Undermodeling-Error Quantification for Quadratically Nonlinear System Identification in the Short-Time Fourier Transform Domain

Yekutiel Avargel, *Member, IEEE*, and Israel Cohen, *Senior Member, IEEE*

***Abstract*—In this paper, we introduce an estimation error analysis for quadratically nonlinear system identification in the short-time Fourier transform (STFT) domain. The identification scheme consists of a parallel connection of a linear component, represented by crossband filters between subbands, and a quadratic component, which is modeled by multiplicative cross-terms. We mainly concentrate on two types of undermodeling errors. The first is caused by employing a purely linear model in the estimation process (i.e., *nonlinear undermodeling*), and the second is a consequence of restricting the number of estimated crossband filters in the linear component. We derive analytical relations between the noise level, nonlinearity strength, and the obtainable mean-square error (mse) in subbands. We show that for low signal-to-noise ratio (SNR) conditions, a lower mse is achieved by allowing for nonlinear undermodeling and utilizing a purely linear model. However, as the SNR increases, the performance can be generally improved by incorporating a nonlinear component into the model. We further show that as the SNR increases, a larger number of crossband filters should be estimated to attain a lower mse, whether a linear or nonlinear model is employed. Experimental results support the theoretical derivations.**

***Index Terms*—Nonlinear systems, Volterra filters, system identification, subband filtering, nonlinear undermodeling, short-time Fourier transform, time-frequency analysis.**

I. Introduction

NONLINEAR system identification has recently attracted great interest in many applications, including acoustic echo cancellation [1]–[4], channel equalization [5]–[7], biological system modeling [8] and image processing [9]. Volterra filters [10]–[15] have been applied for representing a wide range of real-world systems due to their structural generality and versatile modeling capabilities (e.g., [16]–[19]). Traditionally, Volterra-based approaches have been carried out in the time or frequency domains. Time-domain approaches employ conventional linear estimation methods in batch or adaptive forms in order to estimate the Volterra kernels. These approaches, however, often suffer from extremely high computational cost due to the large number of parameters of the Volterra model, especially for long-memory systems [14], [20]. The high complexity of the model together with its severe ill-conditioning, lead to a slow convergence of the adaptive Volterra filter [2], [19]. To ease the computational burden, frequency-domain methods have been introduced [21]–[23]. A discrete frequency-domain model, which approximates the Volterra filter using multiplicative terms, is defined in [22] and [23]. A major limitation of this model is its underlying assumption that the observation data length is relatively large. When the data is of limited size (or when the nonlinear system is not time-invariant), this long duration assumption is very restrictive. Other frequency-domain approaches use cumulants and polyspectra information to estimate the Volterra transfer functions [21]. Although computationally efficient, these approaches often assume a Gaussian input signal, which limits their applicability.

The aforementioned drawbacks of the conventional time- and frequency-domain methods motivate the use of subband (multirate) techniques [24] for improved nonlinear system identification. Such techniques have been successfully applied for identifying linear systems with relatively long impulse responses [25]–[32]. Computational efficiency as well as improved convergence rate can then be achieved due to processing in distinct subbands. Recently, we have proposed nonlinear system identification in the short-time Fourier transform (STFT) domain, based on a time-frequency representation of Volterra filters [33]. We introduced approximate nonlinear STFT models, which consist of a parallel connection of linear and nonlinear components. The linear component is represented by crossband filters between the subbands [25]–[27], while the nonlinear component is modeled by multiplicative cross-terms. We showed that a significant reduction in computational cost as well as a substantial improvement in estimation accuracy can be achieved over time-domain Volterra filters, particularly for long-memory nonlinear systems. Based on this approach, an adaptive estimation algorithm and a detailed convergence analysis of the adaptation process have been recently introduced in [34].

In this paper, we analyze the performance of the model proposed in [33] in estimating quadratically nonlinear systems in the STFT domain using an *offline* identification scheme. We employ a least-squares (LS) criterion for estimating the model parameters and derive explicit expressions for the obtainable mean-square error (mse) in each frequency bin. We mainly concentrate on the error that arises due to undermodeling; that is,

Manuscript received February 07, 2010; accepted August 03, 2010. Date of publication August 19, 2010; date of current version November 17, 2010. The associate editor coordinating the review of this manuscript and approving it for publication was Prof. Maciej Niedzwiecki.

The authors are with the Department of Electrical Engineering, Technion—Israel Institute of Technology, Technion City, Haifa 32000, Israel (e-mail: kutiav@tx.technion.ac.il; icohen@ee.technion.ac.il).

Digital Object Identifier 10.1109/TSP.2010.2068296

when the proposed model does not admit an exact description of the true system. Two types of undermodeling errors are considered. The first is attributable to employing a purely linear model for nonlinear system estimation, which is generally referred to as *nonlinear undermodeling*. This undermodeling has been examined recently in time and frequency domains [35]–[37]. The quantification of this error is of major importance since in many cases a purely linear model is fitted to the data, even though the system is nonlinear (e.g., employing a linear adaptive filter in acoustic echo cancellation applications [38]). The second undermodeling considered in this paper is a consequence of restricting the number of crossband filters in the linear component of the model, such that not all the filters are estimated in each frequency bin. The influence of this undermodeling has been recently investigated for *linear* system identification in the STFT domain [26], [27]. It was shown that the inclusion of more crossband filters in the identification process is preferable only when high signal-to-noise ratio (SNR) or long data are considered.

The analysis in this paper reveals important relations between the undermodeling errors, the noise level and the nonlinear-to-linear ratio (NLR), which represents the power ratio of nonlinear to linear components of the system. Specifically, we show that the inclusion of a nonlinear component in the model is not always preferable. The choice of the model structure (either linear or nonlinear) depends on the noise level and the observable data length. The data length is restricted to enable tracking capability during time variations in the system. We show that for low SNR conditions and rapidly time-varying systems (which restricts the length of the data), a lower mse can be achieved by allowing for nonlinear undermodeling and employing a purely linear model in the estimation process. On the other hand, as the SNR increases or as the time variations in the system become slower (which enables to use longer data), the performance can be generally improved by incorporating a nonlinear component into the model. This improvement in performance becomes larger when increasing the NLR. Moreover, we show that as the nonlinearity becomes weaker (i.e., the NLR decreases), higher SNR should be considered to justify the inclusion of the nonlinear component in the model. Concerning undermodeling in the linear component, we show that similarly to linear system identification [26], the number of crossband filters that should be estimated to attain the minimal mse (mmse) increases as the SNR increases, whether a linear or a nonlinear model is employed. For every noise level there exists an optimal number of useful crossband filters, so increasing the number of estimated crossband filters does not necessarily imply a lower mse. Experimental results demonstrate the theoretical results derived in this paper.

The paper is organized as follows. In Section II, we consider the identification of quadratically nonlinear systems in the STFT domain and formulate an LS optimization criterion for estimating the parameters of the nonlinear STFT model. In Section III, we derive explicit expressions for the mse in subbands using either a linear or a nonlinear model. In Section IV, we analyze the error expressions and investigate the influence of nonlinear undermodeling and the number of estimated crossband filters on the mse performance. Finally, in Section V, we present some experimental results to support the theoretical derivations.

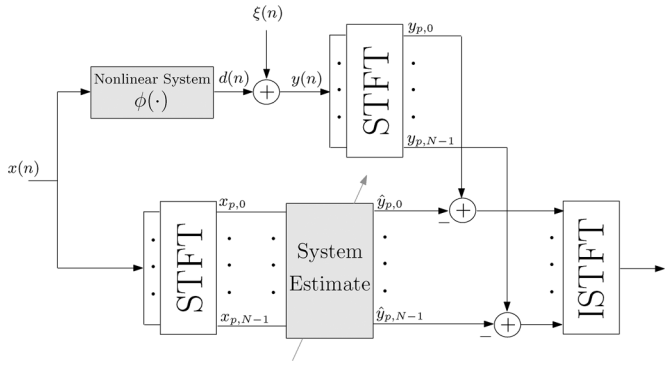


Fig. 1. Nonlinear system identification in the STFT domain. The unknown time-domain nonlinear system $\phi(\cdot)$ is estimated using a given model in the STFT domain.

II. Quadratically Nonlinear System Identification

In this section, we consider an offline scheme for the identification of quadratically nonlinear systems in the STFT domain using an LS optimization criterion for each frequency bin. We assume that the system to be identified can be represented by the nonlinear STFT model proposed in [33]. Although this assumption may not necessarily be valid in real applications, it enables the derivation of explicit expressions for the undermodeling estimation error (see Section III). It should be noted, however, that the derived insights may also be useful when considering more complicated systems, as will be demonstrated later in Section V. Throughout this paper, scalar variables are written with lowercase letters and vectors are indicated with lowercase boldface letters. Capital boldface letters are used for matrices and norms are always ℓ_2 norms.

Consider the STFT-based system identification scheme as illustrated in Fig. 1. The input signal $x(n)$ passes through an unknown quadratic time-invariant system $\phi(\cdot)$, yielding the clean output signal $d(n)$. Together with a corrupting noise signal $\xi(n)$, the system output signal is given by

$$y(n) = \{\phi x\}(n) + \xi(n) = d(n) + \xi(n). \tag{1}$$

The STFT of $y(n)$ is given by [39]

$$y_{p,k} = \sum_n y(n)\,\tilde{\psi}^*_{p,k}(n) = d_{p,k} + \xi_{p,k} \tag{2}$$

where $\tilde{\psi}_{p,k}(n) = \tilde{\psi}(n - pL)\,e^{j(2\pi/N)k(n-pL)}$ denotes a translated and modulated window function, $\tilde{\psi}(n)$ is a real-valued analysis window of length N, p is the frame index, k represents the frequency-bin index $(0 \le k \le N-1)$, L is the translation factor and * denotes complex conjugation. According to the model proposed in [33], the true system is formed as a parallel connection of linear and quadratic components in the time-frequency domain as follows:

$$d_{p,k} = \sum_{k'=0}^{N-1}\sum_{p'=0}^{M-1} x_{p-p',k'} h_{p',k,k'} + \sum_{k'\in\mathcal{F}} x_{p,k'} x_{p,(k-k') \bmod N} c_{k',(k-k') \bmod N} \tag{3}$$

where $h_{p,k,k'}$ denotes the true crossband filter of length M from frequency bin k' to frequency bin k, $c_{k',(k-k') \bmod N}$ is the true quadratic cross-term, and $\mathcal{F} = \{0, 1, \ldots \lfloor k/2 \rfloor, k+1, \ldots, k+1+\lfloor (N-k-2)/2 \rfloor\}$. It was shown [33] that the nonlinear model (3) is more advantageous than the time-domain Volterra model in representing nonlinear systems with relatively long memory (such as in nonlinear acoustic echo cancellation applications). In particular, for relatively high SNR conditions, a substantial improvement of approximately 15–20 dB in the mse is achieved by the proposed model relative to that obtained by the Volterra model.

Let $\mathbf{h}_k$ be the N crossband filters of the true system at frequency bin k

$$\mathbf{h}_k = [\, \mathbf{h}_{k,0}^T \quad \mathbf{h}_{k,1}^T \quad \cdots \quad \mathbf{h}_{k,N-1}^T \,]^T \tag{4}$$

where $\mathbf{h}_{k,k'} = [h_{0,k,k'}\, h_{1,k,k'} \cdots h_{M-1,k,k'}]^T$ is the crossband filter from frequency bin k' to frequency bin k. Let $\mathbf{X}_k$ denote an $P \times M$ Toeplitz matrix whose (m, ℓ)th term is given by

$$(\mathbf{X}_k)_{m,\ell} = x_{m-\ell,k} \tag{5}$$

where P is the observable data length in the STFT domain (i.e., the length of a time-trajectory of $y_{p,k}$ at frequency bin k), and let $\mathbf{\Delta}$ be a concatenation of $\{\mathbf{X}_k\}_{k=0}^{N-1}$ along the column dimension

$$\mathbf{\Delta} = [\, \mathbf{X}_0 \quad \mathbf{X}_1 \quad \cdots \quad \cdots \quad \mathbf{X}_{N-1} \,]. \tag{6}$$

For notational simplicity, let us assume that k and N are both even, such that according to (3), the number of quadratic cross-terms in each frequency bin is $N/2+1$. Let

$$\mathbf{c}_k = \left[\, c_{0,k} \quad \cdots \quad c_{\frac{k}{2},\frac{k}{2}} \quad c_{k+1,N-1} \quad \cdots \quad c_{\frac{N+k}{2},\frac{N+k}{2}} \,\right]^T \tag{7}$$

denote the quadratic cross-terms at the kth frequency bin, and let

$$\mathbf{\Lambda}_k = \left[\mathbf{x}_{0,k} \quad \cdots \quad \mathbf{x}_{\frac{k}{2},\frac{k}{2}} \quad \mathbf{x}_{k+1,N-1} \quad \cdots \quad \mathbf{x}_{\frac{N+k}{2},\frac{N+k}{2}}\right] \tag{8}$$

be an $P \times (N/2+1)$ matrix, where

$$\mathbf{x}_{k,k'} = [\, x_{0,k}x_{0,k'} \quad x_{1,k}x_{1,k'} \quad \cdots \quad x_{P-1,k}x_{P-1,k'} \,]^T$$

is a term-by-term multiplication of the time-trajectories of $x_{p,k}$ at frequency bins k and k', respectively. Then, (2)–(3) can be written in a vector form as

$$\mathbf{y}_k = \mathbf{d}_k + \boldsymbol{\xi}_k \tag{9a}$$

$$\mathbf{d}_k = \mathbf{\Delta}\mathbf{h}_k + \mathbf{\Lambda}_k\mathbf{c}_k \tag{9b}$$

where $\mathbf{y}_k = [\, y_{0,k} \quad y_{1,k} \quad \cdots \quad y_{P-1,k} \,]^T$ is the observable data vector, and $\mathbf{d}_k$ and $\boldsymbol{\xi}_k$ are defined similarly.

