Filter bank optimization with applications in noise suppression and communications.

Thesis by

Sony Akkarakaran

In Partial Fulfillment of the Requirements
for the Degree of
Doctor of Philosophy

California Institute of Technology
Pasadena, California

2001
(Submitted May 4, 2001)

Acknowledgments

First of all I would like to express my heartfelt thanks to my advisor, Dr. Vaidyanathan, for his outstanding guidance and support during my stay at Caltech. His work as a teacher and researcher has been truly inspiring, and his encouragement, enthusiasm, friendship and funding have been vital throughout my research and my stay at Caltech. I am especially thankful to him for the limitless patience and support he has shown during periods when progress was slow.

I would like to thank all members of my defense and candidacy examining committees for their interest: Dr. R.J.McEliece, Dr. Dariush Divsalar, Dr. Babak Hassibi, Dr. Igor Djokovic, Dr. Aaron Kiely, and Dr. See-May Phoong. Generous funding from the National Science Foundation and the Office of Naval Research is also gratefully acknowledged.

I would like to thank my labmates, Murat Mese, Bojan Vrcelj and Andre Tkacenko, who made my stay in the DSP lab enjoyable and entertaining. Murat has been my office–mate for five years now, and we are headed out to jobs at the same destination. His friendship has been very valuable to me. Andre and Bojan brought cheer and humor to the lab. Dr. See-May Phoong, Dr. Yuan-Pei Lin and Dr. Igor Djokovic are frequent visitors to our lab, and I am grateful for their encouragement and interest in my work. See-May served on my candidacy examining committee, and Igor is on my defense committee. I would also like to thank Dr. Ahmet Kirac and Dr. Jamal Tuqan for their company during my starting years here.

I am grateful to many of the Braun House guys — especially Aamod Khandekar, Shachi Gosavi, MA Srinivas, Siddhartha Gadgil, Neelesh Mehta, Pururav Thoutireddy, and others, for the friendship, company and entertainment they provided, that helped a lot in making my stay here pleasant. Discussions with them have often helped me in my research.

Finally, I would like to thank my parents for their sacrifice, constant support and inspiration, and for encouraging me to pursue higher studies.

Abstract

A filter bank (FB) is used to analyze or decompose a signal into several frequency bands, which are processed separately and then combined. This allows us to allocate processing resources in a manner tailored to the distribution of the relevant signal features among the bands. A judicious allocation leads to improved system performance over direct processing of the input signal (without using a FB). FBs have found applications in almost every area of modern digital signal processing, including audio, image and video compression and communications.

The main thrust of this thesis is towards the optimization of FBs based on the statistical properties of their input. We establish the optimality of a type of FB called the **principal component filter bank (PCFB)** for numerous signal processing problems. The PCFB depends on the input power spectrum and on the class of M channel orthonormal FBs over which we seek to optimize the FB. PCFB optimality for compression and progressive transmission has been observed to varying degrees in the past. Our work provides a unified framework for orthonormal FB optimization, that includes these earlier results as special cases. It also covers many other problems not observed earlier, notably in noise suppression and communications.

A central result that we establish is that the PCFB is the optimum orthonormal FB whenever the minimization objective is a concave function of the vector of subband variances of the FB. Many signal processing problems result in such objectives. The earlier results on PCFB optimality for compression can be explained by this framework. Another example not noticed earlier is FB based white noise reduction using zeroth order Wiener filters or hard thresholds in the subbands. Yet another case involves the discrete multitone modulation (DMT) communication system, used in ADSL (asymmetric digital subscriber line) and wireless OFDM (orthogonal frequency division multiplexing) technologies. These systems use the transmultiplexer configuration of a FB which is usually chosen as a DFT or cosine–modulated FB for efficiency of implementation. We show that at increased implementation cost, we can minimize the transmission power requirement (for a given bitrate and error probability) by using the PCFB associated with a certain normalized noise spectrum. We present simulation examples with realistic channel and noise models for the ADSL system to compare the performance of the PCFB against other types of FBs such as the DFT.

We study various extensions of the basic PCFB optimality result. The noise suppression problem becomes more involved when the noise is colored, because the objective then depends on both signal and noise subband variances. For a specific FB class, namely the orthogonal transform coder class, we show that a simultaneous PCFB for the signal and noise is optimal (if it exists). For the class of

unconstrained FBs, this does not hold in general; we develop an algorithm that computes the best FB for piecewise constant spectra. In some cases, PCFB optimality extends to classes of biorthogonal FBs too, although there are many open problems in this area, as we point out. We study the effect of nonexistence of a PCFB on the FB optimizations, and show how they usually become analytically intractable. We show that PCFBs do not exist for the classes of DFT and cosine–modulated FBs. We also study nonuniform FB optimization: We establish the definition of nonuniform PCFBs and study their existence and optimality, which are shown to be much more restricted when compared with uniform PCFBs.

Lastly, we study a related open problem on the parameterization of nonuniform perfect reconstruction (PR) FBs of various classes such as the rational and FIR classes. Not all nonuniform PRFBs can be built by the common method of using tree structures of uniform PRFBs. Given a set of decimators, is there a rational PRFB using them? If so, what are all the PRFBs possible? When are they necessarily derivable from a tree structure? Very little is known about the answers to many such questions. For example, for existence of rational PRFBs with a given set of decimators, certain conditions on the decimators are known to be necessary, while certain others are sufficient. However, conditions that are both necessary and sufficient are unknown. One of our contributions is to strengthen considerably the known conditions. This is an important step towards a complete PR theory for nonuniform filter banks.

Contents

List of Figures

List of Tables

Chapter 1 Introduction

In today's internet era, there is an obvious need for efficient techniques for storage, transmission and processing of digital information of various kinds, such as speech, images and video. The field of multirate digital signal processing (DSP) has produced several versatile practical solutions to these needs. A central idea in this field is the analysis of a signal at different resolutions or scales; achieved by decomposing the signal into different frequency bands. A fundamental building block used to perform this decomposition is the digital filter bank. Filter banks have found applications in almost every area of modern DSP, including image and video compression, digital audio processing, adaptive and statistical signal processing, and communications.

The present thesis is a contribution towards the design and optimization of filter banks adapted to the statistical properties of their input signals. In this introductory chapter, we discuss briefly the basic principles of multirate systems and filter banks, outline the contributions of subsequent chapters, and establish notations to be used later. The thesis is reasonably self contained, and this introduction contributes to making it so, but is by no means an exhaustive treatment of multirate DSP. For further details, history, and applications, the interested reader is referred to several texts on the topic, e.g., [64],[24],[42],[43], [57],[67]. Most of our notation parallels [64].

1.1 Multirate systems and filter banks

1.1.1 Building blocks of multirate systems

The signals of interest in multirate DSP are discrete sequences of real or complex numbers, denoted by $x(n), y(n)$, etc. The sequence $x(n)$ could for instance be obtained by sampling of a continuous–time signal $x_c(t)$, which may represent, for example, the amplitude of a speech signal. In practice usually the signal $x(n)$ is also *quantized*, i.e., discretized not only in time but also in amplitude. However in many situations it suffices and is convenient to regard $x(n)$ as a discrete–time signal with continuous amplitude. It is often of interest to study frequency domain representations of the signal $x(n)$. Commonly used ones are the z-transform, denoted by $X(z)$, defined as $X(z) = \sum_{n=-\infty}^{\infty} x(n)z^{-n}$, and the discrete–time Fourier transform $X(e^{j\omega})$ which is $X(z)$ evaluated on the unit circle $z = e^{j\omega}$ in the z plane.

The basic blocks in a multirate DSP system, operating on a discrete time signal $x(n)$, are the linear time invariant (LTI) filter, the decimator and the expander. An LTI filter, shown in Fig. 1.1, is characterized by its impulse response $h(n)$, or by its transfer function $H(z)$ (which is the z-transform

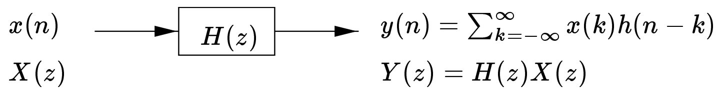


Figure 1.1: A linear time invariant filter.

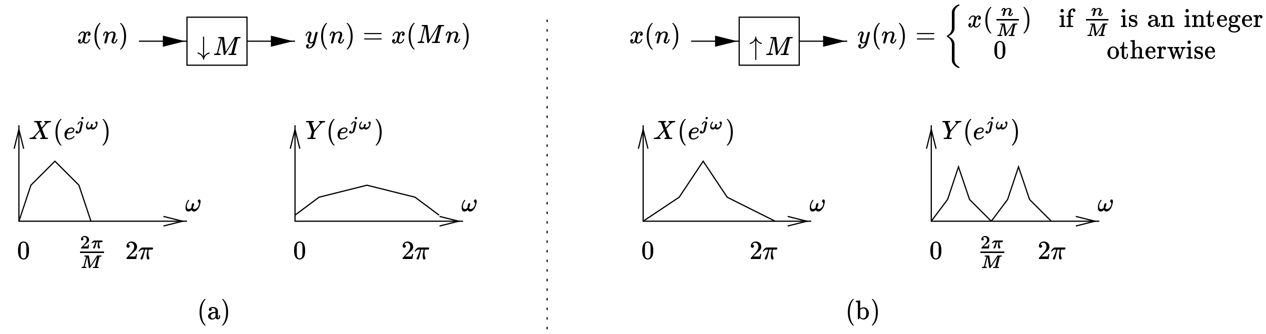


Figure 1.2: M-fold decimator (a) and expander (b).

of $h(n)$), or by its frequency response $H(e^{j\omega})$. The M-fold decimator and expander are shown in Fig. 1.2. Their presence in a system creates signals that are sampled at different rates, giving rise to the term 'multirate' system. Figure 1.2 also shows the frequency domain behaviour of the decimator and expander (the plots illustrate the case when $M = 2$). Their z-domain behavior is described by

$$Y(z) = (X(z))\downarrow_M = \frac{1}{M}\sum_{k=0}^{M-1} X(z^{\frac{1}{M}} e^{\frac{-j2\pi k}{M}}) \quad \text{for } M\text{-fold decimator, and} \tag{1.1}$$

$$Y(z) = (X(z))\uparrow_M = X(z^M) \quad \text{for } M\text{-fold expander.} \tag{1.2}$$

The expander causes 'imaging' of the spectrum $X(e^{j\omega})$, i.e., an M-fold compression along the ω axis resulting in an M-fold increase in periodicity (period changes from 2π to $2\pi/M$). The decimator causes 'aliasing', i.e., $\big(X(e^{j\omega})\big)\downarrow_M$ is obtained by making an M-fold expansion of $X(e^{j\omega})$ along the ω axis, creating M copies of the result each shifted from the next by 2π along the ωaxis, and averaging these waveforms. We say that $X(e^{j\omega})$ has an aliasfree(M) support if these shifted copies do not overlap, i.e., if for any ω, $X(e^{j\omega})$ is nonzero for at most one of the M values $\omega + \frac{2\pi k}{M}$, $k = 0, 1, \ldots, M-1$.

All systems in Figs. 1.1,1.2 are single input single output (SISO) or scalar systems, but we also encounter their multi input multi output (MIMO) or vector counterparts, which are defined by obvious extension: MIMO decimators and expanders operate as banks of SISO ones (i.e., independently on each component of their input vector signal), and a MIMO LTI system is characterized by a (possibly rectangular) transfer matrix rather than a scalar transfer function $H(z)$. Also, all systems in this thesis are one–dimensional (1-D), i.e., all signals $x(n)$ are functions of a single variable n (which usually represents time). An image is an example of a two–dimensional signal $f(i,j)$; it depends

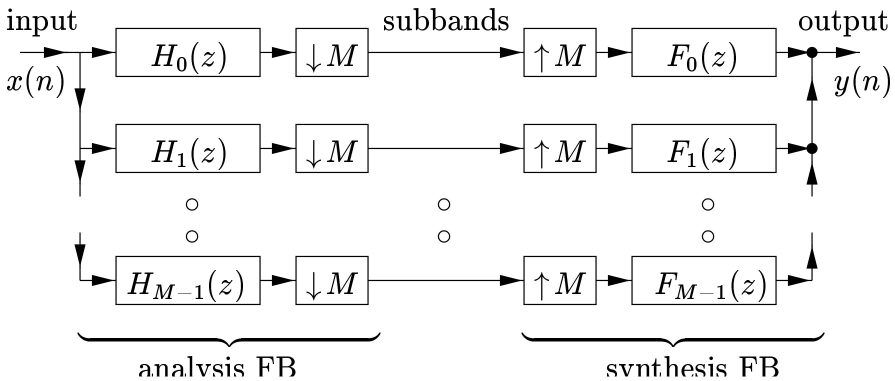


Figure 1.3: Uniform M channel maximally decimated filter bank.

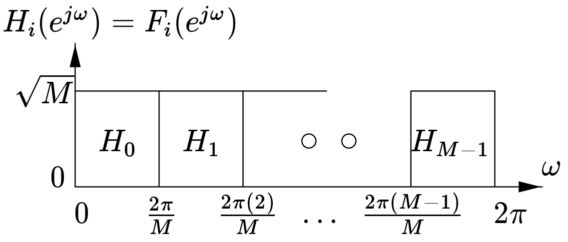


Figure 1.4: Contiguous–stacked brickwall filter bank.

on two position (or space) variables. Most results on 1-D systems generalize in a straightforward manner to *separable* multidimensional (M-D) ones, i.e., those realizable by a cascade of systems each of which acts only on one dimension of the input. Without separability, many issues become more complex and are beyond the scope of this thesis.

1.1.2 Filter Banks

Figure 1.3 shows an M channel filter bank (FB). This FB is said to be uniform, i.e., all channels have the same decimation rate. It is also said to be maximally decimated, i.e., the input sampling rate equals the sum of the subband sampling rates. For uniform FBs this means that the decimation rate equals the number of channels. Figure 1.4 shows a possible choice of filters in this FB. With this choice, the k-th channel analysis filter $H_k(e^{j\omega})$ has a support occupying a fraction $1/M$ of the total signal bandwidth 2π. In accordance with the sampling theorem, its output can be decimated by M without losing information. The decimated output is referred to as the k-th subband signal. The synthesis section of the k-th channel interpolates this signal to undo the effect of the decimator. Since the $H_k(e^{j\omega})$ have nonoverlapping supports, summing the outputs from all the channels thus recovers the original FB input. In other words, $y(n) \equiv x(n)$, which is called the perfect reconstruction (PR) property.

The FB can be thought of as a transform between the input and the subband signals. The

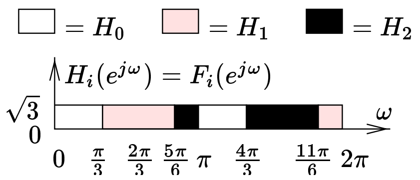


Figure 1.5: Noncontiguous–stacked brickwall filter bank.

transform has the nature of a spectral analyzer. The principle behind the use of FBs is the same as that of any transform domain technique: Various features of interest are displayed more clearly in the transform domain, and hence the transform identifies the areas in which to concentrate the required processing, making it more efficient. For instance, consider the problem of quantization for data compression. The goal is to represent the signal $x(n)$ using an average of b bits per element of the sequence $x(n)$, in a way that minimizes the error between $x(n)$ and its approximation $x_d(n)$ created from this representation. Clearly, the error decreases with increasing b. A direct method uses the uniform quantizer, i.e., sets each $x_d(n)$ to be a b bit approximation of $x(n)$. However, instead of performing this operation on $x(n)$, we can do it on the subbands obtained by analyzing $x(n)$ with a FB. We can then allocate different number of bits b_i to different subbands, keeping the same overall bitrate by ensuring that b is the average of the b_i. By alloting more bits to subbands with more signal energy, we can achieve lower errors than with direct quantization of $x(n)$. This simple principle, implemented with clever schemes to decide the relative importance of the subband coefficients, has resulted in some of the most efficient state of the art image compression algorithms [53],[51].

In the above example, and more generally for most transform domain techniques, it is necessary that the transform itself does not cause any loss of the information in the input. The PR property for a FB means that the transform from the input to the subbands is invertible; the inverse being implemented by the synthesis bank. All FBs in this thesis have PR in absence of additional processing of the subband signals. Fig. 1.4 is by no means the only possible partitioning of the input spectrum that results in a FB with PR: There are infinitely many such partitions, another one is shown in Fig. 1.5 (for $M = 3$). A significant part of this thesis deals with choosing the best FB for a given input and application.

The filters in Figs. 1.4,1.5 are ideal brickwall type, and are unrealizable. Realizable filters are rational, i.e., have transfer functions $H(z)$ that are rational functions of z. The problem of achieving PR with rational filters is solved using the polyphase representation. The M-th order analysis polyphase matrix $\mathbf{E}(z)$ of the filters $H_0(z), H_1(z), \ldots, H_{N-1}(z)$ is defined by

$$\mathbf{h}(z) \stackrel{\triangle}{=} (H_0(z), H_1(z), \ldots, H_{N-1}(z))^T = \mathbf{E}(z^M)\mathbf{d}(z), \quad \text{where} \quad \mathbf{d}(z) = (1, z^{-1}, \ldots, z^{-(M-1)})^T \quad (1.3)$$

is the length M delay vector. Thus, $\mathbf{E}(z)$ is $N \times M$ with i-th column $\big(z^i\mathbf{h}(z)\big)\downarrow_M$. In particular,

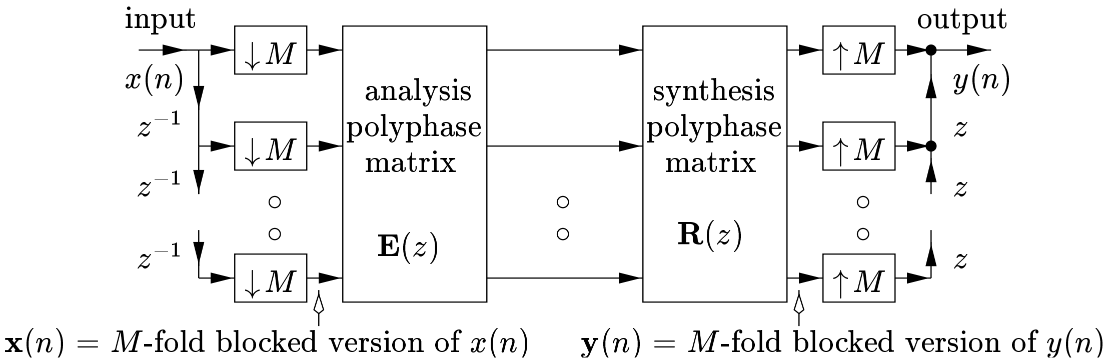


Figure 1.6: Polyphase representation of a uniform FB.

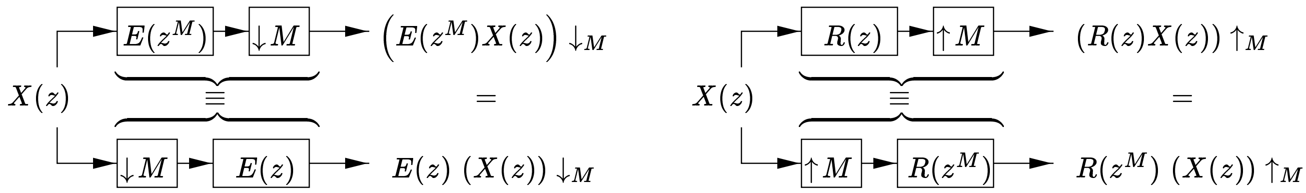


Figure 1.7: The noble identities.

$\mathbf{h}(z)$ could be the vector of analysis filters of the FB of 1.3. These filters can then be implemented using (1.3), i.e., by feeding the input through the delay–chain $\mathbf{d}(z)$ followed by the MIMO system $\mathbf{E}(z^M)$. Since this is followed by M-fold decimation in 1.3, the analysis section of the FB can be redrawn as in Fig. 1.6. Here the decimators have been moved ahead of the MIMO system, which is possible by the noble identities, shown in Fig. 1.7. The synthesis section is similarly redrawn using the analogous definition of the M-th order synthesis polyphase matrix $\mathbf{R}(z)$ of the filters $F_0(z), F_1(z), \ldots, F_{N-1}(z)$: This matrix is such that

$$\mathbf{f}(z) \stackrel{\triangle}{=} (F_0(z), F_1(z), \ldots, F_{N-1}(z)) = \widetilde{\mathbf{d}}(z)\mathbf{R}(z^M), \quad \text{where} \quad \widetilde{\mathbf{d}}(z) = (1, z, \ldots, z^{(M-1)}). \tag{1.4}$$

The polyphase form (Fig. 1.6) is seen to be a simple interposing of two matrices $\mathbf{E}(z)$ and $\mathbf{R}(z)$ between a very elementary PR system called the delay–chain FB. The delay–chain FB simply deinterleaves or 'blocks' the input and then reverses the process. This makes explicit the condition for PR, namely that $\mathbf{R}(z)\mathbf{E}(z)$ is the identity matrix, and allows us to use tools from MIMO LTI system theory in designing PRFBs. If the polyphase matrices $\mathbf{R}(z), \mathbf{E}(z)$ are constant (independent of z), the system operates independently on length M blocks of the input, transforming them using these matrices. Such a system has been variously called a transform coder, block transform, and memoryless transform. FBs are thus seen to be generalizations of such systems. All filtering operations in Fig. 1.6 are performed on signals sampled at a reduced rate as compared to Fig. 1.3. Thus the polyphase form is more efficient to implement.

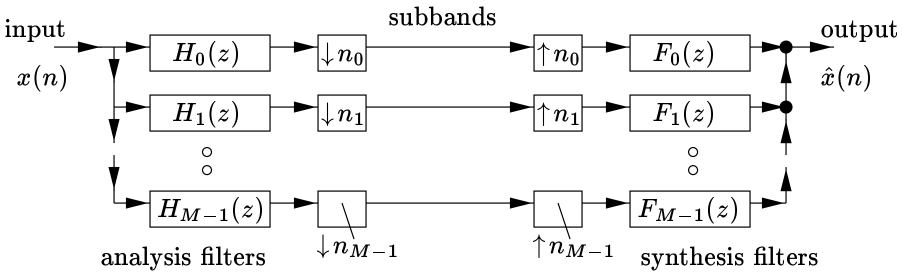


Figure 1.8: General nonuniform filter bank.

1.1.3 Classification of filter banks, and the orthonormality property

Figure 1.8 shows a more general filter bank, where the decimators n_k in different channels may be different. The FB is said to be underdecimated (or redundant), maximally (or critically) decimated, or overdecimated depending on whether $\sum_k (1/n_k)$ exceeds, equals, or is less than unity. This thesis deals mostly with maximally decimated FBs. As mentioned earlier, the FB is said to be uniform if all the decimators n_k are equal, and nonuniform otherwise. We also classify FBs based on the nature of their filters: A FIR FB is one in which all filters are FIR (finite impulse response), and similarly we speak of rational FBs, ideal FBs and unconstrained FBs (where there is no constraint on the filters). A delay–chain FB is one in which all filters are delays. Among uniform M channel PRFBs there is essentially only one delay–chain, i.e., the system in Fig. 1.6 where $\mathbf{E}(z), \mathbf{R}(z)$ are identity matrices. With nonuniform FBs, the situation is more involved, as we will see in Chapter 6.

An important family of (maximally decimated) PRFBs is that of orthonormal FBs. A uniform maximally decimated FB is said to be orthonormal or paraunitary if its analysis polyphase matrix $\mathbf{E}(z)$ is paraunitary, i.e., unitary on the unit circle. Thus, an orthonormal FB is one in which PR is obtained with $\mathbf{R}(e^{j\omega}) = \mathbf{E}^{-1}(e^{j\omega}) = \mathbf{E}^{\dagger}(e^{j\omega})$ (where $(^{\dagger})$ denotes the conjugate transpose). Note that this relation is equivalent to the relation $F_k(e^{j\omega}) = H^*(e^{j\omega})$ between the analysis and synthesis filters (where $(^*)$ denotes the complex conjugate). Many commonly used block transforms such as the DFT and DCT are unitary; paraunitariness is a generalization of the property of being unitary to transforms with memory. Like this property, paraunitariness also is essentially an energy preservation or losslessness property: A MIMO LTI system has a paraunitary transfer matrix if and only if its output energy always equals its input energy. Energy of a (column) vector signal $\mathbf{x}(n)$ here is measured by the usual mean square measure or l_2 norm, as $\sum_n \mathbf{x}^{\dagger}(n)\mathbf{x}(n)$.

Uniform paraunitary FBs are widely used and have been extensively studied. Various results parametrize all such FBs with various properties such as rational, FIR, or linear phase filters. For nonuniform FBs, as the polyphase representation is somewhat more involved as compared to Fig. 1.6, we can define paraunitariness (or orthonormality) as the property of having PR with $F_k(e^{j\omega}) = H^*(e^{j\omega})$. Again, the input energy always equals the sum of the subband energies for an orthonormal FB. Nonuniform orthonormal FBs are usually built by cascading uniform ones in a tree structure. For example, the dyadic wavelet transform is essentially a tree of uniform two channel orthonormal FBs. PRFBs that are not orthonormal are also called biorthogonal. Most of our results on FB optimization focus on optimization over classes of orthonormal FBs.

1.2 Outline and scope of thesis

1.2.1 Optimality of principal component forms

Beginning with Chapter 2, we study the optimization of uniform orthonormal FBs for the given statistics (power spectrum) of their input. This is a unifying theme behind most of this thesis. We show that a type of FB called the **principal component filter bank (PCFB)** is an optimum solution whenever the minimization objective is a concave function of the subband variances produced by the FB. This is one of the central results of this thesis, and shows PCFB optimality for numerous signal processing problems. Some optimality properties of PCFBs, especially for progressive transmission and data compression, have been observed earlier [59],[61],[35],[60]. However, our result provides a unified framework that includes as special cases, all these results as well as many other problems not noticed earlier. Thus, we believe it to be the fundamental explanation for PCFB optimality. We show how the result applies in various white noise reduction systems using Wiener filters or hard thresholds in the subbands. The subsequent chapters apply, extend and generalize the basic result in several directions.

It is important to keep in mind the limitations of the PCFB concept, which also emerge clearly in the course of our study. At the outset, note that the basic concept applies only to orthonormal FBs, and although some results extend to biorthogonal FB optimization, many issues are left open on this front. Even when concerned only with orthonormal FBs, the serious limitations are as follows: (a) PCFBs usually do not exist for many practically implementable FB classes, (b) PCFBs for the unconstrained class (containing *all* M channel orthonormal FBs, including brickwall FBs, etc.) are usually brickwall and unrealizable, and (c) some signal–independent FBs such as the DFT, which have very efficient implementations, often achieve a performance that is asymptotically as good as that of the unconstrained class PCFB in the limit of large number of channels M. Thus, the PCFB is often unattractive from an implementation viewpoint. The importance of the PCFB is that it provides a unified theory for FB optimization, supplies upper bounds on achievable performance, and indicates the direction in which to proceed when it is desired to adapt the FB to its input to improve performance.

1.2.2 Colored noise reduction, PCFB existence issues (Chapter 3)

Here we begin by examining the nature of FB optimization problems of the kind studied in Chapter 2, in situations when PCFBs do not exist. Using the geometry of the optimization search–spaces, we explain exactly why these problems are usually analytically intractable. We show the relation between compaction filter design (i.e., variance maximization) and optimum FBs: A sequential maximization of subband variances produces a PCFB if one exists, but is otherwise suboptimal for several concave objectives. We then make a detailed study of FB optimization and PCFB optimality for *colored* noise suppression. Unlike the case when the noise is white, here the minimization objective is a function of both the signal and the noise subband variances. We show that for the *transform coder* class, if a *common* signal and noise PCFB (KLT) exists, it is optimal for a large class of concave objectives. Common PCFBs for general FB classes have a considerably more restricted optimality, as we show using the class of unconstrained orthonormal FBs. For this class, we also show how to find an optimum FB when the signal and noise spectra are both *piecewise constant* with all discontinuities at rational multiples of π.

1.2.3 PCFB optimality for DMT communication systems (Chapter 4)

Discrete multitone modulation (DMT) is a communication scheme that uses a FB in the 'transmultiplexer' configuration, where the transmitter uses a synthesis FB to combine several data streams and send them on a single channel. In most technologies that use the scheme, the FB used is the DFT. We review the basics of the system, and then formulate the problem of maximizing the transmitted bitrate for a given power, or minimizing the transmission power for a given bitrate, given the acceptable error probability. We show that PCFBs are the best orthonormal FBs for this problem. This is again due to the same general result of Chapter 2, owing to the concavity of the relevant objective. We show a simulation example comparing the performance of various FBs on the ADSL (asymmetric digital subscriber line) channel which uses DMT transmission.

1.2.4 Nonuniform FBs: Optimization and PCFBs (Chapter 5)

Here we extend the notion of PCFBs to classes of nonuniform orthonormal FBs having a fixed set of decimation rates. The nonuniform PCFBs we define generalize the uniform ones by being optimal for a certain family of concave objectives. The optimality is however somewhat more restricted than that for uniform PCFBs. We then study existence of nonuniform PCFBs, showing the following result that is important in this connection: For strictly monotone input power spectra, PCFBs do not exist for the class of unconstrained nonuniform FBs with any given set of decimators that are not all equal. In contrast, the class of unconstrained uniform M channel orthonormal FBs always has a PCFB for any input spectrum. Thus PCFB existence for nonuniform FB classes is much more delicate than that for uniform ones.

1.2.5 Parameterization of nonuniform PRFBs (Chapter 6)

This chapter differs in theme from the earlier ones, in that it does not study optimization of FBs, but the more basic question of their design. In Chapter 5 we studied optimization of nonuniform orthonormal FBs, which are special cases of (nonuniform) FBs with the PR property. Here we ask for a characterization of various classes of FBs having PR, i.e., a 'listing' or parameterization of all FBs in the class. Several results of this kind have been derived for *uniform* FBs, using tools from MIMO system theory, for various classes such as the class of FIR FBs, rational FBs, orthonormal FIR FBs, linear phase FBs etc. Usually these results factorize the polyphase matrix of the FB into a product of elementary matrices each fully characterized by a simple parameter such as an angle or a unit norm vector. Examples are factorizations based on Givens rotations and Householder matrices [64] for paraunitary FBs; and using continued fractions [34] and ladder structures or 'lifting steps' [14],[20] for biorthogonal FBs. However, it is much harder to obtain such parameterizations for nonuniform FBs with a given set of decimators. In fact, we do not even know how to determine whether or not a set S of decimators can be used to build a rational PRFB. There are some obvious sufficient conditions on S that allow building of such a FB, e.g., the condition that all entries in S be equal (so that the FB is uniform). There are also certain conditions that are known to be necessary for existence of such a FB (e.g., no two entries of S can be coprime). However conditions that are both necessary and sufficient are not known. In this chapter we try to find such conditions. The problem in its full generality remains unsolved, but we make an important first step by considerably strengthening many of the known conditions.

1.3 Notation

In general, notation in this thesis closely parallels [64]. Superscripts $(^*)$ and $(^T)$ denote the complex conjugate and matrix (or vector) transpose respectively, while superscript dagger $(^\dagger)$ denotes the conjugate transpose. Boldface letters are used for matrices and vectors. Lower case letters are used for discrete sequences while upper case letters are used for Fourier and z–transforms. Occasionally lowercase boldface letters are used for vector z–transforms. For sequences $\mathbf{h}(n)$ without z–transforms that are rational functions of z, the notation $\mathbf{H}(z)$ is an abbreviation for the Fourier transform $\mathbf{H}(e^{j\omega})$. For LTI transfer matrices $\mathbf{H}(z)$, the 'paraconjugate' $\mathbf{H}^\dagger(1/z^*)$ is denoted by $\widetilde{\mathbf{H}}(z)$; thus $\widetilde{\mathbf{H}}(e^{j\omega}) = \mathbf{H}^\dagger(e^{j\omega})$, and $\mathbf{H}(z)$ is paraunitary iff $\widetilde{\mathbf{H}}(z)\mathbf{H}(z)$ is the identity matrix. Equations (1.1),(1.2) establish notation for decimators and expanders. The L-th root of unity, $e^{-j2\pi/L}$ is denoted by W_L, or by W if the subscript value L is understood. The Kronecker delta function is denoted by δ ($\delta(0) = 1$ and $\delta(x) = 0$ if $x \neq 0$). $\mathcal{R}^M$ denotes the set of M-tuples of real numbers, and $\mathcal{R}_+^M$ denotes that of M-tuples of non-negative real numbers. We denote by $\mathrm{diag}(\mathbf{A})$ the column vector consisting of the diagonal entries of the square matrix $\mathbf{A}$. The convex hull of a set D is denoted by $\mathrm{co}(D)$. The Cartesian product of two sets A, B is denoted by $A \times B$.

Chapter 2 The optimality of principal component filter banks

The problem of optimization of filter banks has been addressed by several authors and many interesting results have been reported in the last five years. Yet there are a number of optimization problems that have not hitherto been addressed. In this chapter, we provide a general framework for the optimization of uniform orthonormal filter banks for given input statistics; which includes many of the known results as special cases. It also produces solutions to a number of problems that have been regarded as difficult or not considered earlier.

A generic signal processing scheme using an M channel uniform perfect reconstruction filter bank (FB) (realized in polyphase form) is shown in Fig. 2.1. The input vector $\mathbf{x}(n)$ is the M-fold blocked version of the scalar input $x(n)$. We assume that $\mathbf{x}(n)$ is a zero mean wide sense stationary (WSS) random process with a given power spectral density (psd) matrix $\mathbf{S_{xx}}(e^{j\omega})$. We are also given a class $\mathcal{C}$ of **orthonormal uniform M channel FBs**. Examples are the class of FBs in which all filters are FIR with a given bound on their order, or the class of unconstrained FBs (where there are no constraints on the filters besides those imposed by orthonormality). The problem this paper is concerned with is that of **finding the best FB from $\mathcal{C}$** for the given input statistics $\mathbf{S_{xx}}(e^{j\omega})$, for use in the system of Fig. 2.1. By 'best FB' we mean one that minimizes a well-defined objective function over the class $\mathcal{C}$. To formulate this objective, we need to describe the purpose or application of the FB in Fig. 2.1, and the nature of the subband processors P_i. This is done in detail in Section 2.3 in a general setting.

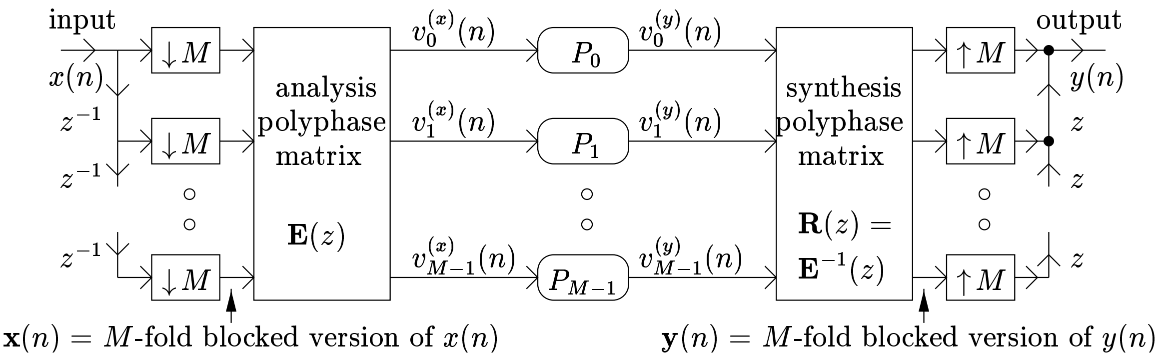


Figure 2.1: Generic FB based signal processing scheme.

