Asymptotic Properties of Subband Identification

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

Abstract—The purpose of this paper is to study the asymptotic properties (i.e., strong convergence and asymptotic convergence rate) of the subband identification method in every subband and in the overall method. The study of strong convergence aims to answer the question whether the "best possible" model is retrieved, on the limit, with probability one. The study of the asymptotic convergence rate aims to give an expression that quantifies how fast the model approaches the "best possible" value as the number of samples goes to infinity. To do this, we need to generalize existing results for fullband identification. In the process of doing so, we come up with a new notion of ergodicity, which we call strong ergodicity. Strongly ergodic signals not only satisfy the assumptions required for our analysis but also enjoy an interesting property, which is that strong ergodicity is invariant under a number of transformations. In particular, the subband components of a strongly ergodic signal are guaranteed to be strongly ergodic, therefore, ergodic, which is not true for an ergodic signal in general.

Index Terms—Ergodicity, multirate signal processing, subband adaptive filtering, subband signal processing, system identification.

I. INTRODUCTION

Fig. 1. Fullband identification block diagram.

THE theory of linear system identification is well developed. Many references are available on the subject; see, e.g., [1] and [2]. 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 process noise, $g(q)$ [q is the forward shift operator (i.e., $qx(t) = x(t+1)$)] is the transfer function of the system, $\hat{g}(q, \theta^n)$ is the model of the system, and $\hat{v}(t, \theta^n)$ is the prediction error, where $\theta^n \in \mathbb{C}^n$ represents the parameters of the model. In most signal processing applications, $\hat{g}(q, \theta^n)$ is chosen to be a finite impulse response (FIR) model. Identification algorithms based on the least-squares technique are commonly employed in practice, and their behaviors are well understood. However, the direct use of these algorithms may not be most suitable for real-time applications where high-order FIR models are required (e.g., speech echo cancellation and channel equalization).

To alleviate the computational problem, the so-called subband identification technique has been proposed; see, e.g., [3] and [4]. Loosely speaking, the subband approach divides the input and output signals into a number of subbands using analysis filterbanks. Each analysis filterbank consists of a bank of M filters whose output is downsampled by a factor of D (i.e., one out of D samples is kept). Then, for each subband channel, a subband m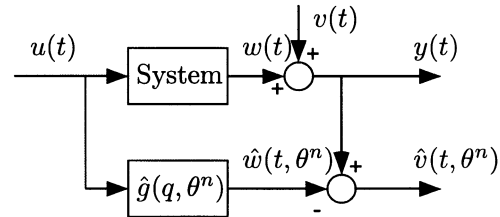odel is identified. These subband models can be combined to give a fullband model. In many applications, the subband models are used directly to estimate certain subband signals that are combined using a synthesis filterbank to form a required signal estimate.

An example of such applications is speech echo cancellation. The reverberation model for a typical video conferencing room requires a tap size in the order of 500–1000 or more. A training signal is often available for estimating the reveberation model. Estimating a fullband model may be very computationally involved, not to mention the numerical stability issues. In this case, subband models of reverberations can be estimated using a more numerically efficient subband identification algorithm. These models can then be used to give an estimate of the source signal (i.e., the speech signal without reverberation) in each subband. Finally, these subband signals are combined to give an estimate of the (fullband) source signal.

Another application where the subband identification technique can be used is broadband wireless channel equalization. Orthogonal frequency division multiplexing is a preferred modulation technique. This involves using a possibly large number of equally-spaced subcarriers to modulate transmit signals. The communication channel involves many (slowly time-varying) multipaths. One main difficulty with broadband wireless equalization is that the multipath channel model may require a large tap size, mainly due to the high data rate. Again, subband identification can be used to solve this problem. There is an extra advantage of the subband technique in this application because the subband signals (i.e., the subcarrier signals) are readily available at the receiver. This advantage yields a major computational saving [5].

The purpose of this paper is to study the asymptotic properties of the subband identification method in every subband and in the overall method. The asymptotic properties include strong convergence and asymptotic convergence rate. The study of strong convergence aims to answer the question whether the "best possible" model is retrieved, on the limit, with probability one. The study of the asymptotic convergence rate aims to give an expression that quantifies how fast the model approaches the "best possible" value as the number of samples goes to infinity.

These properties have been studied in detail in [1] and [6]. In [1], the study is done in terms of the parameters (i.e., coeffi-

cients) in the system model, whereas in [6], the transfer function of the system model is used. However, these approaches turn out to be inadequate for the subband case since the subband models are typically used directly for reconstructing a particular signal (as in the example of echo cancellation) rather than used for forming a fullband model. Motivated by this point, we choose to consider the power of the prediction error as a measure of the performance of the identification. That is, we establish conditions that give satisfactory asymptotic properties for this power.

The technical assumptions on the signals required in [6] is that the signals $u(t)$ and $w(t)$ are assumed to be deterministic signals that do not need to be related in any particular way, and $v(t)$ is a random process. In [1] instead, the signals $u(t)$, $w(t)$, and $v(t)$ are random processes, as required for our approach. However, the signal $u(t)$ needs to be generated by a combination of white noise and a deterministic signal, filtered by linear time-variant systems, and $w(t)$ needs to be generated by $u(t)$ through another linear filter. The stumbling block in generalizing the work of [1] to subband identification is that this required assumption is not preserved after subband decomposition. Hence, we cannot apply the results of [1] directly to subband identification. The approach we take in this paper is to determine the conditions on the signals $u(t)$, $w(t)$, and $v(t)$ such that their subband components will satisfy the conditions compatible with those used in [1]. In the process of doing so, we come up with a new notion of ergodicity that we call strong ergodicity. Roughly speaking, we require the fullband signals $u(t)$, $w(t)$, and $v(t)$ to be strongly ergodic. This guarantees that their subband components satisfy the assumption required to carry out the asymptotic analysis for each subband. An interesting property of strong ergodicity is that it is invariant under a number of transformations. In particular, the subband components of a strongly ergodic signal are guaranteed to be strongly ergodic, and therefore ergodic, which is not true for an ergodic signal in general. The bulk of this paper is devoted to the study of strong ergodicity because its analysis is quite involved mathematically.

The rest of the paper is organized as follows: In Section II, we introduce the notion of strong ergodicity and its key properties. In Section III, we study the asymptotic properties of the fullband method in terms of the power of the prediction error. In Section IV, we study the asymptotic properties of the subband method in each subband and in the overall method.

II. STRONGLY ERGODIC RANDOM PROCESSES

Given a collection of random processes, we can generate another collection by filtering, downsampling, upsampling, and additions of the random processes of the generating collection. We would like that if the generating collection is made of ergodic (in the second-order) random processes, then the generated random processes were also ergodic. This turns out to be not true, but it is true under a stronger condition on the generating random processes, which implies ergodicity. We call this condition strong ergodicity (in the second order).

However, first we need to review the notions of ergodicity and stationarity.

A. Ergodicity and Stationarity

Convention 1: All the random variables, random processes, and linear systems considered are assumed to be scalar and complex, unless explicitly specified. The superscript $*$ denotes complex conjugate. $\mathbb{Z}$ denotes the set of integers, and $\mathbb{N}$ denotes the set of natural numbers (i.e., integers greater or equal to one).

Definition 1: Let $\{x(t)\}$ and $\{y(t)\}$, $t \in \mathbb{Z}$ be two random processes. They are jointly ergodic if, for every $\tau \in \mathbb{Z}$

$$\lim_{T \to \infty} \frac{1}{T} \sum_{t=1}^{T} x^*(t)y(t+\tau) - \mathcal{E}\{x^*(t)y(t+\tau)\} = 0 \quad w.p.1.$$
(1)

Definition 2: Let $\{x(t)\}$ and $\{y(t)\}$, $t \in \mathbb{Z}$ be two random process. They are jointly quasistationary if we have the following.

1) They have uniformly bounded second moments (i.e., there exists $M_x > 0$ such that $\mathcal{E}\left\{|x(t)|^2\right\}^{1/2} < M_x, \forall t \in \mathbb{Z}$ and equivalently for $\{y(t)\}$).
2) For all $\tau \in \mathbb{Z}$, the following limit exists:

$$\lim_{T \to \infty} \frac{1}{T} \sum_{t=1}^{T} \mathcal{E}\{x^*(t)y(t+\tau)\}.$$
(2)

If they further satisfy that, for all $A, B \in \mathbb{N}$ and $a, b \in \mathbb{Z}$, the following limit exists:

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

then, they are 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 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 stationary.

We will extend definitions 1 and 2 to a single random process and to a finite collection of random processes as follows:

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

For a quasistationary random process, it make sense to define.

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

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

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

$$S_x = R_x(0) \tag{4}$$

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

$$\Phi_x(\omega) = \sum_{\tau=-\infty}^{\infty} R_x(\tau) e^{-j\omega\tau} \tag{5}$$

provided the infinite sum exists.

B. Strong Ergodicity

As stated previously, the ergodicity property can be lost after one of the transformations mentioned. To illustrate this point, we introduce two examples. The following example shows that ergodicity can be lost after downsampling.

Example 1: Consider the probability space $([-(1/2), (1/2)], \mathcal{B}, \lambda)$, where $\mathcal{B}$ denotes the Borel σ-algebra on the set $[-(1/2), 1/2]$, and λ denotes the Lebesgue measure. Define the random process $\{x(\omega, t)\}$, $\omega \in [-(1/2), (1/2)]$, $t \in \mathbb{Z}$ as follows: $x(\omega, t) = \begin{cases} 1/2\,(1 + \text{sign}(\omega)), & t \text{ is even} \\ 1/2\,(1 - \text{sign}(\omega)), & t \text{ is odd} \end{cases}$, where $\text{sign}(\omega) = |\omega|/\omega$. Let

$$C_x(\omega, \tau, T) = \frac{1}{T} \sum_{t=1}^{T} x(\omega, t)x(\omega, t+\tau) - \mathcal{E}\{x(\omega, t)x(\omega, t+\tau)\}. \tag{6}$$

Then

$$C_x(\omega, \tau, T)$$
$$= \begin{cases} \frac{1}{T} \sum_{t=1}^{T} (-1)^t \text{sign}(\omega), & \tau \text{ is even} \\ 0, & \tau \text{ is odd} \end{cases} \to 0 \text{ as } T \to \infty$$

so that $\{x(\omega, t)\}$ is ergodic. Let $y(\omega, t) = x(\omega, 2t) = (1/2)(1 + \text{sign}(\omega))$. Then

$$C_y(\omega, \tau, T) = \frac{1}{2}\text{sign}(\omega) \not\to 0 \text{ as } T \to \infty$$

so that $\{y(\omega, t)\}$ is not ergodic.

The following example shows that ergodicity can be lost after filtering with a time-invariant IIR linear filter.

