{g}}_0\|_2^2 + \sum_{l=1}^{k} |\tilde{v}_{g,l}|^2, \tag{8}$$

where $\hat{\mathbf{g}}_0$ denotes the initial condition of the adaptive filter and $\tilde{v}_{g,l} = v_{g,k} + \nu_g f_k$ with $f_k = \mathbf{h}_k^{\mathsf{T}}\mathbf{u}_k - \hat{\mathbf{h}}_k^{\mathsf{T}}\mathbf{u}_k$ comprises the additional noise components. By some straightforward manipulations, (8) leads to

$$\sqrt{\sum_{l=1}^{k} |e_l|^2} \leq \frac{1}{\sqrt{\mu_g}}\|\mathbf{g}-\hat{\mathbf{g}}_0\|_2 + \sqrt{\sum_{l=1}^{k} |v_{g,l}|^2} + \sqrt{\nu_g^2 \sum_{l=1}^{k} |f_l|^2}. \tag{9}$$

For the LMS filter $\hat{\mathbf{h}}_k$, a corresponding expression can be found. Combining latter with (9) provides an upper bound for the square root of the error energy:

$$\sqrt{\sum_{l=1}^{k} |e_l|^2} \leq \frac{1}{1-|\nu_g\nu_h|}\left[\frac{1}{\sqrt{\mu_g}}\|\tilde{\mathbf{g}}_0\|_2 + \sqrt{\sum_{l=1}^{k} v_{g,l}^2} + |\nu_g|\left(\frac{1}{\sqrt{\mu_h}}\|\tilde{\mathbf{h}}_0\|_2 + \sqrt{\sum_{l=1}^{k} v_{h,l}^2}\right)\right]. \tag{10}$$

Considering (10) and the analogous expression for f_k allows for the following interpretations from the robustness analysis:

1. For constant step-sizes $\mu_g \leq \max_k \|\mathbf{x}_k\|_2^{-2}$ and $\mu_h \leq \max_k \|\mathbf{u}_k\|_2^{-2}$, l_2-stability is guaranteed as long as the product of the coupling factors

$$|\nu_g\nu_h| < 1. \tag{11}$$

   Note that in general this bound is rather conservative.

2. Since only the product is relevant, asymmetric coupling with even a relatively strong coupling factor can still result in robust behaviour as long as the second coupling factor compensates for it.

3. If only one coupling factor exists (i.e., the other is zero), both adaptive filters have the same stability bounds as if they are running independently.

3. STEADY-STATE BEHAVIOUR

As an example, consider two adaptive filters which are intended to work independently, but by some undesired design shortcomings the outcomes of the filters turn out to be coupled. We present a simulation example for two random filters of unit norm with length $M_g = M_h = 20$, each excited by an independent real-valued white Gaussian input sequence of unit variance. The output of the filters is disturbed by additive white Gaussian noise of variance $\sigma_{v,g}^2 = \sigma_{v,h}^2 = 10^{-4}$. Typical step-sizes $\mu_g = \mu_h = 0.15/M_g$ are chosen and by simulation the learning curves are determined for equal coupling factors $\nu_g = \nu_h$ in the range of $[-1, 1]$. In general, the coupling appears to be harmful, deteriorating the performance of the adaptive algorithm. Figure 2 depicts the steady-state performance of the relative system misadjustments, given by

$$\frac{\|\mathbf{h}-\hat{\mathbf{h}}_k\|_2^2}{\|\mathbf{h}\|_2^2} \quad \text{respectively} \quad \frac{\|\mathbf{g}-\hat{\mathbf{g}}_k\|_2^2}{\|\mathbf{g}\|_2^2}, \tag{12}$$

and compares them to the results of the classical second order moment analysis. An excellent agreement with theory can be observed (maximal deviation is 0.05dB).

