

Adaptive Nonlinear System Identification in the Short-Time Fourier Transform Domain

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

Abstract—In this paper, we introduce an adaptive algorithm for nonlinear system identification in the short-time Fourier transform (STFT) domain. The adaptive scheme consists of a parallel combination of a linear component, represented by crossband filters between subbands, and a quadratic component, which is modeled by multiplicative cross-terms. We adaptively update the model parameters using the least-mean-square (LMS) algorithm, and derive explicit expressions for the transient and steady-state mean-square error (mse) in frequency bins for white Gaussian inputs. We show that estimation of the nonlinear component improves the mse performance only when the power ratio of nonlinear to linear components is relatively high. Furthermore, as the number of crossband filters increases, a lower steady-state mse may be obtained at the expense of slower convergence. Experimental results support the theoretical derivations.

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

I. INTRODUCTION

IDENTIFICATION of nonlinear systems has recently attracted great interest in many applications, including acoustic echo cancellation [1]–[3], channel equalization [4], [5], biological system modeling [6], and image processing [7]. A popular approach for modeling nonlinear systems is using Volterra filters [8]–[10], which are attractive due to their structural generality and versatile modeling capabilities (e.g., [11] and [12]). An important property of Volterra filters is the linear relation between the system output and the filter coefficients, which enables the use of algorithms from linear estimation theory for estimating the parameters. Adaptation algorithms used for this purpose often employ the least-mean-square (LMS) algorithm [13] due to its robustness and simplicity (e.g., [2], [9], and [12]). However, the LMS algorithm suffers from slow convergence when the input signal to the adaptive filter is correlated, which is extremely problematic when applied to Volterra filters [9]. Another major drawback of the adaptive Volterra filter is the high computational cost caused by the large number of model parameters, especially for long-memory

systems [10], [14]. To speed-up convergence, the affine projection (AP) algorithm and the recursive least-squares (RLS) algorithm were employed for updating the adaptive Volterra filters [10], [15]. These approaches, however, substantially increase the computational complexity of the estimation process. Alternatively, several time-domain approximations, which suggest a less general structure than the Volterra filter, have been proposed, including orthogonalized power filters [16], Hammerstein models [17], parallel-cascade structures [18], and multimemory decomposition [19]. Other adaptive algorithms, which operate in the frequency domain, have been proposed to ease the computational burden [20], [21]. These approaches are based on the discrete frequency-domain model [22], which approximates the Volterra filter using multiplicative terms. Nonetheless, a major limitation of this model is its underlying assumption that the observation frame is sufficiently large compared with the memory length of the system. This assumption may be very restrictive, especially when long and time-varying impulse responses are considered (as in acoustic echo cancellation applications [23]).

The drawbacks of the conventional time- and frequency-domain methods have motivated the use of subband (multirate) techniques [24] for improved nonlinear system identification [25], [26]. As in subband linear system identification [27]–[33], such techniques may achieve computational efficiency as well as improved convergence rate due to processing in distinct subbands. The method developed in [25] for nonlinear system identification in the short-time Fourier transform (STFT) domain is based on a time-frequency representation of Volterra filters. The system model consists of a parallel combination of linear and nonlinear components. The linear component is represented by crossband filters between subbands [28], [31], while the nonlinear component is modeled by multiplicative cross terms. In [25], the parameters of the proposed model were estimated *off-line* using a least-squares (LS) criterion, and it was shown that a significant reduction in computational cost as well as a substantial improvement in estimation accuracy can be achieved over the time-domain Volterra model, particularly when long-memory nonlinear systems are considered. The performance of this off-line scheme has been analyzed in [26] for the quadratic case. A detailed mean-square analysis was presented, and the problem of employing either a linear or a nonlinear model for the estimation process, as well as determining the optimal number of crossband filters, was considered.

In this paper, we introduce an *adaptive* algorithm for the estimation of quadratically nonlinear systems in the STFT domain. The quadratic model proposed in [25] is employed, and its parameters are adaptively updated using the LMS algorithm. We derive explicit expressions for the transient and steady-state

Manuscript received August 29, 2008; accepted March 13, 2009. First published April 28, 2009; current version published September 16, 2009. The associate editor coordinating the review of this manuscript and approving it for publication was Dr. Dennis R. Morgan. This work was supported by the Israel Science Foundation (Grant No. 1085/05).

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).

Color versions of one or more of the figures in this paper are available online at <http://www.ieeexplore.ieee.org>

Digital Object Identifier 10.1109/TSP.2009.2021713

mean-square error (mse) in frequency bins for white Gaussian processes, using different step sizes for the linear and quadratic components of the model. The analysis provides important insights into the influence of nonlinear undermodeling (i.e., employing a purely linear model in the estimation process) and the number of estimated crossband filters on the transient and steady-state performances. We show that as the number of crossband filters increases, a lower steady-state mse is achieved, whether a linear or a nonlinear model is employed; however, the algorithm then suffers from slower convergence. Accordingly, as more data is employed in the adaptation process, additional crossband filters should be estimated to achieve the minimal mse (mmse) at each iteration. Moreover, we show that the choice of the model structure (either linear or nonlinear) is mainly influenced by the nonlinear-to-linear ratio (NLR), which represents the power ratio of nonlinear to linear components of the system. Specifically for high NLR conditions, a lower steady-state mse can be achieved by incorporating a nonlinear component into the model. On the other hand, as the nonlinearity becomes weaker (i.e., the NLR decreases), the steady-state mse associated with the linear model decreases, while the relative improvement achieved by the nonlinear model becomes smaller. Consequently, for relatively low NLR values, utilizing the nonlinear component in the estimation process may not necessarily imply a lower steady-state mse in subbands. Experimental results validate the theoretical results derived in this paper.

The paper is organized as follows. In Section II, we formulate the quadratic STFT model and introduce an adaptive scheme for updating the model parameters. In Section III, we derive explicit expressions for the transient and steady-state mse in subbands. In Section IV, we address the computational complexity of the proposed algorithm and compare it to that of the conventional time-domain Volterra approach. Finally, in Section V, we present some experimental results to support the theoretical derivations.

II. MODEL FORMULATION AND IDENTIFICATION

In this section, we introduce an LMS-based adaptive scheme for the identification of quadratically nonlinear systems in the STFT domain. We assume that the system to be identified can be represented by the nonlinear STFT model proposed in [25]. 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.

Let an input $x(n)$ and output $y(n)$ of an unknown (quadratically) nonlinear system be related by

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

where $\phi(\cdot)$ denotes a discrete-time nonlinear time-invariant system, $\xi(n)$ is a corrupting additive noise signal, and $d(n)$ is the clean output signal. Note that the “noise” signal $\xi(n)$ may sometimes include a useful signal, e.g., the local speaker

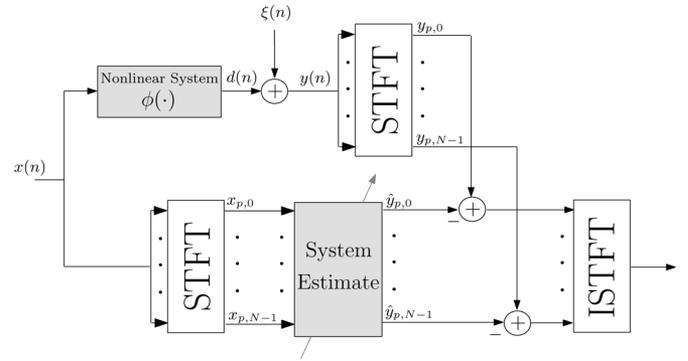


Fig. 1. Nonlinear system identification in the STFT domain.

signal in acoustic echo cancellation [1]–[3]. The STFT of $y(n)$ is given by [34]

$$y_{p,k} = \sum_n y(n) \psi_{p,k}^*(n) \quad (2)$$

where $\psi_{p,k}(n) = \psi(n - pL) e^{j(2\pi/N)k(n-pL)}$ denotes a translated and modulated window function, $\psi(n)$ is a real-valued analysis window of length N , p is the frame index, k represents the frequency-bin index ($0 \leq k \leq N - 1$), L is the translation factor and $*$ denotes complex conjugation. The components of $y(n)$ in (1) are similarly transformed into STFT components

$$y_{p,k} = d_{p,k} + \xi_{p,k}. \quad (3)$$

An adaptive system identification scheme is illustrated in Fig. 1. We assume that the system output signal $d(n)$ arises from the nonlinear STFT model proposed in [25]. Accordingly, the true system is formed as a parallel combination of linear and quadratic components in the time-frequency domain as follows:

$$d_{p,k} = \sum_{k'=0}^{N-1} \sum_{p'=0}^{M-1} x_{p-p',k'} \bar{h}_{p',k,k'} + \sum_{k' \in \mathcal{F}} x_{p,k'} x_{p,(k-k') \bmod N} \bar{c}_{k',(k-k') \bmod N} \quad (4)$$

where $x_{p,k}$ is the STFT of the input $x(n)$, $\bar{h}_{p',k,k'}$ denotes the true linear crossband filter of length M from frequency bin k' to frequency bin k , $\bar{c}_{k',(k-k') \bmod N}$ is the true quadratic cross-term, and $\mathcal{F} = \{0, 1, \dots, \lfloor k/2 \rfloor, k+1, \dots, k+1 + \lfloor (N-k-2)/2 \rfloor\}$. The linear crossband filters are necessary for perfectly representing the linear component of the system in the STFT domain, and are used for canceling the aliasing effects caused by the subsampling factor L [28], [31]. The nonlinear cross-terms $\{\bar{c}_{k',(k-k') \bmod N} | k' \in \mathcal{F}\}$, on the other hand, are used for modeling the quadratic component of the system using a sum over all possible interactions between pairs of input frequencies $x_{p,k'}$ and $x_{p,k''}$, where $k'' = (k - k') \bmod N$.

The goal in adaptive system identification is to define a model for describing the input-output relationship of the true system,

and to adaptively update its parameters according to a given criterion. To do so, let us employ the model in (4) for the adaptive estimation process, using only $2K + 1$ crossband filters, where K controls the undermodeling in the linear component of the model. Denoting the adaptive crossband filters and adaptive cross-terms of the model at frame index p by $h_{p',k,k'}(p)$ and $c_{k',(k-k')\bmod N}(p)$, respectively, the resulting estimate $\hat{y}_{p,k}$ can be written as

$$\hat{y}_{p,k} = \sum_{k'=k-K}^{k+K} \sum_{p'=0}^{M-1} x_{p-p',k'\bmod N} h_{p',k,k'\bmod N}(p) + \sum_{k' \in \mathcal{F}} x_{p,k'} x_{p,(k-k')\bmod N} c_{k',(k-k')\bmod N}(p). \quad (5)$$

Let $\mathbf{h}_{k,k'}(p) = [h_{0,k,k'}(p) \ h_{1,k,k'}(p) \ \cdots \ h_{M-1,k,k'}(p)]^T$ represent the adaptive crossband filter from frequency bin k' to frequency bin k , and let $\mathbf{h}_k(p)$ denote a column-stack concatenation of the $2K + 1$ estimated filters around the k th frequency bin, i.e.,

$$\mathbf{h}_k(p) = \left[\mathbf{h}_{k,(k-K)\bmod N}^T(p) \ \cdots \ \mathbf{h}_{k,(k+K)\bmod N}^T(p) \right]^T. \quad (6)$$

Likewise, let $\mathbf{x}_k(p) = [x_{p,k} \ x_{p-1,k} \ \cdots \ x_{p-M+1,k}]^T$ and

$$\mathbf{x}_{L,k}(p) = \left[\mathbf{x}_{(k-K)\bmod N}^T(p) \ \cdots \ \mathbf{x}_{(k+K)\bmod N}^T(p) \right]^T \quad (7)$$

form the input data vector to the linear component of the model $\mathbf{h}_k(p)$. For notational simplicity, let us assume that k is odd and N is even, such that according to (4), the number of quadratic cross-terms in each frequency bin is $N/2$. Accordingly, let

$$\mathbf{c}_k(p) = \left[c_{0,k}(p) \ \cdots \ c_{\frac{k-1}{2},\frac{k+1}{2}}(p) \right. \\ \left. c_{k+1,N-1}(p) \ \cdots \ c_{\frac{N+k-1}{2},\frac{N+k+1}{2}}(p) \right]^T \quad (8)$$

