STABILITY ANALYSIS OF AN ADAPTIVE WIENER STRUCTURE

Robert Dallinger and Markus Rupp

Vienna University of Technology
Institute of Communications and Radio-Frequency Engineering
Gusshausstrasse 25/389, A-1040 Vienna, Austria
Email: {rdalling, mrupp}@nt.tuwien.ac.at

ABSTRACT

In the context of digital pre-distortion, a typical requirement is to identify the power amplifier with stringently low computational complexity. Accordingly, we consider a simple gradient method which is used to adaptively fit a simplified Wiener model, i.e., a cascade of a linear filter followed by a memoryless nonlinearity, to a dispersive and saturating reference system which represents the power amplifier. For adaptation, the gradient method only relies on the difference between the output of the reference system and the Wiener model. We show that such a structure can be formulated as a proportionate normalised least mean squares (PNLMS) algorithm. As a consequence, conditions for stability in the mean square sense can be deduced. Although not proven in a strict sense, simulation results allow to conjecture robustness.

Index Terms— Nonlinear filters, gradient methods, robustness, stability, identification.

1. INTRODUCTION

In the context of digital pre-distortion for power amplifiers (PA), the use of simple PA models is crucial to keep the necessary processing power as low as possible. For broadband signals, such models are required to describe nonlinear distortions as well as linear dispersive effects [1]. For these reasons, a broadly used structure is the simplified Wiener model which consists of a finite impulse response filter (FIR) followed by a memoryless nonlinearity [1,2]. Considering the equivalent baseband behaviour, not only the output power depends nonlinearly on the input power due to saturation, but also the introduced phase offset is a nonlinear function of the input power [3]. Consequently, the afore mentioned FIR filter as well as the nonlinear function are complex-valued. However, the analyses presented in this work restrict to the real-valued case.

1.1. Simplified Wiener model

Lets first assume that the behaviour of a PA can be described by a memoryless mapping $\mathsf{h} : u_k \mapsto z_k$ which relates the input u_k at time instant k to the corresponding output z_k. In reality, h is a smooth function which allows an approximation around the origin by a polynomial of degree P. Although the PA is a real-valued system which will introduce even as well as odd order nonlinear distortions, due to zonal filtering, in the baseband domain only the influence of the odd order terms remains [4]. Consequently, with $u_k \in \mathbb{C}$ and $z_k \in \mathbb{C}$, such a polynomial will be of the form,

$$z_k = \sum_{\substack{p=1 \\ p \text{ odd}}}^{P} a_p u_k \left|u_k\right|^{p-1} \tag{1}$$

with appropriate coefficients $a_p \in \mathbb{C}$ and the odd polynomial degree P. Note that in [5], it was pointed out that in the context of PA modelling and pre-distortion, including even order terms can be beneficial. In this paper, however, we stick to (1) since it is the commonly used approach and the extension to include even order terms is straightforward.

An equivalent representation of (1) is found by introducing certain polynomials $\psi_p(u_k)$ of degree p as basis functions instead of the powers $u_k \left|u_k\right|^{p-1}$, i.e.,

$$z_k = \sum_{\substack{p=1 \\ p \text{ odd}}}^{P} h_p \psi_p(u_k) \quad \text{with} \quad \psi_p(u_k) = \sum_{\substack{i=1 \\ i \text{ odd}}}^{p} \alpha_{p,i} u_k \left|u_k\right|^{i-1}. \tag{2}$$

As pointed out in [6], choosing an adequate set of basis polynomials (such as Hermite or Chebyshev polynomials [7]), in the context of parameter estimation, the parameters h_p can be estimated with higher numerical stability than the parameters a_p in (1). The actual configuration of the basis is determined by the coefficients $\alpha_{p,i} \in \mathbb{C}$. Introducing the basis vector $\boldsymbol{\psi}(u_k) = \left[\psi_1(u_k), \ldots, \psi_P(u_k)\right]^{\mathrm{T}}$ and the corresponding vector of coefficients $\mathbf{h} = \left[h_1, \ldots, h_P\right]^{\mathrm{T}}$, (2) can be rewritten as

$$z_k = \mathbf{h}^{\mathrm{T}} \boldsymbol{\psi}(u_k). \tag{3}$$

Note that $\mathbf{h}$ and $\boldsymbol{\psi}(u_k)$ have length $M_h = \left\lceil \frac{1}{2} P \right\rceil$ since only odd order entries are contained.