Example 2: Consider the probability space $([-(1/2), (1/2)], \mathcal{B}, \lambda)$. Define the random process $\{x(\omega, t)\}$, $\omega \in [-(1/2), (1/2)]$, $t \in \mathbb{Z}$ by $x(\omega, t) = (1/\sqrt{2})e^{-t}(1 + \text{sign}(\omega))$. Define $C_x(\omega, \tau, T)$ as in (6); then

$$C_x(\omega, \tau, T) = e^{-\tau}\text{sign}(\omega)\frac{1}{T}\sum_{t=1}^{T} e^{-2t} \to 0, \text{ as } T \to \infty.$$

Let $h(k) \in l_1(\mathbb{Z})$ be given by $h(k) = \begin{cases} e^{-k}, & k > 0 \\ 0, & k \le 0 \end{cases}$, and define $y(\omega, t) = h(q)x(\omega, t)$. Then

$$C_y(\omega, \tau, T) = \sum_{k=1}^{\infty} \sum_{l=1}^{\infty} e^{-\tau}\text{sign}(\omega)\frac{1}{T}$$
$$\cdot \sum_{t=1}^{T} e^{-2t} \to \begin{cases} +\infty, & \omega \ge 0 \\ -\infty, & \omega < 0 \end{cases} \text{ as } T \to \infty$$

and therefore, $\{y(\omega, t)\}$ is not ergodic.

Now, we are ready to introduce the notion of strong ergodicity.

Definition 5: Let $p \in \{1, 2\}$, and let $\{x(t)\}$ and $\{y(t)\}$, $t \in \mathbb{Z}$ be two random processes. They are strongly ergodic of pth order if the following holds.

1) They have uniformly bounded $4p$th moments.
2) For any $A, B \in \mathbb{N}$, there exists $C > 0$ such that

$$\left\| C_{xy}^2(A, B, a, b, T) \right\|_p \le \frac{C}{T}, \quad \forall a, b \in \mathbb{Z} \tag{7}$$

where $\|\xi\|_p = \mathcal{E}\{\xi^p\}^{1/p}$, $\mathcal{E}\{\cdot\}$ denotes the expectation operator, and

$$C_{xy}(A, B, a, b, T)$$
$$= \frac{1}{T} \sum_{t=1}^{T} x^*(At+a)y(Bt+b) - \mathcal{E}\{x^*(At+a)y(Bt+b)\}. \tag{8}$$

This definition extends to a single random process and to a set in a way that is similar to Definition 3.

The following proposition states that strong ergodicity implies ergodicity.

Proposition 1: If the collection of random process $\{e_n(t)\}$, $t \in \mathbb{Z}$, $n = 1, \ldots, N$ is strongly ergodic of and pth order, then it is ergodic.

Proof: See Appendix A. ∎

Now, we want to explain what signals can form a strongly ergodic collection. In Proposition 2, we introduce a strongly ergodic collection of random processes, and in Theorem 1, we show how it can be transformed to generate other strongly ergodic collections. In Lemmas 1–4, we introduce four possible transformations on the collection that preserve the strong ergodicity property. They are intermediate steps in the Proof of Theorem 1. Proposition 2 and Theorem 1 together show that the usual signals of interest (signals generated from white noise and deterministic bounded signals) are strongly ergodic.

Proposition 2: Let $\{\mu_n(t)\}$, $t \in \mathbb{Z}$, $n = 1, \ldots, N$ be a collection of N random processes with uniformly bounded $4p$th moments, where $p \in \{1, 2\}$, such that the set of random variables $\{\mu_{n_1}(t_1), \ldots, \mu_{n_k}(t_k)\}$ is independent, for any distinct $(n_1, t_1), \ldots, (n_k, t_k) \in (\{1, \ldots, N\}, \mathbb{Z})$. Let $\{h_{n,t}(q)\}$, $t \in \mathbb{Z}$, $n = 1, \ldots, N$ be a uniformly bounded collection of sequences of linear time-variant systems (i.e., there exists $h(\tau) \in l_1(\mathbb{Z})$ such that $|h_{n,t}(\tau)| < h(\tau)$, for all $\tau \in \mathbb{Z}$, $n = 1, \ldots, N$). If the collection $\{e_n(t)\}$ is generated as follows:

$$e_n(t) = \sum_{\tau=-\infty}^{\infty} h_{n,t}(\tau)\mu_n(t-\tau)$$

then, $\{e_n(t)\}$ is strongly ergodic of pth order.

Proof: See Appendix A. ∎

Remark 1: Note that Proposition 2 includes the possibility that one or more of the signals are a deterministic signal in $l_\infty(\mathbb{Z})$.

Lemma 1: Let $\{e_n(t)\}$, $t \in \mathbb{Z}$, $n = 1, \ldots, N$ be strongly ergodic of pth order, and let $h(t)$, $t \in \mathbb{Z}$ be a time-invariant linear system. If, for some $k \in \{1, \ldots, N\}$

$$e_{N+1}(t) = \sum_{\tau=-\infty}^{\infty} h(\tau)e_k(t-\tau)$$

then $\{e_n(t)\}$, $n = 1, \ldots, N+1$ is strongly ergodic of pth order.

Proof: See Appendix A. ∎

Lemma 2: Let $\{e_n(t)\}$, $t \in \mathbb{Z}$, $n = 1, \ldots, N$ be strongly ergodic of pth order, and let $\alpha \in \mathbb{N}$. If, for some $k \in \{1, \ldots, N\}$

$$e_{N+1}(t) = e_k(\alpha t)$$

then $\{e_n(t)\}$, $n = 1, \ldots, N+1$ is strongly ergodic of pth order.

Proof: See Appendix A. ∎

Lemma 3: Let $\{e_n(t)\}$, $t \in \mathbb{Z}$, $n = 1, \ldots, N$ be strongly ergodic of pth order, and let $\alpha \in \mathbb{N}$. If, for some $k \in \{1, \ldots, N\}$

$$e_{N+1}(t) = \begin{cases} e_k\left(\frac{t}{\alpha}\right), & \frac{t}{\alpha} \in \mathbb{Z} \\ 0, & \frac{t}{\alpha} \notin \mathbb{Z} \end{cases}$$

then $\{e_n(t)\}$, $n = 1, \ldots, N+1$ is strongly ergodic of pth order.

Proof: See Appendix A. ∎

Lemma 4: Let $\{e_n(t)\}$, $t \in \mathbb{Z}$, $n = 1, \ldots, N$ be strongly ergodic of pth order. If, for some $k, l \in \{1, \ldots, N\}$

$$e_{N+1}(t) = e_k(t) + e_l(t)$$

then $\{e_n(t)\}$, $n = 1, \ldots, N+1$ is strongly ergodic of pth order.

Proof: See Appendix A. ∎

By combining Lemmas 1–4, we can prove the following theorem, which we state without proof.

Theorem 1: Let $\{x_m(t)\}$, $m = 1, \ldots, M$ be a pth-order strongly ergodic collection of random processes . If $\{y_n(t)\}$, $n = 1, \ldots, N$ is another collection generated from $\{x_m(t)\}$ by filtering with time-invariant linear filters, downsampling, upsampling and additions, then $\{y_m(t)\}$ is strongly ergodic of pth order.

III. FULLBAND IDENTIFICATION

In this section, we study the asymptotic properties of fullband identification. The system configuration is depicted in Fig. 1. As explained earlier, our approach is similar to [1] and [6], but we focus on the power of the prediction error rather than the parameters or the transfer function of the model. Two results are presented on this section: one on the probability of convergence of the power of the prediction error to its optimal value and one on the rate of the convergence. The results of this section will be used in the analysis of subband identification in the next section.

A. Strong Convergence

Theorem 2 states that the power of the prediction error converges, with probability one, to its optimal value. This result is a generalization of [1, Th. 8.2], stated in terms of the power of the prediction error instead of the set of parameters. That theorem was proved under the assumption that the signal $\{u(t)\}$ is generated by a combination of white noise and a deterministic signal, filtered by uniformly bounded linear systems, and

$\{w(t)\}$ is generated by $\{u(t)\}$ through another linear filter (see [1, Assumption D1]). Our result is derived under a different set of assumption on the signals, i.e., Assumption 1. Indeed, our requirement on the signals $\{u(t)\}$, $\{w(t)\}$, and $\{v(t)\}$ is weaker than [1, Assumption D1].

Two assumption are required for this result.

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 first order and quasistationary.
2) $\{v(t)\}$ is independent of $\{u(t)\}$ and $\{w(t)\}$.

Assumption 2: The model $\hat{g}(q, \theta^n)$ is a parametric linear model. The set of parameters is assumed to satisfy $\theta^n \in \mathcal{D} \subset \mathbb{C}^n$, where $\mathcal{D}$ is assumed to be compact (i.e., close and bounded). There exists $\hat{g}(t) \in l_1(\mathbb{Z})$ such that $|\hat{g}(t, \theta^n)| < \hat{g}(t)$, for all $\theta^n \in \mathcal{D}$ and all $t \in \mathbb{N}$, and there exists $\hat{g}'(t) \in l_1(\mathbb{Z})$, such that $|\hat{g}'_k(t, \theta^n)| < \hat{g}'(t)$, for all $\theta^n \in \mathcal{D}$, and for $k = 1, \ldots, n$, where $\hat{g}'_k(q, \theta^n)$ is the kth component of the vector $\hat{g}'(q, \theta^n) = (\partial/\partial\alpha^n)\hat{g}(q, \alpha^n)|_{\theta^n}$. The identification method is the prediction error method, i.e., the optimal vector of parameters up to time N (which is denoted by θ_N^n) is chosen as follows:

$$\theta_N^n \in \arg\min_{\theta^n \in \mathcal{D}} V_N(\theta^n) \tag{9}$$

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

$$V_N(\theta^n) = \frac{1}{N} \sum_{t=1}^{N} \frac{1}{2} |\hat{v}(t, \theta^n)|^2. \tag{10}$$

(Note that $\min_{\theta^n \in \mathcal{D}} V_N(\theta^n)$ is well defined since $V_N(\theta^n)$ is a continuous function, and $\mathcal{D}$ is compact).

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

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

Theorem 2: Consider the fullband identification method of Fig. 1, together with Assumptions 1 and 2. Then

$$\lim_{N \to \infty} S_{\tilde{w}}(\theta_N^n) = S_{\tilde{w},\text{opt}}^n \quad w.p.1. \tag{11}$$

Proof: See Appendix B. ∎

B. Asymptotic Convergence Rate

The following theorem gives a measure on how fast the power of the prediction error goes to its optimal value. Asymptotic convergence rate studies have been done in [1] and [6]. In [1], the convergence study is done in terms of the speed at which the vector of parameters θ_N^n goes to its optimal value when N goes to infinity. The assumption on the signals is the same as mentioned above, i.e., random processes related as in [1, Assumption D1]. The model is a quite generic one (uniformly stable linear model as defined in [1]). In [6], the convergence study is carried out by analyzing the speed at which the identified transfer function $\hat{g}(e^{j\omega}, \theta_N^n)$ goes to its expected value, instead of its optimal value, as N and n go to infinity. The signals $\{u(t)\}$ and $\{w(t)\}$ are assumed to be deterministic signals, and they do not need to be related in any particular way. The model is assumed to be FIR, and the way N and n go to infinity
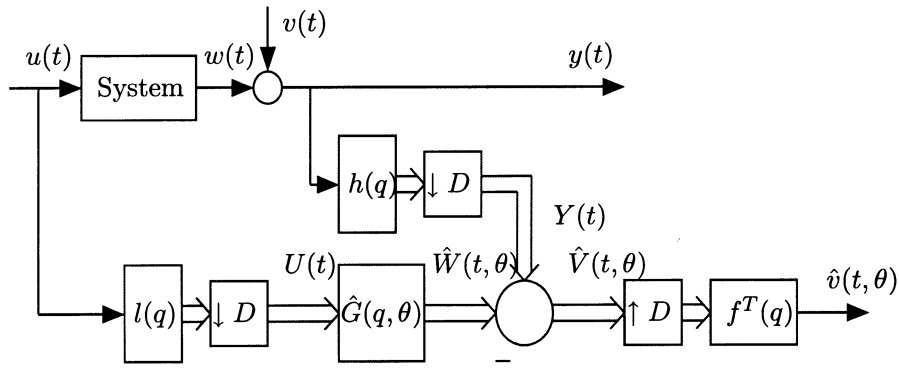
Fig. 2.   Subband identification scheme.

is quite general (see [6, Assumption D1]). Our result states the asymptotic convergence rate in terms of the speed at which the power $S_{\tilde{w}}(\theta_N^n)$ goes to its optimal value, as N and n go to infinity. This approach includes the convergence of the vector of parameters and the identified transfer function as intermediate steps. The analysis is a generalization in the sense that the signals are assumed to be random processes that do not need to be related in any particular way. On the other hand, our assumptions on the system (FIR) and the way N and n go to infinity ($\lim_{n\to\infty}\lim_{N\to\infty}$) are slightly more restrictive. However, the later can be generalized following the technique used in [6].