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

$$\mathbf{x}_{Q,k}(p) = \left[x_{p,0}x_{p,k} \ \cdots \ x_{p,\frac{k-1}{2}}x_{p,\frac{k+1}{2}} \right. \\ \left. x_{p,k+1}x_{p,N-1} \ \cdots \ x_{p,\frac{N+k-1}{2}}x_{p,\frac{N+k+1}{2}} \right]^T \quad (9)$$

be the input data vector to the quadratic component of the model $\mathbf{c}_k(p)$. Then, the output signal estimate $\hat{y}_{p,k}$ from (5) can be rewritten as

$$\hat{y}_{p,k} = \mathbf{x}_{L,k}^T(p) \mathbf{h}_k(p) + \mathbf{x}_{Q,k}^T(p) \mathbf{c}_k(p). \quad (10)$$

The $2K + 1$ adaptive crossband filters and the $N/2$ adaptive cross-terms are updated using the LMS algorithm as

$$\mathbf{h}_k(p+1) = \mathbf{h}_k(p) + \mu_L e_{p,k} \mathbf{x}_{L,k}^*(p) \quad (11)$$

and

$$\mathbf{c}_k(p+1) = \mathbf{c}_k(p) + \mu_Q e_{p,k} \mathbf{x}_{Q,k}^*(p) \quad (12)$$

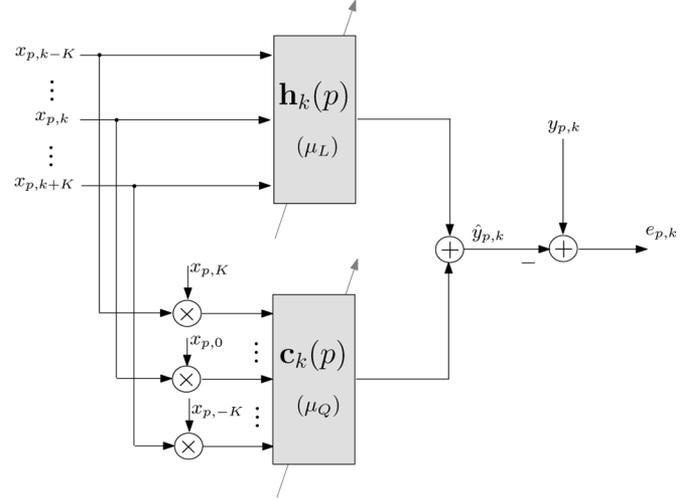


Fig. 2. Block diagram of the proposed adaptive scheme for identifying quadratically nonlinear systems in the STFT domain.

where

$$e_{p,k} = y_{p,k} - \hat{y}_{p,k} \quad (13)$$

is the error signal in the k th frequency bin, $y_{p,k}$ is defined in (2)–(4), and μ_L and μ_Q are the step sizes of the linear and quadratic components of the model, respectively. The separate update equations for $\mathbf{h}_k(p)$ and $\mathbf{c}_k(p)$ enable one to use different step sizes for adaptation of the linear and quadratic components of the model. In case one component varies slower than the other, such adaptation may enhance the tracking capability of the algorithm by utilizing a proper step size for each component. A block diagram of this parallel adaptive scheme is illustrated in Fig. 2. Our objective is to analyze the error attainable in each frequency bin and derive explicit expressions for the transient and steady-state mse.

III. MSE ANALYSIS

In this section, we derive explicit expressions for the transient and steady-state mse obtainable in the k th frequency bin. To make the following analysis mathematically tractable, we use the common independence assumption which states that the current input data vector is statistically independent of the currently updated parameters vector (e.g., [35] and [36]). Specifically, the vector $[\mathbf{x}_{L,k}^T(p) \ \mathbf{x}_{Q,k}^T(p)]$ is independent of $[\mathbf{h}_k^T(p) \ \mathbf{c}_k^T(p)]$. In addition, we assume that $x_{p,k}$ and $\xi_{p,k}$ are statistically independent zero-mean white complex Gaussian signals with variances σ_x^2 and σ_ξ^2 , respectively. The Gaussian assumption of the corresponding STFT signals is often justified by a version of the central limit theorem for correlated signals ([37, Theorem 4.4.2]), and underlies the design of many speech-enhancement systems [38], [39].

A. Transient Performance

The transient mse is defined by

$$\epsilon_k(p) = E\{|e_{p,k}|^2\}. \quad (14)$$

Let us define the misalignment vectors of the linear and quadratic components, respectively, as

$$\mathbf{g}_{L,k}(p) = \mathbf{h}_k(p) - \bar{\mathbf{h}}_k \quad (15)$$

and

$$\mathbf{g}_{Q,k}(p) = \mathbf{c}_k(p) - \bar{\mathbf{c}}_k \quad (16)$$

where $\bar{\mathbf{h}}_k$ and $\bar{\mathbf{c}}_k$ are respectively the $2K + 1$ crossband filters and the $N/2$ cross-terms of the true system [defined similarly to (6) and (8)]. Then, substituting (10) and the definition of $y_{p,k}$ from (2)–(4) into (13), the error signal can be written as

$$e_{p,k} = \tilde{\mathbf{x}}_{L,k}^T(p) \tilde{\mathbf{h}}_k + \mathbf{x}_{L,k}^T(p) \mathbf{g}_{L,k}(p) + \mathbf{x}_{Q,k}^T(p) \mathbf{g}_{Q,k}(p) + \xi_{p,k} \quad (17)$$

where $\tilde{\mathbf{h}}_k$ and $\tilde{\mathbf{x}}_{L,k}(p)$ are the column-stack concatenations of $\{\bar{\mathbf{h}}_{k,k'}\}_{k' \in \mathcal{L}}$ and $\{\mathbf{x}_{k,k'}(p)\}_{k' \in \mathcal{L}}$, respectively, and $\mathcal{L} = \{k' | k' \in [0, N - 1] \text{ and } k' \notin [k - K, k + K]\}$. Substituting (17) into (14) and using our assumptions, the mse can be expressed as (see Appendix I)

$$\epsilon_k(p) = \sigma_\xi^2 + \sigma_x^2 \|\tilde{\mathbf{h}}_k\|^2 + \sigma_x^2 E\{\|\mathbf{g}_{L,k}(p)\|^2\} + \sigma_x^4 E\{\|\mathbf{g}_{Q,k}(p)\|^2\}. \quad (18)$$

In order to find an explicit expression for the transient mse, recursive formulas for $E\{\|\mathbf{g}_{L,k}(p)\|^2\}$ and $E\{\|\mathbf{g}_{Q,k}(p)\|^2\}$ are required. By substituting (17) into (11)–(12), the LMS update equations for the misalignment vectors can be written as

$$\begin{aligned} \mathbf{g}_{L,k}(p+1) &= [\mathbf{I}_{(2K+1)M} - \mu_L \mathbf{x}_{L,k}^*(p) \mathbf{x}_{L,k}^T(p)] \mathbf{g}_{L,k}(p) \\ &\quad - \mu_L \mathbf{x}_{L,k}^*(p) \mathbf{x}_{Q,k}^T(p) \mathbf{g}_{Q,k}(p) \\ &\quad + \mu_L [\tilde{\mathbf{x}}_{L,k}^T(p) \tilde{\mathbf{h}}_k] \mathbf{x}_{L,k}^*(p) + \mu_L \xi_{p,k} \mathbf{x}_{L,k}^*(p) \end{aligned} \quad (19)$$

$$\begin{aligned} \mathbf{g}_{Q,k}(p+1) &= [\mathbf{I}_{N/2} - \mu_Q \mathbf{x}_{Q,k}^*(p) \mathbf{x}_{Q,k}^T(p)] \mathbf{g}_{Q,k}(p) \\ &\quad - \mu_Q \mathbf{x}_{Q,k}^*(p) \mathbf{x}_{L,k}^T(p) \mathbf{g}_{L,k}(p) \\ &\quad + \mu_Q [\tilde{\mathbf{x}}_{L,k}^T(p) \tilde{\mathbf{h}}_k] \mathbf{x}_{Q,k}^*(p) + \mu_Q \xi_{p,k} \mathbf{x}_{Q,k}^*(p) \end{aligned} \quad (20)$$

where \mathbf{I}_P is the identity matrix of size $P \times P$. We proceed with evaluating a recursion for $E\{\|\mathbf{g}_{L,k}(p+1)\|^2\}$. Taking the norm on both sides of (19), and using the fact that odd-order moments of a zero-mean complex Gaussian process are zero [13], we obtain

$$\begin{aligned} E\{\|\mathbf{g}_{L,k}(p+1)\|^2\} &= E\left\{\left\|[\mathbf{I}_{(2K+1)M} - \mu_L \mathbf{x}_{L,k}^*(p) \mathbf{x}_{L,k}^T(p)] \mathbf{g}_{L,k}(p)\right\|^2\right\} \\ &\quad + \mu_L^2 E\left\{\left\|\mathbf{x}_{L,k}^*(p) \mathbf{x}_{Q,k}^T(p) \mathbf{g}_{Q,k}(p)\right\|^2\right\} \\ &\quad + \mu_L^2 E\left\{\left\|[\tilde{\mathbf{x}}_{L,k}^T(p) \tilde{\mathbf{h}}_k] \mathbf{x}_{L,k}^*(p)\right\|^2\right\} \\ &\quad + \mu_L^2 E\left\{\left\|\xi_{p,k} \mathbf{x}_{L,k}^*(p)\right\|^2\right\}. \end{aligned} \quad (21)$$

Using the independence assumption, we obtain after some mathematical manipulations (see Appendix II-A)

$$E\left\{\left\|[\mathbf{I}_{(2K+1)M} - \mu_L \mathbf{x}_{L,k}^*(p) \mathbf{x}_{L,k}^T(p)] \mathbf{g}_{L,k}(p)\right\|^2\right\} = [1 - 2\mu_L \sigma_x^2 + \mu_L^2 \sigma_x^4 (2K + 1)M] E\{\|\mathbf{g}_{L,k}(p)\|^2\}. \quad (22)$$

Furthermore, using the Gaussian sixth-order moment-factoring theorem [13], the second term on the right of (21) can be approximated by (see Appendix II-B)

$$\mu_L^2 E\left\{\left\|\mathbf{x}_{L,k}^*(p) \mathbf{x}_{Q,k}^T(p) \mathbf{g}_{Q,k}(p)\right\|^2\right\} \approx [\mu_L^2 \sigma_x^6 (2K + 1)M] E\{\|\mathbf{g}_{Q,k}(p)\|^2\}. \quad (23)$$

The evaluation of the last two terms in (21) is straightforward, and they can be expressed as

$$\begin{aligned} \mu_L^2 E\left\{\left\|[\tilde{\mathbf{x}}_{L,k}^T(p) \tilde{\mathbf{h}}_k] \mathbf{x}_{L,k}^*(p)\right\|^2\right\} &= \mu_L^2 \sigma_x^4 \|\tilde{\mathbf{h}}_k\|^2 (2K + 1)M \end{aligned} \quad (24a)$$

$$\begin{aligned} \mu_L^2 E\left\{\left\|\xi_{p,k} \mathbf{x}_{L,k}^*(p)\right\|^2\right\} &= \mu_L^2 \sigma_\xi^2 \sigma_x^2 (2K + 1)M. \end{aligned} \quad (24b)$$

Substituting (22)–(24) into (21), we have an explicit recursive expression for $E\{\|\mathbf{g}_{L,k}(p+1)\|^2\}$

$$E\{\|\mathbf{g}_{L,k}(p+1)\|^2\} = \alpha_L E\{\|\mathbf{g}_{L,k}(p)\|^2\} + \beta_L E\{\|\mathbf{g}_{Q,k}(p)\|^2\} + \gamma_L \quad (25)$$

where

$$\alpha_L \triangleq 1 - 2\mu_L \sigma_x^2 + \mu_L^2 \sigma_x^4 (2K + 1)M \quad (26)$$

$$\beta_L \triangleq \mu_L^2 \sigma_x^6 (2K + 1)M \quad (27)$$

$$\gamma_L \triangleq \mu_L^2 \sigma_x^2 (2K + 1)M [\sigma_\xi^2 + \sigma_x^2 \|\tilde{\mathbf{h}}_k\|^2]. \quad (28)$$

