Pseudo Affine Projection Algorithms Revisited: Robustness and Stability Analysis

Markus Rupp, *Senior Member, IEEE*

***Abstract*—The so-called Affine Projection (AP) algorithm is of large interest in many adaptive filters applications due to its considerable speed-up in convergence compared to its simpler version, the LMS algorithm. While the original AP algorithm is well understood, gradient type variants of less complexity with relaxed step-size conditions called pseudo affine projection offer still unresolved problems. This contribution shows i) local robustness properties of such algorithms, ii) global properties of these, concluding l_2-stability conditions that are independent of the input signal statistics, as well as iii) steady-state values of moderate to high accuracy by relatively simple terms when applied to long filters. Of particular interest is the existence of lower step-size bounds for one of the variants, a bound that has not been observed before.**

***Index Terms*—Adaptive filter analysis and design, adaptive gradient type filters, affine projection, error bounds, l_2-stability, mismatch, pseudo affine projection.**

I. Introduction

LET us consider a system identification setup of a linear time-invariant system (plant) with input sequence u_k, noisy output d_k and impulse response of a FIR filter $\mathbf{w}^T = [w_0, w_1, \dots w_{M-1}]$ of order M. By applying a vector notation $\mathbf{u}_k^T = [u_k, u_{k-1}, \dots u_{k-M+1}]$ with k denoting the instantaneous time instant, we will use a reference model $d_k = \mathbf{u}_k^T \mathbf{w} + v_k$ to describe the input-output relation of the plant. Note that we use real-valued signals throughout the paper as we do not expect any particular new insight for complex-valued signals and it is straightforward to adapt the derivations towards the complex case. Table I lists the most important variables.

Ozeki and Umeda [1] have proposed an affine projection (AP) to speed up the convergence by the following update equation:

$$\mathbf{w}_{k+1} = \mathbf{w}_k + \mathbf{U}_{P+1,k}[\mathbf{U}_{P+1,k}^T \mathbf{U}_{P+1,k}]^{-1}[\mathbf{d}_k - \mathbf{U}_{P+1,k}^T \mathbf{w}_k] \quad (1)$$

with the desired signal in vector form

$$\mathbf{d}_k = \mathbf{U}_{P+1,k}^T \mathbf{w} + \mathbf{v}_k \quad (2)$$

$$\mathbf{U}_{P+1,k} = [\mathbf{u}_k, \mathbf{u}_{k-1}, \mathbf{u}_{k-2}, \dots, \mathbf{u}_{k-P}]. \quad (3)$$

TABLE I
List of Commonly Used Variables

Variable	dimension	meaning
$\mathbf{d}_k$	$\mathbb{R}^{P\times 1}$	output of unknown system
$\mathbf{w}$	$\mathbb{R}^{M\times 1}$	impulse response of unknown system
$\mathbf{w}_k$	$\mathbb{R}^{M\times 1}$	estimated impulse response
$\mathbf{v}_k$	$\mathbb{R}^{P\times 1}$	additive noise
$\mathbf{u}_k$	$\mathbb{R}^{M\times 1}$	driving input vector
$\mathbf{U}_{P+1,k}$	$\mathbb{R}^{M\times (P+1)}$	current and past input vectors
ϕ_k	$\mathbb{R}^{M\times 1}$	decorrelated input process
$\hat{\mathbf{a}}_k$	$\mathbb{R}^{P\times 1}$	AR estimates

Here, $\mathbf{u}_k,\ \mathbf{w}_k \in \mathbf{R}^{M\times 1}$ are the estimates of the time-invariant FIR system $\mathbf{w}$, $\mathbf{d}_k^T = [d_k, d_{k-1}, \dots, d_{k-P}]$ the instantaneous and past output observations, $\mathbf{v}_k^T = [v_k, v_{k-1}, \dots, v_{k-P}]$ the additive noise vector, and $\mathbf{U}_{P+1,k} \in \mathbf{R}^{M\times(P+1)}$ a block matrix of current and past inputs. While speeding up convergence in the system identification of time-invariant systems, the algorithms sacrifices tracking capability as soon as the prediction filter order P increases as well as for large filter order M.

This algorithm that also can be interpreted as an underdetermined LS type algorithm [2] can equivalently be brought into the much simpler gradient type form due to its projection property

$$\mathbf{d}_k - \mathbf{U}_{P+1,k}^T \mathbf{w}_{k+1} = \mathbf{0} \quad (4)$$

revealing its underdetermined LS nature.

In [2] it has been shown that the AP algorithm can be equivalently reformulated into

$$\mathbf{w}_{k+1} = \mathbf{w}_k + \frac{\phi_k}{\|\phi_k\|_2^2}\tilde{e}_{a,k};\ k = 0, 1, 2, \dots \quad (5)$$

$$\tilde{e}_{a,k} = d_k - \mathbf{u}_k{}^T \mathbf{w}_k = v_k + \mathbf{u}_k{}^T(\mathbf{w} - \mathbf{w}_k) \quad (6)$$

$$\mathbf{U}_{P,k-1} = [\mathbf{u}_{k-1}, \mathbf{u}_{k-2}, \dots, \mathbf{u}_{k-P}] \quad (7)$$

$$\hat{\mathbf{a}}_k = [\mathbf{U}_{P,k-1}^T \mathbf{U}_{P,k-1}]^{-1} \mathbf{U}_{P,k-1}^T \mathbf{u}_k \quad (8)$$

$$\phi_k = \mathbf{u}_k - \mathbf{U}_{P,k-1}\hat{\mathbf{a}}_k. \quad (9)$$

Here, the prediction property of this algorithm becomes explicit in (7)–(9), showing that the AP algorithm runs the updates $\mathbf{w}_k$ with the decorrelated regression ϕ_k rather than the original $\mathbf{u}_k$. The coefficients $\hat{\mathbf{a}}_k$ can be interpreted as coefficients of a linear predictor of order $P \le M$, perfectly matching the prediction of an autoregressive (AR) random process u_k of maximum order P. Note that (8) can be interpreted as a Wiener estimator for the linear prediction problem: $[\mathbf{U}_{P,k-1}^T \mathbf{U}_{P,k-1}]^{-1}\mathbf{U}_{P,k-1}^T \approx \mathbf{R}_{\mathrm{uu}}^{-1}\mathbf{r}_{\mathrm{u}}$ (see, e.g., [3]). For sufficiently long filters of order M, the estimate will become rather constant but exhibits large fluctuations if M is too small. This aspect is usually not considered

Manuscript received October 27, 2010; accepted February 02, 2011. Date of publication March 24, 2011; date of current version April 13, 2011. The associate editor coordinating the review of this manuscript and approving it for publication was Dr. Suleyman S. Kozat. This work has been funded by the NFN SISE project S10609 (National Research Network Signal and Information Processing in Science and Engineering).