2.1 Relevant earlier work

Consider in particular the case where the P_i are quantizers for signal compression. We use the model of [35] which replaces the quantizer P_i by additive noise of variance $f_i(b_i)\sigma_i^2$. Here b_i is the number of bits alloted to the quantizer, σ_i^2 is its input variance; and f_i is the normalized quantizer function, assumed not to depend on the input statistics. If all quantization noises are jointly stationary random processes, we can show that the overall mean square reconstruction error (which is the minimization objective here) is $g = \sum_{i=0}^{M-1} \frac{1}{M} f_i(b_i)\sigma_i^2$. Kirac and Vaidyanathan show [35] that for any given bit allocation b_i (not necessarily optimum), the best FB for this problem is a **principal component filter bank (PCFB)** for the given class $\mathcal{C}$ and input psd $\mathbf{S}_{\mathbf{xx}}(e^{j\omega})$.

The concept of a PCFB is reviewed in Section 2.4.2. PCFBs provide optimum progressive representations of their input, i.e., they minimize the mean-square error caused by reconstruction after dropping the P weakest (lowest variance) subbands, for any P. PCFBs for certain classes of FBs have been studied earlier. For example, let $\mathcal{C}^t$ denote the class of all M channel orthogonal transform coders, i.e., FBs as in Fig. 2.1 where $\mathbf{E}(z)$ is a constant unitary matrix $\mathbf{T}$. The **Karhunen Loeve transform (KLT)** for the input $\mathbf{x}(n)$ is the transform $\mathbf{T}$ that diagonalizes the autocorrelation matrix of $\mathbf{x}(n)$. It has been well known [30] that the KLT is a PCFB for $\mathcal{C}^t$. For the class $\mathcal{C}^u$ of *all* (unconstrained) orthonormal M channel FBs, construction of the PCFB has been studied by Tsatsanis and Giannakis [59], and independently by Vaidyanathan [61]. The goal of [61] was coding gain maximization for compression under the high bitrate quantizer noise model with optimum bit allocation. This model is in fact a special case of the one described earlier, where $f_i(b_i) = c2^{-2b_i}$. In another work on PCFBs [60], Unser correctly conjectures their optimality for another family of objective functions, of the form $g = \sum_{i=0}^{M-1} h(\sigma_i^2)$ where h is any concave function. (This does not include the earlier objective, since the $f_i(b_i)$ depended on the subband index i.) For this family, optimality has been proved by Mallat [42, Th. 9.8, pg. 398] using a theorem of Hardy et al. In this chapter we consider the more general form $g = \sum_{i=0}^{M-1} h_i(\sigma_i^2)$ where h_i are *possibly different* concave functions. We show optimality of PCFBs for all these objectives. This covers a wider class of applications, as shown in Section 2.7. It includes the conjecture of [60] (proved in [42]) as a special case where $h_i \equiv h$ for all i. It also includes the minimization objective of [35] as a special case when $h_i(x) = f_i(b_i)x$ for all i.

Filter bank design for quantization error minimization has also been studied by Moulin et al. [45],[46]. The earlier stated form $g = \sum_{i=0}^{M-1} \frac{1}{M} f_i(b_i)\sigma_i^2$ of the error requires modification for *biorthogonal* FBs. In an important contribution, [46], Moulin et al. study the minimization of this modified objective over the class of *all* (unconstrained) biorthogonal FBs, for a broad class of $f_i(b_i)$. The authors examine the role of the properties of the PCFB for the unconstrained *orthonormal* FB class $\mathcal{C}^u$ in this problem. It is also claimed that pre and post filters around such a PCFB yield the optimal solution. In [45], an algorithm is proposed for PCFB design for a certain class of FIR orthonormal FBs. It involves a compaction filter design followed by a KLT matrix completion, and will produce the PCFB (which is known to maximize coding gain) *if it exists.* However, it is shown numerically that the designed filters do not always optimize the coding gain (thus showing that in fact the PCFB does not exist). This chapter studies the geometric structure of the optimization search space, and thereby reveals several new optimality properties of PCFBs, especially those connected with noise reduction. The content is drawn mainly from [2], and has been presented in preliminary form in [4],[6].

2.2 Main aims and outline of chapter

This chapter points out a strong connection between orthonormal FB optimization and the principal component property. The main message is as follows: Let σ_i^2 denote the variance of the i-th subband signal. To every FB in the given class $\mathcal{C}$ there then corresponds a set of subband variances σ_i^2. The **PCFB for** $\mathcal{C}$, if it exists, is the **optimum FB in** $\mathcal{C}$ for **all problems** in which the minimization objective can be expressed as a **concave function** of the subband variance vector $\mathbf{v} \stackrel{\triangle}{=} (\sigma_0^2, \sigma_1^2, \ldots, \sigma_{M-1}^2)^T$.

This result has its grounding in majorization and convexity theory, and will be elaborated in detail in later sections. It shows PCFB optimality for all objectives of the form $g = \sum_{i=0}^{M-1} h_i(\sigma_i^2)$ where h_i are any concave functions. For orthonormal FBs, this general form includes as special cases, all the objectives mentioned earlier. We show how such concave objectives arise in many other situations besides coding gain maximization, especially those connected with white noise suppression. Suppose the FB input is a signal buried in noise, and the system of Fig. 2.1 aims to improve the signal to noise ratio. Let each subband processor P_i be a zeroth order Wiener filter. We show that under suitable assumptions on the signal and noise statistics, the problem of FB optimization for such a scheme reduces to the minimization of a concave function of the subband variance vector. So PCFBs if they exist are optimal for such a scheme. PCFB optimality continues to hold even with certain other types of subband processors for noise reduction, e.g., constant multipliers and hard thresholds (defined in Section 2.7).

Thus we have a general problem formulation (Section 2.3) and a unified theory of optimal FBs (Section 2.4), which simultaneously explains the optimality of PCFBs for progressive transmission (Section 2.4.2), compression (Section 2.5.3) and noise suppression (Section 2.7). Known results on PCFB existence are reviewed in Sections 2.5.1 and 2.5.2, and Section 2.6 shows that the classes of DFT and cosine-modulated FBs do not have PCFBs. Section 2.7 also proves certain extensions of the basic PCFB optimality result to biorthogonal FB optimization and *colored* noise suppression. A detailed analysis of the noise reduction problem when the noise is colored is consigned to Chapter 3.

2.3 Problem formulation

We are given a class $\mathcal{C}$ of **orthonormal uniform M channel FBs**. Recall that a FB is fully specified by its analysis polyphase matrix $\mathbf{E}(z)$, or alternatively by the ordered set of analysis and synthesis filter pairs $(H_k(z), F_k(z))$, $k = 0, 1, \ldots, M-1$ (see Figs. 2.1,1.3). We are also given an ordered set of M subband processors P_i, $i = 0, 1, \ldots, M-1$, where P_i denotes the processor acting on the i-th subband. Specific instances of such P_i will be discussed in later sections; in general each P_i is simply a function that maps input sequences to output sequences. The specification of this function may or may not depend on the input statistics.

The system of Fig. 2.1 is built using a FB in $\mathcal{C}$ and the processors P_i. In all problems that we consider, this system is aimed at producing a certain **desired signal** $d(n)$ at the FB output. For example, in context of compression, the processors P_i are quantizers and the desired output equals the input, i.e., $d(n) = x(n)$. In context of noise reduction, the input $x(n) = s(n) + \mu(n)$ where $\mu(n)$ is additive noise, the desired output $d(n) = s(n)$ (the pure signal), and the P_i could for instance be Wiener filters. The FB optimization problem involves finding among all FBs in $\mathcal{C}$, the one minimizing some measure of the error signal

$$e(n) \stackrel{\triangle}{=} d(n) - y(n),$$

where $y(n)$ is the true FB output. To formulate the error measure, we impose random process models on the FB input $x(n)$ and desired signal $d(n)$. We assume that $\mathbf{x}(n)$, the M-fold blocked version of $x(n)$ (see Fig. 2.1) is a WSS vector process with given psd matrix $\mathbf{S_{xx}}(e^{j\omega})$. Equivalently, $x(n)$ is CWSS(M), i.e., wide sense cyclostationary with M as period.[1] All processes are assumed to be zero mean unless otherwise stated. In all our problems, the $d(n)$ and the P_i are such that the error $e(n)$ is also a zero mean CWSS(M) random process. Thus we choose as error measure, the variance of $e(n)$ averaged over its period of cyclostationarity M.

We denote by $v_i^{(x)}(n)$ the i-th subband signal produced when the FB input is the scalar signal $x(n)$ (as in Fig. 2.1). If the error $e(n)$ is CWSS(M), then $v_i^{(e)}(n)$, $i = 0, 1, \ldots, M-1$ are jointly WSS; and orthonormality of the FB can be used to show that the above–mentioned error measure equals

$$\varepsilon = \frac{1}{M}\sum_{i=0}^{M-1} E[|v_i^{(e)}(n)|^2], \quad \text{where} \quad v_i^{(e)}(n) = v_i^{(d)}(n) - v_i^{(y)}(n). \tag{2.1}$$

Thus the processor P_i must try to produce an output 'as close to' $v_i^{(d)}(n)$ as possible, in the sense

[1] In particular, $x(n)$ could be a WSS process with given power spectrum $S(e^{j\omega})$. In this case $\mathbf{S_{xx}}(e^{j\omega})$ is fully determined from $S(e^{j\omega})$, and has the special property of being pseudocirculant[64],[52].

of minimizing $E[|v_i^{(e)}(n)|^2]$. In many situations to be discussed in detail later, the P_i are such that

$$E[|v_i^{(e)}(n)|^2] = h_i(\sigma_i^2). \tag{2.2}$$

Here $\sigma_i^2 = E[|v_i^{(x)}(n)|^2]$ denotes the variance of $v_i^{(x)}(n)$; and h_i is some function whose specification depends only on the nature of the processor P_i and not on the choice of FB from $\mathcal{C}$. Thus for such problems, with $\mathbf{v} = (\sigma_0^2, \sigma_1^2, \ldots, \sigma_{M-1}^2)^T$ denoting the subband variance vector, the objective defined on the class $\mathcal{C}$ becomes

$$g(\mathbf{v}) = \frac{1}{M}\sum_{i=0}^{M-1} h_i(\sigma_i^2). \tag{2.3}$$

Hence the minimization **objective** is **purely a function of the subband variance vector**. This function g of (2.3) is fully specified given the description of the processors P_i. Let $\mathcal{S}$ denote the set of all subband variance vectors corresponding to all FBs in $\mathcal{C}$. The optimization problem thus reduces to that of finding the minima of the real-valued function g on the set $\mathcal{S}$. We will hence refer to $\mathcal{S}$ as the optimization **search space**.

In later sections we show that for a number of FB based signal processing schemes, the above formulation holds and further the objective g is a **concave** function (Section 2.4.1). The central result of the present paper, described in detail in Section 2.4, is that a principal component filter bank (PCFB) is optimal for all such problems where g is concave. The main reason for this is that whenever a PCFB exists, the search space $\mathcal{S}$ has a very special structure: its **convex hull is a polytope** (Section 2.4.1). Since the set $\mathcal{S}$ plays an important role in the further discussion, we summarize the main definitions and facts pertaining to it:

2.3.1 Summary of definitions and facts related to the search space

1. **Definition.** For each FB in the given class $\mathcal{C}$, the **subband variance vector** associated with the input random process $x(n)$ is defined as the vector $\mathbf{v} = (\sigma_0^2, \sigma_1^2, \ldots, \sigma_{M-1}^2)^T$, where σ_i^2 is the variance of the process $v_i^{(x)}(n)$. Here $v_i^{(x)}(n)$ is the i-th subband signal produced by feeding $x(n)$ as the FB input.

2. **Computing the variance vector.** Given the FB analysis polyphase matrix $\mathbf{E}(z)$ and the psd matrix $\mathbf{S_{xx}}(e^{j\omega})$ of the vector input $\mathbf{x}(n)$ in Fig. 2.1, the process $(v_0^{(x)}(n), v_1^{(x)}(n), \ldots, v_{M-1}^{(x)}(n))^T$ has psd matrix $\mathbf{E}(e^{j\omega})\mathbf{S_{xx}}(e^{j\omega})\mathbf{E}^\dagger(e^{j\omega})$. Thus the subband variance vector is

$$\mathbf{v} = \frac{1}{2\pi}\int_0^{2\pi} \operatorname{diag}\left(\mathbf{E}(e^{j\omega})\mathbf{S_{xx}}(e^{j\omega})\mathbf{E}^\dagger(e^{j\omega})\right)\, d\omega. \tag{2.4}$$

3. The optimization **search space** is defined as the set $\mathcal{S}$ of all subband variance vectors corresponding to all FBs in the given class $\mathcal{C}$. So $\mathcal{S}$ is fully specified given the class $\mathcal{C}$ and

the input statistics $\mathbf{S_{xx}}(e^{j\omega})$. All entries of any vector in $\mathcal{S}$ are clearly non-negative. Thus $\mathcal{S} \subset \mathcal{R}_+^M \subset \mathcal{R}^M$.

4. The set $\mathcal{S}$ is **bounded** and **lies entirely on an $M-1$ dimensional hyperplane in $\mathcal{R}^M$**. This follows from (2.4) using the fact that $\mathbf{E}(e^{j\omega})$ is unitary for all ω (orthonormality of the FB): No matter what the class $\mathcal{C}$, there is always an upper bound (depending only on $\mathbf{S_{xx}}(e^{j\omega})$) on all entries of all vectors $\mathbf{v} \in \mathcal{S}$. Thus $\mathcal{S}$ is bounded. Also, the sum of the entries of $\mathbf{v}$ is the same for all $\mathbf{v} \in \mathcal{S}$, i.e., it is the trace of the matrix $\frac{1}{2\pi}\int_0^{2\pi} \mathbf{S_{xx}}(e^{j\omega})\,d\omega$. So $\mathcal{S}$ lies on an $M-1$ dimensional hyperplane in $\mathcal{R}^M$.

5. **Permutation symmetry of $\mathcal{S}$.** A FB is defined by an *ordered* set of analysis and synthesis filters. So a change of this ordering (or equivalently, interchanging of rows of the analysis polyphase matrix) technically produces a different FB, which we will refer to as a *permutation* of the original FB. However, clearly all permutations of a uniform FB are essentially the same, i.e., equally easy to implement. So we make the following very reasonable assumption on the given class $\mathcal{C}$ of FBs: Any permutation of any FB in $\mathcal{C}$ is also in $\mathcal{C}$. This assumption holds for all specific classes $\mathcal{C}$ that we will encounter. Note that if two FBs are permutations of each other, then so are their subband variance vectors; however the minimization objective may attain different values at these vectors. Thus we use the convention of defining a FB as an *ordered* set of filter pairs, because the ordering affects the objective.

2.4 The optimality of PCFBs

We now show that principal component filter banks (PCFBs) are optimal whenever the objective function to be minimized is concave on the optimization search space $\mathcal{S}$. The proof follows from strong connections between the notion of a PCFB and certain results in convexity and majorization theory reviewed in Section 2.4.1. PCFBs are defined and described in Section 2.4.2. In Section 2.4.3 we show the connection between PCFBs and special convex sets called polytopes, and thereby prove the main result of the paper.

2.4.1 Convexity theory [50]

Convex sets: A set $D \subset \mathcal{R}^M$ is defined to be convex if $\mathbf{x}, \mathbf{y} \in D$ implies $\mu\mathbf{x} + (1-\mu)\mathbf{y} \in D$ whenever $0 \le \mu \le 1$. Geometrically, D is convex if any line segment with endpoints in D lies wholly in D; see Fig. 2.2. A **convex combination** of a finite set of vectors $\mathbf{x}_i, i = 1, 2, \ldots, N$ is by definition a vector of the form $\sum_{i=1}^N \alpha_i \mathbf{x}_i$ with $0 \le \alpha_i \le 1$ and $\sum_{i=1}^N \alpha_i = 1$. Thus by definition, D is convex if any convex combination of any *pair* (or equivalently by induction, any *finite set*) of elements of D lies in D.

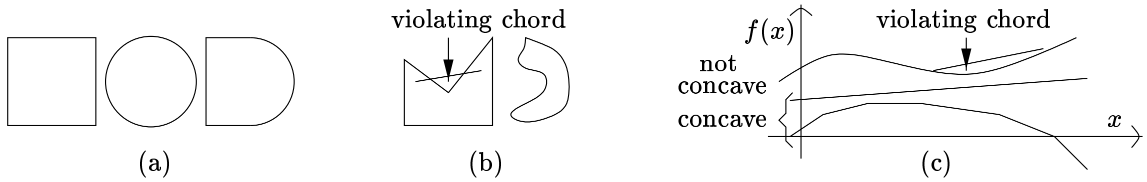


Figure 2.2: Convex sets and concave functions. (a) Convex sets, (b) Nonconvex sets, (c) Concave functions of one variable.

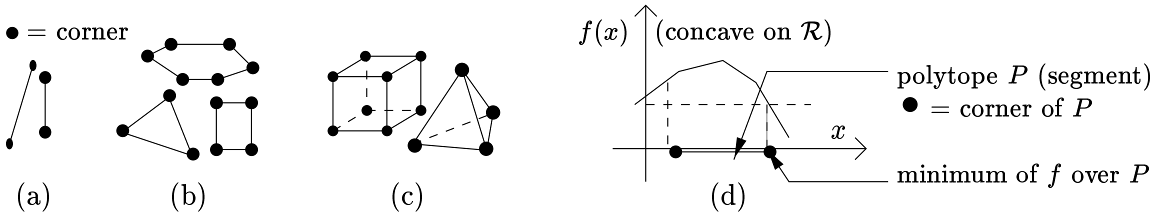


Figure 2.3: M-dimensional polytopes, their extreme points and their optimality. (a) $M = 1$, (b) $M = 2$, (c) $M = 3$, (d) Optimality of extreme points.

Concave functions: Let f be a real-valued function defined on a convex set $D \subset \mathcal{R}^M$. The function f is defined to be concave on the domain D if given any elements $\mathbf{x}, \mathbf{y}$ in D,

$$f(\mu\mathbf{x} + (1-\mu)\mathbf{y}) \geq \mu f(\mathbf{x}) + (1-\mu)f(\mathbf{y}) \quad \text{whenever } 0 \leq \mu \leq 1. \tag{2.5}$$

Graphically, this means that the function f is always above its chord; see Fig. 2.2c. The domain D of f has to be convex to ensure that the argument of f on the left side of (2.5) is in D, i.e., to ensure that the above definition makes sense. For a concave function f, we can use (2.5) to show by induction that for any $\mathbf{x}_i \in D$,

$$f(\sum_{i=1}^{N} \alpha_i \mathbf{x}_i) \geq \sum_{i=1}^{N} \alpha_i f(\mathbf{x}_i) \quad \text{whenever} \quad 0 \leq \alpha_i \leq 1 \;\text{ and }\; \sum_{i=1}^{N} \alpha_i = 1. \tag{2.6}$$

This is known as **Jensen's** inequality. The function f is said to be **strictly concave** if it is concave and further if equality in (2.5) is achieved for distinct $\mathbf{x}, \mathbf{y}$ iff μ is either 0 or 1. For such f, equality is achieved in (2.6) for distinct $\mathbf{x}_i$ iff one of the α_i is unity (and hence all the others are zero).

Convex hulls: The convex hull of a set $D \subset \mathcal{R}^M$ is denoted by $\text{co}(D)$ and is defined as the set of all possible convex combinations of elements of D. Equivalently, it can be defined as the 'smallest' (i.e., *minimal*) convex set containing D, or the intersection of all convex sets containing D. Thus, $D = \text{co}(D)$ iff D is a convex set.

Polytopes: A convex polytope is defined as the convex hull of a finite set. If $E \subset \mathcal{R}^M$ is finite, $P \triangleq \text{co}(E)$ is a polytope. We can assume that no vector in E is a convex combination of other vectors of E, as deleting such vectors from E does not change P. With this condition, the polytope P is said to be generated by the elements of E, and these elements are called the **extreme points** (or vertices, or corners) of P; see Figs. 2.3a,b,c. Notice that an extreme point $\mathbf{v}$ of P cannot be expressed as a nontrivial convex combination of points of P, i.e., $\mathbf{v} = \sum_{i=1}^{J} \alpha_i \mathbf{x}_i$ for $\mathbf{x}_i \in P$ and $0 < \alpha_i < 1$, $\sum_{i=1}^{J} \alpha_i = 1$ implies $\mathbf{x}_1 = \ldots = \mathbf{x}_J$ $(= \mathbf{v})$. The following result on extreme points, illustrated by Fig. 2.3d, is vital in explaining PCFB optimality.

Theorem 2.1: Optimality of extreme points of polytopes. Let a function f have a convex polytope P as domain. If f is concave on P, at least one extreme point of P achieves the minimum of f over P. Further if f is strictly concave, its minimum over P is necessarily at an extreme point of P. ■

Proof : Let E be the set of extreme points of P. Thus E is finite and $P = \text{co}(E)$. Let $E = \{\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_N\}$ and let $\mathbf{v}_j \in E$ attain the minimum of f over the finite set E. Now by definition of a polytope, for any $\mathbf{v} \in P$ we have $\mathbf{v} = \sum_{i=1}^{N} \alpha_i \mathbf{v}_i$ for some α_i such that $0 \leq \alpha_i \leq 1$ and $\sum_{i=1}^{N} \alpha_i = 1$. Thus,

$$\begin{aligned} f(\mathbf{v}) = f(\sum_{i=1}^{N} \alpha_i \mathbf{v}_i) \quad &\geq \quad \sum_{i=1}^{N} \alpha_i f(\mathbf{v}_i) && \text{(by (2.6), i.e., Jensen's inequality)} \\ &\geq \quad \sum_{i=1}^{N} \alpha_i f(\mathbf{v}_j) = f(\mathbf{v}_j) && \text{(by definition of } \mathbf{v}_j \text{ and using } \sum_{i=1}^{N} \alpha_i = 1). \end{aligned}$$

Thus, $f(\mathbf{v}) \geq f(\mathbf{v}_j)$, i.e., the extreme point $\mathbf{v}_j$ of P attains the minimum of f over P. Further, the $\mathbf{v}_i$ are distinct, so if f is strictly concave then Jensen's inequality becomes strict unless one of the α_i is unity. Thus in this case the minimum is necessarily at an extreme point of P. $\bigtriangledown \bigtriangledown \bigtriangledown$

2.4.2 PCFBs and majorization: Definitions and properties

Definition: Majorization. Let $A = \{a_0, a_1, \ldots, a_{M-1}\}$ and $B = \{b_0, b_1, \ldots, b_{M-1}\}$ be two sets each having M real numbers (not necessarily distinct). The set A is defined to **majorize** the set B if the elements of these sets, ordered so that $a_0 \geq a_1 \geq \ldots \geq a_{M-1}$ and $b_0 \geq b_1 \geq \ldots \geq b_{M-1}$, obey the property that

$$\sum_{i=0}^{P} a_i \geq \sum_{i=0}^{P} b_i \qquad \text{for all } P = 0, 1, \ldots, M-1, \text{ with equality holding when } P = M-1. \tag{2.7}$$

Given two vectors $\mathbf{v}_1, \mathbf{v}_2$ in $\mathcal{R}^M$, we will say that $\mathbf{v}_1$ majorizes $\mathbf{v}_2$ when the set of entries of $\mathbf{v}_1$ majorizes that of $\mathbf{v}_2$. Evidently in this case any permutation of $\mathbf{v}_1$ majorizes any permutation of $\mathbf{v}_2$.

Definition: PCFBs. Let $\mathcal{C}$ be the given class of orthonormal uniform M channel FBs, and let $\mathbf{S}_{\mathbf{xx}}(e^{j\omega})$ be the power-spectrum matrix of the vector process input $\mathbf{x}(n)$ (shown in Fig. 2.1). A FB in $\mathcal{C}$ is said to be a **principal component filter bank (PCFB)** for the class $\mathcal{C}$ for the input psd $\mathbf{S_{xx}}(e^{j\omega})$, if its subband variance vector (defined in Section 2.3.1) majorizes the subband variance vector of every FB in the class $\mathcal{C}$.

Remarks on the PCFB definition

1. **PCFB optimality for progressive transmission**. In Fig. 1.3, suppose the FB has subbands numbered in decreasing order of their variances σ_i^2, i.e., $\sigma_0^2 \geq \sigma_1^2 \geq \ldots \geq \sigma_{M-1}^2$, and the P_i are constant multipliers m_i :

$$m_i = \begin{cases} 1 & \text{for } 0 \leq i \leq P-1 \\ 0 & \text{for } P \leq i \leq M-1 \end{cases}$$

for a fixed integer P $(0 \leq P \leq M)$. This system keeps the P strongest (largest variance) subbands, discarding the others. Due to FB orthonormality, the expected mean square error between the output and input is then

$$\varepsilon(P) = \frac{1}{M}\sum_{i=P}^{M-1}\sigma_i^2 = \frac{1}{M}(c - \sum_{i=0}^{P-1}\sigma_i^2), \quad \text{where} \quad c = \sum_{i=0}^{M-1}\sigma_i^2,$$

and c is the same for all orthonormal FBs. Thus, from the definitions of PCFBs and majorization it follows that a PCFB, if it exists, minimizes this error $\varepsilon(P)$ for all P. In fact this property is the origin of the concept of a PCFB [59], and is clearly equivalent to its definition. PCFBs are also optimal for many other problems, as Section 2.4.3 will show.

2. **Existence of PCFB**. Given the class $\mathcal{C}$ of FBs and the input power spectrum $\mathbf{S_{xx}}(e^{j\omega})$, a PCFB for $\mathcal{C}$ may not always exist. The PCFB and its existence depends on both $\mathcal{C}$ and $\mathbf{S_{xx}}(e^{j\omega})$. For example for white input ($\mathbf{S_{xx}}(e^{j\omega})$ = identity matrix), PCFBs always exist: In fact, all FBs in $\mathcal{C}$ are PCFBs, no matter what $\mathcal{C}$ is. Section 2.5 studies certain classes $\mathcal{C}$ for which PCFBs always exist for any input psd $\mathbf{S_{xx}}(e^{j\omega})$ (of course the PCFB will depend on $\mathbf{S_{xx}}(e^{j\omega})$). Section 2.6 studies certain classes $\mathcal{C}$ for which PCFBs do not exist for large families of input spectra.[2]

3. **Nonuniqueness of PCFB**. From the definition of majorization, any permutation of a PCFB is also a PCFB. Further it is possible that two FBs which are not permutations of each other are both PCFBs, i.e., the PCFB need not be unique. However, all PCFBs must have the same subband variance vector up to permutation. This is because **two sets majorizing each other must be identical** – a direct consequence of the definition of majorization. As all our

[2] A question of possible interest is as follows: Given a class $\mathcal{C}$, find all *non-white* input spectra for which a PCFB for $\mathcal{C}$ exists.

FB optimizations involve not the actual FB but only its subband variance vector, we often speak of *the* PCFB even though it may not be unique.

2.4.3 Principal components, convex polytopes and PCFB optimality

Let $\mathcal{C}$ be the given class of orthonormal uniform M channel FBs, and $\mathbf{S_{xx}}(e^{j\omega})$ the psd matrix of the vector input $\mathbf{x}(n)$ of Fig. 2.1. Let $\mathcal{S}$ be the set of all subband variance vectors of all FBs in $\mathcal{C}$ for input $\mathbf{x}(n)$. We have

Theorem 2.2: PCFBs and convex polytopes. A PCFB for the class $\mathcal{C}$ for input psd $\mathbf{S_{xx}}(e^{j\omega})$ exists if and only if the convex hull co($\mathcal{S}$) is a polytope whose extreme points consist of all permutations of a single vector $\mathbf{v}_*$. Under this condition, $\mathbf{v}_*$ is the subband variance vector produced by the PCFB. ■

Theorem 2.3: Optimality of PCFBs. The PCFB for the class $\mathcal{C}$ (if it exists) is the optimum FB in $\mathcal{C}$ whenever the minimization objective is a concave function on the domain co($\mathcal{S}$). Further if this function is strictly concave, the optimum FB is necessarily a PCFB. ■

Theorem 2.3 follows directly from Theorem 2.2 (proved in Section 2.4.4) and Theorem 2.1 of Section 2.4.1. Note that the FB optimization involves choosing the best vector from $\mathcal{S}$, but Theorem 2.1 is used here to find the best vector from co($\mathcal{S}$) $\supset \mathcal{S}$. However, Theorem 2.2 shows that the best vector from co($\mathcal{S}$) in fact lies in $\mathcal{S}$ (and corresponds to the PCFB). Hence it must be optimum over $\mathcal{S}$. Also note that all permutations of a PCFB are PCFBs, and the above theorems do not specify which of these is the optimum. In general they are not all equally good, but as they are finitely many, it is easy to pick the best one. For the objective (2.3), if all h_i are identical then all permutations are equally good, while if $h_i(x) = k_i x$ for all i then we assign the largest σ_i^2 to the least k_i and so on. More generally, finding the best permutation of the PCFB is an instance of the *assignment problem*, well studied in operations research literature [12].

Theorem 2.3 shows optimality of PCFBs for a number of signal processing problems. In Section 2.3, we had a general formulation of the FB optimization problem such that the minimization objective g was purely a function of the subband variance vector, as in (2.3). If the functions h_i in (2.3) are all concave on $\mathcal{R}_+$, then g is concave on the domain co($\mathcal{S}$). This happens in several problems as we will see in later sections. Thus, Theorem 2.3 shows PCFB optimality for all these problems.

2.4.4 Proving the main result

This section aims to prove Theorem 2.2, and hence, the main PCFB optimality result (Theorem 2.3). To do this, we first review some results on majorization theory [29].

Relevant definitions from majorization theory

1. A **doubly stochastic matrix Q** is a square matrix with non-negative real entries q_{ij} satisfying $\sum_i q_{ij} = 1$, $\sum_j q_{ij} = 1$; i.e., the sum of the entries in any row or column of **Q** is unity. All convex combinations and products of $M \times M$ doubly stochastic matrices are also doubly stochastic (Appendix A).

2. **Permutation matrices** are square matrices obtained by permuting rows (or columns) of the identity matrix. Thus they are doubly stochastic. In fact they are the only *unitary* doubly stochastic matrices. (This is because $\sum_{i=1}^{M} p_i = \sum_{i=1}^{M} p_i^2 = 1$ for non-negative p_i iff all but one of the p_i are zero.)

3. An **orthostochastic matrix Q** is one that can be obtained from a unitary matrix **U** by replacing each element u_{ij} by $q_{ij} = |u_{ij}|^2$. We will refer to **Q** as the orthostochastic matrix corresponding to the unitary matrix **U**. Since $\sum_i |u_{ij}|^2 = \sum_j |u_{ij}|^2 = 1$ for unitary **U**, every $M \times M$ orthostochastic matrix is doubly stochastic. The converse is true if $M \leq 2$ but is false if $M > 2$ (Appendix B).

Relevant results from majorization theory

1. **Majorization theorem** [27],[29]: If $\mathbf{a}, \mathbf{b} \in \mathcal{R}^M$, **a** majorizes **b** iff $\mathbf{b} = \mathbf{Q}\mathbf{a}$ for some doubly stochastic **Q**.

2. **Birkhoff's theorem** [29]: A matrix **Q** is doubly stochastic if and only if it is a convex combination of finitely many permutation matrices, i.e., there are finitely many permutation matrices $\mathbf{P}_i$ such that

$$\sum_{i=1}^{N} \alpha_i \mathbf{P}_i = \mathbf{Q}, \quad \text{where} \quad 0 \leq \alpha_i \leq 1, \quad \text{and} \quad \sum_{i=1}^{N} \alpha_i = 1. \tag{2.8}$$

3. **Orthostochastic majorization theorem** [29]: For $\mathbf{a}, \mathbf{b} \in \mathcal{R}^M$, following statements are equivalent:

   (a) **a** majorizes **b**.

   (b) There exists an orthostochastic matrix **Q** (corresponding to a unitary matrix **U**) such that $\mathbf{b} = \mathbf{Q}\mathbf{a}$.

   (c) There is a Hermitian matrix **H** with entries of **a** as its eigenvalues and entries of **b** on its diagonal.

On the proofs: The majorization theorem actually follows from the orthostochastic majorization theorem. See [27] or [69] for an independent proof. Regarding Birkhoff's theorem, as all permutation matrices are doubly stochastic, so is their convex combination $\mathbf{Q}$ of (2.8) (Appendix A). The converse proof is more elaborate [29]. In the orthostochastic majorization theorem, equivalence of (b) and (c) is easily proved: The key idea is that for any diagonal matrix $\mathbf{\Lambda}$ and unitary matrix $\mathbf{T}$, $\text{diag}(\mathbf{T\Lambda T}^\dagger) = \mathbf{Q}\text{diag}(\mathbf{\Lambda})$ where $\mathbf{Q}$ is the orthostochastic matrix corresponding to $\mathbf{T}$. This is because if t_{ij} is the ij-th entry of $\mathbf{T}$ and $\text{diag}(\mathbf{\Lambda}) = (\lambda_0, \lambda_1, \ldots, \lambda_{M-1})^T$, the i-th diagonal entry of $\mathbf{T\Lambda T}^\dagger$ is $\sum_{j=0}^{M-1} |t_{ij}|^2 \lambda_j$, which is exactly the i-th entry of $\mathbf{Q}\text{diag}(\mathbf{\Lambda})$. So given (b), we choose $\text{diag}(\mathbf{\Lambda}) = \mathbf{a}$ and prove (c) by setting $\mathbf{H} = \mathbf{U\Lambda U}^\dagger$. Conversely given (c), we prove (b) by letting $\mathbf{U}$ be a unitary matrix diagonalizing $\mathbf{H}$, i.e., satisfying $\mathbf{H} = \mathbf{U\Lambda U}^\dagger$ for diagonal $\mathbf{\Lambda}$.

That (b) (or (c)) implies (a) follows from the majorization theorem since all $M \times M$ orthostochastic matrices are doubly stochastic. As the converse is false unless $M \leq 2$ (Appendix B), the result that (a) implies (b) (or (c)) is stronger than the corresponding result in the 'plain' majorization theorem. This result is not used until Section 2.5.1. Its proof is more involved [29]. The fact that (c) implies (a) is in fact precisely the statement that the KLT is the PCFB for the class of transform coders, as elaborated in Section 2.5.1.

