Performance Analysis for Subband Identification

Damián Marelli and Minyue Fu, *Senior Member, IEEE*

***Abstract*—The so-called *subband identification* method has been introduced recently as an alternative method for identification of finite-impulse response systems with large tap sizes. It is known that this method can be more numerically efficient than the classical system identification method. However, no results are available to quantify its advantages. This paper offers a rigorous study of the performance of the subband method. More precisely, we aim to compare the performance of the subband identification method with the classical (fullband) identification method. The comparison is done in terms of the asymptotic residual error, asymptotic convergence rate, and computational cost when the identification is carried out using the prediction error method, and the optimization is done using the least-squares method. It is shown that by properly choosing the filterbanks, the number of parameters in each subband, the number of subbands, and the downsampling factor, the two identification methods can have compatible asymptotic residual errors and convergence rate. However, for applications where a high order model is required, the subband method is more numerically efficient. We study two types of subband identification schemes: one using critical sampling and another one using oversampling. The former is simpler to use and easier to understand, whereas the latter involves more design problems but offers further computational savings.**

***Index Terms*—Filterbanks, multirate systems, performance analysis, subband adaptive filtering, system identification.**

I. Introduction

THIS paper studies the use of the so-called *subband identification* method for identification of linear time-invariant systems. This is a relatively new approach and is intended to replace the classical linear system identification technique for applications where the system model is a finite-impulse-response (FIR) filter with a large tap size.

The key idea of the subband identification method is to divide the given input and output signals of the system to be identified into a number of subbands in the frequency domain by using filterbanks and downsamplers. A subband model of the system is then identified in each subband. There are two main types of subband identification schemes: one using critical-sampling and another using oversampling. Critical-sampling refers to the scheme where the number of subbands is equal to the downsampling factor, whereas in the oversampling scheme, the number of subbands is larger than the downsampling factor. For comparative purposes, we will refer to the classical system identification method as fullband identification. See [1] and [2] for an introduction to fullband identification. In comparison with the fullband identification method, the subband method offers two main advantages.

1) *Lower computational cost:* This is mainly due to the fact that the signal rate for each subband is much slower compared with that of the fullband signals, and each subband model requires a much smaller tap size. These features are partly counterbalanced by the extra computation required for forming subband signals, but through a careful choice of the design parameters (including the number of subbands, filterbanks, and subband models), significant computational savings can be achieved using subband identification.
2) *Better numerical properties:* Because each subband model requires a much smaller tap size (compared with the fullband model), a much better level of numerical stability can be obtained in the subband method for the identified model parameters.

These advantages have been recognized in various ways in the literature. Subband identification has been used in speech signal processing applications where long FIR models are often required. For examples, see [3]–[9]. In general, there is a crossing of aliases between subband channels due to filter overlaps; see [3]. There are two main approaches to cope with this problem. The first approach uses critical sampling by applying nonoverlapping filterbanks, which results in spectral gaps between subbands; see [4]. To overcome this problem, [5] used auxiliary channels, with the corresponding extra computational cost. Finally, [6] introduced the use of adaptive cross-terms between subbands. However, these cross-terms increase the computational cost and slow the convergence rate. The second approach uses oversampling. For example, [7] analyzed the existence of exact solutions of the identification problem without cross-terms. In [8], Gabor expansion is used to design the filterbanks, which restricts the flexibility for the filterbanks. Finally, [9] studied the performance of the oversampling case under a number of simplifying assumptions.

However, the research work available in the literature is far from enough to quantify the advantages of the subband identification method. Since there are a number of design parameters available for subband identification (as pointed out earlier) and there are a number of performance indices (such as computational cost, asymptotic residual error and convergence rate), the subband method may or may not be outperformed by the fullband method when measured on a particular performance index. As a result of this, exact comparison with the fullband method is rather difficult. To make the problem more complicated, it is not trivial to know how to optimize the design parameters.

The purpose of this paper is to deal with the aforementioned matters, that is, we aim to quantify the advantages of the subband identification method and study the problem of how to optimize the subband design parameters. To this end, we first give an introduction to the subband identification approach. We

Manuscript received February 5, 2002; revised April 24, 2003. The associate editor coordinating the review of this paper and approving it for publication was Prof. Ioan Tabus.

The authors are with the School of Electrical Engineering and Computer Science, University of Newcastle, N.S.W. 2308, Callaghan, Australia (e-mail: eemf@ee.newcastle.edu.au).

Digital Object Identifier 10.1109/TSP.2003.820083

will show why this approach can be more numerically efficient than the fullband approach for applications where a high order model is required. Second, we compare the performance of the subband identification technique with that of the traditional fullband identification technique in terms of the asymptotic residual error, asymptotic convergence rate, and computational cost. This comparison is done under the assumption that the identification is carried out using the prediction error method and the optimization is done using the least-squares method. This comparison is used to demonstrate the potential power of the subband technique. We study the two types of subband identification schemes, i.e., critical-sampling and oversampling. The former scheme is simpler to use and easier to understand, whereas the latter scheme involves more design problems but offers further computational savings.

This paper is a companion paper of [10], where we studied the asymptotic properties of a subband identification scheme. The results of [10] are used in this paper. Although we have tried to make this paper independent of [10], readers who are interested in detailed studies may find it necessary to read [10] as well. This paper is also partly based on our previous work [11]–[13]. In [11], we studied the critical sampling case. In [12], we investigated the design of the filterbanks that minimize the asymptotic residual error in both critical-sampling and oversampling cases. Finally, [13] is an earlier and simpler version of this paper.

The outline of the paper is as follows. In Section II, we give a review of the fullband identification technique and discuss results from our previous paper [10] that are relevant to subband identification. In Section III, we introduce the subband identification method together with its different analysis approaches. In Section IV, we deal with the conditions for the subbands to be decoupled. In Section V, we introduce the formal assumptions for our study. In Section VI, we provide expressions for the performance indices mentioned above (i.e., asymptotic residual error, asymptotic convergence rate and computational cost). In Section VII, we consider design issues to optimize the performance indices in the critical-sampling case, and in Section VIII, we do the same for the oversampling case. Finally, in Section IX, we give an example to illustrate the performance of the subband identification technique.

II. Fullband Identification

In this section, we give a short review of the well-known theory of linear system identification (see [1] and [2]), which we call the fullband identification method. The setting of the identification problem is illustrated in Fig. 1, where $u(t)$ is the input signal, $w(t)$ is the output of the system, $y(t)$ is the measured output, $v(t)$ is the measurement noise, $g(q)$ (q is the forward shift operator, i.e., $qx(t) = x(t+1)$) is the transfer function of the system, $\hat{g}(q,\theta)$ is the model of the system, and $\hat{v}(t,\theta)$ is the prediction error. The column vector $\theta \in \mathbb{C}^{n_f}$ represents the parameters of the model.

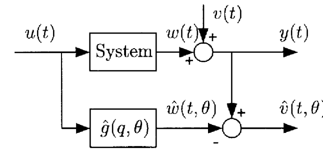

Fig. 1. Block diagram of fullband identification.

A. Some Definitions on Random Processes

The following definitions are needed for our development. See [10] for a more detailed exposition.

Convention 1: All the random processes and linear systems considered in this paper are assumed to be discrete-time, scalar, and complex, unless explicitly specified otherwise. The superscript * denotes complex conjugate, $\mathbb{Z}$ denotes the set of integers, $\mathbb{N}$ denotes the set of positive integers, and $\mathbb{C}$ denotes the set of complex numbers.

Definition 1: Let $p \in \{1,2\}$, and let $\{x(t)\}$ and $\{y(t)\}$, $t \in \mathbb{Z}$ be two random processes. They are said to be *jointly strongly ergodic* of pth order if the following conditions hold.

1) They have uniformly bounded $4p$th moments (i.e., there exists $M_x > 0$ such that $\mathcal{E}\left\{|x(t)|^4\right\}^{1/4} < M_x, \forall\, t \in \mathbb{Z}$, and likewise for $\{y(t)\}$, where $\mathcal{E}\{\cdot\}$ denotes the expectation operator).
2) For any $A, B \in \mathbb{N}$, there exists $C > 0$ such that

$$\left\|C_{xy}^2(A,B,a,b,T)\right\|_p \le \frac{C}{T}, \quad \forall\, a,b \in \mathbb{Z} \tag{1}$$

where $\|\xi\|_p = \mathcal{E}\{\xi^p\}^{1/p}$, and

$$C_{xy}(A,B,a,b,T) = \frac{1}{T}\sum_{t=1}^{T} x^*(At+a)y(Bt+b) - \mathcal{E}\left\{x^*(At+a)y(Bt+b)\right\}. \tag{2}$$

Definition 2: Let $\{x(t)\}$ and $\{y(t)\}$, $t \in \mathbb{Z}$, be two random processes. They are said to be *jointly quasistationary* if there is the following.

1) They have uniformly bounded second-order moments.
2) For all $\tau \in \mathbb{Z}$, the following limit exists:

$$\lim_{T\to\infty} \frac{1}{T}\sum_{t=1}^{T} \mathcal{E}\{x^*(t)y(t+\tau)\}. \tag{3}$$

If in addition, the following limit exists:

$$\lim_{T\to\infty} \frac{1}{T}\sum_{t=1}^{T} \mathcal{E}\{x(At+a)y^*(Bt+b)\}$$

for all $A, B \in \mathbb{N}$ and $a, b \in \mathbb{Z}$, then $\{x(t)\}$ and $\{y(t)\}$ are said to be *jointly quasistationary by phases*. If they further satisfy that for all $D \in \mathbb{N}$ and $\tau \in \mathbb{Z}$

$$\lim_{T\to\infty} \frac{1}{T}\sum_{t=1}^{T} \mathcal{E}\{x^*(Dt)y(Dt+\tau)\} = \lim_{T\to\infty} \frac{1}{T}\sum_{t=1}^{T} \mathcal{E}\{x^*(t)y(t+\tau)\}$$

then they are said to be *almost stationary*. Finally, if they further satisfy that for all $t, \tau \in \mathbb{Z}$

$$\mathcal{E}\{x^*(t)y(t+\tau)\} = \lim_{T\to\infty} \frac{1}{T}\sum_{t=1}^{T} \mathcal{E}\{x^*(t)y(t+\tau)\}$$

then they are said to be *stationary*.

Definition 3: A random process is strongly ergodic (or quasistationary, etc.) if it is jointly strongly ergodic (or jointly quasistationary, etc.) with itself. In addition, a collection of random processes is strongly ergodic (or quasistationary, etc.) if every two random processes in the collection (including a random process with itself) are jointly strongly ergodic (or jointly quasistationary, etc.).

