The Cramer-Rao Bound and DMT Signal Optimisation for the Identification of a Wiener-Type Model

H. Koeppl

Christian Doppler Laboratory for Nonlinear Signal Processing, Graz University of Technology, 8010 Graz, Austria
Email: heinz.koeppl@tugraz.at

A. S. Josan

Department of Electronics and Communication Engineering, Indian Institute of Technology Guwahati,
Guwahati 781039, Assam, India
Email: awlok@iitg.ernet.in

G. Paoli

System Engineering Group, Infineon Technologies, 9500 Villach, Austria
Email: gerhard.paoli@infineon.com

G. Kubin

Christian Doppler Laboratory for Nonlinear Signal Processing, Graz University of Technology, 8010 Graz, Austria
Email: gernot.kubin@tugraz.at

In linear system identification, optimal excitation signals can be determined using the Cramer-Rao bound. This problem has not been thoroughly studied for the nonlinear case. In this work, the Cramer-Rao bound for a factorisable Volterra model is derived. The analytical result is supported with simulation examples. The bound is then used to find the optimal excitation signal out of the class of discrete multitone signals. As the model is nonlinear in the parameters, the bound depends on the model parameters themselves. On this basis, a three-step identification procedure is proposed. To illustrate the procedure, signal optimisation is explicitly performed for a third-order nonlinear model. Methods of nonlinear optimisation are applied for the parameter estimation of the model. As a baseline, the problem of optimal discrete multitone signals for linear FIR filter estimation is reviewed.

Keywords and phrases: Wiener model, Cramer-Rao bound, signal design, nonlinear system identification.

1. INTRODUCTION

In the design of optimal excitation signals for system identification, the Cramer-Rao bound plays a central role. For a given model structure, it gives a lower bound on the variance of the unbiased model parameter estimates for a given perturbation scenario [1]. The problem of signal optimisation for the identification of linear models is considered in [2]. We focus on a nonlinear model structure proposed in [3], which is nonlinear in the parameters and can be considered a generalisation of the classical Wiener model [4, page 143]. For the classical Wiener model, the Cramer-Rao bound was derived in [5]. The goal of this work is to gain further insight into the design of optimal excitation signals for the identification of nonlinear cascade systems. The application that drove our investigations is adaptive nonlinear filtering

for ADSL data transmission systems. The block diagram in Figure 1 shows an application of the nonlinear model as a nonlinear canceler of the hybrid echo for the receive path of an ADSL transceiver system. System distortion analysis revealed that the line-driver circuit is the main source of nonlinearity. In the subsequent simulation experiments, a nonlinear Wiener-type model of this line-driver circuit is used as a reference model. As excitation signal the class of discrete multitone (DMT) signals as used in ADSL data transmission is primarily considered. During the startup phase of the ADSL system, it is possible to send a predetermined DMT training sequence for the nonlinear echo canceler. Thus, the goal of the signal optimisation procedure is to find the DMT training sequence which is optimal in the sense that the most accurate model parameter estimates for the echo canceler can be obtained. Our focus is on the effects of a finite number of
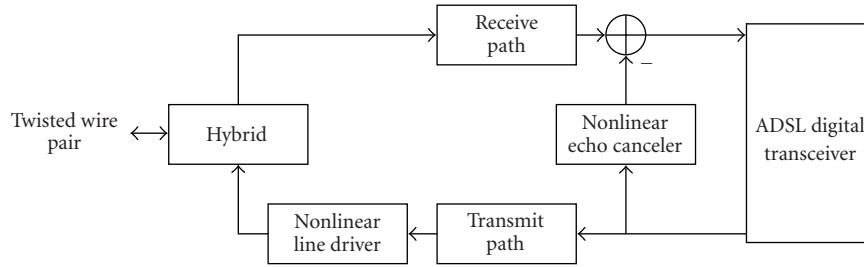
FIGURE 1: Block diagram of the application of a nonlinear canceler of the hybrid echo for an ADSL transceiver system.

tones in the input signal and of a finite number of samples for the estimation of the model parameters.

The work is organised as follows. In Section 2, the considered Wiener-type model is derived from the general Volterra model. The Cramer-Rao bound for this model is computed in Section 3 while Section 4 deals with the parameter estimation algorithm. Verification of the derived Cramer-Rao bound via numerical simulations is performed in Section 5. A discussion, new algorithms, and simulation results concerning the design of optimal excitation signals for the considered model are given in Section 6.

2. VOLTERRA MODEL AND THE WIENER-TYPE MODEL

The multivariate kernel $v_p[k_1, \ldots, k_p]$ of the homogeneous Volterra system of Figure 2 with

$$y[n] = \sum_{k_1=0}^{M_p-1} \cdots \sum_{k_p=0}^{M_p-1} v_p[k_1, \ldots, k_p] u[n-k_1] \cdots u[n-k_p] \tag{1}$$

is factorisable if it can be written as a product of lower-dimensional terms

$$v_p[k_1, \ldots, k_p] = r_p[k_1, \ldots, k_r] w_p[k_{r+1}, \ldots, k_p] \tag{2}$$

shown in Figure 3. The kernel function is fully factorisable if its kernel $v_p[k_1, \ldots, k_p]$ can be written as

$$v_p[k_1, \ldots, k_p] = \prod_{i=1}^{p} h_{pi}[k_i]. \tag{3}$$

The corresponding block diagram is depicted in Figure 4. If all one-dimensional kernels $h_{pi}[k_i]$ are identical, that is, $h_p[k_i] = h_{pi}[k_i]$ for $i = 1, \ldots, p$ with

$$v_p[k_1, \ldots, k_p] = \prod_{i=1}^{p} h_p[k_i], \tag{4}$$

one arrives at the cascade structure of Figure 5, which is recognised as a homogeneous Wiener system. In the case of a general Volterra system of order N for which condition (4) holds for all orders p with $p = 1, \ldots, N$, we obtain the considered simplified factorisable Volterra system.
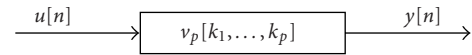

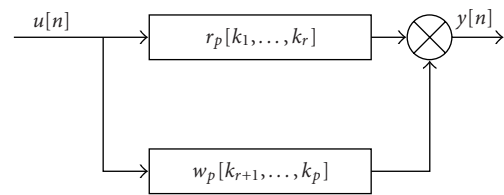FIGURE 2: Homogeneous Volterra system of order p.

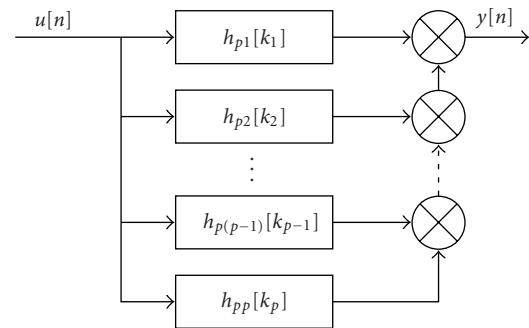FIGURE 3: Partially factorisable homogeneous Volterra system of order p.

FIGURE 4: Fully factorisable homogeneous Volterra system of order p.

This Wiener-type model and the related measurement scenario are depicted in Figure 6. If the N different linear kernels $h_p[k]$ in Figure 6 differ only by a scaling factor, the classical Wiener model is obtained. The measured output $z[n]$ of the considered model can be written as $z[n] = y[n] + \epsilon[n]$ with

$$y[n] = \sum_{p=1}^{N} \left(\sum_{k=0}^{M_p-1} h_p[k] u[n-k] \right)^p, \tag{5}$$

where $u[n]$ is the input signal and $\epsilon[n]$ is assumed to be an additive zero-mean Gaussian noise process with covariance matrix Σ. Subsequently, for the ease of notation and without
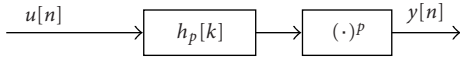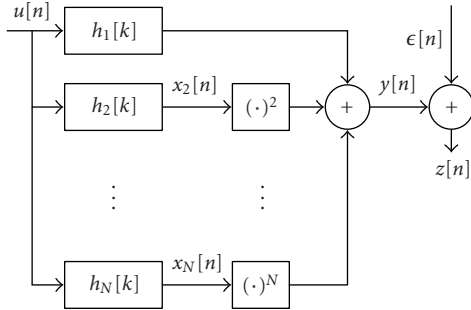
FIGURE 5: Homogeneous Wiener system of order p.

FIGURE 6: The considered nonlinear Wiener-type model.

loss of generality, $M_p = M$ for $p = 1, \ldots, N$ is assumed. For convenience, the following objects are defined. The linear kernel matrix $\mathbf{H} \in \mathbb{R}^{M \times N}$ is defined as

$$\mathbf{H} \equiv \begin{pmatrix} h_1[0] & \cdots & h_N[0] \\ \vdots & & \vdots \\ h_1[M-1] & & h_N[M-1] \end{pmatrix} \tag{6}$$

and the windowed input matrix $\mathbf{U} \in \mathbb{R}^{N_s \times M}$ is defined as

$$\mathbf{U} \equiv \begin{pmatrix} u[1] & u[0] & \cdots & u[-M+2] \\ \vdots & \vdots & & \vdots \\ u[N_s] & u[N_s-1] & \cdots & u[N_s-M+1] \end{pmatrix}, \tag{7}$$

where $u[n]$ for $n < 1$ is assumed to be known and N_s is the considered observation sample length or estimation horizon. To be precise, to build up an $N_s \times M$ data matrix $\mathbf{U}$, one requires the knowledge of $N_s + M - 1$ samples of the input signal $u[n]$, which would actually be the estimation horizon. Nevertheless, in the following, we stick to the convention that the estimation horizon is the number of rows of the data matrix $\mathbf{U}$, that is, N_s. In addition, the power operator $\mathrm{P} : \mathbb{R}^{n \times m} \to \mathbb{R}^n$ with

$$(\mathrm{P}\mathbf{X})_n = \sum_{p=1}^m (X_{np})^p \tag{8}$$

is defined, where the notation $(\cdot)_I$, denoting one element of a nonscalar object with I possibly a multi-index, was used. Making use of the above definitions, the output of the nonlinear model of Figure 6 reads

$$\mathbf{z} = \mathrm{P}\mathbf{X} + \boldsymbol{\epsilon}, \quad \mathbf{X} = \mathbf{U}\mathbf{H}, \tag{9}$$

where the elements of this objects correspond to $z_n \equiv z[n]$, $\epsilon_n \equiv \epsilon[n]$, and $X_{np} \equiv x_p[n]$. The parameter vector $\boldsymbol{\theta} \equiv \mathrm{vec}(\mathbf{H})$ will be needed in the following, where the linear index

j of θ_j corresponds to the matrix indices $[k, p]$ of H_{kp} with $j = (p-1)M + k$ and $k = j \bmod M$, $p = \lceil j/M \rceil$, where $\lceil \cdot \rceil$ denotes the ceiling function.