In addition to Assumptions 1 and 2, we also require the following.

Assumption 3: 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.
2) $\{v(t)\}$ is stationary and has zero mean (i.e., $\mathcal{E}\{v(t)\} = 0, \forall t \in \mathbb{N}$).
3) $\tau R_u(\tau), \tau R_v(\tau) \in l_1(\mathbb{Z})$.
4) There exists $\varepsilon > 0$ such that $\Phi_u(\omega) \geq \varepsilon, \ \forall \omega \in [-\pi, \pi]$.

Remark 2: Note that Assumption 3.3 is satisfied if the signals $\{u(t)\}$ and $\{v(t)\}$ are generated by white noise filtered by a uniformly bounded sequence of rational stable linear filters.

Assumption 4: The model $\hat{g}(q, \theta^n)$ is a parametric FIR model of tap size n. The set $\mathcal{D}$ satisfies

$$\arg \min_{\theta^n \in \mathbb{C}^n} S_{\tilde{w}}(\theta^n) \subset \text{int}(\mathcal{D}) \tag{12}$$

where $\text{int}(\mathcal{D})$ denotes the interior (i.e., excluding the boundary) of $\mathcal{D}$. The optimal parameters are solved using the least-squares (LS) algorithm, i.e., θ_N^n is calculated as follows:

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

where

$$R_N^n = \frac{1}{N} \sum_{t=1}^{N} \varphi_n^*(t) \varphi_n(t) \tag{13}$$

$$\varphi_n(t) = [u(t), u(t-1), \ldots, u(t-(n-1))]. \tag{14}$$

The superscript * denotes transpose conjugate.

Theorem 3: Consider the fullband identification method of Fig. 1, together with Assumptions 1–4. Then

$$\lim_{n\to\infty} \lim_{N\to\infty} \frac{N}{n} \mathcal{E}\{S_{\tilde{w}}^n(N)\} = S_v + \Sigma_{\tilde{w}} \tag{15}$$

where $S_{\tilde{w}}^n(N)$ denotes the power of the signal $\tilde{w}(t, \theta_N^n) = \hat{w}(t, \theta_N^n) - \hat{w}(t, \theta_{\text{opt}}^n)$, S_v is the power of the noise signal $v(t)$, and

$$\Sigma_{\tilde{w}} = \lim_{n\to\infty} \frac{1}{2\pi} \int_{-\pi}^{\pi} \Theta_n(\omega)\Phi_u(\omega)d\omega \tag{16}$$

$$\Theta_n(\omega) = \frac{1}{n}\Omega_n^*(e^{i\omega})[\bar{R}^n]^{-1} \lim_{N\to\infty} \bar{E}_N^n[\bar{R}^n]^{-1}\Omega_n(e^{i\omega}) \tag{17}$$

$$\bar{E}_N^n = \frac{1}{N}\sum_{t=1}^{N}\sum_{s=1}^{N}\mathcal{E}\{\varphi_n^*(t)\varphi_n(s)\tilde{w}(t,\theta_{\text{opt}}^n)\tilde{w}^*(s,\theta_{\text{opt}}^n)\} \tag{18}$$

$$\bar{R}^n = \lim_{N\to\infty} \frac{1}{N}\sum_{t=1}^{N}\mathcal{E}\{\varphi_n^*(t)\varphi_n(t)\} \tag{19}$$

$$\Omega_n^*(e^{i\omega}) = [1, e^{-i\omega}, \ldots, e^{-i\omega(n-1)}]. \tag{20}$$

Proof: See Appendix B.   ■

Remark 3: The residual term $\Sigma_{\tilde{w}}$ in (15) depends on $\hat{w}(t, \theta_{\text{opt}}^n)$. It has a complicated expression. However, this term can be neglected if $S_{\tilde{w},\text{opt}}^n$ is negligible compared with the noise power S_v. Define $S_{\tilde{w},\text{dif}}^n(N) = S_{\tilde{w}}(\theta_N^n) - S_{\tilde{w},\text{opt}}^n \leq S_{\tilde{w}}^n(N)$. Then, informally, (15) can be interpreted as follows: For large n and N and for small $S_{\tilde{w},\text{opt}}^n$

$$\mathcal{E}\{S_{\tilde{w},\text{dif}}^n(N)\} \lesssim \frac{n}{N}S_v. \tag{21}$$

Remark 4: Note that in (15), we consider the limit when N goes to infinity first, and then, we make n go to infinity. If we do not make n go to infinity, then the asymptotic convergence rate has a more complicated expression. It can be easily derived by following the steps of the proof without considering the limit on n. However, the result in (15) is not restrictive since the subband method is used mainly in applications where n is large, which is the case when the subband approach has advantages over the fullband method.

IV. SUBBAND IDENTIFICATION

The subband identification scheme is depicted in Fig. 2.

As we mentioned in introduction, the idea of subband identification is to split both signals $u(t)$ and $y(t)$ into M subbands using two analysis filterbanks $l(q) =$

$[l_1(q), \ldots, l_M(q)]^T$ and $h(q) = [h_1(q), \ldots, h_M(q)]^T$. 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 $\hat{W}(t, \theta) = [\hat{W}_1(t, \theta), \ldots, \hat{W}_M(t, \theta)]^T$, which is the subband equivalent of $\hat{w}(t, \theta)$. The prediction error $\hat{V}(t, \theta) = [\hat{V}_1(t, \theta), \ldots, \hat{V}_M(t, \theta)]^T$ is then formed. Finally, a synthesis filterbank $f(q) = [f_1(q), \ldots, f_M(q)]^T$ is used to reconstruct $\hat{v}(t, \theta)$.

For analysis purposes, we denote the downsampled version of $h(q)w(t)$ by $W(t) = [W_1(t), \ldots, W_M(t)]^T$. We define $\tilde{W}(t, \theta) = W(t) - \hat{W}(t, \theta) = [\tilde{W}_1(t, \theta), \ldots, \tilde{W}_M(t, \theta)]^T$ and denote by $\tilde{w}(t, \theta)$ the signal obtained by upsampling $\tilde{W}(t, \theta)$ and then filtering it with $f(q)$. We also denote the downsampled version of $h(q)v(t)$ by $V(t) = [V_1(t), \ldots, V_M(t)]^T$.

The simplest configuration of this method uses critical sampling (i.e., $D = M$) and a diagonal subband model $\hat{G}(q, \theta)$. It was shown [7], [8] that the filterbanks required for this configuration need to be composed by nonoverlapping filters. In practice, this means that the filters will have very narrow transition bands and, therefore, require a large tap size. There are two approaches to relax the design of the filters. The first uses off-diagonal terms in $\hat{G}(q, \theta)$, as studied in [3]. The second approach uses oversampling ($D < M$). In both approaches, the filters are allowed to be overlapping and, thus, require a low tap size. In this paper, we will study the first case (Assumption 7).

A. Strong Convergence

Theorem 4 studies the strong convergence in every subband, whereas Theorem 5 studies that of the overall prediction error. For Theorem 4, three assumptions are required.

Assumption 5: 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 first order and quasistationary by phases.
2) $\{v(t)\}$ is independent of $\{u(t)\}$ and $\{w(t)\}$.

Remark 5: Note that Assumption 5 is satisfied if the collection formed by $\{u(t)\}$ and $\{v(t)\}$ is strongly ergodic of first order and quasistationary by phases, and $\{w(t)\}$ is generated from $\{u(t)\}$ through a linear system $g(q)$ whose impulse response satisfies $g(t) \in l_1(\mathbb{Z})$.

Assumption 6: The analysis filterbanks $l(q)$ and $h(q)$ are such that $l_m(t)$ and $h_m(t) \in l_1(\mathbb{Z})$ for all $m = 1, \ldots, M$.

Assumption 7: The subband model is a diagonal matrix $\hat{G}(q, \theta^{Mn}) = \text{diag}\{\hat{G}_m(q, \theta_m^n), \ m = 1, \ldots, M\}$, where $\theta^{Mn} = [(\theta_1^n)^T, \ldots, (\theta_M^n)^T]^T$ and all the set of parameters $\theta_m^n, \ m = 1, \ldots M$ have the same order n. Each $\hat{G}_m(q, \theta_m^n)$ satisfies Assumption 2, i.e., for each $m \in \{1, \ldots, M\}$, the model $\hat{G}_m(q, \theta_m^n)$ is a parametric linear model. The set of parameters is assumed to satisfy $\theta_m^n \in \mathcal{D} \subset \mathbb{C}^n$, where $\mathcal{D}$ is assumed to be compact. There exists $\hat{G}(t) \in l_1(\mathbb{Z})$ such that $|\hat{G}_m(t, \theta_m^n)| < \hat{G}(t)$ for all $\theta_m^n \in \mathcal{D}$ and all $t \in \mathbb{N}$, and there exists $\hat{G}'(t) \in l_1(\mathbb{Z})$ such that $|\hat{G}'_{m,k}(t, \theta_m^n)| < \hat{G}'(t)$ for all $\theta_m^n \in \mathcal{D}$, and for $k = 1, \ldots, n$, where $\hat{G}'_{m,k}(q, \theta^n)$ is the kth

component of the vector $\hat{G}'(q, \theta_m^n) = (\partial/\partial\alpha^n)\hat{G}(q, \alpha^n)\big|_{\theta_m^n}$. The identification method is the prediction error method, i.e., the optimal vector of parameters up to time N (which is denoted by $\theta_{m,N}^n$) is chosen as follows:

$$\theta_{m,N}^n \in \arg\min_{\theta_m^n \in \mathcal{D}} V_{m,N}(\theta_m^n)$$

where

$$V_{m,N}(\theta_m^n) = \frac{1}{N}\sum_{t=1}^N \frac{1}{2}\left|\hat{V}_m(t, \theta_m^n)\right|^2.$$

Notation 2: For each subband, we define the power of the prediction error signal $\tilde{W}_m(t, \theta_m^n)$ by $S_{\tilde{W}_m}(\theta_m^n)$. Then, we denote

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

Theorem 4: Consider the subband identification scheme of Fig. 2, together with Assumptions 5–7. Then, for $m = 1, \ldots, M$

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

Proof: See Appendix C. ∎

In every subband, the sequence of random variables $S_{\tilde{W}_m}(\theta_{m,N}^n)$, converges, with probability one to the deterministic constant $S_{\tilde{W}_m,\text{opt}}^n$, which is the global minimum of $S_{\tilde{W}_m}(\theta_m^n)$. The overall error $S_{\tilde{w}}(\theta_N^{Mn})$ also converges, with probability one, to a deterministic constant $S_{\tilde{w},\text{lim}}^n$. The following theorem states this fact formally. One extra assumption is required.