Finally, with the vector of FIR filter coefficients $\mathbf{g}$ and the vector of input samples $\mathbf{x}_k^{\mathrm{T}} = \left[x_k, x_{k-1}, \ldots, x_{k-M_g+1}\right]$ the output z_k of the simplified Wiener model (see e.g., [8]) in the upper branch of Fig. 1 is given by

$$z_k = \mathbf{h}^{\mathrm{T}} \boldsymbol{\psi}\left(\mathbf{g}^{\mathrm{T}} \mathbf{x}_k\right). \tag{4}$$

1.2. Adaptive Wiener structure

A low complexity method to adaptively fit a Wiener structure to the behaviour of a reference system (e.g., a PA) is depicted in Fig. 1. The upper branch corresponds to the reference system, which here is assumed to be a Wiener model itself. It is represented by the FIR weight vector $\mathbf{g}$ and the coefficient vector $\mathbf{h}$ of the nonlinearity (cf. (3)). Analogously, the lower branch represents the adaptive Wiener structure with the time varying parameter vectors $\hat{\mathbf{g}}_k$ and $\hat{\mathbf{h}}_k$.

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

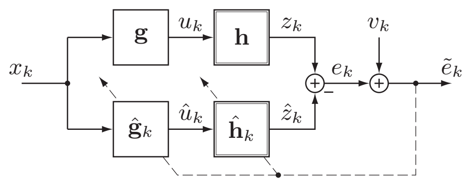

Fig. 1. *The adaptive filter uses a Wiener model. By assumption, the reference system has the same structure.*

Following an update rule as in [9], the same disturbed error $\tilde{e}_k = z_k - \hat{z}_k + v_k$ (the additive noise v_k is assumed to be inherent and not accessible for compensation) is used to update the linear part as well as the nonlinearity, according to

$$\begin{aligned} \hat{\mathbf{g}}_k &= \hat{\mathbf{g}}_{k-1} + \mu_{g,k}\mathbf{x}_k^* \tilde{e}_k & (5)\\ \hat{\mathbf{h}}_k &= \hat{\mathbf{h}}_{k-1} + \mu_{h,k}\boldsymbol{\psi}^*(\hat{u}_k)\,\tilde{e}_k, & (6) \end{aligned}$$

with $\hat{u}_k = \hat{\mathbf{g}}_{k-1}^{\mathrm{T}}\mathbf{x}_k$ and the positive step-sizes $\mu_{g,k}$ and $\mu_{h,k}$. As introduced in (3), $\boldsymbol{\psi}(.)$ is the vector composed by the chosen basis polynomials.

1.3. Proportionate least mean squares algorithm

Originally introduced in [10], the proportionate normalised least mean squares (PNLMS) algorithm was considered and modified in many recent works [11–13]. Compared to other gradient methods (such as the normalised least mean squares algorithm), it shows faster convergence if utilised for the identification of systems with sparse parameter vectors, e.g., echo cancellation.

Assuming that $\mathbf{p}$ denotes the parameter vector of some unknown transversal filter and the vector $\hat{\mathbf{p}}_k$ contains the estimated parameters at time instant k, in the real-valued domain, the update equation of the PNLMS is given by,

$$\hat{\mathbf{p}}_k = \hat{\mathbf{p}}_{k-1} + \mu\mathbf{D}_k\boldsymbol{\eta}_k\tilde{f}_k, \tag{7}$$

with the positive constant step-size μ, the diagonal weighting matrix $\mathbf{D}_k$, the a priori error $\tilde{f}_k = \boldsymbol{\eta}_k^{\mathrm{T}}(\mathbf{p} - \hat{\mathbf{p}}_{k-1}) + n_k$ disturbed by the noise n_k, and the excitation vector $\boldsymbol{\eta}_k$.

2. PNLMS REPRESENTATION OF THE ADAPTIVE WIENER MODEL

For the rest of this paper, all signals and systems are assumed to be real-valued. The extension to the complex domain needs further investigation and will be presented elsewhere.