3. THE CRAMER-RAO BOUND FOR THE WIENER-TYPE MODEL

The Cramer-Rao bound is the theoretical lower bound for the variance of all unbiased estimators $\hat{\boldsymbol{\theta}}$ for the model parameters $\boldsymbol{\theta}$ and is determined by the diagonal elements of the inverse of the Fisher information matrix $\mathbf{F}$:

$$F_{ij} \equiv \mathrm{E}\left(\frac{\partial \ln l(\boldsymbol{\theta}|\mathbf{z})}{\partial \theta_i} \frac{\partial \ln l(\boldsymbol{\theta}|\mathbf{z})}{\partial \theta_j} \right). \tag{10}$$

Here $\mathrm{E}(\cdot)$ denotes the expectation operator with respect to the random vector $\mathbf{z} = \mathrm{P}\mathbf{X} + \boldsymbol{\epsilon}$ and $l(\boldsymbol{\theta}|\mathbf{z})$ is the likelihood function for the parameter vector $\boldsymbol{\theta}$ given the noisy observation vector $\mathbf{z}$ [1]. Thus,

$$\mathrm{cov}\left(\boldsymbol{\theta}\boldsymbol{\theta}^T \right)_{ij} \equiv \mathrm{E}\left([\theta_i - \mathrm{E}(\theta_i)][\theta_j - \mathrm{E}(\theta_j)] \right) \geq \left(\mathbf{F}^{-1} \right)_{ij}. \tag{11}$$

Under the regularity condition [6, page 26]

$$\mathrm{E}\left(\frac{\partial \ln l(\boldsymbol{\theta}|\mathbf{z})}{\partial \boldsymbol{\theta}} \right) = \mathbf{0}, \tag{12}$$

(10) can be written as

$$\mathbf{F} = \mathrm{E}(\mathbf{G}), \tag{13}$$

with

$$G_{ij} \equiv -\frac{\partial^2 \ln l(\boldsymbol{\theta}|\mathbf{z})}{\partial \theta_i \partial \theta_j}, \tag{14}$$

the Hessian matrix of the objective function $-\ln l(\boldsymbol{\theta}|\mathbf{z})$ for the maximum likelihood estimation. For the additive Gaussian noise model of $\boldsymbol{\epsilon}$, the likelihood function $l(\mathbf{H}|\mathbf{z})$ for the parameter matrix $\mathbf{H}$ given the observation vector $\mathbf{z}$ reads as follows:

$$l(\mathbf{H}|\mathbf{z})$$
$$= \left[(2\pi)^{N_s} |\boldsymbol{\Sigma}| \right]^{-1/2} \exp\left\{ -\frac{1}{2} (\mathbf{z} - \mathrm{P}(\mathbf{U}\mathbf{H}))^T \boldsymbol{\Sigma}^{-1} (\mathbf{z} - \mathrm{P}(\mathbf{U}\mathbf{H})) \right\}. \tag{15}$$

The entries of the Fisher information matrix (10) for the considered Wiener-type model (5) are calculated as follows. The log-likelihood function reads as follows:

$$\ln l(\mathbf{H}|\mathbf{z}) = -\frac{1}{2}N_s \log 2\pi - \frac{1}{2}\log|\boldsymbol{\Sigma}| \\ -\frac{1}{2}(\mathbf{z} - \mathrm{P}(\mathbf{U}\mathbf{H}))^T \boldsymbol{\Sigma}^{-1}(\mathbf{z} - \mathrm{P}(\mathbf{U}\mathbf{H})). \tag{16}$$

The derivative of the log-likelihood function with respect to the parameter matrix $\mathbf{H}$ can be decomposed as

$$\frac{\partial \ln l(\mathbf{H}|\mathbf{z})}{\partial H_{rs}} = \frac{\partial \ln l(\mathbf{H}|\mathbf{z})}{\partial \boldsymbol{\epsilon}} \frac{\partial \boldsymbol{\epsilon}}{\partial \mathbf{x}_s} \frac{\partial \mathbf{x}_s}{\partial H_{rs}}, \quad (17)$$

where the columns $\mathbf{x}_s$ of the matrix $\mathbf{X} = [\mathbf{x}_1, \ldots, \mathbf{x}_N]$ have been introduced. The first two terms of the product give

$$\frac{\partial \ln l(\mathbf{H}|\mathbf{z})}{\partial \boldsymbol{\epsilon}} = -\boldsymbol{\epsilon}^T \boldsymbol{\Sigma}^{-1}, \quad (18)$$

$$\frac{\partial \boldsymbol{\epsilon}}{\partial \mathbf{x}_s} \equiv \tilde{\mathbf{X}}_s = s \operatorname{diag}\left(\mathbf{x}_s^{[s-1]}\right), \quad (19)$$

where $(\cdot)^{[p]}$ means elementwise operation. The last term yields

$$\frac{\partial \mathbf{x}_s}{\partial H_{rs}} = \mathbf{u}_r, \quad (20)$$

with the columns $\mathbf{u}_r$ of the matrix $\mathbf{U} = [\mathbf{u}_1, \ldots, \mathbf{u}_M]$. Thus,

$$\frac{\partial \ln l(\mathbf{H}|\mathbf{z})}{\partial H_{rs}} \frac{\partial \ln l(\mathbf{H}|\mathbf{z})}{\partial H_{qp}} = (\tilde{\mathbf{X}}_s \mathbf{u}_r)^T \boldsymbol{\Sigma}^{-1} \boldsymbol{\epsilon} \boldsymbol{\epsilon}^T \boldsymbol{\Sigma}^{-1} \tilde{\mathbf{X}}_p \mathbf{u}_q. \quad (21)$$

Applying the expectation operator to the above expression gives the desired result for the Fisher information matrix, which reads

$$F_{[rs],[qp]} = (\tilde{\mathbf{X}}_s \mathbf{u}_r)^T \boldsymbol{\Sigma}^{-1} \tilde{\mathbf{X}}_p \mathbf{u}_q. \quad (22)$$

The resulting matrix $\mathbf{F} \in \mathbb{R}^{NM \times NM}$ can be thought of as consisting of submatrices $\tilde{\mathbf{F}}_{sp} \in \mathbb{R}^{M \times M}$:

$$\mathbf{F} = \begin{pmatrix} \tilde{\mathbf{F}}_{11} & \cdots & \tilde{\mathbf{F}}_{1N} \\ \vdots & \ddots & \vdots \\ \tilde{\mathbf{F}}_{N1} & \cdots & \tilde{\mathbf{F}}_{NN} \end{pmatrix}, \quad (23)$$

with

$$\tilde{\mathbf{F}}_{sp} = \mathbf{U}^T \tilde{\mathbf{X}}_s \boldsymbol{\Sigma}^{-1} \tilde{\mathbf{X}}_p \mathbf{U}. \quad (24)$$

For the special case of a linear FIR filter, that is, $N = 1$, the Fisher information matrix reads, using (19),

$$\mathbf{F} = \tilde{\mathbf{F}}_{11} = \mathbf{U}^T \boldsymbol{\Sigma}^{-1} \mathbf{U}, \quad (25)$$

which, for $\boldsymbol{\Sigma} = \sigma^2 \mathbf{I}$, gives the familiar result [1, page 86]

$$\mathbf{F}^{-1} = \sigma^2 (\mathbf{U}^T \mathbf{U})^{-1} \quad (26)$$

for the Cramer-Rao bound for linear FIR filters.

4. PARAMETER ESTIMATION

For parameter estimation, the likelihood function $l(\boldsymbol{\theta}|\mathbf{z})$ is maximised with respect to $\boldsymbol{\theta}$ using methods of nonlinear optimisation. The optimisation problem is given as

$$\hat{\boldsymbol{\theta}} = \arg \min_{\boldsymbol{\theta}} J(\boldsymbol{\theta}), \quad J(\boldsymbol{\theta}) \equiv -\ln l(\boldsymbol{\theta}|\mathbf{z}), \quad (27)$$

and $\hat{\boldsymbol{\theta}} \equiv \operatorname{vec}(\hat{\mathbf{H}})$. For the FIR Wiener-type model of (5), the gradient $\mathbf{g} \equiv \partial_{\boldsymbol{\theta}} J(\boldsymbol{\theta})$ as well as the Hessian $\mathbf{G} \equiv \partial_{\boldsymbol{\theta}\boldsymbol{\theta}^T} J(\boldsymbol{\theta})$ of (14) can be computed explicitly. Following the matrix notation for the model parameters, the gradient can be written in matrix form. Define the gradient matrix $\partial_{\mathbf{H}}$ as composed of the gradient vectors for each order of nonlinearity

$$\partial_{\mathbf{H}} \equiv [\partial_{\mathbf{h}_1}, \ldots, \partial_{\mathbf{h}_N}], \quad (28)$$

where $\mathbf{H} \equiv [\mathbf{h}_1, \ldots, \mathbf{h}_N]$ and $\partial_{\boldsymbol{\theta}} = \operatorname{vec}(\partial_{\mathbf{H}})$. Applied to the objective function $J(\boldsymbol{\theta})$, the elements are found to be

$$\partial_{\mathbf{h}_s} J(\mathbf{H}) = -\mathbf{U}^T \tilde{\mathbf{X}}_s \boldsymbol{\Sigma}^{-1} \boldsymbol{\epsilon}. \quad (29)$$

