# Identification of a Nonlinear Power-Amplifier L-N-L Structure for Pre-Distortion Purposes

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*— **Based on power amplifier measurements a low-complex nonlinear model with memory, consisting of a memoryless, but parameterized nonlinearity, sandwiched between two linear FIR-filter is developed (L-N-L structure). Although the signal bandwidth is with 2MHz rather small, memory effects can be noticed. Furthermore, an adaptive algorithm for continuous identification of the model parameters is developed. Simulations show that the algorithm converges in the mean-square sense to the global minimum, given only little knowledge regarding the region where this minimum lies.**

## I. Introduction

In order to linearize the signal path from origin to radiation as required to comply with regulatory issues, for example, spurious radiation in adjacent frequency bands, and not to degrade system performance, digital pre-distortion can be used to compensate for the nonlinearity introduced by the power-amplifier. Due to such power amplifiers and the need to utilize power efficient modes performance loss can be recognized in form of increased bit-error rates. Digital pre-distortion is already in use for GSM where only small bandwidths are to be compensated utilizing simple nonlinear functions realized in form of look-up tables. Such pre-distortion scheme aims to remove the unwanted nonlinear effects at the place of their origin, namely at the transmitter. Due to the large radiated power in the first stage, pre-distortion will be and is installed in base-stations, where additional complexity can be tolerated. The method relies entirely in the digital baseband or low intermediate-frequency (IF) domain, thus becoming a flexible scheme which can also easily upgrade existing transmitters.

For the development of a pre-distortion scheme, a model of the nonlinear system is required, i.e., the power-amplifier with its additional circuitry (mixers, matching-networks). In modern communication systems like UMTS not only the nonlinearity of the power amplifier, which is driven near saturation due to efficiency reasons, becomes a problem, but also memory effects due to the broadband nature of the signals turn up, see [1]. Nonlinear system models without memory exist, e.g. the often used Saleh-model [2] or can be build up more or less easily using e.g. Taylor-series, or orthogonal function-expansions, which are more parsimonious with the parameter usage than Taylor-series. Pre-distortion schemes for memoryless nonlinearities are relatively simple, e.g. by the usage of a look-up table, see e.g. [3]. Incorporating memory effects becomes a task, look-up tables become very large and unmanageable.

A prominent model for nonlinear systems with memory is the Volterra-Series which can approximate a huge class of non-linear systems with mild constraints [4], [5]. The Volterra-series requires a large amount of parameters if the nonlinearity and/or the memory effects are pronounced. The advantage of the Volterra-series lies in the fact that, since the parameters enter linearly in the description, quadratic cost-functions for parameter identification have a unique global minimum. Therefore an adaptive scheme for the parameter-identification will yield good parameter estimates (assuming convergence and low noise). Another possibility are the usage of neural-nets for identification and/or control of dynamical nonlinear systems, [6].

In this contribution a very particular, heuristic structure with low-complexity is considered, namely a memoryless, but parameterized nonlinearity, between linear FIR-filters of variable lengths. Most of the parameters enter the description nonlinearly, causing problems in iterative methods to find the global minimum of a cost function. Simulations show that the derived gradient-type algorithm converges to the global minimum, when a proper initial value for the parameter and a sufficiently small iteration step-size is chosen.

## II. Measurement Procedure

In order to obtain I/O-data to test different structures, measurements on a standard power-amplifier (Minicircuits, Type: ZLH-42W) were performed. Structures of interest are specializations of the L-N-L - sandwich structure, i.e., a pure N, L-N, and N-L system with variable filter-lengths. The chosen test-signal was a multi-tone signal of 2MHz, whereby the individual tones where separated by 20kHz. The phases are uniformly distributed in the interval $[-\pi, \pi)$. The measurement-setup is schematically shown in Fig. 1. Using the Rhode&Schwarz-I/Q-Modulation generator AMIQ I-Q-signals were generated which were up-converted to 2GHz with the Rhode&Schwarz Vector Signal Generator SMIQ. A Power Spectrum Analyzer (Agilent, PSA) down-converted, demodulated and sampled the signal with the (internally chosen) sampling-rate of 10.24MHz. Due to the nonlinearity of the power-amplifier large spurious parts outside of the 2MHz-band could be observed. At the output a total signal-bandwidth of 8MHz was measured.

