Convergence Properties of Subband Identification

Damian Marelli and Minyue Fu
Department of Electrical and Computer Engineering
University of Newcastle, N.S.W. 2308 Australia

Abstract

The so-called *subband identification* method has been introduced recently as an alternative method for identification of finite-impulse response systems with a large tap size. It is known that this method can be more numerically efficient than the classical system identification method, while having compatible performance. In this paper, we assume a probabilistic framework, and we deal with the following two problems: (1) whether or not the identification result depends on the particular realization of the random process under consideration, and (2) whether or not, the identification algorithm converges to the minimum of the error function. We study these properties by considering both the error functions in individual subbands and a combined error function. We study the critical-sampling and the oversampling cases. We show that optimum convergence is not always guaranteed in the oversampling case. A slight modification in the identification algorithm will be proposed to fix this problem.

1 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 an finite-impulse-response (FIR) filter with a large tap size.

The key idea of the subband identification method is to subdivide the given input-output signals of the system into a number of subbands in the frequency domain by using filterbanks and down-samplers and identify subband models of the system. 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 down-sampling factor, whereas in the oversampling scheme the number of subbands is more than the down-sampling factor. For comparative purposes, we will refer the classical system identification method as fullband identification. The reader is referred to [1, 2] for an introduction to fullband identification.

Subband identification has been used in speech signal processing applications where long FIR models are often required. See, for example, [3, 4, 5, 6, 7, 8, 9]. In general, there is 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 non-overlapping filterbanks which result in spectral gaps between subbands; see [4]. In order to cope with this problem, the paper [5] used auxiliary channels, with the corresponding extra computational cost. Finally, the paper [6] introduced the use of adaptive cross-terms between subbands. However, these cross-terms increase the computational cost and the slow the convergence rate. The second approach uses oversampling. For example, the paper [7] analyses the existence of exact solutions of the identification problem without cross-terms. The paper [8] use the gabor expansion to design the filterbanks, which restrict the flexibility for the filterbanks. The paper [9] analyzes the performance of the oversampling case under a number of simplifying assumptions, and [10] does it in a more rigorous way.

In [10], a comparative study is given for subband identification vs. fullband identification. It is shown that the subband method, with a careful choice of design parameters (number of subbands, downsampling factor, filterbanks, and subband models), offers compatible asymptotic residual error and asymptotic convergence rate but with a significantly smaller computational cost. This analysis is done in a probabilistic framework, i.e. under the assumption that, the input signal and the output noise are random processes. When working in a probabilistic framework, an interesting question is whether there is *strong convergence*, i.e. whether the result of the identification depends on a particular realization of the random process. Another question to ask is whether there is *optimum convergence*, i.e. whether or not the identification result converges to the global minimum of the error function. These two properties together will be refered to as *convergence properties*. The purpose of this paper is to give conditions to satisfy these convergence properties for error functions in individual subbands and for a combined error function. We will study both the critical-sampling case and the oversampling case. We will show that optimum convergence is not always guaranteed in the oversampling case. Based on this observation, we will propose a slight modification in the identification algorithm that fixes this problem.

2 Subband Identification

The scheme of subband identification is depicted in figure 1. The idea of subband identification is to split both signals $u(t)$ and $y(t)$ into M subbands using analysis filterbanks $h(q) = [h_1(q), ..., h_M(q)]^T$. These subband signals are down-sampled and the results are denoted by two vector signals $U(t) = [U_1(t), ..., U_M(t)]^T$ and $Y(t) = [Y_1(t), ..., Y_M(t)]^T$. The subband parametric model $\hat{G}(q, \theta) = \text{diag}\{\hat{G}_m(q, \theta_m),\ m = 1, ..., M\}$, where $\theta = [\theta_1^T, ..., \theta_M^T]^T$, is identified in order to reconstruct $\hat{W}(t, \theta) = [\hat{W}_1(t, \theta_1), ..., \hat{W}_M(t, \theta_m)]^T$ which is the subband equivalent of $\hat{w}(t, \theta)$. The prediction error $\hat{V}(t, \theta) = [\hat{V}_1(t, \theta_1), ..., \hat{V}_M(t, \theta_m)]^T = Y(t) - \hat{W}(t, \theta)$ is then formed. Finally, an up-sampler and a synthesis filterbank $f(q) = [f_1(q), ..., f_M(q)]^T$ is used to reconstruct $\hat{v}(t, \theta)$.