The author is with the Vienna University of Technology, Institute of Telecommunications, 1040 Vienna, Austria (e-mail: mrupp@nt.tuwien.ac.at).

This paper has supplementary downloadable multimedia material available at http://ieeexplore.ieee.org provided by the authors. This includes all Matlab code, which shows this simulation examples. This material is 5 kB in size.

Digital Object Identifier 10.1109/TSP.2011.2113346

in past analyses of the algorithm and will also not be considered here, limiting the analysis to sufficiently large values of filter order M. Note however, that we keep including estimates $\hat{\mathbf{a}}_k$ in our further derivations even though we treat them often as time-invariant values $\hat{\mathbf{a}}$. Let us refer in the following the assumption of having sufficiently large filter order M as Assumption A1.

Assumption A1): We assume that the filter order M is sufficiently large, so that the estimates of the linear predictor coefficients $\hat{\mathbf{a}}_k$ can be treated as time-invariant values, say $\hat{\mathbf{a}}$.

Such assumption has further consequences. Take, for example, (9), which can be rewritten as

$$\mathbf{u}_k - \mathbf{U}_{P,k-1}\hat{\mathbf{a}}_k = \mathbf{u}_k - \mathbf{U}_{P,k-1}[\mathbf{U}_{P,k-1}^T \mathbf{U}_{P,k-1}]^{-1}\mathbf{U}_{P,k-1}^T \mathbf{u}_k \tag{10}$$

showing that $||\phi_k||_2^2$ is the prediction error energy and will tend to a constant for large M as long as the statistic of the driving process u_k remains constant.

Further note, that due to the filtering process, driving signal u_k is not only decorrelated but also shaped to become more Gaussian. Statistical analysis methods, relying on Gaussian signals are thus quite reasonable to cover typical scenarios, however extreme non-Gaussian driving signals can be envisioned.

Due to its popularity a fast version has also been proposed by Gay [5], [6] for acoustical and electrical echo compensation, thus making the complexity tractable for long filters. The fast version of the algorithm with some stabilizing variants has been applied successfully in adaptive echo suppression of long distance telephone connections. In [7]–[9] the AP algorithm has been investigated among others and some fundamental decorrelating properties were found.

As the introduction of a step-size $\alpha > 0$ offers more flexibility, variants of the AP algorithms were proposed, one being the Pseudo Affine Projection (PAP) algorithm whose updates read

$$\mathbf{w}_{k+1} = \mathbf{w}_k + \alpha \frac{\phi_k}{||\phi_k||_2^2}\tilde{e}_{a,k};\ k = 0, 1, 2, \ldots \tag{11}$$

applying directions ϕ_k from (9). The challenge to analyze the properties of the algorithm was then set in [2] and finally solved 11 years later (at least in parts) in a classical MSE context [10]. With a step-size unequal to one the affine projection property is lost and thus the name Pseudo AP. Smaller step-sizes are usually of interest as sudden noise bursts as they occur in double-talk situations are typically treated by lowering the step-size. We will show in this contribution that this is not recommendable as depending on the correlation of the driving process even lower step-size stability bounds exist. Although the algorithm in this form resembles a projection type gradient algorithm like NLMS, it behaves differently. Some "strange" properties like its stability bounds could not be explained yet and requires further investigation.

In literature [3], [4] a more general form of the AP algorithm (also referred to as AP algorithm) additionally applies a step-size $\alpha > 0$

$$\mathbf{w}_{k+1} = \mathbf{w}_k + \alpha \mathbf{U}_{P+1,k}[\mathbf{U}_{P+1,k}^T \mathbf{U}_{P+1,k}]^{-1}[\mathbf{d}_k - \mathbf{U}_{P+1,k}^T \mathbf{w}_k]. \tag{12}$$

However, note that property (4) is only satisfied for $\alpha = 1$. We will refer to this variant of the algorithm which is different to the PAP version in the following as Generalized Affine Projection (GAP) Algorithm as it preserves some of the affine projection properties. The classic AP algorithm is thus a special case of GAP. Compared to the PAP algorithm the GAP has much higher complexity as it still requires the involvement of the block matrix $\mathbf{U}_{P+1,k}$.

This paper considers both algorithmic variants with step-size $\alpha > 0$ in the context of robustness [4], [11]–[13], which further allows us to find l_2-stability bounds that:

1) are rather accurate for long filters;
2) take some statistical properties in the context of linear prediction of the driving process into account [Assumption A1) and $P \leq M$] for the PAP algorithm while this is not the case for the GAP algorithm; and further
3) allow to derive also the mismatch of the estimation error.

The paper is further organized as follows. In Section II, some basic properties along with local stability bounds are derived for both algorithms, the PAP as well as for the GAP algorithm. In Section III, we finally derive global l_2-stability bounds for these two algorithms. Section IV elaborates on such bounds by deriving steady-state behavior in terms of the filter mismatch, and finally Section V validates the theoretical results by some numerical examples in form of Matlab simulations. Section VI closes the paper by presenting further open issues.

II. Basic Properties: Local Bounds

A. PAP Algorithm

We start with the PAP algorithm. Due to the computation of ϕ_k we find the following property:

$$\mathbf{U}_{P+1,k}^T \frac{\phi_k}{\phi_k^T \phi_k} = \begin{bmatrix} 1 \\ \mathbf{0} \end{bmatrix}. \tag{13}$$

Note that this property always holds, independent of the selection of α. Reformulating the original AP update in (5) into the parameter error vector form $\tilde{\mathbf{w}}_k = \mathbf{w} - \mathbf{w}_k$ and applying $\mathbf{U}_{P+1,k}$ from the left results into

$$\mathbf{U}_{P+1,k}^T \tilde{\mathbf{w}}_{k+1} = \mathbf{U}_{P+1,k}^T \tilde{\mathbf{w}}_k = -\begin{bmatrix} v_k \\ \mathbf{v}_{k-1} \end{bmatrix} \tag{14}$$

where we introduced a noise vector

$$\mathbf{v}_{k-1}^T = [v_{k-1}, v_{k-2}, \ldots, v_{k-P}]. \tag{15}$$

Further, we introduce the undisturbed *a priori* error and a vector version of it:

$$e_{a,k} = \mathbf{u}_k{}^T \tilde{\mathbf{w}}_k \tag{16}$$

$$\mathbf{e}_{a,k-1}^T = [e_{a,k-1}, e_{a,k-2}, \ldots, e_{a,k-P}]. \tag{17}$$

This allows us to rewrite the PAP algorithm from (11) into a compact form:

$$\begin{aligned} \mathbf{U}_{P+1,k}^T \tilde{\mathbf{w}}_{k+1} &= \begin{bmatrix} \mathbf{u}_k{}^T \\ \mathbf{U}_{P,k-1}^T \end{bmatrix} \tilde{\mathbf{w}}_{k+1}, \\ &= \begin{bmatrix} \mathbf{u}_k{}^T \\ \mathbf{U}_{P,k-1}^T \end{bmatrix} \tilde{\mathbf{w}}_k - \alpha \begin{bmatrix} \tilde{e}_{a,k} \\ \mathbf{0} \end{bmatrix} \\ &= (1-\alpha) \begin{bmatrix} e_{a,k} \\ \mathbf{e}_{a,k-1} \end{bmatrix} - \alpha \begin{bmatrix} v_k \\ \mathbf{v}_{k-1} \end{bmatrix}. \end{aligned} \tag{18}$$