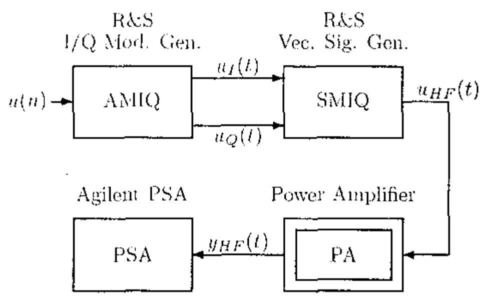


Fig. 1. Measurement setup

## III. Fitting Measurements to Models

The generated I/O-data is the basis for the data-fitting step. Using a L-N-L structure, N memoryless but parameterized, the I/O-mapping can be described by

$$\begin{aligned} y(n) &= L_2 N L_1 u(n) \\ &= L_2 f(\boldsymbol{\theta}, L_1 u(n)) \;, n = 1, 2, \ldots, N \;, \end{aligned} \tag{1}$$

whereby $L_{1,2}$ stands for linear operators (FIR filters with $M_1$ and $M_2$ coefficients, respectively) and $f(\boldsymbol{\theta}, x)$ is a nonlinear function with parameters grouped into the vector $\boldsymbol{\theta}$. The utilized nonlinear memoryless function is defined by

$$f(\theta, x) = \frac{x}{1 + \theta |x|^2} \tag{2}$$

which, for small $|x|$ can well be approximated by $f(\theta, x) \approx x$. For large arguments $|x|$, $f(\theta, x)$ does not behave linear any more but instead turns into a compression mode. There is only one complex-valued parameter $\theta$ describing the nonlinear function. Due to the simplicity of the model the parameter vector $\boldsymbol{\theta}$ in (1) is reduced to a scalar $\theta$. Applying a nonlinear data-fitting procedure (Matlab®'s procedure *lsqnonlin*()), the optimal parameters can be found as well as the corresponding remaining error[1]. The achieved mean-square error

$$\varepsilon_{\mathrm{LNL}} = \frac{1}{N} \sum_{n=1}^{N} |y(n) - \hat{y}_{\mathrm{LNL}}(n)|^2 \tag{3}$$

of the L-N-L and correspondingly the errors of the L-N and the N-L system are compared relative to the the mean-square error

$$\varepsilon_{\mathrm{N}} = \frac{1}{N} \sum_{n=1}^{N} |y(n) - \hat{y}_{\mathrm{N}}(n)|^2 \tag{4}$$

[1]Note that this procedure does not guarantee to find the global minimum.

of a pure nonlinear system without memory. The improvement

$$\eta_{MSE} = \varepsilon_{\mathrm{LNL}} / \varepsilon_{\mathrm{N}} \tag{5}$$

by increasing the filter-lengths of the input and output filter (for L-N-L both simultaneously at each step) is reported in Fig. 2. A significant improvement is noticed only for small numbers of filter parameters. For one additional filter tap in the N-L case the mean-square error is reduced by approx. 3dB, by approx. 3.5dB in case of a L-N or a L-N-L structure. Observe that one tap corresponds to a time interval of 1/10.24MHz ≈ 100ns. In a faster sampled system (with higher signal bandwidth) there would be more taps necessary for an equivalent improvement. As Fig. 2 points out, not very much is gained by using an L-N-L structure, which requires $M_1$ resp. $M_2$ parameters more than the L-N, resp. the N-L structure. Probably this is due to the small signal bandwidth of only 2MHz. Using e.g. a L-N structure with 3 taps would, in this case, be the best choice. Note that a Volterra series of order $P$ with memory length $M$ requires $O(M^P)$ coefficients (without reductions due to symmetry) while the L-N-L structure is only of order $O(M)$.

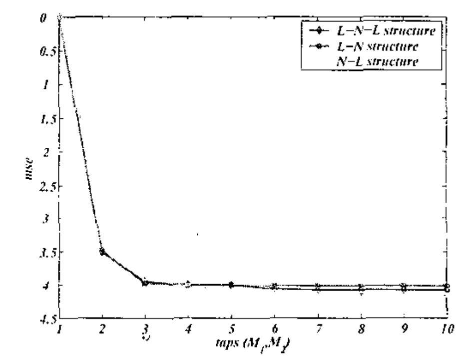


Fig. 2. Improvement of the MSE

In Table I the parameters for the linear filters of the L-N configuration for filter-length four are denoted. In Table II the identified parameters for the N-L structure are shown, while in Table III the parameters for the L-N-L-structure are listed.

