Robust Identification of an L-N-L System

Ernst Aschbacher and Markus Rupp
Institute of Communications and Radio-Frequency Engineering
University of Technology Vienna
Gusshausstrasse 25/389, A-1040 Vienna, Austria
Email: ernst.aschbacher@tuwien.ac.at, markus.rupp@tuwien.ac.at
Web: http://www.nt.tuwien.ac.at/rapid_prototyping

Abstract—A gradient-type algorithm for the parametric identification of a nonlinear structure (L-N-L), comprising of a linear FIR filter in front of a set of nonlinear functions is investigated with respect to its robustness. The set of nonlinear functions is represented by a weighted sum of basis functions. Specifically, a feedback structure, describing the update-part of the adaptive algorithm, is developed. Using the small gain theorem, the stability of this feed-back structure and thus the robustness of the adaptive algorithm is investigated. Step-size conditions for local and global convergence are presented. As an application example, this structure is utilized to model a nonlinear power amplifier for mobile communications.

I. INTRODUCTION

Many technical systems are inherently nonlinear. An example is the power amplifier in mobile communication systems. To increase its efficiency it is driven near saturation. Due to internal effects, like parasitic capacitors, inductors, or simply due to heat flow, the system is dynamic in nature. If relatively narrow-band signals are to be amplified, like in GSM, these memory effects are not significant. For example, pre-distortion units [1], which aim to linearize the signal path, have been designed by using simple look-up table techniques. Pre-distortion is a method to avoid spurious radiation in adjacent frequency bands, required by regulatory issues. With increasing signal bandwidth, memory effects become more and more relevant [2]–[5] and a new model is required that incorporates these effects. A simple extension towards a nonlinear model incorporating memory is to place a linear FIR filter in front of the memoryless nonlinearity. Here, the nonlinear part is represented by a weighted sum of basis functions. Such a model is commonly known as Wiener model. The task for the system identification procedure is to identify the parameters, namely the linear filter weights and the weights of the basis functions. In this contribution, the robustness of a low-complexity gradient-type algorithm is investigated by commonly known methods [6]–[9].

1) Notation: All data is complex-valued; small boldface letters indicate vectors, the symbol $^H$ denotes hermitian transposition, the symbol $^T$ denotes transposition, and the symbol $^*$ stands for conjugation. All vectors are column vectors, except the parameter vectors $\mathbf{h}$ and $\mathbf{w}$, which are row vectors. The time instant $n$ is placed between parentheses for scalar signals, whereas for vectors it is placed as a subscript, e.g., $\mathbf{e}_n(n)$ and $\mathbf{u}_n$.

II. SYSTEM IDENTIFICATION

A. System Model

The following system reference model, see Fig. 1, is utilized in this contribution:

$$d(n) = \sum_{p=1}^{P} w_p^O \phi_p (\mathbf{h}^T \mathbf{u}_n) + v(n) \quad .$$

The (observed) output signal of the system is $d(n)$. The symbol $^O$ denotes the optimal parameters for the system model. The additive noise $v(n)$ includes both, measurement noise and modelling errors. The first stage of the model is a linear FIR filter, the second stage consists of a single-input multiple-output (SIMO) memoryless and parameterless nonlinear map, described via a sum of basis functions $\{\phi_p(\cdot)\}_{P=1}^{P}$. The third stage is a multiple-input single-output (MISO) linear combiner, thus leading to a so called Linear-Nonlinear-Linear (L-N-L) structure. The vector $\mathbf{u}_n$ denotes the input vector for the linear FIR filter for the time instant $n$, $\mathbf{u}_n = [u(n),\ldots,u(n-M+1)]^T$, where a filter length of $M$ is assumed. The weights of the filter are subsumed in the parameter row-vector $\mathbf{h}^O = [h_1^O,\ldots,h_M^O]$. For short notation, the outputs of the basis function-blocks are listed in a vector $\mathbf{\varphi}_n^O = [\varphi_1^O (h_1^O \mathbf{u}_n),\ldots,\varphi_P^O (h_1^O \mathbf{u}_n)]^T$, likewise the parameters $w_p^O \{P=1\}, \mathbf{w}^O = [w_1^O,\ldots,w_P^O]$. In this notation the I/O-relation of the system simply reads

$$d(n) = \mathbf{w}^O \mathbf{\varphi}_n^O + v(n) \quad .$$

This nonlinear mapping can be thought of as an extension of a basis function representation of a memoryless nonlinear function in order to incorporate memory.