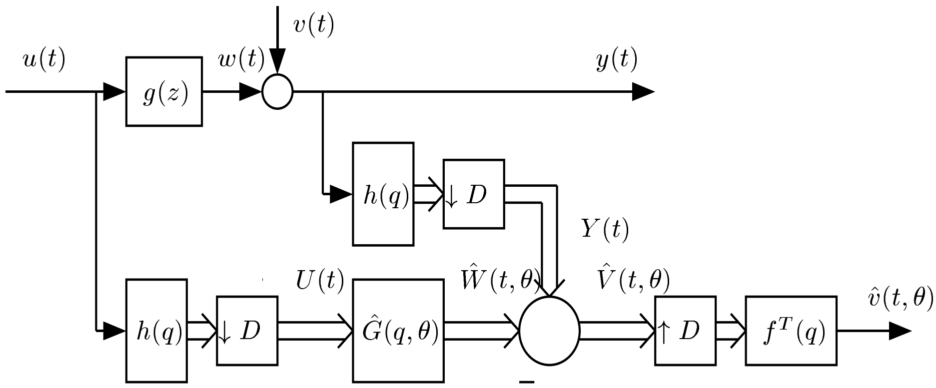

Figure 1: Subband identification, direct representation

For analysis purposes, we denote by $W(t) = [W_1(t), ..., W_M(t)]^T$ the downsampled version of $h(q)w(t)$. We define $\tilde{W}(t, \theta) = W(t) - \hat{W}(t, \theta) = [\tilde{W}_1(t, \theta_1), ..., \tilde{W}_M(t, \theta_m)]^T$ 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 $V(t) = [V_1(t), ..., V_M(t)]^T$.

In [10], it was shown that, with a careful choice of design parameters (number of subbands, downsampling factor, filterbanks, and subband models), the performance of the subband method, in terms of asymptotic residual error and asymptotic convergence rate, can be made compatible to fullband identification. However, the computational cost of the subband method can be significantly smaller. We summarize these results below:

Define the identification error as the power of the signal $\tilde{w}(t, \theta)$, given by

$$S_{\tilde{w}}(\theta) = \lim_{T \to \infty} \frac{1}{T} \sum_{t=1}^{T} \mathcal{E}\{\tilde{w}^*(t, \theta)\tilde{w}(t, \theta)\}$$

where the superscript * denotes complex conjugate, and let θ_N denote the set of parameters computed, by the identification algorithm, up to time N. Then, the asymptotic residual error is defined as

$$S_{\tilde{w},\lim} = \lim_{N \to \infty} S_{\tilde{w}}(\theta_N) \tag{1}$$

Suppose we identify the system with the fullband method, using an FIR model of order n_f, and let $S_{\tilde{w},\lim}^{\text{FB}}$ be the asymptotic residual error obtained in this way. Now, suppose we also identify the same system using the subband method. Assume that the subband models $\hat{G}_m(q, \theta_m)$, $m = 1, ..., M$ are FIR with tap size n_s, and the filters $h_m(q)$, $m = 1, ..., M$ are FIR with tap size l_h. If the input signal $\{u(t)\}$ is white, then the asymptotic residual error of the subband method $S_{\tilde{w},\lim}$ is bounded by

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

where $J(l_h)$ and $K(n_s)$ are decreasing functions of l_h and n_s respectively, whose expressions are irrelevant for the purposes of this paper.

For the convergence rate, we have that, for large n_s and N, and for small $S_{\tilde{w},\lim}$,

$$\mathcal{E}\{S_{\tilde{w}}(\theta_N) - S_{\tilde{w},\lim}\} \lesssim \frac{n_f}{N} S_v \tag{3}$$

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

The asymptotic convergence rate (3) coincides with that of the fullband method. Therefore, by choosing l_h and n_d large enough, we can have compatible performances on both methods. However, by optimizing the values of M and D, the computational cost of the subband method can be made significantly smaller.