TABLE I
Parameters for the L-N structure

| $\mathbf{w}_1$ | $\boldsymbol{\theta}$ |
|---|---|
| $w_{1,1} = 15.712 + j2.8797$ | $\theta = 0.012112 - j0.0040354$ |
| $w_{1,2} = -5.8875 - j1.1768$ | |
| $w_{1,3} = 1.4634 + j0.26428$ | |
| $w_{1,4} = 0.20648 + j0.044805$ | |

## IV. Adaptive Identification of the System

In this section an adaptive algorithm for the continuous identification of the parameters of the L-N-L reference

TABLE II
PARAMETERS FOR THE N-L STRUCTURE

| $\mathbf{w}_2$ | $\theta$ |
|---|---|
| $w_{2,1} = 16.323 + j3.117$ | $\theta = 1.5862 - j0.52638$ |
| $w_{2,2} = -8.1101 - j1.8619$ | |
| $w_{2,3} = 3.9151 + j0.9736$ | |
| $w_{2,4} = -0.71243 - j0.20321$ | |

TABLE III
PARAMETERS FOR THE L-N-L STRUCTURE

| $\mathbf{w}_1$ | $\mathbf{w}_2$ |
|---|---|
| $w_{1,1} = 3.6187 + j0.23288$ | $w_{2,1} = 4.2624 + j0.55122$ |
| $w_{1,2} = -1.0003 - j0.26903$ | $w_{2,2} = -0.28874 + j0.13464$ |
| $w_{1,3} = 0.034828 - j0.10397$ | $w_{2,3} = 0.21805 + j0.25229$ |
| $w_{1,4} = 0.017799 - j0.0037678$ | $w_{2,4} = 0.052459 + j0.072556$ |
| $\theta$ | |
| $\theta = 0.2228 - j0.061253$ | |

structure is developed, see Fig. 3. The task for the adaptive algorithm is to identify the parameters of the reference structure with a small error. Due to its low complexity, a gradient-type algorithm, see e.g. [7], is preferred in a practical scheme and is proposed and developed here.

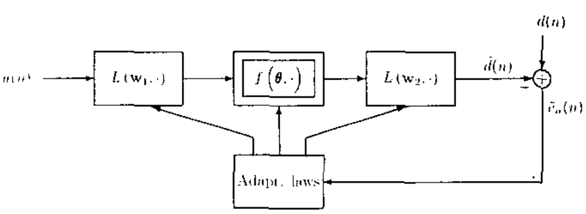


Fig. 3. Adaptive L-N-L structure

The output of the reference L-N-L system is

$$d(n) = \underbrace{L_2 N L_1 u(n)}_{=y(n)} + v(n) \quad . \tag{6}$$

Assume now random processes as input- ($u(n)$), output- ($y(n)$) and noise- ($v(n)$) signals. For shorthand notation the signals will be subsumed in vectors, e.g.

$$\mathbf{u}_n = [u(n), u(n-1), \dots, u(n-M_1+1)]^T, \tag{7}$$

whereby $M_1$ denotes the memory-length of the first linear FIR-filter. The weight-vectors of the linear filters of the reference system are $\mathbf{w}_i = [w_i(1), w_i(2), \dots, w_i(M_i)]^T, i = 1, 2$, while their estimates will be denoted by $\hat{\mathbf{w}}_1$, and $\hat{\mathbf{w}}_2$, respectively. Using this vector notation the estimate of (6) can be written in the form

$$\hat{d}(n) = \mathbf{f}^T\left(\hat{\theta}, \mathbf{u}_n^T \hat{\mathbf{w}}_1\right) \hat{\mathbf{w}}_2 \quad , \tag{8}$$

with

$$\mathbf{f}^T\left(\hat{\theta}, \mathbf{u}_n^T \hat{\mathbf{w}}_1\right) = \left[f\left(\hat{\theta}, \mathbf{u}_n^T \hat{\mathbf{w}}_1\right), \dots, f\left(\hat{\theta}, \mathbf{u}_{n-M_2+1}^T \hat{\mathbf{w}}_1\right)\right] \quad . \tag{9}$$

The cost-function to be minimized by the adaptive algorithm is

$$J\left(\hat{\mathbf{w}}_1, \hat{\mathbf{w}}_2, \hat{\theta}\right) = E\left[\left|d(n) - \hat{d}(n)\right|^2\right] \quad . \tag{10}$$

Thus, the cost function depends on three vectors (one being reduced to a scalar), one for each subsystem.