Proof of Theorem 2.2

Let a PCFB for the class $\mathcal{C}$ exist for the given input psd $\mathbf{S}_{\mathbf{xx}}(e^{j\omega})$. Let $\mathbf{v}_*$ be the PCFB subband variance vector (unique up to permutation; see Section 2.4.2). Let $\mathbf{P}_j$ be the $M \times M$ permutation matrices, for $j = 1, 2, \ldots, J$ (where $J = M!$), and let $\mathbf{v}_j \triangleq \mathbf{P}_j \mathbf{v}_*$. Thus $E \triangleq \{\mathbf{v}_j : j = 1, 2, \ldots, J\}$ is the (finite) set of all permutations of $\mathbf{v}_*$. We have to prove that $\text{co}(\mathcal{S}) = \text{co}(E)$. For this, take any $\mathbf{v} \in \mathcal{S}$. By definition of PCFBs, $\mathbf{v}_*$ majorizes $\mathbf{v}$. So by the majorization theorem, $\mathbf{v} = \mathbf{Q}\mathbf{v}_*$ for some doubly stochastic matrix $\mathbf{Q}$. By Birkhoff's theorem, $\mathbf{Q}$ is some convex combination of the $\mathbf{P}_j$. Thus,

$$\mathbf{v} = \mathbf{Q}\mathbf{v}_* = \sum_{j=1}^{J} \alpha_j \mathbf{P}_j \mathbf{v}_* = \sum_{j=1}^{J} \alpha_j \mathbf{v}_j \text{ for some } \alpha_j \text{ such that } 0 \leq \alpha_j \leq 1, \sum_{j=1}^{J} \alpha_j = 1.$$

So every $\mathbf{v} \in \mathcal{S}$ is a convex combination of the $\mathbf{v}_j$, i.e., $\mathcal{S} \subseteq \text{co}(E)$. So $\text{co}(\mathcal{S}) \subseteq \text{co}(\text{co}(E)) = \text{co}(E)$. But by permutation-symmetry of $\mathcal{S}$ (Section 2.3.1), $E \subset \mathcal{S}$, so $\text{co}(E) \subseteq \text{co}(\mathcal{S})$. Combining, $\text{co}(\mathcal{S}) = \text{co}(E)$ as desired.

Conversely, let $\mathbf{v}_*$ be a vector such that with $\mathbf{v}_j = \mathbf{P}_j \mathbf{v}_*$ and $E \triangleq \{\mathbf{v}_j : j = 1, 2, \ldots, J\}$, we have $\text{co}(\mathcal{S}) = \text{co}(E)$. We then have to prove that a PCFB for the class $\mathcal{C}$ exists for the given input psd, and that $\mathbf{v}_*$ is a PCFB subband variance vector. To do this, note that $\mathcal{S} \subseteq \text{co}(\mathcal{S}) = \text{co}(E)$. Thus if $\mathbf{v} \in \mathcal{S}$ then $\mathbf{v} \in \text{co}(E)$, so $\mathbf{v}$ can be written as a convex combination of the elements $\mathbf{v}_j \in E$. So there are α_j such that

$$0 \leq \alpha_j \leq 1, \ \sum_{j=1}^{J} \alpha_j = 1 \quad \text{and} \quad \mathbf{v} = \sum_{j=1}^{J} \alpha_j \mathbf{v}_j = \sum_{j=1}^{J} \alpha_j \mathbf{P}_j \mathbf{v}_* = \mathbf{Q}\mathbf{v}_* \ \text{ where } \ \mathbf{Q} \triangleq \sum_{j=1}^{J} \alpha_j \mathbf{P}_j.$$

Here $\mathbf{Q}$ is a convex combination of permutation matrices $\mathbf{P}_j$, so it is doubly stochastic (Birkhoff's theorem). As $\mathbf{v} = \mathbf{Q}\mathbf{v}_*$, by the majorization theorem $\mathbf{v}_*$ majorizes $\mathbf{v}$. Thus, a FB with subband variance vector $\mathbf{v}_*$ will be a PCFB for the given class $\mathcal{C}$ and input psd. Thus it only remains to show that there is such a FB in $\mathcal{C}$, i.e., that $\mathbf{v}_* \in \mathcal{S}$. To see this, note that $\text{co}(\mathcal{S}) = \text{co}(E)$ and E is a finite set. Hence, $\text{co}(\mathcal{S})$ is a polytope whose extreme points lie within E. Now the extreme points of $\text{co}(\mathcal{S})$ always lie in $\mathcal{S}$. (Being in $\text{co}(\mathcal{S})$, they are convex combinations of some points in $\mathcal{S}$, but being extreme, these must be trivial combinations.) Thus $\mathbf{v}_j \in \mathcal{S}$ for at least one j, and hence for all j (by the permutation-symmetry of $\mathcal{S}$, Section 2.3.1). Hence indeed $\mathbf{v}_* \in \mathcal{S}$. $\bigtriangledown \bigtriangledown \bigtriangledown$

Note that in general all we can say about the extreme points of a polytope $\text{co}(E)$ is that they lie in E. Here however, with E as the (finite) set of all permutations of $\mathbf{v}_*$, in fact *all* points in E are extreme points of $\text{co}(E)$, i.e., no vector in E is a convex combination of other vectors of E. This is provable by induction on the vector dimension M: Let $\mathbf{v}_1 = \sum_{j=2}^{J} \alpha_j \mathbf{v}_j$ with $0 \leq \alpha_j \leq 1$ and $\sum_{j=2}^{J} \alpha_j = 1$. Then the greatest entry of $\mathbf{v}_1$ is a convex combination of real numbers no greater than itself. So all these numbers must be equal. Deleting from each $\mathbf{v}_j$ the entry corresponding to this number yields the induction hypothesis.

Functions minimized by majorization. Currently known instances of PCFB optimality in signal processing problems arise from minimization objectives of the form (2.3) where the functions h_i are concave on $\mathcal{R}_+$. Theorem 2.3 of course shows PCFB optimality for a more general family of objectives, namely those that are concave in the subband variance vector (and need not necessarily have the special form of (2.3)). In fact even this is not the complete family of objectives minimized by PCFBs. For example, if g is a monotone increasing function on $\mathcal{R}$, then for any concave objective $\phi(\cdot)$, clearly $g(\phi(\cdot))$ is also minimized by PCFBs. Unless g is also concave, in general this new function is not concave. A specific nonconcave example of this kind is generated by $\phi(x_1, \ldots, x_M) = \sum_i \log(x_i)$ and $g(y) = e^y$, giving $g(\phi(\cdot)) = \psi(x_1, \ldots, x_M) = \prod_i x_i$.

If attention is restricted to symmetric functions (i.e., functions ϕ obeying $\phi(\mathbf{P}\mathbf{x}) = \phi(\mathbf{x})$ for all $\mathbf{x}$ if $\mathbf{P}$ is any permutation matrix) then the functions minimized by majorization are said to be **Schur–concave** [44]. To be precise, ϕ is said to be Schur–concave if $\phi(\mathbf{x}) \leq \phi(\mathbf{y})$ whenever $\mathbf{x}$ majorizes $\mathbf{y}$. (This implies symmetry of ϕ since $\mathbf{P}\mathbf{x}$ majorizes $\mathbf{x}$ for any permutation matrix $\mathbf{P}$.) Thus symmetric concave functions are examples of Schur–concave functions, while the function ψ defined earlier is a Schur–concave function that is not concave. Clearly PCFBs minimize all Schur–concave objectives. Full characterizations and several interesting examples of such functions can be found in [44].

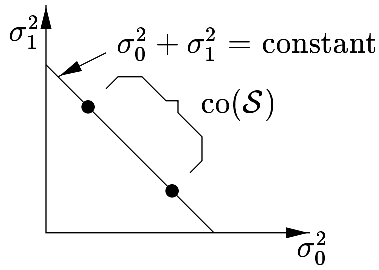


Figure 2.4: Search space $\mathcal{S}$ for a class of two channel FBs.

2.5 Standard class PCFBs, and optimality for compression

This section first shows existence of PCFBs for three special classes of FBs, namely classes with $M = 2$ channels, the class of M channel orthogonal transform coders, and that of all M channel orthonormal FBs. This well-known result is reviewed to show how it fits in the framework of the earlier sections, which have not yet been restricted to any specific class of FBs. We also prove the convexity of the search-space for these classes, which has not been observed earlier. We then review PCFB optimality for data compression.

To begin, let $\mathcal{C}$ be *any* class of uniform orthonormal *two channel* FBs, e.g. that of FIR or IIR FBs with a given bound on the filter order. Irrespective of the input psd matrix, all realizable subband variance vectors $(\sigma_0^2, \sigma_1^2)^T$ in the search-space $\mathcal{S} \subset \mathcal{R}^2$ then have the same value of $\sigma_0^2 + \sigma_1^2$ (Section 2.3.1). Thus $\mathcal{S}$ lies wholly on a line of slope -1 in $\mathcal{R}^2$. So co($\mathcal{S}$) is an interval on this line; see Fig. 2.4. Thus co($\mathcal{S}$) is a polytope with two extreme points, namely the endpoints of the interval. By the definition, a PCFB is simply a FB maximizing one subband variance, thereby minimizing the other. So it always exists for such classes $\mathcal{C}$, and corresponds to the two extreme points of co($\mathcal{S}$), irrespective of the input psd.[3]

2.5.1 The transform coder class

The transform coder class $\mathcal{C}^t$ is defined as the class of uniform M channel orthonormal FBs whose polyphase matrix ($\mathbf{E}(z)$ in Fig. 2.1) is a constant unitary matrix $\mathbf{T}$. So in effect we can speak of $\mathcal{C}^t$ as being the set of all $M \times M$ unitary matrices. Let $\mathbf{R_{xx}}$ be the autocorrelation matrix of the input $\mathbf{x}(n)$ of Fig. 2.1. We have

Theorem 2.4: Transform coders – KLT, PCFBs and polytopes.

1. A PCFB always exists for $\mathcal{C}^t$. Hence the set $\mathcal{S}$ of realizable subband variance vectors for $\mathcal{C}^t$ has a convex hull co($\mathcal{S}$) that is a polytope, as stated by Theorem 2.2, Section 2.4.3.

[3]If co($\mathcal{S}$) is an *open* interval (i.e., one not containing its endpoints), no single FB achieves the maximum subband variance; hence there is no PCFB. However this situation is contrived and does not happen for most natural FB classes and input psds.

2. A unitary matrix $\mathbf{T} \in \mathcal{C}^t$ is a PCFB for $\mathcal{C}^t$ iff it diagonalizes $\mathbf{R_{xx}}$, i.e., $\mathbf{T R_{xx} T}^\dagger$ is diagonal. In other words, $\mathbf{T}$ is a PCFB for $\mathcal{C}^t$ iff it is the Karhunen-Loeve transform (KLT) for the input, i.e., it decorrelates the input (the subband signals $v_i^{(x)}(n), i = 0, 1, \ldots, M-1$ are uncorrelated for each time instant n).

3. $\mathcal{S} = \text{co}(\mathcal{S})$. So $\mathcal{S}$ itself is a polytope with extreme points as permutations of the KLT subband variance vector. ■

Proof: The subband variance vector computation (2.4) becomes

$$\mathbf{v} = \text{diag}(\mathbf{R_{vv}}), \quad \text{where } \mathbf{R_{vv}} = \mathbf{T R_{xx} T}^\dagger, \quad \text{where } \mathbf{R_{xx}} = \frac{1}{2\pi}\int_0^{2\pi} \mathbf{S_{xx}}(e^{j\omega})\, d\omega.$$

Here $\mathbf{R_{vv}}$ is the autocorrelation matrix of the vector process $(v_0^{(x)}(n), v_1^{(x)}(n), \ldots, v_{M-1}^{(x)}(n))^T$ of Fig. 2.1. The input **KLT** is defined as the FB with unitary polyphase matrix $\mathbf{K}$ that diagonalizes $\mathbf{R_{xx}}$, i.e., such that $\mathbf{K R_{xx} K}^\dagger = \mathbf{\Lambda}$ is a diagonal matrix. Thus $\mathbf{v}_* \triangleq \text{diag}(\mathbf{\Lambda})$ is the subband variance vector of the KLT, and has as entries the eigenvalues of $\mathbf{R_{xx}}$. Now the Hermitian matrix $\mathbf{R_{vv}} = \mathbf{T R_{xx} T}^\dagger = \mathbf{T K}^\dagger \mathbf{\Lambda K T}^\dagger$ has entries of $\mathbf{v}$ on its diagonal, and entries of $\mathbf{v}_*$ as its eigenvalues. Hence $\mathbf{v}_*$ majorizes $\mathbf{v}$ by the orthostochastic majorization theorem of Section 2.4.4 (specifically by the fact that (c) implies (a) in its statement). This shows that the KLT is a PCFB – a well-known result. Conversely, if $\mathbf{T}$ is a PCFB for $\mathcal{C}^t$ then $\mathbf{v} = \mathbf{v}_*$ (up to permutation). So the Hermitian matrix $\mathbf{R_{vv}}$ has its eigenvalues as its diagonal elements, and is hence necessarily diagonal; i.e., $\mathbf{T}$ is the KLT for the input.

Lastly, to show that $\mathcal{S}$ is the polytope $\text{co}(\mathcal{S})$, take any $\mathbf{v} \in \text{co}(\mathcal{S})$. Then $\mathbf{v}_*$ majorizes $\mathbf{v}$. We now make a stronger application of the orthostochastic majorization theorem, i.e., that (a) implies (c) in its statement in Section 2.4.4. This shows that there is a Hermitian matrix $\mathbf{R_{vv}}$ with the entries of $\mathbf{v}_*$ as its eigenvalues and those of $\mathbf{v}$ on its diagonal. As $\mathbf{R_{vv}}, \mathbf{R_{xx}}$ have the same eigenvalues, they are 'similar', i.e., $\mathbf{U R_{xx} U}^\dagger = \mathbf{R_{vv}}$ for some unitary matrix $\mathbf{U}$. So the FB $\mathbf{U} \in \mathcal{C}^t$ has subband variance vector $\text{diag}(\mathbf{U R_{xx} U}^\dagger) = \text{diag}(\mathbf{R_{vv}}) = \mathbf{v}$. Thus $\mathbf{v}$ is a realizable subband variance vector for $\mathcal{C}^t$, i.e., $\mathbf{v} \in \mathcal{S}$. This holds for any $\mathbf{v} \in \text{co}(\mathcal{S})$, so $\mathcal{S} = \text{co}(\mathcal{S})$.

2.5.2 The unconstrained class

The class $\mathcal{C}^u$ is defined to contain *all* uniform M channel orthonormal FBs, with no constraints on the filters besides those imposed by orthonormality. So FBs in $\mathcal{C}^u$ could have ideal unrealizable filters. We could in effect think of $\mathcal{C}^u$ as the set of all $M \times M$ (analysis polyphase) matrices $\mathbf{E}(e^{j\omega})$ that are unitary for all ω. An exact analog of Theorem 2.4 holds for this class too. The only difference is in the construction of the PCFB from the given input psd matrix $\mathbf{S_{xx}}(e^{j\omega})$; first described in [60],[61]. This section reviews this construction, and proves the result $\mathcal{S} = \text{co}(\mathcal{S})$ for the class $\mathcal{C}^u$.

PCFB construction: Let $\mathbf{K}(e^{j\omega}) \in \mathcal{C}^u$ diagonalize $\mathbf{S}_{\mathbf{xx}}(e^{j\omega})$ for each ω, i.e., $\mathbf{K}(e^{j\omega})\mathbf{S}_{\mathbf{xx}}(e^{j\omega})\mathbf{K}^\dagger(e^{j\omega}) = \mathbf{\Lambda}(e^{j\omega})$ where $\mathbf{\Lambda}(e^{j\omega})$ is diagonal (for all ω) and $\text{diag}(\mathbf{\Lambda}(e^{j\omega})) = (\lambda_0(e^{j\omega}), \lambda_1(e^{j\omega}), \ldots, \lambda_{M-1}(e^{j\omega}))^T$. Using (2.4), the subband variance vector $\mathbf{v}$ of an arbitrary FB $\mathbf{E}(e^{j\omega}) \in \mathcal{C}^u$ is given by

$$2\pi\mathbf{v} = \int_0^{2\pi} \text{diag}(\mathbf{E}\mathbf{S}_{\mathbf{xx}}\mathbf{E}^\dagger)\ d\omega = \int_0^{2\pi} \mathbf{Q}(e^{j\omega})\ (\lambda_0(e^{j\omega}), \lambda_1(e^{j\omega}), \ldots, \lambda_{M-1}(e^{j\omega}))^T\ d\omega. \tag{2.9}$$

Here at each ω, $\mathbf{Q}(e^{j\omega})$ is the orthostochastic matrix corresponding to the unitary matrix $\mathbf{E}(e^{j\omega})\mathbf{K}^\dagger(e^{j\omega})$. So *at each frequency* ω, the integrand vector of (2.9) produced by the FB $\mathbf{K}(e^{j\omega}) \in \mathcal{C}^u$ majorizes the corresponding vector of any FB in $\mathcal{C}^u$. This holds no matter how we order the eigenvalues $\lambda_i(e^{j\omega})$ in (2.9). The integration operation preserves this majorization relation if and only if the $\lambda_i(e^{j\omega})$ are 'ordered consistently' at all ω. By this we mean that if we number the $\lambda_i(e^{j\omega})$ so that the entries of $\mathbf{v}$ are in descending order, then $\lambda_0(e^{j\omega}) \geq \lambda_1(e^{j\omega}) \geq \ldots \geq \lambda_{M-1}(e^{j\omega})$ *for all* ω. Thus, a FB $\mathbf{K}(e^{j\omega}) \in \mathcal{C}^u$ is a PCFB for $\mathcal{C}^u$ iff it causes two effects : (1) **totally decorrelating** the input, i.e., diagonalizing its psd matrix $\mathbf{S}_{\mathbf{xx}}(e^{j\omega})$, and (2) causing **spectral majorization** [61] – the said ordering of eigenvalues of $\mathbf{S}_{\mathbf{xx}}(e^{j\omega})$. Note that the PCFB for $\mathcal{C}^u$ has uncorrelated subband *processes*, unlike the *instantaneous* decorrelation produced by the KLT (Theorem 2.4).

Proving $\mathcal{S} = \mathbf{co}(\mathcal{S})$: To prove this property for the class $\mathcal{C}^u$, let $\mathbf{v}_*$ be the PCFB subband variance vector, and let $\mathbf{v} \in \text{co}(\mathcal{S})$. Then $\mathbf{v}_*$ majorizes $\mathbf{v}$. So by the orthostochastic majorization theorem (Section 2.4.4), $\mathbf{v} = \mathbf{Q}\mathbf{v}_*$ for some orthostochastic matrix $\mathbf{Q}$ corresponding to a unitary matrix $\mathbf{U}$. Thus, if $\mathbf{K}(e^{j\omega})$ is the polyphase matrix of the PCFB for $\mathcal{C}^u$, (2.9) shows that the FB in $\mathcal{C}^u$ with polyphase matrix $\mathbf{UK}(e^{j\omega})$ produces subband variance vector $\mathbf{v}$, i.e., $\mathbf{v} \in \mathcal{S}$. This shows $\mathcal{S} = \text{co}(\mathcal{S})$.

2.5.3 PCFB optimality for coding/compression

Here we consider the problems of [35],[61], where the processors P_i of Fig. 2.1 are quantizers, and the desired output $d(n)$ equals the input $x(n)$. This situation fits the general problem formulation of Section 2.3 under appropriate quantizer models. The subband error signal $v_i^{(e)}(n)$ of Section 2.3 here represents the i-th subband quantization noise. Under the quantizer model we assume, this noise is zero mean with variance

$$E[|v_i^{(e)}(n)|^2] = f_i(b_i)\sigma_i^2. \tag{2.10}$$

Here b_i is the number of bits allocated to the i-th quantizer, and f_i is a characteristic of the quantizer called the normalized quantizer function [35]. We assume that f_i does not depend on the filter-bank in any way; and that the quantization noise processes in different subbands are jointly stationary. The problem then fits the formulation of Section 2.3. Comparing (2.10) with (2.2) reveals the minimization objective g to be as in (2.3), i.e.,

$$g(\sigma_0^2, \sigma_1^2, \ldots, \sigma_{M-1}^2) = \frac{1}{M} \sum_{i=0}^{M-1} h_i(\sigma_i^2) \qquad \text{with} \qquad h_i(x) = f_i(b_i)x. \tag{2.11}$$

Thus the h_i are linear (and hence concave), so g is indeed concave. So by Theorem 2.3, the PCFB if it exists is optimal for this problem. This is true **no matter what the bit allocation** b_i is.

It is important to note that for the validity of our assumptions of Section 2.3 (and hence for PCFBs to be optimal), the function $h_i(x) = f_i(b_i)x$ **must not depend on the FB** in any way. This is often not the case: In quantizers optimized to their input probability density function (pdf), f_i depends on the i-th subband pdf which in turn is influenced by choice of FB. Even with the model of [61], i.e., uniform quantization under the high bitrate approximation, $f_i(b_i) = c_i 2^{-2b_i}$ where the constant c_i (and hence f_i) depends on the i-th subband pdf. If we further assume the input to be a *Gaussian* random process, then all subbands have Gaussian pdf independent of choice of FB. For this special case, all c_i are equal and constant, and the PCFB is indeed optimal. The need for these assumptions is illustrated by Feng and Effros [23], who demonstrate that the **KLT is not the optimal** orthogonal transform if the input has a uniform distribution.

For the case when $f_i(b_i) = c2^{-2b_i}$ (for which the PCFB *is* optimal), the optimal bit allocation b_i (subject to a constraint on the total bit budget $\sum_{i=0}^{M-1} b_i = B$) is explicitly computable using the arithmetic mean – geometric mean (AM-GM) inequality. The objective under this bit allocation becomes the GM of the subband variances, i.e., $g = (\prod_{i=0}^{M-1} \sigma_i^2)^{(1/M)}$. Minimizing this is equivalent to minimizing $\log(g) = \frac{1}{M} \sum_{i=0}^{M-1} \log(\sigma_i^2)$. This is a concave function of the subband variance vector because $\log(x)$ is concave in x. For general quantizer functions f_i, the optimizations of the FB and the bit allocation have been *decoupled*, since the PCFB is optimum for *all* bit allocations [35]. However, note that different permutations of a PCFB may be optimal for different bit allocations. Also, computing the optimum bit allocation may be more involved. We can however prove one intuitive statement about the optimum b_i in the special case when all f_i are equal to a decreasing function f: In this case, a subband with larger variance receives more bits.

In (2.11), all h_i are linear, i.e., $h_i(x) = m_i x + c_i$ for constants m_i, c_i ($m_i = f_i(b_i)$, $c_i = 0$). In such cases, we can algebraically prove PCFB optimality [35] without using any result on majorization: As c_i are constants, the optimization is unaffected by taking $c_i = 0$. With $m_0 \le m_1 \le \ldots \le m_{M-1}$ and $\sigma_0^2 \ge \sigma_1^2 \ge \ldots \ge \sigma_{M-1}^2$,

$$Mg(\sigma_0^2, \sigma_1^2, \ldots, \sigma_{M-1}^2) = \sum_{i=0}^{M-1} m_i \sigma_i^2 = \sum_{i=0}^{M-2} (m_i - m_{i+1}) \left(\sum_{j=0}^{i} \sigma_j^2 \right) + m_{M-1} \left(\sum_{j=0}^{M-1} \sigma_j^2 \right). \tag{2.12}$$

As the last term is constant for all FBs, and since $m_i - m_{i+1} \le 0$, the above g is minimized by the

PCFB which by definition maximizes all the partial sums $\sum_{j=0}^{i} \sigma_j^2$ for $i = 0, 1, \ldots, M-2$. This proof shows two noteworthy facts not shown by the earlier proof: (1) It exhibits the best permutation of the PCFB to be used, namely that in which the largest subband variance σ_i^2 is associated with the least m_i and so on. (2) It shows that the optimum FB is necessarily a PCFB if the m_i are distinct. However this simple approach works only for *linear* h_i, and thus fails for many of the problems of Section 2.7 that result in nonlinear concave h_i.

2.6 Filter Bank classes having no PCFB

Existence of a PCFB for a class $\mathcal{C}$ of orthonormal FBs implies a very strong condition on the subband variance vectors of the FBs in $\mathcal{C}$. There are many classes $\mathcal{C}$ which do not have PCFBs. Indeed it seems quite plausible that the classes of Section 2.5 are the only ones having PCFBs for all input power spectra. This section reviews some known results on nonexistence of PCFBs, and shows that the classes of ideal DFT and cosine-modulated FBs do not have PCFBs for several input spectra.

If a PCFB for the given class $\mathcal{C}$ of FBs exists, it simultaneously optimizes over $\mathcal{C}$, several functions of the subband variances (Section 2.4). So we can show nonexistence of PCFBs for $\mathcal{C}$ by proving that no single FB in $\mathcal{C}$ can optimize two of such functions. This method is used in [36],[45] for certain classes of FIR FBs, for a fixed input psd. The two functions used are the largest subband variance and the coding gain, both maximized by a PCFB if it exists. However all optimizations are numerical. Nonexistence of PCFBs has not yet been *proved* for any reasonably general FIR class, say the class of all M channel ($M > 2$) FIR orthonormal FBs with polyphase matrix of McMillan degree $\mu > 0$ (though it seems very likely that such classes do not have PCFBs). We now prove nonexistence of PCFBs for the classes of DFT and cosine-modulated FBs.

Definition. The class $\mathcal{C}^{dft}$ of M channel orthonormal **DFT FBs** is the one containing all FBs as in Fig. 1.3 where the analysis filters are related by $H_k(e^{j\omega}) = P(e^{j(\omega - \frac{2\pi k}{M})})$ for some filter $P(e^{j\omega})$ called the *prototype*. For example, any $P(e^{j\omega})$ which has an aliasfree(M) support and has constant magnitude on its support (and is thus Nyquist(M)) produces a FB in $\mathcal{C}^{dft}$.

Definition. The class $\mathcal{C}^{cmfb}$ of M channel orthonormal **cosine-modulated** FBs (CMFBs) is the one containing all FBs as in Fig. 1.3 where $H_k(e^{j\omega}) = P(e^{j(\omega - \frac{k\pi}{M} - \frac{\pi}{2M})}) + P(e^{j(\omega + \frac{k\pi}{M} + \frac{\pi}{2M})})$ for some filter $P(e^{j\omega})$ called the *prototype*. For example, any $P(e^{j\omega})$ having an aliasfree($2M$) support and with constant magnitude on its support, is a valid prototype.

Theorem 2.5: PCFB nonexistence for DFT, cosine-modulated FB classes. There are families of input psds such that the class $\mathcal{C}^{dft}$ defined above does not have a PCFB. The same holds for the class $\mathcal{C}^{cmfb}$. ■

Proof: Consider first the class $\mathcal{C}^{dft}$. Fig. 2.5a shows an input psd, two valid prototypes $P^{(j)}(e^{j\omega})$, and the zeroth filters $H_0^{(j)}(e^{j\omega}) = P^{(j)}(e^{j\omega})$, $j = 1, 2$ in the DFT FBs produced by the prototypes.

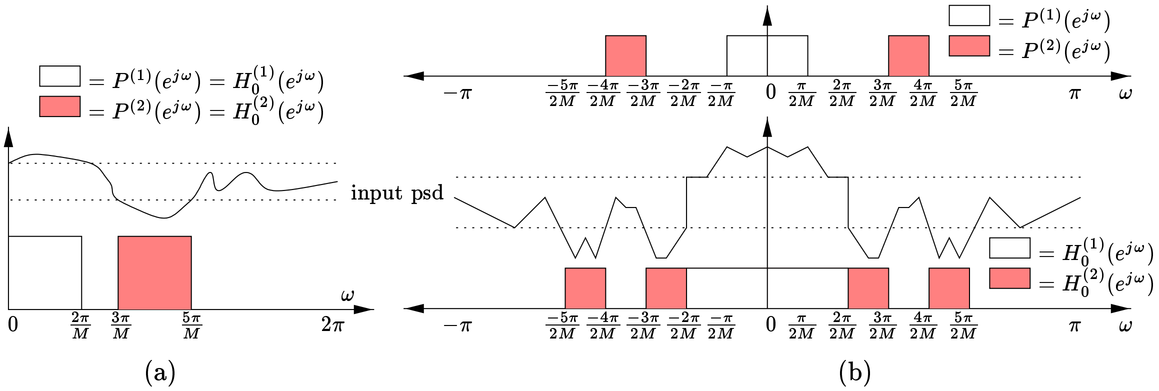


Figure 2.5: Nonexistence of PCFBs. (a) Class of DFT FBs, (b) Class of cosine-modulated FBs.

For the input psd, the filter $H_0^{(1)}(e^{j\omega})$ produces the maximum subband variance achievable by any M channel orthonormal FB, and hence by any FB in $\mathcal{C}^{dft}$. ($H_0^{(1)}(e^{j\omega})$ is the *compaction filter* [61] for the input psd.) Likewise, $H_0^{(2)}(e^{j\omega})$ yields the minimum subband variance possible by any M channel orthonormal FB, and hence by any FB in $\mathcal{C}^{dft}$. Now a PCFB simultaneously maximizes the largest and minimizes the least subband variance. So if a PCFB for $\mathcal{C}^{dft}$ exists, it must contain both filters $H_0^{(j)}(e^{j\omega})$, $j = 1, 2$. This is impossible as these filters are not obtainable from each other by shift of an integer multiple of $\frac{2\pi}{M}$, so a FB having both of them cannot be in the class $\mathcal{C}^{dft}$. Identical arguments hold for the class $\mathcal{C}^{cmfb}$, for the input psd, prototypes $P^{(j)}(e^{j\omega})$ and corresponding filters $H_0^{(j)}(e^{j\omega})$, $j = 1, 2$ shown in Fig. 2.5b. The only difference is that we no longer have $H_0^{(j)}(e^{j\omega}) = P^{(j)}(e^{j\omega})$. Also it takes more effort to show that no FB in $\mathcal{C}^{cmfb}$ can have both filters $H_0^{(j)}(e^{j\omega})$, $j = 1, 2$: We can show that if a CMFB has $H_0^{(1)}(e^{j\omega})$ as one of its filters, then the band–edges of all its filters must be multiples of $\frac{\pi}{M}$, so $H_0^{(2)}(e^{j\omega})$ cannot be a filter in it. In fact [9] a CMFB having $H_0^{(1)}(e^{j\omega})$ of Fig. 2.5b as one of its filters is necessarily the CMFB produced by $P^{(1)}(e^{j\omega})$ of Fig. 2.5b as prototype.

2.7 Optimal noise reduction with filter-banks

Suppose the FB input $x(n)$ of Fig. 2.1 is $x(n) = s(n) + \mu(n)$ where $s(n)$ is a pure signal and $\mu(n)$ is zero mean additive noise. The desired FB output is $d(n) = s(n)$, and the goal of the system of Fig. 2.1 is to produce output $y(n)$ that approximates $s(n)$ as best as possible. We consider the case when all the subband processors P_i are memoryless multipliers k_i, as shown in Fig. 2.6. This problem fits the formulation of Section 2.3 if we assume that $s(n)$ and $\mu(n)$ are uncorrelated, and that $\mu(n)$ is white with a fixed known variance $\eta^2 > 0$. Indeed, using the notation of Section 2.3, the i-th subband process $v_i^{(x)}(n)$ contains a signal component $v_i^{(s)}(n)$ and a zero mean additive noise component $v_i^{(\mu)}(n)$. Orthonormality of the FB ensures that the noise components are uncorrelated

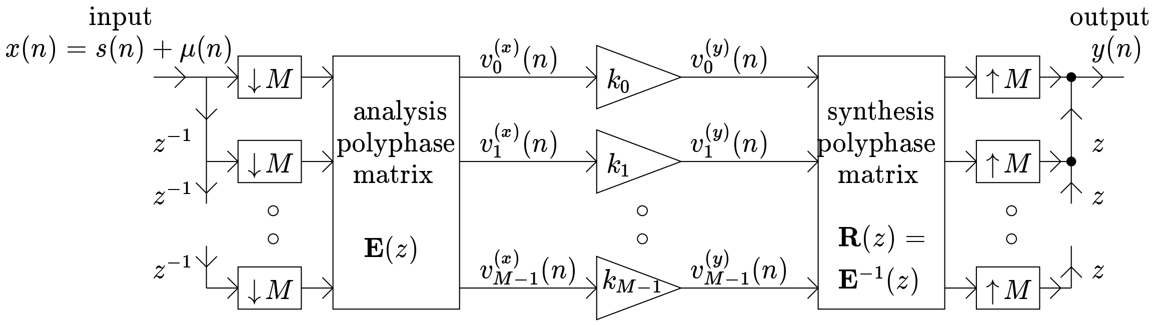


Figure 2.6: Filter bank based noise reduction.

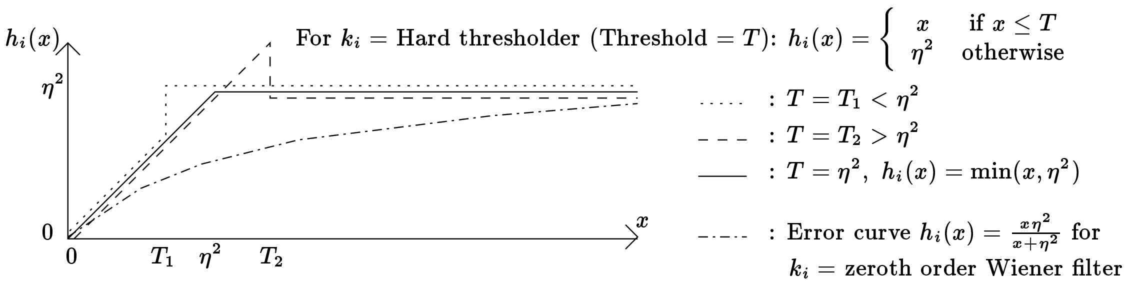


Figure 2.7: Subband error functions in white noise reduction.

to the signal components and have constant variance η^2 (e.g., this is deducible from (2.4)). The subband error process is

$$v_i^{(e)}(n) = v_i^{(d)}(n) - v_i^{(y)}(n) = v_i^{(s)}(n) - k_i v_i^{(x)}(n) = (1 - k_i) v_i^{(s)}(n) - k_i v_i^{(\mu)}(n). \tag{2.13}$$

Thus, the M processes $v_i^{(e)}(n)$ are jointly WSS. As $v_i^{(\mu)}(n)$ is zero mean and uncorrelated to $v_i^{(s)}(n)$,

$$E[|v_i^{(e)}(n)|^2] = |1 - k_i|^2 \sigma_i^2 + |k_i|^2 \eta^2, \tag{2.14}$$

where $\sigma_i^2 = E[|v_i^{(s)}(n)|^2]$ is the i-th signal subband variance. The best choice of multiplier k_i (minimizing the error (2.14)) is the **zeroth order Wiener filter** $k_i = \frac{\sigma_i^2}{\sigma_i^2 + \eta^2}$. This is implementable in practice as η^2 is known and $\sigma_i^2 = E[|v_i^{(x)}(n)|^2] - \eta^2$ can be estimated from the subband signal $v_i^{(x)}(n)$. With this choice, (2.14) becomes $E[|v_i^{(e)}(n)|^2] = \frac{\sigma_i^2 \eta^2}{\sigma_i^2 + \eta^2}$, which is as in (2.2) with

$$h_i(x) = \frac{x\eta^2}{x + \eta^2}. \tag{2.15}$$

This function h_i is plotted in Fig. 2.7, and is easily verified to be concave on $[0, \infty)$. So by Theorem 2.3, PCFBs are optimal if the subband multipliers k_i are zeroth order Wiener filters.