Let us now explain the two convergence properties we are looking for. For a fixed value of θ, $S_{\tilde{w}}(\theta)$ is a deterministic value. However, since θ_N, $N \in \mathbb{N}$ (the set of natural numbers, i.e. integers greater or equal to one) is a random process, it follows that $S_{\tilde{w}}(\theta_N)$ is a random process. Therefore, we hope for the nice property that the limit in (1) does not depend on the particular realization of the random processes $\{u(t)\}$ and $\{v(t)\}$ (strong convergence), i.e. that there exists $S_{\tilde{w},\lim} > 0$, such that

$$S_{\tilde{w},\lim} = \lim_{N\to\infty} S_{\tilde{w}}(\theta_N) \quad w.p.\ 1 \tag{4}$$

where $w.p.$ 1, means *with probability one*. Once this is met, we further hope that $S_{\tilde{w},\lim}$ coincides with the optimum value of the function $S_{\tilde{w}}(\theta_N)$ (optimum convergence), i.e.

$$S_{\tilde{w},\lim} = \min_{\theta\in\mathcal{D}} S_{\tilde{w}}(\theta) \tag{5}$$

where $\mathcal{D}$ denotes the range of θ.

3 Convergence Properties

3.1 In Every Subband

In the m-th subband, let $S_{\tilde{W}_m}(\theta_m)$ denote the power of the signal $\tilde{W}_m(t, \theta_m)$ and let $\theta_{m,N}$ denote the set of parameters, on that subband, computed up to time N. Equation (2) is derived under the assumption that

$$\lim_{N\to\infty} S_{\tilde{W}_m}(\theta_{m,N}) = \min_{\theta_m\in\mathcal{D}_m} S_{\tilde{W}_m}(\theta_m) \quad w.p.\ 1, \quad m = 1, ..., M \tag{6}$$

So, as the first step in the study on convergence properties, we will state the conditions for (6) to hold. The following assumptions are required:

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

1. *$\{u(t)\}$, $\{w(t)\}$ and $\{v(t)\}$ random processes, generated by any arbitrary combination of filtered white noise and deterministic signals.*
2. *$\{v(t)\}$ is independent of $\{u(t)\}$ and $\{w(t)\}$.*
3. *for all $A, B \in \mathbb{N}$ and $a, b \in \mathbb{Z}$ (the set of integers), the following limit exists*

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

where $x(t)$ and $y(t)$ denotes any of $u(t)$, $w(t)$ and $v(t)$.

Assumption 2 *We assume the prediction error identification method, i.e. $\theta_{m,N}$ is chosen as follows*

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

where

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

Assumption 3 *The subband model is a diagonal matrix $\hat{G}(q,\theta) = diag\{\hat{G}_m(q,\theta_m),\ m = 1, ..., M\}$, where $\theta = [(\theta_1)^T, ..., (\theta_M)^T]^T$ and $\theta_m \in \mathcal{D}_m \subset \mathbb{C}^{n_m}$, $m = 1, ..., M$, where $\mathcal{D}_m$ is assumed to be compact (i.e. close and bounded). For each $m \in \{1, ..., M\}$, the model $\hat{G}_m(q,\theta_m)$ is a parametric linear model. There exists $\hat{G}(q) \in l_1(\mathbb{Z})$ such that $|\hat{G}_m(q,\theta_m)| < \hat{G}(q)$, for all $\theta_m \in \mathcal{D}_m$, and there exists $\hat{G}'(q) \in l_1(\mathbb{Z})$, such that $|\hat{G}'_{m,k}(q,\theta_m)| < \hat{G}'(q)$, for all $\theta_m \in \mathcal{D}_m$, and for $k = 1, ..., n_m$, where $\hat{G}'_{m,k}(q,\theta)$ is the k-th component of the vector $\hat{G}'(q,\theta_m) = \frac{\partial}{\partial\alpha}\hat{G}(q,\alpha)\Big|_{\theta_m}$.*

Assumption 4 *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 $m = 1, ..., M$.*

Theorem 1 *Consider the subband identification scheme of figure 1, together with assumptions 1-4. Then, (6) is satisfied.*

Proof: See [11, Theorem 4]. ∎

3.2 Subbands Combined

Now we turn into the convergence of the overall error $S_{\tilde{w}}(\theta_N)$. We treat the critical-sampling and the oversampling cases separately. Note that critical-sampling is a particular case of the oversampling case, so everything that applies to the latter, also applies to the former.