Definition 4: Let $\{x(t)\}$, $t \in \mathbb{Z}$ be a quasistationary random process. The *auto-correlation* of $\{x(t)\}$ is defined by

$$R_x(\tau) = \lim_{T\to\infty} \frac{1}{T} \sum_{t=1}^{T} \mathcal{E}\{x^*(t)x(t+\tau)\}. \tag{4}$$

The *power* of $\{x(t)\}$ is defined by

$$S_x = R_x(0) \tag{5}$$

and the *power spectra* of $\{x(t)\}$ is defined by

$$\Phi_x(\omega) = \sum_{\tau=-\infty}^{\infty} R_x(\tau)e^{-j\omega\tau} \tag{6}$$

provided the infinite sum exists.

B. Formal Assumptions

Notation 1: Define the prediction error $\tilde{w}(t,\theta) = w(t) - \hat{w}(t,\theta)$, and denote its power by $S_{\tilde{w}}(\theta)$. Let

$$S_{\tilde{w},\text{opt}} = \min_{\theta\in\mathcal{D}} S_{\tilde{w}}(\theta).$$

Assumption 1: The signals $\{u(t)\}$, $\{w(t)\}$, and $\{v(t)\}$ satisfy the following.

1) The collection formed by the signals $\{u(t)\}$, $\{w(t)\}$ and $\{v(t)\}$ is strongly ergodic of second order and quasistationary.
2) $\{v(t)\}$ is independent of $\{u(t)\}$ and $\{w(t)\}$.
3) $\{v(t)\}$ is stationary and has zero mean (i.e., $\mathcal{E}\{v(t)\} = 0, \forall t \in \mathbb{N}$).
4) $\tau R_u(\tau), \tau R_v(\tau) \in l_1(\mathbb{Z})$. ($l_1(\mathbb{Z})$ denotes the set of all $x : \mathbb{Z} \to \mathbb{C}$ such that $\sum_{t=-\infty}^{\infty} |x(t)| < \infty$).
5) There exists $\varepsilon > 0$ such that $\Phi_u(\omega) \geq \varepsilon, \forall\, \omega \in [-\pi, \pi]$.

Assumption 2: The model $\hat{g}(q,\theta)$ is an FIR model of tap size n_f. The set of parameters is assumed to satisfy $\theta \in \mathcal{D} \subset \mathbb{C}^{n_f}$, where $\mathcal{D}$ is assumed to be compact (i.e., close and bounded). The set $\mathcal{D}$ satisfies

$$\arg\min_{\theta\in\mathbb{C}^{n_f}} S_{\tilde{w}}(\theta^{n_f}) \subset \text{int}(\mathcal{D}) \tag{7}$$

where int$(\mathcal{D})$ denotes the interior (i.e., excluding the boundary) of $\mathcal{D}$. The identification criterion is the prediction error method, i.e., the optimal vector of parameters up to time N (denoted by θ_N) is chosen as follows:

$$\theta_N \in \arg\min_{\theta\in\mathcal{D}} V_N(\theta) \tag{8}$$

(note that $\arg\min_{\theta\in\mathcal{D}} f(\theta) \subseteq \mathcal{D}$ is a set), where

$$V_N(\theta) = \frac{1}{N}\sum_{t=1}^{N} \frac{1}{2}\left|\hat{v}(t,\theta)\right|^2. \tag{9}$$

Then, θ_N is computed using the least-squares (LS) algorithm, i.e.,

$$\theta_N = [R_N]^{-1}\frac{1}{N}\sum_{t=1}^{N} \varphi^*(t)y(t) \tag{10}$$

where

$$R_N = \frac{1}{N}\sum_{t=1}^{N} \varphi^*(t)\varphi(t) \tag{11}$$

$$\varphi(t) = [u(t), u(t-1), \ldots, u(t-(n_f-1))]. \tag{12}$$

The superscript * above denotes transpose conjugate.

C. Performance Indices

Asymptotic residual error: We know from [10] that under the Assumptions 1 and 2

$$S_{\tilde{w},\lim} := \lim_{N\to\infty} S_{\tilde{w}}(\theta_N) = S_{\tilde{w},\text{opt}} \quad \text{w.p. 1.} \tag{13}$$

In this paper, we are not interested in a bound on $S_{\tilde{w},\text{opt}}$ but in a bound on the difference between the errors of the fullband and the subband methods. This bound is studied in Section VI.

Asymptotic convergence rate: From [10], we have the following result: Suppose Assumptions 1 and 2 are satisfied. If n_f and N are large enough, and $S_{\tilde{w},\lim}$ is small enough, then

$$\mathcal{E}\{S_{\tilde{w},\text{dif}}(N)\} \simeq \frac{n_f}{N} S_v \tag{14}$$

where $S_{\tilde{w},\text{dif}}(N) = S_{\tilde{w}}(\theta_N) - S_{\tilde{w},\lim}$, and S_v is the power of the noise signal $v(t)$.

Computational cost: We assume that the LS solution (10) is implemented using a *recursive least-squares* (RLS) algorithm. The computational cost depends on the particular RLS algorithm used, ranging from $O(n_f)$ to $O(n_f^2)$. Fast algorithms should be used for applications with a high order model, but they tend to be difficult to implement and sensitive numerically; see [14] for a summary of RLS algorithms. For comparative purposes, we consider a reasonably efficient algorithm in [14, Table 6.2, p. 358] called *fast RLS algorithm (version A)*. For that algorithm, the computational cost, measured in terms of number of multiplications per sample, is given by

$$\Psi = 9n_f - 5 \simeq 9n_f. \tag{15}$$

Remark 1: In this paper, we assume that the parameter optimization method is the LS algorithm [(10)–(12)]. This was done in order to guarantee that (8) is satisfied for all $N \in \mathbb{N}$. If a different optimization algorithm such as the LMS algorithm is used, then the results of this paper are still valid in the following sense.

1) The asymptotic residual error will be the same, provided the LMS estimate $\hat{\theta}_N$ converges asymptotically to θ_N with zero error.
2) The asymptotic convergence rate will consist of two terms: one for the convergence of $S_{\tilde{w}}(\theta_N)$ to $S_{\tilde{w},\lim}$, which is the one studied in this paper, and another for the convergence of $S_{\tilde{w}}(\hat{\theta}_N)$ to $S_{\tilde{w}}(\theta_N)$.
3) The computational cost will be obviously reduced further by the use of the LMS algorithm.

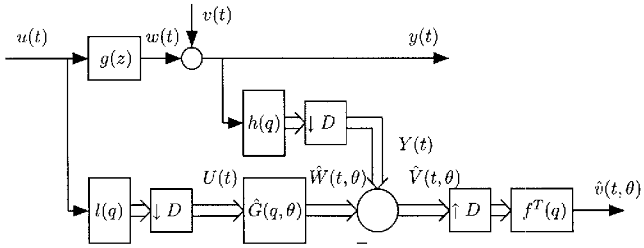

Fig. 2. Subband identification (direct representation).

III. Subband Identification

A. Overview of the Method

The scheme of subband identification is depicted in Fig. 2. As we mentioned in Section I, the idea of subband identification is to split both signals $u(t)$ and $y(t)$ into M subbands using the analysis filterbanks $l(q) = [l_1(q), \ldots, l_M(q)]^T$ and $h(q) = [h_1(q), \ldots, h_M(q)]^T$, respectively. These subband signals are downsampled, and the results are denoted by two vector signals $U(t) = [U_1(t), \ldots, U_M(t)]^T$ and $Y(t) = [Y_1(t), \ldots, Y_M(t)]^T$. The subband parametric model $\hat{G}(q,\theta) = [\hat{G}_{ij}(q,\theta)]_{i,j=1}^M$ is identified in order to reconstruct $\hat{W}(t,\theta) = [\hat{W}_1(t,\theta), \ldots, \hat{W}_M(t,\theta)]^T$: the subband equivalent of $\hat{w}(t,\theta)$. The prediction error $\hat{V}(t,\theta) = [\hat{V}_1(t,\theta), \ldots, \hat{V}_M(t,\theta)]^T = Y(t) - \hat{W}(t,\theta)$ is then formed. Finally, an upsampler and synthesis filterbank $f(q) = [f_1(q), \ldots, f_M(q)]^T$ are used to reconstruct $\hat{v}(t,\theta)$.

We can give now some intuitive explanation about why subband identification can be advantageous compared to fullband identification. To this end, we consider the three key properties mentioned above: asymptotic residual error, convergence rate, and computational cost. For simplicity, we assume the critical-sampling case (i.e., $D = M$). First, we point out that in each subband, the frequency response of the system is much smoother. Therefore, each subband model requires a much lower tap size, compared with the fullband model. It turns out that it is reasonable to take the tap size for each subband model to be $n_s = n_f/M$ plus a small number of taps, which we will ignore here. This guarantees that the subband identification method has a negligible asymptotic residual error. Second, we note that the number of samples in each subband is reduced by a factor of M. This means that the convergence rate remains roughly the same. Third, the computational cost will be M times smaller, because both the tap size and the number of samples in each subband is reduced by a factor of M, and there are M subbands. For RLS algorithms with complexity more than $O(n_f)$, there will be more savings offered by the subband approach. The analysis above clearly shows the advantage of the subband approach. However, we have not taken into account the extra computation required for forming the subband signals. As we will see later, this is a major design issue that determines the efficiency of the subband approach. Nevertheless, we have seen the possibility of using subbands to save computations.

B. Analysis Methods for Subband Identification

There are three representations for analyzing the subband scheme when the signals are assumed to belong to $l_2(\mathbb{Z})$, where $l_2(\mathbb{Z})$ denotes the set of all $x : \mathbb{Z} \to \mathbb{C}$ such that $\sum_{t=-\infty}^{\infty} |x(t)|^2 < \infty$. The first one is the direct representation and corresponds to the scheme in Fig. 2. The other two equivalent representations are the alias representation and the polyphase representation [15]. In this work, we will only use the direct representation and the alias representation, which are introduced below.

Direct representation: This scheme is depicted in Fig. 2. Let the input signals be $u(t)$, $v(t) \in l_2(\mathbb{Z})$ and the impulse responses of the filters be $l(t), h(t), f(t) \in l_2^M(\mathbb{Z})$. We define the following linear maps between fullband and subband domains:

$$\begin{aligned} T_l &: l_2(\mathbb{Z}) \to l_2^M(\mathbb{Z}) : u(t) \mapsto U(t) \\ T_h &: l_2(\mathbb{Z}) \to l_2^M(\mathbb{Z}) : y(t) \mapsto Y(t) \\ T_f^* &: l_2^M(\mathbb{Z}) \to l_2(\mathbb{Z}) : \hat{V}(t,\theta) \mapsto \hat{v}(t,\theta) \end{aligned}$$

where the superscript * denotes the adjoint operator. In addition, we define the norm of T_l by

$$\|T_l\| := \sup_{\|x(t)\|_2 = 1} \|T_l x(t)\|_2 .$$

The norms of $\|T_h\|$ and $\left\|T_f^*\right\|$ are defined in a similar way.