Given an input signal $x(n)$ and noisy observation $y(n)$, the goal in system identification in the STFT domain is to construct a model for describing the input-output relationship, and to select its parameters so that the model output $\hat{y}_{p,k}$ best estimates (or predicts) the measured output signal in the STFT domain. To do so, we employ the model in (3) for the estimation process, but with the use of only $2K+1$ crossband filters. The value of K controls the undermodeling in the linear component of the model by restricting the number of crossband filters. Denoting by $\bar{h}_{p,k,k'}$ and $\bar{c}_{k',(k-k') \bmod N}$ the crossband filters and quadratic cross-terms of the model, the resulting estimate $\hat{y}_{p,k}$ can be written as

$$\hat{y}_{p,k} = \sum_{k'=k-K}^{k+K} \sum_{p'=0}^{M-1} x_{p-p',k' \bmod N} \bar{h}_{p',k,k' \bmod N} + \gamma \sum_{k' \in \mathcal{F}} x_{p,k'} x_{p,(k-k') \bmod N} \bar{c}_{k',(k-k') \bmod N} \tag{10}$$

where the parameter $\gamma \in \{0, 1\}$ controls the nonlinear undermodeling by determining whether the nonlinear component is included in the model. By setting $\gamma = 0$, the nonlinearity is ignored and a purely linear model is fitted to the data, which may degrade the system estimate accuracy. The error caused by nonlinear undermodeling has been studied recently [35]–[37], assuming a certain model for nonlinearity (in the time or frequency domains). In this paper, this error is evaluated in the STFT domain by controlling the value of γ. The influence of the parameters K and γ on the mean-square performance is investigated in Section IV.

Let $\bar{\mathbf{h}}_k$ be the $2K+1$ filters of the model at frequency bin k

$$\bar{\mathbf{h}}_k = \left[\, \bar{\mathbf{h}}_{k,(k-K) \bmod N}^T \quad \cdots \quad \bar{\mathbf{h}}_{k,(k+K) \bmod N}^T \,\right]^T \tag{11}$$

where $\bar{\mathbf{h}}_{k,k'}$ is the crossband filter from frequency bin k' to frequency bin k, and let $\mathbf{\Delta}_k$ be a concatenation of $\{\mathbf{X}_{k'}\}_{k'=(k-K) \bmod N}^{(k+K) \bmod N}$ along the column dimension, i.e.,

$$\mathbf{\Delta}_k = [\, \mathbf{X}_{(k-K) \bmod N} \quad \cdots \quad \mathbf{X}_{(k+K) \bmod N} \,]. \tag{12}$$

Denoting the vector of the model's cross-terms by $\bar{\mathbf{c}}_k$, similarly to (7), the output signal estimate (10) can be written in a vector form as

$$\hat{\mathbf{y}}_{\gamma k}(\boldsymbol{\theta}_k) = \mathbf{\Delta}_k\bar{\mathbf{h}}_k + \gamma\mathbf{\Lambda}_k\bar{\mathbf{c}}_k \triangleq \mathbf{R}_{\gamma k}\boldsymbol{\theta}_k \tag{13}$$

where $\mathbf{\Lambda}_k$ was defined in (8), $\boldsymbol{\theta}_k = [\bar{\mathbf{h}}_k^T \bar{\mathbf{c}}_k^T]^T$ is the model parameters vector, $\hat{\mathbf{y}}_{\gamma k}(\boldsymbol{\theta}_k) = [\, \hat{y}_{0,k} \quad \hat{y}_{1,k} \quad \cdots \quad \hat{y}_{P-1,k} \,]^T$ is the resulting estimate associated with the parameter vector $\boldsymbol{\theta}_k$, and $\mathbf{R}_{\gamma k}$ is defined by

$$\mathbf{R}_{\gamma k} = [\mathbf{\Delta}_k \; \gamma\mathbf{\Lambda}_k]. \tag{14}$$

The subscript γ in $\hat{\mathbf{y}}_{\gamma k}(\boldsymbol{\theta}_k)$ indicates the dependence of the output signal estimate on the model structure, which can be either linear or nonlinear. Finally, using the above notations, the LS estimate of the model parameters at the kth frequency bin is given by

$$\hat{\boldsymbol{\theta}}_{\gamma k} = \arg\min_{\boldsymbol{\theta}_k} ||\mathbf{y}_k - \mathbf{R}_{\gamma k}\boldsymbol{\theta}_k||^2 = \left(\mathbf{R}_{\gamma k}^H\mathbf{R}_{\gamma k}\right)^{-1}\mathbf{R}_{\gamma k}^H\mathbf{y}_k \tag{15}$$

where we assume that $\mathbf{R}_{\gamma k}^H\mathbf{R}_{\gamma k}$ is not singular.[1] Note that both $\hat{\boldsymbol{\theta}}_{\gamma k}$ and $\hat{\mathbf{y}}_{\gamma k}(\boldsymbol{\theta}_k)$ depend also on the parameter K, but for notational simplicity K has been omitted. Substituting the optimal estimate (15) into (13), we obtain the best estimate of the system output signal in the STFT domain $\hat{\mathbf{y}}_{\gamma k}(\hat{\boldsymbol{\theta}}_{\gamma k})$ in the LS sense, for given γ and k values. Our objective is to analyze the mse attainable in each frequency bin, and investigate the influence of the parameters K and γ on the mse performance.

[1] In the ill-conditioned case, when $\mathbf{R}_{\gamma k}^H\mathbf{R}_{\gamma k}$ is singular, matrix regularization is required [40].

III. MSE Analysis

In this section, we derive explicit expressions for the mse obtainable in the kth frequency bin using either a linear ($\gamma = 0$) or a nonlinear ($\gamma = 1$) model. To make the following analysis mathematically tractable we assume that $x_{p,k}$ and $\xi_{p,k}$ are zero-mean white Gaussian signals with variances σ_x^2 and σ_ξ^2, respectively. We also assume that $x_{p,k}$ is statistically independent of $\xi_{p,k}$. The Gaussian assumption of the corresponding STFT signals is often justified by a version of the central limit theorem for correlated signals [41, Theorem 4.4.2], and it underlies the design of many speech-enhancement systems [42], [43]. It should be noted that the whiteness and Gaussianity properties are assumed here only for simplicity, and that the theoretical derivations are also valid for non-Gaussian input excitation signals, as will be shown later in Section V. Note, on the contrary, that this is not the case in some frequency-based approaches (e.g., [21] and [23]), for which the Gaussian assumption is crucial, as it is employed to derive simplified expressions for the Volterra kernels that are not valid for non-Gaussian signals.

The (normalized) mse is defined by[2]

$$\epsilon_{\gamma k}(K) = \frac{1}{E_d} E\{\|\mathbf{d}_k - \hat{\mathbf{y}}_{\gamma k}(\hat{\boldsymbol{\theta}}_{\gamma k})\|^2\} \tag{16}$$

where $E_d \triangleq E\{\|\mathbf{d}_k\|^2\}$. Recall that $\epsilon_{0k}(K)$ denotes the mse obtained by using only a linear model, and $\epsilon_{1k}(K)$ is the mse achieved by incorporating also a quadratic component into the model [see (10)]. Substituting (13) and (15) into (16), the mse can be expressed as

$$\epsilon_{\gamma k}(K) = \frac{1}{E_d} E\left\{\left\|\mathbf{R}_{\gamma k}\left(\mathbf{R}_{\gamma k}^H \mathbf{R}_{\gamma k}\right)^{-1} \mathbf{R}_{\gamma k}^H \boldsymbol{\xi}_k\right\|^2\right\} + \frac{1}{E_d} E\left\{\left\|\left[\mathbf{I}_P - \mathbf{R}_{\gamma k}\left(\mathbf{R}_{\gamma k}^H \mathbf{R}_{\gamma k}\right)^{-1} \mathbf{R}_{\gamma k}^H\right] \mathbf{d}_k\right\|^2\right\} \tag{17}$$

where $\mathbf{I}_P$ is the identity matrix of size $P \times P$. Equation (17) can be rewritten as

$$\epsilon_{\gamma k}(K) = 1 + \frac{\epsilon_1 - \epsilon_2}{E_d} \tag{18}$$

where

$$\epsilon_1 = E\left\{\boldsymbol{\xi}_k^H \mathbf{R}_{\gamma k}\left(\mathbf{R}_{\gamma k}^H \mathbf{R}_{\gamma k}\right)^{-1} \mathbf{R}_{\gamma k}^H \boldsymbol{\xi}_k\right\} \tag{19}$$

and

$$\epsilon_2 = E\left\{\mathbf{d}_k^H \mathbf{R}_{\gamma k}\left(\mathbf{R}_{\gamma k}^H \mathbf{R}_{\gamma k}\right)^{-1} \mathbf{R}_{\gamma k}^H \mathbf{d}_k\right\}. \tag{20}$$

To proceed with the mean-square analysis, we derive simplified expressions for ϵ_1 and ϵ_2. Recall that for any two vectors $\mathbf{a}$ and $\mathbf{b}$ we have $\mathbf{a}^H\mathbf{b} = \mathrm{tr}(\mathbf{a}\mathbf{b}^H)^*$, where the operator $\mathrm{tr}(\cdot)$ denotes the trace of a matrix. Then ϵ_1 can be expressed as

$$\epsilon_1 = \mathrm{tr}\left(E\left\{\boldsymbol{\xi}_k \boldsymbol{\xi}_k^H\right\} E\left\{\mathbf{R}_{\gamma k}\left(\mathbf{R}_{\gamma k}^H \mathbf{R}_{\gamma k}\right)^{-1} \mathbf{R}_{\gamma k}^H\right\}\right)^*. \tag{21}$$

[2]To avoid the well-known *overfitting* problem [44], the mse defined in (16) measures the fit of the optimal estimate $\hat{\mathbf{y}}_{\gamma k}(\hat{\boldsymbol{\theta}}_{\gamma k})$ to the clean output signal $\mathbf{d}_k$, rather than to the measured (noisy) signal $\mathbf{y}_k$. Consequently, the growing model variability caused by increasing the number of model parameters is compensated, and a more reliable measure for the model estimation quality is achieved.