Considering the adaptive Wiener structure in Sec. 1.2, under the assumption that the parameter vectors of the reference system (upper branch in Fig. 1) have the same lengths M_g and M_h as the parameter vectors of the adaptive system (lower branch in Fig. 1), the parameter error vectors $\tilde{\mathbf{h}}_k = \mathbf{h} - \hat{\mathbf{h}}_k$ and $\tilde{\mathbf{g}}_k = \mathbf{g} - \hat{\mathbf{g}}_k$ can be defined. The update error $\tilde{e}_k$ occurring in (5) and (6) can then be expressed as

$$\begin{aligned} \tilde{e}_k &= \mathbf{h}^{\mathrm{T}}\boldsymbol{\psi}(u_k) - \hat{\mathbf{h}}_{k-1}^{\mathrm{T}}\boldsymbol{\psi}(\hat{u}_k) + v_k \\ &= \mathbf{h}^{\mathrm{T}}\left[\boldsymbol{\psi}(u_k) - \boldsymbol{\psi}(\hat{u}_k)\right] + \tilde{\mathbf{h}}_{k-1}^{\mathrm{T}}\boldsymbol{\psi}(\hat{u}_k) + v_k. \end{aligned} \tag{8}$$

Since the nonlinearity is differentiable, the mean value theorem [14] allows to conclude that

$$\mathbf{h}^{\mathrm{T}}\left[\boldsymbol{\psi}(u_k) - \boldsymbol{\psi}(\hat{u}_k)\right] = \nu_{h,k}(u_k - \hat{u}_k), \tag{9}$$

where $\nu_{h,k}$ is defined as the first derivative of the nonlinearity at an adequate point η_k lying in the open interval spanned by $u_k = \mathbf{g}^{\mathrm{T}}\mathbf{x}_k$ and $\hat{u}_k = \hat{\mathbf{g}}_{k-1}^{\mathrm{T}}\mathbf{x}_k$, i.e.,

$$\nu_{h,k} = \mathbf{h}^{\mathrm{T}}\boldsymbol{\psi}'(\eta_k), \tag{10}$$

with the vector $\boldsymbol{\psi}'(\eta_k)$ containing the first derivatives of the basis polynomials evaluated at η_k. Substituting the first term in (8) by (9), the error $\tilde{e}_k$ equivalently reads

$$\tilde{e}_k = \nu_{h,k}\tilde{\mathbf{g}}_{k-1}^{\mathrm{T}}\mathbf{x}_k + \tilde{\mathbf{h}}_{k-1}^{\mathrm{T}}\boldsymbol{\psi}(\hat{u}_k) + v_k. \tag{11}$$

With the definition of the combined parameter vectors $\mathbf{w}$ respectively $\hat{\mathbf{w}}_k$, and the stacked (and weighted) excitation vector $\boldsymbol{\xi}_k$,

$$\mathbf{w}_k = \begin{bmatrix}\mathbf{g}\\ \mathbf{h}\end{bmatrix}, \qquad \hat{\mathbf{w}}_k = \begin{bmatrix}\hat{\mathbf{g}}_k\\ \hat{\mathbf{h}}_k\end{bmatrix}, \qquad \boldsymbol{\xi}_k = \begin{bmatrix}\nu_{h,k}\mathbf{x}_k\\ \boldsymbol{\psi}(\hat{u}_k)\end{bmatrix}, \tag{12}$$

Equ. (11) becomes

$$\tilde{e}_k = \boldsymbol{\xi}_k^{\mathrm{T}}(\mathbf{w} - \hat{\mathbf{w}}_{k-1}) + v_k. \tag{13}$$

Furthermore, (5) and (6) can be combined to

$$\hat{\mathbf{w}}_k = \hat{\mathbf{w}}_{k-1} + \begin{bmatrix}\frac{\mu_{g,k}}{\nu_{h,k}}\nu_{h,k}\mathbf{x}_k\\ \mu_{h,k}\boldsymbol{\psi}(\hat{u}_k)\end{bmatrix}\tilde{e}_k = \hat{\mathbf{w}}_{k-1} + \mathbf{L}_k\boldsymbol{\xi}_k\tilde{e}_k, \tag{14}$$

with the diagonal matrix

$$\mathbf{L}_k = \begin{bmatrix}\frac{\mu_{g,k}}{\nu_{h,k}}\mathbf{I}_{M_g\times M_g} & \mathbf{0}_{M_g\times M_h}\\ \mathbf{0}_{M_h\times M_g} & \mu_{h,k}\mathbf{I}_{M_h\times M_h}\end{bmatrix}, \tag{15}$$

where $\mathbf{I}$ denotes the identity matrix and $\mathbf{0}$ the zero matrix, the subscripts indicate the corresponding dimensions.