In correspondence to the matrix structure of the Fisher information matrix in (24), the "off-diagonal" submatrices of the Hessian matrix are

$$\mathbf{G}_{sp} \equiv \partial_{\mathbf{h}_s \mathbf{h}_p^T} J(\mathbf{H}) = \mathbf{U}^T \tilde{\mathbf{X}}_s \boldsymbol{\Sigma}^{-1} \tilde{\mathbf{X}}_p \mathbf{U} \quad \text{for } s \neq p. \quad (30)$$

The diagonal submatrices given in component notation read

$$\begin{aligned} G_{[rs][qs]} &\equiv \partial_{H_{rs} H_{qs}} J(\mathbf{H}) \\ &= \mathbf{u}_r^T \tilde{\mathbf{X}}_s \boldsymbol{\Sigma}^{-1} \tilde{\mathbf{X}}_s \mathbf{u}_q \\ &\quad + s(s-1) \boldsymbol{\epsilon}^T \boldsymbol{\Sigma}^{-1} \operatorname{diag}\left((\mathbf{x}_s)^{[s-2]}\right) \operatorname{diag}(\mathbf{u}_r) \mathbf{u}_q. \end{aligned} \quad (31)$$

Applying (13) to (30) and (31) and acknowledging the fact that $\boldsymbol{\epsilon}$ is a zero-mean process, the Fisher information matrix (24) is retained. As with (29), (30), and (31), first- and second-order derivatives are available, and it is possible to apply a Newton-like optimisation algorithm [7] for the minimization of (27). This algorithm uses the quadratic approximation of $J(\boldsymbol{\theta})$ around some estimate $\boldsymbol{\theta}^{(k)}$ obtained after k iterations

$$J(\boldsymbol{\theta}^{(k)} + \boldsymbol{\delta}) \approx J(\boldsymbol{\theta}^{(k)}) + \boldsymbol{\delta}^T \mathbf{g}^{(k)} + \frac{1}{2} \boldsymbol{\delta}^T \mathbf{G}^{(k)} \boldsymbol{\delta}, \quad (32)$$

with $\boldsymbol{\delta} = \boldsymbol{\theta} - \boldsymbol{\theta}^{(k)}$. For each iteration k, the quadratic approximation is minimised with respect to $\boldsymbol{\delta}$, where $\mathbf{g}^{(k)}$ and $\mathbf{G}^{(k)}$ denote the gradient and Hessian evaluated at $\boldsymbol{\theta}^{(k)}$, respectively. For this task, the Matlab routine fminunc.m [8] is applied. This procedure requires good initialisation to converge to the global minimum of the objective function $J(\boldsymbol{\theta})$ which is in general multimodal. In this case, the maximum likelihood estimator (27) yields an unbiased estimate. Furthermore, the maximum likelihood estimator is a minimum variance estimator [1], thus the variance of this estimator coincides with the Cramer-Rao bound.

5. VERIFICATION OF THE THEORETICAL RESULT

The above result (24) for the Fisher information matrix of the Wiener-type model is verified by simulation examples. For this purpose, a Wiener-type system is defined and will serve as a reference system for the subsequent simulations. The verification is done by comparing the theoretical parameter

TABLE 1: Model coefficients of the third-order Wiener-type reference model of the line-driver circuit.

Tap	$k = 0$	$k = 1$	$k = 2$	$k = 3$	$k = 4$	$k = 5$
$h_1[k]$	4.2299	1.3909	-1.0805	0.7283	-0.3481	0.0931
$h_3[k]$	0.0511	0.1537	-0.2463	0.1418	-0.0314	0.0009
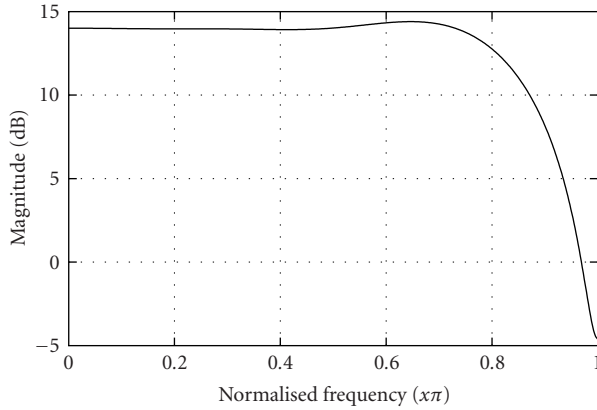
FIGURE 7: Absolute value of the linear transfer function $H_1(e^{j\omega})$ of the Wiener-type reference model of Table 1.
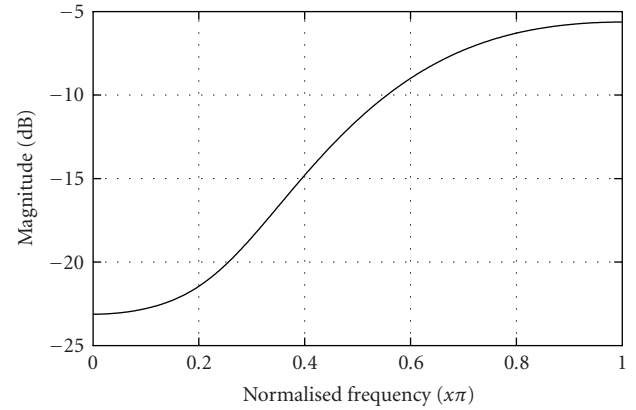
FIGURE 8: Absolute value of the cubic transfer function $H_3(e^{j\omega})$ of the Wiener-type reference model of Table 1.

variance obtained from the Fisher information matrix (24) with the parameter variance obtained by repeated estimation of the model parameters with the algorithm described in Section 4. As this estimator is a minimum variance estimator, the two variances are expected to match. This coincidence is checked for DMT input signals as well as for white Gaussian noise (WGN) input signals over different signal-to-noise (SNR) levels.

5.1. The reference model

For the simulation, a specific reference configuration of the Wiener-type model is chosen. This reference configuration is a simple discrete-time model of an ADSL, G.Lite line-driver circuit [9]. To present reproducible results, the simplest model of the circuit was chosen as the reference model and explicit values of the model coefficients are given. It is a third-order model encompassing 12 coefficients θ_j. Through the differential design of the circuit, the effects of nonlinearities of even orders are negligible compared to the effects of the nonlinearities of odd orders. Thus, the model consists only of a dominating linear part with $M_1 = 6$ and of a small part of third order with $M_3 = 6$. The explicit values of the model coefficients are given in Table 1. They were found originally by identifying the line-driver circuit using a broadband DMT input signal and the estimation algorithm of Section 4.

The model equation for this case reads

$$z[n] = \sum_{k=0}^{5} h_1[k] u[n-k]$$
$$+ \left(\sum_{k=0}^{5} h_3[k] u[n-k] \right)^3 + \epsilon[n]. \tag{33}$$

Written in the compact notation of Section 3, this gives

$$\mathbf{z} = \mathrm{P}(\mathbf{U}\mathbf{H}_r) + \boldsymbol{\epsilon}, \tag{34}$$

with the reference coefficient matrix $\mathbf{H}_r \in \mathbb{R}^{6 \times 2}$. Frequency responses for the linear part $H_1(e^{j\omega}) = \mathcal{F}(h_1[k])$ and for the cubic part $H_3(e^{j\omega}) = \mathcal{F}(h_3[k])$ of the reference model are depicted in Figures 7 and 8, respectively. The linear response shows the typical lowpass characteristic of a power amplifier, while the third-order response reflects the common observation that the nonlinear distortion gets higher for higher frequencies. In Figure 9, the power spectrum of the output signal of the Wiener-type reference model of Table 1 is shown, for a typical downstream ADSL DMT signal as input. The magnitude of the intermodulation products indicates that the nonlinear distortion introduced by the third-order term is 60 dB below the carrier signal. Thus, we are dealing with an extremely weak nonlinear system. Subsequently, the Fisher information matrix of (24) and its inverse are computed for this reference model. In correspondence to the partitioning (24) of the Fisher information matrix

$$\mathbf{F} = \sigma^2 \begin{pmatrix} \mathbf{U}^T\mathbf{U} & \mathbf{U}^T\tilde{\mathbf{X}}_3\mathbf{U} \\ \mathbf{U}^T\tilde{\mathbf{X}}_3\mathbf{U} & \mathbf{U}^T\tilde{\mathbf{X}}_3\tilde{\mathbf{X}}_3\mathbf{U} \end{pmatrix}, \tag{35}$$

the positive-definite covariance matrix can be decomposed into four submatrices:

$$\mathrm{cov}(\boldsymbol{\theta}\boldsymbol{\theta}^T) = \begin{pmatrix} \mathrm{cov}(\mathbf{h}_1\mathbf{h}_1^T) & \mathrm{cov}(\mathbf{h}_1\mathbf{h}_2^T) \\ \mathrm{cov}(\mathbf{h}_2\mathbf{h}_1^T) & \mathrm{cov}(\mathbf{h}_2\mathbf{h}_2^T) \end{pmatrix}. \tag{36}$$
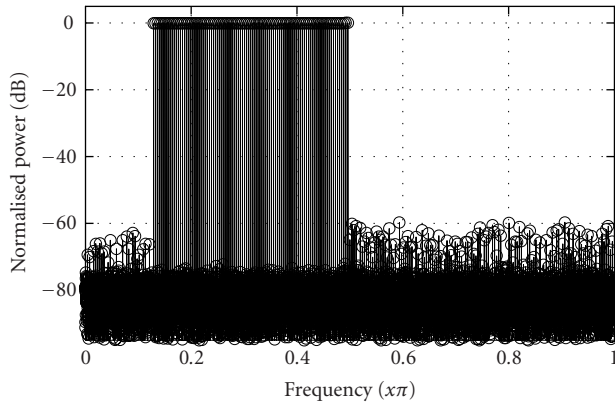
FIGURE 9: Power spectrum of the output of the Wiener-type reference model of Table 1 for the line-driver circuit: DMT input signal with N_c =95 carriers; the perturbation is additive WGN with $\sigma^2 = 1 \times 10^{-5}$.
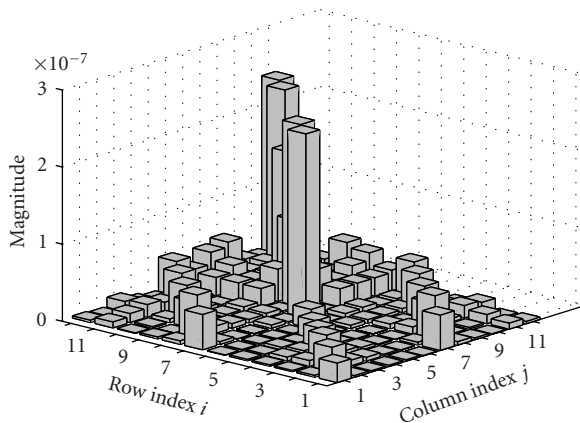