2.7.1 Remarks on PCFB optimality for noise reduction

PCFBs for the pure or the noisy signal? Notice a difference between the argument σ_i^2 of h_i here and in (2.2): In (2.2), σ_i^2 was the variance of the subband signal $v_i^{(x)}(n)$ corresponding to the FB input $x(n)$. Here it is the variance of the subband signal $v_i^{(s)}(n) = v_i^{(x)}(n) - v_i^{(\mu)}(n)$ corresponding to the pure signal $s(n)$. Thus, use of Theorem 2.3 proves the optimality of a PCFB *for the signal* $s(n)$, i.e., a FB that causes the subband variance vector corresponding to $s(n)$ to majorize the variance vectors obtained by using other FBs in the given class $\mathcal{C}$. However, because $v_i^{(\mu)}(n)$ is white with variance η^2 and uncorrelated to $v_i^{(s)}(n)$, we have $E[|v_i^{(s)}(n)|^2] = \sigma_i^2 = E[|v_i^{(x)}(n)|^2] - \eta^2$. Thus any PCFB for $s(n)$ is also a PCFB for $x(n)$ and vice-versa.

Other choices of subband multipliers. The Wiener filter is the optimum choice of the multiplier k_i in Fig. 2.6. However, we may note that there are other choices that also result in an error function (2.14) that is concave in the subband variance σ_i^2. Thus the PCFB will be optimal when the k_i are any combination of such choices. One such other choice is a constant multiplier that is independent of the choice of FB (reminiscent of taps in a graphic equalizer in audio equipment). The error is then (2.14) which is in fact 'linear' in σ_i^2. As the next remark shows, this observation yields an alternative proof of PCFB optimality with subband Wiener filtering. Another possible choice of multiplier k_i is the subband hard threshold:

$$k_i = \begin{cases} 1 & \text{if } \sigma_i^2 \geq T \\ 0 & \text{otherwise} \end{cases} \tag{2.16}$$

The resulting subband error functions h_i are plotted in Fig. 2.7 for different thresholds $T > 0$. For the unique value $T = \eta^2$, which is the optimum threshold in the sense of minimizing $h_i(x)$ pointwise at all x, the resulting $h_i(x) = \min(x, \eta^2)$ is concave on $[0, \infty)$ (though not strictly concave). These choices of multiplier k_i are perhaps not of serious practical interest when compared to the Wiener filter, but certainly demonstrate an academic implication of PCFB optimality. More practical hard thresholding schemes for noise suppression [22] have a threshold that is applied individually to each element of the subband signal sequence (i.e. to each subband or 'wavelet' coefficient) rather than on a subband by subband basis.

PCFB optimality for subband Wiener filtering: Another proof. One can prove PCFB optimality when all subband multipliers k_i are Wiener filters, without using any of the majorization theory arguments of Section 2.4 or the concavity of the function (2.15). To do this, observe that the PCFB is optimal if the subband multipliers are all constants independent of the FB. This was noted in the earlier remark, and can be proved algebraically as in Section 2.5.3 (see Equation (2.12)) without using convexity theory. This is possible since the $h_i(\sigma_i^2)$ in this case are as in (2.14) which is 'linear' (i.e., of the form $m_i\sigma_i^2 + c_i$ where m_i, c_i are constants). Since this optimality for constant multipliers holds irrespective of the multiplier values, it continues to hold if *all* these multipliers are optimized. Zeroth order Wiener filters are the optimum multiplier choices, hence PCFBs are optimal when these are used in all subbands. This alternative proof however fails if some of the multipliers are not Wiener filters, e.g. they are other choices as mentioned in the earlier remark.

We summarize the above-mentioned results on PCFB optimality for noise reduction under

Theorem 2.6: Optimum FB based white noise suppression. In Fig. 2.6, let $s(n)$ be a CWSS(M) random process, and $\mu(n)$ be zero mean additive white noise that has variance η^2 and is uncorrelated to $s(n)$. Let $\mathbf{v} = (\sigma_0^2, \sigma_1^2, \ldots, \sigma_{M-1}^2)^T$ denote the subband variance vector corresponding to $s(n)$. Let each subband multiplier k_i be a zeroth order Wiener filter $k_i = \frac{\sigma_i^2}{\sigma_i^2+\eta^2}$. Consider the FB optimization problem of minimizing the average mean square error between the FB output $y(n)$ and the desired signal $s(n)$. This is equivalent to minimizing $g(\mathbf{v}) = \frac{1}{M}\sum_{i=0}^{M-1} h_i(\sigma_i^2)$, where $h_i(x) = \frac{x\eta^2}{x+\eta^2}$. As these h_i are all concave, a PCFB for $s(n)$ is optimal for this situation. This PCFB is also a PCFB for the input $x(n)$ since the noise is white. This optimality of the PCFB holds even with certain other choices of some or all of the subband multipliers k_i, namely subband hard thresholders (with threshold η^2) and constants (independent of choice of FB); since this merely changes the functional form of the corresponding h_i but preserves its concavity. ■

2.7.2 Subband Wiener filtering: An alternative approach

Since the subband processors studied above were LTI systems, it is possible to take a linear systems approach to the problem, as we elaborate here. While this approach does not prove Theorem 2.6 (derived above) in its entirety, it allows us to generalize some parts of its statement further. In particular, it allows certain extensions to cases when the noise is *colored* and the FBs are *biorthogonal* as opposed to orthonormal.

Consider the system of Fig. 2.8 where the bold-face vectors $\mathbf{s}(n), ^{-}(n), \mathbf{x}(n)$ and $\mathbf{y}(n)$ are all M-fold blocked versions of the corresponding scalar random processes $s(n), \mu(n), x(n)$ and $y(n)$, and the $\mathbf{D}(e^{j\omega})$ represents any $M \times M$ LTI system. We assume that $\mathbf{s}(n)$ and $^{-}(n)$ are uncorrelated WSS vector processes with psd matrices $\mathbf{S}_{\mathbf{ss}}(e^{j\omega})$ and $\mathbf{S}_{--}(e^{j\omega})$ respectively. The blocked version of the error is then $\mathbf{e}(n) = \mathbf{y}(n) - \mathbf{s}(n)$ which is WSS with psd matrix $\mathbf{S}_{\mathbf{ee}}(e^{j\omega})$ as follows: ($\mathbf{I}$ denotes the identity matrix)

$$\mathbf{S}_{\mathbf{ee}}(e^{j\omega}) = \left[\mathbf{A}(e^{j\omega}) - \mathbf{I}\right] \mathbf{S}_{\mathbf{ss}}(e^{j\omega}) \left[\mathbf{A}(e^{j\omega}) - \mathbf{I}\right]^\dagger + \mathbf{A}(e^{j\omega})\mathbf{S}_{--}(e^{j\omega})\mathbf{A}^\dagger(e^{j\omega}), \qquad (2.17)$$

$$\text{where} \quad \mathbf{A}(e^{j\omega}) = \mathbf{R}(e^{j\omega})\mathbf{D}(e^{j\omega})\mathbf{E}(e^{j\omega}), \quad \text{and of course} \quad \mathbf{R}(e^{j\omega}) = \mathbf{E}^{-1}(e^{j\omega}). \qquad (2.18)$$

To see this, note that $\mathbf{e}(n) = \mathbf{e}_s(n) + \mathbf{e}_\mu(n)$, where $\mathbf{e}_s(n), \mathbf{e}_\mu(n)$ are obtained by passing $\mathbf{s}(n), ^{-}(n)$

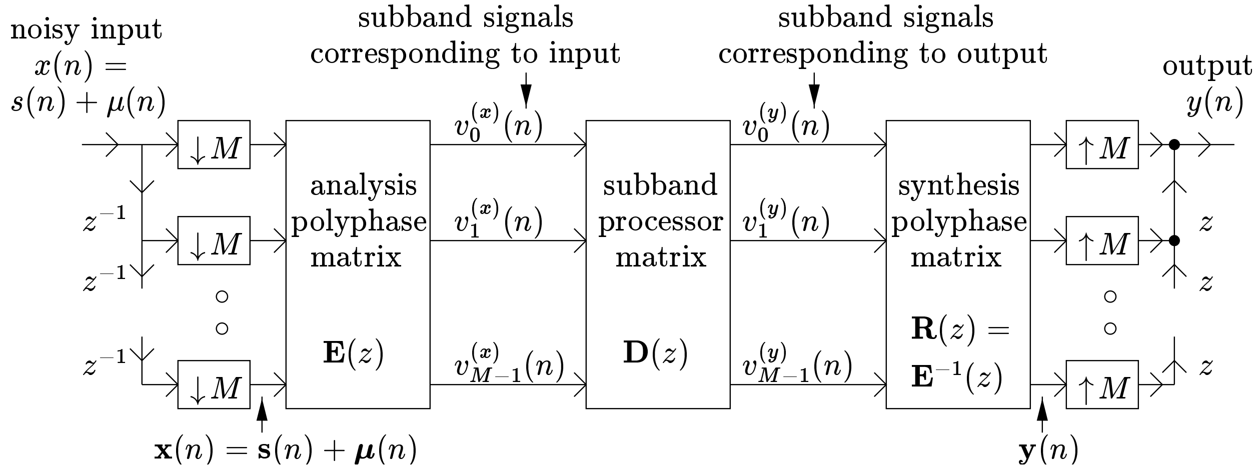


Figure 2.8: Subband noise reduction: System of Section 2.7.2.

through transfer matrices $\left[\mathbf{A}(e^{j\omega}) - \mathbf{I}\right]$ and $\mathbf{A}(e^{j\omega})$ respectively. Since $\mathbf{s}(n), {}^{-}(n)$ are uncorrelated WSS, so are $\mathbf{e}_s(n), \mathbf{e}_\mu(n)$; thus their sum is WSS with psd equal to the sum of their psds, each easy to compute. Note that (2.17),(2.18) do not assume orthonormality of the FB (i.e., that $\mathbf{E}(e^{j\omega})$ is unitary for all ω) or whiteness of the noise (i.e., that $\mathbf{S}^{--}(e^{j\omega})$ is the identity matrix). The average mean-square value of the error $e(n)$ is

$$\varepsilon = \frac{1}{M}\mathrm{trace}\left(\mathbf{\Sigma_{ee}}\right), \text{ where } \mathbf{\Sigma_{ee}} = \frac{1}{2\pi}\int_0^{2\pi} \mathbf{S_{ee}}(e^{j\omega})\,d\omega = \text{ autocorrelation matrix of } \mathbf{e}(n). \quad (2.19)$$

Memoryless $\mathbf{E}, \mathbf{D}, \mathbf{R}$

If the transfer matrices $\mathbf{E}(e^{j\omega})$,$\mathbf{D}(e^{j\omega})$,$\mathbf{R}(e^{j\omega})$ are all memoryless, then so is $\mathbf{A} = \mathbf{RDE}$, and

$$\mathbf{\Sigma_{ee}} = [\mathbf{A} - \mathbf{I}]\,\mathbf{\Sigma_{ss}}\,[\mathbf{A} - \mathbf{I}]^\dagger + \mathbf{A}\mathbf{\Sigma}^{--}\mathbf{A}^\dagger,$$

where $\mathbf{\Sigma_{ss}}, \mathbf{\Sigma}^{--}$ are autocorrelation matrices of $\mathbf{s}(n)$ and ${}^{-}(n)$ respectively. If $\mathbf{D}$ is unconstrained, so is $\mathbf{A}$; and the optimum $\mathbf{A}$ is simply the zeroth order vector Wiener filter for the noisy input $\mathbf{x}(n)$, i.e.,

$$\mathbf{A} = \mathbf{RDE} = \mathbf{\Sigma_{ss}}\left[\mathbf{\Sigma_{ss}} + \mathbf{\Sigma}^{--}\right]^{-1}. \quad (2.20)$$

Suppose the signal and noise have a *common* KLT; i.e., for some unitary $\mathbf{T}$ both $\mathbf{T\Sigma_{ss}T}^\dagger = \mathbf{\Lambda_{ss}}$ and $\mathbf{T\Sigma}^{--}\mathbf{T}^\dagger = \mathbf{\Lambda}^{--}$ are diagonal matrices. Then substitution in (2.20) shows that $\mathbf{A} = \mathbf{RDE} = \mathbf{T}^\dagger\mathbf{WT}$ where $\mathbf{W} = \mathbf{\Lambda_{ss}}\left[\mathbf{\Lambda_{ss}} + \mathbf{\Lambda}^{--}\right]^{-1}$ is diagonal. So the choice $\mathbf{E} = \mathbf{T}$ and $\mathbf{D} = \mathbf{W}$ is optimum under these conditions. Clearly, with this choice the diagonal elements of the (diagonal) matrix $\mathbf{D}$ are the scalar zeroth order Wiener filters for their corresponding inputs. Thus we have proved

Theorem 2.7: Optimum memoryless transform for subband Wiener filtering. In Fig. 2.6, let the pure signal $s(n)$ and the zero mean additive noise $\mu(n)$ be uncorrelated CWSS(M) random processes. The noise $\mu(n)$ could be colored. Let all the subband multipliers k_i be zeroth order Wiener filters for reducing the noise component in their respective input. Suppose there is a common KLT for the signal and noise, namely the unitary matrix $\mathbf{T}$. Then the choice $\mathbf{E}(z) = \mathbf{T}$ in Fig. 2.6 gives optimum noise reduction among all choices where $\mathbf{E}(z)$ is a constant matrix. In other words, the **common KLT** is the **optimum FB among all memoryless biorthogonal transforms** in the sense of maximizing the output signal to noise ratio. ■

Relation between Theorems 2.6 and 2.7. Theorem 2.6 proves optimality of a PCFB for a general class $\mathcal{C}$ of orthonormal FBs, for many white noise suppression problems where the subband multipliers could be any combination of Wiener filters, hard thresholds and constants. On the other hand, Theorem 2.7 focuses on the case when *all* subband multipliers are *Wiener filters*, and on a *special class* of FBs, namely the class $\mathcal{C}^b$ of all FBs with a *constant* (memoryless) polyphase matrix. Notice that $\mathcal{C}^b$ includes the orthogonal transform coder class $\mathcal{C}^t$. All Theorem 2.6 says about this case is that a signal KLT is the optimum FB *within* $\mathcal{C}^t$ when the noise is *white*. Notice that this FB is a common signal and noise KLT, since any orthogonal transform is a KLT for a white input. Thus Theorem 2.7 generalizes the result to the situation when the noise is *colored*, and also shows optimality of the common KLT among a larger class $\mathcal{C}^b$ of all memoryless *biorthogonal* transforms. In summary, Theorems 2.6 and 2.7 have a common element, which they generalize in different directions.

Further generalizations. Attempts to combine Theorems 2.6 and 2.7 yield many interesting further generalizations and open problems. For example, let us restrict attention to *orthogonal* transforms in Theorem 2.7. The common signal and noise PCFB (KLT), if it exists, can then be shown to be optimal even if the subband multipliers are any combination of Wiener filters, hard thresholds and constants (as opposed to all being Wiener filters as in Theorem 2.7). This result is shown in Chapter 3, using the convexity of certain search spaces associated with the signal and noise spectrum. As the input noise is colored, the subband noise variances are no longer constant but depend on choice of FB; hence the approach used to prove Theorem 2.6 needs some modifications. It also appears plausible that the above optimality of the common KLT extends to the class of *all memoryless biorthogonal* transforms.[4] Verifying this is however currently an open problem.

Case when $\mathbf{E}, \mathbf{D}, \mathbf{R}$ have memory; higher order subband Wiener filters.

Suppose the LTI systems $\mathbf{E}, \mathbf{D}, \mathbf{R}$ in Fig. 2.8 have memory. The FB optimization problem then involves choosing from the given class of analysis polyphase matrices $\mathbf{E}(e^{j\omega})$, the one minimizing the

[4] An analogous result is true for the high bitrate coding problem with optimal bit allocation (Section 2.5.3), i.e., the signal KLT is optimal over all memoryless biorthogonal transforms. This is proved using the Hadamard inequality for determinants.

error ε of (2.19) where $\mathbf{S}_{\mathbf{ee}}(e^{j\omega})$ is as in (2.17),(2.18), and $\mathbf{D}(e^{j\omega})$ is an appropriately constrained matrix. For example if N-th order Wiener filters are used in all subbands, then $\mathbf{D}(e^{j\omega})$ is a diagonal matrix depending in an involved manner on $\mathbf{E}(e^{j\omega})$. The FB optimization for such cases appears to be extremely involved, and we do not know any analytical results at this time. N-th order Wiener filters ($N > 0$) cannot be handled like zeroth order ones as in Section 2.7.1. This is because the minimization objective now depends on not just the subband variances, but on N more elements in the autocorrelation sequences of the subband random processes.

If *ideal* (infinite order) Wiener filters are used in each subband, an analog of Theorem 2.7 holds: In this case, any orthonormal FB whose polyphase matrix $\mathbf{E}(e^{j\omega})$ diagonalizes both the signal and noise psd matrices is optimal over the class of all unconstrained *biorthogonal* FBs. This result is obtained by repeating the methods used to prove Theorem 2.7 at each frequency ω. We may note that the optimal FB mentioned here need not be a PCFB for either the signal or the noise: There is no ordering constraint on the subband spectra, and diagonalization of the psd matrices is sufficient. For WSS (as opposed to CWSS(M)) input signal and noise, this is trivially achieved by any brickwall orthonormal FB (having non-overlapping analysis filters). The resulting system is equivalent to an ideal scalar Wiener filter acting directly on the scalar input without use of any FB. (This filter has simply been implemented in M disjoint pieces of the total frequency spectrum $[0, 2\pi]$.)

Thus, the direct *ideal* Wiener filter, as could be expected, represents the lower bound on the achievable mean square error using any FB based noise reduction scheme. Use of FBs yields improved performance only when there is a constraint on the complexity of the filtering. For example, suppose we compare direct zeroth order Wiener filtering of the noisy input against such filtering in each subband of a FB. Note that this is a fair comparison with respect to filtering complexity (excluding the cost of implementing the FB), as the M subband Wiener filters operate at $(1/M)$-th of the sampling rate of the direct filter. In this case, the FB based approach always yields a better performance.

2.8 Conclusion

We have pointed out a strong connection between the optimization of orthonormal filter banks and the principal component property. The main result is that a principal component filter bank (PCFB) is optimal whenever the minimization objective is a concave function of the vector consisting of the subband variances of the FB. We have shown various signal processing systems in which the FB optimization involves minimizing such a concave objective. In particular, the known results on optimality of PCFBs for compression can be explained in this manner. PCFBs are also shown to be optimal for subband domain white noise suppression using any combination of zeroth order Wiener filters and hard thresholds in the subbands. Some extensions have been made to biorthogonal FBs, and to the case when the noise is colored. We have also shown that the classes of ideal DFT and cosine-modulated FBs do not have PCFBs. Chapter 3 extends the study of FB optimization for colored noise suppression, and Chapter 4 extends the PCFB concept to classes of nonuniform FBs.

2.9 Appendices

Appendix A: Doubly stochastic matrices

Here we prove that all convex combinations and products of $M \times M$ doubly stochastic matrices are also doubly stochastic. It suffices to prove this for two matrices, since we can continue by induction. Define the vector $\mathbf{k} \in \mathcal{R}^M$ as $\mathbf{k} = (1, 1, \ldots, 1)^T$. Then by definition, an $M \times M$ matrix $\mathbf{Q}$ is doubly stochastic iff all its entries are non-negative, $\mathbf{Qk} = \mathbf{k}$ and $\mathbf{k}^T\mathbf{Q} = \mathbf{k}^T$. Now consider a convex combination $\mathbf{C} = \alpha\mathbf{A} + (1-\alpha)\mathbf{B}$ (where $0 \leq \alpha \leq 1$) and a product $\mathbf{D} = \mathbf{AB}$ of the $M \times M$ doubly stochastic matrices $\mathbf{A}$ and $\mathbf{B}$. It is required to show that $\mathbf{C}, \mathbf{D}$ are doubly stochastic. Clearly since $\mathbf{A}, \mathbf{B}$ have non-negative entries, so do $\mathbf{C}, \mathbf{D}$. Further,

$$\mathbf{Ck} = \alpha\mathbf{Ak} + (1-\alpha)\mathbf{Bk} = \alpha\mathbf{k} + (1-\alpha)\mathbf{k} = \mathbf{k}, \quad \text{and similarly} \quad \mathbf{k}^T\mathbf{C} = \mathbf{C}.$$

$$\text{Likewise,} \qquad \mathbf{Dk} = \mathbf{ABk} = \mathbf{Ak} = \mathbf{k}, \quad \text{and similarly} \quad \mathbf{k}^T\mathbf{D} = \mathbf{D}.$$

This completes the proof. It also shows that the set of all $M \times M$ doubly stochastic matrices is convex.

Appendix B: Are doubly stochastic matrices orthostochastic?

Evidently **every $M \times M$ orthostochastic matrix is doubly stochastic**. Here we show that the **converse is true if $M \leq 2$ but is false if $M > 2$**. The case $M = 1$ is trivial. For $M = 2$, a 2×2 doubly stochastic matrix must have form $\mathbf{Q} = \begin{bmatrix} p & 1-p \\ 1-p & p \end{bmatrix}$ with $0 \leq p \leq 1$. Now $p = \cos^2(\theta)$ for some real θ, so $\mathbf{Q}$ is indeed the orthostochastic matrix corresponding to the unitary matrix $\begin{bmatrix} \cos(\theta) & \sin(\theta) \\ -\sin(\theta) & \cos(\theta) \end{bmatrix}$. For $M = 3$, take the doubly stochastic $\mathbf{A} = \frac{1}{2}\begin{bmatrix} 1 & 1 & 0 \\ 1 & 0 & 1 \\ 0 & 1 & 1 \end{bmatrix}$. If $\mathbf{A}$ was the orthostochastic matrix corresponding to $\mathbf{U}$ then $\mathbf{U} = \begin{bmatrix} a & b & 0 \\ c & 0 & d \\ 0 & e & f \end{bmatrix}$ for some nonzero a, b, c, d, e, f. Thus $\mathbf{U}$ cannot be unitary as no two of its rows can be orthogonal to each other. So $\mathbf{A}$ is not orthostochastic. Small perturbations of the entries of $\mathbf{A}$ can create other such examples. The doubly stochastic matrix $\begin{bmatrix} \mathbf{A} & \mathbf{0} \\ \mathbf{0} & \mathbf{I} \end{bmatrix}$ gives examples for $M > 3$, where $\mathbf{0}, \mathbf{I}$ are respectively the zero and identity matrices of suitable size. This concludes the proof. We may note here that the set $\mathcal{O}_M$ of $M \times M$ orthostochastic matrices is convex if $M \leq 2$ (as it is then the set of $M \times M$ doubly stochastic matrices), but is *not* convex if $M > 2$. This is because all $(M \times M)$ permutation matrices are in $\mathcal{O}_M$, and every doubly stochastic matrix is a convex combination of these matrices (Birkhoff's theorem). So if $\mathcal{O}_M$ were convex, it would contain all doubly stochastic matrices, but it does not if $M > 2$.

Chapter 3 Colored noise reduction, and PCFB existence issues

This chapter begins by studying the nature of the FB optimizations of Chapter 2 in situations when PCFBs do not exist (Section 3.1). We show that a PCFB exists if and only if there is a single FB that simultaneously minimizes all concave functions of the subband variances. By studying the structure of the search space associated with the optimizations, we show exactly why they are usually analytically intractable in absence of a PCFB. We explain the relation between compaction filter design (or variance maximization) and FB optimization: A sequential maximization of subband variances always yields a PCFB if it exists, but is suboptimum for large classes of concave objectives if a PCFB does not exist.

Next, we examine in detail the problem of FB optimization and PCFB optimality for *colored* noise suppression (Section 3.2). The system is identical to that in Fig. 2.6, but the noise $\mu(n)$ is no longer assumed to be white. With white noise, the minimization objective g is a function of only the signal subband variances. The signal PCFB is optimal if g is concave (Chapter 2). With colored noise however, the objective depends on both the signal and noise subband variances, so the results of Chapter 2 no longer hold. We show here that for the *transform coder* class, if a *common* signal and noise PCFB (KLT) exists, it minimizes a large class of concave objectives. Common PCFBs for a general FB class do not have such optimality, as we show using the unconstrained FB class $\mathcal{C}^u$. We show how to find the optimum FB in $\mathcal{C}^u$ for certain piecewise constant input spectra. We conclude with some open problems, especially on biorthogonal FBs and PCFB existence. The content is drawn mainly from [8], and has been presented in preliminary form in [5],[6].

3.1 What if there is no PCFB?

When a PCFB exists, the search–space $\mathcal{S}$ consisting of all realizable subband variance vectors has a very special structure: Its convex hull $\text{co}(\mathcal{S})$ is a polytope whose extreme points are all permutations of the PCFB subband variance vector (Theorem 2.2). The optimality of PCFBs under concave objectives (Theorem 2.3) follows from this structure and the optimality of extreme points of polytopes (Theorem 2.1). If a PCFB does not exist, $\mathcal{S}$ does not have this structure. Thus $\text{co}(\mathcal{S})$ is a general convex set. For such sets too, there is a notion of extreme points, which coincides with the usual definition when the convex sets are polytopes, and further allows the following generalization of Theorem 2.1: If a function f is concave over a compact convex domain D, at least one extreme

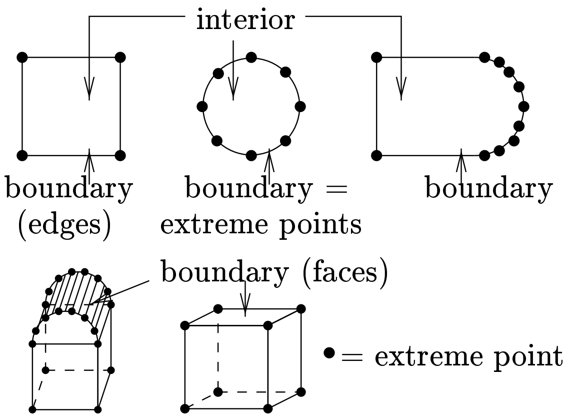


Figure 3.1: Extreme points of compact convex sets.

point of D is a minimum of f over D. Thus in this case, to minimize f over D it suffices to minimize f over the extreme points of D. Polytopes are exactly the compact convex sets having *finitely many* extreme points.

This section uses these observations to study the effect of nonexistence of PCFBs on the FB optimizations. When a PCFB exists, all the (finitely many) extreme points of the set co($\mathcal{S}$) correspond to the PCFB. So the PCFB is always optimal for *all* concave minimization objectives. On the other hand if a PCFB does not exist, co($\mathcal{S}$) could in general have infinitely many extreme points. This explains the analytical intractability of many FB optimizations when PCFBs do not exist. Finally, we explain the relation between PCFBs and 'compaction filters' that maximize their output variance among certain classes of filters.

3.1.1 Arbitrary convex sets: Extreme points and their optimality

Definition [29]. For a convex set $B \subset \mathcal{R}^M$, a point $\mathbf{z} \in B$ is said to be an *extreme point*, or a *corner* of B if

$$\mathbf{z} = \alpha\mathbf{x} + (1-\alpha)\mathbf{y} \quad \text{with} \quad \alpha \in (0,1),\ \mathbf{x},\mathbf{y} \in B \quad \text{implies} \quad \mathbf{x} = \mathbf{y}\,(= \mathbf{z}).$$

Geometrically, no line-segment passing *through* $\mathbf{z}$ (i.e., containing $\mathbf{z}$ but not as an endpoint) can lie wholly in the set B. The interior of B cannot have any extreme points, since around each point in the interior there is a ball lying wholly in B. So all extreme points lie on the boundary, though all boundary points need not be extreme points. If B is a polytope, the above definition can be verified to coincide with the usual definition of extreme points of a polytope. Figure 3.1 illustrates these facts, showing the extreme points of some closed and bounded (or compact) convex sets. It is not hard to show that every (nonempty) compact convex set is the convex hull of its boundary, and that it has at least one extreme point. A stronger result is true:

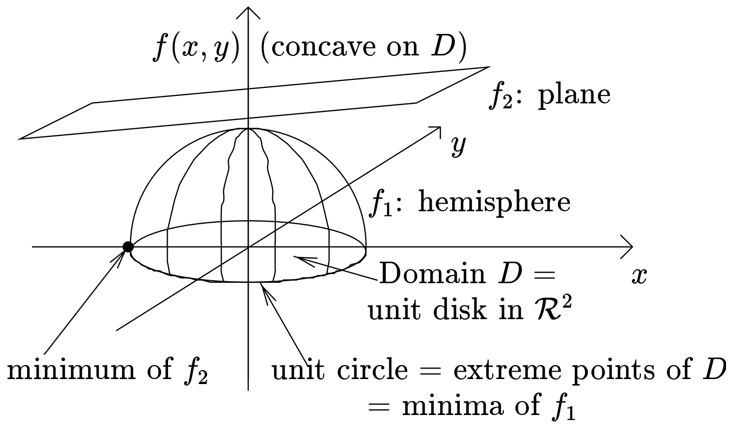


Figure 3.2: Optimality of extreme points of compact convex sets (Theorem 3.1).

Krein–Milman theorem (Internal representation of convex sets) [29, 50]: Every compact convex set D is the convex hull of its extreme points. Hence the set of extreme points of D is the minimal subset of D having D as its convex hull. This fact can serve as an equivalent definition of extreme points of compact convex sets.

This result evidently holds for polytopes, and is verifiable in the examples of Fig. 3.1. Thus it is intuitive, though its formal proof [50] may not be trivial. It is important as it immediately proves the following:

Theorem 3.1: Optimality of extreme points. If a function g is concave on a compact convex set D, at least one of the extreme points of D is a minimum of g over D. Further if g is *strictly* concave, its minimum *has to be* at an extreme point of D. ■

This result reduces to Theorem 2.1 if D is a polytope, and is illustrated in Fig. 3.2 for a compact convex D that is *not* a polytope.

Proof: Let $\mathbf{v}_{opt}$ minimize g over D. (Existence of $\mathbf{v}_{opt}$ is either assumed or follows if g is assumed continuous.) By the Krein–Milman theorem, $\mathbf{v}_{opt}$ is a convex combination of some extreme points of D, i.e.,

$$\mathbf{v}_{opt} = \sum_{j=1}^{J} \beta_j \mathbf{z}_j \quad \text{where} \quad 0 \leq \beta_j \leq 1, \ \sum_{j=1}^{J} \beta_j = 1$$

for some distinct extreme points $\mathbf{z}_j$ of D. If none of these $\mathbf{z}_j$ minimizes g over D, $g(\mathbf{z}_j) > g(\mathbf{v}_{opt})$ for all j, so

$$g(\mathbf{v}_{opt}) = g(\sum_{j=1}^{J} \beta_j \mathbf{z}_j) \geq \sum_{j=1}^{J} \beta_j g(\mathbf{z}_j) > \sum_{j=1}^{J} \beta_j g(\mathbf{v}_{opt}) = g(\mathbf{v}_{opt}),$$

i.e., $g(\mathbf{v}_{opt}) > g(\mathbf{v}_{opt})$, a contradiction. Hence at least one extreme point of D is a minimum of g over D. If g is strictly concave, the first inequality above (Jensen's inequality) is strict unless $\beta_j = 1$ for some j, hence $\mathbf{v}_{opt} = \mathbf{z}_j$, i.e., the minimum is necessarily at an extreme point of D. $\nabla\nabla\nabla$

3.1.2 Filter bank optimization and extreme points of convex sets

In our FB optimizations, the objective is concave on the set $\text{co}(\mathcal{S})$ where $\mathcal{S}$ is the search–space. We seek its minima over $\mathcal{S}$. We assume from now on that $\mathcal{S}$ (and hence $\text{co}(\mathcal{S})$) is compact. This is true for most input power spectra and practical FB classes (Appendix A), and allows use of Theorem 3.1. Let E be the set of extreme points of $\text{co}(\mathcal{S})$. From Theorem 3.1, for any concave objective over $\text{co}(\mathcal{S})$, at least one of its minima lies in E (and all them do if the concavity is strict). From the definition of extreme points, we can show that $E \subset \mathcal{S}$. So the minima over $\text{co}(\mathcal{S}) \supset \mathcal{S}$ found by minimizing over E in fact lie in $\mathcal{S}$, and are hence minima over $\mathcal{S}$ too. Thus, minimization over $\mathcal{S}$ has been reduced to one over the set E of extreme points of $\text{co}(\mathcal{S})$. Now for 'almost every' extreme point $\mathbf{z}$ in E there is a concave (in fact linear) function that is minimized over $\text{co}(\mathcal{S})$ uniquely by $\mathbf{z}$.[1] So without further information on the specific concave objective, nothing can be said about its minima over $\mathcal{S}$ apart from the fact that a search over E will yield at least one of them.

When a PCFB exists, *all* points in E correspond to it. This explains the remarkable optimality of PCFBs for *all* concave objectives. If there is no PCFB, E has at least two points that are not permutations of each other, i.e., that correspond to essentially different FBs. Thus, no single FB can be simultaneously optimal for all concave objectives f. If E is finite, the optimal FB for any given concave f can still be found by a finite exhaustive search over E. Unfortunately in general there is no reason to expect E to be finite, hence a numerical search is required. Any derivation of analytical results on the optimum FB will have to take into account the specific nature of both the concave objective at hand and the set E (which depends on the FB class $\mathcal{C}$ and input psd at hand). This explains why these optimizations are usually analytically intractable.

3.1.3 The sequential compaction algorithm

This is an algorithm that has sometimes been proposed [61],[45] to find a 'good' FB in classes $\mathcal{C}$ that may not have PCFBs. We first state the algorithm in a precise manner that holds for any general class $\mathcal{C}$. We then show that it produces FBs for which the corresponding subband variance vector *is an extreme point of* $\text{co}(\mathcal{S})$. We examine the optimality of the algorithm in this light.

Let $\mathcal{C}$ be the given class of FBs, and $\mathcal{S}$ the corresponding optimization search–space. The algorithm involves rearranging all vectors in $\mathcal{S}$ in decreasing order of their entries, and then picking from these the vector $\mathbf{v}_\alpha \in \mathcal{S}$ defined as the greatest one in the 'dictionary ordering' on $\mathcal{R}^M$. This means that the greatest (first) entry of $\mathbf{v}_\alpha$ is greater than or equal to the greatest entry of any of the other vectors. Among vectors for which equality prevails, the second greatest entry of $\mathbf{v}_\alpha$ is greater than or equal to the second greatest entry of the other vectors, and so on. The output of

[1] This is because for any compact convex set D, the set of extreme points is the closure of the set of *exposed* points [50], which by definition are points $\mathbf{v} \in D$ for which there is a linear function minimized (or maximized) over D uniquely by $\mathbf{v}$.

the algorithm is any FB with subband variance vector $\mathbf{v}_\alpha$ (or any of its permutations). The vector $\mathbf{v}_\alpha$ is well-defined, and finding it involves a sequential maximization of subband variances giving the algorithm its name. (Existence of the maxima follows from compactness of $\mathcal{S}$.)