A recursive expression for $E\{\|\mathbf{g}_{Q,k}(p+1)\|^2\}$ is obtained by taking the norm on both sides of (20) and using the Gaussian odd-order moment-factoring theorem

$$\begin{aligned} E\{\|\mathbf{g}_{Q,k}(p+1)\|^2\} &= E\left\{\left\|[\mathbf{I}_{N/2} - \mu_Q \mathbf{x}_{Q,k}^*(p) \mathbf{x}_{Q,k}^T(p)] \mathbf{g}_{Q,k}(p)\right\|^2\right\} \\ &\quad + \mu_Q^2 E\left\{\left\|\mathbf{x}_{Q,k}^*(p) \mathbf{x}_{L,k}^T(p) \mathbf{g}_{L,k}(p)\right\|^2\right\} \\ &\quad + \mu_Q^2 E\left\{\left\|[\tilde{\mathbf{x}}_{L,k}^T(p) \tilde{\mathbf{h}}_k] \mathbf{x}_{Q,k}^*(p)\right\|^2\right\} \\ &\quad + 2\mu_Q^2 \text{Re}\left\{E\left\{\mathbf{g}_{L,k}^H(p) \mathbf{x}_{L,k}^*(p) \mathbf{x}_{Q,k}^T(p)\right\}\right\} \\ &\quad \times \left\{[\tilde{\mathbf{x}}_{L,k}^T(p) \tilde{\mathbf{h}}_k] \mathbf{x}_{Q,k}^*(p)\right\} \\ &\quad + \mu_Q^2 E\left\{\left\|\xi_{p,k} \mathbf{x}_{Q,k}^*(p)\right\|^2\right\} \end{aligned} \quad (29)$$

where the operator $\text{Re}\{\cdot\}$ takes the real part of its argument. Finding an explicit expression for the first term on the right of

(29) is not straightforward; however, using the independence assumption and the Gaussian eighth-order moment-factoring theorem [13], it can be expressed as (see Appendix III-A)

$$E \left\{ \left\| \left[\mathbf{I}_{N/2} - \mu_Q \mathbf{x}_{Q,k}^*(p) \mathbf{x}_{Q,k}^T(p) \right] \mathbf{g}_{Q,k}(p) \right\|^2 \right\} \\ = \left[1 - 2\mu_Q \sigma_x^4 + \mu_Q^2 \sigma_x^8 \frac{N}{2} \right] E \{ \|\mathbf{g}_{Q,k}(p)\|^2 \}, \quad (30)$$

In addition, using the Gaussian sixth-order moment-factoring theorem, the second term on the right of (29) is approximated by (see Appendix III-B)

$$\mu_Q^2 E \left\{ \left\| \mathbf{x}_{Q,k}^*(p) \mathbf{x}_{L,k}^T(p) \mathbf{g}_{L,k}(p) \right\|^2 \right\} \\ \approx \mu_Q^2 \sigma_x^6 \frac{N}{2} E \{ \|\mathbf{g}_{L,k}(p)\|^2 \} \quad (31)$$

where similarly we get

$$\mu_Q^2 E \left\{ \left\| \left[\tilde{\mathbf{x}}_{L,k}^T(p) \tilde{\mathbf{h}}_k \right] \mathbf{x}_{Q,k}^*(p) \right\|^2 \right\} \approx \mu_Q^2 \sigma_x^6 \frac{N}{2} \|\tilde{\mathbf{h}}_k\|^2. \quad (32)$$

The fourth term on the right of (29) is derived in Appendix III-C as

$$2\mu_Q^2 \text{Re} \left\{ \mathbf{g}_{L,k}^H(p) \mathbf{x}_{L,k}^*(p) \mathbf{x}_{Q,k}^T(p) \left[\tilde{\mathbf{x}}_{L,k}^T(p) \tilde{\mathbf{h}}_k \right] \mathbf{x}_{Q,k}^*(p) \right\} = 0. \quad (33)$$

Moreover, the evaluation of the last term in (29) is straightforward, and it can be expressed as

$$\mu_Q^2 E \{ \|\xi_{p,k} \mathbf{x}_{Q,k}^*(p)\|^2 \} = \mu_Q^2 \sigma_x^4 \sigma_\xi^2 \frac{N}{2}. \quad (34)$$

Finally, substituting (30)–(34) into (29), we have an explicit recursive expression for $E\{\|\mathbf{g}_{Q,k}(p+1)\|^2\}$:

$$E\{\|\mathbf{g}_{Q,k}(p+1)\|^2\} = \alpha_Q E\{\|\mathbf{g}_{Q,k}(p)\|^2\} \\ + \beta_Q E\{\|\mathbf{g}_{L,k}(p)\|^2\} + \gamma_Q \quad (35)$$

where

$$\alpha_Q \triangleq 1 - 2\mu_Q \sigma_x^4 + \mu_Q^2 \sigma_x^8 N/2 \quad (36)$$

$$\beta_Q \triangleq 0.5 \mu_Q^2 \sigma_x^6 N \quad (37)$$

$$\gamma_Q \triangleq 0.5 \mu_Q^2 \sigma_x^4 N \left[\sigma_\xi^2 + \sigma_x^2 \|\tilde{\mathbf{h}}_k\|^2 \right]. \quad (38)$$

Equations (18), (25)–(28), and (35)–(38) represent the mse transient behavior of the proposed adaptive algorithm in the k th frequency bin, using $2K+1$ crossband filters and $N/2$ quadratic cross-terms. As expected from the parallel structure of the model, one can observe the coupling between the recursive (25) and (35). Accordingly, the convergence rate of the linear component of the model depends on that of its quadratic counterpart, and vice versa. This dependency, however, may be controlled by the step-size value of each component.

In this context, it should be noted that the transient behavior of a purely linear model can be obtained as a special case of the above analysis by substituting $\mu_Q = 0$ into (35)–(38),

which yields $\alpha_Q = 1$ and $\beta_Q = \gamma_Q = 0$. Therefore, assuming the adaptive vectors are initialized with zeros, we have $E\{\|\mathbf{g}_{Q,k}(p)\|^2\} = \|\tilde{\mathbf{c}}_k\|^2$, and the resulting mse is given by

$$\epsilon_{k, \text{linear}}(p) = \sigma_\xi^2 + \sigma_x^2 \|\tilde{\mathbf{h}}_k\|^2 + \sigma_x^4 \|\tilde{\mathbf{c}}_k\|^2 + \sigma_x^2 E\{\|\mathbf{g}_{L,k}(p)\|^2\} \quad (39)$$

where

$$E\{\|\mathbf{g}_{L,k}(p+1)\|^2\} = \alpha_{\text{linear}} E\{\|\mathbf{g}_{L,k}(p)\|^2\} + \beta_{\text{linear}} \quad (40)$$

$\alpha_{\text{linear}} = \alpha_L$ [see (26)], and $\beta_{\text{linear}} = \mu_L^2 \sigma_x^2 (2K+1)M[\sigma_\xi^2 + \sigma_x^2 \|\tilde{\mathbf{h}}_k\|^2 + \sigma_x^4 \|\tilde{\mathbf{c}}_k\|^2]$. The error induced by employing a purely linear model for the estimation of nonlinear systems is generally referred to as *nonlinear undermodeling* error [26], [40]–[42]. 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 [23]). In [26], the influence of nonlinear undermodeling in the STFT domain for an off-line estimation scheme was investigated. Next, we analyze the convergence properties of the proposed adaptive algorithm and investigate the influence of the parameter K and the nonlinear undermodeling error on the steady-state mse in each frequency bin.

B. Steady-State Performance

Let us first consider the mean convergence of the misalignment vectors $\mathbf{g}_{L,k}(p)$ and $\mathbf{g}_{Q,k}(p)$. By taking the expected value of both sides of (19) and (20), and by using the Gaussian odd-order moment-factoring theorem, we obtain

$$E\{\mathbf{g}_{L,k}(p+1)\} = [\mathbf{I}_{(2K+1)M} - \mu_L \mathbf{R}_{L,k}^*] E\{\mathbf{g}_{L,k}(p)\} \quad (41)$$

$$E\{\mathbf{g}_{Q,k}(p+1)\} = [\mathbf{I}_{N/2} - \mu_Q \mathbf{R}_{Q,k}^*] E\{\mathbf{g}_{Q,k}(p)\} \quad (42)$$

where $\mathbf{R}_{L,k} = E\{\mathbf{x}_{L,k}(p) \mathbf{x}_{L,k}^H(p)\}$ and $\mathbf{R}_{Q,k} = E\{\mathbf{x}_{Q,k}(p) \mathbf{x}_{Q,k}^H(p)\}$ are the corresponding correlation matrices. Using (71) and (79) from Appendix I, it can be verified that (41) and (42) are convergent if the corresponding step sizes satisfy

$$0 < \mu_L < \frac{2}{\sigma_x^2} \quad (43)$$

$$0 < \mu_Q < \frac{2}{\sigma_x^4} \quad (44)$$

and their steady-state solution is $E\{\mathbf{g}_{L,k}(\infty)\} = E\{\mathbf{g}_{Q,k}(\infty)\} = 0$. Consequently, we get

$$E\{\mathbf{h}_k(\infty)\} = \tilde{\mathbf{h}}_k \quad (45)$$

$$E\{\mathbf{c}_k(\infty)\} = \tilde{\mathbf{c}}_k \quad (46)$$

which indicates that the LMS adaptive vectors $\mathbf{h}_k(p)$ and $\mathbf{c}_k(p)$ converge in the mean to the linear and quadratic components of the true system, respectively. Substituting (45) for $\mathbf{h}_k(p)$ and (46) for $\mathbf{c}_k(p)$ into (18), we find the minimum mse obtainable in the k th frequency bin

$$\epsilon_k^{\min} = \sigma_\xi^2 + \sigma_x^2 \|\tilde{\mathbf{h}}_k\|^2. \quad (47)$$

Note that the unbiased property of the estimators $\mathbf{h}_k(p)$ and $\mathbf{c}_k(p)$ are a consequence of employing a white input signal. However, had the input signal $x_{p,k}$ been correlated, a bias phenomenon could appear, and the adaptive vectors would not converge in mean to the true parameters [43].

We proceed with the mean-square convergence of the adaptive algorithm. Defining

$$\mathbf{q}(p) \triangleq [E\{\|\mathbf{g}_{L,k}(p)\|^2\} \quad E\{\|\mathbf{g}_{Q,k}(p)\|^2\}]^T \quad (48)$$

we combine (25) and (35) and rewrite them in vector form as

$$\mathbf{q}(p+1) = \mathbf{A}\mathbf{q}(p) + \boldsymbol{\gamma} \quad (49)$$

where

$$\mathbf{A} = \begin{bmatrix} \alpha_L & \beta_L \\ \beta_Q & \alpha_Q \end{bmatrix} \quad (50)$$

is a 2×2 matrix, and

$$\boldsymbol{\gamma} = [\gamma_L \quad \gamma_Q]^T. \quad (51)$$

Equation (49) is convergent if and only if the eigenvalues of \mathbf{A} are all within the unit circle. Finding corresponding explicit conditions on the step sizes μ_L and μ_Q is not straightforward. However, sufficient conditions on the step sizes may be derived by assuming that the adaptive vectors $\mathbf{h}_k(p)$ and $\mathbf{c}_k(p)$ are not updated simultaneously. More specifically, assuming that $\mathbf{c}_k(p)$ is constant during the adaptation of $\mathbf{h}_k(p)$ (i.e., $\mu_Q \ll \mu_L$), a sufficient condition for the convergence of (25) is $|\alpha_L| < 1$, which yields

$$0 < \mu_L < \frac{2}{\sigma_x^2(2K+1)M}. \quad (52)$$

Note that since the upper bound of μ_L is inversely proportional to K , a lower step-size value should be utilized with an increasing number of crossband filters, which will result in slower convergence. An optimal step size that results in the fastest convergence of the linear component is then obtained by differentiating α_L with respect to μ_L , which yields $\mu_{L,\text{opt}} = 1/[\sigma_x^2(2K+1)M]$. For the quadratic component, we similarly assume that $\mathbf{h}_k(p)$ is constant during the adaptation of $\mathbf{c}_k(p)$ (i.e., $\mu_L \ll \mu_Q$), which results in the following condition on the step size μ_Q :

$$0 < \mu_Q < \frac{2}{\sigma_x^4 N/2}. \quad (53)$$

The optimal step size for the quadratic component is obtained by differentiating α_Q [see (36)] with respect to μ_Q , which yields $\mu_{Q,\text{opt}} = 1/(\sigma_x^4 N/2)$. It should be noted that when the assumption of separate adaptation of the adaptive vectors does not hold [that is, $\mathbf{h}_k(p)$ and $\mathbf{c}_k(p)$ are updated simultaneously], the convergence of the algorithm is no longer guaranteed by using the derived optimal step sizes. This can easily be shown by substituting $\mu_{L,\text{opt}}$ and $\mu_{Q,\text{opt}}$, respectively, for μ_L and μ_Q in (50), which results in an eigenvalue on the unit circle. Practically, though, the stability of the algorithm can be guaranteed by using the so-called normalized LMS (NLMS) algorithm [13], which also leads to faster convergence.