The adaptive system given by (13)–(15) obviously coincides with the structure of the PNLMS from Sec. 1.3, for the analogies $\mathbf{p} \leftrightarrow \mathbf{w}$, $\hat{\mathbf{p}}_k \leftrightarrow \hat{\mathbf{w}}_k$, $\tilde{f}_k \leftrightarrow \tilde{e}_k$, $\boldsymbol{\eta}_k \leftrightarrow \boldsymbol{\xi}_k$, $n_k \leftrightarrow v_k$ and $\mu\mathbf{D}_k \leftrightarrow \mathbf{L}_k$. The last analogy requires $\mathbf{L}_k$ to be not only diagonal but also positive definite. For positive step-sizes this is achieved if the nonlinearity is strictly increasing, since then the derivative in (10) satisfies $\nu_{h,k} > 0$. Specifically, if $\mathbf{h}^{\mathrm{T}}\boldsymbol{\psi}(u_k)$ represents the amplitude distortion of a PA, under the practical assumption that complete saturation does not occur, $\nu_{h,k}$ is bounded by $0 < G_{\min} \le \nu_{h,k} \le G_{\max} < \infty$, where $G_{\max} > 0$ is the finite maximum gain and $G_{\min} > 0$ is the minimum gain right below the saturation level u_{sat}.

3. STABILITY AND ROBUSTNESS

Sec. 2 shows that in the context of pre-distortion for PAs, the adaptive Wiener structure from Sec. 1.2 can be represented by a PNLMS. Therefore, several results from literature found for the PNLMS can be used to analyse the stability and the robustness of the adaptive Wiener structure.

In [10] conditions are derived under which the PNLMS from (7) behaves stable in the mean square sense for a *constant* step-size matrix $\mathbf{D}_k = \mathbf{D}$. Applying this result to the adaptive Wiener structure described by (14), with the step-size matrix given by (15), the following conditions are found for stability in the mean square sense

$$0 \le \left\{\frac{\mu_g}{\nu_h}, \mu_h\right\} < \frac{1}{\sigma_\xi^2}, \quad \frac{M_g\frac{\mu_g}{\nu_h}}{1 - \frac{\mu_g}{\nu_h}\sigma_\xi^2} + \frac{M_h\mu_h}{1 - \mu_h\sigma_\xi^2} < \frac{2}{\sigma_\xi^2}, \tag{16}$$

where the indices k are omitted to emphasise the assumption of time-independence, and where σ_ξ^2 represents the *uniform* variance of the individual elements in the vector $\boldsymbol{\xi}_k$. However, note that the validity of the conditions in (16) may be more restricted in the here considered case than for an echo canceller, which it was originally derived for in [10]. The reason for this are several assumptions which were made in [10] to make the analysis feasible. The dominating restrictions are listed below:

- Even if the excitation signal is Gaussian, the vector $\boldsymbol{\xi}_k$ will neither be Gaussian nor will its elements have uniform variance. Especially, the elements corresponding to $\boldsymbol{\psi}(\hat{u}_k)$ will have strongly differing mean powers.
- In contrast to echo cancellation, the number of estimated parameters is rather low. Hence, independently of the previous point, the approximation $\|\boldsymbol{\xi}_k\|_2^2 \approx M\sigma_\xi^2$ is very rough.
- Statistical independence between $\hat{\mathbf{w}}_{k-1}$ and $\boldsymbol{\xi}_k$ can never be achieved, even if $\mathbf{g}$ and $\hat{\mathbf{g}}_k$ in Fig. 1 are linear combiners. This is, because $\boldsymbol{\xi}_k$ directly depends on $\hat{\mathbf{g}}_{k-1}$ via $\boldsymbol{\psi}(\hat{u}_k)$.
- The step-size matrix $\mathbf{D}_k$ is assumed to be constant. In general, even for constant step-sizes $\mu_{g,k} = \mu_g$ and $\mu_{h,k} = \mu_h$, this will not be satisfied since $\nu_{h,k}$ introduces an inherent time dependency, except if the nonlinearity degenerates to a linear function.