FIGURE 10: Cramer-Rao lower bound on the parameter covariance matrix $\mathrm{cov}(\boldsymbol{\theta}\boldsymbol{\theta}^T)_{ij}$ with $M = 6$, first- and third-order nonlinearity, and $N_s = 1000$; the pertubation is WGN with $\sigma^2 = 1 \times 10^{-5}$ and $u[n]$ is a WGN input signal with power $\sigma_u^2 = 0.64$.

In Figure 10, the parameter covariance matrix $\mathrm{cov}(\boldsymbol{\theta}\boldsymbol{\theta}^T)_{ij}$ for the Wiener-type reference model is shown for the case $N_s = 1000$ and $\sigma^2 = 1 \times 10^{-5}$ for a WGN input signal with variance $\sigma_u^2 = 0.64$. The figure reveals that there is a high covariance between the linear parameters and the third-order parameters. That corresponds to the known fact that even in the case of a white input signal, the homogeneous first- and third-order responses of a multilinear operator, such as a Volterra model, are correlated [10].

5.2. Parameter estimation and variance comparison

In the following, the derivation of Section 3 is verified using different excitation signals and different perturbation scenarios. These investigations of the Wiener-type reference model of Table 1 are done with an estimation horizon of $N_s = 50$. The variance estimates of the estimators are obtained by repeating the identification procedure of Section 4
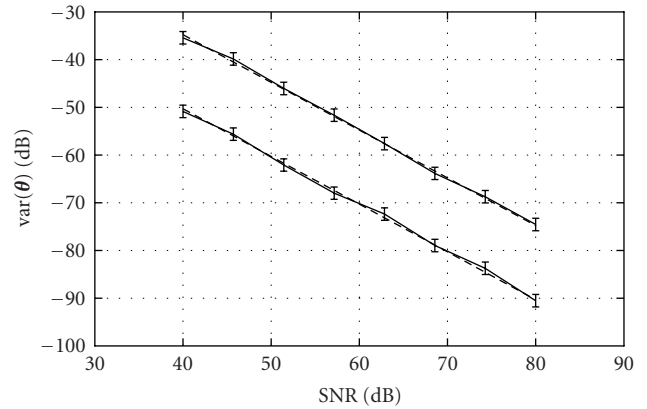

FIGURE 11: Linear dependence of Cramer-Rao bound (dashed) on the SNR and variance of the estimators (solid) over different SNR with 95% confidence intervals shown as vertical bars, plotted for one kernel value for each order p; the two upper curves correspond to parameter $H_{12} = h_3[0]$; the two lower curves correspond to parameter $H_{11} = h_1[0]$; the input signal is WGN.

for $N_r = 100$ i.i.d. realisations of the perturbation process $\epsilon[n]$. Following the asymptotic results of the normality of the maximum likelihood estimator [11, page 52], the parameter estimates pass the Lilliefors test for normality [12]. Thus, the 95% confidence intervals of a normal distribution are indicated in the following figures. To keep these figures simple, the Cramer-Rao bound $\mathrm{diag}(\mathbf{F}^{-1})$ and the variance estimates $\mathrm{var}(\boldsymbol{\theta})$ of only one model parameter per order of nonlinearity p are shown versus different SNR.

5.2.1. WGN input signal

The input signal $u[n]$ to the reference model is taken to be WGN, $u[n] \sim \mathcal{N}(0, \sigma_u^2)$ with σ_u^2 =0.64, while the additive perturbation of the output $y[n]$ is $\epsilon[n] \sim \mathcal{N}(0, \sigma^2)$. The Cramer-Rao bound, the variance estimates of the estimators, and their corresponding confidence regions versus different SNR levels are given in Figure 11. Good agreement between simulation and theory can be observed.

5.2.2. DMT input signal

As a second scenario, the input signal $u[n]$ is taken to be a DMT signal:

$$u[n] = \sum_{k=0}^{N_c-1} a_k \cos\left[(k_s + k)\omega_0 n + \varphi_k\right], \qquad (37)$$

where ω_0 is the normalised grid frequency of the DMT signal. For further use, we define the vector of amplitudes $\mathbf{a} \equiv [a_0, \ldots, a_{N_c-1}]^T$, the corresponding vector of powers of the individual tones $\mathbf{p}$, and the vector of normalised frequencies $\boldsymbol{\omega} \equiv \omega_0 \cdot [k_s, k_s + 1, \ldots, k_s + N_c - 1]^T$. The phase set $\boldsymbol{\varphi} \equiv [\varphi_0, \ldots, \varphi_{N_c-1}]^T$ for this simulation is initialised with random numbers drawn from the uniform distribution $\mathcal{U}[0, 2\pi]$. The identification of the reference model is
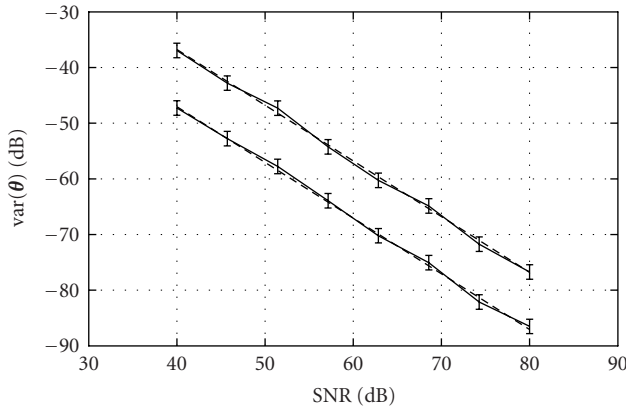
FIGURE 12: Linear dependence of Cramer-Rao bound (dashed) on the SNR and variance of the estimators (solid) over different SNR with 95% confidence intervals shown as vertical bars, plotted for one kernel value for each order p; the two upper curves correspond to parameter $H_{12} = h_3[0]$; the two lower curves correspond to parameter $H_{11} = h_1[0]$; the input signal is a DMT signal with $N_c = 12$.

performed using $N_c = 12$ tones and is done for different SNR levels. The Cramer-Rao bound, the variance estimates of the estimators, and their corresponding confidence regions versus different SNR levels are given in Figure 12. Once again, good agreement between simulation and theory can be observed.

6. DESIGN OF OPTIMAL EXCITATION SIGNALS

Given a model structure with unknown parameters, the accuracy of the parameter estimates of the model depends on the used identification procedure and on the used excitation signal. If the estimator is a minimum variance estimator, then its parameter variance achieves the lower bound, that is, the Cramer-Rao bound. Thus, to even further decrease the variance of the minimum variance estimator of Section 4, one can only optimise the excitation signal in such a way that the corresponding Cramer-Rao bound is decreased. To have an optimality measure, a scalar objective function $\Psi : \mathbb{R}^{MN \times MN} \to \mathbb{R}$ of $\mathbf{F}^{-1}$ has to be found. In the theory of experiment design [13], different types of this objective function $\Psi(\cdot)$ are considered. The most popular criterion of optimality is $\Psi(\mathbf{F}^{-1}) = |\mathbf{F}^{-1}| = |\mathbf{F}|^{-1}$, where $|\cdot|$ denotes the determinant of a matrix.

6.1. Signal design for linear FIR filters

In this section, the well-known problem of optimising the amplitude distribution of a DMT signal subject to a total power constraint so as to achieve minimal variance estimates of the parameters of a linear FIR filter is reviewed. For a WGN perturbation, the Fisher information matrix for the linear FIR filter case is given by (26). As mentioned earlier, one way to minimize the Cramer-Rao bound is to maximize the determinant of $\mathbf{F}$. We apply the inequality $\log x \leq \alpha x - 1 - \log \alpha$ for every $\alpha > 0$ to the M eigenvalues λ_k of the positive-semidefinite matrix $\mathbf{F}$:

$$\sum_{k=1}^{M} \log \lambda_k \leq \alpha \sum_{k=1}^{M} \lambda_k - M(1 + \log \alpha). \qquad (38)$$

Inequality (38) is equivalent to

$$\log |\mathbf{F}| \leq \alpha \operatorname{Tr}(\mathbf{F}) - M(1 + \log \alpha), \qquad (39)$$

with $\operatorname{Tr}(\cdot)$ denoting the trace of a matrix. The quantity $\log |\mathbf{F}|$ reaches its upper bound at $\lambda_k = \lambda = 1/\alpha$ for $k = 1, \ldots, M$. The consequences of this relation for signal optimisation are outlined in the following example. Consider the case N_s is the period of the DMT signal (37). The diagonal elements of $\mathbf{F}$ are all equal and correspond to the constrained total power of the DMT signal, that is, $\operatorname{Tr}(\mathbf{F}) = \sigma^{-2} M N_s \sum p_k$. Thus, for a given power of the DMT signal, the right-hand side of (39) is fixed and gives the upper bound for $\log |\mathbf{F}|$. It reaches its upper bound if the eigenvalues are all equal to $\lambda = 1/\alpha$ with $\alpha = \sigma^2/(N_s \sum p_k)$.