It is straightforward to verify that the conditions for the subband identification scheme to satisfy $\hat{v}(t,\theta) = v(t)$, for all $u(t), v(t) \in l_2(\mathbb{Z})$, are

$$\hat{G}(q,\theta) T_l = T_h g(q) \tag{16}$$

$$T_f^* T_h = I \tag{17}$$

where I is the identity operator. Note that $g(q) : l_2(\mathbb{Z}) \to l_2(\mathbb{Z})$ and $\hat{G}(q,\theta) : l_2^M(\mathbb{Z}) \to l_2^M(\mathbb{Z})$ are operators. Note also that to satisfy (16) and (17), $\hat{G}(q,\theta)$, $h(q)$, and $f(q)$ may need to be noncausal.

Alias representation: In this representation, we use Z-transformed representations of systems and signals. We also use the notation

$$\Omega = e^{-j2\pi/D}. \tag{18}$$

The scheme is depicted in Fig. 3, where

$$U_A(z) = [u(z) u(\Omega z) \cdots u(\Omega^{D-1} z)]^T$$

is the alias representation of the signal $u(z)$ (similar representations for $v(z)$, $w(z)$, and $y(z)$)

$$H_A(z) = \frac{1}{D} \begin{bmatrix} h_1(z) & h_1(\Omega z) & \cdots & h_1(\Omega^{D-1} z) \\ h_2(z) & h_2(\Omega z) & \cdots & h_2(\Omega^{D-1} z) \\ \vdots & \vdots & \ddots & \vdots \\ h_M(z) & h_M(\Omega z) & \cdots & h_M(\Omega^{D-1} z) \end{bmatrix}$$

is the so-called *alias matrix* of the analysis filterbank $h(z)$ (similar notation for $l(z)$), and

$$G_A(z) = \operatorname{diag}\{g(z), g(\Omega z) \ldots, g(\Omega^{D-1} z)\}$$

can be interpreted as the alias representation of the system model $g(z)$.

In this approach, (16) and (17) can be expressed as

$$\hat{G}(z^D, \theta) L_A(z) = H_A(z) G_A(z) \tag{19}$$

$$f^T(z) H_A(z) = [1, 0, \ldots, 0]. \tag{20}$$

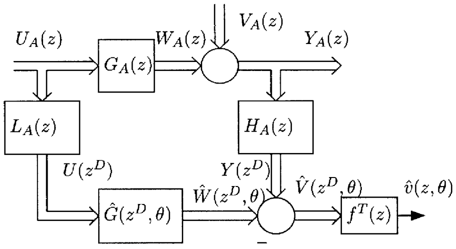

Fig. 3. Alias representation of subband identification.

IV. Decoupling Condition

In general, (16) requires the subband model $\hat{G}(q,\theta)$ to be a full (i.e., nondiagonal) matrix. This complicates the identification process substantially. To simplify the computation, the filterbank $h(q)$ should be designed to reduce the number of nonzero terms in $\hat{G}(q,\theta)$. In this section, we analyze the ideal case where the subband channels are decoupled, which implies that $\hat{G}(q,\theta)$ is a diagonal matrix. We have the following result.

Theorem 1: Consider the subband identification scheme of Fig. 2. Let $h(q)$ be defined such that for each $m = 1,\dots,M$, $h_m(e^{j\omega})$ has its support on a (possibly disjoint) segment $\sigma_m^h \subset [-\pi,\pi]$, i.e.,

$$h_m(e^{j\omega}) \begin{cases} \neq 0, & \omega \in \sigma_m^h \\ = 0, & \omega \notin \sigma_m^h. \end{cases}$$

Let $l(q)$ be defined in a similar way such that $\sigma_m^h \subseteq \sigma_m^l$. If there exists a branch $\beta_m^D(z)$ (see Appendix A for definition) such that $\beta_m^D(e^{j\omega})$ maps $[-\pi,\pi]$ into a segment $\sigma_m \subset [-\pi,\pi]$ satisfying

$$\sigma_m^h \subseteq \sigma_m^l \subseteq \sigma_m \tag{21}$$

then all the diagonal matrices $\hat{G}(q,\theta)$ that satisfy (16) are given by

$$\breve{G}(q) = \mathrm{diag}\{\breve{G}_m(q), (m=1,\dots,M)\} \tag{22}$$

$$\breve{G}_m(e^{j\omega}) = \begin{cases} \frac{h_m(\beta_m^D(e^{j\omega}))}{l_m(\beta_m^D(e^{j\omega}))} g(\beta_m^D(e^{j\omega})), & \omega : \beta_m^D(e^{j\omega}) \in \sigma_m^h \\ 0, & \omega : \beta_m^D(e^{j\omega}) \in \sigma_m^l \setminus \sigma_m^h \\ \text{undefined}, & \omega : \beta_m^D(e^{j\omega}) \in \sigma_m \setminus \sigma_m^l \end{cases} \tag{23}$$

where $\setminus$ denotes the complement set operation, i.e., $A \setminus B = \{x \in A : x \notin B\}$. The converse is also true, i.e., if the subband identification scheme yields a diagonal solution, then the analysis filterbank $h(q)$ satisfies the conditions above, and therefore, all the diagonal solutions satisfy (22) and (23).

This result is a slight generalization of the result in [7]. The only modification we have is that in [7], the filterbanks $l(q)$ and $h(q)$ are assumed to be identical. See Appendix B for proof.

Corollary 1: If $h_m(q) = l_m(q)$, and $\sigma_m = \sigma_m^h (= \sigma_m^l)$, then $\breve{G}_m(t)$ is given by

$$\begin{aligned} \breve{G}_m(t) &= \Upsilon_m(Dt) \\ \Upsilon_m(\tau) &= \Gamma_m(\tau) * g(\tau) \end{aligned} \tag{24}$$

where $*$ denotes the convolution operation, and $\Gamma_m(q)$ is the ideal bandpass filter

$$\Gamma_m(e^{j\omega}) = \begin{cases} D, & \omega \in \sigma_m \\ 0, & \omega \notin \sigma_m. \end{cases} \tag{25}$$

See Appendix B for proof.

Support of ideal analysis filterbanks: From (21), it is clear that the decoupling of subbands implies that the supports σ_m^h and σ_m^l have measure less than or equal to $2\pi/D$. The filters $h_m(e^{j\omega}), m = 1,\dots,M$ are typically approximated using FIR filters. Since the tap size has a negative influence on the computational cost, it is desirable to minimize it. In order to do that, it is required that $\sigma_m^h = \sigma_m^l = \sigma_m$ and that σ_m be a connected subset of $[-\pi,\pi]$ of measure $2\pi/D$. In addition, (20) requires that the union of σ_m equals $[-\pi,\pi]$. In summary, the filters $h(q)$ and $l(q)$ need to meet the following conditions.

C1) $\sigma_m^h = \sigma_m^l = \sigma_m$, for each $m = 1,\dots,M$.
C2) σ_m is a connected subset of $[-\pi,\pi]$, for each $m = 1,\dots,M$.
C3) σ_m has a measure equal to $2\pi/D$, for each $m = 1,\dots,M$.
C4) $\cup_{m=1}^M \sigma_m = [-\pi,\pi]$.

If $D = M$, then these four requirements imply that $h_m(q)$ are simply a set of nonoverlapping ideal bandpass filters with the passband having a bandwidth of $2\pi/D$. In the case where $D < M$, the filters $h_m(q)$ are still ideal bandpass filters with the same bandwidth for the passband, but in this case, their supports are allowed to overlap. It is clear from this discussion that the order of the FIR filters required to approximate the filters $h_m(q)$ is lower in the oversampling case than in the critical-sampling case.

Remark 2: Since, from condition C2, the sets σ_m are connected sets of measure $2\pi/D$, it follows that the filters $h(q)$ and $l(q)$ are complex. If real filters need to be used, condition C2 needs to be violated so the sets σ_m can be split into two symmetric sets of measure π/D. In this case, in terms of computational cost, the extra length of the real filters will be compensated by the fact that the filtering consists of multiplications of real numbers instead of complex. Therefore, no significant computational saving will be achieved, but this is possible to do only in the critical-sampling configuration since the sets σ_m need to equal the range of some branch $\beta_m^D(e^{j\omega})$ and, from Proposition 3 in Appendix A, two symmetric sets that are the range of some branch cannot have partial overlapping as needed for oversampling.

V. Formal Assumptions

In this section, we introduce the formal assumptions required for the analysis of the subband identification method.

Notation 2: Let us denote by $W(t) = [W_1(t),\dots,W_M(t)]^T$ the downsampled version of $h(q)w(t)$. We define $\tilde{W}(t,\theta) = W(t) - \hat{W}(t,\theta) = [\tilde{W}_1(t,\theta),\dots,\tilde{W}_M(t,\theta)]^T$, and denote by $\tilde{w}(t,\theta)$ the signal obtained by upsampling $\tilde{W}(t,\theta)$ followed by filtering with $f(q)$ and summing up the resulting signals. For each subband, we define the prediction error by $\tilde{W}_m(t,\theta) = W_m(t) - \hat{W}_m(t,\theta_m)$, denote its power by $S_{\tilde{w}}(\theta)$, and take

$$\theta_{m,\text{opt}} \in \arg\min_{\theta_m \in \mathcal{D}} S_{\tilde{W}_m}(\theta_m)$$
$$S_{\tilde{W}_m,\text{opt}} = S_{\tilde{W}_m}(\theta_{m,\text{opt}}). \tag{26}$$

We define the fullband equivalent prediction error by $\tilde{w}(t,\theta) = w(t) - \hat{w}(t,\theta)$, denote its power by $S_{\tilde{w}}(\theta)$, and take

$$\theta_{\text{opt}} \in \arg\min_{\theta \in \mathcal{D}} S_{\tilde{w}}(\theta)$$
$$S_{\tilde{w},\text{opt}} = S_{\tilde{w}}(\theta_{\text{opt}}).$$

Assumption 3: The signals $\{u(t)\}$ and $\{v(t)\}$ satisfy Assumption 1 but are strengthened as follows:

1) The collection formed by the signals $\{u(t)\}$ and $\{v(t)\}$ is quasistationary by phases.
2) The signal $\{u(t)\}$ is almost stationary.

Assumption 4: The two analysis filterbanks are the same [i.e., $l(q) = h(q)$]. The filters $h_m(q)$ are FIR with the same tap size l_h, and the filters $f_m(q)$ are FIR with the same tap size l_f. The synthesis filterbank $f(q)$ satisfies the perfect reconstruction condition (17). The impulse responses satisfy $th_m(t) \in l_1(\mathbb{Z})$ and $f_m(t) \in l_1(\mathbb{Z})$ for all $m = 1,\ldots,M$. There exists $\alpha > 0$ such that

$$\frac{1}{D}\sum_{d=0}^{D-1}\left|h_m(\Omega^d e^{j(\omega/D)})\right|^2 \geq \alpha$$
$$\forall\, m = 1,\ldots,M,\ \forall\, \omega \in [-\pi,\pi]$$

where Ω is given by (18).