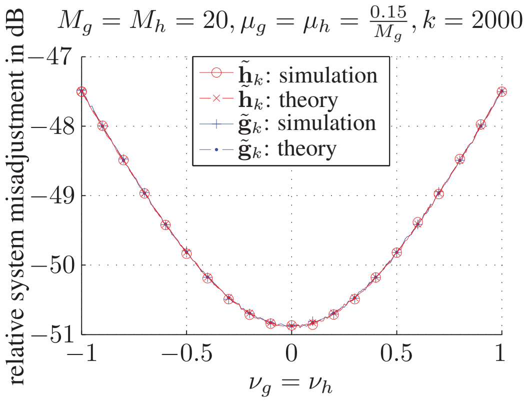

Fig. 2. *Steady-state parameter error energy over $\nu_g = \nu_h$.*

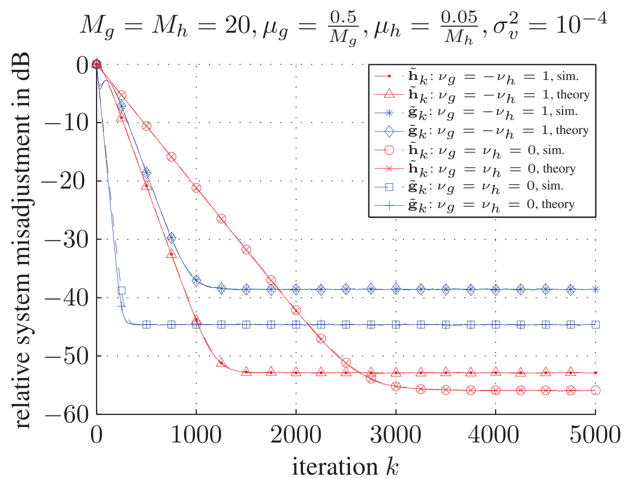

Fig. 3. *Learning curves for zero coupling and $\nu_g = -\nu_h = 1$.*

4. ADAPTIVE SLOW-FAST FIR FILTERS

In [9, 10] an application of adaptive filters was presented where two LMS algorithms run independently in parallel at the same time, one with a small step-size guaranteeing low steady-state values and the other with a large step-size achieving fast convergence. By a clever convex optimization algorithm the best linear combination of both is selected.

In contrast, this section applies the structure of Fig. 1 to identify one unknown system (i.e., $\mathbf{g} = \mathbf{h}$ and $v_{g,k} = v_{h,k} = v_k$) by two adaptive LMS filters with differing step-sizes ($\mu_g > \mu_h$) and coupled errors. Consequently, both filters are excited by the same sequence $\mathbf{x}_k = \mathbf{u}_k$. In the simulations, the results for several random unit norm reference systems were averaged. All filters have same length $M_g = M_h = 20$. The sequences $\mathbf{x}_k$ and v_k are white Gaussian and independent. The step-sizes are chosen to be $\mu_g = 0.5/M_g$ for the fast filter and $\mu_h = 0.05/M_h$ for the slow filter. Fig. 3 presents the corresponding learning curves for (a) zero coupling and (b) $\nu_g = 1, \nu_h = -1$. Again, we observe an excellent agreement between the simulation results and the second order moments analysis. In case (b) the coupling lies just outside the margin of robustness derived in Sec. 2.2. Here, no instabilities occurred which leads to the conclusion that for the considered realisations the bound given by (11) is too restrictive.

5. CASCADED ADAPTIVE FIR FILTERS

It is well known that the longer an adaptive filter is, the slower is its optimum learning rate. Thus, cascaded filters of shorter length could be useful in increasing the convergence speed of adaptive filters. Consider the case of two cascaded FIR filters as depicted in Fig. 4. Note that the structure proposed here is different compared to the cascade structures in [7] and [8]. In [7] the LMS filters were applied for linear prediction and thus not in the context of system identification. In [8] the purpose of the filters is system identification but the update of both is performed by the same error leading to poor quality.

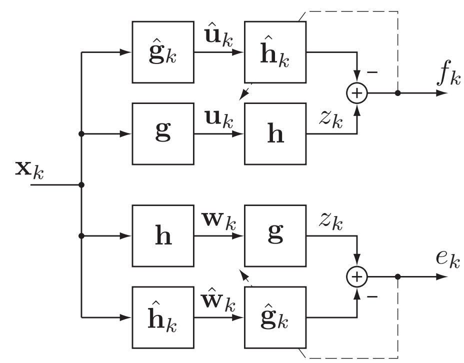

Fig. 4. *Cascaded adaptive filters.*

The update of the filters $\hat{\mathbf{g}}_k$ and $\hat{\mathbf{h}}_k$ is performed according to (again for simplicity reasons fixed step-sizes are assumed)

$$\hat{\mathbf{g}}_{k+1} = \hat{\mathbf{g}}_k + \mu_g \hat{\mathbf{w}}_k^* (z_k - \hat{\mathbf{g}}_k^\mathsf{T} \hat{\mathbf{w}}_k), \tag{13}$$

$$\hat{\mathbf{h}}_{k+1} = \hat{\mathbf{h}}_k + \mu_h \hat{\mathbf{u}}_k^* (z_k - \hat{\mathbf{h}}_k^\mathsf{T} \hat{\mathbf{u}}_k). \tag{14}$$

with

$$\hat{\mathbf{u}}_k = \left[\hat{\mathbf{g}}_k^\mathsf{T} \mathbf{x}_k, \hat{\mathbf{g}}_{k-1}^\mathsf{T} \mathbf{x}_{k-1}, \ldots, \hat{\mathbf{g}}_{k-M_g+1}^\mathsf{T} \mathbf{x}_{k-M_g+1}\right]^\mathsf{T}, \tag{15}$$

$$\hat{\mathbf{w}}_k = \left[\hat{\mathbf{h}}_k^\mathsf{T} \mathbf{x}_k, \hat{\mathbf{h}}_{k-1}^\mathsf{T} \mathbf{x}_{k-1}, \ldots, \hat{\mathbf{h}}_{k-M_h+1}^\mathsf{T} \mathbf{x}_{k-M_h+1}\right]^\mathsf{T}. \tag{16}$$