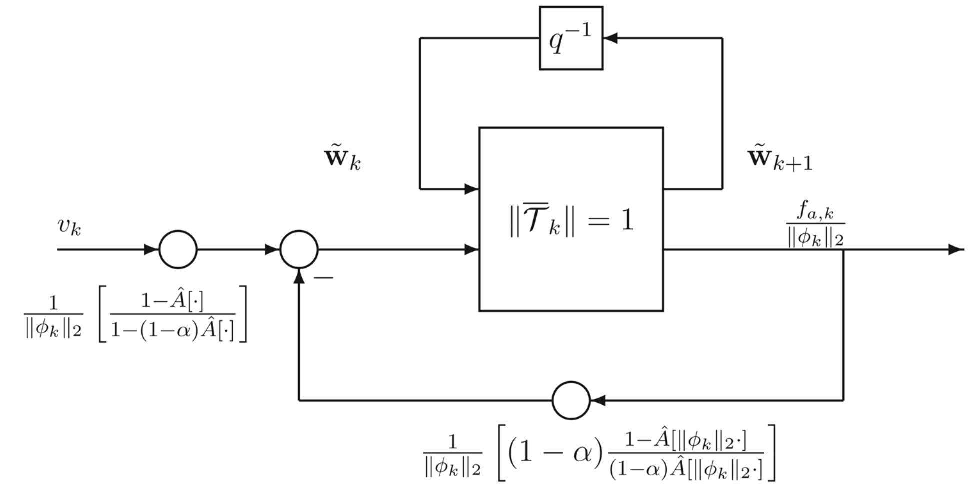

Fig. 1. Feedback structure of the PAP algorithm.

As we are interested in the parameter error vector, we find now

$$\tilde{\mathbf{w}}_{k+1} = \left[\mathbf{I} - \alpha\frac{\phi_k\phi_k^T}{\|\phi_k\|_2^2}\right]\tilde{\mathbf{w}}_k - \alpha\frac{\phi_k}{\|\phi_k\|_2^2}\tilde{v}_k \tag{19}$$

for which we introduced a modified noise term

$$\tilde{v}_k = v_k - \alpha\hat{\mathbf{a}}_k^T\mathbf{v}_{k-1} + (1-\alpha)\hat{\mathbf{a}}_k^T\mathbf{e}_{a,k-1}. \tag{20}$$

Applying the robustness derivation [12], [13] requires to introduce a different *a priori* error, namely

$$f_{a,k} = \phi_k^T\tilde{\mathbf{w}}_k \tag{21}$$

rather than $e_{a,k}$ from (16). It is also an undistorted *a priori* error but in the light of the decorrelated process ϕ_k rather than the original driving process $\mathbf{u}_k$. Following the derivations of [4], [12], and [13], the following theorem then holds immediately, the details of which are elaborated on in Appendix A.

Theorem 2.1 (Local Robustness of PAP): Consider the PAP algorithm in (6)–(9) and its Update (11). It always holds that

$$\frac{\|\tilde{\mathbf{w}}_{k+1}\|^2 + \frac{\alpha|f_{a,k}|^2}{\|\phi_k\|_2^2}}{\|\tilde{\mathbf{w}}_k\|^2 + \frac{\alpha|\tilde{v}_k|^2}{\|\phi_k\|_2^2}} \begin{cases} \le 1, & \text{for } 0 < \alpha < 1 \\ = 1, & \text{for } \alpha = 1 \\ \ge 1, & \text{for } \alpha > 1. \end{cases} \tag{22}$$

The essence of this theorem is that we have now a local robustness condition from time instant k to $k+1$ that shows that the energy of the a-posteriori parameter error vector and the undistorted *a priori* error are smaller than the energy of the *a priori* parameter error vector and the compound noise $\tilde{v}_k$. In Section III this local property will be instrumental to show a global property spanning over a time horizon from $k = 0, 1, \ldots, N$.

B. GAP Algorithm

Reformulating the update in (12) in parameter error vector form, we obtain

$$\begin{aligned}\tilde{\mathbf{w}}_{k+1} = &\left[\mathbf{I} - \alpha\mathbf{U}_{P+1,k}[\mathbf{U}_{P+1,k}^T\mathbf{U}_{P+1,k}]^{-1}\mathbf{U}_{P+1,k}^T\right]\tilde{\mathbf{w}}_k \\ &-\alpha\mathbf{U}_{P+1,k}[\mathbf{U}_{P+1,k}^T\mathbf{U}_{P+1,k}]^{-1}\mathbf{v}_k.\end{aligned} \tag{23}$$

From here, it is straightforward to derive the following theorem (details in Appendix B).

Theorem 2.2 (Local Robustness of GAP): Consider the GAP algorithm in (6)–(9) and its Update (12). It always holds that

$$\frac{\|\tilde{\mathbf{w}}_{k+1}\|^2 + \alpha\mathbf{e}_{a,k}^T[\mathbf{U}_{P+1,k}^T\mathbf{U}_{P+1,k}]^{-1}\mathbf{e}_{a,k}}{\|\tilde{\mathbf{w}}_k\|^2 + \alpha\mathbf{v}_k^T[\mathbf{U}_{P+1,k}^T\mathbf{U}_{P+1,k}]^{-1}\mathbf{v}_k} \begin{cases} \le 1, & \text{for } 0 < \alpha < 1 \\ = 1, & \text{for } \alpha = 1 \\ \ge 1, & \text{for } \alpha > 1. \end{cases} \tag{24}$$

Note that many variations are possible, for example time variant step-sizes α_k rather than a constant α. One can generalize the results for the GAP algorithmic variant [6] with regularization parameter $\beta > 0$:

$$\begin{aligned}\mathbf{w}_{k+1} = \mathbf{w}_k + \alpha_k\mathbf{U}_{P+1,k}&[\mathbf{U}_{P+1,k}^T\mathbf{U}_{P+1,k} + \beta\mathbf{I}]^{-1} \\ &\times[\mathbf{d}_k - \mathbf{U}_{P+1,k}^T\mathbf{w}_k]\end{aligned} \tag{25}$$

by following the lines of [11]. Some interesting results relating to time-variant regularization β_k were presented in [14].