Furthermore, if we assume that M is even and $M = N_s$, with (7) and (26), the matrix $\mathbf{F}$ turns out to be a circulant. Thus, the similarity transformation which diagonalises $\mathbf{F}$ is the discrete Fourier transform (DFT) $\mathbf{T} \in \mathbb{C}^{M \times M}$ and the eigenvalues of $\mathbf{F}$ are the diagonal elements of $\mathbf{S} = \mathbf{T}\mathbf{F}\mathbf{T}^{-1}$ [14, page 379]. If the frequency spacing of the DMT signal (37) is chosen to be $\omega_0 = 2\pi/M$ and $k_s = 0$, the eigenvalues of $\mathbf{F}$ correspond to the discrete power spectrum of the DMT signal. The matrix $\mathbf{F}$ is nonsingular for $k = 0, \ldots, M/2$, which corresponds to $N_c = M/2 + 1$ tones of the DMT signal. The tones at $k = 0$ and $k = M/2$ contribute one spectral component to the discrete power spectrum each, while all other tones contribute two spectral components each. Thus, the eigenvalues of $\mathbf{F}$ are all equal and $\log |\mathbf{F}|$ reaches its upper bound if the $M/2 + 1$ element amplitude vector of the DMT signal has the form $\mathbf{a} = [a/2, a, \ldots, a, a/2]^T$. This is in accordance with the engineering intuition that for a finite number of tones and a predetermined power of the DMT signal, the most accurate parameter estimation is possible if the power is equally distributed over all spectral components. Note that the above example is constructed in such a way that the frequency grid of the DMT signal spans the full bandwidth, that is, $\omega = 2\pi/M \cdot [0, 1, \ldots, M/2]^T$. In general, the circularity of $\mathbf{F}$ is preserved if $N_s = m N_p$ and $M = N_p$, where N_p is the period of the DMT signal and $m \in \mathbb{N}$. In such situations, every mth spectral component of the DMT signal (37) with $\omega_0 = 2\pi/M$ and $N_c = M/2 + 1$ is nonzero and corresponds to an eigenvalue of the matrix $\mathbf{F}$. From above considerations, it is clear that for a frequency spacing $\omega_0 = 2\pi/M$ and $N_c < M/2 + 1$, at least one eigenvalue of $\mathbf{F}$ is exactly zero. Thus, the corresponding estimation problem is an ill-posed one. As soon as the constraints $N_s/M \in \mathbb{N}$ and $\omega_0 = 2\pi/M$ do not hold, the one-to-one correspondence between an eigenvalue of $\mathbf{F}$ and a nonzero spectral component of the DMT signal is lost. Thus, in the general case, one tone of the DMT signal impacts more than one eigenvalue of $\mathbf{F}$. In this case, the amplitude distribution of the DMT signal that maximises $\log |\mathbf{F}|$ has to be found through numerical optimisation methods.
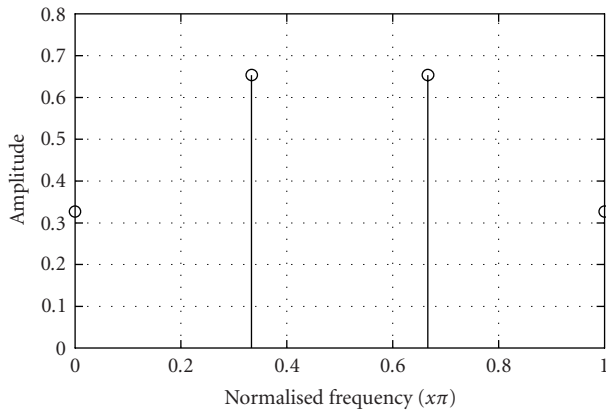
FIGURE 13: Optimal amplitude distribution of a DMT signal over the full bandwidth $[0, \pi]$ encompassing $N_c = 4$ tones for the estimation of an $M = 6$ FIR filter.

In [15], it is shown that, for linear FIR filters, the maximization of $\log |\mathbf{F}|$ subject to the signal power constraint $\sum p_k \leq 1$ leads to a semidefinite programming problem which can be solved efficiently [16]. More explicitly, the semidefinite program takes the form

$$\max_{\mathbf{p}} \log |\mathbf{F}(\mathbf{p})|, \quad \text{subject to } \mathbf{F}(\mathbf{p}) \geq 0, \ \tilde{\mathbf{p}} \geq 0, \quad (40)$$

with $\tilde{\mathbf{p}} \equiv [1 - \sum p_k, p_0, \ldots, p_{N_c-1}]^T$. The key observation that allows this elegant formulation is that the Fisher information matrix for a period of a DMT signal is the weighted sum of partial Fisher information matrices corresponding to each tone of the DMT signal. The weights turn out to be the powers p_k of the individual tones. Following this approach, the optimal excitation signals for a linear FIR filter are found subsequently. From (25), it is clear that the amplitude distribution of the optimal DMT signal does not depend on the model parameters. In correspondence to the linear part of the reference model of Table 1, the optimal amplitude distribution for an $M = 6$ linear FIR filter is computed.

6.1.1. DMT signal with bandwidth $[0, \pi]$

To guarantee that the matrix $\mathbf{F}$ is nonsingular, above considerations suggest that at least $N_c = M/2 + 1 = 4$ tones are required if tones at $\omega = 0$ and $\omega = \pi$ are included. The optimised amplitude distribution found by semidefinite programming is given in Figure 13. This amplitude distribution corresponds to a flat signal spectrum because the spectral components for $\omega_k = 0$ and $\omega_k = \pi$ scale differently (by a factor of 2) than the other components. Thus for a finite number of tones and finite sample length N_s equal to the period of the signal and for full bandwidth, the spectrum of the optimal DMT signal turns out to be flat. For many applications, the number of tones of the excitation signal is not exactly $N_c = M/2 + 1$, but higher. Also for such a case with $N_c > M/2+1$, the optimal amplitude distribution over the full bandwidth $[0, \pi]$ is found to be spectrally flat. More interesting observations can be made for a bandpass DMT signal in the next section.
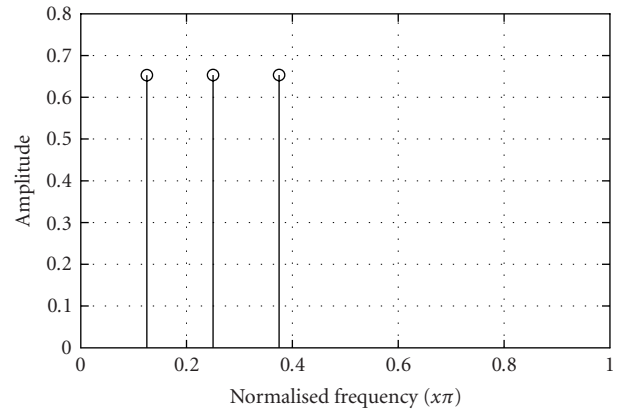

FIGURE 14: Optimal amplitude distribution for a bandpass DMT signal encompassing $N_c = 3$ tones for the estimation of an $M = 6$ FIR filter.
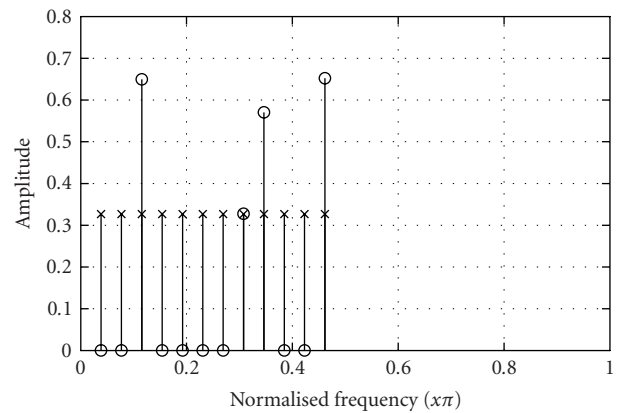

FIGURE 15: Amplitude distribution of a bandpass DMT signal encompassing $N_c = 12$ tones for the estimation of an $M = 6$ FIR filter: optimised signal (circles) and, for reference, the spectrally flat signal (crosses).

6.1.2. DMT signal with bandwidth $(0, \pi/2)$

In the case of a bandpass signal, where neither the frequency $\omega = 0$ nor $\omega = \pi$ is included, each tone contributes two spectral components and thus the minimum number of tones required for the estimation of the linear FIR filter is $N_c = \lceil M/2 \rceil$. The optimal amplitude distribution for an $N_c = 3$ bandpass signal using semidefinite programming is depicted in Figure 14. Thus, for the bandpass signal with $N_c = \lceil M/2 \rceil$, the optimal spectral distribution is flat over the given bandwidth $(0, \pi/2)$. But, if more than $\lceil M/2 \rceil$ tones are contained in the DMT signal, the optimal amplitude distribution is no longer spectrally flat. This is exemplified for the case $N_c = 12$ in Figure 15. The figure shows, in addition to the optimal amplitude distribution, the spectrally flat amplitude distribution as a reference. Thus, for general bandpass DMT signals, it turns out that the optimal spectral distribution is not flat over the given bandwidth $(0, \pi/2)$. In the next section, this result is verified through estimation runs using the
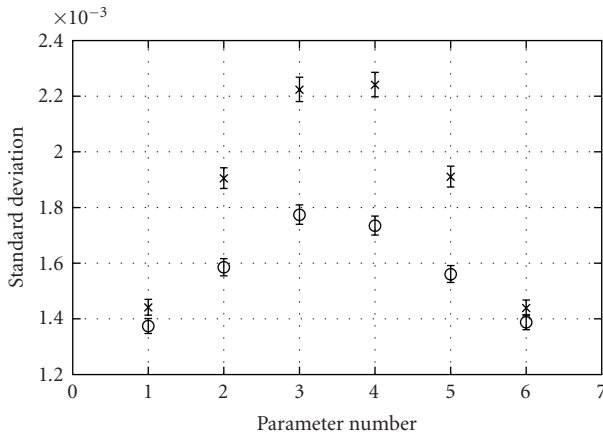
FIGURE 16: Mean and 95% confidence region of the estimated standard deviation of the linear FIR filter parameter estimates for a bandpass input signal with $N_c = 12$ tones: spectrally flat amplitude distribution (crosses), optimised amplitude distribution (circles); the perturbation is WGN with $\sigma^2 = 1 \times 10^{-5}$ and the estimation horizon is $N_s = 56$.

optimal $N_c = 12$ DMT signal and the spectrally flat $N_c = 12$ DMT signal.