Assumption 5: The subband model is a diagonal matrix $\hat{G}(q,\theta) = \text{diag}\{\hat{G}_m(q,\theta_m),\ m = 1,\ldots,M\}$, where $\theta = [\theta_1^T,\ldots,\theta_M^T]^T$. Each $\hat{G}_m(q,\theta_m)$ satisfies Assumption 2, i.e., $\hat{G}_m(q,\theta_m)$ is an FIR model of tap size n_s with $\theta_m \in \mathcal{D} \subset \mathbb{C}^{n_s}$, where $\mathcal{D}$ is assumed to be compact. The set $\mathcal{D}$ satisfies

$$\arg\min_{\theta_m \in \mathbb{C}^{n_s}} S_{\tilde{W}_m}(\theta_m) \subset \text{int}(\mathcal{D}).$$

As in the fullband case, the identification criterion is the prediction error method [i.e., (8) and (9)], and the parameter optimization method is the LS algorithm [i.e., (10)–(12)].

Assumption 6: The system $g(q)$ is linear and time-invariant with impulse response satisfying $tg(t) \in l_2(\mathbb{Z})$.

Remark 3: Note that Assumption 6 is satisfied if $g(z)$ is rational and stable.

VI. Performance Indices

As mentioned in Section I, we will evaluate the performance of the subband identification method by comparing it with that of the fullband method. The comparison will be based on three performance indices: asymptotic residual error, asymptotic convergence rate, and computational cost. In this section, we will provide expressions for these three indices. These are general expressions in the sense that they are valid for both critical sampling and oversampling. In Sections VII and VIII, we will provide the final expressions of the performance indices for each particular case.

Asymptotic Residual Error: In the fullband method, the sequence of random variables $S_{\tilde{w}}(\theta_N)$ converges, with probability one, to the deterministic constant $S_{\tilde{w},\text{opt}}$, which is the global minimum of $S_{\tilde{w}}(\theta)$. In the subband method, $S_{\tilde{w}}(\theta_N)$ still converges, with probability one, to a deterministic constant $S_{\tilde{w},\lim}$, but it does not equal $S_{\tilde{w},\text{opt}}$ in general. The following theorem states this fact formally, and its corollary gives the conditions that guarantee $S_{\tilde{w},\lim} = S_{\tilde{w},\text{opt}}$.

Theorem 2 [10, Th. 5]: Consider the subband identification scheme of Fig. 2, together with Assumptions 3–6. Then, there exists a deterministic constant $S_{\tilde{w},\lim} \geq 0$ such that

$$\lim_{N\to\infty} S_{\tilde{w}}(\theta_N) = S_{\tilde{w},\lim} \quad \text{w.p. 1.}$$

Corollary 2: If $D = M$ (critical-sampling case), the analysis filterbank $h(q)$ is paraunitary (i.e., $T_h^* T_h = cI$), and the synthesis filterbank $f(q)$ is given by

$$f(t) = \frac{1}{c}h^*(-t) \tag{27}$$

(i.e., $T_f = 1/cT_h$), then

$$S_{\tilde{w},\lim} = S_{\tilde{w},\text{opt}} = \min_{\theta \in \mathcal{D}} S_{\tilde{w}}(\theta). \tag{28}$$

Proof: See Appendix C. ■

Remark 4: If $D < M$, (28) cannot be guaranteed in general. However, it can be guaranteed by introducing a modification in the identification criterion. This modification was proposed in [16].

As stated in Section II, we are not interested in a bound on $S_{\tilde{w},\lim}$ but in a bound on the difference between $S_{\tilde{w},\lim}$ and the asymptotic residual error of the fullband method. With the abuse of notation, we will denote the fullband asymptotic residual error (13) by $S_{\tilde{w},\lim}^{\text{FB}}$. For technical simplicity, we will provide this bound under the assumption that the input signal $\{u(t)\}$ has a flat power spectra.

Theorem 3: Consider the subband identification scheme of Fig. 2 together with Assumptions 3–6. Suppose that $g(q)$ is identified using the fullband method with an FIR fullband model of tap size n_f. Denote the asymptotic residual error of the fullband method by $S_{\tilde{w},\lim}^{\text{FB}}$. Let the FIR subband models $\hat{G}_m(q,\theta),\ m = 1,\ldots,M$ be given by

$$\hat{G}_m(q,\theta) = \sum_{i=-n_d}^{n_p+n_d-1} \hat{G}_{m,i}q^i \tag{29}$$

where $n_p = \lceil n_f - 1/D \rceil + 1$, and $n_d > 0$. (Here, $\lceil x \rceil$ denotes the smallest integer greater than or equal to x.) If $\{u(t)\}$ has a flat power spectra, then

$$S_{\tilde{w},\lim} \lesssim \|T_f\|^2 \|T_h\|^2 (J + R^2K) \|g(t)\|_1^2 S_u + S_{\tilde{w},\lim}^{\text{FB}} \tag{30}$$

where

$$J = 4\frac{M}{D}F^2; \quad K = \frac{M}{D}C^2 \tag{31}$$

and

$$R = \frac{1}{\sqrt{D}\,\|T_h\|} \max_{1\le m\le M} \{\|h_m(z)\|_\infty\} \tag{32}$$

$$F = \frac{1}{\sqrt{D}\,\|T_h\|} \times \max_{1\le m\le M} \left\{ \frac{1}{2\pi} \int_{[-\pi,\pi]\setminus\sigma_m} |h_m(e^{j\omega})|^2 d\omega \right\}^{1/2} \tag{33}$$

$$C = \left(\sum_{t\notin\{-n_d,\ldots,n_p+n_d-1\}} \operatorname{sinc}^2\left(t-\frac{1}{2}\right)\right)^{1/2} \tag{34}$$

where σ_m is defined in Theorem 1, and $\operatorname{sinc}(t) = \sin(\pi t)/(\pi t)$.

Proof: See Appendix C. ■

Corollary 3: In order to minimize $S_{\tilde{w},\lim}$, the analysis filterbank $h(q)$ needs to be the paraunitary filterbank that minimizes F, and the synthesis filterbank has to be given by (27).

Proof: See Appendix C.

Remark 5: From (29), the tap size of the subband models is given by

$$n_s = n_p + 2n_d \tag{35}$$

where the first n_d parameters are noncausal. Note that this noncausality can be removed in implementation by inserting delays in the subband models.

Remark 6: From Theorem 1 and Corollary 1, we know that if the filterbanks satisfy conditions C1–C4 and the subband models are given by $\hat{G}(q,\theta) = \breve{G}(q)$ [where $\breve{G}(q)$ is given by (24)], then the asymptotic residual error $S_{\tilde{w},\lim}$ is zero. However, these filterbanks and subband models have an infinite tap size. Since we will use FIR approximations (Assumptions 4 and 5), (30) indicates that the bound of $S_{\tilde{w},\lim} - S^{\mathrm{FB}}_{\tilde{w},\lim}$ is made of two components.

E1) error due to FIR approximation of the subband filters $\|T_f\|^2 \|T_h\|^2 J \|g(t)\|_1^2 S_u$;
E2) error due to FIR approximation in the diagonal subband terms $\|T_f\|^2 \|T_h\|^2 R^2 K \|g(t)\|_1^2 S_u$.

Asymptotic Convergence Rate: As in the fullband case, we define $S_{\tilde{w},\mathrm{dif}}(N) = S_{\tilde{w}}(\theta_N) - S_{\tilde{w},\lim}$.

Theorem 4 [10, Th. 7]: Consider the subband identification scheme of Fig. 2 together with Assumptions 3–6. Then, for large n_s and N and small $S_{\tilde{w},\lim}$

$$\mathcal{E}\{S_{\tilde{w},\mathrm{dif}}(N)\} \lesssim \frac{Dn_s}{N} \|T_f\|^2 \|T_h\|^2 S_v \tag{36}$$

where S_v is the power of the noise signal $v(t)$.

Corollary 4: In order to maximize the convergence rate, the analysis filterbank $h(q)$ needs to be paraunitary, and the synthesis filterbank has to be given by (27). In this case, for large n_s and N and small $S_{\tilde{w},\lim}$

$$\mathcal{E}\{S_{\tilde{w},\mathrm{dif}}(N)\} \lesssim \frac{Dn_s}{N} S_v. \tag{37}$$

Proof: The proof is similar to the proof of Corollary 3. ■

Computational cost: Recall that we use an RLS implementation for computing $\theta_{m,N}$. Then, from (15), we have that the computational cost, measured in the amount of multiplications per fullband sample, is

$$\Psi = \frac{M}{D}\left[2l_h + l_f + 9n_s\right]. \tag{38}$$

In Sections VII and VIII, we will use the expressions of the performance indices given in this section to give design issues for the critical-sampling case and the oversampling case. In each case, we provide the folowing:

1) the ideally required analysis $h(q)$ and synthesis $f(q)$ filterbanks;
2) issues for practical implementation: optimal values for M, D, n_s, l_h, and l_f and optimization criteria for the design of the FIR filterbanks.

VII. Critical-Sampling Case

In the critical-sampling case ($M = D$), the number of subbands is minimal. This eliminates the information redundancy, which saves computational cost. In addition, it can be guaranteed that the minimum asymptotic residual is achieved (see Corollary 2).

Ideal filterbanks: Since $M = D$, given an analysis filterbank $h(q)$, the option for the synthesis filterbank $f(q)$ that satisfies the perfect reconstruction condition (17) is unique. Therefore, we just need to provide a choice for the analysis filterbank $h(q)$.

From Corollaries 3 and 4, in order to minimize the asymptotic residual error and maximize the asymptotic convergence rate, the filterbank $h(q)$ has to be paraunitary and minimize F, and $f(q)$ has to be given by (27) (which, in this case, is the only possibility). Ideally, we want to choose $h(q)$ such that $F = 0$. This is achieved if $h(q)$ satisfies conditions C1–C4.

Proposition 1: Consider the subband identification scheme of Fig. 2. Let $h_m(q), m = 1, \ldots, M$ satisfy conditions C1–C4. Then, $h(q)$ is paraunitary if and only if there exists $c > 0$ such that

$$\sum_{m=1}^{M} \left|h_m\left(e^{j\omega}\right)\right|^2 = c. \tag{39}$$

Proof: See Appendix C. ■

In view of Proposition 1, if C1–C4 are satisfied, then $h(q)$ is paraunitary if and only if (39) is satisfied. It is straightforward to verify that C1–C4 and (39) are satisfied by the filters in Fig. 4.