Provided that μ_L and μ_Q satisfy the convergence conditions of the LMS algorithm, the steady-state mse can be expressed as

$$\epsilon_k(\infty) = \epsilon_k^{\min} + \sigma_x^2 E\{\|\mathbf{g}_{L,k}(\infty)\|^2\} + \sigma_x^4 E\{\|\mathbf{g}_{Q,k}(\infty)\|^2\} \quad (54)$$

where ϵ_k^{\min} is defined in (47), and $E\{\|\mathbf{g}_{L,k}(\infty)\|^2\}$ and $E\{\|\mathbf{g}_{Q,k}(\infty)\|^2\}$ are the steady-state solutions of (25) and (35), which can be derived using (49) as

$$\mathbf{q}(\infty) = \begin{bmatrix} E\{\|\mathbf{g}_{L,k}(\infty)\|^2\} \\ E\{\|\mathbf{g}_{Q,k}(\infty)\|^2\} \end{bmatrix} = [\mathbf{I} - \mathbf{A}]^{-1}\boldsymbol{\gamma}. \quad (55)$$

Finally, substituting (50), (26)–(28), and (36)–(38) into (55), we obtain explicit expressions for $E\{\|\mathbf{g}_{L,k}(\infty)\|^2\}$ and $E\{\|\mathbf{g}_{Q,k}(\infty)\|^2\}$, which we substitute into (54) to obtain, after some manipulation,

$$\epsilon_k(\infty) = f(\mu_L, \mu_Q)\epsilon_k^{\min} \quad (56)$$

where

$$f(\mu_L, \mu_Q) = \frac{2}{2 - \mu_L \sigma_x^2(2K+1)M - \mu_Q \sigma_x^4 N/2}. \quad (57)$$

Equations (47) and (56)–(57) provide an explicit expression for the steady-state mse in the k th frequency bin. Note that since μ_L is inversely proportional to K [see (52)], we expect $f(\mu_L, \mu_Q)$ to be independent of K . Consequently, based on the definition of ϵ_k^{\min} from (47), a lower steady-state mse is expected by increasing the number of estimated crossband filters, as will be further demonstrated in Section V.

Following a similar analysis, the steady-state mse of a purely linear model can be derived by finding a steady-state solution of (39)–(40), which yields

$$\epsilon_{k,\text{linear}}(\infty) = f(\mu_L, 0)\epsilon_{k,\text{linear}}^{\min} \quad (58)$$

where

$$\epsilon_{k,\text{linear}}^{\min} = \sigma_\xi^2 + \sigma_x^2 \|\tilde{\mathbf{h}}_k\|^2 + \sigma_x^4 \|\bar{\mathbf{c}}_k\|^2 \quad (59)$$

represents the minimum mse that can be obtained by employing a linear model in the estimation process. It can be verified from (47), (57) and (59) that $\epsilon_k^{\min} \leq \epsilon_{k,\text{linear}}^{\min}$ and $f(\mu_L, \mu_Q) \geq f(\mu_L, 0)$, which implies that in some cases, a lower steady-state mse might be achieved by using a linear model, rather than a nonlinear one. A similar phenomenon was also indicated in [26] in the context of off-line system identification, where it was shown that the nonlinear undermodeling error is mainly influenced by the NLR. Specifically in our case, let

$$\rho = \sigma_{d_Q}^2 / \sigma_{d_L}^2 \quad (60)$$

denote the NLR, where $\sigma_{d_L}^2 = \sigma_x^2(\|\bar{\mathbf{h}}_k\|^2 + \|\tilde{\mathbf{h}}_k\|^2)$ and $\sigma_{d_Q}^2 = \sigma_x^4 \|\bar{\mathbf{c}}_k\|^2$ are the powers of the output signals of the linear and quadratic components, respectively, and the vectors $\bar{\mathbf{h}}_k$, $\tilde{\mathbf{h}}_k$ and $\bar{\mathbf{c}}_k$ are defined in (15)–(17). Then, the ratio between $\epsilon_{k,\text{linear}}^{\min}$ and ϵ_k^{\min} from (47) can be written as

$$\frac{\epsilon_{k,\text{linear}}^{\min}}{\epsilon_k^{\min}} = 1 + \frac{\|\bar{\mathbf{h}}_k\|^2 + \|\tilde{\mathbf{h}}_k\|^2}{\sigma_\xi^2 / \sigma_x^2 + \|\tilde{\mathbf{h}}_k\|^2} \cdot \rho. \quad (61)$$

Equation (61) indicates that as the nonlinearity becomes stronger (i.e., ρ increases), the minimum mse attainable by the full nonlinear model (ϵ_k^{\min}) would be much lower than that obtained by the purely linear model ($\epsilon_{k,\text{linear}}^{\min}$), such that $\epsilon_k(\infty) < \epsilon_{k,\text{linear}}(\infty)$. On the other hand, the purely linear model may achieve a lower steady-state mse when low NLR values are considered. In the limit, for $\rho \rightarrow 0$, we get $\epsilon_k^{\min} = \epsilon_{k,\text{linear}}^{\min}$, and consequently $\epsilon_{k,\text{linear}}(\infty) < \epsilon_k(\infty)$. Note, however, that since more parameters need to be estimated in the nonlinear model, we expect to obtain (for any NLR value) slower convergence than that of a linear model.

In this context, the close relation to the problems of model-structure selection and model-order selection [44]–[50] should be mentioned. In our case, the model structure is determined by μ_Q , the step size of the nonlinear component of the model. By setting $\mu_Q = 0$, the nonlinearity is ignored and a purely linear model is fitted to the data; whereas for $\mu_Q \neq 0$, the vector $\mathbf{c}_k(p)$ is also updated and a full nonlinear model is employed. Generally (for sufficiently high NLR), as more data is available in the estimation process, a richer structure can be used, and correspondingly, a better estimation can be achieved by incorporating a nonlinear model rather than a linear one. Therefore, the purely linear model is associated with faster convergence, but suffers from higher steady-state mse, compared to using a nonlinear model. 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, at the beginning of the adaptation process, the length of the data is short, and only a few crossband filters are estimated, whether a linear or a nonlinear model is employed. As the adaptation process proceeds, more data can be used, additional crossband filters can be estimated, and lower mse can be achieved. These points will be demonstrated in Section V.

IV. COMPUTATIONAL COMPLEXITY

In this section, we consider the computational complexity of the proposed subband approach, and compare it to that of the conventional time-domain Volterra method.

For subband system identification, the adaptation formulas given in (11) and (12) require $(2K + 1)M + N/2 + 2$ complex multiplications, $(2K + 1)M + N/2$ complex additions, and one complex subtraction to compute the error signal. Moreover, computing the desired signal estimate in (10) results in an additional $2(2K + 1)M + 2N/2 - 1$ arithmetic operations. Note that each arithmetic operation is not carried out every input sample, but only once for every L input samples, where L denotes the decimation factor of the STFT representation. Thus, the adaptation process requires $4(2K + 1)M + 2N + 2$ arithmetic operations for every L input samples and each frequency bin. Finally, repeating the process for each frequency bin, and neglecting the computations required for the forward and inverse STFTs, the complexity associated with the proposed subband approach is given by

$$O_s \sim O\left(\frac{N}{L} \{4[(2K + 1)M + N/2] + 2\}\right). \quad (62)$$

As expected, we observe that the computational complexity increases as K increases. Note that the complexity of the proposed approach may be reduced if the signals are assumed real valued in the time domain, since in this case it is sufficient to consider only the first $N/2 + 1$ frequency bins.

For time-domain system identification, we apply a second-order Volterra model [8] for estimating the quadratically nonlinear system, expressed as

$$\hat{y}(n) = \sum_{m=0}^{N_1-1} h_1(m)x(n-m) + \sum_{m=0}^{N_2-1} \sum_{\ell=m}^{N_2-1} h_2(m,\ell)x(n-m)x(n-\ell) \quad (63)$$

where $h_1(m)$ and $h_2(m,\ell)$ are the linear and quadratic Volterra kernels, respectively, with N_1 and N_2 being their corresponding memory lengths. Note that the triangular Volterra representation is used in (63) for the quadratic kernel [8], [10]. Since the model output depends linearly on the filter coefficients, it can be written in vector form as

$$\hat{y}(n) = \mathbf{h}_1^T(n)\mathbf{x}_1(n) + \mathbf{h}_2^T(n)\mathbf{x}_2(n) \quad (64)$$

where

$$\begin{aligned} \mathbf{h}_1(n) &= [h_1(0) \quad h_1(1) \quad \cdots \quad h_1(N_1 - 1)]^T \\ \mathbf{h}_2(n) &= [h_2(0,0) \quad \cdots \quad h_2(0,N_2 - 1) \quad h_2(1,1) \cdots \\ &\quad h_2(1,N_2 - 1) \quad \cdots \quad h_2(N_2 - 1, N_2 - 1)]^T \end{aligned}$$

are the coefficient vectors of the adaptive linear and quadratic kernels, respectively, and $\mathbf{x}_1(n)$ and $\mathbf{x}_2(n)$ are their corresponding input data vectors. The adaptive vectors are updated using the LMS algorithm as

$$\mathbf{h}_1(n+1) = \mathbf{h}_1(p) + \mu_1 e(n)\mathbf{x}_1^*(p) \quad (65)$$

and

$$\mathbf{h}_2(n+1) = \mathbf{h}_2(p) + \mu_2 e(n)\mathbf{x}_2^*(p) \quad (66)$$

where $e(n) = y(n) - \hat{y}(n)$ is the error signal, $y(n)$ is the system output in the time domain, and μ_1 and μ_2 are the step sizes of the linear and quadratic components of the Volterra model, respectively. Similarly to the subband approach, updating the vectors $\mathbf{h}_1(p)$ and $\mathbf{h}_2(p)$ using (65)–(66), and computing the output signal estimate (64), the computational complexity of the fullband approach can be expressed as

$$O_f \sim O[4(N_1 + \bar{N}_2) + 1] \quad (67)$$

where $\bar{N}_2 = N_2(N_2 + 1)/2$ is the dimension of the vector $\mathbf{h}_2(p)$. Rewriting the subband approach complexity (62) in terms of the fullband parameters (by using the relation $M \approx N_1/L$ [31]), the ratio between the subband and fullband complexities can be written as

$$\frac{O_s}{O_f} \sim \frac{N}{L} \cdot \frac{2N_1(2K + 1)/L + N}{2N_1 + N_2^2}. \quad (68)$$

According to (68), the complexity of the proposed subband approach would be typically lower than that of the conventional fullband approach. This computational efficiency becomes even more significant when systems with relatively large second-order memory length are considered (e.g., nonlinear acoustic echo cancellation applications [1]–[3]). This is because these systems necessitate an extremely large memory length N_2 for the quadratic kernel of the time-domain Volterra model, such that $N \ll N_2^2$ and consequently $O_s \ll O_f$. For instance, with $N = 128$, $L = 64$ (50% overlap), $N_1 = 1024$, $N_2 = 80$, and $K = 2$, computational complexity of the proposed approach is smaller by a factor of 15 compared to that of the fullband approach. Note that the computational efficiency of the proposed approach was also shown in the context of off-line nonlinear system identification [25].