Nevertheless, for equal and constant step-sizes $\mu_{g,k} = \mu_{h,k} = \mu$, Equ. (16) leads to the stability bound

$$\mu < \frac{4G_{\min}}{\sigma_\xi^2\left(\gamma + \sqrt{\gamma^2 - 8G_{\min}\,(2+M)}\right)} \approx \frac{2G_{\min}}{\sigma_\xi^2\,(2+M_g)}, \quad (17)$$

where $\gamma = 2 + M_g + G_{\min}\,(2 + M_h)$ was introduced for reasons of brevity. The approximation on the right-hand side is valid under the practical assumption that the minimum gain near total saturation is rather small, i.e., $G_{\min} \ll 1$.

In [15] a robustness analysis for algorithms of the form (14) is presented. Accordingly, under the condition $0 < \boldsymbol{\xi}_k^{\mathrm{T}}\mathbf{L}_k\boldsymbol{\xi}_k \leq 1$, local passivity is achieved. Meaning in loose terms that then, from one time instant k to the next $k+1$, for bounded disturbances, the estimation error will be bounded as well. An extension of this local passivity to (global) robustness requires not only bounded noise energy but also rigorous restrictions on the step-size matrix $\mathbf{L}_k$. Therefore, [15] presumes the step-size matrix to be quasi-constant, i.e., $\mathbf{L}_k = \mu_k\mathbf{L}$ with some time varying step-size μ_k. Following an alternative approach in [16], allows to show robustness for constant step-sizes $\mu_{g,k} = \mu_g$ and $\mu_{h,k} = \mu_h$; however, under the very restrictive constraint $\sum_{k=1}^{\infty}|\nu_{h,k}^{-2} - \nu_{h,k-1}^{-2}| < \infty$. Eventually, [17] shows that for time-varying $\mathbf{L}_k$ the decision on robustness depends on the actual sequence $\{\mathbf{L}_k\}$, in general non-robustness has to be expected.

Finally, an important detail for the stability of the structure in Fig. 1 has to be mentioned. Since, the output of the linear part $\hat{\mathbf{g}}_k$ is fed through the nonlinearity before the update error $\tilde{e}_k$ is determined, the nonlinearity has to fulfill a strict positive realness condition (see e.g., [18]). Accordingly, the nonlinearity is required to satisfy $\mathbf{h}^{\mathrm{T}}\boldsymbol{\psi}(\hat{u}_k)\,\mathrm{sign}(\hat{u}_k) > 0$ for any occurring input $\hat{u}_k \neq 0$, which is satisfied in the context of pre-distortion.

4. SIMULATION RESULTS

In this section, two adaptive systems are investigated by simulations. In Sec. 4.1, based on (14), a simplification of the Wiener structure in Fig. 1 is considered. The actual Wiener structure is analysed in Sec. 4.2. For all simulations, the excitation signals and the noise signals are random zero-mean white Gaussian and mutually independent. The excitation sequences (approximately) have a variance of 0.3 and the noise sequences have variance 10^{-6}. The vectors $\mathbf{g}$, $\hat{\mathbf{g}}_k$ and $\mathbf{h}$, $\hat{\mathbf{h}}_k$ have the lengths $M_g = 10$ and $M_h = 3$ respectively. The simulation results are averaged over $N_{\mathrm{avg}} = 100$ passes. All linear systems are transversal filters, and for each averaging pass randomly re-generated with independent and zero-mean Gaussian elements of variance $\sigma_g^2 = \frac{1}{M_g}$ resp. $\sigma_h^2 = \frac{1}{M_h}$ (latter is used only in Sec. 4.1). Moreover, during the simulations it was ensured that the local passivity condition ($0 < \boldsymbol{\xi}_k^{\mathrm{T}}\mathbf{L}_k\boldsymbol{\xi}_k \leq 1$) mentioned in Sec. 3 was fulfilled for each adaptation step, except from only a few sparse outliers.

4.1. Pseudo Wiener system

Using (14), we first consider a simplification of the Wiener structure in Fig. 1, which will here be referred to by pseudo Wiener structure. The simplification is made in the sense that the nonlinearities ($\mathbf{h}$, $\hat{\mathbf{h}}_k$) are replaced by simple transversal filters. The filters $\mathbf{g}$, $\hat{\mathbf{g}}_k$ are excited by the random sequence $\{x_k\}$, the filters $\mathbf{h}$, $\hat{\mathbf{h}}_k$ are both excited by the sequence $\{u_k\}$, which is independent from $\{x_k\}$. Consequently, the excitation vector $\boldsymbol{\xi}_k$ is given by

$$\boldsymbol{\xi}_k = \left[x_k, x_{k-1}, \ldots, x_{k-M_g+1}, u_k, u_{k-1}, \ldots, u_{k-M_h+1}\right]^{\mathrm{T}}. \quad (18)$$