6.1.3. Comparison of the estimation performance of bandpass DMT signals

Now that the optimal bandpass input signal for a linear FIR filter is found, the signal can be applied to the identification of a given linear FIR filter. The result is then compared with the identification result obtained by applying the bandpass signal with a flat spectral distribution for the given bandwidth $(0, \pi/2)$. For this, the linear part of the Wiener-type model of Table 1 is used as the reference linear FIR filter and input-output data, that is, $\{u[n], z[n]\}$, are measured. For identification the unbiased minimum variance estimator (UMVE) [1, page 87] for the linear FIR filter case,

$$\hat{\boldsymbol{\theta}} = (\mathbf{U}^T \mathbf{U})^{-1} \mathbf{U}^T \mathbf{z} \qquad (41)$$

is applied both for the optimal bandpass sequence and for the spectrally flat bandpass sequence. The variance of the estimate $\hat{\boldsymbol{\theta}}$ is computed by performing the estimation (41) over $N_r = 1000$ i.i.d. noise realisations of the perturbation process $\epsilon[n] \sim \mathcal{N}(0, \sigma^2)$ with $\sigma^2 = 1 \times 10^{-5}$ and $N_s = 56$. The estimated standard deviations of each FIR filter parameter are shown for these signals in Figure 16. In addition, the Cramer-Rao bounds for both signals and each parameter are computed. All bounds lie in the indicated 95% confidence region. To keep the figure simple, the bounds are not shown in Figure 16. The result shows clearly that the optimised DMT signal which is not spectrally flat outperforms the spectrally flat reference DMT signal. The relative reduction of the parameter variance averaged over all FIR filter parameters comes out to be 26.01% or 1.45 dB. The following remarks can be made.

(1) To be able to apply semidefinite programming, the estimation horizon N_s has to match multiples of the period of the DMT signal. In this case, the phase distribution $\boldsymbol{\varphi}$ falls out of the optimisation problem.

(2) The characteristic shape of the variance as a function of the parameter index as plotted in Figure 16 can be explained by the spectral decomposition of the matrix $\mathbf{F}$. Due to the band limitation, the eigenvalue spread of the matrix $\mathbf{F}$ is of the order 1×10^3. Therefore, $\mathbf{F}^{-1}$ is governed by the smallest eigenvalue λ_k of $\mathbf{F}$ and can be approximated by $\mathbf{F}^{-1} \approx \lambda_k^{-1} \mathbf{v}_k \mathbf{v}_k^T$, where $\mathbf{v}_k$ is the corresponding eigenvector of $\mathbf{F}$. Thus, the characteristic shape in Figure 16 is primarily determined by the shape of the eigenvector corresponding to the smallest eigenvalue of $\mathbf{F}$.

6.2. DMT signal design for the Wiener-type model

As the Wiener-type model of (5) is a nonlinear-in-the-parameters model, its Fisher information matrix (24) depends on the model parameters. In contrast to the FIR filter case, for each model parameter set, an optimal excitation signal can be defined. Furthermore, the entries of the Fisher information matrix correspond to higher-order moments of the input signal. Therefore, the optimal DMT signal is not only determined by its amplitude distribution but also by its phase distribution $\boldsymbol{\varphi}$. This implies that, even in the case where the estimation horizon N_s is the period of the DMT signal, the entire Fisher information matrix cannot be written as a weighted sum of the partial Fisher information matrices for each tone of the DMT signal. Due to this, the formulation of the signal optimisation problem by a semidefinite program is not possible for the case of the Wiener-type model. The optimisation problem reads

$$\max_{\mathbf{p}, \boldsymbol{\varphi}} \log |\mathbf{F}(\mathbf{p}, \boldsymbol{\varphi})|, \quad \text{subject to } \mathbf{F}(\mathbf{p}, \boldsymbol{\varphi}) \geq 0, \ \tilde{\mathbf{p}} \geq 0, \quad (42)$$

where the objective function $\log |\mathbf{F}(\mathbf{p}, \boldsymbol{\varphi})|$ and the constraint for the positive semidefiniteness $\mathbf{F}(\mathbf{p}, \boldsymbol{\varphi}) \geq 0$ are now nonlinear functions of the optimisation variables $\mathbf{p}$ and $\boldsymbol{\varphi}$. To the best of the authors' knowledge, no optimisation algorithm is available that combines a nonlinear objective function with a nonlinear semidefinite matrix constraint. Furthermore, for the above optimisation problem and for the rest of Section 6.2, it is assumed that the reference model coefficients of Table 1 are known, where as in reality they are not. In Section 6.3, a practical solution to circumvent this unrealistic assumption is presented.

6.2.1. Design of optimal QAM-DMT signals

To still be able to illustrate the role of optimal signal design for the Wiener-type model, we restrict the considered signal class to a subclass of DMT signals with a finite number of members. The determination of the optimal excitation signal from this subclass can now be tackled by a complete search over all members of the subclass. A realistic subclass is the class of DMT signals that are modulated according to a specific QAM (quadrature amplitude modulation) scheme. The amplitudes and phases of the tones can now vary only on
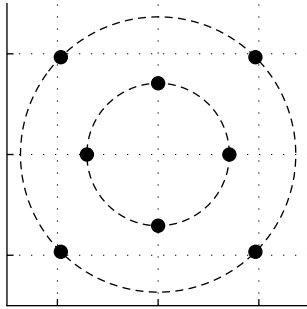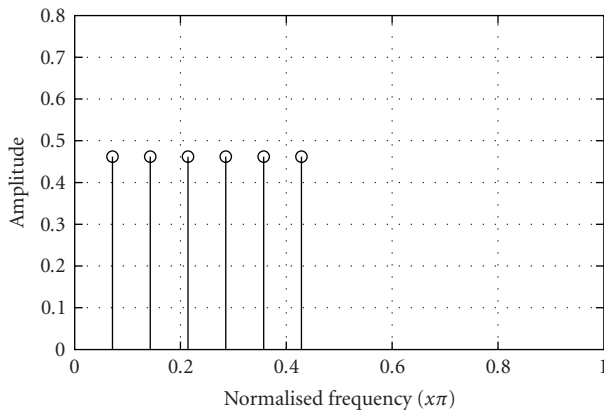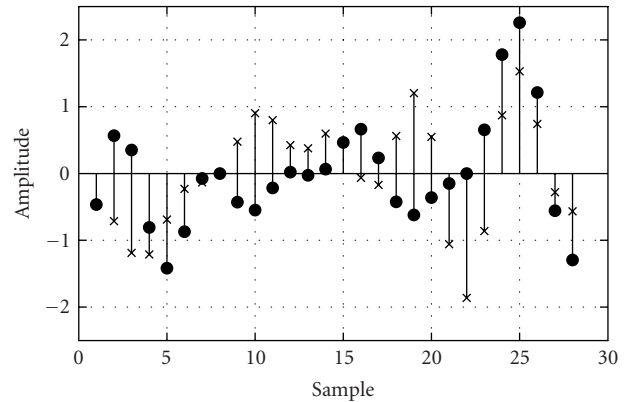
FIGURE 17: Eight-point QAM signal constellation.

FIGURE 18: Optimal amplitude distribution of the bandpass eight-point QAM-DMT signal encompassing $N_c = 6$ tones for the estimation of the Wiener-type model of Table 1.

FIGURE 19: One period $N_s = 28$ of two discrete-time input signals for the Wiener-type model: signal with optimal QAM constellation (circles) and suboptimal signal (crosses) with the same amplitude but different phase distribution than the optimal signal.

the quantised levels of the QAM constellation. In the following simulation experiments, an eight-point QAM for each of the N_c tones is applied. The amplitude quantisation is done in such a way that if all N_c tones occupy the outer ring of the QAM constellation, the signal power is $\sum p_k = 0.64$. In Figure 17, the used QAM constellation is depicted schematically. The optimal amplitude distribution for an eight-point QAM-DMT bandpass signal with $\omega \in (0, \pi/2)$, $N_c = 6$, which maximises $\log |\mathbf{F}(\mathbf{p}, \boldsymbol{\varphi})|$, found through a complete search for the nonlinear reference model of Table 1, is shown in Figure 18. For the 12-parameter Wiener-type reference model, a DMT signal with at least $N_c = 6$ tones has to be applied to prevent an ill-posedness of the estimation problem. From the insight gained through the simulation experiments, the following remarks can be made.

(1) Due to the experiment setup, it comes at no surprise that the amplitude distribution of the optimal excitation signal for the Wiener-type model is spectrally flat. The reason for that is that, roughly speaking, the Cramer-Rao bound can be seen as a noise-to-signal power ratio and thus the bound gets lowered if more signal power is applied to the corresponding system. Therefore, for the optimal signal, all of the $N_c = 6$ tones occupy the outer QAM constellation points of Figure 17.

(2) In contrast to the linear FIR filter case, the phase constellation turns out to be of crucial importance even for N_s being the signal period. It is observed that even input signals with the same amplitude distribution but different phase sets $\boldsymbol{\varphi}$ than the optimal input signal can lead not only to very high Cramer-Rao bounds but even to biased estimates. These biased estimates are caused by the practical problem that, for these special phase sets $\boldsymbol{\varphi}$, the Hessian matrix of the estimator of Section 4 gets near to a singular matrix and thus the optimisation algorithm fails to converge.

Note that these observations have severe implications for the methodology of nonlinear system identification. An improper choice of the phase set of the DMT excitation signal can lead to an extremely ill-posed estimation problem.