B. Update Equations

Using the cost function $J(n) = \mathbb{E} [|\mathbf{e}(n)|^2]$, $\mathbb{E}[\cdot]$ denoting the expectation operator, with the disturbed error

$$\tilde{e}(n) = d(n) - \mathbf{w} \varphi_n = d(n) - y(n) \quad ,$$

in which $\mathbf{w} = [w_1,\ldots,w_P]$ and $\varphi_n^T = [\varphi_1(h_1 \mathbf{u}_n),\ldots,\varphi_P(h_1 \mathbf{u}_n)]$ denote estimates for the parameter vector $\mathbf{w}^O$ and the regression vector $\varphi_n^O$, respectively, the update equations for the parameter vectors using a steepest
After evaluating the derivatives and simplification of the expectation operator this leads to the following gradient-type algorithm

\[
\begin{align*}
\mathbf{w}_n &= \mathbf{w}_{n-1} + \mu_w(n) \left( -\frac{\partial}{\partial \mathbf{w}} J(n) \right)^H \bigg|_{\mathbf{w}_{n-1}, \mathbf{h}_{n-1}} \quad \text{(4)} \\
\mathbf{h}_n &= \mathbf{h}_{n-1} + \mu_h(n) \left( -\frac{\partial}{\partial \mathbf{h}} J(n) \right)^H \bigg|_{\mathbf{w}_{n-1}, \mathbf{h}_{n-1}} \quad \text{(5)}
\end{align*}
\]

where \( \varphi_{n-1} = [\varphi_1(\mathbf{h}_{n-1} \mathbf{u}_n), \ldots, \varphi_p(\mathbf{h}_{n-1} \mathbf{u}_n)]^T \), \( \varphi'_{n-1} = [\varphi'_1(\mathbf{h}_{n-1} \mathbf{u}_n), \ldots, \varphi'_p(\mathbf{h}_{n-1} \mathbf{u}_n)]^T \). Here, \( \gamma \) stands for the first derivative of the respective function, \( \varphi'(x) = \partial_x \varphi(x) \). The signal \( \tilde{e}_a(n) = d(n) - \mathbf{w}_{n-1} \varphi_{n-1} \) denotes the disturbed a-priori error. The term \( \varphi_{n-1}^H \mathbf{w}_{n-1}^H \) is a scalar, modulating the step-size \( \mu_h(n) \). It is assimilated further on in the modified step-size

\[
\tilde{\mu}_h(n) = \mu_h(n) \varphi_{n-1}^H \mathbf{w}_{n-1}^H \quad \text{(7)}
\]

The undisturbed a-priori error \( e_a(n) \) can be decomposed in the following way:

\[
\begin{align*}
e_a(n) &= \mathbf{w}^o \varphi_n^o - \mathbf{w}_{n-1} \varphi_{n-1} = \\
&= \mathbf{w}_{n-1} \varphi_{n-1} + \mathbf{w}^o (\varphi_n^o - \varphi_{n-1}) \quad \text{(8)}
\end{align*}
\]

where the parameter-error vector \( \tilde{\mathbf{w}}_{n-1} = \mathbf{w}^o - \mathbf{w}_{n-1} \) is introduced. Further, expanding

\[
\mathbf{w}^o (\varphi_n^o - \varphi_{n-1}) = e_{a,h}(n) \xi(n) \quad \text{(9)}
\]

with \( \xi(n) = \frac{\mathbf{w}^o (\varphi_n^o - \varphi_{n-1})}{e_{a,h}(n)} = \frac{e_a(n) - e_{a,w}(n)}{e_{a,h}(n)} \) leads to the expression

\[
\tilde{e}_a(n) = e(n) + e_{a,w}(n) + \xi(n) e_{a,h}(n) \quad \text{(10)}
\]

for the disturbed a-priori error. Unfortunately, \( \xi(n) \) cannot be observed, since \( \mathbf{w}^o, \varphi_n^o \) and \( e_{a,h}(n) \) are not accessible.