Issues for practical implementation: The ideal filters of Fig. 4 will be approximated by using linear-phase FIR filters whose frequency responses cannot have zero amplitude in their stopband. As a consequence, $F \neq 0$. Then, $h(q)$ needs to be the paraunitary filterbank that minimizes F. The synthesis filterbank $f(q)$ is given by (27), and the number of parameters of the subband models n_s is taken as in (35).

Since the filters $h_m(q)$, $m = 1, \ldots, M$ will be approximations of the filters in Fig. 4, we have $R \simeq 1$ and $\|T_f\|\,\|T_h\| \simeq 1$. Therefore, if $\{u(t)\}$ has a flat power spectra, we obtain from (30) that

$$S_{\tilde{w},\lim} \lesssim (J+K)\,\|g(t)\|_1^2 S_u + S^{\mathrm{FB}}_{\tilde{w},\lim}.$$

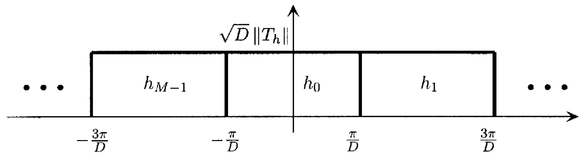

Fig. 4. Analysis filterbank for the critical-sampling case.

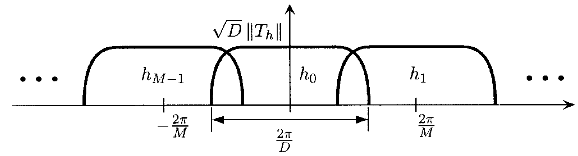

Fig. 5. Analysis filterbank for the oversampling case.

From (37) and (35), for large n_f and N, small $S_{\tilde{w},\lim}$, and $n_f/D \gg n_d$

$$\mathcal{E}\{S_{\tilde{w},\text{dif}}(N)\} \simeq \frac{n_f}{N} S_v. \tag{40}$$

For given values of D and l_h, the required $h(q)$ is uniquely specified. Hence, we can express $F(D, l_h)$ as a function of D and l_h. From (34) and $n_f/D \gg n_d$, we have

$$C \simeq C(n_d) = \left(\sum_{t=n_d}^{\infty} \text{sinc}^2 \left(t - \frac{1}{2} \right) \right)^{1/2}.$$

From (31), we see that if we choose some desired values for J and K, the required tap size $l_h(D, J)$ and subband parameters $n_d(K)$ can be computed numerically. In addition, from (27), we have $l_f = l_h$. Then, (38) becomes

$$\Psi = 3l_h(D, J) + 9 \left(\left\lceil \frac{n_f - 1}{D} \right\rceil + 1 + 2n_d(K) \right). \tag{41}$$

Let us summarize the analysis above. We can see that the convergence rate (40) is independent of the design parameters, and it equals that of the fullband method. The choices of J and K influence the asymptotic residual error for the subband method. Hence, they need to be small. However, they should not be too small, or the computational cost will be too high. Then, for fixed values of J and K, we can numerically optimize D to minimize the computational cost while keeping the asymptotic residual error and convergence rate compatible with those of the fullband method.

VIII. Oversampling Case

The disadvantage of the critical-sampling case is that the filters $h_m(q)$ need a sharp transition band. Consequently, the required filter length is considerably long, which contributes negatively to the computational cost. The idea of oversampling is to increase the value of M to allow the use of filters that are easier to approximate. Of course, more computational cost is required to cope with the extra subbands caused by oversampling, but it turns out that this extra cost can be overweighed by the savings on the filterbank approximations.

Ideal filterbanks: In contrast to the critical-sampling case, for a given analysis filterbank $h(q)$, there are infinitely many synthesis filterbanks that satisfy the perfect reconstruction condition (17).

Following the reasoning introduced for the critical-sampling case, in order to minimize the asymptotic residual error, maximize the asymptotic convergence rate, and making $F = 0$, $h(q)$ needs to satisfy C1–C4 and (39), and $f(q)$ has to be given by (27). In view of Proposition 1, it can be verified that the two conditions on $h(q)$ are satisfied by the filters in Fig. 5, where the shape of the transition bands is proportional to $\sqrt{\omega}$.

Issues for practical implementation: As in the critical-sampling case, the ideal analysis filters $h(q)$ will be approximated by linear-phase FIR filters that are the paraunitary filters that minimize F. The synthesis filters $f(q)$ are given by (27), and n_s is also taken as in (35).

As in the critical-sampling case, if $\{u(t)\}$ has a flat power spectra, then

$$S_{\tilde{w},\lim} \lesssim (J + K) \|g(t)\|_1^2 S_u + S_{\tilde{w},\lim}^{\text{FB}}.$$

Furthermore, for large n_f and N, small $S_{\tilde{w},\lim}$, and $n_f/D \gg n_d$, we have

$$\mathcal{E}\{S_{\tilde{w},\text{dif}}(N)\} \simeq \frac{n_f}{N} S_v.$$

In addition, in this case, we can numerically evaluate the required tap size $l_h(M, D, J)$ and subband parameters $n_d(M, D, K)$. Then, (38) becomes

$$\Psi = \frac{M}{D} \Bigg(3l_h(M, D, J) + 9 \left(\left\lceil \frac{n_f - 1}{D} \right\rceil + 1 + 2n_d(M, D, K) \right) \Bigg). \tag{42}$$

Therefore, for given values of J and K, we can numerically optimize M and D to minimize the computational cost while having asymptotic residual error and convergence rate compatible with those of the fullband method.

IX. Simulations

In Sections VII and VIII, we state that both the critical sampling case and the oversampling case can have the same performance as the fullband method, in terms of asymptotic residual error and convergence rate, but with less computational cost. Furthermore, we expect that the computational savings are more significant in the oversampling case since it includes the critical-sampling case as a particular case. In order to illustrate these points, we identify a linear, time-invariant system using the three methods.

The transfer function of the system is shown in Fig. 6, with $\|g(t)\|_1 = 141$. The power of the input signal is $S_u = 0.0119$, the output power is $S_w = 1$, and the noise power is $S_v = 1$.

We use a tap size of $n_f = 200$ for the fullband method. This choice of n_f means that the fullband model ignores the part of the impulse response after $t = 200$, resulting in an asymptotic residual error of approximately $S_{\tilde{w},\text{opt}} = 0.05$. In order to bound the error of the subband method, we adopt $J = 0.001$ and $K = 0.05$.

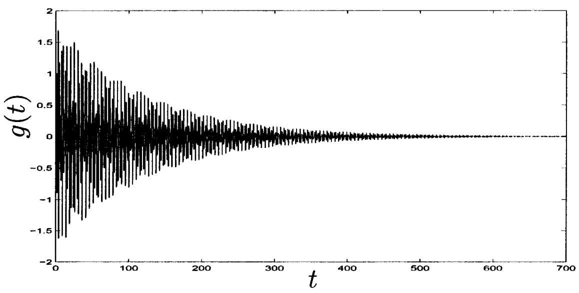

Fig. 6. System impulse response.

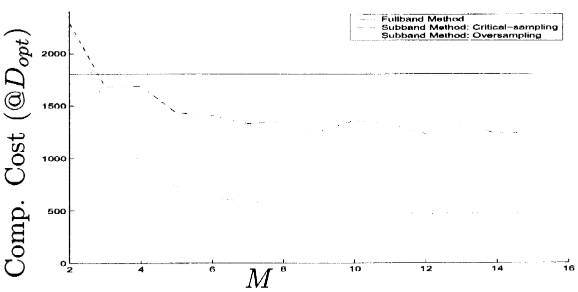

Fig. 7. Computational cost versus M.

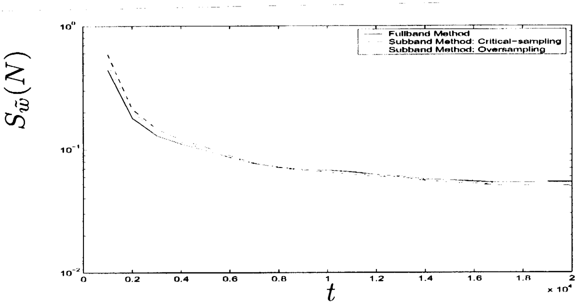

Fig. 8. Evolution of the identification error power.

With these parameters, we optimize the values of M and D to minimize the computational cost for the critical-sampling case (41) and the oversampling case (42). The computational cost as a function of M is shown in Fig. 7. The plot of the computational cost of the oversampling case is obtained by using the optimal value of D for each value of M.

The optimal values are $M = D = 12$ for the critical-sampling case and $M = 12$, $D = 10$ for the oversampling case. With these values we have, for the critical-sampling case $l_h = 354$, $n_s = 19$, $n_d = 1$, and for the oversampling case: $l_h = 63$, $n_s = 22$, $n_d = 1$.

The evolution of $S_{\tilde{w}}(N)$ is shown in Fig. 8. We see that the asymptotic residual error and asymptotic convergence rate are compatible. However, the computational costs are 1800 (multiplications per fullband sample) for the fullband method, 1233 for the critical-sampling subband method, and 464 for the oversampling subband method.

X. Conclusion

In this paper, we have analyzed the performance of the decoupled subband identification method in its two versions (the critical-sampling case and the oversampling case) by comparing them with the classical time-domain identification method (the fullband method). The comparison is based on three performance indices: asymptotic residual error, asymptotic convergence rate, and computational cost. We have provided selection criteria for the filterbanks, the number of parameters of each subband model, the number of subbands, and the downsampling factor. We have shown that the asymptotic residual error and asymptotic convergence rate of the two versions of the subband method are compatible with those of the fullband method. However, if the impulse response of the system to be identified is large enough, the computational costs of the subband methods are smaller. Furthermore, more computational savings can be achieved in the oversampling case. We expect that subband identification methods find applications in various acoustic and speech signal processing problems as well as in wideband communications problems, where high-order FIR models are required.

Appendix A
Roots of z and Its Branches

While using the alias representation, sometimes we meet the expression $z^{1/D}$ (the D-th root of z). We know that if $z \in \mathbb{C}$ is given by $z = re^{j\theta}$, $r > 0$, $-\pi \leq \theta < \pi$, then there are D values for $z^{1/D}$ given by

$$z^{1/D} = r^{1/D} e^{j(2\pi d + \theta/D)}, \quad d = 0, \ldots, D-1.$$

In order to avoid ambiguities, we need the notion of *branches*.

Definition 5: Let $\gamma : [-\pi, \pi) \to \{0, \ldots, D-1\}$ be a map. Then, the γ-branch of the Dth root of $z = re^{j\theta}$ is the map $\beta_\gamma^D : \mathbb{C} \to \mathbb{C}$ defined by

$$\beta_\gamma^D(z) = r^{1/D} e^{j(2\pi\gamma(\theta) + \theta/D)}.$$

Therefore, a branch is a map that sends every z to one of its D possible Dth roots, and the decision is based on the angle θ. We have the following two useful equalities.