Relation to compaction filters. The *ideal compaction filter* [61] for an input process is defined as the filter maximizing its output variance among all filters $H(e^{j\omega})$ whose magnitude squared $\left|H(e^{j\omega})\right|^2$ is Nyquist(M). The Nyquist(M) constraint is imposed because these filters are used to build an orthonormal M channel FB, and any filter in such a FB obeys this constraint [64]. For WSS inputs, a procedure from [61] finds the compaction filter given the input psd. It always yields a 'brickwall' filter, i.e., one with constant magnitude on its support. If such a filter is to be an analysis filter in an orthonormal FB, its support cannot overlap with that of any other analysis filter. Thus the FB can be built by a sequential design of compaction filters: The next filter maximizes its output variance among all filters that have a Nyquist(M) magnitude squared and a support that does not overlap with the supports of the previously designed filters.[2]

This FB design method from [61] is exactly the sequential algorithm described above, applied to the unconstrained FB class $\mathcal{C}^u$ when the input is WSS (as distinct from CWSS(M)). The variance maximization in the algorithm corresponds to an ideal compaction filter design. This connection has motivated the study and design of *FIR compaction filters* [38]. These are defined as filters maximizing their output variance among all filters of order not exceeding N whose magnitude squared is Nyquist(M). It was believed that such filters would play a role in PCFB design for the class $\mathcal{C}^{fir}$ of all M channel orthonormal FBs in which all filters are FIR with order not exceeding N $(N \geq M)$. Indeed, it may seem that the first step in the sequential algorithm for the class $\mathcal{C}^{fir}$ is to design a FIR compaction filter. However this is not true for a general M and input psd, as there may not even be a FB in $\mathcal{C}^{fir}$ having the FIR compaction filter as one of its filters. The correct first step in the sequential algorithm for a general FB class $\mathcal{C}$ is to design a filter maximizing its output variance *among all filters belonging to FBs in* $\mathcal{C}$. It seems quite infeasible to propose any variant of the sequential algorithm or the class $\mathcal{C}^{fir}$ in which FIR compaction filters will play any serious role. The only notable exception is when $M = 2$, where the FB is fully determined by any one of its filters. Thus, a clear relation between the sequential algorithm and compaction filters exists only for the unconstrained class $\mathcal{C}^u$ when the input is WSS (as opposed to CWSS(M)).

3.1.4 Is the sequential algorithm optimal?

The optimality properties of the sequential algorithm of Section 3.1.3 follow easily from the following:
Assertion 3.1. The subband variance vector $\mathbf{v}_\alpha = (\alpha_0, \alpha_1, \ldots, \alpha_{M-1})^T \in \mathcal{S}$ (with $\alpha_0 \geq \ldots \geq \alpha_{M-1}$) produced by the sequential algorithm **is an extreme point of** $\mathrm{co}(\mathcal{S})$.

[2] Equivalently, it is an ideal compaction filter for the psd that is obtained by setting to zero the bands of the original input psd falling within the supports of the previously designed filters.

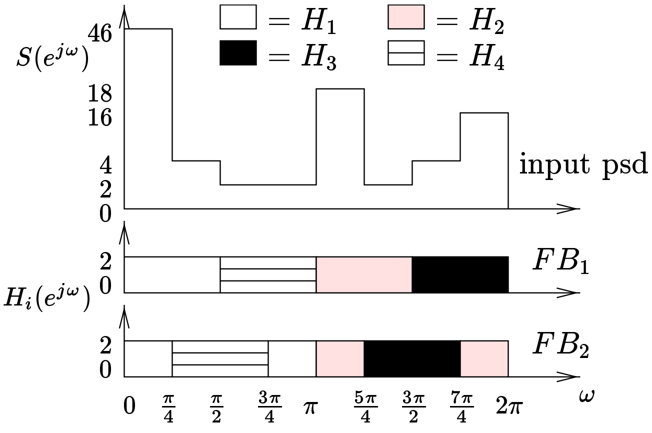


Figure 3.3: Suboptimality of sequential compaction algorithm.

Proof. Let $\mathbf{v}_\alpha = \gamma\mathbf{x} + (1-\gamma)\mathbf{y}$ for $\gamma \in (0,1)$ and $\mathbf{x}, \mathbf{y} \in \mathrm{co}(\mathcal{S})$. By definition of an extreme point (Section 3.1.1), showing that $\mathbf{x} = \mathbf{y} = \mathbf{v}_\alpha$ will complete the proof. Now by definition of the convex hull $\mathrm{co}(\mathcal{S})$, we see that $\mathbf{x}, \mathbf{y}$ and hence $\mathbf{v}_\alpha$ can be written as convex combinations of elements of $\mathcal{S}$, i.e., $\mathbf{v}_\alpha = \sum_{j=1}^{J} \beta_j \mathbf{v}^j$ for some $\mathbf{v}^j = (v_0^j, v_1^j, \ldots, v_{M-1}^j) \in \mathcal{S}$ and $\beta_j \in (0,1]$ satisfying $\sum_{j=1}^{J} \beta_j = 1$. We now show $\mathbf{x} = \mathbf{y} = \mathbf{v}_\alpha$ by showing $\mathbf{v}^j = \mathbf{v}_\alpha$ for all $j = 1, 2, \ldots, J$. To this end, since $\mathbf{v}_\alpha$ exceeds (or equals) all the $\mathbf{v}^j$ in the dictionary ordering on $\mathcal{R}^M$, we have $\alpha_0 \geq v_0^j$, but α_0 is a convex combination of the v_0^j. Hence $\alpha_0 = v_0^j$ for all j. This in turn leads to $\alpha_1 \geq v_1^j$, and hence to $\alpha_1 = v_1^j$ and so on; until finally $\mathbf{v}_\alpha = \mathbf{v}^j$ for all $j = 1, 2, \ldots, J$. $\bigtriangledown\bigtriangledown\bigtriangledown$

When the class $\mathcal{C}$ has a PCFB, all extreme points of $\mathrm{co}(\mathcal{S})$ correspond to the PCFB. Hence the sequential algorithm always yields the PCFB and is thus optimal for many problems (Chapter 2). The subband variance vector $\mathbf{v}_\alpha$ produced by the algorithm here has an additional property: If its entries are arranged in *increasing order*, then in fact it becomes the *least* vector in $\mathcal{S}$ in the dictionary ordering.[3] On the other hand, if a PCFB does not exist, then there will be at least two extreme points that do not correspond to essentially the same FB, i.e., whose coordinates are not permutations of each other. The algorithm of Section 3.1.3 produces one extreme point, but the minimum could easily be at another one. Thus the algorithm could be suboptimum.

The following hypothetical example with $M = 4$ channels illustrates this point: Let $\mathrm{co}(\mathcal{S}) = \mathrm{co}(E)$ where E is the set of all permutations of $\mathbf{v}_1 = (25, 10, 10, 2)^T$ and $\mathbf{v}_2 = (24, 17, 3, 3)^T$. This would happen for a WSS input with psd shown in Fig. 3.3, when the class $\mathcal{C}$ has exactly the two FBs in the figure. As E is finite, $\mathrm{co}(\mathcal{S})$ is a polytope whose extreme points lie in E. In fact all points in E are extreme points of $\mathrm{co}(\mathcal{S})$ as neither of $\mathbf{v}_1, \mathbf{v}_2$ *majorizes* the other. A PCFB does not exist, as $\mathbf{v}_1$ is not a permutation of $\mathbf{v}_2$. Now consider the high bitrate coding problem of [61]. Here the objective to

[3] However, the fact that $\mathbf{v}_\alpha$ has this property does not imply that a PCFB exists, unless the number of channels is $M \leq 3$. *Majorization* is a stronger requirement. For example, $\mathbf{v}_1 = (25, 10, 10, 2)$ exceeds $\mathbf{v}_2 = (24, 17, 3, 3)$ and its permutations, and also becomes less than them if its entries are rearranged in increasing order; but still $\mathbf{v}_1$ does not majorize $\mathbf{v}_2$.

be minimized over $\mathcal{S}$ is $\pi(\mathbf{v})$, the geometric mean of the entries of $\mathbf{v} \in \mathcal{S}$. (As noted in Section 2.5.3, this is equivalent to minimizing an objective that is concave on $\text{co}(\mathcal{S})$.) Now $[\pi(\mathbf{v}_1)]^4 = 5000 > [\pi(\mathbf{v}_2)]^4 = 3672$, so $\mathbf{v}_2$ is the minimum. However, the algorithm of Section 3.1.3 yields $\mathbf{v}_\alpha = \mathbf{v}_1$, and is thus suboptimum. Further it remains so even if it is run to sequentially minimize rather than maximize variances (again, giving in general some extreme point of $\text{co}(\mathcal{S})$, in this case $\mathbf{v}_1$).

In fact one can even create a family of (necessarily artificial) concave objectives that the algorithm actually maximizes instead of minimizing. Let $P \subseteq \text{co}(\mathcal{S})$ be the polytope with extreme points as permutations of the vector $\mathbf{v}_\alpha$ output by the algorithm, so $P = \text{co}(\mathcal{S})$ iff a PCFB exists. Let $f(\mathbf{v}) = -d(\mathbf{v}, P)$, where $d(\mathbf{v}, P) = \min\{\|\mathbf{v} - \mathbf{x}\| : \mathbf{x} \in P\}$ is the minimum distance from $\mathbf{v}$ to P (well–defined since P is compact) using any valid norm $\|\cdot\|$ on $\mathcal{R}^M$. Now f is continuous and concave on $\mathcal{R}^M$ (Appendix B). Its definition shows that f is constant (zero) on P, and that if a PCFB does not exist, P is actually the set of *maxima* of f over $\text{co}(\mathcal{S})$. Thus FBs with subband variance vector $\mathbf{v}_\alpha \in P$ or its permutations (output by the sequential algorithm) perform the worst. Even if these examples may seem artificial, they should convince the reader of the total absence of intrinsic connection between FB optimality and variance maximization / compaction filter design *except if a PCFB exists*, in which case the sequential algorithm yields exactly the PCFBs.

3.2 Optimum FBs for colored noise suppression

In Section 2.7, we had studied FB optimization for the noise reduction system of Fig. 1.3, where the noise $\mu(n)$ was assumed white. Here we study the same system without making this assumption. The problem is then significantly more complicated, because the optimization objective now depends on both signal and noise subband variances. We will begin by describing the general form of the optimization objective, and relating it to the special case studied in Chapter 2. We then define various search spaces akin to the set $\mathcal{S}$ of Section 2.3.1, that are useful in solving the problem. The main results are presented in Section 3.2.3, and are proved in the later sections. All results here deal exclusively with orthonormal FB optimization.

3.2.1 Form of objective, applicable problems, and relation to Chapter 2

Suppose the input $v_i^{(x)}(n)$ to each subband processor P_i in Fig. 2.1 contains a signal component $v_i^{(s)}(n)$ and a zero mean noise component $v_i^{(\mu)}(n)$, which are uncorrelated to each other. Their variances are denoted by σ_i^2, η_i^2 respectively, and all the $v_i^{(s)}(n)$ and $v_i^{(\mu)}(n)$ are assumed jointly WSS. The subband processors P_i are multipliers k_i which may or may not depend on the input statistics. The system aims to reject the noise components in the subbands, i.e., the desired output of P_i is $v_i^{(d)}(n) = v_i^{(s)}(n)$.

Following the analysis of Sections 2.3 and 2.7, we see that the subband errors $v_i^{(e)}(n)$ of (2.1) are all jointly WSS, and the minimization objective, i.e., the mean square error (ε of (2.1)) between the true and desired FB output, is of the form

$$f(\mathbf{v}_\sigma, \mathbf{v}_\eta) = \frac{1}{M} \sum_{i=0}^{M-1} f_i(\sigma_i^2, \eta_i^2). \tag{3.1}$$

Here $\mathbf{v}_\sigma = (\sigma_0^2, \sigma_1^2, \ldots, \sigma_{M-1}^2)^T$, $\mathbf{v}_\eta = (\eta_0^2, \eta_1^2, \ldots, \eta_{M-1}^2)^T$ are respectively the signal and noise subband variance vectors, and

$$f_i(x,y) = \begin{cases} |1-k_i|^2 x + |k_i|^2 y & \text{for constant multiplier } k_i \\ \frac{xy}{x+y} & \text{for 0-th order Wiener filter } k_i \ \left(= \frac{\sigma_i^2}{\sigma_i^2+\eta_i^2}\right) \\ \min(x,y) & \text{for hard threshold } k_i \ \left(= \begin{cases} 1 & \text{if } \sigma_i^2 \geq \eta_i^2 \\ 0 & \text{otherwise} \end{cases} \right) \end{cases} \tag{3.2}$$

Notice that the functional form of f_i is independent of choice of FB, and depends only on the type of subband multiplier k_i. Also, all the functions f_i of (3.2) are concave on $\mathcal{R}_+^2$ (Appendix A). We mention two ways in which the above noise suppression problem could arise:

1. When the FB input is $x(n) = s(n)$, and the corresponding subband signals $v_i^{(s)}(n)$ are transmitted on separate communication lines. Here $v_i^{(\mu)}(n)$ represents the noise in the i-th line.

2. When the FB input is $x(n) = s(n) + \mu(n)$ where $s(n)$ is the pure signal desired at the FB output, and $\mu(n)$ is zero mean additive noise uncorrelated to $s(n)$. Both $s(n)$ and $\mu(n)$ are assumed CWSS(M) random processes. Thus $v_i^{(\mu)}(n)$ is the i-th subband signal corresponding to $\mu(n)$. This is the main problem of interest later in this section.

The objectives studied in Chapter 2 had the form (2.3), which we repeat here:

$$g(\mathbf{v}) = \frac{1}{M} \sum_{i=0}^{M-1} h_i(\sigma_i^2), \tag{3.3}$$

where $\mathbf{v} = (\sigma_0^2, \sigma_1^2, \ldots, \sigma_{M-1}^2)^T$ is the subband variance vector, and the function h_i has a form that depends only on the type of the i-th subband processor and not on the choice of FB. In Chapter 2 we saw that if all the h_i are concave (on $\mathcal{R}_+$), a PCFB for the signal to which the subband variances σ_i^2 correspond will be optimal. We now study cases where the objective form (3.1),(3.2) for the two problems mentioned earlier in this section reduces to the above form (3.3).

When the $v_i^{(\mu)}(n)$ are communication line noises: Here, as long as the noise variances η_i^2 do not depend on the choice of FB, the form (3.3) holds with $h_i(x) = f_i(x, \eta_i^2)$, which is concave on $\mathcal{R}_+$ for the f_i of (3.2). If η_i^2 depends on the FB, then we do not in general have the form (3.3) as h_i then depends on the FB. Even here however, in some special cases this dependence can be accounted for by a modified h_i that is independent of the FB. For example, suppose the noise $v_i^{(\mu)}(n)$ arises due to quantization of the signal $v_i^{(s)}(n)$. By the usual quantizer model (2.10), $\eta_i^2 = f_i(b_i)\sigma_i^2$, which depends on the FB only through the subband signal variance σ_i^2. Substituting this in the expressions for the error variances $E[|v_i^{(e)}(n)|^2]$ shows that the form (3.3) still holds, with the modified h_i still being concave on $\mathcal{R}_+$ and given by

$$h_i(x) = \begin{cases} \left[|1-k_i|^2 + |k_i|^2 f_i(b_i)\right] x & \text{for constant multiplier } k_i \\ \left[\frac{f_i(b_i)}{1+f_i(b_i)}\right] x & \text{for 0-th order Wiener filter } k_i \\ \left[\min(1, f_i(b_i))\right] x & \text{for hard threshold } k_i \end{cases} \tag{3.4}$$

When the $v_i^{(\mu)}(n)$ come from additive noise $\mu(n)$ at the FB input: Let $\mathbf{S}_{ss}(e^{j\omega}), \mathbf{S}_{\mu\mu}(e^{j\omega})$ be the psd matrices of the M-fold blocked versions of $s(n)$ and $\mu(n)$ respectively. The subband signal variances σ_i^2 depend on $\mathbf{S}_{ss}(e^{j\omega})$ and the FB, and similarly the noise variances η_i^2 depend on $\mathbf{S}_{\mu\mu}(e^{j\omega})$ and the FB. Thus, in general, the objective form (3.1) cannot be reduced to (3.3) which depends only on one set of subband variances. However, such a reduction is possible for certain special types of $\mathbf{S}_{ss}(e^{j\omega}), \mathbf{S}_{\mu\mu}(e^{j\omega})$. Examples are as follows:

1. When the input noise $\mu(n)$ is *white*, i.e., $\mathbf{S}_{\mu\mu}(e^{j\omega}) = \eta^2\mathbf{I}$, in which case $\eta_i^2 = \eta^2$ for all i independent of choice of FB (deducible, e.g., from (2.4)). Thus, (3.3) holds with $h_i(x) = f_i(x, \eta^2)$ (which is concave); this is the case studied in Section 2.7, where a PCFB for the signal $s(n)$ is optimal. Notice that this PCFB is also a PCFB for the noise $\mu(n)$ (and in fact even for $x(n) = s(n) + \mu(n)$), since any FB is a PCFB for a white input.

2. When $\eta_i^2 = c\sigma_i^2$ for some c independent of FB, in which case (3.3) holds with the h_i modified in a manner very similar to that in (3.4). Again the new h_i are concave, and a PCFB for $s(n)$ is optimal. In this case too, this PCFB is also a PCFB for $\mu(n)$ and $x(n)$: The definition of PCFBs and the relation $\eta_i^2 = c\sigma_i^2$ ensures that a PCFB for any one of the signals $s(n), \mu(n), x(n)$ is also a PCFB for the others. This case happens when $\mathbf{S}_{\mu\mu}(e^{j\omega}) = c\mathbf{S}_{ss}(e^{j\omega})$. If all FBs in the given class $\mathcal{C}$ (over which we are seeking the optimum FB) have memoryless polyphase matrices, it suffices that the corresponding autocorrelation matrices $\mathbf{R}_{ss}, \mathbf{R}_{\mu\mu}$ obey $\mathbf{R}_{\mu\mu} = c\mathbf{R}_{ss}$. These are of course very contrived situations. Even if we did have $\mathbf{S}_{\mu\mu}(e^{j\omega}) = c\mathbf{S}_{ss}(e^{j\omega})$, we would then rather not use FB based schemes at all: The ideal (infinite order) Wiener filter in this case reduces to a constant $\frac{1}{1+c}$, and as noted at the end of Section 2.7.2, yields the best possible performance a subband filtering based scheme can attain.

In the rest of this section, we will study the colored noise suppression problem (where the $v_i^{(\mu)}(n)$ come from additive noise at the FB input). We have seen two special cases where its objective reduces to the form (3.3) which has been studied in Chapter 2. In both these cases, we have seen that the optimum solution turns out to be a *common* PCFB for the signal $s(n)$ and then noise $\mu(n)$. **Is this true in greater generality?** We answer this question in detail. We show that *for the transform coder class* $\mathcal{C}^t$, the common PCFB (if it exists) is indeed optimal; while the same is not always true for other classes of FBs, specifically for the unconstrained FB class $\mathcal{C}^u$. We also show how to find the optimum FB in $\mathcal{C}^u$ when the input signal and noise spectra are both *piecewise constant* with all discontinuities at rational multiples of π.

3.2.2 Notation and study of search–spaces

To study the issues mentioned above, we need notations for certain sets associated with the optimization problem. We now introduce these notations, which will hold throughout Section 3.2.

1. *Signal and noise variance spaces* S_σ, S_η. The set of all realizable subband signal variance vectors $\mathbf{v}_\sigma$ is denoted by S_σ. Similarly, the set of all realizable subband noise variance vectors $\mathbf{v}_\eta$ is denoted by S_η.

2. *Optimization search space* S_v. We denote by S_v the set of all realizable *pairs* of signal and noise variance vectors $\begin{pmatrix} \mathbf{v}_\sigma \\ \mathbf{v}_\eta \end{pmatrix}$. The minimization objectives for the problems studied here have the form (3.1), i.e., they are real-valued functions on S_v. Thus S_v is the 'search–space' for these problems, just as S_σ is for those of Chapter 2 (see Section 2.3.1). As both $\mathbf{v}_\sigma$ and $\mathbf{v}_\eta$ have entries whose sum is independent of the FB, the set S_v is bounded and lies on a $2M-2$ dimensional hyperplane in $\mathcal{R}_+^{2M}$. It also has a permutation symmetry, slightly different from that of S_σ but arising from the same reason (see Section 2.3.1). It is expressed as $\begin{pmatrix} \mathbf{v}_\sigma \\ \mathbf{v}_\eta \end{pmatrix} \in S_v \Rightarrow \begin{pmatrix} \mathbf{P}\mathbf{v}_\sigma \\ \mathbf{P}\mathbf{v}_\eta \end{pmatrix} \in S_v$ for any permutation matrix $\mathbf{P}$. Also, $\mathbf{v}_\sigma \in S_\sigma, \mathbf{v}_\eta \in S_\eta$ does not always imply $\begin{pmatrix} \mathbf{v}_\sigma \\ \mathbf{v}_\eta \end{pmatrix} \in S_v$; i.e., S_v is some subset of the Cartesian product $S_\sigma \times S_\eta$, usually a proper subset.[4] We also assume S_v (and hence $\mathrm{co}(S_v)$) to be compact, for similar reasons as in Section 3.1.2.

3. *Objective function domain* T. We study general minimization objectives concave over the set

$$T \triangleq \mathrm{co}(S_\sigma) \times \mathrm{co}(S_\eta) = \mathrm{co}(S_\sigma \times S_\eta) \supset \mathrm{co}(S_v). \tag{3.5}$$

(We have used above, and will freely use the set identity $\mathrm{co}(A \times B) = \mathrm{co}(A) \times \mathrm{co}(B)$.) Note that if all the f_i in (3.1) are concave on $\mathcal{R}_+^2$, the objective f of (3.1) is concave on $\mathcal{R}_+^{2M}$ and hence on T. Also, the f_i of (3.2) arising for the noise suppression problems above are indeed concave

[4] $S_v = S_\sigma \times S_\eta$ only in artificial / degenerate cases, e.g. if $\mu(n)$ (or $s(n)$) is white. (For white $\mu(n)$, S_η has only one element.)

on $\mathcal{R}_+^2$ (Appendix B). We know that minimizing a concave function over S_v is reducible to minimizing it over the set of extreme points of $\text{co}(S_v)$ (Section 3.1). So we will try to study the structure of this set of extreme points.

4. *Extreme point sets* E_σ, E_η, E_v. We denote by $E_\sigma \subset S_\sigma$, $E_\eta \subset S_\eta$, $E_v \subset S_v$ the sets of extreme points of $\text{co}(S_\sigma), \text{co}(S_\eta), \text{co}(S_v)$ respectively. (Extreme points of $\text{co}(A)$ always lie in A.) From definitions it is easily shown that $E_\sigma \times E_\eta$ is the set of extreme points of the set T of (3.5). In all problems in this section we assume that separate PCFBs for the signal and noise psds always exist (otherwise most optimizations are analytically intractable for similar reasons as explained in Section 3.1). Thus E_σ, E_η are both finite sets, each one being the set of all permutations of a single vector that corresponds to the relevant PCFB. Also T is a polytope, as its set of extreme points $E_\sigma \times E_\eta$ is also finite.

5. *Common PCFB point set* E_c. We denote by E_c the set of all points in S_v that correspond to a *common* signal and noise PCFB for the given FB class $\mathcal{C}$. (E_c is empty iff there is no such PCFB.) From earlier discussions, a FB in $\mathcal{C}$ will be such a common PCFB iff its corresponding point in the search space S_v lies in the finite set $E_\sigma \times E_\eta$. However, even when a common PCFB exists, in general all points of $E_\sigma \times E_\eta$ will not correspond to such PCFBs: In fact usually many of them will be unrealizable, i.e., outside the search space S_v. Thus, $E_c = (E_\sigma \times E_\eta) \cap S_v$, i.e., E_c consists of the extreme points of the polytope T that lie in $S_v \subset T$. Points in E_c are hence also extreme points of $\text{co}(S_v)$, i.e., $E_c \subset E_v$.

From the above definitions and discussions, the optimum FB for minimizing functions that are concave on the domain T of (3.5) can be found by a search over the FBs corresponding to points in $E_v \subset S_v$. On the other hand, common signal and noise PCFBs correspond to points in the finite set $E_c \subset E_v$. Now as noted in Section 3.1.2, for almost every $\mathbf{z} \in E_v$ there is a concave objective minimized over S_v uniquely by $\mathbf{z}$. Thus, **the common signal and noise PCFB will minimize all concave objectives over T if and only if $E_c = E_v$.** For the transform coder class $\mathcal{C}^t$, it turns out that indeed $E_c = E_v$ whenever a common signal and noise PCFB (KLT) exists. For the unconstrained class $\mathcal{C}^u$ on the other hand, even when a common PCFB exists (i.e., E_c is nonempty), $E_c \neq E_v$ in general, except for some very restricted input spectra (for example with *constant* signal and noise psd matrices, in which case the PCFBs are the corresponding KLTs). We formally state results on PCFB optimality for colored noise suppression in the next section; their proofs follow from the above comments on the relation between E_c and E_v which will be proved later. Figure 3.4 shows the various geometries of S_v as a subset of T arising in the different situations discussed above. (The figure only serves as illustration: Actually T lies in $\mathcal{R}^{2M}$ and not $\mathcal{R}^3$ as the figure shows.)

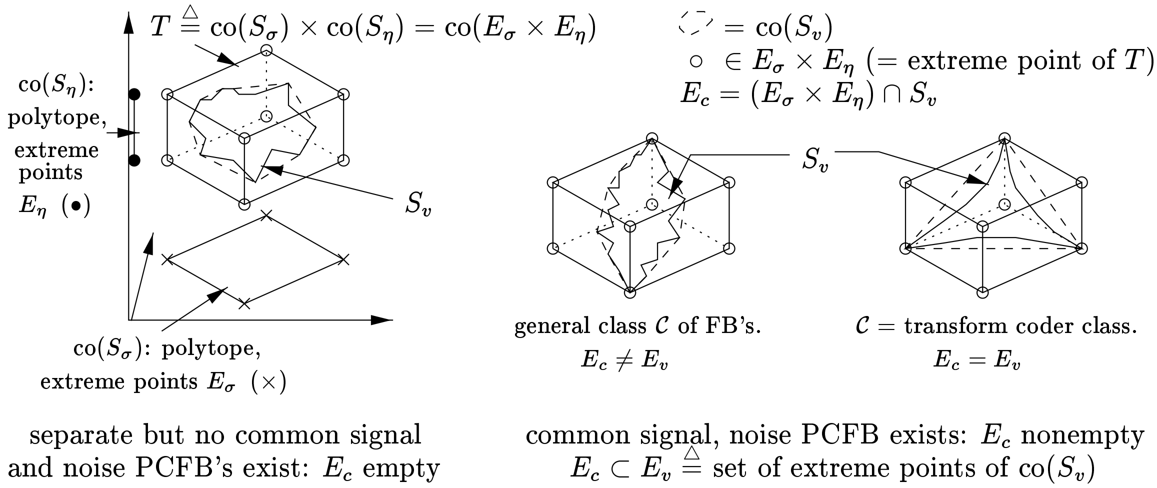


Figure 3.4: Colored noise suppression: Geometry of search–space.

3.2.3 Statement and discussion of results

Theorem 3.2: Optimality of common KLT. Consider any minimization objective that is concave on the set T of (3.5). The common signal and noise PCFB for the transform coder class $\mathcal{C}^t$ (i.e., the common KLT) if it exists, is the optimum FB in $\mathcal{C}^t$ for all these problems. Thus, it is mean square sense optimum for the noise-suppression system using any combination of constant multipliers, zeroth order Wiener filters and hard-thresholds (Section 3.2.1) in the subbands. ■

Theorem 3.3: Suboptimality of common PCFB. The optimality of the common signal and noise PCFB *for the transform coder class* $\mathcal{C}^t$ (Theorem 3.2) *does not hold* for all classes of FBs. In particular it is violated for large families of input signal and noise spectra for the class $\mathcal{C}^u$ of unconstrained two channel FBs. ■

Theorem 3.4: Optimality for a restricted class of concave objectives. For any FB class $\mathcal{C}$, the common signal and noise PCFB if it exists is always optimal for a certain well-defined subset of the minimization objectives that are concave over the domain T of (3.5). There is a finite procedure to identify whether or not a given concave objective falls in this subset. ■

Theorem 3.4 is easily proved: As long as separate PCFBs exist for the signal and noise, the set T of (3.5) is a polytope, and a search over the finite set $E_\sigma \times E_\eta$ of its extreme points will yield a minimum $\mathbf{z}_f$ of any concave objective f over T. If $\mathbf{z}_f$ *lies in the true search–space* $S_v \subset T$, then it also minimizes f over S_v, and is in E_c, i.e., corresponds to a common signal and noise PCFB. In general $\mathbf{z}_f$ does not lie in S_v, but the common PCFB minimizes all concave objectives f for which it does, thus proving Theorem 3.4.

As explained in Section 3.2.2, we will complete the proof of Theorem 3.2 (in Section 3.2.4) by showing that if a common signal and noise KLT exists, $E_c = E_v$ for the class $\mathcal{C}^t$. Section 3.2.4 also proves Theorem 3.3, using a specific example of PCFB suboptimality. We may also note here another speciality of the class $\mathcal{C}^t$ besides that shown by Theorems 3.2 and 3.3: The common signal and noise PCFB for $\mathcal{C}^t$ (i.e., common KLT) is also the PCFB (i.e., KLT) for the noisy FB input $x(n) = s(n) + \mu(n)$. This need not be true for a general FB class $\mathcal{C}$ (for example, for the unconstrained class $\mathcal{C}^u$). For the noise suppression problems, we have already shown some restricted cases of Theorem 3.2: Theorem 2.6 (Section 2.7.1) shows the case when the noise is white, while Theorem 2.7 (Section 2.7.2) shows the case when all subbands use zeroth order Wiener filters. In the latter case in fact the optimality of the common KLT has been shown even over the class of memoryless biorthogonal transforms; it is an open problem as to whether this stronger optimality still holds with other subband operations such as constant

The above results show that PCFB optimality for noise reduction is considerably restricted for colored (as opposed to white) noise. If the PCFB is not optimal, what is? We know that searching the extreme point set E_v will yield an optimal FB, but in general E_v may be infinite, making analytical solutions difficult. However for one special case involving unconstrained FBs and *piecewise constant spectra*, E_v is finite and easily characterized, as shown by the next result (proved in Section 3.2.5):

Theorem 3.5: Optimum unconstrained FB for piecewise constant spectra. Consider the problem of finding within the unconstrained M channel orthonormal FB class $\mathcal{C}^u$, a FB minimizing an objective function f that is concave on the set T of (3.5). From Section 3.2.2, this is reducible to a minimization of f over the set E_v of extreme points of the convex hull $\text{co}(S_v)$ (where S_v is the search space, defined in Section 3.2.2). Suppose the input signal and noise are WSS with psds that are constant on all intervals $(\frac{2\pi k}{MN}, \frac{2\pi(k+1)}{MN})$ for all integers k for some fixed positive integer N. Then,

1. S_v is a polytope, i.e., $S_v = \text{co}(S_v)$ and E_v is finite. Further, let $\mathcal{F}$ be the set of all *brickwall* FBs in $\mathcal{C}^u$ having all filter band–edges at integer multiples of $\frac{2\pi}{MN}$. Then the size of $\mathcal{F}$ is $|\mathcal{F}| = (M!)^N$, and for each point of E_v there is a FB in $\mathcal{F}$ corresponding to it.

2. For fixed M, $|\mathcal{F}|$ is exponential in N, but the number of FBs in $\mathcal{F}$ actually corresponding to points in E_v is polynomial: $|E_v| \leq K_M N^{2M-3}$, where $K_M = 4(2M-3)(M!(M!-1)/2)^{2M-3}$. These FBs can be extracted from $\mathcal{F}$ in $C_1 N^{2M-2}(M!)^{4M-5}$ arithmetic operations if $M > 2$ and in $C_2 N \log N$ operations if $M = 2$ (where constants C_1, C_2 are independent of M, N), again polynomial in N. ■

Discussion on Theorem 3.5

1. *On brickwall orthonormal M channel FBs* [61, 64]: In these FBs, all filters have piecewise constant responses $H_i(e^{j\omega}) \in \{0, \sqrt{M}\}$ for all ω. Their supports are nonoverlapping and aliasfree(M), i.e., for any ω, exactly one of the M numbers $H_i(e^{j(\omega+\frac{2\pi k}{M})})$, $k = 0, 1, \ldots, M-1$

is nonzero. If further all filter band–edges (i.e., points of discontinuity of $H_i(e^{j\omega})$) are integer multiples of $\frac{2\pi}{MN}$, the number of such FBs is evidently finite and not hard to compute; our proof (Section 3.2.5) gives a way to compute it.

2. *Result appeals but is not obvious:* The theorem shows that the optimum FB can always be chosen to be in $\mathcal{F}$, i.e., brickwall with nonoverlapping filter responses having shapes similar to the input spectra (piecewise constant with the same allowed discontinuities). While intuitively appealing, this is by no means obvious; e.g., it is in general false if the objective is not concave.

3. *Bounds on* $|E_v|$*:* Items (1) and (2) of the theorem statement give two different bounds $(M!)^N (= |\mathcal{F}|)$ and $K_M N^{2M-3}$ respectively on the size of E_v. The latter bound is stronger when $N \gg M$, while the former is when $M \gg N$. There are no bounds that are polynomial in both M and N.

4. *Common PCFBs and the case of* $N = 1$*:* Theorem 3.5 holds whether or not a common signal and noise PCFB for $\mathcal{C}^u$ exists for the given spectra. If such a PCFB exists, it also corresponds to points of E_v (often it is also in $\mathcal{F}$). However it need not always be optimal (Theorem 3.3), as E_v could in general have other points as well. In the special case when $N = 1$ however, $|\mathcal{F}| = M!$, and all elements of $\mathcal{F}$ are permutations of the same FB, namely the usual contiguous–stacked brickwall FB, which is hence always optimal. This FB is a common signal and noise PCFB in this case: It produces *white* and totally decorrelated signal and noise subband processes. The comments after the proof of Theorem 3.2 in Section 3.2.4 provide an independent proof of the optimality of FBs producing such subband processes.

5. *Approximating optimum FBs for arbitrary spectra:* Most spectra can be approximated by the piecewise constant ones in the premise of Theorem 3.5, to any desired accuracy by sufficiently increasing M and/or N. Thus Theorem 3.5 in principle allows approximation of the optimum FB in $\mathcal{C}^u$ for any input spectra to any desired accuracy. However the complexity of the algorithm for this is polynomial in N but super–exponential in M. Thus, we have good algorithms for low M (especially $M = 2$, where the complexity is linear in N). For sufficiently large M, we get good enough approximations of the true spectra by taking $N = 1$. The earlier remark then gives at no cost, the optimum FB in $\mathcal{C}^u$, i.e., the usual contiguous–stacked brickwall FB. There are no good algorithms if both M and N are large.