The whiteness assumption for $\xi_{p,k}$ yields $E\{\boldsymbol{\xi}_k \boldsymbol{\xi}_k^H\} = \sigma_\xi^2 \mathbf{I}_P$. Then, using the property that $\mathrm{tr}(AB) = \mathrm{tr}(BA)$ for any two matrices A and B, we have

$$\begin{aligned}\epsilon_1 &= \sigma_\xi^2 E\left\{\mathrm{tr}\left(\mathbf{R}_{\gamma k}^H \mathbf{R}_{\gamma k}\left(\mathbf{R}_{\gamma k}^H \mathbf{R}_{\gamma k}\right)^{-1}\right)^*\right\} \\ &= \sigma_\xi^2 E\left\{\mathrm{tr}\left(\mathbf{I}_{(2K+1)M+\gamma(N/2+1)}\right)^*\right\} \\ &= \sigma_\xi^2\left[(2K+1)M + \gamma\left(\frac{N}{2}+1\right)\right].\end{aligned} \tag{22}$$

To evaluate ϵ_2, we first use (14) to express the inverse of $\mathbf{R}_{\gamma k}^H \mathbf{R}_{\gamma k}$ in (20) as

$$\left(\mathbf{R}_{\gamma k}^H \mathbf{R}_{\gamma k}\right)^{-1} = \begin{bmatrix} \boldsymbol{\Delta}_k^H \boldsymbol{\Delta}_k & \gamma \boldsymbol{\Delta}_k^H \boldsymbol{\Lambda}_k \\ \gamma \boldsymbol{\Lambda}_k^H \boldsymbol{\Delta}_k & \gamma \boldsymbol{\Lambda}_k^H \boldsymbol{\Lambda}_k \end{bmatrix}^{-1}. \tag{23}$$

Let us assume that $x_{p,k}$ is an ergodic process [45], such that its frame-index average $(1/P)\sum_{p=0}^{P-1} x_{p,k} x_{p,k'} x_{p+s,k''}^*$ tends to the ensemble average $E\{x_{p,k} x_{p,k'} x_{p+s,k''}^*\}$ as $P \to \infty$ (in the mean-square sense). Accordingly, assuming a relatively large observed-data length P, the (m,ℓ)th term of $\boldsymbol{\Delta}_k^H \boldsymbol{\Lambda}_k$ can be accurately approximated by

$$\begin{aligned}\left(\boldsymbol{\Delta}_k^H \boldsymbol{\Lambda}_k\right)_{m,\ell} &= \sum_n x_{n-m \bmod M,\left(k-K+\left\lfloor \frac{m}{M} \right\rfloor\right) \bmod N}^* \\ &\quad \times x_{n,\ell_k} x_{n,(k-\ell_k) \bmod N} \\ &\approx P E\left\{x_{n-m \bmod M,\left(k-K+\left\lfloor \frac{m}{M} \right\rfloor\right) \bmod N}^*\right. \\ &\quad \left. \times x_{n,\ell_k} x_{n,(k-\ell_k) \bmod N}\right\}\end{aligned} \tag{24}$$

where $\ell_k = \ell$ if $\ell \le k/2$, and $\ell_k = \ell + k/2$ otherwise. Since odd-order moments of a zero-mean complex Gaussian process are zero [46], we get $(\boldsymbol{\Delta}_k^H \boldsymbol{\Lambda}_k)_{m,\ell} \approx 0$, and (23) reduces to

$$\left(\mathbf{R}_{\gamma k}^H \mathbf{R}_{\gamma k}\right)^{-1} = \begin{bmatrix} \left(\boldsymbol{\Delta}_k^H \boldsymbol{\Delta}_k\right)^{-1} & \mathbf{0}_{(2K+1)M \times N/2+1} \\ \mathbf{0}_{N/2+1 \times (2K+1)M} & \gamma\left(\boldsymbol{\Lambda}_k^H \boldsymbol{\Lambda}_k\right)^{-1} \end{bmatrix} \tag{25}$$

where $\mathbf{0}_{N \times M}$ is a zero matrix of size $N \times M$. Substituting (25) and (14) into (20), we obtain

$$\epsilon_2 = \epsilon_{12} + \gamma \epsilon_{22} \tag{26}$$

where

$$\epsilon_{12} = E\left\{\mathbf{d}_k^H \boldsymbol{\Delta}_k\left(\boldsymbol{\Delta}_k^H \boldsymbol{\Delta}_k\right)^{-1} \boldsymbol{\Delta}_k^H \mathbf{d}_k\right\} \tag{27}$$

$$\epsilon_{22} = E\left\{\mathbf{d}_k^H \boldsymbol{\Lambda}_k\left(\boldsymbol{\Lambda}_k^H \boldsymbol{\Lambda}_k\right)^{-1} \boldsymbol{\Lambda}_k^H \mathbf{d}_k\right\}. \tag{28}$$

We proceed with evaluating ϵ_{12} and ϵ_{22}. Using the ergodicity and whiteness properties of $x_{p,k}$, the (m,ℓ)th term of $\boldsymbol{\Delta}_k^H \boldsymbol{\Delta}_k$ can be approximated as (see [26, eq. (39)])

$$\left(\boldsymbol{\Delta}_k^H \boldsymbol{\Delta}_k\right)_{m,\ell} \approx P \sigma_x^2 \delta_{m-\ell} \tag{29}$$

where δ_n denotes the Kronecker delta function. Substituting (29), and the definition of $\mathbf{d}_k$ from (9b) into (27), we obtain

$$\epsilon_{12} = \frac{1}{P\sigma_x^2}\left[\mathbf{h}_k^H \boldsymbol{\Omega}_1 \mathbf{h}_k + 2\,\mathrm{Re}\left\{\mathbf{h}_k^H \boldsymbol{\Omega}_2 \mathbf{c}_k\right\} + \mathbf{c}_k^H \boldsymbol{\Omega}_3 \mathbf{c}_k\right] \tag{30}$$

where $\mathbf{\Omega}_1 \triangleq E\{\mathbf{\Delta}^H\mathbf{\Delta}_k\mathbf{\Delta}_k^H\mathbf{\Delta}\}$, $\mathbf{\Omega}_2 \triangleq E\{\mathbf{\Delta}^H\mathbf{\Delta}_k\mathbf{\Delta}_k^H\mathbf{\Lambda}_k\}$, $\mathbf{\Omega}_3 \triangleq E\{\mathbf{\Lambda}_k^H\mathbf{\Delta}_k\mathbf{\Delta}_k^H\mathbf{\Lambda}_k\}$, and the operator $\mathrm{Re}\{\cdot\}$ takes the real part of its argument. An explicit expression for $\mathbf{\Omega}_1$ was derived in [26] using the Gaussian fourth-order moment-factoring theorem [46], and its (m,ℓ)th term is given by [26, eq. (41)]

$$(\mathbf{\Omega}_1)_{m,\ell} = \sigma_x^4 P\delta_{m-\ell}[M(2K+1) + P\delta_{m\in\mathcal{L}_0}] \tag{31}$$

where $\mathcal{L}_0 = \{[(k-K+n_1) \bmod N]M + n_2\}$ with $n_1 \in \{0,\ldots,2K\}$ and $n_2 \in \{0,\ldots,M-1\}$. In addition, the (m,ℓ)th term of $\mathbf{\Omega}_2$ can be written as

$$\begin{aligned}(\mathbf{\Omega}_2)_{m,\ell} &= \sum_{n,r,q} E\Big\{x^*_{r-m \bmod M,\lfloor\frac{m}{M}\rfloor} x^*_{q-n \bmod M,\left(k-K+\lfloor\frac{n}{M}\rfloor\right) \bmod N} \\ &\quad \times x_{r-n \bmod M,\left(k-K+\lfloor\frac{n}{M}\rfloor\right) \bmod N} x_{q,\ell_k} x_{q,(k-\ell_k) \bmod N}\Big\} \\ &= 0 \end{aligned} \tag{32}$$

where ℓ_k is defined in (24), and the last equation is due to the definition of odd-order moments of Gaussian process. Furthermore, using the Gaussian sixth-order moment-factoring theorem [46], the (m,ℓ)th term of $\mathbf{\Omega}_3$ can be expressed as (see Appendix I)

$$(\mathbf{\Omega}_3)_{m,\ell} = \sigma_x^6 P\delta_{m-\ell}\left[M(2K+1)\left(1+\delta_{m_k\in\left\{\frac{k}{2},\frac{k+N}{2}\right\}}\right) + \sum_{i=1}^{4}\delta_{m_k\in\mathcal{L}_i}\right] \tag{33}$$

where m_k is defined similarly to ℓ_k in (24), and $\mathcal{L}_1 = \mathcal{B}\cap\mathcal{A}_k$, $\mathcal{L}_2 = \mathcal{B}\cap\mathcal{A}_0$, $\mathcal{L}_3 = \mathcal{C}\cap\mathcal{A}_k$, and $\mathcal{L}_4 = \mathcal{C}\cap\mathcal{A}_0$, with $\mathcal{A}_k \triangleq \{[(k-K+n_1) \bmod N]M | n_1 \in \{0,\ldots,2K\}\}$, $\mathcal{B} \triangleq \{k/2, (k+N)/2\}$, and $\mathcal{C} \triangleq \{[0,k/2]\cup[k+1,(k+N)/2]\}$. Substituting (31), (32), and (33) into (30), we obtain

$$\begin{aligned}\epsilon_{12} &= \sigma_x^2 M(2K+1)\|\mathbf{h}_k\|^2 + \sigma_x^2 P\sum_{m=k-K}^{k+K}\|\mathbf{h}_{k,m \bmod N}\|^2 \\ &\quad + \sigma_x^4 M(2K+1)\left(\|\mathbf{c}_k\|^2 + \left|c_{\frac{k}{2},\frac{k}{2}}\right|^2 + \left|c_{\frac{k+N}{2},\frac{k+N}{2}}\right|^2\right) \\ &\quad + \sigma_x^4\sum_{i=1}^{4}\sum_{m\in\mathcal{L}_i}|c_{m,(k-m) \bmod N}|^2. \end{aligned} \tag{34}$$

An expression for ϵ_{22} is obtained by substituting $\mathbf{d}_k$ from (9b) into (28):

$$\epsilon_{22} = \mathbf{h}_k^H\mathbf{\Theta}_1\mathbf{h}_k + 2\,\mathrm{Re}\left\{\mathbf{h}_k^H\mathbf{\Theta}_2\mathbf{c}_k\right\} + \mathbf{c}_k^H\mathbf{\Theta}_3\mathbf{c}_k \tag{35}$$

where $\mathbf{\Theta}_1 \triangleq E\{\mathbf{\Delta}^H\mathbf{\Lambda}_k(\mathbf{\Lambda}_k^H\mathbf{\Lambda}_k)^{-1}\mathbf{\Lambda}_k^H\mathbf{\Delta}\}$, $\mathbf{\Theta}_2 \triangleq E\{\mathbf{\Delta}^H\mathbf{\Lambda}_k\}$ and $\mathbf{\Theta}_3 \triangleq E\{\mathbf{\Lambda}_k^H\mathbf{\Lambda}_k\}$. Finding an explicit expression for $\mathbf{\Theta}_1$ is not straightforward. Nonetheless, using the ergodicity of $x_{p,k}$ and the Gaussian sixth-order moment-factoring theorem, we obtain after some mathematical manipulations (see Appendix II-A)

$$(\mathbf{\Theta}_1)_{m,\ell} = \sigma_x^2\delta_{m-\ell} \times\left[1+\frac{N}{2}+\delta_{m\in\left\{\frac{k}{2},\frac{N+k}{2}\right\}M}+\delta_{m\in\{0,\ldots,N-1\}M}\right]. \tag{36}$$

The (m,ℓ)th term of $\mathbf{\Theta}_2$ consists of a third-order moment of $x_{p,k}$, and as such is equal to zero. The (m,ℓ)th term of $\mathbf{\Theta}_3$ can be expressed as (see Appendix II-B)

$$(\mathbf{\Theta}_3)_{m,\ell} = \sigma_x^4 P\delta_{m-\ell}\left[1+\delta_{m\in\left\{\frac{k}{2},\frac{N}{2}\right\}}\right]. \tag{37}$$

Substituting (36) and (37) into (35), we obtain

$$\begin{aligned}\epsilon_{22} &= \sigma_x^2\left[\left(1+\frac{N}{2}\right)\|\mathbf{h}_k\|^2 + \left|h_{0,k,\frac{k}{2}}\right|^2 \right. \\ &\quad \left. + \left|h_{0,k,\frac{k+N}{2}}\right|^2 + \sum_{k'=0}^{N-1}|h_{0,k,k'}|^2\right] \\ &\quad + \sigma_x^4 P\left(\|\mathbf{c}_k\|^2 + \left|c_{\frac{k}{2},\frac{k}{2}}\right|^2 + \left|c_{\frac{k+N}{2},\frac{k+N}{2}}\right|^2\right). \end{aligned} \tag{38}$$

Finally, substituting (34) and (38) into (26), we obtain an explicit expression for ϵ_2, which together with ϵ_1 from (22) is substituted into (18) to yield

$$\begin{aligned}&\epsilon_{\gamma k}(K) \\ &= 1 + \frac{\sigma_\xi^2}{E_d}\left[M(2K+1)+\gamma\left(\frac{N}{2}+1\right)\right] \\ &\quad - \frac{\sigma_x^2}{E_d}\|\mathbf{h}_k\|^2\left[M(2K+1)+\gamma\left(\frac{N}{2}+1\right)\right] \\ &\quad - \frac{\sigma_x^4}{E_d}(M(2K+1)+\gamma P) \\ &\quad \times\left(\|\mathbf{c}_k\|^2 + \left|c_{\frac{k}{2},\frac{k}{2}}\right|^2 + \left|c_{\frac{k+N}{2},\frac{k+N}{2}}\right|^2\right) \\ &\quad - \frac{1}{E_d}\sigma_x^2 P\sum_{m=0}^{2K}\left\|\mathbf{h}_{k,(k-K+m) \bmod N}\right\|^2 \\ &\quad - \frac{\sigma_x^4}{E_d}\sum_{i=1}^{4}\sum_{m\in\mathcal{L}_i}\left|c_{m,(k-m) \bmod N}\right|^2 \\ &\quad - \frac{\gamma}{E_d}\sigma_x^2\left[\left|h_{0,k,\frac{k}{2}}\right|^2 + \left|h_{0,k,\frac{k+N}{2}}\right|^2\sum_{k'=0}^{N-1}|h_{0,k,k'}|^2\right]. \end{aligned} \tag{39}$$

Equation (39) provides an explicit expression for the mse obtained in the kth frequency bin as a function of γ, using LS estimates of $2K+1$ crossband filters and $N/2+1$ quadratic cross-terms. Next, we analyze this error expression in order to provide important insights into the system identifier performance.