Proposition 2: Let $f : \mathbb{C} \to \mathbb{C}$. If $\beta_{\gamma_1}^D, \beta_{\gamma_2}^D$ are two branches of $z^{1/D}$, then

$$\sum_{d=0}^{D-1} f(\Omega^d \beta_{\gamma_1}^D(z)) = \sum_{d=0}^{D-1} f(\Omega^d \beta_{\gamma_2}^D(z)) \tag{43}$$

$$\sum_{d=0}^{D-1} f(\Omega^d \beta_{\gamma_1}^D(z^D)) = \sum_{d=0}^{D-1} f(\Omega^d \beta_{\gamma_2}^D(z^D)) = \sum_{d=0}^{D-1} f(\Omega^d z) \tag{44}$$

where $\Omega = e^{-j(2\pi/D)}$.

Equation (43) means that if we are adding up all the Dth roots, the sum is independent of the branch. Equation (44) means that we can replace $(z^D)^{1/D}$ by z for a similar summation (note that $(z^D)^{1/D} \neq z$ in general).

Notation 3: In view of (43), we will denote the Dth root of z by $z^{1/D}$.

Proposition 3: Let $\beta^D(z)$ be a branch such that $\beta^D(e^{j\omega}) : [-\pi, \pi) \to \sigma$. If σ is symmetric (i.e., if $\omega \in \sigma$, then $-\omega \in \sigma$)

and $\sigma \cap [0,\pi)$ is a connected subset of $[-\pi,\pi)$, then there exists $d \in \{0,\dots,D-1\}$ such that

$$\sigma = \left\{\omega \in [-\pi,\pi) : \pi\frac{d}{D} \le |\omega| < \pi\frac{d+1}{D}\right\}. \tag{45}$$

Proof: Let $\omega_o \in \sigma$. Since σ is the range of a branch, then $\omega_o + 2\pi d/D \notin \sigma$ for $d = 1,\dots D-1$. In addition, because σ is symmetric, $-\omega_o \in \sigma$ and $-\omega_o + 2\pi(d/D) \notin \sigma$ for $d = 1,\dots D-1$. Since $\sigma \cap [0,\pi)$ is connected, it is straightforward to see that the only possibilities for σ are the ones given by (45) ■.

APPENDIX B
PROOFS FOR SECTION IV

Proof of Theorem 1: Denote $\hat{G}(q,\theta)$ by $\hat{G}(q)$ because its dependence on θ is irrelevant in this theorem. Let $\hat{G}(q) = [\hat{G}_{ij}(q)]_{i,j=1}^{M}$. Using the alias representation (Fig. 3), we have that for $m = 1,\dots,M$

$$\begin{aligned} W_m(z) &= \frac{1}{D}\sum_{d=0}^{D-1} h_m(\Omega^d z^{1/D}) \\ &\quad\times g(\Omega^d z^{1/D}) u(\Omega^d z^{1/D}) \end{aligned} \tag{46}$$

$$\begin{aligned} \hat{W}_m(z) &= \sum_{i=1}^{M} \hat{G}_{mi}(z)\frac{1}{D} \\ &\quad\times \sum_{d=0}^{D-1} l_i(\Omega^d z^{1/D}) u(\Omega^d z^{1/D}). \end{aligned} \tag{47}$$

Suppose (21) holds. Recall that by using $z^{1/D}$, we implicitly mean that the expression is independent of the branch used (in view of (43)). This means that in (46) and (47), we can replace $z^{1/D}$ by the branch $\beta_m^D(z)$. Then

$$W_m(z) = \frac{1}{D} g(\beta_m^D(z)) h_m(\beta_m^D(z)) u(\beta_m^D(z)) \tag{48}$$

$$\hat{W}_m(z) = \frac{1}{D}\sum_{i=1}^{M} \hat{G}_{mi}(z) l_i(\beta_m^D(z)) u(\beta_m^D(z)). \tag{49}$$

If we impose the subband model $\hat{G}(z)$ to be a diagonal matrix, then (49) becomes

$$\hat{W}_m(z) = \frac{1}{D}\hat{G}_m(z) l_m(\beta_m^D(z)) u(\beta_m^D(z)) \tag{50}$$

where we used the notation $\hat{G}_m(z) = \hat{G}_{mm}(z)$. By comparing (48) and (50), we conclude that if we want $W_m(e^{j\omega}) = \hat{W}_m(e^{j\omega})$, $\forall\ \omega \in [-\pi,\pi]$, then all the possible diagonal solutions are given by (23).

For the converse, suppose that there exists a diagonal $\hat{G}(z)$ such that $W_m(z) = \hat{W}_m(z)$, $m = 1,\dots,M$. Then, from (46) and (47), we have

$$\begin{aligned} &\frac{1}{D}\sum_{d=0}^{D-1} g(\Omega^d z^{1/D}) h_m(\Omega^d z^{1/D}) u(\Omega^d z^{1/D}) \\ &\qquad = \hat{G}_m(z)\frac{1}{D}\sum_{d=0}^{D-1} l_m(\Omega^d z^{1/D}) u(\Omega^d z^{1/D}) \end{aligned}$$

which can be true only if (21) holds.

Proof of Corollary 1: If $h_m(q) = l_m(q)$ and $\sigma_m^\beta = \sigma_m^h (= \sigma_m^l)$, then

$$\breve{G}_m(e^{j\omega}) = g(\beta_m^D(e^{j\omega})). \tag{51}$$

In addition, from [15, eq. (4.1.4), p. 102], we know that for a given $f(t) \in l_2(\mathbb{Z})$

$$\mathcal{Z}\{f(Dt)\} = \frac{1}{D}\sum_{d=0}^{D-1} f(\Omega^d z^{1/D}). \tag{52}$$

Then

$$\begin{aligned} \{\mathcal{Z}\{(\Gamma_m(\tau) * g(\tau))(Dt)\} &= \frac{1}{D}\sum_{d=0}^{D-1} \Gamma_m(\Omega^d \beta_m^D(z)) \\ &\quad\times g(\Omega^d \beta_m^D(z)) \\ &= \frac{1}{D}\sum_{d=0}^{D-1} \Gamma_m(\Omega^d \beta_m^D(z)) \\ &\quad\times g(\Omega^d \beta_m^D(z)) \\ &= g(\beta_m^D(z)). \end{aligned} \tag{53}$$

Hence, (24) follows from (51) and (53).

APPENDIX C
PROOFS FOR SECTIONS VI AND VII

Lemma 1 [10, Lemma 8]: Let $\{X(t) = [X_1(t),\dots,X_M(t)]^T\}$, $t \in \mathbb{Z}$ be an array of quasistationary random processes, and let $H(q) = [H_{ij}(q)]$, $i,j = 1,\dots,M$ satisfy $H_{ij}(t) \in l_1(\mathbb{Z})$. Let $\breve{X}(t) = [\breve{X}_1(t),\dots,\breve{X}_M(t)]^T$ be generated from $\{X(t)\}$ by upsampling by a factor of U. Let $Y(t) = [Y_1(t),\dots,Y_M(t)]^T$ be defined by $Y(t) = \sum_{k=-\infty}^{\infty} H(k)\breve{X}(Dt-k)$ (i.e., $\{Y(t)\}$ is generated from $\{\breve{X}(t)\}$ by filtering followed by downsampling by a factor D). Let $x(t) = [x_1(t),\dots,x_M(t)]^T \in l_2^M(\mathbb{Z})$, and let $y(t)$ be generated from $x(t)$ in the same way as $\{Y(t)\}$ is generated from $\{X(t)\}$. If there exists $T > 0$ such that $\|y(t)\|_2 \le T\|x(t)\|_2$, where $\|x(t)\|_2^2 = \sum_{m=1}^{M}\|x_m(t)\|_2^2$, then

$$S_Y \le \frac{D}{U}T^2 S_X$$

where $S_X = \sum_{m=1}^{M} S_{X_m}$. Further, if $\|y(t)\|_2 = T\|x(t)\|_2$, then $S_Y = (D/U)T^2 S_X$.

Proof of Corollary 2: Since $h(q)$ is paraunitary, then $(1/\sqrt{c})T_h$ and $\sqrt{c}T_f$ are isometric isomorphisms. Using Lemma 1 and (13), we get

$$\begin{aligned} S_{\tilde{w},\lim} = cS_{\tilde{W},\lim} &= c\sum_{m=1}^{M} S_{\tilde{W}_m,\lim} = c\sum_{m=1}^{M}\min_{\theta\in\mathcal{D}} S_{\tilde{W}_m}(\theta) \\ &= \min_{\theta\in\mathcal{D}} cS_{\tilde{W}}(\theta) = \min_{\theta\in\mathcal{D}} S_{\tilde{w}}(\theta). \end{aligned}$$

Proof of Theorem 3: From [10, Th. 4], we have in every subband

$$\lim_{N\to\infty} S_{\tilde{W}_m}(\theta_{m,N}) = S_{\tilde{W}_m,\mathrm{opt}} \text{ w.p. 1.} \tag{54}$$

Therefore, in the context of this proof, we will assume that the set of parameters θ is given by $\theta = [\theta_{1,\mathrm{opt}}^T,\dots,\theta_{M,\mathrm{opt}}^T]^T$, and we will eliminate from the notation the dependence on the set of parameters. We split the proof into seven steps.

Step 1) Let $\hat{g}(q,\theta)$ be the parametric model used for the full-band method, and denote $\hat{g}^{\mathrm{FB}}(q) = \hat{g}(q,\theta_{\mathrm{opt}})$. In order to prove (30), we will provide a bound of the asymptotic residual error $S_{\tilde{w},\lim}^{\mathrm{dif}}$ obtained when the subband method is used to identify $\hat{g}^{\mathrm{FB}}(q)$. Then, it will be straightforward that

$$S_{\tilde{w},\lim} \le S^{\mathrm{dif}}_{\tilde{w},\lim} + S^{\mathrm{FB}}_{\tilde{w},\lim}$$

Step 2) Assume for a moment that $u(t) \in l_2(\mathbb{Z})$. In view of the alias representation (Fig. 3), we have, for $m = 1,\ldots,M$

$$W_m(z) = \frac{1}{D}\sum_{d=0}^{D-1} h_m(\Omega^d z^{1/D})\hat{g}^{\mathrm{FB}}(\Omega^d z^{1/D})u(\Omega^d z^{1/D})$$
$$\hat{W}_m(z) = \hat{G}_m(z)\frac{1}{D}\sum_{d=0}^{D-1} h_m(\Omega^d z^{1/D})u(\Omega^d z^{1/D}).$$

Then

$$\begin{aligned}\tilde{W}_m(z) &= W_m(z) - \hat{W}_m(z)\\ &= \frac{1}{D}\sum_{d=0}^{D-1} h_m(\Omega^d z^{1/D})\left(\hat{g}^{\mathrm{FB}}(\Omega^d z^{1/D}) - \hat{G}_m(z)\right)\\ &\quad\times u(\Omega^d z^{1/D})\\ &= \frac{1}{D}\sum_{d=0}^{D-1} h_m(\Omega^d z^{1/D})\\ &\quad\times\left(\hat{g}^{\mathrm{FB}}(\Omega^d z^{1/D})\hat{G}_m((\Omega^d z^{1/D})^D)\right)\\ &\quad\times u(\Omega^d z^{1/D}).\end{aligned}$$

Therefore, in the time domain

$$\begin{aligned}\tilde{W}_m(t) &= \downarrow_D \left\{u_m^{\#}(t)\right\} \text{ with}\\ u_m^{\#}(t) &= h_m(t) * \left(\hat{g}^{\mathrm{FB}}(t) - \uparrow_D \left\{\hat{G}_m(t)\right\}\right) * u(t)\end{aligned}$$

where $\downarrow_D, \uparrow_D: l_2(\mathbb{Z}) \to l_2(\mathbb{Z})$ denote the downsampling and upsampling operators, respectively.