3.2.4 Proof and comments on Theorems 3.2,3.3

Proof of Theorem 3.2. Using the notations and discussion of Section 3.2.2, we need to show that for the transform coder class $\mathcal{C}^t$, $E_c = E_v$ whenever a common signal and noise PCFB (KLT) exists. Let $\mathbf{R}_\sigma, \mathbf{R}_\eta$ be the autocorrelation matrices of the M-fold blocked versions of the signal $s(n)$ and noise $\mu(n)$ respectively. Let the unitary $\mathbf{K}$ be a common KLT. Its subband signal and noise variance vectors are thus $\mathbf{z}_\sigma = \mathrm{diag}(\mathbf{\Lambda}_\sigma)$ and $\mathbf{z}_\eta = \mathrm{diag}(\mathbf{\Lambda}_\eta)$ respectively, where $\mathbf{\Lambda}_\sigma = \mathbf{K}\mathbf{R}_\sigma\mathbf{K}^\dagger$, $\mathbf{\Lambda}_\eta = \mathbf{K}\mathbf{R}_\eta\mathbf{K}^\dagger$ are both diagonal. The set of points in S_v corresponding to the KLT $\mathbf{K}$ and its permutations is thus

$$E'_c = \left\{ \begin{bmatrix} \mathbf{P}_j & \mathbf{0} \\ \mathbf{0} & \mathbf{P}_j \end{bmatrix} \begin{bmatrix} \mathbf{z}_\sigma \\ \mathbf{z}_\eta \end{bmatrix} \quad : j = 1, 2, \ldots, M! \right\} \tag{3.6}$$

where $\mathbf{P}_j$ are the $M \times M$ permutation matrices. Now E_c is the set of points in S_v corresponding to any common KLT, so $E'_c \subset E_c$. (It will turn out that $E'_c = E_c$, but this needs proof due to possible nonuniqueness of the KLT $\mathbf{K}$.) We now compute S_v: Note that $(\mathbf{v}_\sigma^T, \mathbf{v}_\eta^T)^T \in S_v$ iff there is a transform coder producing $\mathbf{v}_\sigma, \mathbf{v}_\eta$ as signal and noise subband variance vectors respectively, i.e., iff there is a unitary matrix $\mathbf{T}$ such that $\mathrm{diag}(\mathbf{T}\mathbf{R}_\sigma\mathbf{T}^\dagger) = \mathbf{v}_\sigma = \mathrm{diag}(\mathbf{T}\mathbf{K}^\dagger\mathbf{\Lambda}_\sigma\mathbf{K}\mathbf{T}^\dagger)$ and $\mathrm{diag}(\mathbf{T}\mathbf{R}_\eta\mathbf{T}^\dagger) = \mathbf{v}_\eta = \mathrm{diag}(\mathbf{T}\mathbf{K}^\dagger\mathbf{\Lambda}_\eta\mathbf{K}\mathbf{T}^\dagger)$. Let $\mathbf{Q}$ be the orthostochastic matrix (defined in Section 2.4.4) corresponding to $\mathbf{T}\mathbf{K}^\dagger$, i.e., the doubly stochastic matrix formed by replacing each entry of the unitary $\mathbf{T}\mathbf{K}^\dagger$ by the square of its absolute value. Then $\mathbf{v}_\sigma = \mathbf{Q}\mathbf{z}_\sigma$ and $\mathbf{v}_\eta = \mathbf{Q}\mathbf{z}_\eta$. Thus

$$S_v = \left\{ \begin{bmatrix} \mathbf{Q} & \mathbf{0} \\ \mathbf{0} & \mathbf{Q} \end{bmatrix} \begin{bmatrix} \mathbf{z}_\sigma \\ \mathbf{z}_\eta \end{bmatrix} : \mathbf{Q} \text{ orthostochastic} \right\} \subseteq \left\{ \begin{bmatrix} \mathbf{Q} & \mathbf{0} \\ \mathbf{0} & \mathbf{Q} \end{bmatrix} \begin{bmatrix} \mathbf{z}_\sigma \\ \mathbf{z}_\eta \end{bmatrix} : \mathbf{Q} \text{ doubly stochastic} \right\} \triangleq A. \tag{3.7}$$

By Birkhoff's theorem (Section 2.4.4) we can express $\mathbf{Q}$ above as a convex combination of permutation matrices, thus obtaining $A = \mathrm{co}(E'_c)$. Since $E'_c \subset S_v \subseteq A = \mathrm{co}(E'_c)$, we have $\mathrm{co}(E'_c) \subseteq \mathrm{co}(S_v) \subseteq \mathrm{co}(A) = \mathrm{co}(E'_c)$, i.e., $\mathrm{co}(S_v) = \mathrm{co}(E'_c)$, which is thus a polytope whose extreme points lie in E'_c. But E_v is by definition the set of these extreme points, so $E_v \subset E'_c$. Together with $E'_c \subset E_c \subset E_v$, this gives $E_c = E_v$ as desired. $\bigtriangledown\bigtriangledown\bigtriangledown$

We may note here that the set S_σ of realizable subband signal variance vectors $\mathbf{v}_\sigma$ is convex (Section 2.5.1), and that

$$S_\sigma = \{\mathbf{Q}\mathbf{z}_\sigma : \mathbf{Q} \text{ orthostochastic}\} = \{\mathbf{Q}\mathbf{z}_\sigma : \mathbf{Q} \text{ doubly stochastic}\} = \mathrm{co}(S_\sigma). \tag{3.8}$$

Is S_v convex too? For dimension $M \leq 2$, every doubly stochastic matrix is orthostochastic (Section 2.9). So from (3.7), $S_v = A = \mathrm{co}(S_v)$, i.e., S_v is indeed convex, as we also verify in Section 3.2.6 by explicitly computing S_v. Even for general M, the same argument that proves convexity of S_σ also shows that S_v is convex in two very special cases: (1) if all entries of $\mathbf{z}_\eta$ (or $\mathbf{z}_\sigma$) are equal, i.e., $\mathbf{R}_\eta$ (respectively, $\mathbf{R}_\sigma$) is the identity matrix upto scale – the 'white noise' case, and (2) if $\mathbf{z}_\eta = c\mathbf{z}_\sigma$ (i.e., $\mathbf{R}_\eta = c\mathbf{R}_\sigma$). However if $M > 2$, S_v is *not* convex for several pairs of values of $\mathbf{z}_\sigma, \mathbf{z}_\eta$ (some shown in Appendix C).

We can try to modify the above proof to show that $E_c = E_v$ for the class $\mathcal{C}^u$ too. To do this we re-

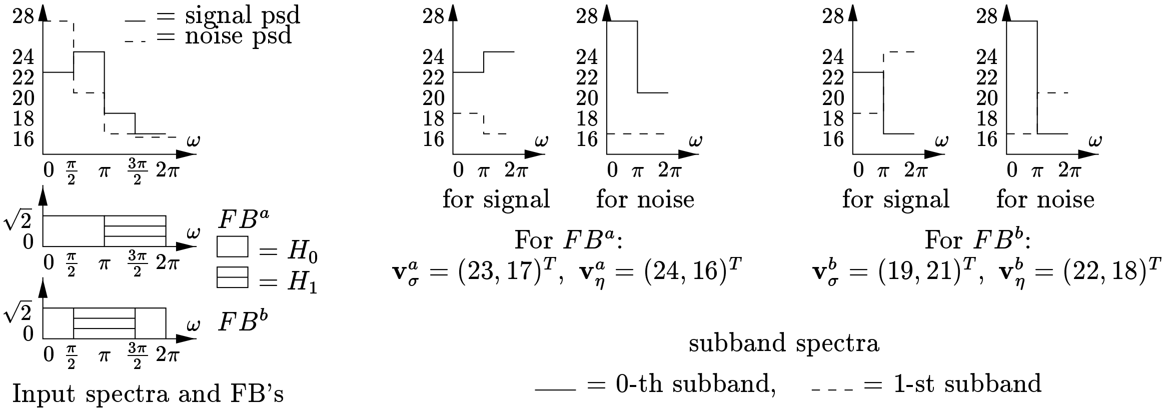


Figure 3.5: Suboptimality of common unconstrained signal and noise PCFB.

place the autocorrelation matrices $\mathbf{R}_\sigma, \mathbf{R}_\eta$ with psd matrices $\mathbf{S}_\sigma(e^{j\omega}), \mathbf{S}_\eta(e^{j\omega})$ and try to use the earlier arguments at each ω. We cannot complete the proof for *all* psd matrices, for else a common signal and noise PCFB would always be optimal for the class $\mathcal{C}^u$ too, contradicting Theorem 3.3. However, we can in fact complete the proof for some restricted classes of psds: (1) If $\mathbf{S}_\eta(e^{j\omega})$ (or $\mathbf{S}_\sigma(e^{j\omega})$) is the identity matrix upto scale – the 'white noise' case, (2) if $\mathbf{S}_\eta(e^{j\omega}) = c\mathbf{S}_\sigma(e^{j\omega})$, and (3) if the diagonalized versions of $\mathbf{S}_\sigma(e^{j\omega}), \mathbf{S}_\eta(e^{j\omega})$ are both *constant* (independent of ω). We have seen cases (1) and (2) earlier, as situations where a signal PCFB is automatically also a noise PCFB and minimizes all concave objectives of the form (3.1). In case (3), the common PCFB for $\mathcal{C}^u$ has *white* and uncorrelated subband signal and noise components. Examples of this case are (a) if $\mathbf{S}_\sigma(e^{j\omega}), \mathbf{S}_\eta(e^{j\omega})$ are themselves independent of ω– the PCFBs for $\mathcal{C}^u$ are then the corresponding KLTs, and (b) if $N = 1$ in Theorem 3.5 – the common PCFB for $\mathcal{C}^u$ is then the usual contiguous–stacked brickwall FB.

Proof of Theorem 3.3. We show a specific example of PCFB suboptimality. For the class $\mathcal{C}^u$ of unconstrained two channel FBs, consider the input signal and noise spectra and the two FBs from $\mathcal{C}^u$ shown in Fig. 3.5. The figure also shows the resulting subband spectra and signal and noise variance vectors. As the analysis filters are non-overlapping, the subbands are totally decorrelated. From Fig. 3.5, the subbands of FB^a also obey spectral majorization (Section 2.5.2), while those of FB^b do not. Thus FB^a is a common signal and noise PCFB while FB^b is neither a signal PCFB nor a noise PCFB for the class $\mathcal{C}^u$. However, consider the concave objectives of the noise suppression problem with either zeroth order Wiener filters or hard-thresholds in *both* subbands (see (3.1),(3.2)). By evaluation using the subband variances in Fig. 3.5, FB^b achieves a lower value than FB^a for these objectives. Thus the common PCFB is not always optimal. More examples of PCFB suboptimality can be created by slight perturbations of the spectra of Fig. 3.5. $\triangledown\triangledown\triangledown$

The spectra in Fig. 3.5 are piecewise constant and Theorem 3.5 can be applied to them (with $M = N = 2$). This shows that one of the two FBs (FB^a and FB^b) in Fig. 3.5 is optimum in $\mathcal{C}^u$ for any concave minimization objective. Thus, in the example in the proof of Theorem 3.3, not only is FB^b better than the common signal and noise PCFB (FB^a), but it is in fact the best possible two channel (unconstrained orthonormal) FB.

3.2.5 Proof of Theorem 3.5

Let $H_i(e^{j\omega}), i = 0, 1, \ldots, M-1$ be the analysis filters of a M channel orthonormal FB (i.e., a FB from $\mathcal{C}^u$). For $i = 0, 1, \ldots, M-1$ and $k = 0, 1, \ldots, MN-1$, define

$$f_{ik} \triangleq \frac{N}{2\pi} \int_{\frac{2\pi k}{MN}}^{\frac{2\pi(k+1)}{MN}} \left|H_i(e^{j\omega})\right|^2 \, d\omega. \tag{3.9}$$

Let the constant values of the input signal and noise psds $D_\sigma(e^{j\omega})$ and $D_\eta(e^{j\omega})$ on the interval $(\frac{2\pi k}{MN}, \frac{2\pi(k+1)}{MN})$ be a_k, b_k respectively. Let σ_i^2, η_i^2 respectively be the signal and noise variances in the i-th subband. Then

$$\sigma_i^2 = \frac{1}{2\pi} \int_0^{2\pi} \left|H_i(e^{j\omega})\right|^2 D_\sigma(e^{j\omega}) \, d\omega = \frac{1}{N} \sum_{k=0}^{MN-1} f_{ik} a_k. \tag{3.10}$$

Similarly $N\eta_i^2 = \sum_{k=0}^{MN-1} f_{ik} b_k$. Thus all subband variances are linear functions of the f_{ik}. So the search space S_v (Section 3.2.2) is the image under a linear transformation of the set of all possible arrays f_{ik} corresponding to all FBs in $\mathcal{C}^u$. Hence we now proceed to study this set. By FB orthonormality, from [64],

$$\sum_{i=0}^{M-1} \left|H_i(e^{j\omega})\right|^2 = M \quad \text{(power complementarity), and} \tag{3.11}$$

$$\sum_{k=0}^{M-1} \left|H_i(e^{j(\omega+\frac{2\pi k}{M})})\right|^2 = M \quad \text{for } i = 0, 1, \ldots, M-1 \text{ (Nyquist}(M)\text{ constraint), hence} \tag{3.12}$$

$$0 \le f_{ik} \le 1 \qquad \text{for all } i, k \text{ (for which } f_{ik} \text{ is defined),} \tag{3.13}$$

$$\sum_{i=0}^{M-1} f_{ik} = 1 \qquad \text{for all } k, \tag{3.14}$$

$$\sum_{k=0}^{M-1} f_{i(l+Nk)} = 1 \qquad \text{for all } i, \text{ for each } l = 0, 1, \ldots, N-1. \tag{3.15}$$

Here (3.13) follows from $0 \le \left|H_i(e^{j\omega})\right|^2 \le M$ for all i, ω (which follows from (3.11) or (3.12)), while (3.14),(3.15) follow from (3.11),(3.12) respectively. Define the $M \times M$ matrix $\mathbf{G}^{(l)}$ to have entries $g_{ik}^{(l)} = f_{i(l+Nk)}$ (where $i, k \in \{0, 1, \ldots, M-1\}$), for $l = 0, 1, \ldots, N-1$. Then (3.13),(3.14),(3.15) are equivalent to the following:

$$\mathbf{G}^{(l)} \text{ is } \textbf{doubly stochastic} \text{ for all } l = 0, 1, \ldots, N-1. \tag{3.16}$$

Let $\mathcal{G}$ be the collection of all ordered sets $(\mathbf{G}^{(0)}, \mathbf{G}^{(1)}, \ldots, \mathbf{G}^{(N-1)})$ corresponding to all FBs in $\mathcal{C}^u$. Instead of studying the set of all arrays f_{ik}, we can study $\mathcal{G}$ (as S_v is also the image of $\mathcal{G}$ under a linear transform). Let $\mathcal{Q}, \mathcal{P}$ respectively denote the sets of all $M \times M$ doubly stochastic matrices and permutation matrices. From (3.16), $\mathcal{G} \subset \mathcal{Q}^N$ $(= \mathcal{Q} \times \mathcal{Q} \times \ldots \times \mathcal{Q})$.

Claim: $\mathcal{G} = \mathcal{Q}^N$, which (by Birkhoff's theorem, Section 2.4.4) is a polytope with $\mathcal{P}^N$ as its set of extreme points. Also, FBs in the set $\mathcal{F}$ (defined in stating Theorem 3.5) correspond directly (one–to–one) with points in $\mathcal{P}^N$.

Showing this claim will prove item (1) in the statement of Theorem 3.5. Recall that S_v is the image of $\mathcal{G}$ under a linear transform $\mathcal{L}$. So if $\mathcal{G}$ is a polytope, so is S_v; further all its extreme points are images of some extreme points of $\mathcal{G}$ under $\mathcal{L}$. The claim above thus means that there is a FB in $\mathcal{F}$ for every extreme point of S_v. The correspondence between $\mathcal{F}$ and $\mathcal{P}^N$ also means that $\mathcal{F}$ has $|\mathcal{F}| = |\mathcal{P}^N| = |\mathcal{P}|^N = (M!)^N$ FBs (counting separately all permutations of each FB in $\mathcal{F}$ – else we must divide the number by $M!$).

Proof of Claim: We show that $\mathcal{G} = \mathcal{Q}^N$ by building a brickwall FB in $\mathcal{C}^u$ corresponding to any given $\underline{\mathbf{G}} = (\mathbf{G}^{(0)}, \mathbf{G}^{(1)}, \ldots, \mathbf{G}^{(N-1)}) \in \mathcal{Q}^N$. To do this, let σ_m, $m = 0, 1, \ldots, M!-1$ be the $M!$ permutation functions on the set $\{0, 1, \ldots, M-1\}$. Now there is a one–to–one correspondence between brickwall FBs and functions ϕ mapping each $\omega \in [0, \frac{2\pi}{M})$ to one of the σ_m. This is described by the following construction of the analysis filters $H_i(e^{j\omega}), i = 0, 1, \ldots, M-1$ of the FB given the function ϕ: Let $\omega \in [0, \frac{2\pi}{M})$ and $\phi(\omega) = \sigma_m$. Then $H_{\sigma_m(k)}(e^{j(\omega + \frac{2\pi k}{M})}) = \sqrt{M}$ for $k = 0, 1, \ldots, M-1$. In other words, the permutation $\sigma_m = \phi(\omega)$ decides which of the M filter responses is nonzero at the M frequencies $\omega + \frac{2\pi k}{M}$. The construction ensures nonoverlapping aliasfree(M) filter responses resulting in a valid FB in $\mathcal{C}^u$. Now for each $l = 0, 1, \ldots, N-1$, let $x_m^{(l)}$ be the fraction of length of the interval $[\frac{2\pi l}{MN}, \frac{2\pi(l+1)}{MN})$ that is mapped by ϕ to σ_m, for $m = 0, 1, \ldots, M!-1$. For a brickwall FB, f_{ik} of (3.9) is the fraction of length of the interval $[\frac{2\pi k}{MN}, \frac{2\pi(k+1)}{MN})$ on which $H_i(e^{j\omega})$ is nonzero (i.e., $= \sqrt{M}$). Thus, the chosen ϕ yields a FB corresponding to the given $\underline{\mathbf{G}} \in \mathcal{Q}^N$ (i.e., given set of f_{ik} obeying (3.13),(3.14),(3.15)) iff for $i, k = 0, 1, \ldots, M-1$ and $l = 0, 1, \ldots, N-1$ we have

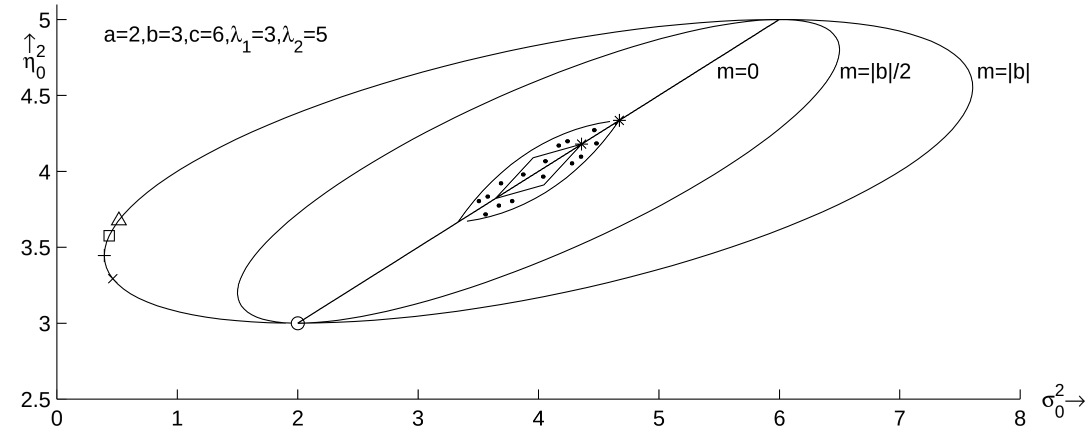
$$\sum_{\text{all } m \text{ obeying } \sigma_m(k) = i} x_m^{(l)} \quad = \quad f_{i(l+Nk)} \;\; (\stackrel{\triangle}{=} g_{ik}^{(l)}) \tag{3.17}$$

Thus, given $\underline{\mathbf{G}}$, we must find $x_m^{(l)} \in [0, 1]$ obeying (3.17). This is easy if $\underline{\mathbf{G}} \in \mathcal{P}^N$: Here for each l, $\mathbf{G}^{(l)}$ (with entries $g_{ik}^{(l)}$) is a permutation matrix, i.e., there is a $m_*(l)$ such that $g_{ik}^{(l)}$ is 1 if $\sigma_{(m_*(l))}(k) = i$

and is 0 otherwise. We then simply set $x_m^{(l)}$ to be 1 for $m = m_*(l)$ and 0 for all other m. Note that this yields a FB in the set $\mathcal{F}$ defined in stating Theorem 3.5. For a general $\underline{\mathbf{G}} \in \mathcal{Q}^N$, we use Birkhoff's theorem to write $\mathbf{G}^{(l)} \in \mathcal{Q}$ as a convex combination of elements of $\mathcal{P}$. The same convex combination of the solution vectors $(x_0^{(l)}, x_1^{(l)}, \ldots, x_{M!-1}^{(l)})$ corresponding to each element of $\mathcal{P}$ yields the corresponding solution for $\mathbf{G}^{(l)}$. Repeating the process for $l = 0, 1, \ldots, N-1$ completes the solution. This shows that $\mathcal{G} = \mathcal{Q}^N$, a polytope with $\mathcal{P}^N$ as its set of extreme points. The proof has also associated to each of these points a unique FB in $\mathcal{F}$. Conversely for any FB in $\mathcal{F}$, the f_{ik} of (3.9), and hence all entries of the doubly stochastic $\mathbf{G}^{(l)}$, are either 0 or 1. Hence the corresponding point in $\mathcal{G}$ is in $\mathcal{P}^N$. This proves the one–to–one correspondence between $\mathcal{F}$ and $\mathcal{P}^N$. $\bigtriangledown\bigtriangledown\bigtriangledown$

Proof of item (2) in Theorem 3.5: By (3.10), the map $\mathcal{L}$ from $\underline{\mathbf{G}} = (\mathbf{G}^{(0)}, \mathbf{G}^{(1)}, \ldots, \mathbf{G}^{(N-1)}) \in \mathcal{G}$ to the associated point in S_v is $\mathcal{L}(\underline{\mathbf{G}}) = \sum_{l=0}^{N-1} \begin{pmatrix} \mathbf{G}^{(l)}\mathbf{a}^{(l)} \\ \mathbf{G}^{(l)}\mathbf{b}^{(l)} \end{pmatrix}$, where $\mathbf{a}^{(l)} = (a_l, a_{l+N}, \ldots, a_{l+(M-1)N})^T/N$, $\mathbf{b}^{(l)} = (b_l, b_{l+N}, \ldots, b_{l+(M-1)N})^T/N$. For any fixed $l = 0, 1, \ldots, N-1$, as $\mathbf{G}^{(l)}$ can be any element of $\mathcal{Q}$, the set of possible values of $\begin{pmatrix} \mathbf{G}^{(l)}\mathbf{a}^{(l)} \\ \mathbf{G}^{(l)}\mathbf{b}^{(l)} \end{pmatrix}$ is itself a polytope $T^{(l)}$. It lies on a $2M-2$ dimensional hyperplane in $\mathcal{R}^{2M}$, and its extreme points correspond to the $M!$ possible choices of $\mathbf{G}^{(l)} \in \mathcal{P}$. Thus, $S_v = \left\{\sum_l \mathbf{x}^{(l)} \mid \mathbf{x}^{(l)} \in T^{(l)}\right\}$, which is known as the *Minkowski sum* of the polytopes $T^{(l)}$. Minkowski sums have been well studied in computational geometry [26],[49], e.g. in context of robot motion planning algorithms in 2 and 3 dimensions [49]. Theorem 2.1.10 and Corollary 2.1.11 of [26] bound the number of extreme points of the Minkowski sum of k polytopes of dimension d with not more than p extreme points each. Theorem 2.3.7′ and Proposition 2.3.9 of [26], with their proofs, outline algorithms to find the extreme points of this Minkowski sum, thus bounding the number of arithmetic operations needed for the same. Applying these bounds with $k = N$, $d = 2M-2$ and $p = M!$ yields item (2) of the statement of Theorem 3.5. $\bigtriangledown\bigtriangledown\bigtriangledown$

Note that like S_v, the set S_σ of realizable signal subband variance vectors is also the image of $\mathcal{G}$ under a linear map $\mathcal{L}_\sigma$ given by $\mathcal{L}_\sigma(\underline{\mathbf{G}}) = \sum_{l=0}^{N-1} \mathbf{G}^{(l)}\mathbf{a}^{(l)}$. However, while $\mathcal{G}$ has $(M!)^N$ extreme points (i.e., points in $\mathcal{P}^N$), and S_v has $|E_v| \leq K_M N^{2M-3}$ of them, we know from Chapter 2 that S_σ has at most $M!$ of them – its extreme points are the permutations of the signal PCFB subband variance vector $\mathbf{v}_*$. Indeed, here $\mathbf{v}_* = \mathcal{L}_\sigma(\underline{\mathbf{G}})$ when each $\mathbf{G}^{(l)}$ is a permutation matrix rearranging the entries of $\mathbf{a}^{(l)}$ in decreasing order. It is not hard to see (by definition of majorization) that $\mathbf{v}_*$ majorizes all points $\mathcal{L}_\sigma(\underline{\mathbf{G}})$ for $\underline{\mathbf{G}} \in \mathcal{P}^N$ (i.e., for all choices of $\underline{\mathbf{G}}$ as extreme points of $\mathcal{G}$). Hence, from Section 2.4, all these points are some convex combinations of the permutations of $\mathbf{v}_*$. Thus these permutations are the only extreme points of S_σ. Note that S_σ too is expressible as a Minkowski sum of polytopes $T_\sigma^{(l)}$, where $T_\sigma^{(l)}$ is the set of all permutations of $\mathbf{a}^{(l)}$. Using [26] to bound the number of extreme points of S_σ gives a bound that grows with both M and N, whereas the true



Δ = optimum for hard thresholding, $\square$ = optimum for Wiener filtering.
$+$ = signal KLT, $\times$ = signal + noise KLT, $\circ$ = noise KLT.
= search space for which common PCFB exists ($*$), but is not always optimal.

Figure 3.6: Search–space $S_{\sigma\eta}$ for two channel transform coders.

number is independent of N. The tightening of the bound of [26] has been achieved by exploiting the special structure of the summand polytopes $T_\sigma^{(l)}$. The summands $T^{(l)}$ of S_v also have a special structure (different from that of the $T_\sigma^{(l)}$), but it is not clear whether the bound on $|E_v|$ can be similarly tightened using this structure.

3.2.6 Study of S_v for two channel transform coders

Here we explicitly compute for the class $\mathcal{C}^t$ of two channel transform coders, the set $S_v \subset \mathcal{R}^4$ of all realizable pairs of subband signal and noise variance vectors $(\sigma_0^2, \sigma_1^2)^T$ and $(\eta_0^2, \eta_1^2)^T$ respectively. In fact it suffices to study instead, the set $S_{\sigma\eta} \subset \mathcal{R}^2$ of all realizable pairs $(\sigma_0^2, \eta_0^2)^T$ (which can be plotted, unlike S_v). This is because the properties of these sets are directly related by the fact that $\sigma_0^2 + \sigma_1^2 = k_\sigma$ and $\eta_0^2 + \eta_1^2 = k_\eta$ are constants (independent of choice of FB from $\mathcal{C}^t$). For example S_v is convex iff $S_{\sigma\eta}$ is convex, there is an obvious correspondence between the extreme points of the sets $\text{co}(S_v)$ and $\text{co}(S_{\sigma\eta})$, and the permutation symmetry of S_v is equivalently restated as $(\sigma_0^2, \eta_0^2)^T \in S_{\sigma\eta} \Rightarrow (k_\sigma - \sigma_0^2, k_\eta - \eta_0^2)^T \in S_{\sigma\eta}$. The result of computing $S_{\sigma\eta}$ is summarized as follows:

Theorem 3.6: Search space for two channel transform coders. Consider the class $\mathcal{C}^t$ of two channel transform coders, and the associated set $S_{\sigma\eta} \subset \mathcal{R}^2$ defined above. If a common signal and noise PCFB (KLT) exists for $\mathcal{C}^t$, then $S_{\sigma\eta}$ is a **line-segment** whose endpoints correspond to the common PCFB. Otherwise, it is an **elliptical disk**. ■

Discussion. When a common KLT exists, $\text{co}(S_{\sigma\eta})$, and hence $\text{co}(S_v)$, is a polytope ($\text{co}(S_{\sigma\eta})$ is a 1–dimensional polytope, i.e., a line-segment). Further the extreme points of the polytope are precisely the points corresponding to the common KLT. This corroborates for two channel FBs, the result $E_c = E_v$ proved in Section 3.2.4 for any transform coder class with a common signal and noise KLT. Recall (from Section 3.2.2) that this result was the key to the optimality of the common KLT (Theorem 3.2). Also note that $S_{\sigma\eta} = \text{co}(S_{\sigma\eta})$, i.e., $S_{\sigma\eta}$ is convex, and hence so is S_v. This also was independently proved earlier for *two channel* transform coders, though it does not always hold with more than two channels (Section 3.2.4).

If there is no common KLT, $S_{\sigma\eta}$ is an elliptical disk – a compact convex set whose extreme points are the points on its elliptical boundary. Thus S_v is a compact convex set with infinitely many extreme points. The minima over S_v of different concave objectives are at different extreme points. Figure 3.6 shows a plot of $S_{\sigma\eta}$; the parameters a, b, c, etc. are constants depending on the input spectra (defined shortly). The individual signal and noise KLTs are extreme points of $S_{\sigma\eta}$ – respectively the points at which the vertical and horizontal tangents to the disk $S_{\sigma\eta}$ touch it. This verifies a general fact: The individual signal PCFB for any class of FBs corresponds to boundary points of $\text{co}(S_v)$, and further if it uniquely defines the subband noise variance vector, it actually corresponds to an extreme point of $\text{co}(S_v)$. However, the individual signal and noise KLTs need not be optimum: Figure 3.6 shows that different concave objectives yield different minima, all lying on the disk boundary. The figure also shows contrived examples of FB classes for which common signal and noise PCFBs exist but do not minimize all concave objectives. The classes are defined as sets of all FBs in $\mathcal{C}^t$ whose variance pairs $(\sigma_0^2, \eta_0^2)^T$ lie in well–chosen subsets of $S_{\sigma\eta}$, marked as dotted areas in the figure. Note that these subsets obey the required permutation symmetry property. These examples are artificial due to the choice of these subsets, and also because the FB class definition depends on the input spectra.

Proof of Theorem 3.6. Let the input signal and noise autocorrelation matrices $\mathbf{R}_\sigma, \mathbf{R}_\eta$, and a general element $\mathbf{T} \in \mathcal{C}^t$ (i.e., a general 2×2 unitary matrix), respectively be given by

$$\mathbf{R}_\sigma = \begin{bmatrix} a & b \\ b^* & c \end{bmatrix}, \quad \mathbf{R}_\eta = \begin{bmatrix} \lambda_1 & 0 \\ 0 & \lambda_2 \end{bmatrix}, \quad \mathbf{T} = \begin{bmatrix} \cos\theta & e^{-j\phi}\sin\theta \\ e^{j\phi}\sin\theta & -\cos\theta \end{bmatrix}. \tag{3.18}$$

Here $a, c, \lambda_1, \lambda_2 \geq 0$ and $ac \geq |b|^2$ as $\mathbf{R}_\sigma, \mathbf{R}_\eta$ are positive semidefinite. By initially passing the noisy input through the KLT for the noise, $\mathbf{R}_\eta$ can be assumed diagonal without loss of generality. A common signal and noise KLT exists iff one (or both) of the following hold: (1) $\mathbf{R}_\sigma$ is diagonal too, i.e., $b = 0$, or (2) $\mathbf{R}_\eta$ is the identity matrix upto scale (so that any unitary matrix diagonalizes it), i.e., $\lambda_1 = \lambda_2$ (e.g. this happens with white input noise). Also in (3.18), $\theta, \phi \in [0, 2\pi)$, and the unitary $\mathbf{T}$ is fully general upto multiplication by a *diagonal* unitary matrix, which does not affect its subband variances. By direct computation, the subband signal and noise variance vectors

$(\sigma_0^2, \sigma_1^2)^T = \mathrm{diag}(\mathbf{T}\mathbf{R}_\sigma\mathbf{T}^\dagger)$ and $(\eta_0^2, \eta_1^2)^T = \mathrm{diag}(\mathbf{T}\mathbf{R}_\eta\mathbf{T}^\dagger)$ respectively are:

$$\begin{pmatrix} \sigma_0^2 \\ \sigma_1^2 \end{pmatrix} = \begin{pmatrix} a\cos^2\theta + c\sin^2\theta + \mathrm{Re}(be^{j\phi})\sin 2\theta \\ a\sin^2\theta + c\cos^2\theta - \mathrm{Re}(be^{j\phi})\sin 2\theta \end{pmatrix}, \quad \begin{pmatrix} \eta_0^2 \\ \eta_1^2 \end{pmatrix} = \begin{pmatrix} \lambda_1\cos^2\theta + \lambda_2\sin^2\theta \\ \lambda_1\sin^2\theta + \lambda_2\cos^2\theta \end{pmatrix}. \tag{3.19}$$

Here $\mathrm{Re}(z)$ is the real part of z. Note that $\mathbf{T}$ is the signal KLT iff $be^{j\phi}$ is real (i.e., $\left|\mathrm{Re}(be^{j\phi})\right|$ is maximized) and the choice of θ then maximizes (or minimizes) σ_0^2. Of course $\mathbf{T}$ is the noise KLT iff it is diagonal or antidiagonal, i.e., iff $\cos\theta\sin\theta = 0$.

From (3.19), $S_{\sigma\eta}$ is the set of all $(\sigma_0^2, \eta_0^2)^T$ satisfying for some θ, ϕ, the equation

$$\begin{pmatrix} \sigma_0^2 \\ \eta_0^2 \end{pmatrix} - \mathbf{e} = \mathbf{A}_\phi \begin{pmatrix} \cos 2\theta \\ \sin 2\theta \end{pmatrix}, \quad \text{where} \quad \mathbf{A}_\phi = \frac{1}{2}\begin{bmatrix} a-c & 2\mathrm{Re}(be^{j\phi}) \\ \lambda_1 - \lambda_2 & 0 \end{bmatrix}, \quad \mathbf{e} = \frac{1}{2}\begin{pmatrix} a+c \\ \lambda_1 + \lambda_2 \end{pmatrix}. \tag{3.20}$$

For each fixed ϕ, let $\mathcal{I}_\phi$ be the set of vectors in $\mathcal{R}^2$ given by the right side as θ varies. Then $S_{\sigma\eta}$ is the union of these sets $\mathcal{I}_\phi$ as ϕ varies, with origin shifted to $\mathbf{e}$. As $\mathbf{e}$ is constant, it suffices to prove Theorem 3.6 replacing $S_{\sigma\eta}$ by the union $\bigcup_\phi \mathcal{I}_\phi$. From (3.20), $\mathcal{I}_\phi$ is the image of the unit circle under a linear map $\mathbf{A}_\phi$. So $\mathcal{I}_\phi$ is a line segment with midpoint at the origin if $\mathbf{A}_\phi$ is singular, and an ellipse centered at the origin otherwise. Suppose a common signal and noise KLT exists, i.e., $\lambda_1 = \lambda_2$ or $b = 0$ (or both). Then $\mathbf{A}_\phi$ is singular for all ϕ. If $\lambda_1 = \lambda_2$, $\mathcal{I}_\phi$ is horizontal, while if $b = 0$, it lies along the line $(\lambda_1 - \lambda_2)x = (a - c)y$, for all ϕ. So in either case, $\bigcup_\phi \mathcal{I}_\phi$ is a line-segment with midpoint at the origin. Its endpoints correspond to extremum (maximum or minimum) values of both σ_0^2 and η_0^2, i.e., to the common KLT.