V. EXPERIMENTAL RESULTS

In this section, we present experimental results that verify the mean-square theoretical derivations. The influence of nonlinear undermodeling and the number of crossband filters on the mse performance is also demonstrated. The adaptive algorithm performance is evaluated under the assumption of white Gaussian signals in the STFT domain, for given SNR and NLR values, where the SNR is defined by σ_d^2/σ_ξ^2 and $\sigma_d^2 = E\{|d_{p,k}|^2\}$ denotes the power of the system output signal in the STFT domain. Results are obtained by averaging over 1000 independent runs.

The system to be identified is formed as a parallel combination of linear and quadratic components as described in (2)–(4). The input signal $x_{p,k}$ is a zero-mean white complex Gaussian process with variance σ_x^2 . Note that $x_{p,k}$ is not necessarily a valid STFT signal, as a sequence whose STFT is given by $x_{p,k}$ may not always exist [51]. Similarly, the corrupting noise signal $\xi_{p,k}$ is also a zero-mean white Gaussian process with variance σ_ξ^2 , which is uncorrelated with $x_{p,k}$. We use a Hamming analysis window of length $N = 128$ with 50% overlap (i.e., $L = 0.5N$), and a corresponding minimum-energy synthesis window of length $N = 128$ that satisfies the completeness condition [52]. Note that the true crossband filters of the system $\bar{h}_{p,k,k'}$ are related to the time-domain linear impulse response $\bar{h}(n)$ by [31]

$$\bar{h}_{p,k,k'} = \{\bar{h}(n) * \phi_{k,k'}(n)\}_{n=pL} \quad (69)$$

where the function $\phi_{k,k'}(n)$ depends on the analysis and synthesis windows. Here, we model the linear impulse response $\bar{h}(n)$ as a nonstationary stochastic process with an exponential-decay envelope, i.e., $\bar{h}(n) = u(n)\beta(n)e^{-0.009n}$, where $u(n)$ is the unit step function and $\beta(n)$ is a unit-variance zero-mean white Gaussian noise. The length of the impulse response is set to 768 samples. For the quadratic component, the cross-terms of the system $\{\bar{c}_{k',(k-k') \bmod N} | k' \in \mathcal{F}\}$ are modeled here as a unit-variance zero-mean white Gaussian process.

First, we employ several values of K in order to determine the influence of the number of crossband filters on the mse performance. Since the step size of the linear kernel μ_L should be inversely proportional to K [see (52)], we choose $\mu_L = 0.25/[\sigma_x^2(2K+1)M]$, which ensures convergence. Similarly, the nonlinear component of the model is estimated with a step

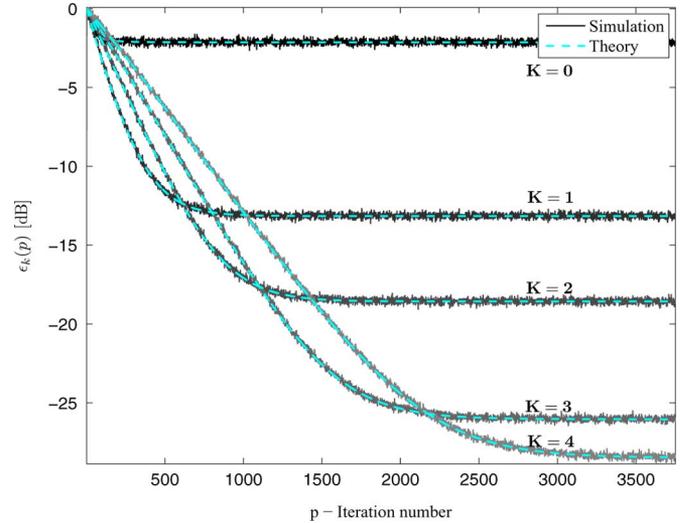


Fig. 3. Comparison of simulation and theoretical convergence of the mse (14) for frequency bin $k = 11$ and white Gaussian signals, as obtained for an SNR of 40 dB, and a nonlinear-to-linear ratio (NLR) of -10 dB.

size of $\mu_Q = 0.25/(\sigma_x^4 N/2)$ [see (53)]. Fig. 3 shows the resulting (normalized) mse curves $\epsilon_k(p)$ [see (18)] for frequency bin $k = 11$, an SNR of 40 dB, and an NLR of -10 dB, as obtained from simulation results and from the theoretical derivations [see (18), (25)–(28), and (35)–(38)]. Clearly, the theoretical analysis accurately describes both the transient and steady-state performance of the adaptive algorithm. The results confirm that as more data is employed in the adaptation process, a lower mse is obtained by estimating additional crossband filters. As expected from (56)–(57), a lower steady-state mse $\epsilon_k(\infty)$ is achieved as K increases; however, the algorithm then suffers from slower convergence. For instance, ignoring the crossband filters and estimating only the band-to-band filters ($K = 0$) yields the fastest convergence, but also results in the highest steady-state mse. Including five crossband filters ($K = 2$), on the other hand, enables a decrease of approximately 16 dB in the steady-state mse, while not greatly slowing convergence. Similar results are obtained for the other frequency bins.

Next, we examine the influence of nonlinear undermodeling on the mse performance. A purely linear model is fitted to the data by setting the step size of the quadratic component to zero (i.e., $\mu_Q = 0$); whereas, a full nonlinear model is employed by updating the quadratic component with a step size of $\mu_Q = 0.25/(\sigma_x^4 N/2)$. For both cases, the linear kernel is updated with step size $\mu_L = 0.25/[\sigma_x^2(2K+1)M]$ for two different values of K ($K = 1$ and 3). Fig. 4 shows the resulting mse curves $\epsilon_k(p)$ and $\epsilon_{k,\text{linear}}(p)$, as obtained from simulation results and from the theoretical derivations [see (18), (25)–(28) and (35)–(38) for the full nonlinear model; and (39)–(40) for the purely linear model], for frequency bin $k = 11$, an SNR of 40 dB, and an NLR of -10 dB [Fig. 4(a)] and -30 dB [Fig. 4(b)]. It can be seen that the experimental results are accurately described by the theoretical mse curves. We observe from 4(a) that for a -10 dB NLR, a lower steady-state mse is achieved by using the nonlinear model. Specifically for $K = 3$, a significant improvement of 12 dB can be achieved over a purely linear model. On the contrary, Fig. 4(b) shows that for a lower NLR value (-30 dB), the inclusion of the nonlinear component in the model is not necessarily

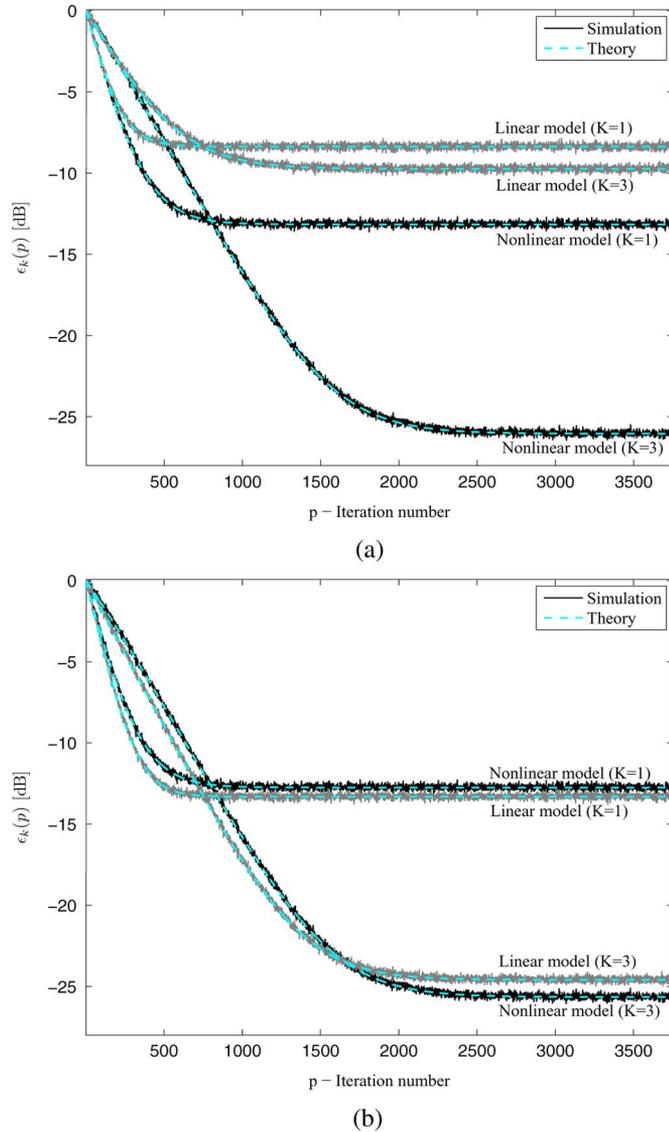


Fig. 4. Comparison of simulation and theoretical curves of the transient mse (14) for frequency bin $k = 11$ and white Gaussian signals, as obtained by using a purely linear model ($\mu_Q = 0$; light) and a nonlinear model ($\mu_Q \neq 0$; dark). (a) Nonlinear-to-linear ratio (NLR) of -10 dB. (b) NLR of -30 dB.

preferable. For example when $K = 1$, the linear model achieves the lowest steady-state mse, while for $K = 3$, the improvement achieved by the nonlinear model is insignificant, and apparently does not justify the substantial increase in model complexity. In general, by further decreasing the NLR, the steady-state mse associated with the linear model decreases, while the relative improvement achieved by the nonlinear model becomes smaller. These results, which were accurately described by the theoretical error analysis in Section III-B [see (56)–(61)], are attributable to the fact that the linear model becomes more accurate as the nonlinearity strength decreases. As a result, the advantage of the nonlinear model due to its improved modeling capability becomes insignificant (i.e., $\epsilon_k^{\min} \approx \epsilon_{k,\text{linear}}^{\min}$), and therefore cannot compensate for the additional adaptation noise caused by also updating the nonlinear component of the model. Another interesting point that can be concluded from the comparison of Fig. 4(a) and (b) is the strategy of controlling the model structure and the model order. Specifically, for high NLR conditions

[Fig. 4(a)], a linear model with a small K should be used at the beginning of the adaptation. Then, the model structure should be changed to nonlinear at an intermediate stage of the adaptation, and the number of estimated crossband filters should increase as the adaptation process proceeds in order to achieve the minimum mse at each iteration. On the other hand, for low NLR conditions [Fig. 4(b)], one would prefer to initially update a purely linear model in order to achieve faster convergence, and then to gradually increase the number of crossband filters. In this case, switching to a different model structure and also incorporating the nonlinear component into the model would be preferable only at an advanced stage of the adaptation process.

VI. CONCLUSION

We have proposed an adaptive scheme for the estimation of quadratically nonlinear systems in the STFT domain, based on the quadratic model proposed in [25]. The proposed model consists of a parallel combination of a linear component, which is represented by crossband filters between subbands, and a quadratic component, modeled by multiplicative cross-terms. We adaptively updated the model parameters using the LMS algorithm and derived explicit expressions for the transient and steady-state mse in frequency bins for white Gaussian inputs. We showed that as more data is employed in the adaptation process, whether a purely linear or a nonlinear model is employed, additional crossband filters should be estimated to achieve the minimum mse at each iteration. We further showed that incorporating the nonlinear component into the model may not necessarily imply a lower steady-state mse in subbands. In fact, the estimation of the nonlinear component improves the mse performance only for high NLR conditions. This improvement in performance becomes smaller as the nonlinearity becomes weaker. It was also shown that the proposed adaptive algorithm is more advantageous in terms of computational complexity than the conventional time-domain Volterra approach.