Step 3) Now, let $\{u(t)\}$ be the random process satisfying Assumption 3. Since $\{u(t)\}$ is almost stationary, it is straightforward to verify that $\{u_m^{\#}(t)\}$ is also almost stationary. It follows that

$$R_{\tilde{W}_m}(\tau) = R_{u_m^{\#}}(D\tau). \tag{55}$$

From (52), (55), and the fact that $\{u(t)\}$ is white, we get

$$\begin{aligned}\Phi_{\tilde{W}_m}(\omega) &= \mathcal{F}\{R_{\tilde{W}_m}(\tau)\} = \frac{1}{D}\sum_{d=0}^{D-1}\Phi_{u_m^{\#}}\left(2\pi\frac{d}{D} + \frac{\omega}{D}\right)\\ &= \frac{1}{D}\sum_{d=0}^{D-1}\left|h_m(\Omega^d e^{j(\omega/D)})\right|^2\\ &\quad\times\left|\hat{g}^{\mathrm{FB}}(\Omega^d e^{j(\omega/D)}) - \hat{G}_m((\Omega^d e^{j(\omega/D)})^D)\right|^2\\ &\quad\times\Phi_u\left(2\pi\frac{d}{D} + \frac{\omega}{D}\right)\\ &= \frac{1}{D}\sum_{d=0}^{D-1}\left|h_m(\Omega^d e^{j(\omega/D)})\right|^2\\ &\quad\times\left|\hat{g}^{\mathrm{FB}}(\Omega^d e^{j(\omega/D)}) - \hat{G}_m((\Omega^d e^{j(\omega/D)})^D)\right|^2 S_u\end{aligned}$$

and from (43), we have

$$\begin{aligned}\Phi_{\tilde{W}_m}(\omega) &= \frac{1}{D}\sum_{d=0}^{D-1}\left|h_m(\Omega^d\beta_m^D(e^{j\omega}))\right|^2\\ &\quad\times\left|g(\Omega^d\beta_m^D(e^{j\omega})) - \hat{G}_m((\Omega^d\beta_m^D(e^{j\omega}))^D)\right|^2 S_u\\ &= \frac{1}{D}\sum_{d=0}^{D-1}\left|h_m(\Omega^d\beta_m^D(e^{j\omega}))\right|^2\\ &\quad\times\left|g(\Omega^d\beta_m^D(e^{j\omega})) - \hat{G}_m(e^{j\omega})\right|^2 S_u\end{aligned}\tag{56}$$

where $\beta_m^D(z)$ is the branch defined in Theorem 1.

Step 4) Consider the ideal subband (IIR) model $\breve{G}_m(e^{j\omega})$ given by (24), and suppose we truncate it to have the same support as $\hat{G}_m(t)$ does. Denote this truncated version by $\breve{G}_m^{\mathrm{tr}}(t)$. From (54) and (56), we have that

$$\begin{aligned}S_{\tilde{W}_m} &= \frac{1}{2\pi}\int_{-\pi}^{\pi}\left(\frac{1}{D}\sum_{d=0}^{D-1}\left|h_m(\Omega^d\beta_m^D(e^{j\omega}))\right|^2\right.\\ &\qquad\left.\times\left|\hat{g}^{\mathrm{FB}}(\Omega^d\beta_m^D(e^{j\omega})) - \hat{G}_m(e^{j\omega})\right|^2 S_u\right)d\omega\\ &\le \frac{1}{2\pi}\int_{-\pi}^{\pi}\left(\frac{1}{D}\sum_{d=0}^{D-1}\left|h_m(\Omega^d\beta_m^D(e^{j\omega}))\right|^2\right.\\ &\qquad\left.\times\left|\hat{g}^{\mathrm{FB}}(\Omega^d\beta_m^D(e^{j\omega})) - \breve{G}_m^{\mathrm{tr}}(e^{j\omega})\right|^2 S_u\right)d\omega.\end{aligned}$$

Then

$$\begin{aligned}S_{\tilde{W}_m} &\le \frac{1}{D}\sum_{d=0}^{D-1}\left\|h_m(\Omega^d\beta_m^D(e^{j\omega}))\right.\\ &\quad\left.\times\left(\hat{g}^{\mathrm{FB}}(\Omega^d\beta_m^D(e^{j\omega})) - \breve{G}_m^{\mathrm{tr}}(e^{j\omega})\right)\right\|_2^2 S_u\\ &\le \frac{1}{D}\left\|h_m(\beta_m^D(e^{j\omega}))\right\|_\infty^2\\ &\quad\times\left\|\hat{g}^{\mathrm{FB}}(\beta_m^D(e^{j\omega})) - \breve{G}_m^{\mathrm{tr}}(e^{j\omega})\right\|_2^2 S_u +\\ &\quad\frac{1}{D}\sum_{d=1}^{D-1}\left\|h_m(\Omega^d\beta_m^D(e^{j\omega}))\right\|_2^2\\ &\quad\times\left\|\hat{g}^{\mathrm{FB}}(\Omega^d\beta_m^D(e^{j\omega})) - \breve{G}_m^{\mathrm{tr}}(e^{j\omega})\right\|_\infty^2 S_u.\end{aligned}\tag{57}$$

Step 5) Since $\breve{G}_m(e^{j\omega}) = \hat{g}^{\mathrm{FB}}(\beta_m^D(e^{j\omega}))$, we have

$$\begin{aligned}&\left\|\hat{g}^{\mathrm{FB}}(\beta_m^D(e^{j\omega})) - \breve{G}_m^{\mathrm{tr}}(e^{j\omega})\right\|_2^2\\ &\quad= \left\|\breve{G}_m(t) - \breve{G}_m^{\mathrm{tr}}(t)\right\|_2^2\\ &\quad= \sum_{t\notin\{-n_d,\ldots,n_p+n_d-1\}}\left|(\Gamma_m * \hat{g}^{\mathrm{FB}})(Dt)\right|^2\end{aligned}\tag{58}$$

where $\Gamma_m(t)$ is given by (25). Now, by Holder's inequality, we get

$$\begin{aligned}|(\Gamma_m * \hat{g}^{\mathrm{FB}})(Dt)| &\leq \sum_{\tau=0}^{n_f-1} |\hat{g}^{\mathrm{FB}}(\tau)\Gamma_m(Dt-\tau)| \\ &= \sum_{\tau=0}^{n_f-1} |\hat{g}^{\mathrm{FB}}(\tau)\Gamma_m(Dt-\tau)^2|^{1/2} \\ &\quad \times |\hat{g}^{\mathrm{FB}}(\tau)|^{1/2} \\ &\leq \left(\sum_{\tau=0}^{n_f-1} |\hat{g}^{\mathrm{FB}}(\tau)||\Gamma_m(Dt-\tau)|^2\right)^{1/2} \\ &\quad \times \left(\sum_{\tau=0}^{n_f-1} |\hat{g}^{\mathrm{FB}}(\tau)|\right)^{1/2} \\ &= \left\|\hat{g}^{\mathrm{FB}}(t)\right\|_1^{1/2} \\ &\quad \times \left(\sum_{\tau=0}^{n_f-1} |\hat{g}^{\mathrm{FB}}(\tau)||\Gamma_m(Dt-\tau)|^2\right)^{1/2}.\end{aligned}$$

From (58), we obtain

$$\begin{aligned}&\left\|\hat{g}^{\mathrm{FB}}(\beta_m^D(e^{j\omega})) - \breve{G}_m^{\mathrm{tr}}(e^{j\omega})\right\|_2^2 \\ &\leq \left\|\hat{g}^{\mathrm{FB}}(t)\right\|_1 \\ &\quad \times \sum_{t\notin\{-n_d,\ldots,n_p+n_d-1\}} \sum_{\tau=0}^{n_f-1} |\hat{g}^{\mathrm{FB}}(\tau)||\Gamma_m(Dt-\tau)|^2 \\ &= \left\|\hat{g}^{\mathrm{FB}}(t)\right\|_1 \sum_{\tau=0}^{n_f-1} |\hat{g}^{\mathrm{FB}}(\tau)| \\ &\quad \times \sum_{t\notin\{-n_d,\ldots,n_p+n_d-1\}} |\Gamma_m(Dt-\tau)|^2 \leq C_m^2 \left\|\hat{g}^{\mathrm{FB}}(t)\right\|_1^2\end{aligned}$$

where

$$C_m^2 = \max_{0\leq\tau\leq n_f-1}\left(\sum_{t\notin\{-n_d,\ldots,n_p+n_d-1\}} |\Gamma_m(Dt-\tau)|^2\right).$$

In view of conditions C1–C4, the only possibility for $\Gamma_m(t)$ is $\Gamma_m(t) = \mathrm{sinc}(t/D)e^{j\omega_m t}$ for some ω_m. It is straightforward to show that

$$C_m^2 \leq C^2 = \sum_{t\notin\{-n_d,\ldots,n_p+n_d-1\}} \mathrm{sinc}^2\left(t-\frac{1}{2}\right), \quad m = 1,\ldots,M. \tag{59}$$

Finally

$$\left\|\hat{g}^{\mathrm{FB}}(\beta_m^D(e^{j\omega})) - \breve{G}_m^{\mathrm{tr}}(e^{j\omega})\right\|_2^2 \leq C^2 \left\|\hat{g}^{\mathrm{FB}}(t)\right\|_1^2. \tag{60}$$