IV. Discussion

In this section, we investigate the influence of nonlinear undermodeling (controlled by γ) and the number of crossband filters (controlled by K) on the mse performance, and derive explicit relations in terms of the SNR and the NLR.

Let $\sigma_d^2 = E\{|d_{p,k}|^2\}$ denote the power of the system output signal in the STFT domain. Using (3) and the whiteness property of $x_{p,k}$, σ_d^2 can be written as

$$\sigma_d^2 = \sigma_{d_L}^2 + \sigma_{d_Q}^2 \tag{40}$$

where $\sigma_{d_L}^2 = \sigma_x^2 \|\mathbf{h}_k\|^2$ and $\sigma_{d_Q}^2 = \sigma_x^4(\|\mathbf{c}_k\|^2 + |c_{k/2,k/2}|^2 + |c_{(k+N)/2,(k+N)/2}|^2)$ are the powers of the output signals of the linear and quadratic components, respectively. Note that the separable notation in (40) is possible since the linear and quadratic components of a system represented by (3) are orthogonal to each other for Gaussian inputs (analogously to the first- and second-order Volterra operators [10]). Since σ_d^2 is independent of p, we can express E_d from (16) as $E_d = \sum_{p=0}^{P-1} E\{|d_{p,k}|^2\} = P\sigma_d^2$. Then, denoting the SNR by $\eta = \sigma_d^2/\sigma_\xi^2$ and the NLR by $\rho = \sigma_{d_Q}^2/\sigma_{d_L}^2$, (39) can be rewritten as[3]

$$\epsilon_{\gamma k}(K) = \frac{\alpha_{\gamma k}(K)}{\eta} + \beta_{\gamma k}(K) \tag{41}$$

where

$$\alpha_{\gamma k}(K) \triangleq \frac{(2K+1)M}{P} + \gamma\frac{N/2+1}{P} \tag{42a}$$

$$\begin{aligned}\beta_{\gamma k}(K) \triangleq\ & 1 - \frac{(2K+1)M}{P} \\ & - \|\mathbf{h}_k\|^{-2}\left[h_1(K) + \frac{\sigma_x^2 c(K)}{P}\right]\frac{1}{1+\rho} \\ & - \gamma\left[\frac{1+N/2+\|\mathbf{h}_k\|^{-2}h_2}{P} + \rho\right]\frac{1}{1+\rho}\end{aligned} \tag{42b}$$

and $h_1(K) \triangleq \sum_{m=0}^{2K} \|\mathbf{h}_{k,(k-K+m) \bmod N}\|^2$, $h_2 \triangleq |h_{0,k,(k/2)}|^2 + |h_{0,k,(k+N/2)}|^2 + \sum_{k'=0}^{N-1} |h_{0,k,k'}|^2$ and $c(K) \triangleq \sum_{i=1}^{4} \sum_{m\in\mathcal{L}_i} |c_{m,(k-m) \bmod N}|^2$. Note that both η and ρ depend on the powers of the linear and quadratic components, and as such they may have mutual influence on each other. However, to properly analyze the error, we will assume in the following that variations in the SNR value η does not influence the value of ρ. We observe from (41) that the mse $\epsilon_{\gamma k}(K)$, for fixed values of γ, k and K, is a monotonically decreasing function of η, which expectedly indicates that a better estimation of the model parameters is enabled by increasing the SNR. Moreover, substituting $\gamma = 0$, $\rho = 0$ and $c(K) = 0$ into (41)–(42b), the mse degenerates to that derived in [26]:

$$\epsilon_{k,\text{linear}}(K) = \frac{(2K+1)M}{P}\cdot\frac{1}{\eta} + 1 - \frac{(2K+1)M}{P} - \frac{h_1(K)}{\|\mathbf{h}_k\|^2} \tag{43}$$

which represents the mse achieved by estimating a linear system with a purely linear model.

A. Influence of Nonlinear Undermodeling

From (42a) and (42b), it can be verified that $\alpha_{1k}(K) > \alpha_{0k}(K)$ and $\beta_{1k}(K) < \beta_{0k}(K)$, which implies that $\epsilon_{1k}(K) > \epsilon_{0k}(K)$ for low SNR ($\eta \ll 1$), and $\epsilon_{1k}(K) \le \epsilon_{0k}(K)$ for high SNR ($\eta \gg 1$). As a result, since $\epsilon_{1k}(K)$ and $\epsilon_{0k}(K)$ are monotonically decreasing functions of η, they must intersect at a certain SNR value, denoted by $\bar{\eta}$. Accordingly, for SNR values lower than $\bar{\eta}$, we get $\epsilon_{0k}(K) < \epsilon_{1k}(K)$, and correspondingly a lower mse is achieved by allowing for nonlinear undermodeling (i.e., employing only a linear model). On the other hand, as the SNR increases, the mse performance can be generally improved by incorporating also the nonlinear component into the model ($\gamma = 1$).

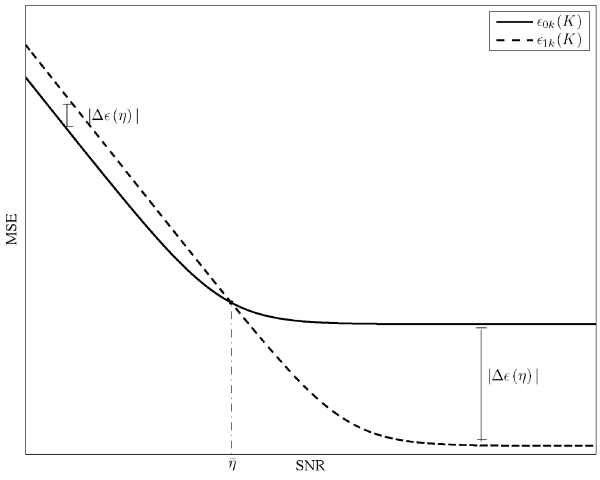
The SNR intersection point $\bar{\eta}$ is obtained by requiring that $\epsilon_{1k}(K) = \epsilon_{0k}(K)$, which yields

$$\bar{\eta} = \frac{1+\rho}{1 + 2\|\mathbf{h}_k\|^{-2}h_2(N+2)^{-1} + 2P(N+2)^{-1}\rho}. \tag{44}$$

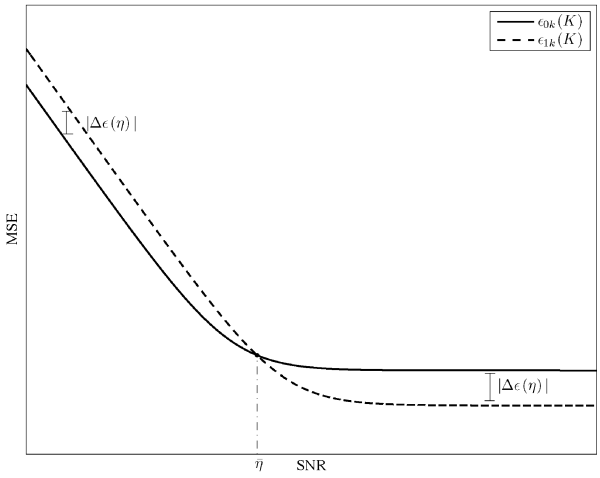
Equation (44) implies that $\bar{\eta}$ is a monotonically decreasing function of the observable data length in the STFT domain (P). Therefore, for a fixed SNR value, as more data is available in the identification process, a lower mse is achieved by estimating also the parameters of the nonlinear component. Recall that the system is assumed time invariant during P frames (its estimate is updated every P frames), in case the time variations in the system are relatively fast, we should decrease P and correspondingly allow for nonlinear undermodeling to achieve lower mse. Another interesting point that can be concluded from (44) is that $\bar{\eta}$ is a monotonically decreasing function of ρ (assuming $P > N/2+1$, which holds in our case due to the ergodicity assumption made in the previous section). Consequently, as the nonlinearity becomes weaker (i.e., ρ decreases), higher SNR values should be considered to justify the estimation of the nonlinear component.

Equations (41)–(42b) also provide an insight into the mutual influence of ρ and γ on the mse performance. Specifically for high SNR conditions, it can be verified that when a purely linear model is employed ($\gamma = 0$), the mse increases with increasing ρ [since $\beta_{0k}(K)$ increases]. On the other hand, including a nonlinear component into the model ($\gamma = 1$) decreases the mse for high SNR values, and improves the accuracy of the system estimate. This improvement in performance becomes larger as ρ increases, as the last term in $\beta_{1k}(K)$ increases with increasing ρ. This stems from the fact that the error induced by the undermodeling in the linear component (i.e., by not considering all of the crossband filters) is less substantial as the nonlinearity strength increases, such that the true system can be more accurately estimated by the full model. To summarize the above discussion, Fig. 2 shows typical mse curves of $\epsilon_{1k}(K)$ and $\epsilon_{0k}(K)$ as a function of the SNR, obtained for a high NLR ρ_1 [Fig. 2(a)] and a lower one $0.2\rho_1$ [Fig. 2(b)], where $|\Delta\epsilon(\eta)|$ denotes the nonlinear undermodeling error. Note that as the NLR ρ increases, the intersection point $\bar{\eta}$ decreases, while the undermodeling error $|\Delta\epsilon(\eta)|$ increases (for high SNR conditions).

[3]In general, η and ρ depend on the frequency-bin index k since the input-signal energy (or the true-system energy) may often not be uniformly distributed in frequency (e.g., speech signals [47]). However, for notational simplicity k has been omitted.



(a)



(b)

Fig. 2. Illustration of typical mse curves as a function of the SNR, showing the relation between $\epsilon_{0k}(K)$ (solid) and $\epsilon_{1k}(K)$ (dashed) for (a) high NLR ρ_1 and (b) low NLR $0.2\rho_1$. $|\Delta\epsilon(\eta)|$ denotes the nonlinear undermodeling error.

B. Influence of the Number of Crossband Filters

The number of estimated crossband filters in the linear component also influences the system identifier performance. It was shown in [26] and [27] that when a linear model is employed for estimating a purely linear system, the mse in subbands not necessarily decreases by increasing the number of crossband filters. The inclusion of more crossband filters in the identification process is preferable only when high SNR or long data are considered. The same applies also in our case, when the system to be identified is nonlinear. This can easily be verified from (41)–(42b), which indicate that $\epsilon_{\gamma k}(K+1) > \epsilon_{\gamma k}(K)$ for low SNR ($\eta \ll 1$), and $\epsilon_{\gamma k}(K+1) \leq \epsilon_{\gamma k}(K)$ for high SNR ($\eta \gg 1$). Therefore, for every noise level there exists an optimal number of crossband filters, which increases as the SNR increases. In the limit, for a sufficiently large SNR value and infinitely long data, we would prefer to employ a nonlinear model and to estimate all the crossband filters. This can be shown from (41)–(42b) by taking η and P to infinity, obtaining

$$\lim_{\eta,P\to\infty} \epsilon_{\gamma k}(K) = \frac{1 - h_1(K)\|\mathbf{h}_k\|^{-2}}{1+\rho} + \frac{(1-\gamma)\,\rho}{1+\rho}. \quad (45)$$

Equation (45) represents the bias error of the model, which can be decomposed into two terms. The first term is attributable to the undermodeling caused by restricting the number of crossband filters. It reduces to zero when $K = N/2$ [since then $h_1(K) = \|\mathbf{h}_k\|^2$], and is monotonically decreasing as a function of ρ. On the other hand, the second term is due to nonlinear undermodeling, and vanishes when $\gamma = 1$. This term is a monotonically decreasing function of ρ. Clearly, the asymptotic error in (45) reduces to zero when employing a nonlinear model and estimating all the crossband filters.