In the sequel, we assume that the reference system can be separated into two transversal filters $\mathbf{g}$ and $\mathbf{h}$ of same length $M_g = M_h = M$. Then, the update errors are given by

$$e_k = z_k - \hat{\mathbf{g}}_k^\mathsf{T} \hat{\mathbf{w}}_k = (\mathbf{g} - \hat{\mathbf{g}}_k)^\mathsf{T} \hat{\mathbf{w}}_k + \mathbf{g}^\mathsf{T} (\mathbf{w}_k - \hat{\mathbf{w}}_k) \tag{17}$$

$$f_k = z_k - \hat{\mathbf{h}}_k^\mathsf{T} \hat{\mathbf{u}}_k = (\mathbf{h} - \hat{\mathbf{h}}_k)^\mathsf{T} \hat{\mathbf{u}}_k + \mathbf{h}^\mathsf{T} (\mathbf{u}_k - \hat{\mathbf{u}}_k). \tag{18}$$

Assuming small step-sizes, the variation with time of $\mathbf{g}_k$ and $\mathbf{h}_k$ can be approximatively discarded. Hence,

$$\mathbf{g}^\mathsf{T} (\mathbf{w}_k - \hat{\mathbf{w}}_k) \approx \mathbf{g}^\mathsf{T} \mathbf{X}_k \left(\mathbf{h} - \hat{\mathbf{h}}_k\right) = \mathbf{u}_k^\mathsf{T} \left(\mathbf{h} - \hat{\mathbf{h}}_k\right) \tag{19}$$

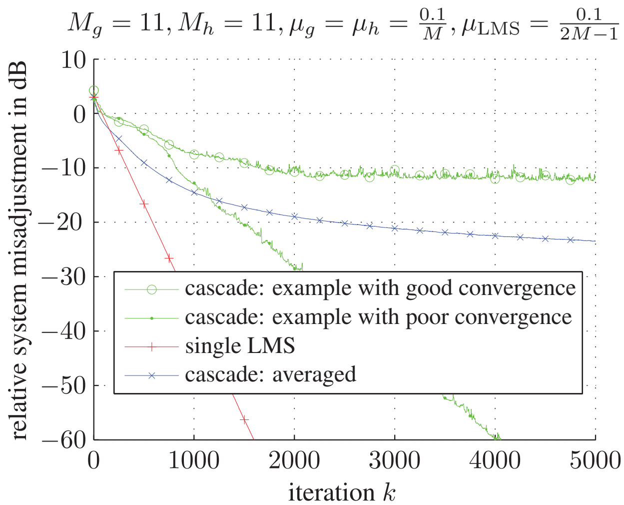

Fig. 5. *Parameter error energy for cascaded LMS filters.*

where $\mathbf{X}_k = [\mathbf{x}_k, \mathbf{x}_{k-1}, \ldots, \mathbf{x}_{k-M_g+1}]^\mathsf{T}$ was used. An analogous expression can be found for $\mathbf{h}^\mathsf{T}(\mathbf{u}_k - \hat{\mathbf{u}}_k)$. Substituting these approximations in (17) respectively (18) leads to

$$e_k \approx \mathbf{g}_k^\mathsf{T}\hat{\mathbf{w}}_k - \hat{\mathbf{g}}_k^\mathsf{T}\hat{\mathbf{w}}_k + \mathbf{h}_k^\mathsf{T}\hat{\mathbf{u}}_k - \hat{\mathbf{h}}_k^\mathsf{T}\hat{\mathbf{u}}_k, \tag{20}$$

$$f_k \approx \mathbf{h}_k^\mathsf{T}\hat{\mathbf{u}}_k - \hat{\mathbf{h}}_k^\mathsf{T}\hat{\mathbf{u}}_k + \mathbf{g}_k^\mathsf{T}\hat{\mathbf{w}}_k - \hat{\mathbf{g}}_k^\mathsf{T}\hat{\mathbf{w}}_k. \tag{21}$$

Comparing (20) with (3) respectively (21) with (4) (note that here the $\tilde{\cdot}$ is suppressed since no noise terms are considered) leads to the conclusion that the cascaded structure in Fig. 4 can be approximatively analysed by the system in Fig. 1 with $\nu_g = \nu_h = 1$, $\mathbf{u}_k = \hat{\mathbf{u}}_k$ and $\mathbf{x}_k = \hat{\mathbf{w}}_k$. Thus, from a robustness point of view such filtering is right on the stability bound given by (11), which may be rather conservative.