III. Global Properties

Note that the theorems of the previous section have only delivered a so-called local robustness in terms of boundedness, that is when going from time instant k to the next $k+1$. They do not include such property when we observe the algorithms over a finite horizon, say from $k = 0, 1 \ldots, N$. In order to include robustness with respect to all uncertainties like noise $\{v_k\}_{0\le k\le N}$ but also the initial parameter estimation error $\tilde{\mathbf{w}}_0$ we have to extend our investigation. The procedure follows the concept of [12] in which the adaptive algorithm is split in two parts: a feed-forward structure that is an allpass and a feedback structure (see for example Fig. 1 further ahead), typically containing the step-size of the algorithm. Applying the small gain theorem [15], [16], only the feedback part is responsible for the l_2-stability of the algorithm and conservative stability bounds, depending on the step-size can be derived. The interested reader is referred to some textbooks [4, Ch. 17], [13] for more details.

A. PAP Algorithm

The derivation of the global properties of the PAP algorithm requires one further step in order to apply the small gain theorem. We have to find the relation between $f_{a,k}$ and $e_{a,k}$ as $\sqrt{\alpha} f_{a,k}$ is the output of the allpass and $e_{a,k}$ is part of the feedback signal $\tilde{v}_k$. Recalling that

$$\begin{aligned} f_{a,k} - e_{a,k} &= [\phi_k^T - \mathbf{u}_k{}^T]\tilde{\mathbf{w}}_k \\ &= -(1-\alpha)\hat{\mathbf{a}}_k^T \mathbf{e}_{a,k-1} + \alpha \hat{\mathbf{a}}_k^T \mathbf{v}_{k-1} \\ &= -(1-\alpha)\hat{A}(q^{-1})[e_{a,k}] + \alpha \hat{A}(q^{-1})[v_k] \end{aligned} \quad (26)$$

we are able to find

$$e_{a,k} = \frac{1}{1-(1-\alpha)\hat{A}(q^{-1})}[f_{a,k}] - \alpha \frac{\hat{A}(q^{-1})}{1-(1-\alpha)\hat{A}(q^{-1})}[v_k]. \quad (27)$$

Note that we have used the simplifying notation with a linear operator $\hat{A}(q^{-1})[\cdot]$ for which we even dropped the iteration index k which originates from our Assumption A1). The final step to obtain a global relation is to split the adaptive form into an allpass and a feedback structure. This is achieved by starting with (19):

$$\begin{aligned} \tilde{\mathbf{w}}_{k+1} &= \left[\mathbf{I} - \alpha \frac{\phi_k \phi_k^T}{\|\phi_k\|_2^2}\right] \tilde{\mathbf{w}}_k - \alpha \frac{\phi_k}{\|\phi_k\|_2^2} \tilde{v}_k \\ &= \left[\mathbf{I} - \frac{\phi_k \phi_k^T}{\|\phi_k\|_2^2}\right] \tilde{\mathbf{w}}_k - \frac{\phi_k}{\|\phi_k\|_2^2}(\alpha \tilde{v}_k + (\alpha - 1) f_{a,k}). \end{aligned} \quad (28)$$

Substituting $\tilde{v}_k$ from (20), we find the feedback part to consist of

$$\alpha \tilde{v}_k - (1-\alpha) f_{a,k} = \frac{1-\hat{A}(q^{-1})}{1-(1-\alpha)\hat{A}(q^{-1})}\left[\alpha v_k - (1-\alpha) f_{a,k}\right]. \quad (29)$$

Due to the small gain theorem, we are able now to formulate the desired conservative stability condition. Fig. 1 depicts the feedback structure of the PAP algorithm.

Theorem 3.1 (Global Stability of PAP): Consider the PAP algorithm in (6)–(9) and its Update (11) with sufficiently long filter order M (A1). It always holds that l_2-stability from its uncertainties $\{\tilde{\mathbf{w}}_0, \{\sqrt{\alpha}/\|\phi_k\|_2 v_k\}_{0 \le k \le N}\}$ to its errors $\{\tilde{\mathbf{w}}_{N+1}, \{\sqrt{\alpha}/\|\phi_k\|_2 f_{a,k}\}_{0 \le k \le N}\}$ is guaranteed if

$$\left|(1-\alpha)\frac{1-\hat{A}(q^{-1})}{1-(1-\alpha)\hat{A}(q^{-1})}\right| < 1. \quad (30)$$

Note that the term "sufficiently long filter order M" is somewhat vague relating to a stable estimate of the prediction coefficients $\hat{\mathbf{a}}_k$ which in turn also results in $\|\phi_k\|_2^2 = \mathrm{const}$ as already argued in the introduction. If the filter order is too small, these coefficients fluctuate considerably and the above condition only holds in the more impractical form of impulse response matrices (see for example [12]). In Appendix C, we provide the derivation without Assumption A1). In simulations (see Section V), we found a filter order of $M \ge 20$ to be sufficient for A1) to hold. For smaller values of M, the derived conditions lack precision.

In simulations, the derived upper bound appears surprisingly sharp. One reason may be in the derived methodology using an allpass filter in the forward path. As an allpass is unitary, it is energy preserving and the entire stability condition is concentrated in the gain of the feedback path. A second reason for obtaining relatively sharp upper bounds is that the additive noise as well as the feedback signal are filtered by the same transfer function as shown in (29). Thus, if the potential stability border is reached, that is the filter's maximum gain is one, the maximum noise energy occurs at exactly the same frequency as the driving process, ensuring that there is sufficient excitation at this point.

The result of Theorem 3.1 does not only provide an upper bound for the step-size. For some input signal statistics, there exists indeed also a lower bound (see Example 2 further ahead). This discovery teaches us that the PAP algorithm is not recommended to be applied with low step-sizes unless the statistics of the driving process are well known *a priori*.