The adaptive algorithm presented in this paper may be further improved by incorporating adaptive control methods [53]–[57], which dynamically adjust the number of model parameters to provide a balance between complexity, convergence rate, and steady-state performance. Accordingly, by adaptively controlling the model structure (employing either a linear or a nonlinear model) and the model order (determining the number of crossband filters), a full adaptive-control scheme may be constructed to achieve faster convergence without compromising for higher steady-state mse.

APPENDIX I DERIVATION OF (18)

Substituting (17) into (14), and using the independence assumption and the whiteness property of the input signal, the mse can be expressed as

$$\begin{aligned} \epsilon_k(p) = & \sigma_\xi^2 + \sigma_x^2 \|\tilde{\mathbf{h}}_k\|^2 + E \{ \mathbf{g}_{L,k}^T(p) \mathbf{R}_{L,k} \mathbf{g}_{L,k}^*(p) \} \\ & + 2 \text{Re} \{ E \{ \mathbf{g}_{L,k}^T(p) \mathbf{R}_{LQ,k} \mathbf{g}_{Q,k}^*(p) \} \} \\ & + \tilde{\mathbf{h}}_k^T \tilde{\mathbf{R}}_{LQ,k} \mathbf{g}_{Q,k}^*(p) \} \\ & + E \{ \mathbf{g}_{Q,k}^T(p) \mathbf{R}_{Q,k} \mathbf{g}_{Q,k}^*(p) \} \end{aligned} \quad (70)$$

where $\mathbf{R}_{L,k} = E\{\mathbf{x}_{L,k}(p)\mathbf{x}_{L,k}^H(p)\}$, $\mathbf{R}_{Q,k} = E\{\mathbf{x}_{Q,k}(p)\mathbf{x}_{Q,k}^H(p)\}$, $\mathbf{R}_{LQ,k} = E\{\mathbf{x}_{L,k}(p)\mathbf{x}_{Q,k}^H(p)\}$ and $\tilde{\mathbf{R}}_{LQ,k} = E\{\tilde{\mathbf{x}}_{L,k}(p)\mathbf{x}_{Q,k}^H(p)\}$ are correlation matrices, and the operator $\text{Re}\{\cdot\}$ takes the real part of its argument. From (7), the (m, ℓ) th term of $\mathbf{R}_{L,k}$ is given by

$$\begin{aligned} (\mathbf{R}_{L,k})_{m,\ell} &= E \left\{ x_{p-m \bmod M, (k-K+\lfloor \frac{m}{M} \rfloor) \bmod N} \right. \\ &\quad \left. \times x_{p-\ell \bmod M, (k-K+\lfloor \frac{\ell}{M} \rfloor) \bmod N}^* \right\} \\ &= \sigma_x^2 \delta_{m-\ell} \end{aligned} \quad (71)$$

where the last equation is due to the whiteness property of $x_{p,k}$ (see [31], Appendix I.A). In addition, from (9), the (m, ℓ) th term of $\mathbf{R}_{LQ,k}$ can be written as

$$(\mathbf{R}_{LQ,k})_{m,\ell} = E \left\{ x_{p-m \bmod M, (k-K+\lfloor \frac{m}{M} \rfloor) \bmod N} \right. \\ \left. \times x_{p,\ell_k}^* x_{p,(k-\ell_k) \bmod N}^* \right\} \quad (72)$$

where $\ell_k = \ell$ if $\ell \leq (k-1)/2$, and $\ell_k = \ell + (k+1)/2$ otherwise. Since odd-order moments of a zero-mean complex Gaussian process are zero ([13], p. 68), we get

$$(\mathbf{R}_{LQ,k})_{m,\ell} = (\tilde{\mathbf{R}}_{LQ,k})_{m,\ell} = 0. \quad (73)$$

The (m, ℓ) th term of $\mathbf{R}_{Q,k}$ can be written as

$$(\mathbf{R}_{Q,k})_{m,\ell} = E \left\{ x_{p,m_k} x_{p,(k-m_k) \bmod N} \right. \\ \left. \times x_{p,\ell_k}^* x_{p,(k-\ell_k) \bmod N}^* \right\} \quad (74)$$

where m_k is defined similarly to ℓ_k in (72). By using the fourth-order moment factoring theorem for zero-mean complex Gaussian samples ([13], p. 68]), (74) reduces to products of second-order moments as follows:

$$\begin{aligned} (\mathbf{R}_{Q,k})_{m,\ell} &= E \left\{ x_{p,\ell_k}^* x_{p,m_k} \right\} \\ &\quad \times E \left\{ x_{p,(k-\ell_k) \bmod N}^* x_{p,(k-m_k) \bmod N} \right\} \\ &\quad + E \left\{ x_{p,\ell_k}^* x_{p,(k-m_k) \bmod N} \right\} \\ &\quad \times E \left\{ x_{p,(k-\ell_k) \bmod N}^* x_{p,m_k} \right\}. \end{aligned} \quad (75)$$

Using the whiteness property of $x_{p,k}$, we can write (75) as

$$(\mathbf{R}_{Q,k})_{m,\ell} = r_1 + r_2 \quad (76)$$

where

$$r_1 = \sigma_x^4 \delta_{m_k - \ell_k} \delta_{(k-m_k) \bmod N - (k-\ell_k) \bmod N} \quad (77)$$

and

$$r_2 = \sigma_x^4 \delta_{(k-m_k) \bmod N - \ell_k} \delta_{m_k - (k-\ell_k) \bmod N}. \quad (78)$$

Clearly, r_1 is nonzero only if $m_k = \ell_k$ and $(k-m_k) \bmod N = (k-\ell_k) \bmod N$. Using the definitions of m_k and ℓ_k , it is easy to verify that these conditions reduce to $m = \ell$, and therefore $r_1 = \sigma_x^4 \delta_{m-\ell}$. In addition, r_2 is nonzero only if $\ell_k = (k-m_k) \bmod N$ and $m_k = (k-\ell_k) \bmod N$. Note, however, that since $m_k \in \mathcal{T}_1 = \{[0, (k-1)/2] \cup [k+1, (N+k-1)/2]\}$,

the possible values of $(k-m_k) \bmod N$ belong to the set $\mathcal{T}_2 = \{[(k+1)/2, k] \cup [(N+k+1)/2, N-1]\}$. Therefore, since $\mathcal{T}_1 \cap \mathcal{T}_2 = \emptyset$ (an empty set), the conditions imposed in r_2 cannot be satisfied, and we get $r_2 = 0$. Consequently, (76) reduces to

$$(\mathbf{R}_{Q,k})_{m,\ell} = \sigma_x^4 \delta_{m-\ell}. \quad (79)$$

Substituting (71), (73), and (79) into (70) yields (18).

APPENDIX II

EVALUATION OF $E\{\|\mathbf{g}_{L,k}(p+1)\|^2\}$

A. Derivation of (22)

Using the independence assumption of $\mathbf{x}_{L,k}(p)$ and $\mathbf{h}_k(p)$, the first term on the right of (21) can be expressed as

$$\begin{aligned} &E \left\{ \left\| [\mathbf{I}_{(2K+1)M} - \mu_L \mathbf{x}_{L,k}^*(p) \mathbf{x}_{L,k}^T(p)] \mathbf{g}_{L,k}(p) \right\|^2 \right\} \\ &= E \left\{ \|\mathbf{g}_{L,k}(p)\|^2 \right\} - 2\mu_L E \left\{ \mathbf{g}_{L,k}^H(p) \mathbf{A}_k(p) \mathbf{g}_{L,k}(p) \right\} \\ &\quad + \mu_L^2 E \left\{ \mathbf{g}_{L,k}^H(p) \mathbf{B}_k(p) \mathbf{g}_{L,k}(p) \right\} \end{aligned} \quad (80)$$

where

$$\mathbf{A}_k(p) = E \left\{ \mathbf{x}_{L,k}^*(p) \mathbf{x}_{L,k}^T(p) \right\} \quad (81)$$

and

$$\mathbf{B}_k(p) = E \left\{ \mathbf{x}_{L,k}^*(p) \mathbf{x}_{L,k}^T(p) \mathbf{x}_{L,k}^*(p) \mathbf{x}_{L,k}^T(p) \right\}. \quad (82)$$

Using the whiteness property of $x_{p,k}$, $\mathbf{A}_k(p)$ reduces to [see (71)]

$$\mathbf{A}_k(p) = \sigma_x^2 \mathbf{I}_{(2K+1)M} \quad (83)$$

where $\mathbf{I}_{(2K+1)M}$ is the identity matrix of size $(2K+1)M \times (2K+1)M$. The (m, ℓ) th term of $\mathbf{B}_k(p)$ from (82) can be written as

$$\begin{aligned} [\mathbf{B}_k(p)]_{m,\ell} &= \sum_n E \left\{ x_{p-m \bmod M, (k-K+\lfloor \frac{m}{M} \rfloor) \bmod N}^* \right. \\ &\quad \times x_{p-\ell \bmod M, (k-K+\lfloor \frac{\ell}{M} \rfloor) \bmod N} \\ &\quad \times x_{p-n \bmod M, (k-K+\lfloor \frac{n}{M} \rfloor) \bmod N}^* \\ &\quad \left. \times x_{p-n \bmod M, (k-K+\lfloor \frac{n}{M} \rfloor) \bmod N} \right\} \end{aligned} \quad (84)$$

where the index n sums over integer values for which the subscripts of x are defined. By using the fourth-order moment factoring theorem for zero-mean complex Gaussian samples, (84) can be rewritten as

$$\begin{aligned} [\mathbf{B}_k(p)]_{m,\ell} &= \sum_n E \left\{ x_{p-m \bmod M, (k-K+\lfloor \frac{m}{M} \rfloor) \bmod N}^* \right. \\ &\quad \times x_{p-\ell \bmod M, (k-K+\lfloor \frac{\ell}{M} \rfloor) \bmod N} \\ &\quad \times E \left\{ x_{p-n \bmod M, (k-K+\lfloor \frac{n}{M} \rfloor) \bmod N}^* \right. \\ &\quad \left. \times x_{p-n \bmod M, (k-K+\lfloor \frac{n}{M} \rfloor) \bmod N} \right\} \\ &\quad \left. + \sum_n E \left\{ x_{p-m \bmod M, (k-K+\lfloor \frac{m}{M} \rfloor) \bmod N}^* \right. \right. \end{aligned}$$

$$\begin{aligned}
& \times x_{p-n \bmod M, (k-K+\lfloor \frac{n}{M} \rfloor) \bmod N} \} \\
& \times E \left\{ x_{p-n \bmod M, (k-K+\lfloor \frac{n}{M} \rfloor) \bmod N}^* \right. \\
& \left. \times x_{p-\ell \bmod M, (k-K+\lfloor \frac{\ell}{M} \rfloor) \bmod N} \right\} \quad (85)
\end{aligned}$$

where by using the whiteness property of $x_{p,k}$, we obtain [see (71)]

$$[\mathbf{B}_k(p)]_{m,\ell} = \sigma_x^4 \sum_n \delta_{m-\ell} + \sigma_x^4 \sum_n \delta_{m-n} \delta_{\ell-n}. \quad (86)$$

Since n ranges from 0 to $(2K+1)M-1$, (86) reduces to

$$\mathbf{B}_k(p) = \sigma_x^4 [(2K+1)M+1] \mathbf{I}_{(2K+1)M}. \quad (87)$$

Assuming $(2K+1)M \gg 1$, and substituting (83) and (87) into (80) yields (22).