It is worthwhile noting that the results in this section are closely related to model-structure selection and model-order selection, which are fundamental problems in many system identification applications [44], [48]–[53]. In our case, the model structure may be either linear $(\gamma = 0)$ or nonlinear $(\gamma = 1)$, where a richer and larger structure is provided by the latter. The larger the model structure, the better the model fits to the data, at the expense of an increased variance of parametric estimates [44]. Generally, the structure to be chosen is affected by the level of noise in the data and the length of the observable data. As the SNR increases or as more data is employable, a richer structure can be used, and correspondingly a better estimation can be achieved by incorporating a nonlinear model rather than a linear one. Once a model structure has been chosen, its optimal order (i.e., the number of estimated parameters) should be selected, where in our case the model order is determined by the number of crossband filters. Accordingly, as the SNR increases, whether a linear or a nonlinear model is employed, more crossband filters should be utilized to achieve a lower mse. These points will be further demonstrated in the next section. Note that the above analysis does not concern the problem of controlling the model structure (linear or nonlinear) or the model order (number of crossband filters). Nonetheless, based on these theoretical derivations, a fully control scheme, which chooses the optimal structure and order of a model, can be construct for the estimation of real-world non-stationary systems.

V. Experimental Results

In this section, we present experimental results which support the theoretical derivations. The influence of nonlinear undermodeling on the mse performance is demonstrated by fitting both linear and nonlinear models to the observable data and comparing the resulting mse values. In order to evaluate the results under more realistic assumptions than those made in Section III, we use time-domain white Gaussian input signals as well as real speech signals. We also validate the theoretical results in the case of systems that cannot be exactly described by the STFT model, such as second-order Volterra systems. For a demonstration of the applicability of the STFT-model in a real nonlinear acoustic echo cancellation scenario, the interested reader is referred to [33].

Consider a nonlinear system of the following form:

$$y(n) = \sum_{m=0}^{N_1-1} h(m)x(n-m) + \{\mathcal{L}x\}(n) + \xi(n) \quad (46)$$

where $h(n)$ is the impulse response of the linear component, and $\{\mathcal{L}x\}(n)$ denotes the output of the nonlinear component. We model the linear impulse response as a nonstationary stochastic process with an exponential decay envelope, i.e., $h(n) = u(n)\beta(n)e^{-\alpha n}$, where $u(n)$ is the unit step function, $\beta(n)$ is a unit-variance zero-mean white Gaussian noise, and α is the decay exponent. The length of the impulse response is $N_1 = 768$, $\alpha = 0.009$, and the data contains $N_x = 24\,000$ samples. In addition, we model the input signal $x(n)$ and the additive noise signal $\xi(n)$ as uncorrelated zero-mean white Gaussian processes with variances σ_x^2 and σ_ξ^2, respectively. For evaluating the quality of the system estimate, we define the time-domain mse as

$$\epsilon_\gamma(K) = \frac{E\{|d(n) - \hat{y}_\gamma(K;n)|^2\}}{E\{|d(n)|^2\}} \tag{47}$$

where $d(n)$ is the clean output signal [i.e., $d(n) = y(n) - \xi(n)$], and $\hat{y}_\gamma(K;n)$ is the inverse STFT of the model output signal $\hat{y}_{p,k}$ [see (10)], as obtained for a given γ value, and by estimating $2K+1$ crossband filters. Finally, for the STFT, we use a Hamming analysis window of length $N = 256$ with 50% overlap (i.e., $L = 0.5N$), and a corresponding minimum-energy synthesis window that satisfies the completeness condition [54].

In the first experiment, the nonlinear component $\{\mathcal{L}x\}(n)$ from (46) is generated according to the quadratic model (3), i.e.,

$$\{\mathcal{L}x\}(n) = S^{-1}\sum_{k'\in\mathcal{F}} x_{p,k'}x_{p,(k-k') \bmod N}c_{k',(k-k') \bmod N} \tag{48}$$

where S^{-1} denotes the inverse STFT operator and $\{c_{k',(k-k') \bmod N}|k' \in \mathcal{F}\}$ are the true quadratic cross-terms of the system. These terms are modeled here as a unit-variance zero-mean white Gaussian process. Initially, a fixed value of $K = 0$ is assumed (i.e., the crossband filters are ignored and only the band-to-band filters of the model $\{\bar{h}_{p,k,k}\}_{k=0}^{N-1}$ are estimated). Fig. 3 shows the resulting mse curves $\epsilon_0(0)$ and $\epsilon_1(0)$ as a function of the SNR, as obtained for an NLR of 0 dB [Fig. 3(a)] and -20 dB [Fig. 3(b)]. The results confirm that for relatively low SNR values, a lower mse is achieved by estimating the system using a purely linear model ($\gamma = 0$) and allowing for nonlinear undermodeling. For instance, Fig. 3(a) shows that for a -20-dB SNR, employing only a linear model reduces the mse by approximately 11 dB, when compared to that achieved by using a nonlinear model ($\gamma = 1$). On the other hand, when considering high SNR values, the performance can be generally improved by incorporating a nonlinear component into the model, as expected from (41)–(42b). For an SNR of 20 dB, for instance, a nonlinear model enables a decrease of 13 dB in the mse. Furthermore, a comparison of Fig. 3(a) and (b) indicates that the SNR intersection point between the corresponding mse curves increases as we decrease the NLR [as expected from (44)]. Clearly, for high SNR conditions, as the NLR increases, the mse associated with the linear model increases, while the relative improvement achieved by the nonlinear model becomes larger. This was accurately described by the theoretical error analysis in Section IV [see Fig. 2]. It should be noted that similar results are obtained for other (fixed) values of K.

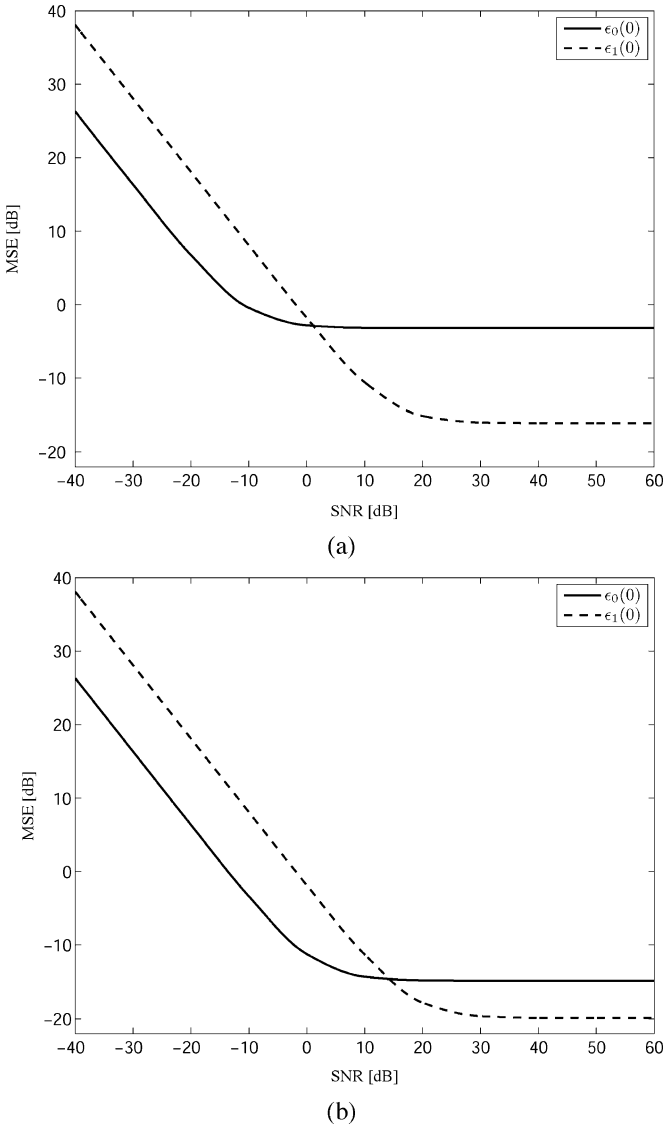
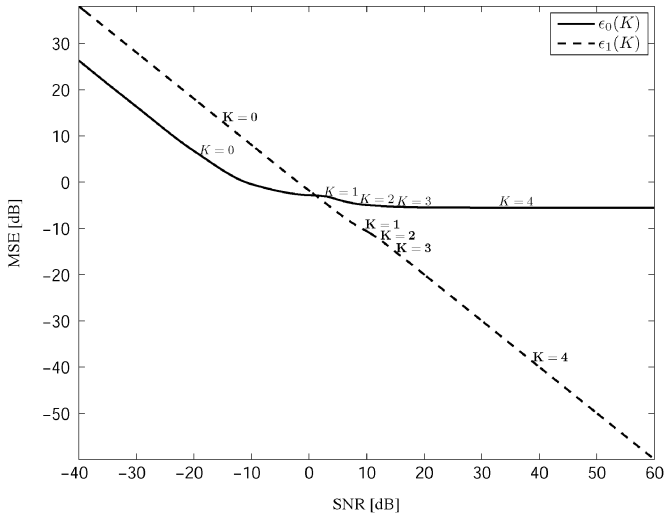
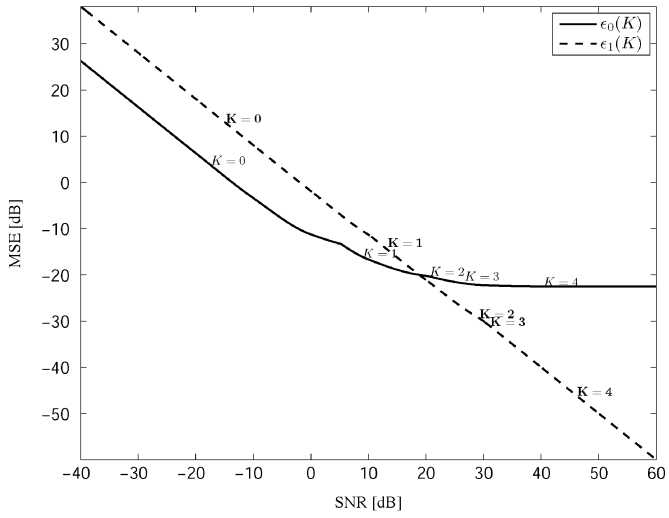


Fig. 3. MSE curves as a function of the SNR for white Gaussian signals, as obtained by the STFT model (10) using a purely linear model [$\epsilon_0(0)$; solid] and a nonlinear one [$\epsilon_1(0)$; dashed]. For both models, a fixed value of $K = 0$ is assumed for the linear component (i.e., only the band-to-band filters are estimated). The true system is formed as a combination of linear and quadratic components, where the latter is modeled according to (48). (a) Nonlinear-to-linear ratio (NLR) of 0 dB. (b) NLR of -20 dB.

Next, in order to determine the influence of the number of estimated crossband filters on the mse performance, we employ several values of K and seek for the optimal one that achieves the mmse for every SNR value. Fig. 4 shows the resulting mse curves $\epsilon_0(K)$ and $\epsilon_1(K)$ as a function of the SNR, as obtained for an NLR of 0 dB [Fig. 4(a)] and -20 dB [Fig. 4(b)]. The optimal value of K is indicated above the corresponding mse curves. Note that each K is indicated at a certain SNR value that represents the center of the segment for which the mmse is attained by that specific K. For example, $K = 0$ written above



(a)



(b)

Fig. 4. MSE curves as a function of the SNR for white Gaussian signals, as obtained by the STFT model (10) using a purely linear model [$\epsilon_0(K)$; solid] and a nonlinear one [$\epsilon_1(K)$; dashed]. The optimal value of K is indicated above the corresponding mse curves (light and dark fonts, respectively). The true system is formed as a combination of linear and quadratic components, where the latter is modeled according to (48). (a) Nonlinear-to-linear ratio (NLR) of 0 dB. (b) NLR of -20 dB.

the curve $\epsilon_0(K)$ in Fig. 4(a) indicates that for $-40 < \text{SNR} < 0$, the mmse is achieved by ignoring all crossband filters ($K = 0$). Expectedly, Fig. 4 confirms that as the SNR increases, the optimal K increases, and consequently a larger number of crossband filters should be estimated to attain the mmse, both for the linear [$\epsilon_0(K)$] and nonlinear [$\epsilon_1(K)$] models. Clearly, for high SNR conditions, the nonlinear model is considerably more advantageous. Specifically for a 30-dB SNR, Fig. 4(a) shows that a substantial decrease of 25 dB is achieved by the nonlinear model, relative to that obtained by the linear one. The mse values obtained by each value of K for a 0-dB NLR and for various SNR conditions are specified in Table I. One can observe that for an SNR value of 35 dB, for instance, a significant improvement of approximately 11 dB over a linear model with $K = 4$ is achieved by a nonlinear model with only $K = 0$, which substantially reduces the complexity of the model. Note that similar

TABLE I
MSE OBTAINED BY A LINEAR MODEL [$\epsilon_0(K)$] AND A NONLINEAR MODEL [$\epsilon_1(K)$] FOR SEVERAL K VALUES, AND UNDER VARIOUS SNR CONDITIONS. THE NONLINEAR-TO-LINEAR RATIO (NLR) IS 0 dB

K	$\epsilon_0(K)$ [dB]		$\epsilon_1(K)$ [dB]	
	SNR:-10 dB	35 dB	SNR:$= -10$ dB	35 dB
0	-0.42	-3.17	8.08	-16.06
1	2.41	-3.82	8.75	-18.78
2	3.98	-4.29	9.35	-21.54
3	5.36	-4.91	9.89	-28.59
4	6.36	-5.51	10.03	-34.96

results are obtained for a smaller NLR value [Fig. 4(b)], with the only difference is that the two curves intersect at a higher SNR value. Decreasing the NLR expectedly improves the mse achieved by the linear model at high SNR values, and correspondingly decreases the nonlinear undermodeling error.