Assumption 8: In every subband, the set $\arg\min_{\theta_m^n \in \mathcal{D}} S_{\tilde{W}_m}(\theta_m^n)$ has only one element.

Remark 6: Note that Assumption 8 is satisfied if Assumptions 9.4, 9.5, 10, and 11 below, are satisfied.

Theorem 5: Consider the subband identification scheme of Fig. 2, together with Assumptions 5–8. Then, there exists a deterministic constant $S_{\tilde{w},\text{lim}} \geq 0$ such that

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

Proof: See Appendix C. ∎

The constant $S_{\tilde{w},\text{lim}}^n$ is not equal to $S_{\tilde{w},\text{opt}}^n$ in general. However, a modification in the identification method, which was introduced in [9], guarantees that $S_{\tilde{w},\text{lim}}^n = S_{\tilde{w},\text{opt}}^n$. In addition, with this modification, Assumption 8 is not needed for strong convergence.

B. Asymptotic Convergence Rate

As with the study of strong convergence, we have a theorem for the asymptotic convergence rate in every subband (Theorem 6) and a theorem for the overall asymptotic convergence rate (Theorem 7). For Theorem 6, in addition to Assumptions 5–7, we also require the following.

Assumption 9: 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.
2) $\{v(t)\}$ is stationary and has zero mean.

3) $\tau R_u(\tau), \tau R_v(\tau) \in l_1(\mathbb{Z})$.
4) There exists $\varepsilon > 0$ such that $\Phi_u(\omega) \geq \varepsilon$, $\forall \omega \in [-\pi, \pi]$.
5) The signal $\{u(t)\}$ is almost stationary.

Remark 7: Note that Assumption 9.1 is satisfied if the collection formed by $\{u(t)\}$, and $\{v(t)\}$ is strongly ergodic of second order and quasistationary by phases, and $\{w(t)\}$ is generated from $\{u(t)\}$ through a linear system $g(q)$ whose impulse response satisfies $g(t) \in l_1(\mathbb{Z})$.

Assumption 10: The analysis filterbanks $l(q)$ and $h(q)$ are such that $tl_m(t)$ and $th_m(t) \in l_1(\mathbb{Z})$, for $m = 1, \ldots, M$. There exists $\alpha > 0$ such that

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

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

Assumption 11: Each $\hat{G}_m(q, \theta_m^n)$ satisfies Assumption 4, i.e., for each $m \in \{1, \ldots, M\}$, the model $\hat{G}_m(q, \theta_m^n)$ is a parametric FIR model of tap size n. The set $\mathcal{D}$ satisfies

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

The optimal parameters are solved using the least-squares (LS) algorithm, i.e., $\theta_{m,N}^n$ is calculated as follows:

$$\theta_{m,N}^n = [R_{m,N}^n]^{-1} \frac{1}{N} \sum_{t=1}^{N} \varphi_{m,n}^*(t) Y_m(t)$$

where

$$R_{m,N}^n = \frac{1}{N} \sum_{t=1}^{N} \varphi_{m,n}^*(t) \varphi_{m,n}(t)$$
$$\varphi_{m,n}(t) = [U_m(t), U_m(t-1), \ldots, U_m(t-(n-1))].$$

Theorem 6: Consider the subband identification scheme of Fig. 2, together with Assumptions 5–11. Then, for each $m = 1, \ldots, M$, for large n and N, and for small $S_{\tilde{W}_m,\text{opt}}^n$

$$\mathcal{E}\{S_{\tilde{W}_m,\text{dif}}^n(N)\} \simeq \frac{n}{N} S_{V_m} \tag{22}$$

where $S_{\tilde{W}_m,\text{dif}}^n(N) = S_{\tilde{W}_m}(\theta_N^n) - S_{\tilde{W}_m,\text{opt}}^n$, and S_{V_m} is the power of the noise signal $V_m(t)$.

Proof: See Appendix C. ∎

Notation 3: Consider the subband identification scheme of Fig. 2. Let $u(t), w(t), v(t) \in l_2(\mathbb{Z})$. We define the following linear maps:

$$T_h : l_2(\mathbb{Z}) \to l_2^M(\mathbb{Z}) : y(t) \mapsto Y(t)$$
$$\text{and } T_f^* : l_2^M(\mathbb{Z}) \to l_2(\mathbb{Z}) : \hat{V}(t, \theta) \mapsto \hat{v}(t, \theta)$$

where T_f^* denotes the adjoint of some operator T_f. We denote their induced norms by

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

and $\left\| T_f^* \right\|$ in a similar way.

The following result gives a bound for the norms of T_h and T_f^*.

Proposition 3: Consider the subband identification scheme in Fig. 2. Then

$$\|T_h\| \leq \frac{1}{D}$$
$$\cdot \left[\sup_{-\pi \leq \omega \leq \pi} \left\{ \sum_{m=1}^{M} \left| h_m(e^{j\omega}) \right|^2 \right\} + \sum_{d=1}^{D-1} (\beta(-d)\beta(d))^{1/2} \right]$$

with equality if $\beta(d) = 0$, $d = 1, \ldots, D-1$, where $\beta(d) = \sup_{-\pi \leq \omega \leq \pi} \left\{ \sum_{m=1}^{M} \left| h_m(e^{j\omega}) \right| \left| h_m(\Omega^d e^{j\omega}) \right| \right\}$, and $\Omega = e^{-j(2\pi/D)}$. The norm of T_f^* is bounded in the same way.

Proof: See Appendix C. ∎

Theorem 7: Consider the subband identification scheme of Fig. 2, together with Assumptions 5–11. Then, for large n and N, and for small $S_{\tilde{w},\text{lim}}$

$$\mathcal{E}\{S_{\tilde{w},\text{dif}}^n(N)\} \lesssim \frac{n}{N} \|T_f\|^2 \|T_h\|^2 S_v$$

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

Proof: See Appendix C. ∎

V. CONCLUSIONS

We have studied the asymptotic properties of the subband identification method. The study was done in terms of the power of the prediction error and was carried in every subband and in the overall method. In the process of doing this, we came up with a new notion of ergodicity called strong ergodicity. The key properties of strongly ergodic signals is that they are invariant under a number of transformations, i.e., filtering, downsampling, upsampling, and addition. Using the notion of strong ergodicity, the well-known asymptotic properties of the fullband identification method are generalized to subband identification.

In a companion paper [10], we will give a rigorous study of the performance of the subband identification method to explain where the advantages of the method come from and how to optimize the performance of the method.

APPENDIX A
PROOFS OF SECTION II

Lemma 5: Let $\{x(t)\}, t \in \mathbb{Z}$ be a sequence of random variables with uniformly bounded pth moments, and let $M_x > 0$ be the bound of the pth moments. Let $h_t(\tau)$ be a uniformly bounded linear system bounded by $h(\tau)$. If $\{y(t)\}$ is defined by $y(t) = \sum_{\tau=-\infty}^{\infty} h_t(\tau) x(t-\tau)$, then the pth moments of $\{y(t)\}$ are uniformly bounded by $\|h(\tau)\|_1 M_x$.

Proof: We can interpret the pth moment of $y(t)$ as a norm operation, i.e., $\|y(t)\|_p = \mathcal{E}\{|y(t)|^p\}^{1/p}$. Then

$$\mathcal{E}\{|y(t)|^p\}^{1/p} = \|y(t)\|_p \leq \sum_{\tau=-\infty}^{\infty} |h_t(\tau)| \|x(t-\tau)\|_p$$
$$\leq \sum_{\tau=-\infty}^{\infty} h(\tau) M_x = \|h(\tau)\|_1 M_x.$$

∎

Notation 4: Let $\{x(t)\}$ and $\{y(t)\}$, $t \in \mathbb{Z}$ be two random processes $A, B \in \mathbb{N}$ and $a, b \in \mathbb{Z}$. The random process $\{\xi_{xy}(A, B, a, b, t)\}$ is defined by

$$\xi_{xy}(A, B, a, b, t) = x^*(At + a)y(Bt + b) - \mathcal{E}\{x^*(At + a)y(Bt + b)\}. \quad (23)$$

The following proof is based on the technique used in the proof of the Strong Law of Large Numbers [11, Th. 5.1.2].

Proof of Proposition 1: Let $x(t), y(t) \in \{e_n(t)\}$. Let $(\Omega, \mathcal{A}, \mathbb{P})$ be the probability space on which the stochastic process $\{x(t)\}$ is defined. We split the proof into three steps.

1) From (7) and Lyapunov's inequality, we have that

$$\sum_{L=1}^{\infty} \mathcal{E}\{C_{xy}^2(A, B, a, b, L^2)\} \leq \sum_{L=1}^{\infty} \mathcal{E}\{C_{xy}^{2p}(A, B, a, b, L^2)\}^{1/p}$$
$$\leq \sum_{L=1}^{\infty} \frac{C}{L^2} \leq \frac{\pi^2}{6}C, \ \forall a, b \in \mathbb{Z}. \quad (24)$$

In addition, from Chevychev's inequality [11, pp. 48]

$$\mathbb{P}\{|C_{xy}(A, B, a, b, L^2)| > \varepsilon\} \leq \frac{1}{\varepsilon^2}\mathcal{E}\{C_{xy}^2(A, B, a, b, L^2)\}$$
$$\forall a, b \in \mathbb{Z}, \forall \varepsilon > 0. \quad (25)$$

Then, from (24) and (25)

$$\sum_{L=1}^{\infty} \mathbb{P}\{|C_{xy}(A, B, a, b, L^2)| > \varepsilon\}$$
$$\leq \sum_{L=1}^{\infty} \frac{1}{\varepsilon^2}\mathcal{E}\{C_{xy}^2(A, B, a, b, L^2)\}$$
$$\leq \frac{\pi^2}{6}\frac{C}{\varepsilon^2}, \forall a, b \in \mathbb{Z}, \forall \varepsilon > 0. \quad (26)$$

By the Borel–Cantelli lemma [11, Th. 4.2.1], (26) implies that

$$\mathbb{P}\left\{\limsup_{L \to \infty}\{|C_{xy}(A, B, a, b, L^2)| > \varepsilon\}\right\} = 0,$$
$$\forall a, b \in \mathbb{Z}, \forall \varepsilon > 0$$

which, in turn, by [11, Th. 4.2.2], implies that

$$\lim_{L \to \infty} C_{xy}(A, B, a, b, L^2) = 0 \ \ w.p.1, \ \ \forall a, b \in \mathbb{Z}. \quad (27)$$