\[Fig. 1. System model\]
where $\gamma(n) = \sqrt{\frac{\mu_w(n)}{\mu_h(n)}}$ and $\beta_h(n) = \xi(n)\alpha_h(n)$, are introduced. The complete local system equations (cf. (18),(16) and (17)) read now

$$\begin{bmatrix} \hat{\theta}_{n,1} \\ \vdots \\ \hat{\theta}_{n,N} \end{bmatrix} = \begin{bmatrix} T_n^{-1} \\ \vdots \\ T_n^{-1} \end{bmatrix} \begin{bmatrix} \hat{\theta}_{n-1,1} \\ \vdots \\ \hat{\theta}_{n-1,N} \end{bmatrix} \text{ forward path}$$

$$v_n = A_n v_n - G_n e_{a,n} \text{ feedback path}$$

revealing the feedback structure of the algorithm, as shown in Fig. 2.

![Feedback structure of the adaptive algorithm](image1)

The two lossless mappings in (18) can now compactly be written using the vectors in (19):

$$\frac{\|\hat{\theta}_n\|^2 + \|e_{n,n}\|^2}{\|\hat{\theta}_{n-1}\|^2 + \|v_n\|^2} = 1 .$$

(22)

Using the small gain theorem, cf. e.g. [11], [12], local stability is guaranteed if the product of the gains of the feed-forward and the feedback system is strictly less than one. Since the map $T_n$ is lossless, $\|T_n\|_2 = 1$, local stability can be guaranteed if $\|G_n\|_2 < 1$. Evaluating the $L_2$-norm of the feedback matrix $G_n$, $\|G_n\|_2 = \sqrt{\max_{\lambda \in \lambda_n(G_n^T G_n)}}$, $\max_{\lambda \in \lambda_n}$ being the maximum eigenvalue of the matrix in parentheses, it can be shown, that if $0 < \alpha_w(n), \beta_h(n) < 1$, both eigenvalues are less than one in magnitude, guaranteeing local stability, see Fig. 3 and Fig. 4.

![First eigenvalue of $G_n^T G_n$ if $0 < \alpha_w(n), \beta_h(n) < 1$.](image2)

![Second eigenvalue of $G_n^T G_n$ if $0 < \alpha_w(n), \beta_h(n) < 1$.](image3)

Unfortunately, $\beta_h(n)$ cannot be observed, since the signal $\xi(n)$ is not accessible, and therefore hardly set to the optimal value, although it can be controlled via the parameter $\alpha_h(n) = \frac{\mu_h(n)}{\mu_h(n)}$ in which the step-size $\mu_h(n)$ is freely adjustable.

**D. Global Passivity Relation**

So far only local stability of the adaptive algorithm has been analyzed. For global stability the behavior of the algorithm over the time horizon $n = 1, \ldots, N$ is to investigate. Numerator and denominator in (22) are summed over the entire time horizon, yielding

$$\frac{\|\hat{\theta}_N\|^2 + \sum_{n=1}^N \|e_{n,n}\|^2}{\|\hat{\theta}_0\|^2 + \sum_{n=1}^N \|v_n\|^2} = 1 .$$

(23)

By stacking the vectors and matrices in (19) and (20) for $n = 1, \ldots, N$

$$\bar{e}_n = \begin{bmatrix} e_{a,1}^T \\ \vdots \\ e_{a,N}^T \end{bmatrix}^T$$

$$\bar{v} = \begin{bmatrix} v_1^T \\ \vdots \\ v_N^T \end{bmatrix}^T$$

$$\bar{\theta} = \begin{bmatrix} \theta_1^T \\ \vdots \\ \theta_N^T \end{bmatrix}^T$$

$$A = \text{ diag } \{ A_1, \ldots, A_N \}$$

$$G = \text{ diag } \{ G_1, \ldots, G_N \}$$

the global system equations can be written in the form

$$\begin{bmatrix} \bar{\theta}_N \\ \bar{e}_n \end{bmatrix} = \begin{bmatrix} T \bar{\theta}_0 \\ \bar{v} \end{bmatrix}$$