Step 6) Since, for $m = 1,\ldots,M$, the identified model $\breve{G}_m^{\mathrm{tr}}(e^{j\omega})$ is a truncation of the ideal model $\breve{G}_m(e^{j\omega}) = \hat{g}^{\mathrm{FB}}(\beta_m^D(e^{j\omega}))$, we can take the following approximation:

$$\begin{aligned}&\left\|\hat{g}^{\mathrm{FB}}(\Omega^d\beta_m^D(e^{j\omega})) - \breve{G}_m^{\mathrm{tr}}(e^{j\omega})\right\|_\infty \\ &\simeq \left\|\hat{g}^{\mathrm{FB}}(\Omega^d\beta_m^D(e^{j\omega})) - \hat{g}^{\mathrm{FB}}(\beta_m^D(e^{j\omega}))\right\|_\infty \\ &\leq 2\left\|\hat{g}^{\mathrm{FB}}(e^{j\omega})\right\|_\infty \\ &\leq 2\left\|\hat{g}^{\mathrm{FB}}(t)\right\|_1.\end{aligned} \tag{61}$$

Step 7) Finally, putting together (33), (57), (60), and (61), we have

$$S_{\tilde{W}_m} \lesssim \|T_h\|^2 \left(4F^2 + R_m^2 C^2\right) \left\|\hat{g}^{\mathrm{FB}}(t)\right\|_1^2 S_u$$

where $R_m = \|h_m(z)\|_\infty / (\sqrt{D}\,\|T_h\|)$. Then

$$S_{\tilde{W}} = \sum_{m=1}^{M} S_{\tilde{W}_m} \lesssim M\|T_h\|^2 \left(4F^2 + R^2C^2\right) \left\|\hat{g}^{\mathrm{FB}}(t)\right\|_1^2 S_u.$$

We know that if $\tilde{W}(t) \in l_2^M(\mathbb{Z})$, then $\|\tilde{w}(t)\|_2 \leq \|T_f\| \left\|\tilde{W}(t)\right\|_2$. Then, from Lemma 1, we get

$$\begin{aligned}S_{\tilde{w},\lim}^{\mathrm{dif}} &\lesssim \frac{M}{D}\|T_f\|^2\|T_h\|^2\left(4F^2 + R^2C^2\right)\left\|\hat{g}^{\mathrm{FB}}(t)\right\|_1^2 S_u \\ &\simeq \|T_f\|^2\|T_h\|^2\left(J + R^2K\right)\|g(t)\|_1^2 S_u.\end{aligned}$$

Lemma 2: Consider the subband identification scheme in Fig. 2. Then, we have that

$$\begin{aligned}\sup_{\|x(t)\|_2=1} \|T_h x(t)\|_2^2 &\geq A \\ &:= \frac{1}{D}\left[\inf_{|z|=1}\left\{\sum_{m=1}^{M}|h_m(z)|^2\right\} - \sum_{d=1}^{D-1}\left(\beta(-d)\beta(d)\right)^{1/2}\right]\end{aligned} \tag{62}$$

$$\begin{aligned}\inf_{\|x(t)\|_2=1} \|T_h x(t)\|_2^2 &\leq B \\ &:= \frac{1}{D}\left[\sup_{|z|=1}\left\{\sum_{m=1}^{M}|h_m(z)|^2\right\} + \sum_{d=1}^{D-1}\left(\beta(-d)\beta(d)\right)^{1/2}\right].\end{aligned} \tag{63}$$

Further, the inequalities in (62) and (63) will become equalities if $\beta(d) = 0$, $d = 1$, and $D-1$, where

$$\beta(d) = \sup_{|z|=1}\left\{\sum_{m=1}^{M}|h_m(z)|\left|h_m\left(\Omega^d z\right)\right|\right\}. \tag{64}$$

Proof: Filterbanks can be interpreted as a particular case of the so-called frame decomposition. The frame theory is developed in [17], and its connection with filterbanks is treated in [18]. The proof of this Lemma follows the proof of [17, Sec. 3.3.2, p. 67] which is done for frame decomposition. ■

Proof of Corollary 3: We will assume that $u(t) \in l_2(\mathbb{Z})$. Then, we will look at the analysis $h(q)$ and synthesis $f(q)$ filterbanks whose associated operators T_h and T_f minimize the 2-norm of $\tilde{w}(t)$. In view of Lemma 1, this is equivalent to minimizing the power $S_{\tilde{w}}$.

From the perfect reconstruction condition (16), T_f^* must be a left inverse of T_h. In order to minimize the 2-norm of $\hat{v}$, we must have that T_f^* cancels the orthogonal complement of the range of T_h. It follows that

$$T_f^* = T_h^+ \text{ with } T_h^+ = (T_h^* T_h)^{-1} T_h^*. \tag{65}$$

From (62) and (63), $\|T_h\| \leq B^{1/2}$ and $\|T_f\| \leq A^{-1/2}$. In addition, $A \leq B$; then, from (30), we must have that $A = B$ in order to minimize the asymptotic residual error. This, in turn, means that T_h is a scaled isometry, or equivalently, $h(q)$ is paraunitary. In addition, from (30), we have that $h(q)$ needs to minimize F.

Finally, since T_h is a scaled isometry, (65) implies $T_f = (1/c)T_h$, or equivalently, $f(t) = (1/c)h^*(-t)$.

Proof of Proposition 1: We define $\beta(d)$ as in (64). Conditions C1–C4 imply that $\beta(d) = 0$, and $d = 1, \ldots, D-1$. Then, from Lemma 2

$$\sup_{\|x(t)\|_2=1} \|T_h x(t)\|_2^2 = \frac{1}{D} \inf_{|z|=1} \left\{ \sum_{m=1}^{M} |h_m(z)|^2 \right\}$$
$$\inf_{\|x(t)\|_2=1} \|T_h x(t)\|_2^2 = \frac{1}{D} \sup_{|z|=1} \left\{ \sum_{m=1}^{M} |h_m(z)|^2 \right\}.$$

Then, (39) follows since $h(q)$ being paraunitary is equivalent to T_h being a scaled isometry.

REFERENCES

[1] L. Lennart, *System Identification: Theory for the User*, second ed. Upper Saddle River, NJ: Prentice-Hall, 1999.

[2] T. Söderström and P. Stoica, *System Identification*. New York: Prentice-Hall, 1989.

[3] A. Gilloire, "Experiments with sub-band acoustic echo cancellers for teleconferencing," in *Proc. IEEE Int. Conf. Acoustics, Speech, Signal Process.*, 1987, pp. 2141–2144.

[4] H. Yasukawa, S. Shimada, and I. Furukawa, "Acoustic echo canceller with high speech quality," in *Proc. IEEE Int. Conf. Acoustics, Speech, Signal Process.*, 1987, pp. 2125–2128.

[5] V. S. Somayazulu, S. K. Mitra, and J. J. Shynk, "Adaptive line enhancement using multirate techniques," in *Proc. IEEE Int. Conf. Acoustics, Speech, Signal Process.*, vol. 2, May 1989, pp. 928–931.

[6] A. Gilloire and M. Vetterlli, "Adaptive filtering in subbands with critical sampling: Analysis, experiments, and application to acoustic echo cancellation," *IEEE Trans. Signal Processing*, vol. 40, pp. 1862–1875, Aug. 1992.

[7] W. Kellermann, "Analysis and design of multirate systems for cancellation of acoustical echoes," in *Proc. IEEE Int. Conf. Acoustics, Speech, Signal Process.*, 1988, pp. 2570–2573.

[8] Y. Lu and J. Morris, "Gabor expansion for adaptive echo cancellation," *IEEE Signal Processing Mag.*, vol. 16, pp. 68–80, Mar. 1999.

[9] M. R. Petraglia and S. K. Mitra, "Performance analysis of adaptive filter structures based on subband decomposition," in *Proc. IEEE Int. Symp. Circuits Syst.*, vol. 1, 1993, pp. 60–63.

[10] D. Marelli and M. Fu, "Asymptotic properties of subband identification," *IEEE Trans. Signal Processing*, vol. 51, pp. 3128–3142, Dec. 2003.

[11] ——, "A comparative study on subband identification," in *Proc. IEEE Conf. Decision Contr.*, vol. 3, Dec. 2000, pp. 2409–2415.

[12] ——, "Optimized filterbank design for subband identification with oversampling," in *Proc. IEEE Int. Conf. Acoustics, Speech, Signal Process.*, 2001.

[13] ——, "System identification using subband signal processing," in *Proc. IEEE Conf. Decision Contr.*, vol. 4, Dec. 2001, pp. 3485–3490.

[14] J. G. Proakis, C. M. Raider, F. Ling, and C. L. Nikias, *Advanced Digital Signal Processing*. New York: Maxwell Macmillan Int., 1992.

[15] P. P. Vaidyanathan, *Multirate Systems and Filterbanks*. Englewood Cliffs, N.J: Prentice-Hall, 1993.

[16] D. Marelli and M. Fu, "Convergence properties of subband identification," in *Proc. Asian Contr. Conf.*, 2002.

[17] I. Daubechies, *Ten Lectures on Wavelets*. Philadelphia, PA: SIAM, 1992.

[18] M. Vetterli and J. Kovačević, *Wavelets and Subband Coding*. Englewood Cliffs, N.J: Prentice-Hall, 1995.

Damián Marelli received the Bachelors degree in electronics engineering from the Universidad Nacional de Rosario, Rosario, Argentina, in 1995. He received the Ph.D. degree from the School of Electrical Engineering and Computer Science, University of Newcastle, Newcastle, Australia, in 2003.

He currently holds a research fellowship at the University of Newcastle. He worked as a software engineer at Tesis Ingeniería Informática S.R.L., Rosario, from 1995 to 1996 and at BLC S.A., Rosario, in 1998. In 1997, he held a teaching assistantship and a research assistantship at the Universidad Nacional de Rosario. His main research interests include joint time-frequency signal analysis, system identification, and adaptive filtering.

Minyue Fu (SM'94) received the Bachelors degree in electrical engineering from the China University of Science and Technology, Hefei, China, in 1982 and the M.S. and Ph.D. degrees in electrical engineering from the University of Wisconsin, Madison, in 1983 and 1987, respectively.

From 1983 to 1987, he held a teaching assistantship and a research assistantship at the University of Wisconsin. He was a Computer Engineering Consultant at Nicolet Instruments, Inc., Madison, during 1987. From 1987 to 1989, he was an Assistant Professor with the Department of Electrical and Computer Engineering, Wayne State University, Detroit, MI. During the summer of 1989, he was with the Universite Catholoque de Louvain, Louvain-la-Neuve, Belgium, as an invited lecturer. He joined the Department of Electrical and Computer Engineering, the University of Newcastle, Newcastle, Australia, in 1989. He was the Head of Department from 1998 to 2001. Currently, he is a Chair Professor of Electrical Engineering. He was also a Visiting Associate Professor at the University of Iowa, Iowa City, from 1995 to 1996 and a Visiting Professor at the Nanyang Technological University, Singapore, in 2002. His main research interests include control systems, signal processing, and applications to communications.

Dr. Fu has been an Associate Editor for the IEEE TRANSACTIONS ON AUTOMATIC CONTROL and the *Journal of Optimization and Engineering*.