In the second experiment, we demonstrate the applicability of the theoretical derivations in a more realistic scenario, where the STFT model does not admit an exact description of the unknown system. Specifically, the system to be identified is represented by a second-order Volterra filter [10], [14], which can be expressed by (46) with a nonlinear component of the form

$$\{\mathcal{L}x\}(n) = \sum_{m=0}^{N_2-1}\sum_{\ell=0}^{N_2-1} h_2(m,\ell)x(n-m)x(n-\ell) \qquad (49)$$

where $h_2(m,\ell)$ is the quadratic Volterra kernel and N_2 is its memory length. The quadratic kernel used in this experiment was taken from a real nonlinear acoustic echo cancellation scenario, where a loudspeaker is fed with a far-end signal $x(n)$ at high volume, thus introducing non-negligible nonlinear distortion (for further details, see [33]). In the following, in addition to the STFT model, we also evaluate the undermodeling errors in the time domain by fitting a second-order Volterra model with a reduced-size quadratic kernel. Accordingly, the resulting estimator for the system output is given by

$$\hat{y}_\gamma(n) = \sum_{m=0}^{\bar{N}_1-1} \bar{h}_1(m)x(n-m) + \gamma\sum_{m=0}^{\bar{N}_2-1}\sum_{\ell=m}^{\bar{N}_2-1} \bar{h}_2(m,\ell)x(n-m)x(n-\ell) \qquad (50)$$

where $\bar{h}_1(m)$ and $\bar{h}_2(m,\ell)$ are the linear and quadratic Volterra kernels of the model, respectively, with $\bar{N}_1$ and $\bar{N}_2 (< N_2)$ being their corresponding memory lengths. Note that for the quadratic kernel of the model in (50), the triangular Volterra representation is used [10]. As in the STFT model, the parameter $\gamma \in \{0,1\}$ in (50) controls the nonlinear undermodeling and determines whether the nonlinear component is also included in the model. The parameters of the Volterra model are estimated offline using an LS criterion (see e.g., [14] and [33]), and the resulting mse is compared to that achieved by the STFT model. We use $N_2 = 80$ for the true system (49), and $\bar{N}_1 = N_1 = 768$ and $\bar{N}_2 = 40$ for the model (50).

Fig. 5 shows the resulting mse curves as a function of the SNR for a 0 dB NLR, as obtained for a white Gaussian input signal [Fig. 5(a)] and a real speech signal [Fig. 5(b)]. For both time-

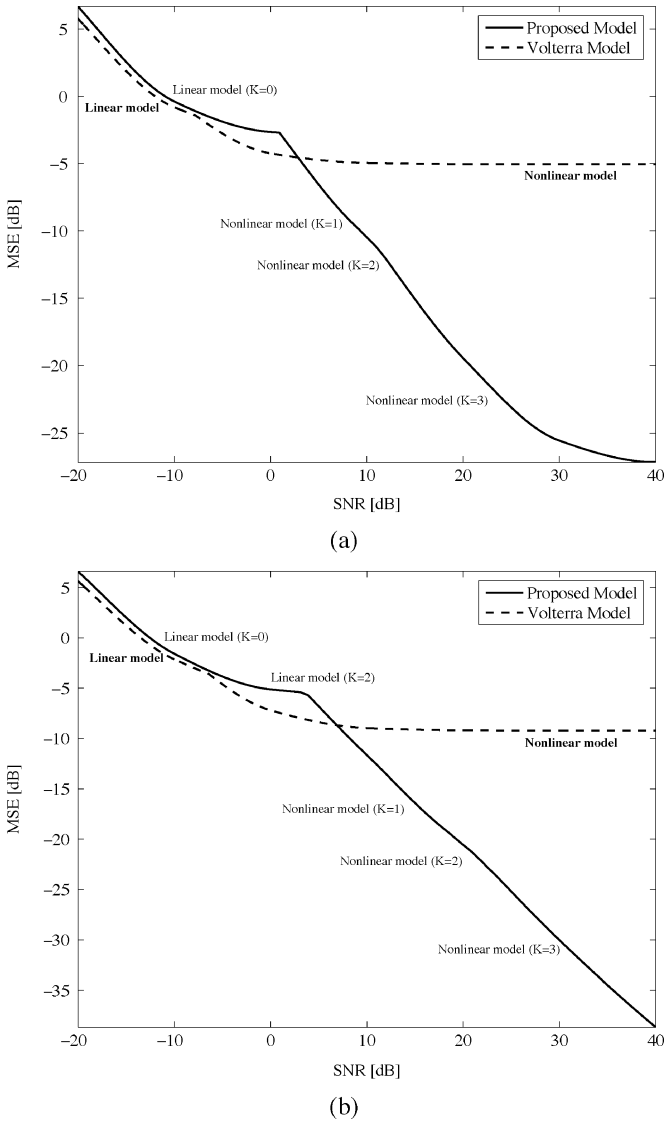


Fig. 5. MSE curves as a function of the SNR for (a) a white Gaussian input signal and (b) a real speech signal, as obtained by the STFT model (10) and the conventional time-domain Volterra model (50). The optimal model structure (linear or nonlinear) and model order (K) are indicated near the corresponding mse curves (light and dark fonts, respectively). The true system is formed as a second-order Volterra filter according to (46) and (49).

and STFT-based approaches and for every SNR value, we utilize either a linear or a nonlinear model and specify near each curve the optimal model structure and order (K) that achieve the mmse. Fig. 5(a) confirms that the theoretical results derived in Sections III and IV are valid also in this more realistic scenario when the system cannot be accurately described by the model. Specifically, we observe that for both Volterra and STFT approaches, the inclusion of a nonlinear component in the model is preferable only for high SNR conditions. For the STFT approach, such high SNR conditions enable the estimation of more crossband filters for achieving the mmse. In addition, we observe that for relatively high SNRs, the STFT model is considerably more advantageous than the time-domain Volterra model. For instance, for an SNR of 20 dB, the STFT model enables a decrease of approximately 14 dB in the mse using $K = 3$. Note that similar results are obtained for real speech input signals [Fig. 5(b)], which verifies that the theoretical derivations are not restricted by the Gaussian assumption made in Section III, but also applicable for non-Gaussian excitation signals. Clearly, the Volterra model is not sufficient for identification of long-memory systems, as the one represented by (46) and (49), since such systems necessitate long-Volterra kernels, which cannot be practically employed in real applications. Besides its improved estimation accuracy, the STFT-based approach also provides a significant reduction in computational cost over the time-domain Volterra approach. A detailed comparison between these two approaches in terms of estimation accuracy and computational cost is given in [33], and the reader is referred to there for further details.

VI. Conclusion

We have provided an explicit analysis for undermodeling errors in quadratically nonlinear system identification in the STFT domain. The model employed for system identification consists of a parallel connection of a linear component, which is represented by crossband filters between subbands, and a quadratic component, modeled by multiplicative cross-terms. We showed that the inclusion of the quadratic component in the model is preferable only for high SNR conditions and slowly time-varying systems (which enables to use longer observable data). A significant improvement in mse performance is then achieved compared to using a purely linear model. This improvement in performance becomes larger as the nonlinearity becomes stronger. On the other hand, as the SNR decreases or as the time variations in the system become faster, a lower mse is attained by allowing for nonlinear undermodeling and employing only the linear component in the estimation process. Furthermore, we showed that increasing the number of crossband filters in the linear component does not necessarily imply a lower mse. For every noise level, whether a linear or a nonlinear model is employed, there exists an optimal number of crossband filters, which increases as the SNR increases. Experimental results have supported the theoretical derivations.

It should be mentioned that the analysis provided in this paper does not concern the problem of controlling the model structure (linear or nonlinear) or the model order (number of crossband filters). It rather concentrates on proving the existence of such optimal values. Accordingly, future research will concentrate on constructing a fully control scheme, which chooses the optimal structure of the model as well as its optimal order, and may consequently be more efficient in estimating real-world nonstationary systems.

Appendix I
Derivation of (33)

Denoting $g(n) = (k - K + \lfloor (n/M) \rfloor) \bmod N$, the (m,ℓ)th term of $\mathbf{\Omega}_3$ from (30) can be written as

$$(\mathbf{\Omega}_3)_{m,\ell} = \sum_{n,r,q} E\left\{ x^*_{r,m_k} x_{r-n \bmod M, g(n)} x^*_{r,(k-m_k) \bmod N} x_{q,\ell_k} \times x^*_{q-n \bmod M, g(n)} x_{q,(k-\ell_k) \bmod N} \right\} \quad (51)$$

where $m_k = m$ if $m \le k/2$, and $m_k = m + k/2$ otherwise, and ℓ_k is defined similarly. By using the sixth-order moment factoring theorem for zero-mean complex Gaussian samples [46, p. 68], (51) reduces to products of different combinations of second-order moments, as follows:

$$(\mathbf{\Omega}_3)_{m,\ell} = \sum_{i=1}^{6} \omega_i \tag{52}$$

where

$$\begin{aligned}
\omega_1 = \sum_{n,r,q} & E\left\{x^*_{r,m_k} x_{r-n \bmod M, g(n)}\right\} \\
& \times E\left\{x^*_{r,(k-m_k) \bmod N} x_{q,\ell_k}\right\} \\
& \times E\left\{x^*_{q-n \bmod M, g(n)} x_{q,(k-\ell_k) \bmod N}\right\} \qquad (53a) \\
\omega_2 = \sum_{n,r,q} & E\left\{x^*_{r,m_k} x_{r-n \bmod M, g(n)}\right\} \\
& \times E\left\{x^*_{r,(k-m_k) \bmod N} x_{q,(k-\ell_k) \bmod N}\right\} \\
& \times E\left\{x^*_{q-n \bmod M, g(n)} x_{q,\ell_k}\right\} \qquad (53b) \\
\omega_3 = \sum_{n,r,q} & E\left\{x^*_{r,m_k} x_{q,\ell_k}\right\} \\
& \times E\left\{x^*_{r,(k-m_k) \bmod N} x_{r-n \bmod M, g(n)}\right\} \\
& \times E\left\{x^*_{q-n \bmod M, g(n)} x_{q,(k-\ell_k) \bmod N}\right\} \qquad (53c) \\
\omega_4 = \sum_{n,r,q} & E\left\{x^*_{r,m_k} x_{q,\ell_k}\right\} \\
& \times E\left\{x^*_{r,(k-m_k) \bmod N} x_{q,(k-\ell_k) \bmod N}\right\} \\
& \times E\left\{x^*_{q-n \bmod M, g(n)} x_{r-n \bmod M, g(n)}\right\} \qquad (53d) \\
\omega_5 = \sum_{n,r,q} & E\left\{x^*_{r,m_k} x_{q,(k-\ell_k) \bmod N}\right\} \\
& \times E\left\{x^*_{r,(k-m_k) \bmod N} x_{r-n \bmod M, g(n)}\right\} \\
& \times E\left\{x^*_{q-n \bmod M, g(n)} x_{q,\ell_k}\right\} \qquad (53e) \\
\omega_6 = \sum_{n,r,q} & E\left\{x^*_{r,m_k} x_{q,(k-\ell_k) \bmod N}\right\} \\
& \times E\left\{x^*_{r,(k-m_k) \bmod N} x_{q,\ell_k}\right\} \\
& \times E\left\{x^*_{q-n \bmod M, g(n)} x_{r-n \bmod M, g(n)}\right\}. \qquad (53f)
\end{aligned}$$

Due to the whiteness property of $x_{p,k}$, each term ω_i in (53) can be written as products of delta-functions, which impose certain conditions on both the matrix indexes m and ℓ, and the summation indexes n, r and q. Specifically for ω_1, the whiteness property of $x_{p,k}$ yields

$$\begin{aligned}
\omega_1 = \sigma_x^6 \sum_{n,r,q} & \delta_{n \bmod M} \delta_{m_k - g(n)} \delta_{r-q} \delta_{\ell_k - (k-m_k) \bmod N} \\
& \times \delta_{(k-\ell_k) \bmod N - g(n)}. \qquad (54)
\end{aligned}$$