2) Using the notation (23), let

$$D_{xy}(A, B, a, b, L) = \max_{L^2 \leq \tau < (L+1)^2} \left|\frac{1}{\tau}\sum_{t=L^2}^{\tau} \xi_{xy}(A, B, a, b, t)\right|.$$

Then

$$\mathcal{E}\{D_{xy}^2(A, B, a, b, L)\}$$
$$\leq \mathcal{E}\left\{\left(\frac{1}{L^2}\sum_{t=L^2}^{(L+1)^2-1} |\xi_{xy}(A, B, a, b, t)|\right)^2\right\}.$$

We can interpret the operation $\mathcal{E}\{\cdot^2\}^{1/2}$ as a norm operation; then

$$\mathcal{E}\left\{\left(\sum_{t=L^2}^{(L+1)^2-1} |\xi_{xy}(A, B, a, b, t)|\right)^2\right\}^{1/2}$$
$$\leq \sum_{t=L^2}^{(L+1)^2-1} \mathcal{E}\{|\xi_{xy}(A, B, a, b, t)|^2\}^{1/2} \leq 2L\,M_x^{1/2}$$

where M_x is the bound of the fourth moments of $\{x(t)\}$; then, $\mathcal{E}\{D_{xy}^2(A, B, a, b, L)\} \leq 4(M_x/L^2)$, and therefore, there exists $D > 0$, such that

$$\sum_{L=1}^{\infty} \mathcal{E}\{D_{xy}^2(A, B, a, b, L)\} \leq D, \quad \forall a, b \in \mathbb{Z}.$$

Now, following the same reasoning as in step 1, we have that

$$\lim_{L \to \infty} D_{xy}(A, B, a, b, L) = 0 \ \ w.p.1, \quad \forall a, b \in \mathbb{Z}. \quad (28)$$

3) For every $T \in \mathbb{N}$, we have that $C_{xy}(A, B, a, b, T) \leq C_{xy}(A, B, a, b, L^2) + D_{xy}(A, B, a, b, L)$, where L is the largest integer such that $L^2 \leq T$. Then, from (27) and (28)

$$\lim_{T \to \infty} C_{xy}(A, B, a, b, T) = 0 \ \ w.p.1, \quad \forall a, b \in \mathbb{Z}$$

which implies (1). ∎

Proof of Proposition 2: We consider the case where $p = 1$. To prove the case where $p = 2$ is similar.

To simplify the notation, let $n_1, n_2 \in \{1, \ldots, N\}$, and let $x(t) = e_{n_1}(t)$, $y(t) = e_{n_2}(t)$, $u(t) = \mu_{n_1}(t)$, $v(t) = \mu_{n_2}(t)$, $f_t(q) = h_{n_1, t}(q)$, and $g_t(q) = h_{n_2, t}(q)$. Using the notation (23), we can write

$$\mathcal{E}\{C_{xy}^2(A, B, a, b, T)\}$$
$$= \frac{1}{T^2}\sum_{t=1}^{T}\sum_{s=1}^{T} \mathcal{E}\{\xi_{xy}(A, B, a, b, t)\xi_{xy}(A, B, a, b, s)\}$$

where $\xi_{xy}(A, B, a, b, t) = \sum_{k=-\infty}^{\infty}\sum_{l=-\infty}^{\infty} f_t^*(k)g_t(l)\xi_{uv}(A, B, a-k, b-l, t)$. Then, we have the first equation shown at the bottom of the next page, where the exchange of the expectation with the infinite sums, in the second inequality, is valid in view of Fubini's Theorem since $f_t(t)$ and $g_t(t)$ are uniformly bounded, and the random processes have uniformly bounded fourth moments. It can be easily verified that

$$\mathcal{E}\{\xi_{uv}(A, B, a, b, t)\xi_{uv}(A, B, a, b, s)\} \begin{cases} \leq M_x, & \text{if} \quad t = s \\ = 0, & \text{if} \quad t \neq s \end{cases}$$

where M_x is the bound of the fourth moments of $\{x(t)\}$. Then

$$\mathcal{E}\{C_{xy}^2(A, B, a, b, T)\}$$
$$\leq \sum_{k=-\infty}^{\infty}\sum_{l=-\infty}^{\infty}\sum_{m=-\infty}^{\infty}\sum_{n=-\infty}^{\infty} h(k)h(l)h(m)h(n)\frac{1}{T^2}\sum_{t=1}^{T} M_x$$
$$= \|h(t)\|_1^4 \frac{M_x}{T}.$$

∎

Proof of Lemma 1: We consider the case where $p = 1$ to prove the case where $p = 2$ is similar.

Let $x(t) = e_k(t)$, $z(t) = e_{N+1}(t)$ and $y(t) \in \{e_n(t)\}$. Using the notation (23), we have that

$$\mathcal{E}\left\{C_{zy}^2(A, B, a, b, T)\right\}$$
$$= \frac{1}{T^2} \sum_{t=1}^{T} \sum_{s=1}^{T} \mathcal{E}\left\{\xi_{zy}(A, B, a, b, t)\xi_{zy}(A, B, a, b, s)\right\}$$

where $\xi_{zy}(A, B, a, b, t) = \sum_{k=-\infty}^{\infty} h^*(k)\xi_{xy}(A, B, a - k, b, t)$. Then, we have the second equation shown at the bottom of the page, where the exchange of the expectation, in the second equality, is valid in view of Fubini's Theorem since $h(t) \in l_1(\mathbb{Z})$, and the random processes have uniformly bounded fourth moments. Then

$$\mathcal{E}\left\{C_{zy}^2(A, B, a, b, T)\right\} \leq \|h(t)\|_1^2 \frac{C}{T}.$$

To prove that $\{y(t)\}$ has uniformly bounded fourth moments, we apply Lemma 5. ∎

Proof of Lemma 2: Let $x(t) = e_k(t)$, $z(t) = e_{N+1}(t)$, and $y(t) \in \{e_n(t)\}$. We have that

$$C_{zy}(A, B, a, b, T) = C_{xy}(\alpha A, B, \alpha a, b, T)$$

and the rest of the proof is straightforward. ∎

Proof of Lemma 3: This proof is similar to the proof of Proposition 2. ∎

Proof of Lemma 4: Let $x(t) = e_k(t)$, $w(t) = e_l(t)$, $z(t) = e_{N+1}(t)$, and $y(t) \in \{e_n(t)\}$. We have that

$$C_{zy}(A, B, a, b, T) = C_{xy}(A, B, a, b, T) + C_{wy}(A, B, a, b, T)$$

and therefore

$$\begin{aligned}
C_{zy}^2(A, B, a, b, T) =& C_{xy}^2(A, B, a, b, T) + C_{wy}^2(A, B, a, b, T)] \\
&+ 2C_{xy}(A, B, a, b, T)C_{wy}(A, B, a, b, T) \\
\leq& C_{xy}^2(A, B, a, b, T) + C_{wy}^2(A, B, a, b, T) \\
&+ 2\left|C_{xy}(A, B, a, b, T)C_{wy}(A, B, a, b, T)\right| \\
\leq& 2\left(C_{xy}^2(A, B, a, b, T) + C_{wy}^2(A, B, a, b, T)\right)
\end{aligned}$$

and the rest of the proof is straightforward. ∎

APPENDIX B
PROOFS OF SECTION III

Proof of Theorem 2: We split the proof into four steps.

1) Let

$$\bar{V}_N(\theta^n) = \mathcal{E}\left\{V_N(\theta^n)\right\}. \tag{29}$$

From Assumption 1 and Theorem 1, we have that $\{\hat{v}(t, \theta^n)\}$ is ergodic for all $\theta^n \in \mathcal{D}$; then

$$\lim_{N \to \infty} V_N(\theta^n) - \bar{V}_N(\theta^n) = 0 \quad w.p.1. \tag{30}$$

Since $\{u(t)\}$, $\{w(t)\}$, and $\{v(t)\}$ form a quasistationary collection, and $\hat{g}(q, \theta^n)$ is an FIR filter, $\hat{v}(t, \theta^n) = w(t) -$

$$\begin{aligned}
\mathcal{E}\left\{C_{zy}^2(A, B, a, b, T)\right\} \leq& \frac{1}{T^2} \sum_{t=1}^{T} \sum_{s=1}^{T} \left|\mathcal{E}\left\{\xi_{xy}(A, B, a, b, t)\xi_{xy}(A, B, a, b, s)\right\}\right| \\
\leq& \sum_{k=-\infty}^{\infty} \sum_{l=-\infty}^{\infty} \sum_{m=-\infty}^{\infty} \sum_{n=-\infty}^{\infty} h(k)h(l)h(m)h(n) \cdot \\
& \cdot \frac{1}{T^2} \sum_{t=1}^{T} \sum_{s=1}^{T} \left|\mathcal{E}\left\{\xi_{uv}(A, B, a - k, b - l, t)\xi_{uv}(A, B, a - m, b - n, s)\right\}\right|
\end{aligned}$$

$$\begin{aligned}
\mathcal{E}\left\{C_{zy}^2(A, B, a, b, T)\right\} =& \frac{1}{T^2} \sum_{t=1}^{T} \sum_{s=1}^{T} \mathcal{E}\left\{\sum_{k=-\infty}^{\infty} \sum_{l=-\infty}^{\infty} h^*(k)h^*(l)\xi_{xy}(A, B, a - k, b, t)\xi_{xy}(A, B, a - l, b, s)\right\} \\
=& \sum_{k=-\infty}^{\infty} \sum_{l=-\infty}^{\infty} h^*(k)h^*(l)\mathcal{E}\left\{C_{xy}(A, B, a - k, b, T)C_{xy}(A, B, a - l, b, T)\right\} \\
\leq& \sum_{k=-\infty}^{\infty} \sum_{l=-\infty}^{\infty} |h(k)||h(l)|\mathcal{E}\left\{|C_{xy}(A, B, a - k, b, T)C_{xy}(A, B, a - l, b, T)|\right\} \\
\leq& \frac{1}{2} \sum_{k=-\infty}^{\infty} \sum_{l=-\infty}^{\infty} |h(k)||h(l)|\left(\mathcal{E}\left\{C_{xy}^2(A, B, a - k, b, T)\right\} + \mathcal{E}\left\{C_{xy}^2(A, B, a - l, b, T)\right\}\right)
\end{aligned}$$

$\hat{w}(t, \theta) + v(t)$ is quasistationary for all $\theta^n \in \mathcal{D}$. It follows that the following limit exists:

$$
\begin{aligned}
\lim_{N \to \infty} & \bar{V}_N(\theta^n) \\
&= \lim_{N \to \infty} \mathcal{E}\{V_N(\theta^n)\} \\
&= \frac{1}{2} \lim_{N \to \infty} \frac{1}{N} \sum_{t=1}^{N} \mathcal{E}\left\{|\hat{v}(t, \theta^n)|^2\right\} \triangleq \bar{V}(\theta^n).
\end{aligned} \tag{31}
$$

From (30) and (31)

$$
\lim_{N \to \infty} V_N(\theta^n) = \bar{V}(\theta^n) \quad w.p.1. \tag{32}
$$