B. GAP Algorithm

We now derive the global properties of the GAP algorithm. We start again with the parameter error vector form, that we reformulate for obtaining an allpass in the feedforward path.

$$\begin{aligned} \tilde{\mathbf{w}}_{k+1} &= \left[\mathbf{I} - \alpha \mathbf{U}_{P+1,k}[\mathbf{U}_{P+1,k}^T \mathbf{U}_{P+1,k}]^{-1}\mathbf{U}_{P+1,k}^T\right]\tilde{\mathbf{w}}_k \\ &\quad - \alpha \mathbf{U}_{P+1,k}[\mathbf{U}_{P+1,k}^T \mathbf{U}_{P+1,k}]^{-1}\mathbf{v}_k \\ &= \left[\mathbf{I} - \mathbf{U}_{P+1,k}[\mathbf{U}_{P+1,k}^T \mathbf{U}_{P+1,k}]^{-1}\mathbf{U}_{P+1,k}^T\right]\tilde{\mathbf{w}}_k \\ &\quad - \mathbf{U}_{P+1,k}[\mathbf{U}_{P+1,k}^T \mathbf{U}_{P+1,k}]^{-1}\tilde{\mathbf{v}}_k \end{aligned} \quad (31)$$

where we introduced a modified noise term containing the feedback part

$$\tilde{\mathbf{v}}_k = \alpha \mathbf{v}_k - (1-\alpha)\mathbf{U}_{P+1,k}^T \tilde{\mathbf{w}}_k. \quad (32)$$

From here, we recognize that the only term in the feedback path is $1-\alpha$ which immediately allows the formulation of a global l_2-stability condition.

Theorem 3.2 (Global Stability of GAP): Consider the GAP algorithm with (6)–(9) and its Update (12). It always holds that l_2-stability from its uncertainties $\{\tilde{\mathbf{w}}_0, \{\sqrt{\alpha}\sqrt{\mathbf{v}_k^T[\mathbf{U}_{P+1,k}^T \mathbf{U}_{P+1,k}]^{-1}\mathbf{v}_k}\}_{0 \le k \le N}\}$ to its errors $\{\tilde{\mathbf{w}}_{N+1}, \{\sqrt{\alpha}\sqrt{\mathbf{e}_{a,k}^T[\mathbf{U}_{P+1,k}^T \mathbf{U}_{P+1,k}]^{-1}\mathbf{e}_{a,k}}\}_{0 \le k \le N}\}$ is guaranteed if

$$0 < \alpha < 2.$$

It is worth comparing with stability results in [3, Ch. 6.5] and [4, Ch. 6.11] for this algorithm as the results are being derived only from the weaker mean square analysis.

IV. Steady-State Behavior

Due to the feedback structure it is now also possible to compute the steady-state error of the PAP algorithm.

Theorem 4.1 (Steady-State of PAP): Consider the PAP algorithm in (6)–(9) with its Update (11) with sufficiently long filter order M (A1). Under the assumption that the driving process ϕ_k is perfectly decorrelated by the adaptive filter (P is sufficiently large) and the additive noise is white, the mismatch $\mathcal{M}_{\mathrm{PAP}}$ is given by

$$\mathcal{M}_{\mathrm{PAP}} = \alpha^2\left[\frac{G^2 H^2}{1-(1-\alpha)^2 H^2} + F^2\right] \quad (33)$$

with the abbreviations

$$H^2 = \frac{1}{2\pi}\int \left| \frac{1 - A\left(e^{j\Omega}\right)}{1-(1-\alpha)A\left(e^{j\Omega}\right)} \right|^2 d\Omega \tag{34}$$

$$G^2 = \frac{1}{2\pi}\int \left| \frac{1}{1-(1-\alpha)A\left(e^{j\Omega}\right)} \right|^2 d\Omega \tag{35}$$

$$F^2 = \frac{1}{2\pi}\int \left| \frac{A\left(e^{j\Omega}\right)}{1-(1-\alpha)A\left(e^{j\Omega}\right)} \right|^2 d\Omega. \tag{36}$$

Proof: If we can assume that P is sufficiently large, then the *a priori* error $f_{a,k} = \phi_k^T \tilde{\mathbf{w}}_k$ is perfectly decorrelated and further assuming that the additive noise v_k is also white, we find the energy relation on the feedback part to be

$$\begin{aligned} E[|f_{a,k}|^2] &= \alpha^2 H^2 E[|v_k|^2] + (1-\alpha)^2 H^2 E[|f_{a,k}|^2] \\ &= \sigma_v^2 \frac{\alpha^2 H^2}{1-(1-\alpha)^2 H^2} \end{aligned} \tag{37}$$

with the abbreviation H^2 given above and σ_v^2 denoting the variance of the additive noise. Applying further (27), we also can relate $E[|f_{a,k}|^2]$ to $E[|e_{a,k}|^2]$ and obtain as mismatch $\mathcal{M}_{\text{PAP}}$

$$\begin{aligned} \mathcal{M}_{\text{PAP}} &= \frac{E[|e_{a,k}|^2]}{\sigma_v^2} \\ &= \alpha^2 \left[\frac{E[|f_{a,k}|^2]}{\sigma_v^2} G^2 + F^2 \right] \\ &= \alpha^2 \left[\frac{G^2 H^2}{1-(1-\alpha)^2 H^2} + F^2 \right] \end{aligned} \tag{38}$$

with the two additional abbreviations G and F defined in the theorem. ■

The mismatch of the GAP algorithm is straightforward to compute and is only given explicitly here to round up the topic.

Theorem 4.2 (Steady-State of GAP): Consider the GAP algorithm (6)–(9) with its Update (12). Under the assumption that P is sufficiently large so that the driving process ϕ_k is perfectly decorrelated by the adaptive filter and the additive noise is white, it always holds that

$$\mathcal{M}_{\text{GAP}} = \frac{\alpha}{2-\alpha}.$$

Note that this result is somewhat surprisingly simple in particular when compared to the elaborate results in [18] that includes tracking behavior. However, we assumed here that the driving process is perfectly decorrelated which may not be the case. Also note that the result is equivalent to the result from [4, Lemma 6.10.1].

V. Simulation Examples

The following simulation examples show the algorithmic behavior for two systems of length $M = 64$ with different sets of AR coefficients $\mathbf{a}$. We are thus applying an AR process of the form

$$u_k = \sum_{i=1}^{P} a_i u_{k-i} + n_k = \mathbf{a}^T \mathbf{u}_k + n_k \tag{39}$$

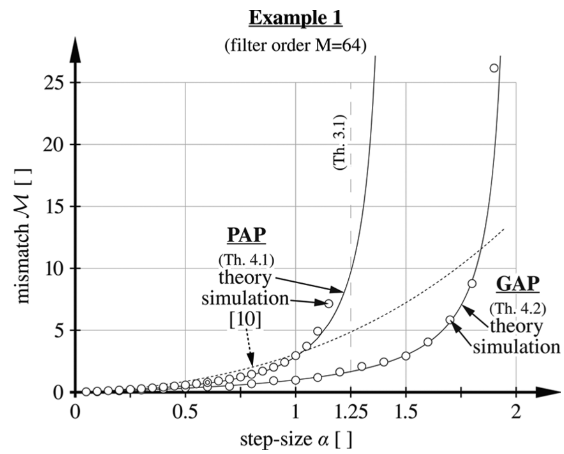

Fig. 2. Mismatch of Example 1, PAP and GAP algorithms.

with white noise n_k. All runs are performed by Gaussian sequences n_k for driving as well as additive noise sequences v_k. The noise was set to $E[u_k]/E[v_k] = 40$ dB (SNR). The results were averaged over 50 runs each in order to obtain values of sufficient quality. We apply a very small regularization parameter ($\beta = 0.0001$) in (25). We use two sets of prediction coefficients $\mathbf{a}$ in the following:

Example 1: The first set $\mathbf{a}^T = [0.8, -0.5, 0.2]$ is an example, taken from [10]. From the stability condition of Theorem 3.1 only an upper stability border results at $\alpha = 1.25$ for the PAP algorithm.