In the following experiment we compare the learning behaviour of the cascaded LMS algorithm in Fig. 4 with $M = 11$ to a single LMS with a filter length of $2M-1$. For the cascaded structure, the step-sizes $\mu_g = \mu_h = 0.1/M$ are chosen (step-sizes of $0.15/M$ lead to instable behaviour). The filters are initialised by $\hat{\mathbf{g}}_0 = [1, 0, \ldots, 0]^\mathsf{T}$ and $\hat{\mathbf{h}}_0 = [0, 1, 0, \ldots, 0]^\mathsf{T}$. If they were not different, both adaptive filters would converge to the same solution, since $\mu_g = \mu_h$. An alternative would be to start with identical initial values but different step-sizes. In any case, the initialisation has high impact on the learning behaviour. To obtain comparable results, the single LMS has a step-size of $\mu_{\mathrm{LMS}} = 0.1/(2M-1)$. As in the previous experiments, the excitation signal is a zero-mean white Gaussian random sequence.

Fig. 5 depicts the relative system misadjustment for the cascade structure as well as for the single LMS, both in the noisefree case. The simulations average the results for 500 random unit norm reference systems. It can be observed that in the mean the cascade structure performs very poor compared to the single LMS. The convergence behaviour strongly depends on the initial value of the adaptive filters as well as on the actual configuration of the reference system. This can also be seen in Fig. 5, which additionally shows for the cascade structure the fastest and the slowest converging results out of all performed averaging passes.

We conclude that it does not pay off to apply cascaded techniques in order to gain learning speed or precision. Nevertheless, the cascade mode can be formulated in a stable form.

6. CONCLUSION

In this paper we proposed a coupled adaptive gradient type algorithm. An l_2-stability bound was derived based on a robustness analysis. Simulations in conjunction with second order moments analyses were presented. They revealed that in general coupling does neither improve the steady-state behaviour nor the convergence speed. However, applying the presented robustness analysis to cascaded adaptive filters allows for finding an l_2-stability bound in the case of small step-sizes. The treated cascaded structure is of further interest in the context of digital pre-distortion, when another cascaded structure, a Wiener-Hammerstein system, needs to be matched to an unknown dynamic nonlinear systems. An investigation in this context is left to future work.

7. REFERENCES

[1] A. H. Sayed and M. Rupp, "A time-domain feedback analysis of adaptive gradient algorithms via the small gain theorem," in *Proc. SPIE 1995*, San Diego, USA, Jul. 1995, pp. 458–469.

[2] M. Rupp and A. H. Sayed, "A time-domain feedback analysis of filtered-error adaptive gradient algorithms," *IEEE Trans. On SP.*, vol. 44, no. 6, pp. 1428–1439, Jun. 1996.

[3] A. H. Sayed and M. Rupp, "Error-energy bounds for adaptive gradient algorithms," *IEEE Trans. on SP.*, vol. 44, no. 8, pp. 1982–1989, Aug. 1996.

[4] A. H. Sayed and M. Rupp, "Robustness issues in adaptive filtering," in *The DSP Handbook*. CRC Press, 1998.

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

[6] W. A. Sethares and I. M. Mareels, "Nonlinear dynamics in adaptive echo cancellation," *32nd Midwest Symp. on Circuits & Systems*, pp. 873–876 vol.2, Aug. 1989.

[7] D. Y. Huang, X. Su, and A. Nallanathan, "Characterization of a cascade lms predictor," *ICASSP '05*, vol. 3, pp. iii/173–iii/176 Vol. 3, Mar. 2005.

[8] R. T. Flanagan and J.-J. Werner, "Cascade echo canceler arrangement," *U.S. Patent 6,009,083*, Dec. 28, 1999.

[9] J. Arenas-García, A. R. Figueiras-Vidal, and A. H. Sayed, "Mean-square performance of a convex combination of two adaptive filters," *IEEE Trans. on SP.*, vol. 54, no. 3, pp. 1078–1090, Mar. 2006.

[10] M. T. M. Silva and V. H. Nascimento, "Improving the tracking capability of adaptive filters via convex combination," *IEEE Trans. on SP.*, vol. 56, no. 7, pp. 3137–3149, Jul. 2008.