2) Let $\theta^n, \Delta\theta^n \in \mathcal{D}$; then

$$
\begin{aligned}
& \left|\bar{V}_N(\theta^n + \Delta\theta^n) - \bar{V}_N(\theta^n)\right| \\
& \leq \frac{1}{N} \sum_{t=1}^{N} \mathcal{E}\left\{\left|\left|\frac{1}{2}\hat{v}^2(t, \theta^n + \Delta\theta^n)\right| - \left|\frac{1}{2}\hat{v}^2(t, \theta^n)\right|\right|\right\} \\
& \leq \frac{1}{N} \sum_{t=1}^{N} \mathcal{E}\left\{\left|\frac{1}{2}\hat{v}^2(t, \theta^n + \Delta\theta^n) - \frac{1}{2}\hat{v}^2(t, \theta^n)\right|\right\} \\
& \leq \frac{1}{N} \sum_{t=1}^{N} M \|\Delta\theta^n\|_2 = M \|\Delta\theta^n\|_2
\end{aligned} \tag{33}
$$

where

$$
M = \mathcal{E}\left\{\sup_{\theta^n \in \mathcal{D}} \left\|\left.\frac{\partial}{\partial\alpha^n} \frac{1}{2}\hat{v}^2(t, \alpha^n)\right|_{\theta^n}\right\|_2\right\}.
$$

Consider the kth component of the gradient of $(1/2)\hat{v}^2(t, \theta^n)$

$$
\begin{aligned}
& \left|\left[\left.\frac{\partial}{\partial\alpha^n} \frac{1}{2}\hat{v}^2(t, \alpha^n)\right|_{\theta^n}\right]_k\right| \\
&= |\hat{v}(t, \theta^n)\hat{g}'_k(q, \theta)u(t)| \\
&= |v(t) + w(t) - \hat{g}(q, \theta)u(t)| \, |\hat{g}'_k(q, \theta)u(t)| \\
&\leq (|v(t)| + |w(t)| + \hat{g}(q)|u(t)|) \, \hat{g}'(q)|u(t)|.
\end{aligned}
$$

Therefore

$$
\begin{aligned}
& \left\|\left.\frac{\partial}{\partial\alpha^n} \frac{1}{2}\hat{v}^2(t, \alpha^n)\right|_{\theta^n}\right\|_2 \\
& \leq \sqrt{n}\,(|v(t)| + |w(t)| + \hat{g}(q)|u(t)|)\,\hat{g}'(q)|u(t)|.
\end{aligned}
$$

Now

$$
\begin{aligned}
M \leq & \sqrt{n}\left(\mathcal{E}\{v^2(t)\}^{1/2} + \mathcal{E}\{w^2(t)\}^{1/2}\right. \\
& \left. + \mathcal{E}\left\{(\hat{g}(q)u(t))^2\right\}^{1/2}\right)\mathcal{E}\left\{(\hat{g}'(q)u(t))^2\right\}^{1/2} \\
\leq & \sqrt{n}\,(M_v + M_w + \|\hat{g}(t)\|_1 M_u)\,\|\hat{g}'(t)\|_1 M_u < \infty
\end{aligned}
$$

where, for the last inequality, we applied Lemma 5. Therefore, for any $\theta^n \in \mathcal{D}$ and any $\varepsilon > 0$, if $\|\Delta\theta^n\|_2 < \delta = \sqrt{\varepsilon/L}$, then from (33), for all $N \in \mathbb{N}$

$$
\left|\bar{V}_N(\theta^n + \Delta\theta^n) - \bar{V}_N(\theta^n)\right| \leq \varepsilon
$$

and taking the limit as N goes to infinity

$$
\left|\bar{V}(\theta^n + \Delta\theta^n) - \bar{V}(\theta^n)\right| \leq \varepsilon
$$

which shows that $\bar{V}(\theta^n)$ is continuous for all $\theta^n \in \mathcal{D}$. Following the same steps, it is straightforward to show that $V_N(\theta^n)$ is also continuous. Then, since $\mathcal{D}$ is compact, the convergence in (32) is uniform on $\theta^n \in \mathcal{D}$.

3) Let $\mathcal{D}_C = \arg\min_{\theta^n \in \mathcal{D}} \bar{V}(\theta^n)$, which is well defined and compact since $\bar{V}(\theta^n)$ is continuous and $\mathcal{D}$ is compact. Since the limit in (32) is uniform on $\theta^n \in \mathcal{D}$, then

$$
\lim_{N \to \infty} \min_{\theta^n \in \mathcal{D}_C} \|\theta_N^n - \theta^n\|_2 = 0 \quad w.p.1. \tag{34}
$$

Note that the operation $\min$ is well defined since $\mathcal{D}_C$ is compact. To show (34), let $(\Omega, \mathcal{A}, \mathbb{P})$ be the underlying probability space of $\{u(t)\}, \{w(t)\}$, and $\{v(t)\}$. Let $\Omega_0 \subseteq \Omega$ be the set where (34) holds. Consider a particular $\omega_0 \in \Omega_0$. Suppose that there exists a sequence θ_N^n such that

$$
\lim_{N \to \infty} \min_{\theta^n \in \mathcal{D}_C} \|\theta_N^n - \theta^n\|_2 = L. \tag{35}
$$

Define the set

$$
\mathcal{D}_F = \left\{\alpha^n \in \mathcal{D} : \min_{\theta^n \in \mathcal{D}_C} \|\alpha^n - \theta^n\|_2 \geq \frac{L}{2}\right\} \tag{36}
$$

which is obviously compact. From (35) and (36), there exists $N_1 \in \mathbb{N}$ such that

$$
\theta_N^n \in \mathcal{D}_F, \quad \forall N \geq N_1. \tag{37}
$$

Define $\bar{V}_F = \min_{\theta^n \in \mathcal{D}_F} \bar{V}(\theta^n)$, which is well defined since $\mathcal{D}_F$ is compact, and define $\delta = \bar{V}_F - \min_{\theta^n \in \mathcal{D}} \bar{V}(\theta^n) > 0$. Since the convergence in (32) is uniform on $\theta^n \in \mathcal{D}$, then, there exists $N_2 \in \mathbb{N}$, such that

$$
\max_{\theta^n \in \mathcal{D}} \left|V_N(\theta^n) - \bar{V}(\theta^n)\right| \leq \frac{\delta}{3}, \quad \forall N \geq N_2. \tag{38}
$$

From (9), (37), and (38), for all $\theta^n \in \mathcal{D}_C$ and $N \geq \max\{N_1, N_2\}$

$$
V_N(\theta^n) \geq \bar{V}_C + \frac{2}{3}\delta = \bar{V}(\theta^n) + \frac{2}{3}\delta
$$

which is impossible in view of (38). Finally, from (34), and since $\bar{V}(\theta^n)$ is continuous, it follows that

$$
\lim_{N \to \infty} \bar{V}(\theta_N^n) = \min_{\theta^n \in \mathcal{D}} \bar{V}(\theta^n) \quad w.p.1. \tag{39}
$$

4) Now, since $\hat{v}(t, \theta_N^n) = \tilde{w}(t, \theta_N^n) + v(t)$ and $\{\tilde{w}(t, \theta^n)\}$ and $\{v(t)\}$ are independent, then

$$
\bar{V}(\theta^n) = \frac{1}{2}S_{\tilde{w}}(\theta^n) + \frac{1}{2}S_v
$$

hence, from (39)

$$
\lim_{N \to \infty} S_{\tilde{w}}(\theta_N^n) = \min_{\theta^n \in \mathcal{D}} S_{\tilde{w}}(\theta^n) = S_{\tilde{w}, \text{opt}}^n.
$$

∎

To prove Theorem 3, we need the following lemma.

Lemma 6: Consider the fullband identification method of Fig. 1. Let the collection formed by the signals $\{u(t)\}$ and $\{v(t)\}$ be quasistationary, and satisfy Assumptions 3.2–3.4. Let

$$\bar{P}_N^n = \frac{1}{N} \sum_{t=1}^{N} \sum_{s=1}^{N} \mathcal{E}\{\varphi_n^*(t)\varphi_n(s)v(t)v^*(s)\} \tag{40}$$

where $\varphi_n(t)$ is given by (14). Then

$$\lim_{N\to\infty} \bar{P}_N^n = \bar{P}^n$$

where

$$[\bar{P}^n]_{k,l} = (R_u * R_v)(k-l), \quad \forall k,l \in \{1,\dots,n\}. \tag{41}$$

Further

$$\lim_{n\to\infty} \frac{1}{n}\Omega_n^*(e^{i\omega})[\bar{R}^n]^{-1} \lim_{N\to\infty} \bar{P}_N^n[\bar{R}^n]^{-1}\Omega_n(e^{i\omega}) = \frac{\Phi_v(\omega)}{\Phi_u(\omega)}$$

where $\bar{R}^n$ and $\Omega_n(e^{i\omega})$ are given by (19) and (20), respectively.

Proof: This proof follows the steps of the [6, proof of th. 4.1], in spite of the fact that we have different assumptions. ∎

Proof of Theorem 3: We split the proof into six steps.

1) From Assumption 2, the set of parameters $\theta^n \in \mathcal{D}$, which is a compact set of $\mathbb{C}^n$. In this proof, we will allow that $\theta^n \in \mathbb{C}^n$. The result obtained will not be affected by this fact in view of (12) in Assumption 4.

2) From (10), it is straightforward to prove that

$$\theta_N^n - \theta_{\text{opt}}^n = [R_N^n]^{-1} f_N$$

where

$$f_N = \frac{1}{N}\sum_{t=1}^{N} \varphi_n^*(t)\hat{v}(t,\theta_{\text{opt}}^n) \text{ and}$$

$$R_N^n = \frac{1}{N}\sum_{t=1}^{N} \varphi_n^*(t)\varphi_n(t).$$

Then

$$N(\theta_N^n - \theta_{\text{opt}}^n)(\theta_N^n - \theta_{\text{opt}}^n)^* = [R_N^n]^{-1}Q_N^n[R_N^n]^{-1} \tag{42}$$

with

$$Q_N^n = \frac{1}{N}\sum_{t=1}^{N}\sum_{s=1}^{N} \varphi_n^*(t)\varphi_n(s)\hat{v}(t,\theta_{\text{opt}}^n)\hat{v}^*(s,\theta_{\text{opt}}^n).$$

3) Now, due to Assumption 3.1, we have that there exists $M_1 > 0$ such that

$$\mathcal{E}\left\{\left\|\sqrt{N}f_N\right\|_2^4\right\} < M_1, \quad \forall N \in \mathbb{N}. \tag{43}$$

In addition, by following an argument similar to [1, App. 9B], there exists $M_2 > 0$ such that

$$\mathcal{E}\left\{\left\|\sqrt{N}(\theta_N^n - \theta_{\text{opt}}^n)\right\|_2^4\right\} < M_2, \quad \forall N \in \mathbb{N}. \tag{44}$$

In turn, (44) implies that there exists $M_3 > 0$ such that

$$\mathcal{E}\left\{N(\theta_N^n - \theta_{\text{opt}}^n)(\theta_N^n - \theta_{\text{opt}}^n)^*\right\} < M_3, \quad \forall N \in \mathbb{N}. \tag{45}$$

4) Let $\bar{Q}_N^n = \mathcal{E}\{Q_N^n\}$. Since $\hat{v}(t,\theta_{\text{opt}}^n) = \tilde{w}(t,\theta_{\text{opt}}^n)+v(t)$, $\{v(t)\}$ is independent of $\{\tilde{w}(t,\theta_{\text{opt}}^n)\}$ and has zero mean, we have that

$$\bar{Q}_N^n = \bar{P}_N^n + \bar{E}_N^n \tag{46}$$

where $\bar{P}_N^n$ and $\bar{E}_N^n$ are given by (40) and (18), respectively.