Example 2: The second example takes only two AR coefficients $\mathbf{a}^T = [-1.8, -0.95]$. Different to the previous example and all examples from [10] is that here the unusual situation occurs that also a lower stability bound for the step-size exists. The stability bounds according to Theorem 3.1 are 0.45 for the lower bound and 1.16 for the upper bound of the PAP algorithm.

Fig. 2 depicts the mismatch $\mathcal{M}$ for PAP as well as GAP algorithm over various step-sizes employing the set of coefficients from Example 1. For the PAP algorithm an even better agreement compared to [10] is found [compare to Fig. 4(a) in [10] repeated as dashed line in Fig. 2] with a much simpler formula according to Theorem 4.1. Only at the stability border the prediction becomes poor. The predicted stability border following Theorem 3.1, however, is in excellent agreement with the experimental results. It appears as rather sharp limit.

The GAP algorithm on the other hand does not show stability problems before reaching $\alpha = 2$ as predicted by Theorem 3.2. The mismatch follows exactly the simple formula of Theorem 4.2 even close to the stability bound. As the figure reveals theory and simulation are in excellent agreement even close to the stability border.

Fig. 3 depicts the obtained mismatch for both algorithms when employing the set of prediction coefficients from Example 2. As in the previous example the estimates are only poor on the stability border. For most of the stable step-size range the agreement with the PAP theory is excellent. For the GAP algorithm the theory delivers an excellent fit through the entire range of step-sizes. According to Theorem 3.1, we predict a lower stability bound at $\alpha = 0.45$ which was found in the

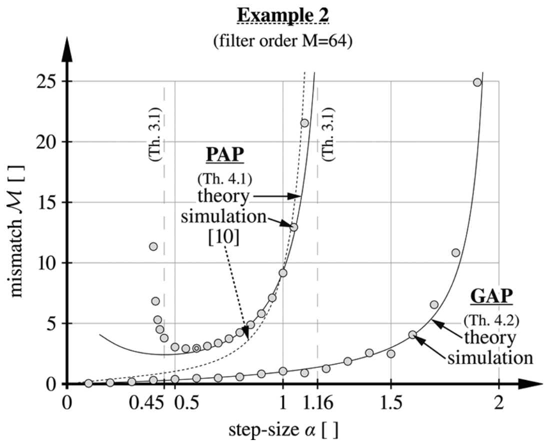

Fig. 3. Mismatch of Example 2, PAP and GAP algorithms.

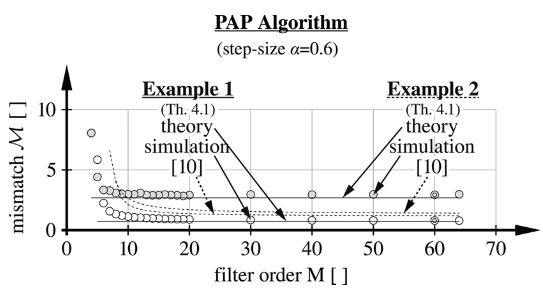

Fig. 4. Mismatch of Example 1 and Example 2, PAP algorithm under various filter order M, $\alpha = 0.6$.

simulation at $\alpha = 0.39$. The theory in [10] is not predicting it at all (see dashed line in the figure). The upper stability bound at $\alpha = 1.16$ appears very sharp.

In Fig. 4, the mismatch of the PAP algorithm is presented for various filter order $M = [5, 6, \ldots, 20, 30, 40, 50, 60, 64]$, repeating the sets of Experiment 1 and $M = [4, 5, \ldots, 20, 30, 40, 50, 60, 64]$ for Experiment 2. We selected the normalized step-size in both cases to $\alpha = 0.6$ for which we expect stable behavior of the algorithm. We plot the misadjustment according to Theorem 4.1 (denoted theory in the figure). As the figure reveals, our assumption that M is sufficiently large seems to hold for even small values $M > 10$, sacrificing precision but definitely for practical values of $M \geq 20$. A comparison to the misadjustment from [10] reveals large discrepancies for small as well as large values of the filter order M. The Matlab code for these experiments is available under https://www.nt.tuwien.ac.at/downloads/featured-downloads.

VI. Conclusion

Although many substantial questions relating stability and steady state of pseudo affine projection algorithms could be answered in this contribution, some issues remain open.

1) If the filter length M becomes small ($M - P + 2 \leq 20$), the estimates of the prediction coefficients fluctuate and have considerable impact on stability as well as on steady-state. This impact in qualitative and quantitative terms remains very difficult to describe with simple terms [see (53) in Appendix C]. A simpler description not including actual values of ϕ_k and $\hat{\mathbf{a}}_k$ is desired. Intuitively, the mismatch curves can be corrected by a factor $M/[M-P-1]$ which would offer good agreement even for small values of M but its justification remains an open issue.
2) A stability bound at low step-sizes for the PAP algorithm was correctly predicted in our contribution. However, the bound is an upper bound due to conservative arguments, a more correct and sharp bound is desired.