Since the dependence of ω_1 on r and q is only via δ_{r-q}, and since r and q range from 0 to $P-1$, the double summation over r and q may be replaced by a multiplication of ω_1 by P. Moreover, it is easy to verify from the conditions imposed on m and ℓ that the condition $m = \ell$ must be satisfied. Nonetheless, the range of m (and ℓ) that contributes to ω_1 is bounded by the conditions imposed on the index n. Specifically, n should satisfy $n \bmod M = 0$ [recall that n ranges from 0 to $(2K+1)M-1$], and m satisfies the following:

$$m_k = g(n) = \left(k - K + \left\lfloor \frac{n}{M} \right\rfloor\right) \bmod N \tag{55a}$$

$$m_k = (k - \ell_k) \bmod N \tag{55b}$$

$$\ell_k = (k - m_k) \bmod N. \tag{55c}$$

Using the definitions of m_k and ℓ_k, it can be shown that the last two conditions reduce to $m_k = \ell_k \in \{k/2, (k+N)/2\}$. Also, since $n \in \{0, 1, \ldots, (2K+1)M-1\}$, (55a) implies that $m_k \in \mathcal{A}_k$ where

$$\mathcal{A}_k \triangleq \{[(k - K + n_1) \bmod N]M | n_1 \in \{0, \ldots, 2K\}\}. \tag{56}$$

Then, from the above discussion, ω_1 from (53a) may be written as

$$\omega_1 = \sigma_x^6 P \delta_{m-\ell} \delta_{m_k \in \mathcal{L}_1} \tag{57}$$

where

$$\mathcal{L}_1 = \left\{\frac{k}{2}, \frac{k+N}{2}\right\} \cap \mathcal{A}_k. \tag{58}$$

Following a similar analysis, it can be verified that

$$\omega_i = \sigma_x^6 P \delta_{m-\ell} \delta_{m_k \in \mathcal{L}i}; \quad \text{for } i = 2, 3, 5 \tag{59}$$

where

$$\mathcal{L}_2 = \left\{\left[0, \frac{k}{2}\right] \cup \left[k+1, \frac{k+N}{2}\right]\right\} \cap \mathcal{A}_k \tag{60}$$

$$\mathcal{L}_3 = \left\{\left[0, \frac{k}{2}\right] \cup \left[k+1, \frac{k+N}{2}\right]\right\} \cap \mathcal{A}_0 \tag{61}$$

and

$$\mathcal{L}_5 = \left\{\frac{k}{2}, \frac{k+N}{2}\right\} \cap A_0. \tag{62}$$

Concerning ω_4 and ω_6, the whiteness property of $x_{p,k}$ straightforwardly implies that their values do not depend on the index n. Consequently, the summation over n in (53d) and (53f) can be replaced by a multiplication by $(2K+1)M$, obtaining

$$\omega_4 = \sigma_x^6 P (2K+1) M \delta_{m-\ell} \tag{63}$$

$$\omega_6 = \sigma_x^6 P (2K+1) M \delta_{m-\ell} \delta_{m \in \{\frac{k}{2}, \frac{N}{2}\}}. \tag{64}$$

Substituting (57)–(64) into (52) yields (33).

Appendix II
Evaluation of ϵ_{22}

A. Derivation of (36)

Similarly to (24), we can use the ergodicity property of $x_{p,k}$ to approximate the (m,ℓ)th term of $\mathbf{\Lambda}_k^H \mathbf{\Lambda}_k$ as

$$\begin{aligned}
& \left(\mathbf{\Lambda}_k^H \mathbf{\Lambda}_k\right)_{m,\ell} \\
& = \sum_n x^*_{n,m_k} x_{n,\ell_k} x^*_{n,(k-m_k) \bmod N} x_{n,(k-\ell_k) \bmod N} \\
& \approx PE\left\{x^*_{n,m_k} x_{n,\ell_k} x^*_{n,(k-m_k) \bmod N} x_{n,(k-\ell_k) \bmod N}\right\}. \qquad (65)
\end{aligned}$$

By using the fourth-order moment factoring theorem for zero-mean complex Gaussian samples [46, p. 68], (65) can be rewritten as

$$\begin{aligned}\left(\mathbf{\Lambda}_k^H\mathbf{\Lambda}_k\right)_{m,\ell} &= PE\{x^*_{n,m_k}x_{n,\ell_k}\}\\ &\quad\times E\left\{x^*_{n,(k-m_k)\bmod N}x_{n,(k-\ell_k)\bmod N}\right\}\\ &\quad+PE\left\{x^*_{n,m_k}x_{n,(k-\ell_k)\bmod N}\right\}\\ &\quad\times E\left\{x^*_{n,(k-m_k)\bmod N}x_{n,\ell_k}\right\}\end{aligned}\tag{66}$$

which reduces to

$$\begin{aligned}\left(\mathbf{\Lambda}_k^H\mathbf{\Lambda}_k\right)_{m,\ell} &= P\sigma_x^4\delta_{m_k-\ell_k}\delta_{(k-m_k)\bmod N-(k-\ell_k)\bmod N}\\ &\quad+P\sigma_x^4\delta_{m_k-(k-\ell_k)\bmod N}\delta_{\ell_k-(k-m_k)\bmod N}\end{aligned}\tag{67}$$

due to the whiteness property of $x_{p,k}$. The first term in (67) is nonzero only if $m_k=\ell_k$, and the second term is nonzeros only if $m_k=(k-\ell_k)\bmod N$ and $\ell_k=(k-m_k)\bmod N$. Recall that $m_k=m$ if $m\le k/2$, and $m_k=m+k/2$ otherwise (ℓ_k is defined similarly), then (67) reduces to

$$\left(\mathbf{\Lambda}_k^H\mathbf{\Lambda}_k\right)_{m,\ell}=P\sigma_x^4\delta_{m-\ell}\left[1+\delta_{m\in\{\frac{k}{2},\frac{N}{2}\}}\right].\tag{68}$$

Let $\tilde{\mathbf{I}}_{N/2+1}$ denote an $(N/2+1)\times(N/2+1)$ diagonal matrix whose (m,m)th term satisfies

$$\left(\tilde{\mathbf{I}}_{N/2+1}\right)_{m,m}=\begin{cases}0.5, & m\in\left\{\frac{k}{2},\frac{N}{2}\right\}\\ 1, & \text{otherwise.}\end{cases}\tag{69}$$

Then, substituting (68) into $\mathbf{\Theta}_1$ from (35), we obtain

$$\begin{aligned}&(\mathbf{\Theta}_1)_{m,\ell}\\ &\quad=\frac{1}{P\sigma_x^4}\sum_{n,r,q}E\left\{(\mathbf{\Delta}^*)(\mathbf{\Lambda}_k\tilde{\mathbf{I}}_{N/2+1})_{rn}(\mathbf{\Lambda}_k^*)_{qn}(\mathbf{\Delta})_{q\ell}\right\}\\ &\quad=\frac{1}{P\sigma_x^4}\sum_{n,r,q}E\left\{x^*_{r-m\bmod M,\lfloor\frac{m}{M}\rfloor}\left[x_{r,n_k}x_{r,(k-n_k)\bmod N}\right.\right.\\ &\qquad\left.-0.5x_{r,\frac{k}{2}}\delta(n-k/2)-0.5x_{r,\frac{N+k}{2}}\delta(n-N/2)\right]\\ &\qquad\left.\times\,x^*_{q,n_k}x^*_{q,(k-n_k)\bmod N}x_{q-\ell\bmod M,\lfloor\frac{\ell}{M}\rfloor}\right\}\end{aligned}\tag{70}$$

which can be expressed as

$$(\mathbf{\Theta}_1)_{m,\ell}=\theta_1-0.5\left[\theta_2\left(\frac{k}{2}\right)+\theta_2\left(\frac{N+k}{2}\right)\right]\tag{71}$$

where

$$\begin{aligned}\theta_1=\frac{1}{P\sigma_x^4}\sum_{n,r,q}E\Big\{&x^*_{r-m\bmod M,\lfloor\frac{m}{M}\rfloor}x_{r,n_k}\,x^*_{q,n_k}x_{r,(k-n_k)\bmod N}\\ &\times\,x^*_{q,(k-n_k)\bmod N}x_{q-\ell\bmod M,\lfloor\frac{\ell}{M}\rfloor}\Big\}\end{aligned}\tag{72}$$

and

$$\begin{aligned}\theta_2(k)=\frac{1}{P\sigma_x^4}\sum_{r,q}E\Big\{&x^*_{r-m\bmod M,\lfloor\frac{m}{M}\rfloor}x_{r,k}x^*_{q,k}x_{r,k}x^*_{q,k}\\ &\times\,x_{q-\ell\bmod M,\lfloor\frac{\ell}{M}\rfloor}\Big\}.\end{aligned}\tag{73}$$

Equations (72) and (73) may be evaluated by using the Gaussian sixth-order moment factoring theorem, as applied in (33) for deriving the (m,ℓ)th term of $\mathbf{\Omega}_3$ (see Appendix I). Then, following a similar analysis to that given in Appendix I, we obtain explicit expressions for both θ_1 and $\theta_2(k)$:

$$\begin{aligned}\theta_1&=\sigma_x^2\delta_{m-\ell}\\ &\quad\times\left[3\delta_{m\in\left\{\frac{k}{2},\frac{k+N}{2}\right\}M}+\delta_{m\in\{0,\ldots,N-1\}M}+\frac{N}{2}+3\right]\end{aligned}\tag{74}$$

and

$$\theta_2(k)=\sigma_x^2\delta_{m-\ell}[4\delta_{m-kM}+2].\tag{75}$$

Substituting (74) and (75) into (71) yields (36).

B. Derivation of (37)

The (m,ℓ)th term of $\mathbf{\Theta}_3$ from (35) can be written as

$$\begin{aligned}(\mathbf{\Theta}_3)_{m,\ell}&=\sum_n E\\ &\quad\times\left\{x^*_{n,m_k}x_{n,\ell_k}x^*_{n,(k-m_k)\bmod N}x_{n,(k-\ell_k)\bmod N}\right\}.\end{aligned}\tag{76}$$

The forth-order moment in (76) was already derived in Appendix II-A, and is given by [see (65)–(67)]

$$\begin{aligned}&E\left\{x^*_{n,m_k}x_{n,\ell_k}x^*_{n,(k-m_k)\bmod N}x_{n,(k-\ell_k)\bmod N}\right\}\\ &\quad=\sigma_x^4\delta_{m_k-\ell_k}\delta_{(k-m_k)\bmod N-(k-\ell_k)\bmod N}\\ &\qquad+\sigma_x^4\delta_{m_k-(k-\ell_k)\bmod N}\delta_{\ell_k-(k-m_k)\bmod N}\\ &\quad=\sigma_x^4\delta_{m-\ell}\left[1+\delta_{m\in\left\{\frac{k}{2},\frac{N}{2}\right\}}\right]\end{aligned}\tag{77}$$

where the last equation follows from (68). Since (77) does not depend on n, and n ranges from 0 to $P-1$, the summation in (76) can be replaced by multiplication by P, which yields (37).

ACKNOWLEDGMENT

The authors thank the anonymous reviewers for their constructive comments and helpful suggestions.

REFERENCES

[1] A. Stenger and W. Kellermann, "Adaptation of a memoryless preprocessor for nonlinear acoustic echo cancelling," *Signal Process.*, vol. 80, no. 9, pp. 1747–1760, 2000.

[2] A. Guérin, G. Faucon, and R. L. Bouquin-Jeannès, "Nonlinear acoustic echo cancellation based on Volterra filters," *IEEE Trans. Speech Audio Process.*, vol. 11, no. 6, pp. 672–683, Nov. 2003.

[3] H. Dai and W. P. Zhu, "Compensation of loudspeaker nonlinearity in acoustic echo cancellation using raised-cosine function," *IEEE Trans. Circuits Syst. II*, vol. 53, no. 11, pp. 1190–2006, Nov. 2006.

[4] K. Shi, X. Ma, and G. T. Zhou, "Acoustic echo cancellation using a pseudocoherence function in the presence of memoryless nonlinearity," *IEEE Trans. Circuits Syst. I*, vol. 55, no. 9, pp. 2639–2649, Oct. 2008.

[5] S. Benedetto and E. Biglieri, "Nonlinear equalization of digital satellite channels," *IEEE J. Sel. Areas Commun.*, vol. SAC-1, pp. 57–62, Jan. 1983.