5) We have that

$$\frac{1}{n}\Omega_n^*(e^{i\omega})[\bar{R}^n]^{-1} \lim_{N\to\infty} \bar{P}_N^n[\bar{R}^n]^{-1}\Omega_n(e^{i\omega})$$
$$\leq \left\|[\bar{R}^n]^{-1}\right\|^2 \left\|\lim_{N\to\infty} \bar{P}_N^n\right\|.$$

Since $\bar{R}^n$ and $\lim_{N\to\infty} P_N^n$ are Toeplitz matrices, then, from (41) and [12, (6), p. 64], we have that, for all $n \in \mathbb{N}$

$$\left\|[\bar{R}^n]^{-1}\right\| \leq \left\|\Phi_u^{-1}(\omega)\right\|_\infty \leq \frac{1}{\varepsilon} \text{ and}$$
$$\left\|\lim_{N\to\infty} \bar{P}_N^n\right\| \leq \left\|\Phi_u(\omega)\Phi_v(\omega)\right\|_\infty.$$

Assumption 3.3 guarantees that $\|\Phi_u(\omega)\Phi_v(\omega)\|_\infty < \infty$; then, there exists $M > 0$ such that

$$\frac{1}{n}\Omega_n^*(e^{i\omega})[\bar{R}^n]^{-1} \lim_{N\to\infty} \bar{P}_N^n[\bar{R}^n]^{-1}\Omega_n(e^{i\omega}) \leq M$$
$$\forall n \in \mathbb{N}, \ \forall \omega \in [-\pi,\pi]. \tag{47}$$

6) We have that

$$\lim_{n\to\infty}\lim_{N\to\infty} \frac{N}{n}\mathcal{E}\{S_{\tilde{w}}^n(N)\}$$
$$= \lim_{n\to\infty}\lim_{N\to\infty} \frac{1}{2\pi}\int_{-\pi}^{\pi} \frac{N}{n}\mathcal{E}|\hat{g}(e^{i\omega},\theta_N^n)-\hat{g}(e^{i\omega},\theta_{\text{opt}}^n)|^2\Phi_u(\omega)d\omega \tag{48}$$

where the exchange of the expectation is justified by Tonelli's theorem since the integrand is positive. Now, the model $\hat{g}(e^{i\omega},\theta^n)$ can be expressed as $\hat{g}(e^{i\omega},\theta^n) = \Omega_n^*(e^{i\omega})\theta^n$, where $\Omega_n^*(e^{i\omega})$ is given by (20). Then, we have (49), shown at the bottom of the next page. From (45), it follows that the integrand in (49) is bounded, and we can apply the Lebesgue's dominated convergence theorem (LDCT) to exchange the limit on N with the integral. In addition, from (44) and [11, th. 4.5.2], we can exchange this limit with the expectation. Now, from (42) and since $\{u(t)\}$ is ergodic

$$\lim_{N\to\infty} N(\theta_N^n - \theta_{\text{opt}}^n)(\theta_N^n - \theta_{\text{opt}}^n)^*$$
$$= \lim_{N\to\infty}[R_N^n]^{-1}Q_N^n[R_N^n]^{-1}$$
$$= [\bar{R}^n]^{-1} \lim_{N\to\infty} Q_N^n[\bar{R}^n]^{-1} \quad w.p.1.$$

Then, we have (50), shown at the bottom of the next page. Now, we can write

$$Q_N^n = \left[\sqrt{N}f_N\right]\left[\sqrt{N}f_N\right]^*$$

and from (43) and [11, th. 4.5.2], we can exchange the limit with the expectation in (50). Then, we have (51), shown at the bottom of the page, where the last equality follows from (16) and (46). In addition, by (47), the LDCT allows us to exchange the limit on n with the integral; then, we have (52), shown at the bottom of the page, and by Lemma 6

$$(52) = \frac{1}{2\pi} \int_{-\pi}^{\pi} \frac{\Phi_v(\omega)}{\Phi_u(\omega)} \Phi_u(\omega) d\omega + \Sigma_{\tilde{w}} = S_v + \Sigma_{\tilde{w}}.$$

■

APPENDIX C
PROOFS OF SECTION IV

Lemma 7: Let $f, g : \mathbb{Z} \to \mathbb{C}$. If $tf(t) \in l_1(\mathbb{Z})$ and $tg(t) \in l_1(\mathbb{Z})$, then $t(f * g)(t) \in l_1(\mathbb{Z})$.

Proof: We have the last equation at the bottom of the page, which is straightforward to verify that is finite. ■

Proof of Theorem 4: We need to verify that the condition in Assumption 1 holds in every subband. Then, the result follows from Assumption 7 and Theorem 2. Fix $m \in \{1, \dots, M\}$.

Assumption 1.1: The fact that the collection formed by the signals $\{U_m(t)\}$, $\{V_m(t)\}$, and $\{W_m(t)\}$ is strongly ergodic follows from Theorem 1.

We will prove that the same collection is quasistationary by phases. This implies that it is also quasistationary. To be rigorous, we should follow the steps of the proof of Theorem 1, i.e., show that a quasistationary by phases collection is invariant under the transformations under consideration. Instead, we will give a sketch of the idea in the following.

1) We have that the collection $\{\{u(t)\}, \{w(t)\}, \{v(t)\}\}$ is quasistationary by phases (Assumption 5).
2) The operations of filtering by a time-invariant filter and downsampling preserve the quasistationary by phases property. Then, it follows that $\{\{U_m(t)\}, \{W_m(t)\}, \{V_m(t)\}\}$ is quasistationary by phases.

$$(48) = \lim_{n\to\infty} \lim_{N\to\infty} \frac{1}{2\pi} \int_{-\pi}^{\pi} \frac{1}{n} \Omega_n^*(e^{i\omega}) \mathcal{E}\left\{ N(\theta_N^n - \theta_{\mathrm{opt}}^n)(\theta_N^n - \theta_{\mathrm{opt}}^n)^* \right\} \Omega_n(e^{i\omega}) \Phi_u(\omega) d\omega. \tag{49}$$

$$(49) = \lim_{n\to\infty} \frac{1}{2\pi} \int_{-\pi}^{\pi} \frac{1}{n} \Omega_n^*(e^{i\omega}) [\bar{R}^n]^{-1} \mathcal{E}\left\{ \lim_{N\to\infty} Q_N^n \right\} [\bar{R}^n]^{-1} \Omega_n(e^{i\omega}) \Phi_u(\omega) d\omega. \tag{50}$$

$$(50) = \lim_{n\to\infty} \frac{1}{2\pi} \int_{-\pi}^{\pi} \frac{1}{n} \Omega_n^*(e^{i\omega}) [\bar{R}^n]^{-1} \lim_{N\to\infty} \bar{Q}_N^n [\bar{R}^n]^{-1} \Omega_n(e^{i\omega}) \Phi_u(\omega) d\omega$$
$$= \lim_{n\to\infty} \frac{1}{2\pi} \int_{-\pi}^{\pi} \frac{1}{n} \Omega_n^*(e^{i\omega}) [\bar{R}^n]^{-1} \lim_{N\to\infty} \bar{P}_N^n [\bar{R}^n]^{-1} \Omega_n(e^{i\omega}) \Phi_u(\omega) d\omega + \Sigma_{\tilde{w}} \tag{51}$$

$$(51) = \frac{1}{2\pi} \int_{-\pi}^{\pi} \lim_{n\to\infty} \frac{1}{n} \Omega_n^*(e^{i\omega}) [\bar{R}^n]^{-1} \lim_{N\to\infty} \bar{P}_N^n [\bar{R}^n]^{-1} \Omega_n(e^{i\omega}) \Phi_u(\omega) d\omega + \Sigma_{\tilde{w}} \tag{52}$$

$$\sum_{t=-\infty}^{\infty} |t(f*g)(t)| = \sum_{t=-\infty}^{\infty} |t| \left| \sum_{\tau=-\infty}^{\infty} f(\tau) g(t-\tau) \right| \le \sum_{\tau=-\infty}^{\infty} \sum_{t=-\infty}^{\infty} |t| |f(\tau)| |g(t-\tau)|$$
$$= |f(0)| \sum_{t=-\infty}^{\infty} |tg(t)| + \sum_{\tau=-\infty, \tau\neq 0}^{\infty} |\tau f(\tau)| \left\{ |g(0)| + \sum_{t=-\infty, t\neq\tau}^{\infty} \left| \frac{t}{\tau(t-\tau)} \right| \|(t-\tau)g(t-\tau)| \right\}$$

Assumption 1.2: The fact that $\{V_m(t)\}$ is independent of $\{U_m(t)\}$ and $\{W_m(t)\}$ follows since $\{v(t)\}$ is independent of $\{u(t)\}$ and $\{w(t)\}$. ∎

Proof of Theorem 5: In view of Assumption 8, in every subband, $\theta_{m,\text{opt}}^n$ is unique. Let $\theta_N^{Mn} = [\theta_{1,N}^n, \ldots, \theta_{M,N}^n]^T$ and $\theta_{\text{opt}}^{Mn} = [\theta_{1,\text{opt}}^n, \ldots, \theta_{M,\text{opt}}^n]^T$. Then

$$\lim_{N \to \infty} S_{\tilde{w}}(\theta_N^{Mn}) = S_{\tilde{w}}\left(\lim_{N \to \infty} \theta_N^{Mn}\right) = S_{\tilde{w}}(\theta_{\text{opt}}^{Mn}) := S_{\tilde{w},\text{lim}}$$

where the exchange of the limit is valid since $S_{\tilde{w}}(\theta^{Mn})$ is a continuous function. ∎

Proof of Theorem 6: We need to verify that the condition in Assumption 3 holds in every subband. Then, the result follows from Assumption 11 and (21). Fix $m \in \{1, \ldots, M\}$.

Assumption 3.1: This assumption follows immediately from Lemma 5.

Assumption 3.2: We have that $V_m(t) = \sum_{k=-\infty}^{\infty} h_m(k)v(Dt-k)$. Then, $\{V_m(t)\}$ has zero-mean because $\{v(t)\}$ has zero-mean. In addition

$$\mathcal{E}\{V_m^*(t)V_m(t+\tau)\} = \sum_{k=-\infty}^{\infty}\sum_{l=-\infty}^{\infty} h_m^*(k)h_m(l)R_v(D\tau+k-l)$$

and therefore, $\mathcal{E}\{V_m(t)V_m(t+\tau)\}$ does not depend on t, and it follows that $\{V_m(t)\}$ is stationary.

Assumption 3.3: Let $u_m^\#(t) = l_m(q)u(t)$. It is easy to verify that $R_{u_m^\#}(\tau) = (l_m(t) * l_m^*(-t) * u(t))(\tau)$. Two applications of Lemma 7 gives that $\tau R_{u_m^\#}(\tau) \in l_1(\mathbb{Z})$. Now, since $\{u(t)\}$ is almost stationary (Assumption 9), so is $\{u_m^\#(t)\}$; then

$$R_{U_m}(\tau) = R_{u_m^\#}(D\tau) \tag{53}$$

from where it is straightforward to verify that $\tau R_{U_m}(\tau) \in l_1(\mathbb{Z})$.