Applying the steepest-descent method for each parameter vector and simplifying the expectation-operator, a gradient-type algorithm is developed. The update part of the algorithm for the three parameter vectors is given by

$$\hat{\mathbf{w}}_{1,n} = \hat{\mathbf{w}}_{1,n-1} + \mu_{w_1} \tilde{e}_a(n) \left(-\frac{\partial}{\partial \hat{\mathbf{w}}_1} \hat{d}(n)\right)^H \tag{11}$$

$$\hat{\mathbf{w}}_{2,n} = \hat{\mathbf{w}}_{2,n-1} + \mu_{w_2} \tilde{e}_a(n) \left(-\frac{\partial}{\partial \hat{\mathbf{w}}_2} \hat{d}(n)\right)^H \tag{12}$$

$$\hat{\theta}_n = \hat{\theta}_{n-1} + \mu_\theta \tilde{e}_a(n) \left(-\frac{\partial}{\partial \hat{\theta}} \hat{d}(n)\right)^H \tag{13}$$

$$\tilde{e}_a(n) = d(n) - \mathbf{f}^T\left(\hat{\theta}_{n-1}, \mathbf{u}_n^T \hat{\mathbf{w}}_{1,n-1}\right) \hat{\mathbf{w}}_{2,n-1} \quad , \tag{14}$$

where $\hat{\boldsymbol{\theta}}_{n-1}, rv\hat{e}cw_{i,n-1}, i = 1, 2$ denotes the estimates at iteration-step $n-1$. The derivatives are

$$\left(-\frac{\partial}{\partial \hat{\mathbf{w}}_1} \hat{d}(n)\right)^H = \mathbf{A}_{n-1}^H \hat{\mathbf{w}}_{2,n-1}^* \tag{15}$$

$$\left(-\frac{\partial}{\partial \hat{\mathbf{w}}_2} \hat{d}(n)\right)^H = \mathbf{f}^*\left(\hat{\theta}_{n-1}, \mathbf{u}_n^T \hat{\mathbf{w}}_{1,n-1}\right) \tag{16}$$

$$\left(-\frac{\partial}{\partial \hat{\theta}} \hat{d}(n)\right)^H = \mathbf{b}_{n-1}^H \hat{\mathbf{w}}_{2,n-1}^* \quad . \tag{17}$$

where the $M_2 \times M_1$ dimensional matrix $\mathbf{A}_{n-1}$

$$\mathbf{A}_{n-1} = \begin{pmatrix} \mathbf{u}_n^T f_x\left(\hat{\theta}_{n-1}, \hat{x}_n\right) \\ \mathbf{u}_{n-1}^T f_x\left(\hat{\theta}_{n-1}, \hat{x}_{n-1}\right) \\ \vdots \\ \mathbf{u}_{n-M_2+1}^T f_x\left(\hat{\theta}_{n-1}, \hat{x}_{n-M_2+1}\right) \end{pmatrix} \tag{18}$$

and the $M_2 \times 1$ vector

$$\mathbf{b}_{n-1} = \begin{bmatrix} f_\theta\left(\hat{\theta}_{n-1}, \hat{x}_n\right) \\ f_\theta\left(\hat{\theta}_{n-1}, \hat{x}_{n-1}\right) \\ \vdots \\ f_\theta\left(\hat{\theta}_{n-1}, \hat{x}_{n-M_2+1}\right) \end{bmatrix} \quad , \tag{19}$$

with $\hat{x}_n = \mathbf{u}_n^T \hat{\mathbf{w}}_{1,n-1}$ and $f_x(\dots x_i, \dots) = \frac{\partial}{\partial x} f(\dots, x, \dots)|_{x=x_i}$.