(24)

$$\bar{v} = A \bar{v} - G \bar{e}_n .$$

(25)
Since the forward path is lossless, \(|\|T\|_2 = 1\), cf. (23), stability can be assured if the feedback system is passive, \(|\|G\|_2 < 1\). Rewriting (23) yields

\[
\sum_{n=1}^{N} |\tilde{e}_{a,n}|^2 = |\tilde{\theta}_0|^2 + |\tilde{\theta}_N|^2 + \sum_{n=1}^{N} |\tilde{v}_n|^2 \leq \sum_{n=1}^{N} |\tilde{e}_n|^2 + \sum_{n=1}^{N} |\tilde{v}_n|^2 \leq \delta \|e\|^2_2 \quad \text{(26)}
\]

The term \(|\|v_n\|_2^2\), see (21), can be upper bounded using the triangle inequality and the definition of norms by

\[
|\|v_n\|_2^2 = |\|A_n v_n - G_n e_{a,n}\|_2^2 \leq |\|A_n v_n\|_2^2 + |\|G_n e_{a,n}\|_2^2 \leq |\|A_n\|_2^2 |\|v_n\|_2^2 + |\|G_n\|_2^2 |\|e_{a,n}\|_2^2 |\|v_n\|_2^2 \leq \max_{n=1,\ldots,N} (|\|A_n\|_2^2 |\|v_n\|_2^2 + \max_{n=1,\ldots,N} (|\|G_n\|_2^2 |\|e_{a,n}\|_2^2)
\]

\[
\Delta_N \triangleq \|v_n\|_2^2 \quad \text{and} \quad \|\tilde{e}_n\|_2^2 \quad \text{the} \quad \delta \|e\|^2_2 \quad \text{for} \quad n = 1, \ldots, N \quad \text{is required. Upper bounding in (26) and several times in (27) leads naturally to more and} \quad \text{more conservative bounds for the step-sizes} \quad \mu_h(n) \quad \text{and} \quad \mu_w(n) \quad \text{in order to fulfill} \quad \Delta_N < 1.
\]

### III. APPLICATION EXAMPLES

#### A. Example 1

In this example, a system, consisting of a linear FIR filter of length four with the tap-weights \(h^o = [0.93, 0.34, 0.12, 0.05]\) and a nonlinear function, consisting of the first three uneven Hermite polynomials \(H_1(\cdot), H_3(\cdot), H_5(\cdot)\), see e.g. [13], with the weights \(w^o = [1, 2, 7, 7, 4]\) is to identify by algorithm (6). The system of this example is real-valued. The exciting signal is a two-tone signal with frequencies \(\Omega_1 = 0.1, \Omega_2 = 0.2\). The amplitudes of the two cosine-signals are identical, \(A_1 = A_2 = 0.5\). The phases of the two signals are independent random variables, equally distributed in \((-\pi, \pi]\) and constant for each simulation run. Fifty simulation runs are performed. The output of the system is disturbed by independent, additive, white, gaussian noise, with variance \(\sigma_e^2 = 10^{-2}\). Four cases with different, fixed step-sizes are simulated, and the signals \(\alpha_w(n)\) and \(\beta_h(n)\) are observed. In a real setting with an unknown system, the signal \(\beta_h(n)\) is not observable. The initial guess in all cases is \(h_0 = [1, 0, 0, 0]\) and \(w_0 = [1, 0, 0]\).

---

**Table I**

<table>
<thead>
<tr>
<th>Case</th>
<th>(\mu_h)</th>
<th>(\mu_w)</th>
<th>(\lambda_{1, \text{max}})</th>
<th>(\lambda_{1, \text{min}})</th>
<th>(\lambda_{2, \text{max}})</th>
<th>(\lambda_{2, \text{min}})</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>(10^{-7})</td>
<td>(10^{-6})</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>B</td>
<td>(10^{-8})</td>
<td>(10^{-7})</td>
<td>2.0 \times 10^{-5}</td>
<td>2.0 \times 10^{-5}</td>
<td>2.0 \times 10^{-5}</td>
<td>2.0 \times 10^{-5}</td>
</tr>
<tr>
<td>C</td>
<td>(3\times 10^{-6})</td>
<td>(3\times 10^{-6})</td>
<td>3.0 \times 10^{-6}</td>
<td>3.0 \times 10^{-6}</td>
<td>3.0 \times 10^{-6}</td>
<td>3.0 \times 10^{-6}</td>
</tr>
<tr>
<td>D</td>
<td>(5\times 10^{-6})</td>
<td>(5\times 10^{-6})</td>
<td>5.0 \times 10^{-6}</td>
<td>5.0 \times 10^{-6}</td>
<td>5.0 \times 10^{-6}</td>
<td>5.0 \times 10^{-6}</td>
</tr>
</tbody>
</table>