[6] M. Bellafemina and S. Benedetto, "Identication and equalization of nonlinear channels for digital transmission," in *Proc. Int. Symp. Circuits Syst.*, Kyoto, Japan, Jun. 1985, pp. 1477–1480.

[7] D. G. Lainiotis and P. Papaparaskeva, "A partitioned adaptive approach to nonlinear channel equalization," *IEEE Trans. Commun.*, vol. 46, no. 10, pp. 1325–1336, Oct. 1998.

[8] D. T. Westwick and R. E. Kearney, "Separable least squares identification of nonlinear Hammerstein models: Application to stretch reflex dynamics," *Ann. Biomed. Eng.*, vol. 29, no. 8, pp. 707–718, Aug. 2001.
[9] G. Ramponi and G. L. Sicuranza, "Quadratic digital filters for image processing," *IEEE Trans. Acoust., Speech, Signal Process.*, vol. 36, no. 6, pp. 937–939, Jun. 1988.
[10] W. J. Rugh, *Nonlinear System Theory: The Volterra-Wiener Approach*. Baltimore, MD: The John Hopkins Univ. Press, 1981.
[11] T. Koh and E. J. Powers, "Second-order Volterra filtering and its application to nonlinear system identification," *IEEE Trans. Acoust., Speech, Signal Process.*, vol. ASSP-33, no. 6, pp. 1445–1455, Dec. 1985.
[12] M. Schetzen, *The Volterra and Wiener Theories of Nonlinear Systems*. New York: Krieger, 1989.
[13] V. J. Mathews, "Adaptive polynomial filters," *IEEE Signal Process. Mag.*, vol. 8, no. 3, pp. 10–26, Jul. 1991.
[14] G. O. Glentis, P. Koukoulas, and N. Kalouptsidis, "Efficient algorithms for Volterra system identification," *IEEE Trans. Signal Process.*, vol. 47, no. 11, pp. 3042–3057, Nov. 1999.
[15] V. J. Mathews and G. L. Sicuranza, *Polynomial Signal Processing*. New York: Wiley, 2000.
[16] E. Biglieri, A. Gersho, R. D. Gitlin, and T. L. Lim, "Adaptive cancellation of nonlinear intersymbol interference for voiceband data transmission," *IEEE J. Sel. Areas Commun.*, vol. 2, no. 5, pp. 765–777, Sep. 1984.
[17] J. C. Stapleton and S. C. Bass, "Adaptive noise cancellation for a class of nonlinear, dynamic reference channels," *IEEE Trans. Circuits Syst. II*, vol. 32, no. 2, pp. 143–150, Feb. 1985.
[18] J.-Y. Lin and C.-H. Wei, "Adaptive nonlinear decision feedback equalization with channel estimation and timing recovery in digital magnetic recording systems," *IEEE Trans. Circuits Syst. II*, vol. 42, no. 3, pp. 196–206, Mar. 1995.
[19] A. Stenger, L. Trautmann, and R. Rabenstein, "Nonlinear acoustic echo cancellation with 2nd order adaptive Volterra filters," in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Processing*, Phoenix, AZ, Mar. 1999, pp. 877–880.
[20] R. D. Nowak, "Penalized least squares estimation of Volterra filters and higher order statistics," *IEEE Trans. Signal Process.*, vol. 46, no. 2, pp. 419–428, Feb. 1998.
[21] P. Koukoulas and N. Kalouptsidis, "Nonlinear system identification using Gaussian inputs," *IEEE Trans. Signal Process.*, vol. 43, no. 8, pp. 1831–1841, Aug. 1995.
[22] K. I. Kim and E. J. Powers, "A digital method of modeling quadratically nonlinear systems with a general random input," *IEEE Trans. Acoust., Speech, Signal Process.*, vol. 36, no. 11, pp. 1758–1769, Nov. 1988.
[23] C. H. Tseng and E. J. Powers, "Batch and adaptive Volterra filtering of cubically nonlinear systems with a Gaussian input," in *IEEE Int. Symp. Circuits Syst. (ISCAS)*, 1993, vol. 1, pp. 40–43.
[24] P. P. Vaidyanathan, *Multirate Systems and Filters Banks*. Englewood Cliffs, NJ: Prentice-Hall, 1993.
[25] A. Gilloire and M. Vetterli, "Adaptive filtering in subbands with critical sampling: Analysis, experiments, and application to acoustic echo cancellation," *IEEE Trans. Signal Process.*, vol. 40, no. 8, pp. 1862–1875, Aug. 1992.
[26] Y. Avargel and I. Cohen, "System identification in the short-time Fourier transform domain with crossband filtering," *IEEE Trans. Audio Speech Lang. Process.*, vol. 15, no. 4, pp. 1305–1319, May 2007.
[27] Y. Avargel and I. Cohen, "Linear system identification in the short-time Fourier transform domain," in *Speech Processing in Modern Communication: Challenges and Perspectives*, I. Cohen, J. Benesty, and S. Gannot, Eds. New York: Springer, 2010, ch. 1, pp. 1–32.
[28] B. E. Usevitch and M. T. Orchard, "Adaptive filtering using filter banks," *IEEE Trans. Circuits Syst. II*, vol. 43, no. 3, pp. 255–265, Mar. 1996.
[29] W. Kellermann, "Analysis and design of multirate systems for cancellation of acoustical echoes," in *Proc. IEEE Int. Conf. Acoustics, Speech, Signal Processing (ICASSP)*, New York, Apr. 1988, pp. 2570–2573.
[30] Y. Avargel and I. Cohen, "On multiplicative transfer function approximation in the short-time Fourier transform domain," *IEEE Signal Process. Lett.*, vol. 14, no. 5, pp. 337–340, May 2007.
[31] Y. Avargel and I. Cohen, "Adaptive system identification in the short-time Fourier transform domain using cross-multiplicative transfer function approximation," *IEEE Trans. Audio Speech Lang. Process.*, vol. 16, no. 1, pp. 162–173, Jan. 2008.
[32] S. S. Pradhan and V. U. Reddy, "A new approach to subband adaptive filtering," *IEEE Trans. Signal Process.*, vol. 47, no. 3, pp. 655–664, Mar. 1999.
[33] Y. Avargel and I. Cohen, "Modeling and identification of nonlinear systems in the short-time Fourier transform domain," *IEEE Trans. Signal Process.*, vol. 58, no. 1, pp. 291–304, Jan. 2010.
[34] Y. Avargel and I. Cohen, "Adaptive nonlinear system identification in the short-time Fourier transform domain," *IEEE Trans. Signal Process.*, vol. 57, no. 10, pp. 3891–3904, Oct. 2009.
[35] A. E. Nordsjo, B. M. Ninness, and T. Wigren, "Quantifying model error caused by nonlinear undermodeling in linear system identification," in *Preprints 13th World Congr. IFAC*, San Francisco, CA, 1996, vol. I, pp. 145–149.
[36] B. Ninness and S. Gibson, "Quantifying the accuracy of hammerstein model estimation," *Automatica*, vol. 38, no. 12, pp. 2037–2051, 2002.
[37] J. Schoukens, R. Pintelon, T. Dobrowiecki, and Y. Rolain, "Identification of linear systems with nonlinear distortions," *Automatica*, vol. 41, no. 3, pp. 491–504, 2005.
[38] C. Breining, P. Dreiseitel, E. Hänsler, A. Mader, B. Nitsch, H. Puder, T. Schertler, G. Schmidt, and J. Tlip, "Acoustic echo control," *IEEE Signal Process. Mag.*, vol. 16, no. 4, pp. 42–69, Jul. 1999.
[39] M. R. Portnoff, "Time-frequency representation of digital signals and systems based on short-time Fourier analysis," *IEEE Trans. Signal Process.*, vol. ASSP-28, no. 1, pp. 55–69, Feb. 1980.
[40] A. Neumaier, "Solving ill-conditioned and singular linear systems: A tutorial on regularization," *SIAM Rev.*, vol. 40, no. 3, pp. 636–666, Sep. 1998.
[41] D. R. Brillinger, *Time Series: Data Analysis and Theory*. Philadelphia, PA: SIAM, 2001.
[42] Y. Ephraim and D. Malah, "Speech enhancement using a minimum mean-square error short-time spectral amplitude estimator," *IEEE Trans. Acoust., Speech, Signal Process.*, vol. 32, no. 6, pp. 1109–1121, Dec. 1984.
[43] Y. Ephraim and I. Cohen, "Recent advancements in speech enhancement," in *The Electrical Engineering Handbook*, R. C. Dorf, Ed., 3rd ed. Boca Raton, FL: CRC Press, 2006.
[44] L. Ljung, *System Identification: Theory for the User*. Upper Saddle River, NJ: Prentice-Hall, 1999.
[45] A. Papoulis, *Probability, Random Variables, and Stochastic Processes*. Singapore: McGraw-Hill, 1991.
[46] S. Haykin, *Adaptive Filter Theory*. Englewood Cliffs, NJ: Prentice-Hall, 2002.
[47] T. F. Quatieri, *Discrete-Time Speech Signal Processing*. Englewood Cliffs, NJ: Prentice-Hall, 2002.
[48] F. D. Ridder, R. Pintelon, J. Schoukens, and D. P. Gillikin, "Modified AIC and MDL model selection criteria for short data records," *IEEE Trans. Instrum. Meas.*, vol. 54, no. 1, pp. 144–150, Feb. 2005.
[49] G. Schwarz, "Estimating the dimension of a model," *Ann. Statist.*, vol. 6, no. 2, pp. 461–464, 1978.
[50] P. Stoica and Y. Selen, "Model order selection: A review of information criterion rules," *IEEE Signal Process. Mag.*, vol. 21, no. 4, pp. 36–47, Jul. 2004.
[51] G. C. Goodwin, M. Gevers, and B. Ninness, "Quantifying the error in estimated transfer functions with application to model order selection," *IEEE Trans. Autom. Control*, vol. 37, no. 7, pp. 913–928, Jul. 1992.
[52] H. Akaike, "A new look at the statistical model identification," *IEEE Trans. Autom. Control*, vol. AC-19, no. 6, pp. 716–723, Dec. 1974.
[53] J. Rissanen, "Modeling by shortest data description," *Automatica*, vol. 14, no. 5, pp. 465–471, 1978.
[54] J. Wexler and S. Raz, "Discrete Gabor expansions," *Signal Process.*, vol. 21, pp. 207–220, Nov. 1990.

Yekutiel Avargel (S'06) received the B.Sc., M.Sc., and Ph.D. degrees in electrical engineering from the Technion—Israel Institute of Technology, Haifa, in 2004, 2007, and 2008, respectively.

From 2003 to 2004, he was a Research Engineer with RAFAEL Research Laboratories, Haifa, Israel Ministry of Defense. From 2004 to 2008, he was a Research Assistant and a Project Supervisor with the Signal and Image Processing Lab (SIPL), Electrical Engineering Department, Technion. Since 2009, he has been a Research Scientist with ELTA Research Laboratories, Israel Ministry of Defense. His research interests are radar signal processing, parameter estimation, statistical signal processing, system identification, and adaptive filtering.

Dr. Avargel received in 2008 the Jury award for distinguished graduate students and the SIPL Excellent Supervisor award.

Israel Cohen (M'01–SM'03) received the B.Sc. (*summa cum laude*), M.Sc., and Ph.D. degrees in electrical engineering from the Technion—Israel Institute of Technology, Haifa, Israel, in 1990, 1993 and 1998, respectively.

From 1990 to 1998, he was a Research Scientist with RAFAEL Research Laboratories, Haifa, Israel Ministry of Defense. From 1998 to 2001, he was a Postdoctoral Research Associate with the Computer Science Department, Yale University, New Haven, CT. In 2001, he joined the Electrical Engineering Department of the Technion, where he is currently an Associate Professor. His research interests are statistical signal processing, analysis and modeling of acoustic signals, speech enhancement, noise estimation, microphone arrays, source localization, blind source separation, system identification, and adaptive filtering.

Dr. Cohen served as a Guest Editor of the *EURASIP Journal on Advances in Signal Processing* Special Issue on Advances in Multimicrophone Speech Processing and the *EURASIP Speech Communication Journal* Special Issue on Speech Enhancement. He is a Co-Editor of the Multichannel Speech Processing section of the *Springer Handbook of Speech Processing* (Springer, 2007), a coauthor of *Noise Reduction in Speech Processing* (Springer, 2009), and a Co-Chair of the 2010 International Workshop on Acoustic Echo and Noise Control. In 2009, he received the Muriel and David Jacknow Award for Excellence in Teaching, and in 2010, the Alexander Goldberg Prize for Excellence in Research. He served as Associate Editor of the IEEE TRANSACTIONS ON AUDIO, SPEECH, AND LANGUAGE PROCESSING and the IEEE SIGNAL PROCESSING LETTERS.