The derivatives of $f(\theta, x)$ with respect to $\theta$, resp. $x$, are

$$f_x(\theta, x) = \frac{1}{\left(1 + \theta |x|^2\right)^2} \tag{20}$$

$$f_\theta(\theta, x) = -f^2(\theta, x) x^* . \tag{21}$$

Since the parameter-vectors $\hat{\theta}$ and $\hat{\mathbf{w}}_1$ enter in a nonlinear fashion in the system-equation (8), the cost-function can be expected to exhibit local minima. Without prior knowledge regarding the region in which the global minimum lies, the adaptive algorithm is in general not expected to converge to the global minimum. Therefore a smart choice of the starting values $\mathbf{w}_{1,0}$ and $\theta_0$ for the adaptive algorithm is essential.

### *A. Special cases: L-N, N-L - models*

By setting either $L_2$ or $L_1$ equal to the identity operator (11) to (19) can be specialized to an adaptive L-N resp. an adaptive N-L configuration. Such systems are known in literature under the name Wiener- (L-N) and Hammerstein-model (N-L). In the following, simulation results will be presented for all three nonlinear adaptive systems with memory, the L-N, the N-L, as well as the L-N-L structure.

## V. Simulation Results

In order to gain insight in the learning behavior of the derived algorithm, simulations were performed since an analytical treatment regarding convergence seems not easily feasible due to the occurrence of a non-linear subsystem. In order to adapt to a global minimum a smart choice for the initial values of the parameter-vectors $\hat{\theta}_0$ and $\hat{\mathbf{w}}_1$ must be available. The initial guess for the parameter vectors which is used in all three cases described below is $\mathbf{w}_{i,0} = [1+j, 0, 0, \ldots, 0]^T, i = 1, 2.$ $\hat{\theta}_0 = 0$. By starting with initial values reflecting a purely linear system without memory (the first tap of the linear filters equal $1+j$), but not knowing even an approximate value, it turns out that all three adaptive structures find the global minimum.

In the following, the parameters from Tables I to III were used for the reference structures.

### *A. Convergence behavior of the L-N-L structure*

The convergence in the mean-square sense was simulated by performing $R = 50$ runs and averaging the resultant squared a-priori error $|\tilde{e}_a(n,r)|^2$ over the runs $r$ ($n$ denoting the iteration step), resulting in the mean-squared error $mse(n) = \frac{1}{R}\sum_{r=1}^{R} |\tilde{e}_a(n,r)|^2$. The system was fed with a zero-mean complex white gaussian input signal $u(n)$, both real and imaginary part of unit variance. The variance of the added noise $v(n)$ at the output of the system to identify was set to $10^{-4}$ for the real and the imaginary part, respectively.

The step-sizes were chosen heuristically in order to achieve convergence and are $\mu_{w_1} = 10^{-2}, \mu_{w_2} = 10^{-2}$ and $\mu_\theta = 5\cdot 10^{-3}$. In Fig. 4 the mean-squared error of the adaptive L-N-L-structure is depicted. Up to now no limits are known for the step-sizes in order to achieve convergence, nor optimal values for fastest convergence available. Convergence to an undesired local minimum was observed when no prior knowledge was used, e.g. when setting the initial value of the parameters equal to zero. Furthermore, a large uncertainty regarding the convergence rate is noticed. Even with a lot more simulation runs (1000 runs were performed), no "smoother" curve could be achieved.

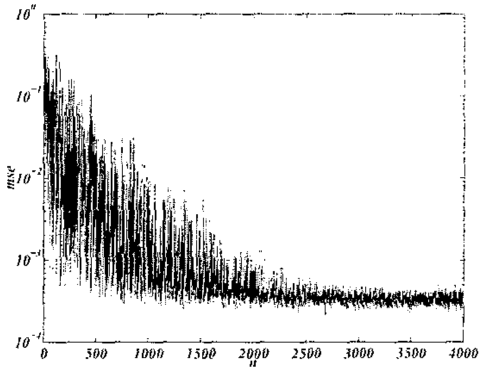


Fig. 4. Learning behavior of the adaptive L-N-L structure

### *B. Convergence behavior of the L-N structure*

In Fig. 5 the learning behavior of an L-N structure is depicted. The chosen step-sizes where $\mu_{w_1} = 0.4$ for the linear part and $\mu_\theta = 10^{-5}$ for the nonlinear part. The added noise $v(n)$ at the output was, like in the previous L-N-L case, drawn from $\mathcal{N}(0, 10^{-4})$ for real and imaginary part, respectively. Also in this configuration, the convergence rate differs significantly from realization to realization.