Now suppose a common signal and noise KLT does not exist. Then $\mathcal{I}_\phi$ is an ellipse centered at the origin for general ϕ. It degenerates into a line segment for exactly two values of ϕ in $[0, 2\pi)$ at which $\mathrm{Re}(be^{j\phi}) = 0$, i.e., $\mathbf{A}_\phi$ is singular. To compute $\bigcup_\phi \mathcal{I}_\phi$, we write (using (3.20)) the nonparametric equation of the ellipse $\mathcal{I}_\phi$:

$$\left(\frac{x - (\frac{a-c}{\lambda_1-\lambda_2})y}{m}\right)^2 + \left(\frac{2y}{\lambda_1 - \lambda_2}\right)^2 = 1, \quad \text{where} \quad m = \mathrm{Re}(be^{j\phi}) \text{ and } \begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} \sigma_0^2 \\ \eta_0^2 \end{pmatrix} - \mathbf{e}.$$

This shows that (1) the ellipses for $+m$ and $-m$ are the same, (2) the ellipse for $m_1 \geq 0$ lies inside that for $m_2 > m_1$, and (3) for every point in the interior of the ellipse for $m_2 > 0$ there is another ellipse for $m_1 \geq 0$ (for some $m_1 < m_2$) passing through it (the 'ellipse' for $m = 0$ is the line segment $x = \frac{a-c}{\lambda_1-\lambda_2}y$ with endpoints having $y = \pm\left|\lambda_1 - \lambda_2\right|/2$). Since the range of values of m is $[-|b|, |b|]$, we conclude that $\bigcup_\phi \mathcal{I}_\phi$ is an elliptical disk whose boundary is the ellipse corresponding to $m = |b|$. $\bigtriangledown\bigtriangledown\bigtriangledown$

For the present example, with certain concave objectives of the form (3.1), it is easy to explicitly compute the optimum FB $\mathbf{T}$ of (3.18), by inserting the variances of (3.19) into the objective and analytically optimizing θ and ϕ. For example it can be done for noise suppression using constant multipliers in both subbands (see (3.2)). This will verify that the optimum FB indeed corresponds to a boundary point of $S_{\sigma\eta}$, and further that the common signal and noise KLT is optimum if it exists.

3.3 Conclusion

We have extended the study of principal component filter banks in many ways. A central theme in our analysis is to study the geometry of the relevant search spaces of realizable subband variances, and to exploit concavity of the minimization objective on these spaces. However, many interesting issues are still unresolved.

An important question is whether there are any useful classes of FBs for which PCFBs exist for all (or large families of) input spectra. Indeed it seems possible that the two channel, the transform coder and the unconstrained classes may be the only such classes (ruling out contrived situations where the class definition depends on the input spectrum). However this has not been proved. Analytical study of PCFB existence and FB optimization for classes of FIR FBs has proven to be very complicated. The problem stated in most general form is as follows: Given a class of orthonormal FBs, find all input spectra for which a PCFB exists.

Regarding the FIR classes, we could reach a partial solution by solving the following problem: Find a family of input spectra for which there is no PCFB for some general FIR class, say that of all FBs with a given bound N on the McMillan degree or order of the polyphase matrix. At present, a few such results are known for specific low values of the bound N, for isolated input spectra [45],[36]. Even in these cases, the proofs of PCFB nonexistence need numerical optimizations. Further one of these, from [36], is suspect due to the assumption that the FB maximizing its largest subband variance must contain a FIR compaction filter. Some insight may possibly be obtained by analytical computation of the search-spaces for simple examples of these classes (e.g. the class of all 3 channel FIR FBs with polyphase matrices of McMillan degree unity).

Another area of open issues involves biorthogonal FBs. The compression and noise reduction systems of this paper remain well-defined if the FB used is biorthogonal rather than orthonormal; however the FB optimization objective no longer depends purely on the subband variances. We have seen certain cases where the best orthonormal FB is also the best biorthogonal one. For example, the KLT is not only the best orthogonal transform but also the best memoryless biorthogonal one for both the high bitrate coding problem with optimal bit allocation and for noise reduction with Wiener filters in all subbands. However it is not known whether this is true with other subband operations, e.g. low bitrate coding and noise reduction by hard-thresholds. For the unconstrained biorthogonal FB class, even for the high bitrate coding problem the best FB was known only in certain cases [62] until recently when [46] has claimed a full solution.

With regard to noise suppression, we have considered only Wiener filters of order $N = 0$ in the subbands. If $N > 0$, the objective depends not only on the subband variances but also on other coefficients in the autocorrelation sequences of the subband processes. In this case, analytical results on the optimum FB are not known. The performance gain due to increasing the order of the subband Wiener filters could instead be obtained by using a FB with more channels, however the exact nature of this tradeoff is not known.

3.4 Appendices

Appendix A: Compactness of search–space

Here we justify the assumption of Section 3.1.2 that the search–space $\mathcal{S}$ is compact, i.e., closed and bounded. (In fact we already know from Section 2.3.1 that it is bounded.) Many FB classes $\mathcal{C}$ are parameterized by a vector of real numbers, that is free to take any values in a set P which may be called the parameter space. It often happens that P is compact, and that for any bounded non-impulsive input spectrum, there is a continuous function mapping parameter vectors (from P) to the subband variance vectors (in $\mathcal{S}$) produced by the corresponding FB. Thus $\mathcal{S}$ is the continuous image of the compact set P, and is hence compact. This reasoning works for example when $\mathcal{C}$ is the set of all FIR orthonormal M channel FBs with a given McMillan degree – Here $\mathcal{C}$ is parameterized by a finite set of unit norm vectors in $\mathcal{R}^M$ and a unitary matrix [64]. Thus P is compact, being the Cartesian product of finitely many sphere-surfaces in $\mathcal{R}^M$ and the set of $M \times M$ unitary matrices.

Appendix B: Concavity proofs for some functions in the paper

$f(\mathbf{v}) = -d(\mathbf{v}, P)$ *(Section 3.1.4)*: Continuity of f follows from that of the norm. To show concavity of f, we must show that $d(\mathbf{z}, P) \leq \alpha d(\mathbf{x}, P) + (1-\alpha) d(\mathbf{y}, P)$, where $\mathbf{z} = \alpha\mathbf{x} + (1-\alpha)\mathbf{y}$, for $0 \leq \alpha \leq 1$, $\mathbf{x}, \mathbf{y} \in \mathcal{R}^M$. Let $\mathbf{a}, \mathbf{b}$ be the points in P that are closest to $\mathbf{x}, \mathbf{y}$ respectively. (They exist because P is compact.) Thus

$$\alpha d(\mathbf{x}, P) + (1-\alpha) d(\mathbf{y}, P) = \alpha \|\mathbf{x} - \mathbf{a}\| + (1-\alpha) \|\mathbf{y} - \mathbf{b}\| \geq \|\alpha(\mathbf{x} - \mathbf{a}) + (1-\alpha)(\mathbf{y} - \mathbf{b})\| = \|\mathbf{z} - \mathbf{c}\|$$

where $\mathbf{c} = \alpha\mathbf{a} + (1-\alpha)\mathbf{b} \in P$ since P is convex. Thus $\|\mathbf{z} - \mathbf{c}\| \geq d(\mathbf{z}, P)$, which completes the proof.

Functions of (3.2): Linear functions and the minimum of concave functions are concave [50], so $f_i(x, y) = |1 - k_i|^2 x + |k_i|^2 y$ and $f_i(x, y) = \min(x, y)$ are concave on $\mathcal{R}^2$. We now show that

$f_i(x,y) = \frac{xy}{x+y}$ is concave on $\mathcal{R}_+^2$, i.e., that for $x, y, a, b \geq 0$,

$$\frac{[\alpha x + (1-\alpha)a]\,[\alpha y + (1-\alpha)b]}{\alpha x + (1-\alpha)a + \alpha y + (1-\alpha)b} \geq \alpha\left(\frac{xy}{x+y}\right) + (1-\alpha)\left(\frac{ab}{a+b}\right) \quad \text{when} \quad 0 \leq \alpha \leq 1. \tag{3.21}$$

By cross-multiplying and defining $L = \alpha^2 xy + (1-\alpha)^2 ab$, this is equivalent to proving that

$$[L + \alpha(1-\alpha)(xb+ya)]\,(x+y)(a+b) \geq [\alpha xy(a+b) + (1-\alpha)ab(x+y)]\,[\alpha(x+y) + (1-\alpha)(a+b)]\,.$$

The right side is $L(x+y)(a+b) + \alpha(1-\alpha)\left[ab(x+y)^2 + xy(a+b)^2\right]$, hence as $\alpha \in [0,1]$, it suffices to show that $(xb+ya)(x+y)(a+b) \geq ab(x+y)^2 + xy(a+b)^2$, i.e., (expanding and simplifying) that $x^2b^2 + y^2a^2 - 2xyab = (xb-ya)^2 \geq 0$, which is true. Thus $f_i(x,y) = \frac{xy}{x+y}$ arising in colored noise reduction (see (3.2)) is concave on $\mathcal{R}_+^2$. However it is not *strictly* concave as equality holds in (3.21) when $xb = ya$. Note that fixing x (or y) in $f_i(x,y)$ yields univariate functions that appear in white noise reduction and are *strictly* concave on $\mathcal{R}_+$.

Appendix C: Nonconvexity of search–space S_v (Section 3.2.4)

For $M = 3$, let $\mathbf{Q}_* = \begin{bmatrix} 0.5 & 0.5 & 0 \\ 0.5 & 0 & 0.5 \\ 0 & 0.5 & 0.5 \end{bmatrix}$ which is doubly stochastic but not orthostochastic (Section 2.9). Let $\mathbf{z}_\sigma = (a,b,c)^T, \mathbf{z}_\eta = (k,0,0)^T$ where $k(b-c) \neq 0$. By (3.7), $\mathbf{v}_* \triangleq \begin{bmatrix} \mathbf{Q}_* & \mathbf{0} \\ \mathbf{0} & \mathbf{Q}_* \end{bmatrix} \begin{bmatrix} \mathbf{z}_\sigma \\ \mathbf{z}_\eta \end{bmatrix} = \begin{bmatrix} \mathbf{v}_\sigma \\ \mathbf{v}_\eta \end{bmatrix} \in A = \text{co}(S_v)$, where $\mathbf{v}_\sigma = 0.5(a+b, a+c, b+c)^T, \mathbf{v}_\eta = k(0.5, 0.5, 0)^T$. Now $\mathbf{v}_* \in S_v$ iff $\mathbf{Q}\mathbf{z}_\sigma = \mathbf{v}_\sigma$ and $\mathbf{Q}\mathbf{z}_\eta = \mathbf{v}_\eta$ for some orthostochastic $\mathbf{Q}$; but it can be verified that a doubly stochastic $\mathbf{Q}$ satisfies these equations iff $\mathbf{Q} = \mathbf{Q}_*$ which is not orthostochastic. So $\mathbf{v}_* \notin S_v$, proving that S_v is not convex. A somewhat more restricted class of pairs $\mathbf{z}_\sigma = (a, 0, \ldots, 0)^T$, $\mathbf{z}_\eta = (0, b, 0, \ldots, 0)^T$ (with $ab \neq 0$) also produces nonconvex S_v *for any* $M > 2$. To show this we use the earlier argument replacing $\mathbf{Q}_*$ by $\mathbf{Q}_{**} = \begin{bmatrix} \mathbf{Q}_* & \mathbf{0} \\ \mathbf{0} & \mathbf{I} \end{bmatrix}$, where $\mathbf{I}$ is the identity and the $\mathbf{0}$'s are zero matrices of suitable sizes. Here a doubly stochastic $\mathbf{Q}$ satisfying $\mathbf{Q}\mathbf{z}_\sigma = \mathbf{v}_\sigma$ and $\mathbf{Q}\mathbf{z}_\eta = \mathbf{v}_\eta$ need not be $\mathbf{Q}_{**}$, but must agree with it in the first two columns. This already prevents it from being orthostochastic.

Chapter 4 Optimum filter banks for communications

Filter banks are used in communications in a configuration called the transmultiplexer, where the synthesis bank precedes the analysis bank as shown in Fig. 4.1. The transmitter uses the synthesis bank to combine the signals from several channels into a single signal which is sent on the communications channel. The channel is often modelled as a LTI system with additive Gaussian noise at its output. The receiver uses the analysis bank to decompose the received signal into its constituent channels. The discrete multitone modulation (DMT) system uses such a configuration, and has found applications in ADSL (asymmetric digital subscriber line) transmission schemes that use the twisted pair copper telephone cable to send data (as opposed to voice). The system has also been proposed for wireless transmission, where it is called orthogonal frequency division multiplexing (OFDM). Most of these systems use the DFT or cosine–modulated FBs, due to their implementation efficiency. Here we study the problem of optimizing the FB for use in these systems, and show that principal component FBs are also optimal in this case. We present simulation examples with fairly realistic channel and noise models for the ADSL system, to compare the performance of various FBs.

4.1 Introduction

The DMT system, shown in Fig. 4.1, is an attractive method for communication over a non-flat channel with possibly colored additive noise. It has been extensively studied [31],[13],[1], with applications in DSL technology [16],[56] and wireless OFDM transmission for cellular radio and digital audio broadcasting [58]. The system is the 'dual' of the filter bank system of Fig. 1.3, and many results for this system parallel those for FBs.

In the extreme case when the FB in Fig. 4.1 is a delay–chain, the system performs time–division multiplexing (TDM) of the M input data streams. Thus these streams are 'orthogonal in time', and are perfectly recovered at the receiver in absence of channel impairments ($C(z) = 1, S_{ee}(e^{j\omega}) = 0$ in Fig. 4.1). In the other extreme where the FB is the ideal brickwall FB of Fig. 1.4, we get a frequency division multiplexing (FDM), and the orthogonality is now in the frequency domain. If we choose a general orthonormal FB, we again have orthogonal subcarriers, and the system is a tradeoff between the two extremes mentioned above. The polyphase representation (Fig. 4.1b) makes it clear that the condition for perfect transmultiplexing ($\hat{x}_i(n) = x_i(n)$ in absence of channel imperfections) is the same as that for perfect reconstruction (in the absence of subband processing) for the FB of

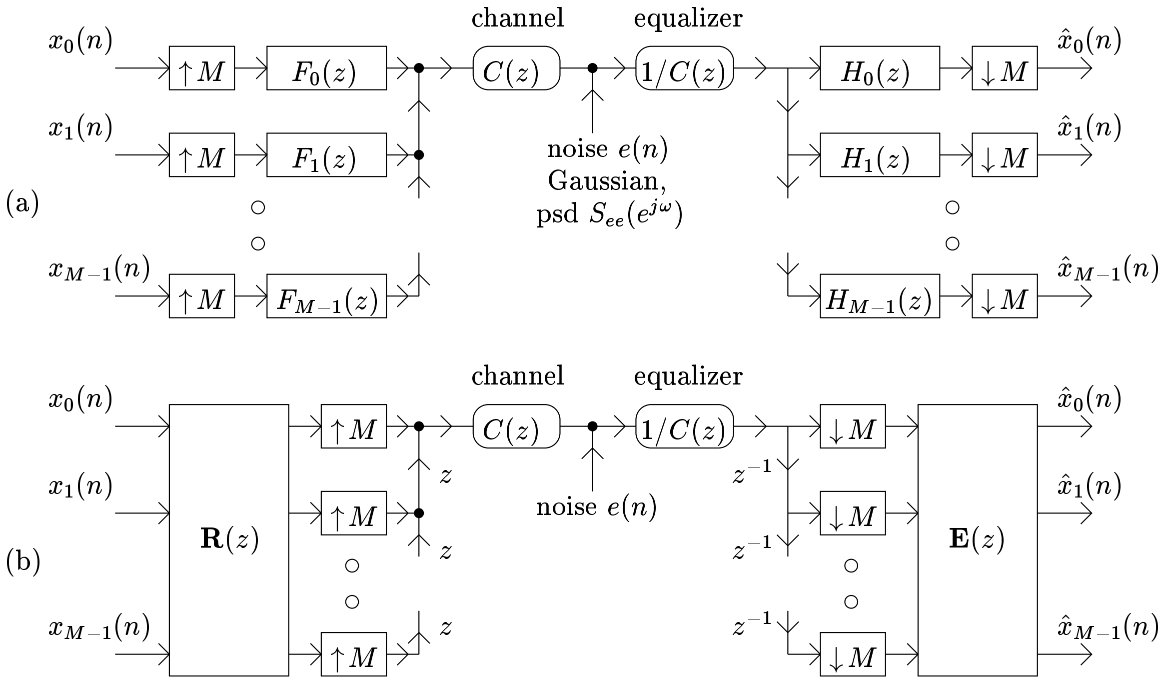


Figure 4.1: The DMT system: Filter bank used as transmultiplexer. (a) Showing analysis and synthesis filters, (b) polyphase representation.

Figs. 1.3,1.6.

The receiver in Fig. 4.1 contains an ideal channel equalizer which inverts the LTI channel $C(z)$. Practical systems approximate such an equalizer as follows: $C(z)$ is modelled as a rational function, and its denominator is cancelled out using a FIR filter. This step is called 'channel shortening', and the resulting effective channel is a FIR transfer function of reasonably low order L. This is equalized by adding redundancy, i.e., using an underdecimated FB. This is described in detail in [17] for DFT based DMT, where $\mathbf{E}(z), \mathbf{R}(z)$ in Fig. 4.1b are respectively the DFT matrix and its inverse. Briefly, the order L FIR channel performs a convolution on the interleaved samples coming from the output of the inverse DFT matrix. Making a periodic extension of these samples using a cyclic prefix of length L will cause the effect of the channel to appear like a *circular* convolution. Now circular convolution on a vector corresponds to pointwise multiplication on its DFT. Thus the channel can be equalized by simply using a single constant multiplier k_i on each of the signals $\hat{x}_i(n)$ received at the output of the DFT matrix. This system is hence called the frequency domain equalizer (FEQ), as opposed to the FIR 'channel shortening' filter which is called the time domain equalizer (TEQ). Generalizations of the cyclic prefix and FEQ to DMT using other FBs (distinct from the DFT) have also been advanced [41].

To elaborate on the parallels between FBs and transmultiplexers, recall the motivation for using FBs in data compression (Section 1.1.2): We could allocate bits to the subbands depending on the strengths of their signals. Chapter 2 showed that we reap the maximum benefit from this process when the FB used is the PCFB for the input spectrum, which in some sense maximizes the disparity between the signal strengths in the various subbands. Similarly, in Fig. 4.1, we could spend less power for a given bitrate, or achieve a higher bitrate for the same power, on the subchannels $x_i(n)$ that 'see' a more favorable communications channel (low noise $S_{ee}(e^{j\omega})$ and high channel gain $C(e^{j\omega})$). Again, the maximum benefit in terms of power or bitrate is realized by maximizing the disparity between the effective channels seen by the different streams $x_i(n)$. More precisely, the PCFB for the effective noise spectrum $\frac{S_{ee}(e^{j\omega})}{|C(e^{j\omega})|^2}$ turns out to be the optimum orthonormal FB. In this chapter, we will formally state the FB optimization problem, obtain its objective, prove PCFB optimality and present simulation examples comparing the performance of various FBs.

4.2 Problem formulation and PCFB optimality

In the DMT system of Fig. 4.1, the channel signals $x_i(n)$ consist of symbols from some digital modulation constellation [48]. To be specific, we will assume that the $x_i(n)$ are b_i bit PAM (pulse amplitude modulation) symbols, i.e., each $x_i(n)$ is a random variable with 2^{b_i} equiprobable values $\pm S_i, \pm 3S_i, \ldots, \pm(2^{b_i}-1)S_i$. The average transmitted power in the i-th channel (i.e., $E[|x_i(n)|^2]$) is denoted by P_i. Since we assume perfect channel equalization and the FB used has PR in absence of channel impairments, the received signal $\hat{x}_i(n)$ consists of $x_i(n)$ corrupted by additive Gaussian noise. This noise is obtained by passing the channel noise $e(n)$ through the equalizer and the receive filter $H_i(z)$. We denote its variance by σ_i^2. The detector at the receiver estimates $x_i(n)$ from its 2^{b_i} possible values, based on $\hat{x}_i(n)$. This is done simply by finding the possible value that is closest in Euclidean distance to the received value. The probability of error $P_e(i)$ in this detection is purely a function of P_i and σ_i^2, and is calculated for various modulation constellations in [48]. Specifically for PAM,

$$P_e(i) = 2(1-2^{-b_i})\mathcal{Q}\left(\sqrt{\frac{3P_i}{(2^{2b_i}-1)\sigma_i^2}}\right), \tag{4.1}$$

where $\mathcal{Q}(x) = \frac{1}{\sqrt{2\pi}}\int_x^\infty e^{-v^2/2}\,dv$. Observe that the σ_i^2 are the subband variances of a FB in response to an input with power spectrum

$$S_{qq}(e^{j\omega}) = \frac{S_{ee}(e^{j\omega})}{|C(e^{j\omega})|^2}, \tag{4.2}$$

which we call the effective noise spectrum. Given the acceptable error probability $P_e(i)$, (4.1) relates the power requirement P_i to the achieved bitrate b_i. The relation involves the effective noise variance σ_i^2, which we can control by choice of the FB. The problem of optimizing the (orthonormal) FB for DMT can now be posed in two ways, as we describe next.

Minimizing average transmitted power P for given bitrate: From (4.1), we have

$$P = \frac{1}{M}\sum_{i=0}^{M-1} P_i = \frac{1}{M}\sum_{i=0}^{M-1} \beta(P_e(i), b_i)\ \sigma_i^2, \tag{4.3}$$

where the exact nature of the function β is not of immediate interest. Given the b_i and $P_e(i)$, what is the FB that minimizes P? We see that (4.3) is a concave (in fact linear) function of the variances σ_i^2. Thus, by Chapter 2, a PCFB corresponding these variances, i.e., to the spectrum $\frac{S_{ee}(e^{j\omega})}{|C(e^{j\omega})|^2}$, will be optimal. This is true for any fixed choice of the b_i. In particular, they can be chosen optimally, i.e., to minimize P subject to a constraint on the average bitrate $B = \frac{1}{M}\sum_i b_i$. This choice of course depends on the σ_i^2, i.e., on the FB. The best FB to use in conjunction with this choice is still the above–mentioned PCFB. The optimal choice of b_i can be explicitly computed if we make the approximation $1 - 2^{-b_i} \approx 1$ in (4.1). This is done using the AM–GM inequality in a manner similar to that used in Section 2.5.3 for Equation (2.11) for the case where $f_i(b_i) = c2^{-2b_i}$. This is possible because the objective (4.3) takes a similar form, i.e.,

$$\beta(P_e(i), b_i) = K(P_e(i))(2^{2b_i} - 1), \quad \text{where} \quad K(P_e(i)) = [\mathcal{Q}^{-1}(P_e(i)/2)]^2/3. \tag{4.4}$$

Maximizing average transmission bitrate B for given power: From (4.1), we have

$$B = \frac{1}{M}\sum_{i=0}^{M-1} b_i = \frac{1}{M}\sum_{i=0}^{M-1} \gamma_{(P_e(i),P_i)}(\sigma_i^2). \tag{4.5}$$

The function γ here is well defined from (4.1), but has no closed form (due to the way in which b_i occurs at two places in (4.1)), unless we take $1 - 2^{-b_i} \approx 1$, in which case

$$b_i = \gamma_{(P_e(i),P_i)}(\sigma_i^2) = 0.5\log_2\left(1 + \frac{P_i}{K(P_e(i))}\frac{1}{\sigma_i^2}\right). \tag{4.6}$$

(Here $K(P_e(i))$ is as in (4.4).) This is easily verified to be convex in σ^2 for any $P_e, P \geq 0$. In fact, this convexity of the function γ is provable even without the above approximation [66]. Thus, again the PCFB for the effective noise spectrum (4.2) is optimal, i.e., maximizes the bitrate B of (4.5) for any given P_i and $P_e(i)$. Here too, one can choose the powers P_i optimally, so as to maximize (4.5) subject to an average power constraint $\frac{1}{M}\sum_{i=0}^{M-1} P_i = P$. A closed form for this allocation is possible using the approximation (4.6); it takes the form of a water-filling solution as we describe next.

Capacity, and water-filling. The DMT system of Fig. 4.1 can be represented by

$$\hat{x}_i(n) = x_i(n) + q_i(n) \tag{4.7}$$

where $q_i(n)$ are the effective subband noise processes which are Gaussian with variances σ_i^2 as described earlier. In general these noises are not white and uncorrelated, but this is approximately true with large number of channels M if the FB is a reasonable approximation of the ideal brickwall FB (of Fig. 1.4). In this case, the system is identical to the **parallel Gaussian channel**, well studied in information theory [18], with Shannon capacity

$$\mathcal{C} = 0.5 \sum_{i=0}^{M-1} \log_2 \left(1 + \frac{P_i}{\sigma_i^2}\right). \tag{4.8}$$

This is the upper bound on the achievable bitrate with any channel coding strategy with error probability tending to zero as the coding block size increases. Equations (4.5),(4.6) on the other hand represent the actual achieved rate with fixed (nonzero) error probabilities $P_e(k)$ without any error control coding. Notice that the terms in the sum in (4.8) have exactly the same functional form as (4.6), except for the factor $K(P_e(i))$. Thus, if the terms in (4.8) have to equal the corresponding b_i of (4.6) (making $B = \mathcal{C}$), the actual power P_i used must be more than that in (4.8) by a factor of $K(P_e(i))$ which amounts to an increase of **9.75 dB** at $P_e(i) = 10^{-7}$. Channel coding is included in many DMT systems to reduce this gap.

The capacity (4.8) depends on the σ_i^2 and hence on the choice of FB. The identical functional forms of (4.8) and (4.6) show that the PCFB maximizes not only the achieved rate but also the information theoretic capacity (as (4.8) is also convex in the vector of the variances σ_i^2). The problem of allocating the powers P_i to maximize (4.8) given the σ_i^2 and a constraint on the average power $P = \frac{1}{M}\sum_{i=0}^{M-1} P_i$ has been well studied [18]. The solution takes the 'water-filling' form

$$P_i = \begin{cases} \lambda - \sigma_i^2 & \text{if this is non-negative} \\ 0 & \text{otherwise} \end{cases} \tag{4.9}$$

where λ is chosen to satisfy the power constraint P. Clearly a similar solution also works for power allocation to maximize B of (4.5),(4.6).

4.3 The twisted pair channel

The twisted pair copper wire reaches every household with a telephone connection. DSL (digital subscriber line) technology [56] harnesses this bandwidth resource for high speed data transmission. Asymmetric DSL (ADSL), which uses the DMT system, has a downstream (central office to customer) signal band ranging approximately from 25 KHz to 1.6 MHz, and an upstream band from 25 KHz to 138 KHz. There has been a great deal of study, both theoretical and measurement based, on the nature of the twisted pair channel – both its transfer function and its noise sources [56],[68].

The channel transfer function $\left|C(e^{j\omega})\right|^2$ decreases with line length, and is in general decreasing

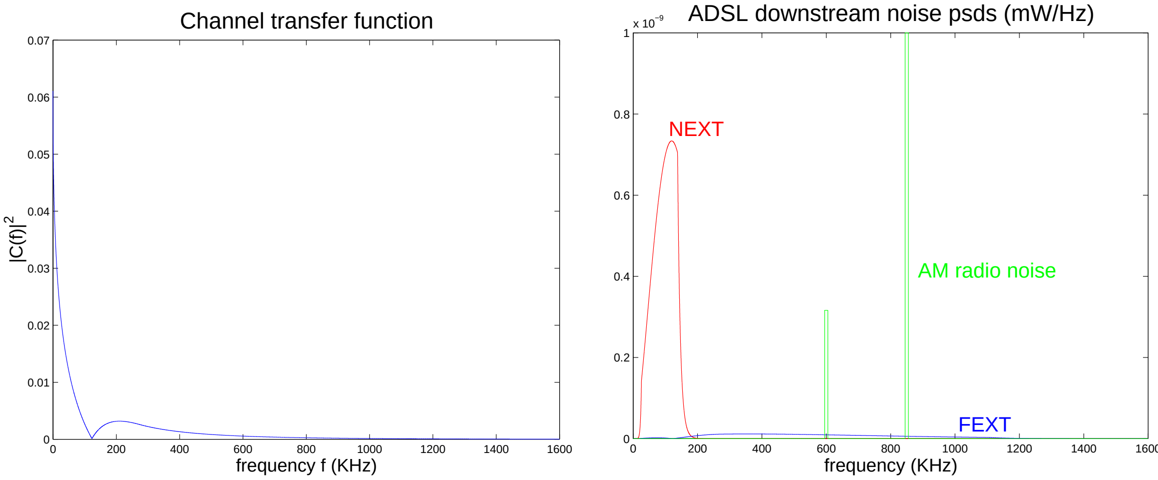


Figure 4.2: The ADSL downstream twisted pair channel transfer function and noise spectra.

in frequency. Bridged taps are typically attached to telephone lines in the US for service flexibility; these add nulls in the channel transfer function, at frequencies that depend on the lengths of the taps. The dominant noise sources are near end crosstalk or *next*, and far end crosstalk or *fext*. These are interferences that arise because several twisted pairs are usually placed in a single cable and hence suffer from electromagnetic coupling between them. The *next* noise is that coming from transmitters at the same end of the cable as the receiver, while *fext* is due to transmitters at the other end. Though these are interferences, they have characteristics of Gaussian noise [33] (which is reasonable by the central limit theorem). The power spectra of various DSL signals are documented, and given that all the interfering signals have the same psd, there are standard procedures [56] to compute the psd of the crosstalk noise. Another noise source is that from AM radio waves, entering in the band from 560 KHz to 1.6 MHz, and from amateur radio (HAM) stations, which lies in a higher band which is outside the standard ADSL bandwidth.

Figure 4.2 plots the ADSL downstream channel transfer function and noise spectra that we use to compare the performance of various FBs in the DMT system. The channel gain was obtained from an expression in [32] with a simulated effect of a bridged tap. The noise psds were obtained from [56] as outlined earlier. The system bandwidth chosen was from 0 to 1.6 MHz, and sampling rate for analog to digital conversion was the Nyquist rate of 3.2 MHz. A detailed description of the system specifications and the computation of these plots can be found in [66]. The resulting effective noise spectrum (4.2) is shown in Fig. 4.3. The ideal (unconstrained class) PCFB for a monotone spectrum is the traditional contiguous stacked brickwall FB of Fig. 1.4; however the spectrum of Fig. 4.3 is far from monotone, and the corresponding PCFB is quite different. Figure 4.4 compares the performance of various FBs, in terms of required power for a bitrate of 3.2Mb/s at error probability of 10^{-9} on all channels, for varying number of channels M in the DMT system. As expected, the

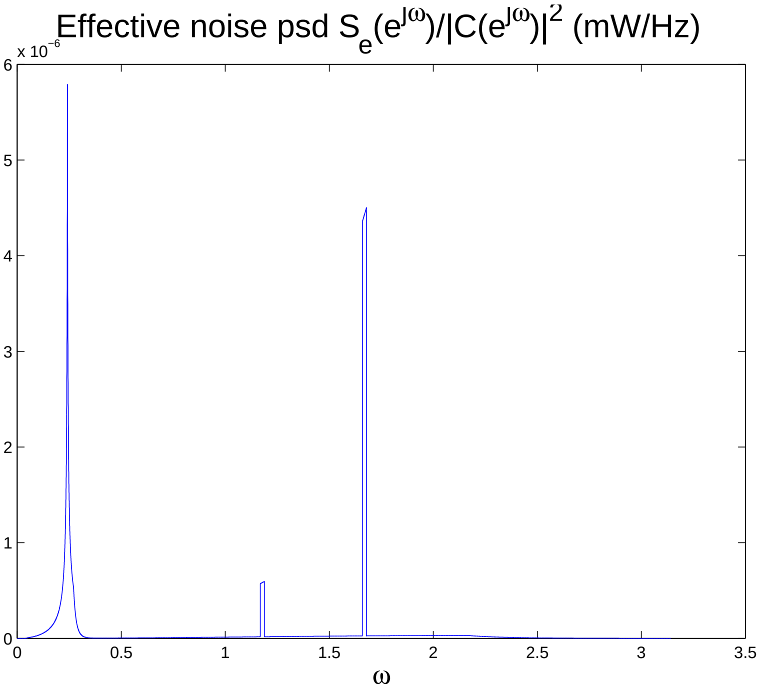


Figure 4.3: Effective noise spectrum for the ADSL channel.

ideal PCFB outperforms all other FBs, and the KLT, being a PCFB for the transform coder class, outperforms the other transform coders considered, namely the DFT and DCT.

For 8 channel FBs, the power saving by using the ideal (unconstrained) PCFB as opposed to the usual DFT is about a factor of five. This significant saving is however obtained at the cost of increased complexity of implementing the FB and continuously adapting its design to variations in the channel characteristics such as the noise spectrum. The ideal PCFB has unrealizable brickwall filters, which would have to be approximated in practice. Figure 4.5 shows two of the 8 filters in the ideal PCFB. We see that the filters can have multiple bands (owing to the 'bumpy' nature of the effective noise spectrum), and can be very hard to implement. Also, Fig. 4.4 shows that the performance gain from the PCFB reduces significantly as the number of channels is increased. Many practical ADSL systems use 512 or more channels. Thus the PCFB currently appears to be unattractive from the viewpoint of a practical implementation. Its value is more as an benchmark and upper bound for the achievable performance. Another issue that needs study is the role of PCFBs in DMT systems that use redundant FBs so that frequency domain equalization is possible. Some results for this case have been derived in [41]. Yet another issue is that of 'spectral compatibility': Using FBs different from the DFT will cause a change in the transmitted signal spectrum, which in turn has an effect on the interference noise spectrum. These effects need to be analyzed if FBs other than the DFT are to be considered for use in DMT.

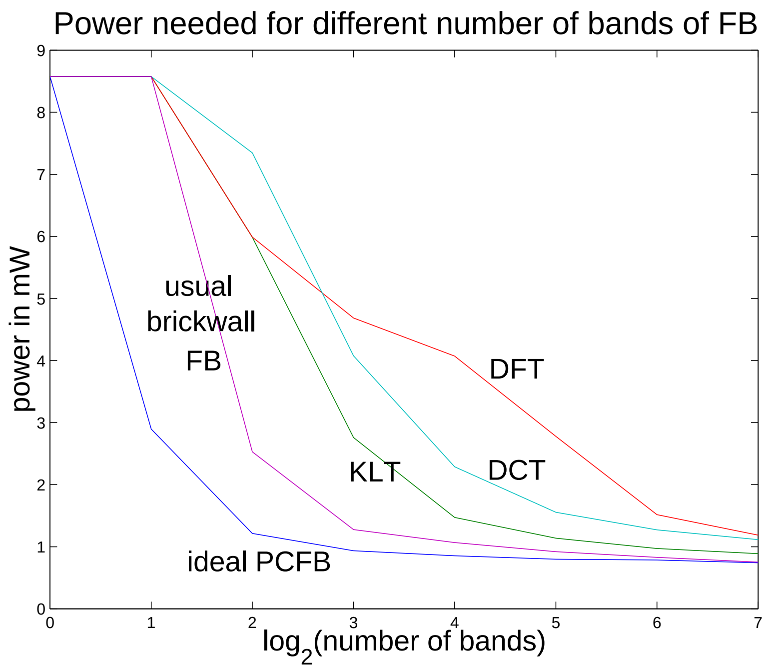


Figure 4.4: Performance of various filter banks on the ADSL channel.