B. Derivation of (23)

Using the independence assumption, the second term on the right of (21) can be expressed as

$$\begin{aligned}
\mu_L^2 E \left\{ \left\| \mathbf{x}_{L,k}^*(p) \mathbf{x}_{Q,k}^T(p) \mathbf{g}_{Q,k}(p) \right\|^2 \right\} \\
= \mu_L^2 E \left\{ \mathbf{g}_{Q,k}^H(p) \mathbf{C}_k(p) \mathbf{g}_{Q,k}(p) \right\} \quad (88)
\end{aligned}$$

where

$$\mathbf{C}_k(p) = E \left\{ \mathbf{x}_{Q,k}^*(p) \mathbf{x}_{L,k}^T(p) \mathbf{x}_{L,k}^*(p) \mathbf{x}_{Q,k}^T(p) \right\}. \quad (89)$$

The (m, ℓ) th term of $\mathbf{C}_k(p)$ can be written as

$$\begin{aligned}
[\mathbf{C}_k(p)]_{m,\ell} \\
= \sum_n E \left\{ x_{p,m_k}^* x_{p-n \bmod M, (k-K+\lfloor \frac{n}{M} \rfloor) \bmod N} \right. \\
\left. \times x_{p,(k-m_k) \bmod N}^* x_{p,\ell_k} x_{p,(k-\ell_k) \bmod N} \right. \\
\left. \times x_{p-n \bmod M, (k-K+\lfloor \frac{n}{M} \rfloor) \bmod N}^* \right\} \quad (90)
\end{aligned}$$

where ℓ_k is defined below (72), and m_k is defined similarly. A similar expression to (90) was derived in [26] using the sixth-order moment factoring theorem for zero-mean complex Gaussian samples [13, p. 68]. Then, following the analysis given in [26, Appendix I-B], we obtain

$$[\mathbf{C}_k(p)]_{m,\ell} = \sigma_x^6 [(2K+1)M + \delta_{m_k \in \mathcal{S}}] \delta_{m-\ell} \quad (91)$$

where $\mathcal{S} = \mathcal{A} \cap \{\mathcal{B}_k \cup \mathcal{B}_0\}$, with $\mathcal{A} \triangleq \{[0, (k-1)/2] \cup [k+1, (N+k-1)/2]\}$ and $\mathcal{B}_k \triangleq \{[(k-K+n_1) \bmod N]M | n_1 \in \{0, \dots, 2K\}\}$. Substituting (91) into (88), and using the definition of $\mathbf{g}_{Q,k}(p)$ from (16), we obtain

$$\begin{aligned}
\mu_L^2 E \left\{ \left\| \mathbf{x}_{L,k}^*(p) \mathbf{x}_{Q,k}^T(p) \mathbf{g}_{Q,k}(p) \right\|^2 \right\} \\
= [\mu_L^2 \sigma_x^6 (2K+1)M] E \left\{ \left\| \mathbf{g}_{Q,k}(p) \right\|^2 \right\} \\
+ \mu_L^2 \sigma_x^6 \sum_{m \in \mathcal{S}} E \left\{ \left| c_{m,(k-m) \bmod N}(p) \right. \right. \\
\left. \left. - \bar{c}_{m,(k-m) \bmod N} \right|^2 \right\}. \quad (92)
\end{aligned}$$

In order to simplify the above expression, let us assume that

$$\begin{aligned}
\sum_{m \in \mathcal{S}} E \left\{ \left| c_{m,(k-m) \bmod N}(p) - \bar{c}_{m,(k-m) \bmod N} \right|^2 \right\} \\
\ll (2K+1)ME \left\{ \left\| \mathbf{g}_{Q,k}(p) \right\|^2 \right\}. \quad (93)
\end{aligned}$$

This assumption is reasonable and can be justified by noting that $\dim \mathcal{S} \leq 4K+2$ and $\max K \ll \dim \mathbf{g}_{Q,k}(p) = N/2$, where the latter is due to fact that most of the energy of the STFT representation of a real-world linear system is concentrated around a few number of crossband filters [31]. Then, neglecting the last term in (92), we obtain (23).

APPENDIX III

EVALUATION OF $E\{\|\mathbf{g}_{Q,k}(p+1)\|^2\}$

In this Appendix, we evaluate the terms in (29), defining $E\{\|\mathbf{g}_{Q,k}(p+1)\|^2\}$.

A. Derivation of (30)

Using the independence assumption of $\mathbf{x}_{Q,k}(p)$ and $\mathbf{c}_k(p)$, the first term on the right of (29) can be expressed as

$$\begin{aligned}
E \left\{ \left\| [\mathbf{I}_{N/2} - \mu_Q \mathbf{x}_{Q,k}^*(p) \mathbf{x}_{Q,k}^T(p)] \mathbf{g}_{Q,k}(p) \right\|^2 \right\} \\
= E \left\{ \left\| \mathbf{g}_{Q,k}(p) \right\|^2 \right\} - 2\mu_Q E \left\{ \mathbf{g}_{Q,k}^H(p) \mathbf{D}_k(p) \mathbf{g}_{Q,k}(p) \right\} \\
+ \mu_Q^2 E \left\{ \mathbf{g}_{Q,k}^H(p) \mathbf{F}_k(p) \mathbf{g}_{Q,k}(p) \right\} \quad (94)
\end{aligned}$$

where

$$\mathbf{D}_k(p) = E \left\{ \mathbf{x}_{Q,k}^*(p) \mathbf{x}_{Q,k}^T(p) \right\} \quad (95)$$

and

$$\mathbf{F}_k(p) = E \left\{ \mathbf{x}_{Q,k}^*(p) \mathbf{x}_{Q,k}^T(p) \mathbf{x}_{Q,k}^*(p) \mathbf{x}_{Q,k}^T(p) \right\}. \quad (96)$$

Using the whiteness property of $x_{p,k}$, $\mathbf{D}_k(p)$ reduces to [see (79)]

$$\mathbf{D}_k(p) = \sigma_x^4 \mathbf{I}_{N/2} \quad (97)$$

where $\mathbf{I}_{N/2}$ is the identity matrix of size $N/2 \times N/2$. The (m, ℓ) th term of $\mathbf{F}_k(p)$ from (96) can be written as

$$\begin{aligned}
[\mathbf{F}_k(p)]_{m,\ell} \\
= \sum_n E \left\{ x_{p,m_k} x_{p,(k-m_k) \bmod N}^* x_{p,\ell_k}^* x_{p,(k-\ell_k) \bmod N}^* \right. \\
\left. \times x_{p,n_k} x_{p,(k-n_k) \bmod N}^* x_{p,n_k}^* x_{p,(k-n_k) \bmod N}^* \right\} \quad (98)
\end{aligned}$$

where ℓ_k is defined below (72), and m_k is defined similarly. Using the Gaussian eighth-order moment-factoring theorem [13, p. 68], (98) can be expressed as

$$\begin{aligned}
[\mathbf{F}_k(p)]_{m,\ell} \\
= \sum_n E \left\{ x_{p,m_k} x_{p,(k-m_k) \bmod N} x_{p,\ell_k}^* x_{p,(k-\ell_k) \bmod N}^* \right. \\
\left. \times E \left\{ x_{p,n_k} x_{p,(k-n_k) \bmod N} x_{p,n_k}^* x_{p,(k-n_k) \bmod N}^* \right\} \right\}
\end{aligned}$$

$$\begin{aligned}
& + \sum_n E \left\{ x_{p,m_k} x_{p,(k-m_k) \bmod N} x_{p,n_k}^* \right. \\
& \times \left. x_{p,(k-n_k) \bmod N}^* \right\} \\
& \times E \left\{ x_{p,n_k} x_{p,(k-n_k) \bmod N} x_{p,\ell_k}^* x_{p,(k-\ell_k) \bmod N}^* \right\} \\
& + \sum_n E \left\{ x_{p,m_k} x_{p,(k-m_k) \bmod N} x_{p,\ell_k}^* x_{p,n_k}^* \right\} \\
& \times E \left\{ x_{p,n_k} x_{p,(k-n_k) \bmod N} x_{p,(k-\ell_k) \bmod N}^* \right. \\
& \times \left. x_{p,(k-n_k) \bmod N}^* \right\} \\
& + \sum_n E \left\{ x_{p,m_k} x_{p,(k-m_k) \bmod N} x_{p,\ell_k}^* \right. \\
& \times \left. x_{p,(k-n_k) \bmod N}^* \right\} \\
& \times E \left\{ x_{p,n_k} x_{p,(k-n_k) \bmod N} x_{p,n_k}^* x_{p,(k-\ell_k) \bmod N}^* \right\}. \quad (99)
\end{aligned}$$

Each term in (99) can be decomposed into products of different combinations of second-order moments, imposing certain conditions on both the matrix indices m and ℓ , and the summation index n . It can be verified that the possible structures of the resulting conditions are $\alpha = \beta$, $(k - \alpha) \bmod N = (k - \beta) \bmod N$, and $\alpha = (k - \beta) \bmod N$, where $\alpha, \beta \in \{m_k, \ell_k, n_k\}$. Then, since the last condition cannot be satisfied [see (78)], and the first two conditions reduce to $m = \ell$ [see (77)], (99) reduces to

$$\begin{aligned}
& [\mathbf{F}_k(p)]_{m,\ell} \\
& = \sigma_x^8 \sum_n (\delta_{m-\ell} + \delta_{n-m} \delta_{\ell-m} + 0 + \delta_{\ell-m} \delta_{n-m} \delta_{\ell-n}) \quad (100)
\end{aligned}$$

and since n ranges from 0 to $N/2 - 1$, we get

$$[\mathbf{F}_k(p)]_{m,\ell} = \sigma_x^8 \left(\frac{N}{2} + 2 \right) \delta_{m-\ell}. \quad (101)$$

Assuming $N \gg 4$, and substituting (97) and (101) into (94) yields (30).

B. Derivation of (31)

Using the independence assumption, the second term on the right of (29) can be expressed as

$$\begin{aligned}
& \mu_Q^2 E \left\{ \left\| \mathbf{x}_{Q,k}^*(p) \mathbf{x}_{L,k}^T(p) \mathbf{g}_{L,k}(p) \right\|^2 \right\} \\
& = \mu_Q^2 E \left\{ \mathbf{g}_{L,k}^H(p) \mathbf{G}_k(p) \mathbf{g}_{Q,k}(p) \right\} \quad (102)
\end{aligned}$$

where

$$\mathbf{G}_k(p) = E \left\{ \mathbf{x}_{L,k}^*(p) \mathbf{x}_{Q,k}^T(p) \mathbf{x}_{Q,k}^*(p) \mathbf{x}_{L,k}^T(p) \right\}. \quad (103)$$

The (m, ℓ) th term of $\mathbf{G}_k(p)$ can be written as

$$\begin{aligned}
& [\mathbf{G}_k(p)]_{m,\ell} \\
& = \sum_n E \left\{ x_{p-m \bmod M, (k-K + \lfloor \frac{m}{M} \rfloor) \bmod N}^* \right. \\
& \times \left. x_{p-\ell \bmod M, (k-K + \lfloor \frac{\ell}{M} \rfloor) \bmod N} x_{p,n_k} \right. \\
& \times \left. x_{p,(k-n_k) \bmod N} x_{p,n_k}^* x_{p,(k-n_k) \bmod N}^* \right\} \quad (104)
\end{aligned}$$

where n_k is defined similarly to ℓ_k in (72). Using the Gaussian sixth-order moment factoring theorem, and following a similar analysis to that given in ([26, Appendix I.B]), we obtain

$$[\mathbf{G}_k(p)]_{m,\ell} = \sigma_x^6 \left[\frac{N}{2} + \delta_{m \in \mathcal{U}} \right] \delta_{m-\ell} \quad (105)$$

where

$$\begin{aligned}
\mathcal{U} & = M \{ [(K - n_1) \bmod N] \\
& \times \cup [(n_1 - k + K) \bmod N] | n_1 \in \mathcal{A} \} \\
\mathcal{A} & \triangleq \{ [0, (k-1)/2] \cup [k+1, (N+k-1)/2] \}.
\end{aligned}$$

Using the definition of $\mathbf{g}_{L,k}(p)$ from (15), and substituting (105) into (102), we obtain

$$\begin{aligned}
& \mu_Q^2 E \left\{ \left\| \mathbf{x}_{Q,k}^*(p) \mathbf{x}_{L,k}^T(p) \mathbf{g}_{L,k}(p) \right\|^2 \right\} \\
& = \mu_Q^2 \sigma_x^6 \left[\frac{N}{2} E \{ \left\| \mathbf{g}_{L,k}(p) \right\|^2 \} + \sum_{m \in \mathcal{U}} E \{ \left\| \mathbf{g}_{L,k}(p) \right\|_m^2 \} \right] \quad (106)
\end{aligned}$$

where $[\mathbf{g}_{L,k}(p)]_m$ denotes the m th term of $\mathbf{g}_{L,k}(p)$. Assuming that $N \gg 2$ and noting that $\dim \mathcal{U} \leq 4K + 2 \ll \dim \mathbf{g}_{L,k}(p)$, we may neglect the last term in (106) to obtain (31).