The step-sizes are set to the same value $\mu_{g,k} = \mu_{h,k} = \mu\|\boldsymbol{\xi}_k\|_2^{-2}$, with the positive constant μ. Referring to the structure of the matrix in (15), the time-varying inverse of the first derivative $\nu_{h,k}^{-1}$ is modelled by a random sequence, uniformly distributed on the interval $[G_{\max}^{-1}, G_{\min}^{-1}] = [0.1, 1.9]$. Fig. 2 shows the results for two different values of μ and compares them to the theoretical behaviour according to [10]. Due to the strongly idealised conditions, the simulation agrees very well with theory.

As shown in [17], a system described by (14) is potentially non-robust. Hence, in a second experiment this concern was investigated for the pseudo Wiener structure. For each adaptation step, the new excitation value x_k was selected out of 100 randomly generated values, such that the norm of the parameter error vector $\tilde{\mathbf{w}}_k = \mathbf{w} - \hat{\mathbf{w}}_k$ was maximised. However, the new value u_k was simply chosen randomly. In Fig. 2, the corresponding simulation results are denoted by "worst-case". After a short period of convergence, for $\mu = 0.02$, the a priori error starts to diverge. Although less visible, the same applies to the case with $\mu = 0.002$. Thus, the pseudo Wiener structure is revealed to be non-robust.

4.2. Wiener system with Chebyshev base

Eventually, this section analyses the structure in Fig. 1. The nonlinear basis is composed by the Chebyshev polynomials [7] of the orders 1, 3 and 5. The parameter vector $\mathbf{h}$ was chosen such that the nonlinearity resembles the amplitude distortion of a Saleh model [1]. This Saleh model was parameterised in a way that complete saturation occurred for $u_k \geq u_{\mathrm{sat}} = 1$ with the corresponding saturation level at the output $z_k = z_{\mathrm{sat}} = 1$. For adaptation, the same normalised step-sizes were used as for the pseudo Wiener case. Fig. 3 presents the simulation results with simple random excitation as well as with worst-case excitation. Similarly to Sec. 4.1, the worst-case excitation was generated by a random search over 200 realisations, such that $\|\tilde{\mathbf{w}}_k\|_2$ was maximised. It can be observed that for simple random excitation, the behaviour is qualitatively similar to the results of the simplified case. However, for worst-case excitation,

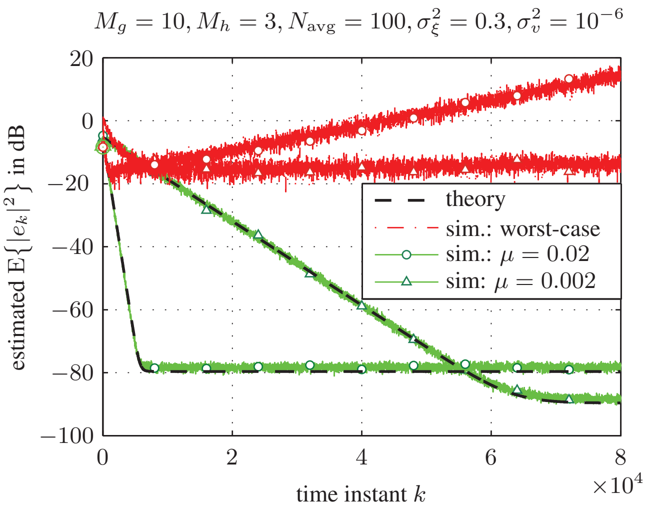

Fig. 2. Estimated mean square of the undisturbed a priori error for the pseudo Wiener system with and without worst-case excitation.