---

**Fig. 6.** Learning curves (rel. system mismatch) for case A

The best result is achieved in case A, see Fig. 6 and Table I, where the relative system mismatch \(m_{rel,h} = |\|h_n\|_2^2 / |\|h^o\|_2^2\| = 1\) (solid line), and \(m_{rel,w} = |\|w_n\|_2^2 / |\|w^o\|_2^2\| = 1\) (dashed line) are plotted against the iterations. The signals \(\alpha_w(n), \beta_h(n)\) are in a very small region around zero, meaning that both eigenvalues are very close to one. This behavior is similar to the linear system case where small step-sizes lead to small errors.

---

**Fig. 7.** Learning curves (rel. system mismatch) for case B

Increasing the step-sizes leads to worse results, the signals \(\alpha_w(n), \beta_h(n)\) oscillate in a larger region, and so do the eigenvalues \(\lambda_1, \lambda_2\) of the feedback matrix \(G_n\). Since the eigenvalues becomes larger than one, the system becomes more and more unstable, cf. Fig. 7 where the relative misalignment for the vector \(w\) is slightly increasing. Results worsen further in cases C and D, cf. Figs. 8 and 9, where the signals \(\alpha_w(n), \beta_h(n)\) leave the stability region \((0, 1)\) for \(\alpha_w(n)\) and \(\beta_h(n)\), cf. also the Figs. 3 and 4. Practically no system identification...
is performed.

![Fig. 8. Learning curves (rel. system mismatch) for case C](image)

![Fig. 9. Learning curves (rel. system mismatch) for case D](image)

**B. Example 2**

In this example measured output-data from a power amplifier is used as the desired signal $d(n)$. The exciting signal $u(n)$ is a multi-tone signal with bandwidth 2MHz. The tone separation is 2kHz, each tone has a random phase, drawn from a uniform distribution $U(-\pi, \pi)$. The nonlinear model is

$$y(n) = \sum_{p=1,3,5,7} \left( w_p H_p(hu_n) + w'_p H'_p(hu_n) \right), \quad (29)$$

the linear filter $h$ being of length four, $H_p(\cdot)$ denoting the Hermite polynomial of order $p$. The system parameter $\{w_p, w'_p\}_{p=1}^{P}$ and $h$ are complex valued. In Fig. 10 the squared a-priori error over the number of iterations is shown. Three different step-sizes $\mu_h = \mu_w$ are chosen. The step-size $\mu_h = \mu_w = 10^{-4}$ showed best behavior. Increasing the step-sizes only slightly to $\mu_h = \mu_w = 1.45 \cdot 10^{-4}$ results in an unstable system. Decreasing the step-sizes, a stable system is obtained, although with a poorer behavior - the convergence is slower, but also the steady-state error is larger. The closer to the instability bound the step-sizes are chosen, the better the identification is. In this case, the eigenvalues of the feedback matrix $G_b$ are both close to one, and, consistent with Section III-A, this showed optimal results.

**IV. CONCLUDING REMARKS**

A gradient type algorithm for the adaptive identification of an L-N-L system is investigated with respect to robustness. This structure results from a simple extension of a basis function representation for a nonlinear function by placing a linear FIR filter in front of such basis function representation. An equivalent feedback structure for the update part of the adaptive algorithm is derived. This feedback structure can be analyzed with the small-gain theorem, from which step-size conditions for local stability can be devised. Further, stability for the time horizon $n = 1, \ldots, N$, is investigated, allowing to find step-size conditions for a globally stable system.

**ACKNOWLEDGMENT**

The authors would like to thank Holger Arthaber and Prof. Gottfried Magerl for their kind support performing the measurements.

**REFERENCES**