3.2.1 Critical-sampling

Definition 1 *Consider the subband identification scheme of figure 1. Suppose we make $u(t) = 0$ and*

$$f(t) = \frac{1}{c} h^*(-t) \tag{7}$$

If $\hat{v}(t,\theta) = v(t)$, then the filterbank $h(q)$ is called paraunitary [12].

In the critical-sampling case, both (4) and (5) are satisfied by requiring the following extra assumption:

Assumption 5 *The analysis filterbank $h(q)$ is paraunitary, and the synthesis filterbank $f(q)$ is given by (7).*

Theorem 2 *Consider the subband identification scheme of figure 1, in the critical-sampling case (i.e. $D = M$), together with assumptions 1-5. Then, both (4) and (5) are satisfied*

Proof: See [10, Corollary 2]. ∎

3.2.2 Oversampling

The following theorem gives the conditions for (4) to be satisfied. One extra assumption is required:

Assumption 6 *In every subband, the set* $\arg\min_{\theta_m \in \mathcal{D}_m} S_{\tilde{W}_m}(\theta_m)$ *has only one element.*

Remark 1 *Note that assumption 6 is easy to satisfy. It is satisfyed if, for example the subband models* $\hat{G}_m(q, \theta_m)$ *are FIR and each subband input signal* $U_m(t)$ *have its spectrum bounded from below by a positive constant.*

Theorem 3 *Consider the subband identification scheme of figure 1, together with assumptions 1-4 and 6. Then, (4) is satisfied.*

Proof: See [11, Theorem 5]. ■

In the oversampling case, (5) is not satisfied in general. However, both (4) and (5) can be guaranteed by a slight modification in the identification method of assumption 2. We state the modification in the following assumption:

Assumption 7 $\theta_{m,N}$ *is chosen as follows:*

$$\theta_{m,N} \in \arg \min_{\theta_m \in \mathcal{D}_m} V_N(\theta_m)$$

where

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

and $\hat{v}(t, \theta_m)$ *denotes the random process* $\hat{v}(t, \theta)$ *expressed as a function of the parameters of the* m*-th parametric model* $\hat{G}_m(q, \theta_m)$*.*

Theorem 4 *Consider the subband identification scheme of figure 1, together with assumptions 1-4 and 7. Then, both (4) and (5) are satisfied*

Proof: To be provided in the final version of the paper. ■

4 Conclusion

In this work we have studied the convergence properties, i.e. strong convergence and optimum convergence, of the subband identification method. We have provided the conditions required to satisfy both properties for both the error functions in individual subbands and a combined error function.

References

[1] L. Ljung, *System Identification: Theory for the User.* Prentice Hall, second ed., 1999.

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

[3] A. Gilloire, "Experiments with sub-band acoustic echo cancellers for teleconferencing," in *Proceedings of the IEEE International Conference on Acoustics, Speech, and Signal Processing*, 1987.

[4] H. Yasukawa, S. Shimada, and I. Furukawa, "Acoustic echo canceller with high speech quality," in *Proceedings of the IEEE International Conference on Acoustics, Speech, and Signal Processing*, 1987.

[5] V. S. Somayazulu, S. K. Mitra, and J. J. Shynk, "Adaptive line enhancement using multirate techniques," in *Proceedings of the IEEE International Conference on Acoustics, Speech, and Signal Processing*, 1989.

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

[7] W. Kellermann, "Analysis and design of multirate systems for cancellation of acoustical echoes," in *Proceedings of the IEEE International Conference on Acoustics, Speech, and Signal Processing*, 1988.

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

[9] M. R. Petraglia and S. K. Mitra, "Performance analysis of adaptive filter structures based on subband decomposition," in *Proceedings of the IEEE International Symposium on Circuits and Systems*, 1993.

[10] D. Marelli and M. Fu, "Performance analysis for subband identification." Submitted to IEEE Transactions on Signal Processing.

[11] D. Marelli and M. Fu, "Asymptotic properties of subband identification." Submitted to IEEE Transactions on Signal Processing.

[12] P. P. Vaidyanathan, *Multirate Systems and Filterbanks.* Prentice Hall, 1993.