6.2.2. Comparison of the estimation performance for QAM-DMT signals

As a consequence of the above remarks, we present an estimation performance comparison between the optimal input signal (determined by its phase and amplitude distribution) and an input signal with the same amplitude but different phase distribution, which still allows an unbiased estimation, that is, allows convergence of the optimisation algorithm. The two discrete-time signals which are compared in the estimation performance are shown in Figure 19. The performance is evaluated by repeated identification of the reference Wiener-type model of Table 1 over $N_r = 500$ i.i.d. realisations of the perturbation process $\epsilon[n] \sim \mathcal{N}(0, \sigma^2)$ with $\sigma^2 = 1 \times 10^{-5}$. The resulting standard deviations of the estimates for the two excitation signals are shown in Figures 20 and 21 for the linear and cubic part of the Wiener-type model, respectively. In addition, the Cramer-Rao bounds for both signals and each model parameter are computed. All bounds lie in the indicated 95% confidence region. To keep the figures simple, they are not shown in Figures 20 and 21. The mean parameter variance and the variance gain for the two signals of Figure 19 are given in Table 2.
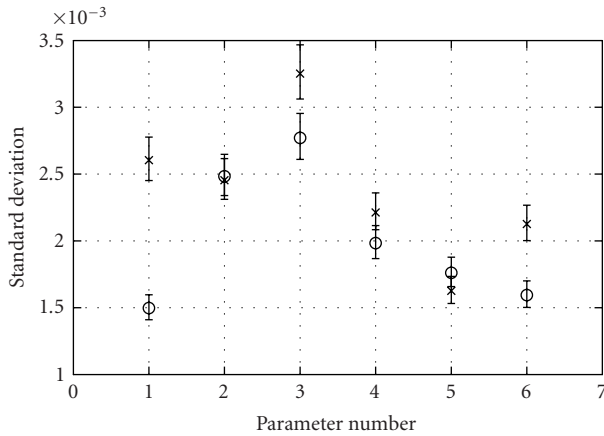
Figure 20: Mean and 95% confidence region of estimated standard deviation of the estimators for the linear part of the Wiener-type model for a bandpass QAM-DMT signal with $N_c = 6$: optimal input signal (circles) and suboptimal input signal (crosses); the perturbation is WGN with $\sigma^2 = 1 \times 10^{-5}$ and the estimation horizon is $N_s = 28$.
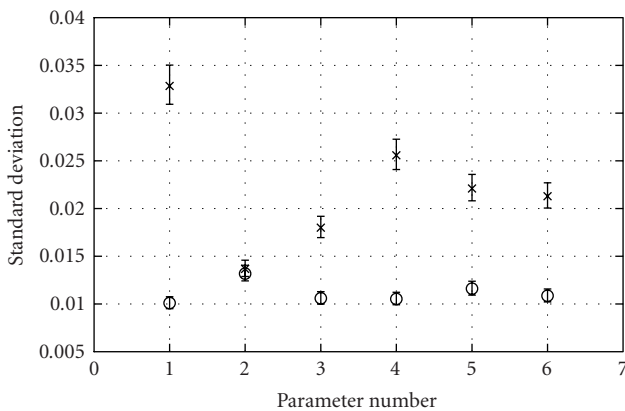

Figure 21: Mean and 95% confidence region of estimated standard deviation of the estimates for the cubic part of the Wiener-type model for a bandpass QAM-DMT signal with $N_c = 6$: optimal input signal (circles) and suboptimal input signal (crosses); the perturbation is WGN with $\sigma^2 = 1 \times 10^{-5}$ and the estimation horizon is $N_s = 28$.

Table 2: Result of the estimation comparison for optimal and suboptimal input signals of Figure 19 for the identification of the Wiener-type model of Table 1.

Mean variance for optimal signal	6.46×10^{-5}
Mean variance for suboptimal signal	2.68×10^{-4}
Mean variance gain	6.18 dB

One can draw the important conclusion that, for a signal with optimal amplitude distribution but suboptimal phase distribution, the variances of the parameter estimates can be an order of magnitude larger than for the optimal signal.

6.3. Three-step identification procedure

For nonlinear-in-the-parameters models, the Fisher information matrix is a function of the model parameters. However, to find optimal excitation signals, the unknown model parameters $\mathbf{H}$ cannot be assumed to be known. But, if there is some a priori knowledge concerning the values of the model parameters in form of a probability density function $p(\mathbf{H})$ [11, page 127], then one could optimise the criterion

$$\log E_{\mathbf{H}} \left(\Psi(\mathbf{F}(\mathbf{H})) \right), \tag{43}$$

where $E_{\mathbf{H}}(\cdot)$ denotes the expectation with respect to $\mathbf{H}$. In the proposed three-step identification procedure, the expectation operator over $\mathbf{H}$ is replaced with a point estimate $\hat{\mathbf{H}}$. Thus, the new criterion reads

$$\log \Psi(\mathbf{F}(\hat{\mathbf{H}})). \tag{44}$$

The point estimate is generated by a first parameter estimation run with the algorithm of Section 4, applying an admissible excitation signal $u_1[n]$. It was shown in the previous section that, in contrast to the linear FIR filter case, an admissible signal is not only determined by its amplitude distribution. For the Wiener-type model, a heuristic to find such admissible input signals without knowledge of the parameter values has not been found and remains an open research issue.

Following the above considerations, a three-step identification procedure is introduced.

(1) Given a fixed estimation horizon, determine a preliminary estimate $\hat{\mathbf{H}}$ of the model parameters using an admissible DMT input signal $u_1[n]$.

(2) Use the estimate $\hat{\mathbf{H}}$ to find the optimal DMT input signal $u_2[n]$, using an optimality criterion on the Fisher information matrix, which in this work is

$$\max_{\mathbf{p}, \boldsymbol{\varphi}} \log \left| \mathbf{F}(\mathbf{p}, \boldsymbol{\varphi}, \hat{H}) \right|. \tag{45}$$

(3) Perform a second estimation of the model parameters using the concatenation of the admissible DMT signal of step (1) $u_1[n]$ and the optimal DMT signal from step (2) $u_2[n]$.

An illustration of this procedure is given in Figure 22. From this block diagram, it becomes clear that one could iterate this procedure of preliminary estimation and signal optimisation several times such that the signal found using the intermediate parameter estimate converges to the optimal signal found using the true parameters.

This general procedure is subsequently exemplified with the same problem as in Section 6.2.2. We assume that the data transmission application allows an estimation horizon of $N_s = 56$, that is, two periods of the considered $N_c = 6$ bandpass QAM-DMT signals in Figure 19. To get the parameter estimates for the Wiener-type model of Table 1, one can either apply two periods of the admissible, suboptimal signal in Figure 19 to the Wiener-type model or one could use the introduced three-step identification procedure. The
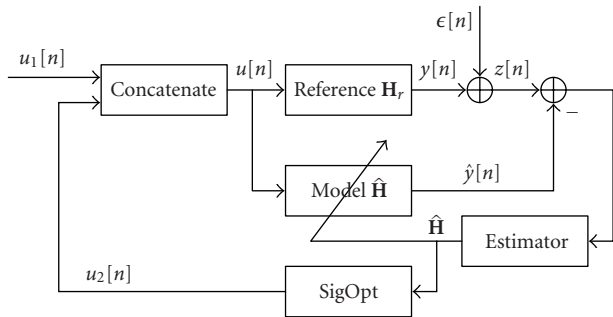
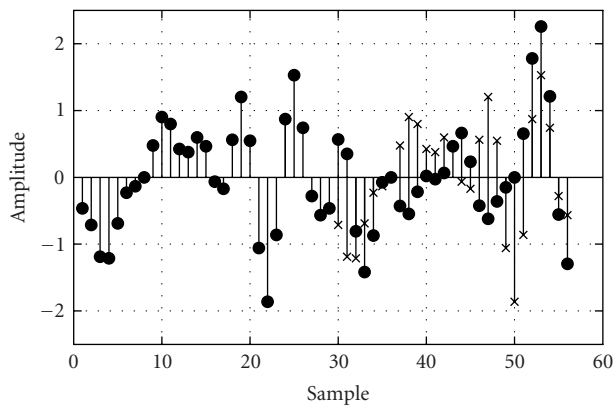FIGURE 22: Block diagram of the three-step identification procedure.

FIGURE 23: The $N_s = 56$ QAM-DMT input signals for the identification of the Wiener-type model, obtained via the three-step procedure (circles) and two periods of the suboptimal input signal (crosses).

$N_s = 56$ discrete-time input signal obtained by the three-step identification procedure is depicted in Figure 23. In addition, the two periods of the suboptimal input signal are also shown. When looking at the second period of the signals in Figure 23, one recognises that the optimal input signal found using the preliminary parameter estimates coincides with the optimal input signal of Section 6.2.2 found using the true parameters (cf. Figure 19).

The performance is once again evaluated by repeated identification of the reference Wiener-type model of Table 1 over $N_r = 500$ i.i.d. realisations of the perturbation process $\epsilon[n] \sim \mathcal{N}(0, \sigma^2)$ with $\sigma^2 = 1 \times 10^{-5}$ with $N_s = 56$. The resulting standard deviations of the estimates for the two excitation signals are shown in Figures 24 and 25 for the linear and cubic parts of the Wiener-type model, respectively.

The estimated variances for the two signals averaged over all parameters are given in Table 3. The following remarks can be made.