Appendix

A. Derivation of Local Robustness Properties in Theorem 2.1

We start with (19) that can equivalently be written as

$$\tilde{\mathbf{w}}_{k+1} = \tilde{\mathbf{w}}_k - \alpha \frac{\phi_k}{\|\phi_k\|_2^2}(f_{a,k} + \tilde{v}_k). \tag{40}$$

Computing the squared l_2-norm on both sides of the equation, we arrive at

$$\begin{aligned}\|\tilde{\mathbf{w}}_{k+1}\|_2^2 &= \|\tilde{\mathbf{w}}_k\|_2^2 + \frac{\alpha^2}{\|\phi_k\|_2^2}|f_{a,k} + \tilde{v}_k|^2 \\ &\quad - \frac{\alpha}{\|\phi_k\|_2^2}[f_{a,k}(f_{a,k} + \tilde{v})^* + f_{a,k}^*(f_{a,k} + \tilde{v}_k)] \\ &= \|\tilde{\mathbf{w}}_k\|_2^2 + \frac{\alpha^2}{\|\phi_k\|_2^2}|f_{a,k} + \tilde{v}_k|^2 \\ &\quad + \frac{\alpha}{\|\phi_k\|_2^2}\left[|\tilde{v}_k|^2 - |f_{a,k} + \tilde{v}_k|^2 - |f_{a,k}|^2\right]. \end{aligned} \tag{41}$$

Rearranging terms leads to

$$\begin{aligned}\|\tilde{\mathbf{w}}_{k+1}\|_2^2 + \frac{\alpha}{\|\phi_k\|_2^2}|f_{a,k}|^2 &= \|\tilde{\mathbf{w}}_k\|_2^2 + \frac{\alpha}{\|\phi_k\|_2^2}|\tilde{v}_k|^2 \\ &\quad + \frac{\alpha(\alpha-1)}{\|\phi_k\|_2^2}|f_{a,k} + \tilde{v}_k|^2. \end{aligned} \tag{42}$$

The last term is negative for $\alpha < 1$ and positive for $\alpha > 1$, explaining the bounds of Theorem 2.1. We like to remark here that many other local bounds in terms of *a priori* and/or *a posteriori* errors can be defined. Also the extended range of $1 < \alpha < 2$ can be shown to have local robustness. The interested reader is referred here to [4], [12] and [17] for more details.

B. Derivation of Local Robustness Properties in Theorem 2.2

We start with (23) in the form

$$\tilde{\mathbf{e}}_{a,k} = \left[\mathbf{U}_{P+1,k}^T \tilde{\mathbf{w}}_k + \mathbf{v}_k\right] \tag{43}$$

$$\tilde{\mathbf{w}}_{k+1} = \tilde{\mathbf{w}}_k - \alpha \mathbf{U}_{P+1,k}[\mathbf{U}_{P+1,k}^T \mathbf{U}_{P+1,k}]^{-1}\tilde{\mathbf{e}}_{a,k}. \tag{44}$$

Squaring at both sides of the equation leads to

$$\begin{aligned}\|\tilde{\mathbf{w}}_{k+1}\|_2^2 &= \|\tilde{\mathbf{w}}_k\|_2^2 + \alpha^2 \tilde{\mathbf{e}}_{a,k}^T[\mathbf{U}_{P+1,k}^T \mathbf{U}_{P+1,k}]^{-1}\tilde{\mathbf{e}}_{a,k} \\ &\quad - 2\alpha \mathbf{e}_{a,k}^T[\mathbf{U}_{P+1,k}^T \mathbf{U}_{P+1,k}]^{-1}\tilde{\mathbf{e}}_{a,k}. \end{aligned} \tag{45}$$

With a similar mathematical reformulation as before $\tilde{\mathbf{e}}_{a,k} - \mathbf{e}_{a,k} = \mathbf{v}_k$, we obtain

$$\begin{aligned}\|\tilde{\mathbf{w}}_{k+1}\|_2^2 &+ \alpha \mathbf{e}_{a,k}^T[\mathbf{U}_{P+1,k}^T \mathbf{U}_{P+1,k}]^{-1}\mathbf{e}_{a,k} \\ &= \|\tilde{\mathbf{w}}_k\|_2^2 + \alpha \mathbf{v}_k^T[\mathbf{U}_{P+1,k}^T \mathbf{U}_{P+1,k}]^{-1}\mathbf{v}_k \\ &\quad + \alpha(\alpha-1)\tilde{\mathbf{e}}_{a,k}^T[\mathbf{U}_{P+1,k}^T \mathbf{U}_{P+1,k}]^{-1}\tilde{\mathbf{e}}_{a,k}. \end{aligned} \tag{46}$$

Depending on $\alpha < 1$ or $\alpha > 1$, we can neglect the last term or not and the theorem is proven.

C. Derivation of Theorem 3.1 Without Assumption A1

We start with (28) in the following form:

$$\begin{aligned}\tilde{\mathbf{w}}_{k+1} &= \left[\mathbf{I} - \frac{\phi_k\phi_k^T}{\|\phi_k\|_2^2}\right]\tilde{\mathbf{w}}_k - \frac{\phi_k}{\|\phi_k\|_2^2}\underbrace{(\alpha\tilde{v}_k + (\alpha-1)f_{a,k})}_{\bar{v}_k} \\ &= \tilde{\mathbf{w}}_k - \frac{\phi_k}{\|\phi_k\|_2^2}[f_{a,k} + \bar{v}_k]. \end{aligned} \quad (47)$$

Here, the new compound noise $\bar{v}_k$ is given by

$$\begin{aligned}\bar{v}_k &= \alpha\tilde{v}_k - (1-\alpha)f_{a,k} \\ &= \frac{1-\hat{A}_k(q^{-1})}{1-(1-\alpha)\hat{A}_k(q^{-1})}\left[\alpha v_k - (1-\alpha)f_{a,k}\right] \end{aligned} \quad (48)$$

where we applied time-variant operators $\hat{A}_k(q^{-1})$ now. Formulating this in matrix form, we obtain

$$\begin{aligned}[\mathbf{I} - (1-\alpha)\mathbf{A}]\bar{\mathbf{v}}_k &= [\mathbf{I} - \mathbf{A}]\left[\alpha\mathbf{v}_k - (1-\alpha)\mathbf{f}_{a,k}\right] \\ \bar{\mathbf{v}}_k &= \alpha\underbrace{[\mathbf{I} - (1-\alpha)\mathbf{A}]^{-1}[\mathbf{I} - \mathbf{A}]}_{\mathbf{F}}\mathbf{v}_k \\ &\quad - (1-\alpha)[\mathbf{I} - (1-\alpha)\mathbf{A}]^{-1}[\mathbf{I} - \mathbf{A}]\mathbf{f}_{a,k}, \\ &= \alpha\mathbf{F}\mathbf{v}_k - (1-\alpha)\mathbf{F}\mathbf{f}_{a,k} \end{aligned} \quad (49)$$

where we used the following impulse response matrix of dimension $(N+1)\times(N+1)$ as in [12]:

$$\mathbf{A} = \begin{bmatrix} 0 & a_{1,0} & a_{2,0} & \cdots \\ 0 & 0 & a_{1,1} & a_{2,1} \\ & & & \\ 0 & 0 & \ddots & \end{bmatrix} \quad (50)$$

as well as vector forms of v_k, $\bar{v}_k$ and $f_{a,k}$ of appropriate dimensions, starting at time instant $k=0$ until N. The first row of $\mathbf{A}$ contains the filter coefficients of the predictor at time instant 0, the next at time instant 1 and so on. Squaring both sides of (48), we obtain

$$\|\tilde{\mathbf{w}}_{k+1}\|_2^2 + \frac{|f_{a,k}|^2}{\|\phi_k\|_2^2} = \|\tilde{\mathbf{w}}_k\|_2^2 + \frac{|\bar{v}_k|^2}{\|\phi_k\|_2^2}. \quad (51)$$