Assumption 3.4: We have that $\Phi_{u_m^\#}(\omega) = \left|l_m(e^{j\omega})\right|^2 \Phi_u(\omega)$. From [13, eq. (4.1.4), p. 102], (53), and Assumption 10, we have the equation at the bottom of the page, where $\mathcal{F}\{\cdot\}$ denotes the Fourier transform. ∎

Proof of Proposition 3: The map $T_h : y(t) \mapsto Y(t) = [Y_1(t), \ldots, Y_M(t)]^T$ is given by

$$Y_m(t) = [(T_h y(t))(\tau)]_m$$
$$= \langle y(t), h_m^*(\tau D - t)\rangle, \; \forall m = 1, \ldots, M, \; \tau \in \mathbb{Z}$$

where $[\alpha]_m$ denotes the mth component of the vector α, and $\langle \cdot, \cdot \rangle$ denotes the inner product on $l_2(\mathbb{Z})$. Then, the result can be reached following the proof of [14, Sect. 3.3.2, p. 67].

For the norm of T_f^*, we have that $T_f^* : \hat{V}(t,\theta) = [\hat{V}_1(t,\theta), \ldots, \hat{V}_M(t,\theta)] \mapsto \hat{v}(t,\theta)$ is given by

$$\hat{v}(t,\theta) = T_f^* \hat{V}(t,\theta) = \sum_{m=1}^{M}\sum_{\tau=-\infty}^{\infty} \hat{V}_m(t,\theta)f_m(t-\tau D).$$

From [14, Sect. 3.2, p. 57, (3.2.3)], it follows that $T_f : \hat{v}(t,\theta) \mapsto \hat{V}(t,\theta)$ is given by

$$\hat{V}_m(t,\theta) = [(T_f\hat{v}(t,\theta))(\tau)]_m$$
$$= \langle \hat{v}(t,\theta), f_m^*(\tau D - t)\rangle, \; \forall m = 1, \ldots, M, \; \tau \in \mathbb{Z}$$

and therefore, the result follows since $\left\|T_f^*\right\| = \|T_f\|$. ∎

Lemma 8: Let $\{X(t) = [X_1(t), \ldots, X_M(t)]^T\}$, $t \in \mathbb{Z}$ be an array of quasistationary random processes, and let $H(q) = [H_{ij}(q)]$, $i,j = 1, \ldots, M$, satisfying $H_{ij}(t) \in l_1(\mathbb{Z})$. Let $\breve{X}(t) = [\breve{X}_1(t), \ldots, \breve{X}_M(t)]^T$ be generated from $\{X(t)\}$ by upsampling by a factor of U. Let $Y(t) = [Y_1(t), \ldots, Y_M(t)]^T$ 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 $x(t) = [x_1(t), \ldots, x_M(t)]^T \in l_2^M(\mathbb{Z})$, 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 $\|y(t)\|_2 \leq T\|x(t)\|_2$, where $\|x(t)\|_2^2 = \sum_{m=1}^{M}\|x_m(t)\|_2^2$, then

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

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: Let $\Sigma_Y = \lim_{N \to \infty}(1/N)\sum_{t=1}^{N}\mathcal{E}\{Y(t)Y^*(t)\}$; then

$$\Sigma_Y$$
$$= \sum_{k=-\infty}^{\infty}\sum_{l=-\infty}^{\infty} H(k)\lim_{N \to \infty}\frac{1}{N}\sum_{t=1}^{N}\mathcal{E}\left\{\breve{X}(Dt-k)\breve{X}^*(Dt-l)\right\}H^*(l).$$
$$\tag{54}$$

Without details, we will say that the exchange of the expectation with the infinite sums is valid in view of Fubini's Theorem since $\{X(t)\}$ has uniformly bounded second moments (recall the definition of quasistationary), and the exchange of the limit with the infinite sums is valid in view of Lebesgue's dominated

$$\Phi_{U_m}(\omega) = \mathcal{F}\{R_{U_m}(\tau)\} = \frac{1}{D}\sum_{d=0}^{D-1}\Phi_{u_m^\#}\left(2\pi\frac{d}{D} + \frac{\omega}{D}\right)$$
$$= \frac{1}{D}\sum_{d=0}^{D-1}\left|l_m(\Omega^d e^{j(\omega/D)})\right|^2 \Phi_u\left(2\pi\frac{d}{D} + \frac{\omega}{D}\right) \geq \alpha\varepsilon, \; \forall \omega \in [-\pi, \pi]$$

$$(54) = \frac{D}{U} \sum_{k=-\infty}^{\infty} \sum_{l=-\infty}^{\infty} H(k) \lim_{n \to \infty} \frac{1}{n} \sum_{t=-\infty}^{\infty} \mathcal{E} \left\{ \breve{X}_n(Dt - k)\breve{X}_n^*(Dt - l) \right\} H^*(l)$$

$$= \frac{D}{U} \lim_{n \to \infty} \frac{1}{n} \sum_{t=-\infty}^{\infty} \mathcal{E} \{Y_n(t)Y_n^*(t)\}$$

---

convergence theorem since $H(t) \in l_1^{M \times M}(\mathbb{Z})$. Now, define the sequence of truncations $\{X_n(t)\}$ as follows:

$$X_n(t) = \begin{cases} X(t), & 1 \le t \le n \\ 0, & \text{otherwise} \end{cases}$$

and let $\{\breve{X}_n(t)\}$ and $\{Y_n(t)\}$ be generated from $\{X_n(t)\}$ in the same way as $\{\breve{X}(t)\}$ and $\{Y(t)\}$ are generated from $\{X(t)\}$. It is quite straightforward to show that for all $k, l \in \mathbb{Z}$

$$\lim_{N \to \infty} \frac{1}{N} \sum_{t=1}^{N} \mathcal{E}\{\breve{X}(Dt - k)\breve{X}^*(Dt - l)\}$$

$$= \frac{D}{U} \lim_{n \to \infty} \frac{1}{n} \sum_{t=-\infty}^{\infty} \mathcal{E} \left\{ \breve{X}_n(Dt - k)\breve{X}_n^*(Dt - l) \right\}.$$

Then, we have the equation at the top of the page, where the exchange of the limit with the infinite sums is valid for the same argument given above.

Now, let $(\Omega, \mathcal{A}, \mathbb{P})$ be the underlying probability space of $\{X(t)\}$. We have that for every $\omega \in \Omega$, $X_n(t, \omega) \in l_2^M(\mathbb{Z})$, and $\|Y_n(t, \omega)\|_2 \le T \|X_n(t, \omega)\|_2$. Then

$$S_Y = \text{Tr}\{\Sigma_Y\} = \frac{D}{U} \lim_{n \to \infty} \frac{1}{n} \sum_{t=-\infty}^{\infty} \mathcal{E}\{Y_n^*(t)Y_n(t)\}$$

$$= \frac{D}{U} \lim_{n \to \infty} \frac{1}{n} \mathcal{E}\left\{\|Y_n(t)\|_2^2\right\}$$

$$\le \frac{D}{U}T^2 \lim_{n \to \infty} \frac{1}{n} \mathcal{E}\left\{\|X_n(t)\|_2^2\right\} = \frac{D}{U}T^2 S_X.$$

To prove the last part, replace the inequality with an equality. ∎

Proof of Theorem 7: Let $S_{\breve{w}}^n(N)$ denote the power of the signal $\breve{w}(t, \theta_N^n) = \hat{w}(t, \theta_N^n) - \hat{w}(t, \theta_{\text{opt}}^n)$, and for each $m = 1, \dots, M$, let $S_{\breve{W}_m}^n(N)$, denote the power of the signal $\breve{W}_m(t, \theta_N^{Mn}) = \hat{W}_m(t, \theta_N^{Mn}) - \breve{W}_m(t, \theta_{\text{opt}}^{Mn})$. We have that $S_{\breve{w}, \text{dif}}^n(N) \le S_{\breve{w}}^n(N)$. Then, from Lemma 8

$$\lim_{n \to \infty} \lim_{N \to \infty} \frac{N}{n}\mathcal{E}\{S_{\breve{w}, \text{dif}}^n(N)\}$$

$$\le \lim_{n \to \infty} \lim_{N \to \infty} \frac{N}{n}\mathcal{E}\{S_{\breve{w}}^n(N)\}$$

$$\le \lim_{n \to \infty} \lim_{N \to \infty} \frac{N}{n}\mathcal{E}\left\{\frac{1}{D}\|T_f\|^2 \sum_{m=1}^{M} S_{\breve{W}_m}^n(N)\right\}$$

$$= \|T_f\|^2 \sum_{m=1}^{M} \lim_{n \to \infty} \lim_{N \to \infty} \frac{N}{Dn}\mathcal{E}\left\{S_{\breve{W}_m}^n(N)\right\}. \quad (55)$$

From (22), we have that, in every subband

$$\lim_{n \to \infty} \lim_{N \to \infty} \frac{N}{n}\mathcal{E}\left\{S_{\breve{W}_m}^n(N)\right\} = \sum_{m=1}^{M} S_{V_m}. \quad (56)$$

From (55), (56) and Lemma 8

$$(55) = \frac{1}{D}\|T_f\|^2 \sum_{m=1}^{M} S_{V_m} \le \|T_f\|^2 \|T_h\|^2 S_v. \quad (57)$$

∎

REFERENCES

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

[2] M. L. Honig and D. G. Messerschmit, *Adaptive Filters: Structures, Algorithms, and Applications*. Boston, MA: Kluwer, 1984.

[3] 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.

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

[5] D. Marelli and M. Fu, "Subband methods for ofdm equalization," Proc. IEEE Int. Conf. Commun., 2003, to be published.

[6] L. Ljung and Z.-D. Yuan, "Asymptotic properties of black-box identification of transfer functions," *IEEE Trans. Automat. Contr.*, vol. AC-30, pp. 514–530, June 1985.

[7] W. Kellermann, "Analysis and design of multirate systems for cancellation of acoustical echoes," *Proc. IEEE Int. Conf. Acoust., Speech, Signal Processing*, 1988.

[8] D. Marelli and M. Fu, "Optimized filterbank design for subband identification with oversampling," *Proc. IEEE Int. Conf. Acoust., Speech, Signal Processing*, 2001.

[9] ——, "Convergence properties of subband identification," in *Proc. Asian Control Conf.*, 2002.

[10] ——, "Performance analysis for subband identification," *IEEE Trans. Signal Processing*, submitted for publication.

[11] K. L. Chung, *A Course in Probability Theory*. New York: Academic, 1974.

[12] U. Grenander and G. Szeg, *Toeplitz Forms and Their Applications*. Berkley, CA: Univ. California Press, 1958.

[13] P. P. Vaidyanathan, *Multirate Systems and Filterbanks*. Englewood Cliffs, NJ: Prentice-Hall, 1993.

[14] I. Daubechies, "Discrete wavelet transform: Frames," in *Ten Lectures on Wavelets*. Philadelphia, PA: SIAM, 1992, ch. 3.

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 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*.