C. Derivation of (33)

Using the independence assumption, the fourth term on the right-hand side of (29) can be expressed as

$$\begin{aligned}
& 2\mu_Q^2 \text{Re} \left\{ E \left\{ \mathbf{g}_{L,k}^H(p) \mathbf{x}_{L,k}^*(p) \mathbf{x}_{Q,k}^T(p) \right. \right. \\
& \times \left. \left. \left[\tilde{\mathbf{x}}_{L,k}^T(p) \tilde{\mathbf{h}}_k \right] \mathbf{x}_{Q,k}^*(p) \right\} \right\} \\
& = 2\mu_Q^2 \text{Re} \left\{ \sum_{n,\ell,m} f_{nm\ell}(x_{p,k}) E \{ [\mathbf{g}_{L,k}^*(p)]_n \} (\tilde{\mathbf{h}}_k)_\ell \right\} \quad (107)
\end{aligned}$$

where

$$\begin{aligned}
& f_{nm\ell}(x_{p,k}) \\
& = E \left\{ x_{p-n \bmod M, g_1(n)}^* x_{p-\ell \bmod M, g_2(\ell)} \right. \\
& \times \left. x_{p,m_k} x_{p,(k-m_k) \bmod N} x_{p,m_k}^* x_{p,(k-m_k) \bmod N}^* \right\} \quad (108)
\end{aligned}$$

and the functions $g_1(n)$ and $g_2(\ell)$ determine the frequency-bin indices that correspond to the n th term of $\mathbf{x}_{L,k}(p)$ and the ℓ th term of $\tilde{\mathbf{x}}_{L,k}(p)$, respectively. It is easy to verify from the definitions of $\mathbf{x}_{L,k}(p)$ and $\tilde{\mathbf{x}}_{L,k}(p)$ that $g_1(n) \neq g_2(\ell)$ for any pair of indices (n, ℓ) , which implies that $E \{ x_{p-n \bmod M, g_1(n)}^* x_{p-\ell \bmod M, g_2(\ell)} \} = 0$. Consequently, using the Gaussian sixth-order moment factoring theorem and following a similar analysis to that given in ([26, Appendix I.B]), (108) can be written as

$$\begin{aligned}
& f_{nm\ell}(x_{p,k}) \\
& = \sigma_x^6 \delta_{n \bmod M} \delta_{\ell \bmod M} \delta_{g_1(n) - g_2(\ell)} \\
& \times \left[\delta_{m_k - g_1(n)} + \delta_{(k-m_k) \bmod N - g_1(n)} \right]. \quad (109)
\end{aligned}$$

However, since $g_1(n) \neq g_2(\ell)$ we get $f_{nm\ell}(x_{p,k}) = 0$, which can be substituted into (107) to obtain (33).

ACKNOWLEDGMENT

The authors thank the Associate Editor, Dr. D. Morgan, and the anonymous reviewers for their constructive comments and helpful suggestions.

REFERENCES

- [1] A. Stenger and W. Kellermann, "Adaptation of a memoryless pre-processor for nonlinear acoustic echo cancelling," *Signal Process.*, vol. 80, no. 9, pp. 1747–1760, 2000.
- [2] A. Guerin, G. Faucon, and R. L. Bouquin-Jeannés, "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. 1190–1194, Nov. 2006.
- [4] S. Benedetto and E. Biglieri, "Nonlinear equalization of digital satellite channels," *IEEE J. Sel. Areas Commun.*, vol. SAC-1, pp. 57–62, Jan. 1983.
- [5] D. G. Lainiotis and P. Papaparaska, "A partitioned adaptive approach to nonlinear channel equalization," *IEEE Trans. Commun.*, vol. 46, no. 10, pp. 1325–1336, Oct. 1998.
- [6] 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.
- [7] 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, June 1988.
- [8] W. J. Rugh, *Nonlinear System Theory: The Volterra-Wiener Approach*. Baltimore, MD: The Johns Hopkins Univ. Press, 1981.
- [9] 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.
- [10] 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.
- [11] 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, Sept. 1984.
- [12] 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 Process.*, Phoenix, AZ, Mar. 1999, pp. 877–880.
- [13] S. Haykin, *Adaptive Filter Theory*. Englewood Cliffs, NJ: Prentice-Hall, 2002.
- [14] 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.
- [15] A. Fermo, A. Carini, and G. L. Sicuranza, "Simplified Volterra filters for acoustic echo cancellation in GSM receivers," in *Proc. Eur. Signal Process. Conf.*, Tampere, Finland, 2000, pp. 2413–2416.
- [16] F. Kuech and W. Kellermann, "Orthogonalized power filters for nonlinear acoustic echo cancellation," *Signal Process.*, vol. 86, no. 6, pp. 1168–1181, Jun. 2006.
- [17] E. W. Bai and M. Fu, "A blind approach to Hammerstein model identification," *IEEE Trans. Signal Process.*, vol. 50, no. 7, pp. 1610–1619, July 2002.
- [18] T. M. Panicker, "Parallel-cascade realization and approximation of truncated Volterra systems," *IEEE Trans. Signal Process.*, vol. 46, no. 10, pp. 2829–2832, Oct. 1998.
- [19] W. A. Frank, "An efficient approximation to the quadratic Volterra filter and its application in real-time loudspeaker linearization," *Signal Process.*, vol. 45, no. 1, pp. 97–113, Jul. 1995.
- [20] S. W. Nam, S. B. Kim, and E. J. Powers, "On the identification of a third-order Volterra nonlinear systems using a frequency-domain block RLS adaptive algorithm," in *Proc. IEEE Int. Conf. Acoust. Speech, Signal Process.*, Albuquerque, NM, Apr. 1990, pp. 2407–2410.
- [21] C. H. Tseng and E. J. Powers, "Batch and adaptive Volterra filtering of cubically nonlinear systems with a Gaussian input," in *Proc. IEEE Int. Symp. Circuits Syst. (ISCAS)*, 1993, vol. 1, pp. 40–43.
- [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. Breining, P. Dreiseitel, E. Hiinsler, 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.
- [24] P. P. Vaidyanathan, *Multirate Systems and Filter Banks*. Englewood Cliffs, NJ: Prentice-Hall, 1993.
- [25] Y. Avargel and I. Cohen, "Modeling and identification of nonlinear systems in the short-time Fourier transform domain," *IEEE Trans. Signal Process.*, to be published.
- [26] Y. Avargel and I. Cohen, "Performance analysis of nonlinear system identification in the short-time Fourier transform domain," *IEEE Trans. Signal Process.*, submitted for publication.
- [27] W. Kellermann, "Analysis and design of multirate systems for cancellation of acoustical echoes," in *Proc. Int. Conf. Acoust., Speech Signal Process. (ICASSP)*, New York, Apr. 1988, pp. 2570–2573.
- [28] 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.
- [29] 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.
- [30] 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.
- [31] 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.
- [32] 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.
- [33] 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.
- [34] 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.
- [35] L. L. Horowitz and K. D. Senne, "Performance advantage of complex LMS for controlling narrow-band adaptive arrays," *IEEE Trans. Acoust., Speech, Signal Process.*, vol. ASSP-29, no. 3, pp. 722–736, Jun. 1981.
- [36] K. Mayyas, "Performance analysis of the deficient length LMS adaptive algorithm," *IEEE Trans. Signal Process.*, vol. 53, no. 8, pp. 2727–2734, Aug. 2005.
- [37] D. R. Brillinger, *Time Series: Data Analysis and Theory*. Philadelphia, PA: SIAM, 2001.
- [38] 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.
- [39] Y. Ephraim and I. Cohen, R. C. Dorf, Ed., "Recent advancements in speech enhancement," in *The Electrical Engineering Handbook, Circuits, Signals, and Speech and Image Processing*. Boca Raton, FL: CRC, 2006.
- [40] A. E. Nordsjo, B. M. Ninness, and T. Wigren, "Quantifying model errors caused by nonlinear undermodeling in linear system identification," in *Preprints 13th IFAC World Congr.*, San Francisco, CA, July 1996, pp. 145–149.
- [41] B. Ninness and S. Gibson, "Quantifying the accuracy of Hammerstein model estimation," *Automatica*, vol. 38, no. 12, pp. 2037–2051, 2002.
- [42] 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.
- [43] Y. Avargel and I. Cohen, "Performance analysis of cross-band adaptation for subband acoustic echo cancellation," in *Proc. Int. Workshop Acoust. Echo Noise Control (IWAENC)*, Paris, France, Sept. 2006, pp. 1–4, paper no. 8.
- [44] H. Akaike, "A new look at the statistical model identification," *IEEE Trans. Autom. Control*, vol. AC-19, no. 6, pp. 716–723, Dec. 1974.
- [45] J. Rissanen, "Modeling by shortest data description," *Automatica*, vol. 14, no. 5, pp. 465–471, 1978.

- [46] G. Schwarz, "Estimating the dimension of a model," *Ann. Stat.*, vol. 6, no. 2, pp. 461–464, 1978.
- [47] 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.
- [48] L. Ljung, *System Identification: Theory for the User*. Upper Saddle River, NJ: Prentice-Hall, 1999.
- [49] P. Stoica and Y. Selen, "Model order selection: A review of information criterion rules," *IEEE Signal Process. Mag.*, vol. 21, no. 4, pp. 36–A7, Jul. 2004.
- [50] 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.
- [51] D. W. Griffin and J. S. Lim, "Signal estimation from modified short-time Fourier transform," *IEEE Trans. Acoust., Speech, Signal Process.*, vol. ASSP-32, no. 2, pp. 236–243, Apr. 1984.
- [52] J. Wexler and S. Raz, "Discrete Gabor expansions," *Signal Process.*, vol. 21, pp. 207–220, Nov. 1990.
- [53] F. Riera-Palou, J. M. Noras, and D. G. M. Cruickshank, "Linear equalisers with dynamic and automatic length selection," *Electron. Lett.*, vol. 37, no. 25, pp. 1553–1554, Dec. 2001.
- [54] R. C. Bilcu, P. Kuosmanen, and K. Egiuzarian, "A new variable length LMS algorithm: Theoretical analysis and implementations," in *Proc. 9th Int. Conf. Electron., Circuits, Syst.*, Sep. 2002, pp. 1031–1034.
- [55] Y. Gu, K. Tang, H. Cui, and W. Du, "Convergence analysis of a deficient-length LMS filter and optimal-length sequence to model exponential decay impulse response," *IEEE Signal Process. Lett.*, vol. 10, no. 1, pp. 4–7, Jan. 2003.
- [56] Y. Gu, K. Tang, and H. Cui, "LMS algorithm with gradient descent filter length," *IEEE Signal Process. Lett.*, vol. 11, no. 3, pp. 305–307, Mar. 2004.
- [57] Y. Gong and C. F. N. Cowan, "An LMS style variable tap-length algorithm for structure adaptation," *IEEE Trans. Signal Process.*, vol. 53, no. 7, pp. 2400–2407, Jul. 2005.



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. Since 2004, he has been a Research Assistant and a Project Supervisor with the Signal and Image Processing Lab (SIPL), Electrical Engineering Department, Technion. His research interests are statistical signal processing, system identification, adaptive filtering, and digital speech processing.

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, 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, 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 received the Technion Excellent Lecturer awards in 2005 and 2006. He served as Associate Editor of the IEEE TRANSACTIONS ON AUDIO, SPEECH, AND LANGUAGE PROCESSING and the IEEE SIGNAL PROCESSING LETTERS, and as Guest Editor of a Special Issue of the *EURASIP Journal on Advances in Signal Processing* on Advances in Multimicrophone Speech Processing and a Special Issue of the *EURASIP Speech Communication Journal* on Speech Enhancement. He is a coeditor of the Multichannel Speech Processing section of the *Springer Handbook of Speech Processing* (New York: Springer, 2007), and a Co-Chair of the 2010 International Workshop on Acoustic Echo and Noise Control.