We thus find $\bar{v}_k/\|\phi_k\|_2$ being the input of the allpass and $f_{a,k}/\|\phi_k\|_2$ the output of it which are in vector notation $\mathbf{M}\bar{\mathbf{v}}$ and $\mathbf{M}\bar{\mathbf{f}}_{a,k}$, respectively, with

$$\mathbf{M} = \mathrm{diag}\left\{\frac{1}{\|\phi_0\|_2}, \frac{1}{\|\phi_1\|_2}, \ldots, \frac{1}{\|\phi_N\|_2}\right\}. \quad (52)$$

With this the feedback part reads $(1-\alpha)\mathbf{M}\mathbf{F}\mathbf{M}^{-1}$ and the l_2-stability condition reads

$$\left\|(1-\alpha)\mathbf{M}[\mathbf{I} - (1-\alpha)\mathbf{A}]^{-1}[\mathbf{I} - \mathbf{A}]\mathbf{M}^{-1}\right\|_{2,\mathrm{ind}} < 1. \quad (53)$$

In order to apply the condition for each update the norm of a growing matrix has to be computed and due to the nonlinear dependency on α numerically evaluated.

Acknowledgment

The author would like to thank S. Caban for preparing the figures. The author is grateful for the numerous suggestions and comments of the anonymous reviewers.

References

[1] K. Ozeki and T. Umeda, "An adaptive filtering algorithm using orthogonal projection to an affine subspace and its properties," *Electron. Commun. Japan*, vol. 67-A, no. 5, pp. 19–27, 1984.

[2] M. Rupp, "A family of adaptive filter algorithms with decorrelating properties," *IEEE Trans. Signal Process.*, vol. 46, no. 3, pp. 771–775, 1998.

[3] S. Haykin, *Adaptive Filter Theory*, 4th ed. Englewood Cliffs, NJ: Prentice-Hall, 2002.

[4] A. H. Sayed, *Fundamentals of Adaptive Filtering*. New York: Wiley, 2003.

[5] S. L. Gay, "A fast converging, low complexity adaptive filtering algorithm," presented at the IEEE Acoust., Speech, Signal Process. Workshop, New Paltz, NY, Oct. 1993.

[6] S. L. Gay and S. Tavathia, "The fast affine projection algorithm," in *Proc. Int. Conf. Acoust., Speech, Signal Process. (ICASSP)*, Detroit, MI, 1995, pp. 3023–3026.

[7] D. Slock, "The block underdetermined covariance (BUC) fast transversal filter (FTF) algorithm for adaptive filtering," in *Proc. 26th. Asilomar Conf.*, Pacific Grove, CA, 1992, pp. 550–554.

[8] D. Slock, "Underdetermined growing and sliding covariance fast transversal filter RLS algorithms," in *Proc. EUSIPCO*, Brussels, Belgium, 1992, pp. 1169–1172.

[9] B. Baykal, D. Tarikulo, and A. Constantinides, "Asymptotic analysis of the underdetermined RLS algorithm," in *Proc. EUSIPCO*, Trieste, Italy, Sep. 1996, pp. 1397–1400.

[10] S. J. M. de Almeida, J. C. M. Bermudez, and N. J. Bershad, "A stochastic model for a pseudo affine projection algorithm," *IEEE Trans. Signal Process.*, vol. 57, no. 1, pp. 107–118, Jan. 2009.

[11] M. Rupp and A. H. Sayed, "A robustness analysis of Gauss–Newton recursive methods," *Signal Process.*, vol. 50, no. 3, pp. 165–188, Jun. 1996.

[12] M. Rupp and A. H. Sayed, "A time-domain feedback analysis of filtered-error adaptive gradient algorithms," *IEEE Trans. Signal Process.*, vol. 44, pp. 1428–1440, Jun. 1996.

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

[14] H. Rey, L. Rey-Vega, S. Tressens, and J. Benesty, "Variable explicit regularization in affine projection algorithm: Robustness issues and optimal choice," *IEEE Trans. Signal Process.*, vol. 55, no. 5, pp. 2096–2109, May 2007.

[15] H. K. Khalil, *Nonlinear Systems*. New York: MacMillan, 1992.

[16] M. Vidyasagar, *Nonlinear Systems Analysis*, 2nd ed. Englewood Cliffs, NJ: Prentice-Hall, 1993.

[17] A. H. Sayed and M. Rupp, "Error-energy bounds for adaptive gradient algorithms," *IEEE Trans. Signal Process.*, vol. 44, pp. 1982–1989, Aug. 1996.

[18] S. Hyun-Chool and A. H. Sayed, "Transient behavior of affine projection algorithms," in *Proc. Int. Conf. Acoust., Speech, Signal Process. (ICASSP)*, Hong Kong, Apr. 2003, pp. 353–356.

Markus Rupp (M'03–SM'06) received the Dipl.-Ing. degree from the University of Saarbrücken, Saarbrücken, Germany, in 1988 and the Dr.Ing. degree from Technische Universität Darmstadt, Darmstadt, Germany, in 1993, where he worked with E. Hänsler on designing new algorithms for acoustical and electrical echo compensation.

From November 1993 to July 1995, he held, with S. Mitra, a postdoctoral position with the University of California, Santa Barbara, where he worked with A. H. Sayed on a robustness description of adaptive filters with impact on neural networks and active noise control. From October 1995 to August 2001, he was a Member of Technical Staff with the Wireless Technology Research Department, Bell Laboratories, Crawford Hill, NJ, where he worked on various topics related to adaptive equalization and rapid implementation for IS-136, 802.11 and the Universal Mobile Telecommunications System. Since October 2001, he has been a Full Professor of digital signal processing in mobile communications with the Vienna University of Technology, where he founded the Christian Doppler Laboratory for Design Methodology of Signal Processing Algorithms, at the Institute of Communications and Radio-Frequency Engineering, in 2002. He served as Dean from 2005 to 2007. He is the author or a coauthor of more than 350 papers and is the holder of 15 patents on adaptive filtering, wireless communications and rapid prototyping, as well as automatic design methods.

Dr. Rupp is currently an Associate Editor for the JASP *EURASIP Journal of Advances in Signal Processing* and the JES *EURASIP Journal on Embedded Systems*. He was an Associate Editor for the IEEE TRANSACTIONS ON SIGNAL PROCESSING from 2002 to 2005. He has been an Administrative Committee Member of EURASIP since 2004 and served as the President of EURASIP from 2009 to 2010.