In contrast to the L-N-L structure, convergence to the optimal parameters was observed even with no "smart" choices for the parameters (zero-vectors), but the convergence was slower (approx. $5\cdot 10^3$ iterations), using the same step-sizes. In this special case of an L-N structure the zero initial condition seems to be "smart" enough to achieve good performance.

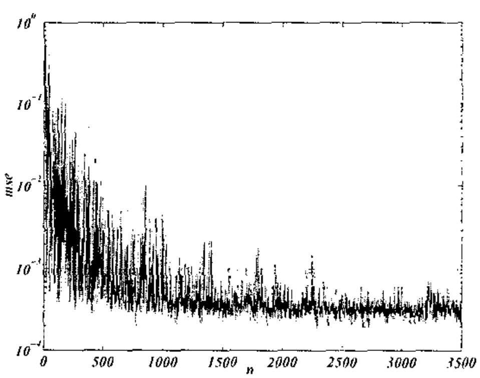


Fig. 5. Learning behavior of the adaptive L-N structure

### *C. Convergence behavior of the N-L structure*

Also in this case, using the "smart" initial guess for the parameters, the algorithm converged to the global minimum, thus being able to identify the nonlinear power-amplifier model. The convergence behavior is very similar to the former described two cases, with rather large uncertainty regarding the convergence rate. The step-sizes used in this case are

$\mu_\theta = 10^{-2}$ and $\mu_{w_2} = 0.4$ and were selected heuristically like in the previous cases. The input-signal and noise parameters are identical to the former cases.

Using zero vectors as an initial guess the LMS-algorithm converged to the global minimum, but, using the same step-sizes, a more iteration steps (about $3 \cdot 10^3$) were required.

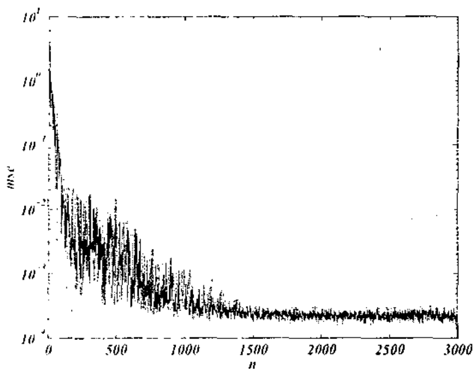


Fig. 6. Learning behavior of the adaptive N-L structure

## VI. Conclusion

A power amplifier for UMTS was measured in order to obtain parameters for a nonlinear model. The measured amplifier showed memory effects, however, due to the rather limited bandwidth they were not very pronounced. Low complexity nonlinear models with memory to describe such a nonlinear power-amplifier were tested, based on the measured I/O-data. An adaptive system identification of the power amplifier, based on the low complex LMS-algorithm behaved convergent in the mean-square sense, analyzed via simulations. Bounds or optimal values for the step-sizes to guarantee convergence, and fastest convergence, remain open problems.

## Acknowledgment

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

## References

[1] E. Aschbacher, M. Rupp, "Modelling and Identification of a Nonlinear Power-Amplifier with Memory for Nonlinear Digital Adaptive Pre-Distortion," in *to appear in Proceedings of SPAWC03, Roma*, 2003.

[2] A. A. M. Saleh, "Frequency-Independent and Frequency-Dependent Nonlinear Models of TWT Amplifiers," *IEEE Transactions on Communications*, vol. COM-29, no. 11, pp. 1715–1720, Nov. 1981.

[3] K. J. Muhonen, M. K., R. Krishnamoorthy, "Look-Up Table Techniques for Adaptive Digital Predistortion: A Development and Comparison," *IEEE Transactions on Vehicular Technology*, vol. 49, no. 5, pp. 1995–2002, Sep. 2000.

[4] S. Boyd, L. O. Chua, "Fading Memory and the Problem of Approximationg Nonlinear Operators with Volterra Series," *IEEE Transactions on Circuits and Systems*, vol. CAS-32, no. 11, pp. 1150–1161, Nov. 1985.

[5] M. Schetzen, *The Volterra and Wiener Theories of Nonlinear Systems*, John Wiley, 1980.

[6] K. S. Narendra, K. Parthasarathy, "Identification and Control of Dynamical Systems Using Neural Networks," *IEEE Transactions on Neural Networks*, vol. 1, pp. 4–27, March 1990.

[7] S. Haykin, *Adaptive Filter Theory*, Prentice Hall, third edition, 1996.