(1) One observes that the variance gain in Table 3 is larger than the gain in Table 2 obtained by the exact method of Section 6.2.2. An explanation of this counterintuitive effect is that the concatenation of two periods of one signal is just a scaling of the Cramer-Rao bound for one period by 1/2,
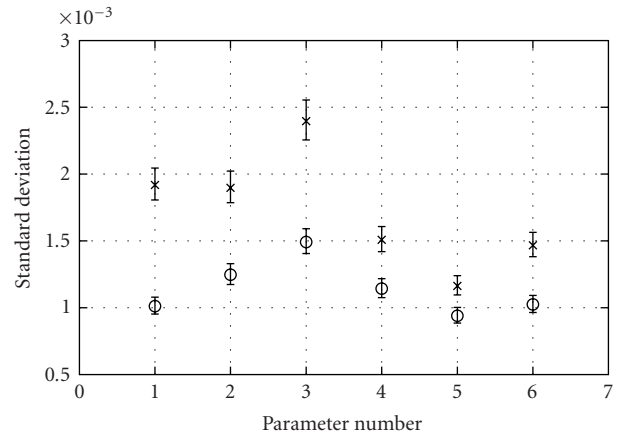

FIGURE 24: Mean and 95% confidence region of estimated standard deviations of the estimators for the linear part of the Wiener-type model for a bandpass QAM-DMT signal with $N_c = 6$: optimal input signal via the three-step procedure (circles) and suboptimal input signal (crosses); the perturbation is WGN with $\sigma^2 = 1 \times 10^{-5}$ and the estimation horizon is $N_s = 56$.
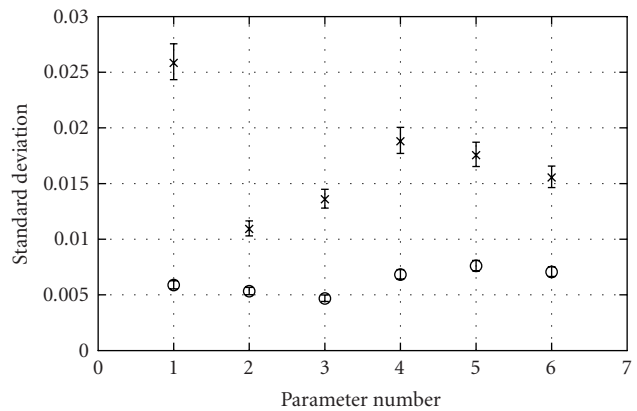

FIGURE 25: Mean and 95% confidence region of estimated standard deviations of the estimates for the cubic part of the Wiener-type model for a bandpass QAM-DMT signal with $N_c = 6$: optimal input signal via the three-step procedure (circles) and suboptimal input signal (crosses); the perturbation is WGN with $\sigma^2 = 1 \times 10^{-5}$ and the estimation horizon is $N_s = 56$.

while the concatenation of two periods of two distinct signals impacts the Cramer-Rao bound in a more complicated way. Thus, even if one applies two periods of the optimal input signal of Section 6.2.2, the obtained mean variance turns out to be 3.16×10^{-5}, which is still higher than the mean variance obtained via the three-step procedure of Table 3.

(2) For weakly nonlinear analog circuits, such as the considered line-driver circuit, reasonably accurate model parameter estimates $\hat{\mathbf{H}}$ can already be obtained when the circuit is simulated using an analog circuit simulator such as Spice. In such situations, the signal optimisation applying the preliminary parameter estimates of the three-step procedure is not required to run online, but can be precomputed offline.

Table 3: Result of the estimation comparison for the three-step input signal and for the suboptimal input signal of Figure 23 for the identification of the Wiener-type model of Table 1.

Mean variance, three-step	2.06×10^{-5}
Mean variance, suboptimal	1.58×10^{-4}
Mean variance gain	8.85 dB

7. CONCLUSION

The Cramer-Rao bound for a Wiener-type nonlinear model has been derived. The parameter estimation algorithm maximises the likelihood function using a standard Newton-like algorithm. Signal optimisation based on the Fisher information matrix is introduced for the Wiener-type model and the selection of optimal DMT signals for linear FIR filter estimation using a finite number of tones is reviewed.

For the Wiener-type model, it turned out that in contrast to the linear FIR filter case, the amplitude as well as the phase distribution of the DMT excitation signal are of vital importance. A three-step procedure to obtain parameter estimates with the lowest possible variance is outlined, even for models which are nonlinear in the parameters. Good agreement between theory and simulation is shown.

REFERENCES

[1] S. M. Kay, *Fundamentals of Statistical Signal Processing: Estimation Theory*, Prentice Hall, Upper Saddle River, NJ, USA, 1993.

[2] J. Schoukens and R. Pintelon, *Identification of Linear Systems—A Practical Guideline for Accurate Modeling*, Pergamon Press, Oxford, UK, 1991.

[3] H. Koeppl and G. Paoli, "Non-linear modeling of a broadband SLIC for ADSL-Lite-over-POTS using harmonic analysis," in *Proc. IEEE Int. Symp. Circuits and Systems (ISCAS '02)*, vol. 2, pp. 133–136, Phoenix-Scottsdale, Ariz, USA, May 2002.

[4] L. Ljung, *System Identification: Theory for the User*, Information and System Sciences Series. Prentice Hall, Upper Saddle River, NJ, USA, 2nd edition, 1999.

[5] A. E. Nordsjo, "Cramer-Rao bounds for a class of systems described by Wiener and Hammerstein models," *International Journal of Control*, vol. 68, no. 5, pp. 1067–1084, 1997.

[6] S. Kullback, *Information Theory and Statistics*, Dover Publications, New York, NY, USA, 1997.

[7] R. Fletcher, *Practical Methods of Optimization*, John Wiley & Sons, New York, NY, USA, 2nd edition, 1987.

[8] The MathWorks, *Optimization Toolbox User's Guide*, Natick, Mass, USA, 2nd edition, 2002.

[9] B. Zojer, R. Koban, J. Pichler, and G. Paoli, "A broadband high-voltage SLIC for a splitter- and transformerless combined ADSL-Lite/POTS linecard," *IEEE Journal of Solid-State Circuits*, vol. 35, no. 12, pp. 1976–1987, 2000.

[10] C. L. Nikias and A. P. Petropulu, *Higher-Order Spectra Analysis: A Nonlinear Signal Processing Framework*, Prentice Hall Signal Processing Series. Prentice-Hall, Englewood Cliffs, NJ, USA, 1993.

[11] G. C. Goodwin and R. L. Payne, *Dynamic System Identification: Experiment Design and Data Analysis*, vol. 136 of *Mathematics in Science and Engineering*, Academic Press, New York, NY, USA, 1977.

[12] W. J. Conover, *Practical Nonparametric Statistics*, John Wiley & Sons, New York, NY, USA, 3rd edition, 1998.

[13] V. V. Fedorov and P. Hackl, *Model-Oriented Design of Experiments*, vol. 125 of *Lecture Notes in Statistics*, Springer-Verlag, New York, NY, USA, 1997.

[14] C. D. Meyer, *Matrix Analysis and Applied Linear Algebra*, Society for Industrial and Applied Mathematics, Philadelphia, Pa, USA, 2000.

[15] G. B. Javorzky, I. Kollar, L. Vandenberghe, S. Boyd, and S. P. Wu, "Optimal excitation signal design for frequency domain system identification using semidefinite programming," in *Proc. 8th IMEKO TC4 Symposium on Recent Advances in Electrical Measurements*, pp. 192–197, Budapest, Hungary, September 1996.

[16] S.-P. Wu and S. Boyd, "sdpsol: A parser/solver for semidefinite programs with matrix structure," in *Recent Advances in LMI Methods for Control*, L. El Ghaoui and S.-I. Niculescu, Eds., Chapter 4, pp. 79–91, Society for Industrial and Applied Mathematics, Philadelphia, Pa, USA, 2000.

H. Koeppl was born in Friesach, Austria, in 1975. He received his M.S. degree in Physics from the Institute of Theoretical Physics, Graz University of Technology (TU Graz), in 2001. He currently works toward his Ph.D. degree in electrical engineering at the Christian Doppler Laboratory for Nonlinear Signal Processing, TU Graz, Austria, in the area of nonlinear system identification for mixed signal processing applications. His interests include Volterra models and adaptive systems. He has authored peer-reviewed publications and two patents.

A. S. Josan was born in Ludhiana, India, in 1982. He is currently studying for his B. Tech. degree in electronics and communication engineering at Indian Institute of Technology Guwahati. During May–July 2003, he worked as a Research Trainee at the Christian Doppler Laboratory for Nonlinear Signal Processing, Graz University of Technology (TU Graz), Austria. His interests include signal processing and its application to communications.

G. Paoli was born in Bregenz, Austria, in 1966. He received the Dipl.-Ing. (M.S.) degree and the Ph.D. degree in theoretical electrical engineering from the Graz University of Technology (TU Graz) in 1994 and 1998, respectively. Writing his diploma thesis on plane electromagnetic waves in multilayered media, his main focus during his research assistantship at the IGTE, TU Graz, was on computing electromagnetics and theory. In 1998, Dr. Paoli joined Infineon, starting as an Analog Designer for broadband high-voltage Telecom ICs. Since 1999, he is with the Concept Engineering Group, working on ADSL systems. He is responsible for nonlinear system identification and modeling of analog integrated circuitries. G. Paoli is the author of several journal and conference papers.

G. Kubin was born in Vienna, Austria, on June 24, 1960. He received his Dipl.-Ing. and Dr. Techn. (sub auspiciis praesidentis) degrees in electrical engineering from Vienna University of Technology (TU Vienna) in 1982 and 1990, respectively. He is a Professor of nonlinear signal processing and Head of the Signal Processing and Speech Communication Laboratory (SPSC) at Graz University of Technology (TU Graz), Austria, since September 2000. Earlier international appointments include CERN, Geneva, in 1980; TU Vienna from 1983 till 2000; Erwin Schroedinger Fellowship at Philips Research Laboratories, Eindhoven, the Netherlands, in 1985; AT&T Bell Labs, Murray Hill, USA, in 1992–1993 and 1995; KTH Stockholm, Sweden, in 1998; Vienna Telecommunications Research Centre (FTW) from 1999 up till now (Key Researcher and Member of the Board); Global IP Sound, Sweden and USA (Scientific Consultant), in 2000–2001; Christian Doppler Laboratory for Nonlinear Signal Processing from 2002 up till now (Founding Director). Dr. Kubin is a Member of the Board of Austrian Acoustics Association and Vice Chair for the European COST Action 277, nonlinear speech processing. He has authored or coauthored over eighty peer-reviewed publications and three patents.