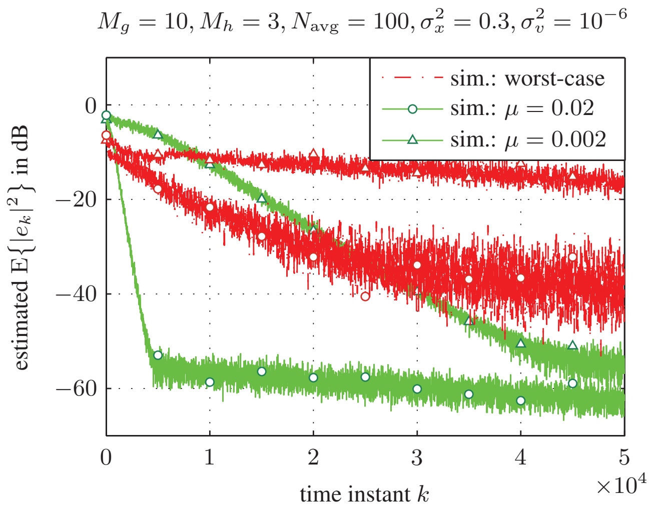

Fig. 3. Estimated mean square of the undisturbed a priori error for the Wiener system with and without worst-case excitation.

the actual Wiener structure does not diverge, it keeps converging but with reduced convergence speed.

Since the robustness is not examined exhaustively, but only by a finite number of sample sequences, the result does not lead to a strict statement. Nevertheless, taking into account that for the pseudo Wiener case the worst-case sequences could be found in spite of a smaller search space, the results allow to conjecture that for appropriate settings the Wiener structure is robust.

5. CONCLUSIONS

Analysing the stability of the here considered adaptive structure is challenging due to several reasons. (i) It is actually a combination of two adaptive algorithms. (ii) The adaptation of the nonlinearity is aggravated since the excitation signal of the nonlinear reference is not accessible. (iii) From the perspective of the linear part, the update error is distorted by a time-varying nonlinearity. By interpreting it as a PNLMS, the system is revealed to belong to a known class of algorithms which allows to resort to already available results on stability and robustness and to provide a better insight.

A more general analysis of baseband models requires to consider the complex-valued case. Since this needs to be investigated in further detail, it is deferred to future work.

6. REFERENCES

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

[2] M. Y. Cheong, E. Aschbacher, P. Brunmayr, M. Rupp, and T. Laakso, "Comparison and experimental verification of two low-complexity digital predistortion methods," in *Record of the 39th Asilomar Conference on Signals, Systems and Computers*, Pacific Grove (CA), USA, Nov. 2005, pp. 432–436.

[3] P. B. Kenington, *High linearity RF amplifier design*, Artech House, Inc., Boston (MA), USA, 2000.

[4] G. T. Zhou, H. Qian, L. Ding, and R. Raich, "On the baseband representation of a bandpass nonlinearity," *IEEE Trans. on Signal Processing*, vol. 53, no. 8, pp. 2953–2957, Aug. 2005.

[5] L. Ding and G. T. Zhou, "Effects of even-order nonlinear terms on power amplifier modeling and predistortion linearization," *IEEE Trans. on Vehicular Technology*, vol. 53, no. 1, pp. 156–162, Jan. 2004.

[6] R. Raich and G. T. Zhou, "Orthogonal polynomials for complex Gaussian processes," *IEEE Trans. on Signal Processing*, vol. 52, no. 10, pp. 2788–2797, Oct. 2004.

[7] M. Abramowitz, *Handbook of mathematical functions*, National Bureau of Standards, Washington (DC), USA, 10th edition, 1972.

[8] L. Ljung, *System Identification*, Prentice–Hall, Upper Saddle River (NJ), USA, 2nd edition, 1999.

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

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

[11] J. Benesty and S. L. Gay, "An improved PNLMS algorithm," in *Proc. of the IEEE International Conference on Acoustics, Speech, and Signal Processing*, Orlando (FL), USA, May 2002, vol. 2, pp. 1881–1884.

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

[13] F. de Souza, O. Tobias, R. Seara, and D. Morgan, "Alternative approach for computing the activation factor of the PNLMS algorithm," in *Proc. of the 17th European Signal Processing Conference*, Glasgow, Scotland, Aug. 2009.

[14] T. K. Moon and W. C. Stirling, *Mathematical methods and algorithms for signal processing*, Prentice–Hall, Inc., Upper Saddle River (NJ), USA, 2000.

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

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

[17] R. Dallinger and M. Rupp, "A strict stability limit for adaptive gradient type algorithms," in *Record of the 42nd Asilomar m on Signals, Systems and Computers*, Pacific Grove (CA), USA, Nov. 2009.

[18] V. Solo and X. Kong, *Adaptive signal processing algorithms*, Prentice–Hall, Englewood Cliffs (NJ), USA, 2nd edition, 1995.