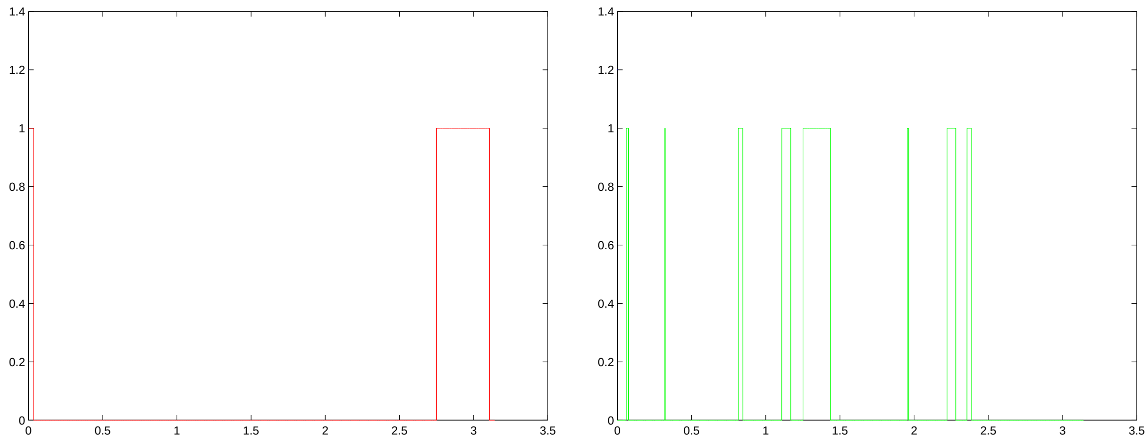


Figure 4.5: Two filters in the unconstrained 8 channel PCFB for the spectrum of Fig. 4.3.

4.4 Conclusion

We have seen that PCFBs are optimum orthonormal FBs for minimizing the transmitted power or maximizing both the achieved bitrate and the information theoretic capacity in DMT systems. The optimality follows from the same general framework of Chapter 2, due to the concavity of the relevant objective functions. The performance gain from the PCFB is obtained at the expense of complexity of implementing and adapting the FB to the channel and noise spectrum. The gain can be significant when the FB has a small number of channels M, but decreases for large M. PCFBs pose tough challenges from an implementation standpoint, but are useful as benchmarks for the achievable performance.

Chapter 5 Nonuniform filter banks: Optimization and PCFBs

In this chapter, we define the notion of a PCFB for a class of nonuniform orthonormal FBs with a fixed set of decimation rates. We then show how it generalizes the uniform PCFBs by being optimal for a certain family of concave objectives. We show that compared to previously attempted definitions, ours is the more natural generalization of the PCFB concept to nonuniform FBs. The earlier ones did not involve the concept of majorization which is crucial to uniform PCFB optimality, and hence the resulting PCFBs could not be shown to have any interesting optimality properties. We then study the issue of existence of nonuniform PCFBs, showing the following result that is important in this connection: For strictly monotone input power spectra, PCFBs do not exist for the class of unconstrained nonuniform FBs with any given set of decimators that are not all equal. In contrast, the class of unconstrained uniform M channel orthonormal FBs always has a PCFB for any input spectrum. Thus PCFB existence for nonuniform FB classes is much more delicate than that for uniform ones. The results are mostly drawn from [10],[11].

5.1 Problem formulation

Figure 5.1 shows a general subband signal processing scheme using an M channel filter bank (FB). We always assume the FB to be *maximally decimated*, i.e. its decimators n_i obey $\sum_{i=0}^{M-1} \frac{1}{n_i} = 1$. The FB is uniform if all decimators n_i are equal, i.e. (by maximal decimation) $n_i = M$ for all i. We will only study *orthonormal* FBs, i.e. those having PR with $F_i(e^{j\omega}) = H_i^*(e^{j\omega})$ (explained in Section 1.1.3). The subband domain processing systems (e.g., for compression and noise suppression) considered in Chapter 2 remain well defined even if they use such a nonuniform FB with a given set

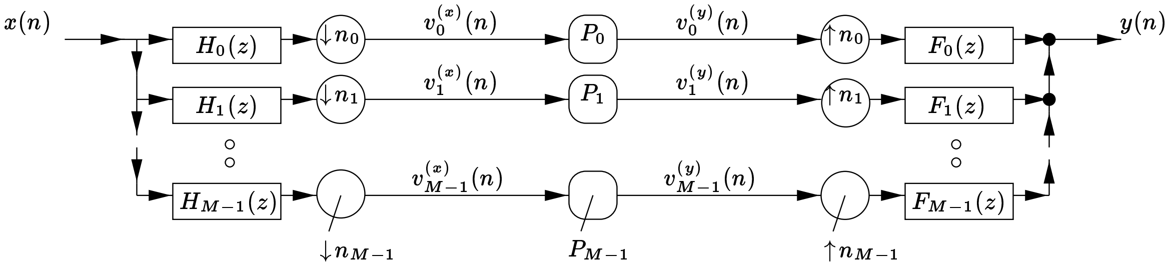


Fig. 1. General subband signal processing scheme using M channel filter bank.

Figure 5.1: General subband signal processing scheme using M channel filter bank.

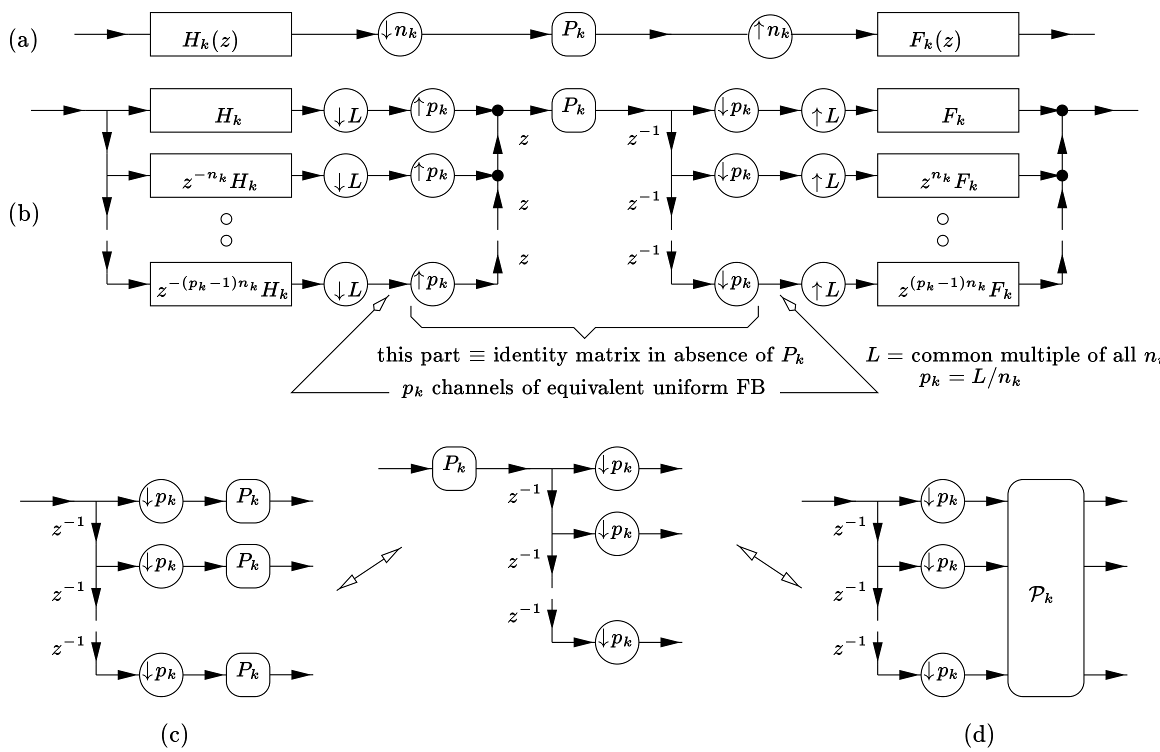


Figure 5.2: Transforming a nonuniform FB to an equivalent uniform one. (a) A single subband. (b) Corresponding subbands of equivalent uniform FB. (c) Moving processor P_k from nonuniform to uniform subbands — special case, (d) general case.

of decimators instead of a uniform FB. Here we aim to generalize Chapter 2 to such a situation, i.e., to find the best FB in a given class containing such nonuniform FBs. To do this, the first step is to study the new form of the minimization objective, generalizing the analysis of Section 2.3 that led to the corresponding uniform FB optimization objective form (2.3) which we repeat here:

$$g(\mathbf{v}) = \frac{1}{M}\sum_{i=0}^{M-1} h_i(\sigma_i^2), \tag{5.1}$$

where $\mathbf{v} = (\sigma_0^2, \sigma_1^2, \ldots, \sigma_{M-1}^2)^T$ is the subband variance vector.

The key step in making this generalization, and indeed in many dealings with nonuniform FB classes, is the transformation [21, 3, 39] from the nonuniform FB to a uniform L channel FB where L is any common multiple of the decimators n_i (usually $L = \text{lcm}\{n_i\}$). This transformation is shown in Fig. 5.2a,b. The k-th channel of the nonuniform FB, with decimator n_k, corresponds to $p_k = L/n_k$ channels of the uniform one. The filters in these p_k channels are delayed versions of each other. Due to these dependencies between the filters of the equivalent uniform FB, it is not possible to redraw every L channel uniform FB as a nonuniform FB with the given decimators n_i (unless we allow the nonuniform FB to have filters that are *periodically time varying* with period L, see Chapter 6).

Note from Fig. 5.2 that many properties such as maximal decimation, perfect reconstruction and orthonormality are shared in common by the nonuniform FB and its equivalent uniform FB (i.e. each has the property if and only if the other does).

The above transformation makes it easy to formulate the minimization objective for nonuniform FBs starting from the earlier formulation for uniform ones. Let $w_i^{(s)}(n)$ denote the signals in the equivalent uniform FB subbands when the FB input is $s(n)$. The notation $v_i^{(s)}(n)$ from Section 2.3 is used for the nonuniform FB subbands. Both (2.1) and (2.2) can be generalized to nonuniform FBs; the issue to be resolved is whether the subband signals in these equations should now refer to the uniform subbands $w_i^{(e)}(n)$ or to the nonuniform ones $v_i^{(e)}(n)$. Evidently (2.1) generalizes as $\varepsilon = \sum_{i=0}^{L-1} E[|w_i^{(e)}(n)|^2]$, under the assumption that the signals $w_i^{(e)}(n)$ are jointly WSS. On the other hand, (2.2) reflects the action of processor P_i on its WSS input (a subband signal), and hence generalizes as it is, under the assumption that the signal $v_i^{(x)}(n)$ is WSS with variance σ_i^2. Finally, examine the signals $w_i^{(e)}(n)$ for the group of p_k uniform FB subbands in Fig. 5.2b. By linearity of the systems in Fig. 5.2, these signals are obtained by passing the signal $v_k^{(e)}(n)$ of the corresponding nonuniform subband through the p_k channel delay–chain. Thus if $v_k^{(e)}(n)$ is WSS, all these $w_i^{(e)}(n)$ have the same variance equal to that of $v_k^{(e)}(n)$. Combining these observations yields the minimization objective for nonuniform FB classes:

$$g(\mathbf{v}) = \sum_{i=0}^{M-1} \frac{p_i}{L} h_i(\sigma_i^2) = \sum_{i=0}^{M-1} \frac{1}{n_i} h_i(\sigma_i^2) \tag{5.2}$$

This is a very natural generalization of (5.1), obtained by appropriate generalization of the choice of stationarity assumptions.

Notice that (5.2) is exactly the objective that would result for the equivalent uniform FB optimization if the processor P_k in Fig. 5.2b were replaced by p_k identical copies of itself, one in each of the corresponding uniform FB subbands as demonstrated in Fig. 5.2c. For a general processor P_k, this operation does not preserve the system, as that requires a p_k-input p_k-output processor in the uniform FB subbands as shown in Fig. 5.2d. However the two systems are identical from the viewpoint of the minimization objective.[1] In spite of this, there is a difference between the equivalent uniform FB optimization and the optimizations considered in Chapter 2: In Chapter 2, we were allowed to couple any subband to any subband processor (see Section 2.3.1, permutation symmetry of search space), whereas here the coupling is constrained by the fact that subbands derived from the same nonuniform FB subband must have identical processors. This constraint makes PCFB optimality somewhat more restricted for nonuniform FBs, as we will seen in Section 5.2.2.

We now briefly review the FB based signal processing systems studied in Chapter 2, and describe

[1] For the compression and noise suppression problems studied in Chapter 2, in fact it can be seen that the systems are also equivalent, e.g., this is clear for the noise suppression system because P_k is a constant multiplier.

how they satisfy the statistical assumptions leading up to (5.2). The functions h_i in these systems are the same as they were in Chapter 2, and thus they are concave on $[0, \infty)$, so that in the special case of uniform FB classes ($n_i = M$ for all i), PCFBs are optimal (Chapter 2). The later sections of this chapter seek to generalize this to nonuniform FB classes.

Compression (Section 2.5.3): Here the processor P_i is a b_i bit quantizer, modelled as an additive noise source with variance $f_i(b_i)\sigma_i^2$, where σ_i^2 is the variance of its (WSS) input. The function f_i is assumed to be independent of choice of FB. Thus (5.2) holds with $h_i(x) = f_i(b_i)x$. For uniform FBs, we need the assumption that the quantizer noises are all jointly WSS. For nonuniform FBs, we require the corresponding noises generated in the subbands of the equivalent uniform FB to be jointly WSS. The stronger assumption of WSS and *uncorrelated* quantizer noises implies both these.

The special case where the quantizers are assumed to obey the high bitrate assumption is worth mention: Here $f_i(b_i) = c_i 2^{-2b_i}$ where c_i depends on the i-th subband probability density function (pdf), and hence on the choice of FB. This is not allowed under our formulation of (5.2) (the functions h_i are assumed to be independent of FB); however we circumvent the problem by assuming a Gaussian input. This makes all subband pdfs Gaussian, hence forcing all the c_i to be equal and constant, independent of the FB. In this case, under optimal allocation of the bits b_i (subject to a total bitrate constraint $\sum_{i=0}^{M-1} b_i = B$), it can be shown (using the arithmetic mean - geometric mean inequality) that minimizing (5.2) is equivalent to minimizing the weighted geometric mean $GM = \prod_{i=0}^{M-1}(\sigma_i^2)^{(1/n_i)}$ of the subband variances σ_i^2 (with the decimators n_i as weights). Equivalently, we can minimize $\log(GM)$, which again has the form of (5.2) with $h_i(x) = \log(x)$ for all i.

Noise reduction (Section 2.7): Here the FB input is $x(n) = s(n) + \mu(n)$ where $s(n)$ is the pure input and $\mu(n)$ is white noise uncorrelated to $s(n)$. Each processor P_i is a multiplier k_i that is either constant or adapted to its input statistics as (a) a Wiener filter: $k_i = \frac{\sigma_i^2}{\sigma_i^2+\eta^2}$, or (b) a subband hard threshold operator: $k_i = 0$ if $\sigma_i^2 < \eta^2$ and $k_i = 1$ otherwise. Here σ_i^2 is the variance of the pure signal component $v_i^{(s)}(n)$ in the i-th subband, and η^2 is the noise variance. In all these cases, (5.2) applies with

$$h_i(x) = \begin{cases} |1-k_i|^2 x + |k_i|^2 \eta^2 & \text{for constant subband multiplier } k_i \\ \frac{x\eta^2}{x+\eta^2} & \text{for zeroth order Wiener filter } k_i \\ \min(x, \eta^2) & \text{for hard threshold } k_i \end{cases} \tag{5.3}$$

In Section 2.7, we allowed $s(n)$ to be CWSS(M) (for M channel uniform FB optimization). Here we need all the nonuniform FB subbands to be WSS, for which the correct generalization is to allow $s(n)$ to be CWSS(g) where g is the gcd of all the decimators n_i (in particular, $s(n)$ could be WSS).

5.2 Nonuniform PCFBs: Definitions and optimality

As seen in Chapter 2, uniform PCFBs are optimum orthonormal FBs whenever the minimization objective is a concave function of the subband variance vector. This basic result leads to PCFB optimality for many FB based signal processing schemes, motivating us to generalize the PCFB concept to nonuniform FBs. To do this, we begin by noting, from Chapter 2, that uniform PCFB optimality is solely dependent on the following facts:

1. The objective function has form (5.1) where all h_i are *FB–independent* functions concave on the interval $[0, \infty)$ (Section 2.3).
2. The PCFB subband variance vector majorizes the subband variance vectors of all FBs in the given class $\mathcal{C}$. (Definition of PCFBs, Section 2.4.2.)
3. The entries of the variance vector are allowed to be inserted in the objective (5.1) in any order. In other words, if $\mathbf{v}$ is a variance vector realizable by the given class $\mathcal{C}$, so is any permutation of $\mathbf{v}$ (Section 2.3.1). This is because permuting $\mathbf{v}$ simply corresponds to permuting the subbands, or to redistributing the processors among them. The subbands have no intrinsic ordering – their numbering purely denotes their association with the subband processor. The best of the (finitely many) permutations of the PCFB for the class $\mathcal{C}$ is optimal for the entire class.

5.2.1 Nonuniform PCFBs: Definition

We seek to generalize the definition of PCFBs to a general class $\mathcal{N}$ of nonuniform M channel orthonormal FBs with a fixed set of decimators n_i. It is desirable that the generalization reduce to the usual definition for uniform FBs if all the n_i are equal, and that the defined PCFBs be optimal for some reasonable class of objectives of the form (5.2). From the discussion of uniform PCFB optimality, a natural approach to generalization is to define some suitable variance vector for all FBs in the class, and define the PCFB to be the one whose variance vector majorizes all the variance vectors of all FBs in the class $\mathcal{N}$. This PCFB would then be optimal for all objectives of the form $\sum_i f_i(\alpha_i)$ where α_i are the entries of the variance vector, provided f_i are FB–independent concave functions and the appropriate permutation–symmetry condition (analogous to fact 3 mentioned earlier) hold. The majorization requirement demands that $\sum_i \alpha_i$ be FB–independent; this should follow automatically from FB orthonormality just as it happens for uniform FBs. Otherwise it would impose too severe a constraint on the FB class $\mathcal{N}$, and PCFBs would hardly ever exist.

Let σ_i^2 be the variance of the i-th subband (with decimator n_i). FB orthonormality implies that $\sum_i \frac{\sigma_i^2}{n_i}$ is FB–independent, but in general $\sum_i \sigma_i^2$ is not (unless $n_i = M$ for all i). Thus, the usual way of defining the subband variance vector to have the σ_i^2 as its entries is unsuitable (except for uniform FBs). In fact the above considerations leave only two reasonable definitions of the variance vector:

- The *normalized* subband variance vector defined as $(\frac{\sigma_0^2}{n_0}, \frac{\sigma_1^2}{n_1}, \ldots, \frac{\sigma_{M-1}^2}{n_{M-1}})$.

- The *equivalent uniform* subband variance vector, defined as the (usual) subband variance vector of the equivalent uniform L channel FB (for any fixed L that is a multiple of all the n_i). From the construction of this FB (Fig. 5.2), with $p_i = L/n_i$, this vector has the form

$$(\underbrace{\sigma_0^2, \sigma_0^2, \ldots, \sigma_0^2,}_{p_0 \text{ elements}} \quad \underbrace{\sigma_1^2, \sigma_1^2, \ldots, \sigma_1^2,}_{p_1 \text{ elements}} \quad \ldots \ldots, \quad \underbrace{\sigma_{M-1}^2, \sigma_{M-1}^2, \ldots, \sigma_{M-1}^2}_{p_{M-1} \text{ elements}})^T \tag{5.4}$$

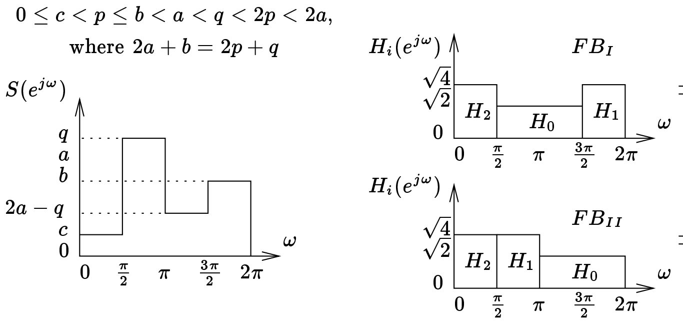
Definition: Nonuniform PCFBs. Let $\mathcal{N}$ be a class of orthonormal M channel FBs having a fixed set of decimation rates n_i, not necessarily equal. A FB in $\mathcal{N}$ is said to be a PCFB for $\mathcal{N}$ for the given input power spectrum if its suitably defined subband variance vector **majorizes** the subband variance vector of every FB in $\mathcal{C}$. The variance vector in this definition could be either the normalized or the equivalent uniform subband variance vector (defined earlier). Thus we have two definitions of nonuniform PCFBs, one based on each of these variance vectors.

Remarks on the nonuniform PCFB definitions.

1. *Reduction to uniform PCFBs.* For the special case of uniform FBs, where $n_i = M$ for all i, both the above definitions reduce to the usual one by choosing $L = M$ for the equivalent uniform variance vector and omitting the constant scale M for the normalized variances.

2. *Ambiguity in L does not matter.* Let $L_1 = \text{lcm}\{n_i\}$ and $L_2 = kL_1$ for any positive integer k. Take two FBs in the given class, having subband variance vectors $\mathbf{v}_1, \mathbf{v}_1^*$ for their equivalent L_1 channel uniform FBs, and $\mathbf{v}_2, \mathbf{v}_2^*$ for their equivalent L_2 channel uniform FBs. From (5.4), if the entries of $\mathbf{v}_1$ in descending order are $(a_0, a_1, \ldots, a_{L_1-1})$ then the entries of $\mathbf{v}_2$ in descending order are obtained by arranging the a_i in the same order and repeating each a_i k times. Thus by definition (2.7) of majorization, $\mathbf{v}_1^*$ majorizes $\mathbf{v}_1$ if and only if $\mathbf{v}_2^*$ majorizes $\mathbf{v}_2$. Hence, the PCFB definition using the equivalent L channel uniform subband variance vector is independent of which common multiple of the n_i we fix L to be.

3. *The two PCFB definitions are distinct.* It may seem that extending the above argument will also prove that using normalized variance vectors is equivalent to using the equivalent uniform ones. However this is not true. The reason is that the definition of majorization (2.7) demands arranging the entries of the vectors in decreasing order. This order could be different for the σ_i^2 (i.e. for the entries of (5.4)) and for the $\frac{\sigma_i^2}{n_i}$. The distinctness of the two definitions is shown by the example of Fig. 5.3: Consider the WSS input with spectrum as in Fig. 5.3a, and the class $\mathcal{N}$ having exactly the two filter banks FB_I, FB_{II} of Fig. 5.3b. The resulting subband variances are shown in Fig. 5.3c. From the relations among the parameters in Fig. 5.3a, we can verify



Subband variance vectors

usual	normalized	equivalent uniform
(a, b, c) $a > b > c$	$(\frac{a}{2}, \frac{b}{4}, \frac{c}{4})$ $\frac{a}{2} > \frac{b}{4} > \frac{c}{4}$	(a, a, b, c) $a = a > b > c$
(p, q, c) $q > p > c$	$(\frac{p}{2}, \frac{q}{4}, \frac{c}{4})$ $\frac{p}{2} > \frac{q}{4} > \frac{c}{4}$	(p, p, q, c) $q > p = p > c$

(a) (b) (c)

Figure 5.3: Distinctness of two nonuniform PCFB definitions: Class having two FBs with decimators $n_0 = 2, n_1 = n_2 = 4$. (a) Input spectrum $S(e^{j\omega})$, (b) the FBs, (c) their subband variance vectors.

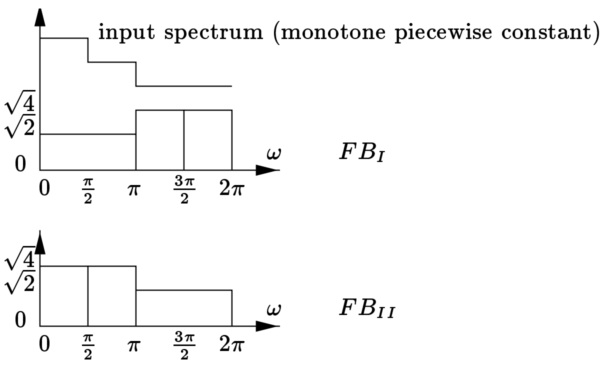


Figure 5.4: Distinctness of two PCFB definitions: Unconstrained FB class with decimators 2, 4, 4.

that FB_I is the unique PCFB by the normalized variances, while FB_{II} is the unique PCFB by the equivalent uniform ones. Figure 5.4 shows another example for a more natural FB class, namely that of all (unconstrained) orthonormal FBs with decimators 2, 4, 4. Here FB_I is a PCFB by normalized variances due to monotonicity of the input spectrum (Theorem 5.3, Section 5.3.3), while FB_{II} is a PCFB by equivalent uniform variances due to its brickwall nature and the whiteness of its subbands (Theorem 5.1, Section 5.3).

5.2.2 Nonuniform PCFBs: Optimality

For uniform FBs, there is no intrinsic ordering of the subbands. Permuting the subbands, or redistributing the processors P_i among the subbands, was simply represented by evaluating the same objective function $g(\mathbf{v})$ of (5.1) for a different argument $\mathbf{v}$, i.e. a permutation of the original subband variance vector. On the other hand, nonuniform FB subbands are indexed by their decimation rates n_i. If we interchange the processors for two subbands with different decimators, the new performance measure must be computed by interchanging not just the corresponding variances but also the corresponding decimators in (5.2). This changes the functional form of the objective, since the objective in (5.2) is viewed as a function of the variances σ_i^2 with the n_i as parameters. Free redistribution of processors among subbands without changing the functional form of the objective is possible only among groups of subbands with equal decimators. Thus, it is necessary to distinguish two different nonuniform FB optimization problems:

1. Finding the best FB for a fixed ordering of the subbands (i.e. of the n_i). We refer to this problem as $OP1$.

2. Optimizing both the FB (i.e. its subband variances) and the decisions as to which subband should have which processor P_i, i.e the choice of ordering of the subbands (or the n_i). We refer to this problem as $OP2$.

The problem $OP1$ does not exist for uniform FBs since there is no intrinsic ordering of the subbands. If $OP1$ is solvable for all the (finitely many) possible orderings of the subbands, then clearly $OP2$ is solved by picking from these solutions the one optimizing the performance measure. Note that $OP2$ is usually of much more interest than $OP1$, since fixing apriori which channel decimator should be associated with which processor P_i is quite restrictive. However $OP1$ is easier to analyze, due to the fixed functional form of its objective. A summary of all the optimality results for nonuniform PCFBs (to be derived next) can be found at the end of the section, and optimality there refers only to the (more interesting) problem $OP2$. The derivations involve study of both problems.

Optimality of PCFB defined using equivalent uniform variance vector: We generically denote the entries of this variance vector by α_i. The objective (5.2) has the form $\sum_i f_i(\alpha_i)$ with the f_i chosen as follows: All the $p_k = L/n_k$ equivalent uniform FB subbands derived from the k-th nonuniform FB subband (with decimator n_k) have the corresponding f_i equal to h_k of (5.2) (upto constant scale L). Thus this PCFB is a possible candidate for solving $OP1$ when all h_k (and hence all f_i) are FB–independent and concave. However for optimality, we must be allowed to insert the α_i into the f_i in any order in forming the sum $\sum_i f_i(\alpha_i)$ (condition 3 in Section 3.1). In actual fact we are only allowed certain permutations of the α_i, corresponding to groups of nonuniform subbands with equal decimator value. Thus the PCFB solves $OP1$ if the best ordering of the α_i happens to be one that is actually allowed.

Recall that solving $OP1$ for all permutations of the n_k solves $OP2$. For a given permutation of the n_k, suppose the PCFB does not actually solve $OP1$, i.e. the best ordering of the α_i above is not allowed. Then this ordering of the α_i gives an unattainable upper bound on the performance for the problem $OP1$. Sometimes we may be able to solve $OP2$ using just these bounds (instead of the true optima). Indeed if a known solution to $OP1$ for one permutation of the n_k outperforms all these bounds computed for all the other permutations, then it also solves $OP2$.

In the noteworthy special case when $h_k = h$ for all i, i.e. all h_k and hence all f_i are identical, we see that all orderings of the α_i give the same performance measure $\sum_i f_i(\alpha_i)$. Thus in this case the PCFB will solve $OP1$ if the function h is FB–independent and concave. This is true for all permutations of the n_k, and the corresponding optimal performance measures are identical too. Thus the PCFB also solves $OP2$. In fact the distinction between $OP1$ and $OP2$ vanishes in this case. Identical h_k usually results from similar or identical subband processors P_k.

Optimality of PCFB defined using normalized variance vector: We generically denote the entries of this variance vector by $\beta_i = \sigma_i^2/n_i$. The objective (5.2) has the form $\sum_i \hat{f}_i(\beta_i)$ where $\hat{f}_i(x) = (1/n_i)h_i(n_i x)$. Thus, this PCFB is a candidate for solving $OP1$ when the h_i (and hence the $\hat{f}_i$) are FB–independent and concave. However, again optimality is assured only if the β_i can be inserted in the sum $\sum_i \hat{f}_i(\beta_i)$ in any order. In actual fact we are only permitted to permute the β_i within groups of channels having the same decimator n_i. Thus, as with the earlier definition, these PCFBs solve $OP1$ if the best ordering of the β_i turns out to be one that is permitted. The special case where all $\hat{f}_i$ are identical (which would make all orderings of the β_i equally good) is not of much interest here: It does not commonly happen since $\hat{f}_i(x) = (1/n_i)h_i(n_i x)$ which depends on n_i.

However, another special case is of interest, namely that where $h_i(x) = k_i x$ for some constants k_i (for all i). Here $\hat{f}_i(x) = k_i x$ too, which has the speciality of being (concave and) independent of the decimators n_i. In general, to compute the performance measure after permuting two nonuniform subbands required not just permuting of the corresponding β_i, but also corresponding modifications of the $\hat{f}_i$ since they depended on the n_i. (This changed the functional form of the objective, and gave rise to the distinction between problems $OP1$ and $OP2$.) However if $\hat{f}_i$ are independent of the n_i, permuting the nonuniform subbands is fully equivalent to permuting the β_i. Thus, the nonuniform FB defined using the variances β_i will solve $OP2$ (for a suitable permutation of its subbands, i.e. ordering of the β_i).

Nonuniform PCFB optimality: Summary.

- Nonuniform PCFBs defined using the equivalent uniform subband variance vector are optimal for objectives of the form (5.2) when $h_i = h$ for all i, where h is a FB–independent concave function. Such objectives occur with similar processing in the subbands, e.g. in the high bitrate coding problem with optimal bit allocation ($h(x) = \log(x)$) or in the noise suppression problem using either zeroth order Wiener filters ($h(x) = \frac{x\eta^2}{x+\eta^2}$) or hard thresholds ($h(x) = \min(x, \eta^2)$) in *all* subbands (see Section 5.1).

- Nonuniform PCFBs defined using the normalized variance vector are optimal for objectives of the form (5.2) where for all i, $h_i(x) = k_i x$ for constant k_i. Currently we do not know of general signal processing schemes resulting in such objectives. The only example known is the highly degenerate case of colored noise suppression discussed in Section 3.2.1, where the noise spectrum is obtained from the signal spectrum by a constant scaling c. Here the h_i are linear and given by (3.4) with $f_i(b_i)$ replaced by c. The objective for the coding problem with fixed bit allocation (Section 5.1 or 2.5.3) may seem to be another such example (with $k_i = f_i(b_i)$), but it is not. This is because in optimizing this objective, we have an overall bitrate constraint $\sum_i (b_i/n_i) = B$. In the case of uniform FBs, where all n_i are equal, this constraint was symmetric in the b_i, whereas now it is no longer so. Thus, the free permutation of subband processors within the subbands (necessary to PCFB optimality) is no longer possible.

- For general objectives of the form (5.2) where h_i are FB–independent and concave, it may be possible to show optimality of either of the PCFBs in a specific case, using the actual specifications of the h_i and numerical values of the PCFB subband variances. Finite procedures to do this follow from the earlier discussion. However in general the procedure will be unsuccessful, and PCFB optimality is not assured.

- The above discussion indicates that the PCFB defined using equivalent uniform subband variances is optimal for a broader class of problems. It should thus be the preferred one when it comes to choosing between the two definitions we have provided for nonuniform PCFBs.

5.2.3 Previous attempts at defining nonuniform PCFBs

To our knowledge, the works [37, 65] are the only earlier attempts to define nonuniform PCFBs. These can be described as follows: Let $\mathcal{N}$ be the given class of M channel orthonormal FBs, all having the same set of decimators n_i. First an ordering of the subbands of all FBs (i.e. of the n_i) is decided upon. Then with σ_i^2 as the variance of the i-th subband (with decimator n_i), the PCFB is defined as one whose variances $\hat{\sigma}_i^2$ satisfy

$$\sum_{i=0}^{P} \frac{\hat{\sigma}_i^2}{n_i} \geq \sum_{i=0}^{P} \frac{\sigma_i^2}{n_i} \quad \text{for } P = 0, 1, \ldots, M-1, \quad \text{with (automatic) equality for } P = M-1 \tag{5.5}$$

While this equation is superficially similar to (2.7) that defines majorization, there is a very important difference: The a_i and b_i of (2.7) are arranged in decreasing order, whereas the analogous quantities $\frac{\hat{\sigma}_i^2}{n_i}$ and $\frac{\sigma_i^2}{n_i}$ in (5.5) are arranged according to the predefined ordering of the n_i. In [65] the PCFB is defined only for classes of dyadic (or 'wavelet style') tree–structured FBs, and the n_i are chosen in decreasing order. A construction procedure is outlined for the PCFB thus defined, when the FB class $\mathcal{N}$ is the unconstrained one. In [37] the PCFB is defined with respect to the ordering of the n_i, i.e. there are several PCFBs defined as above, one for each ordering of the n_i.

Consider the case when the minimization objective is as in (5.2) with $h_i(x) = k_i x$ for constant k_i (for all i). Consider the FB optimization problem $OP1$ defined in Section 3.3, where the n_i are ordered apriori and the association between the n_i and k_i is thus forcibly fixed. Let j_i be the permutation such that $k_{j_0} \leq k_{j_1} \leq \ldots k_{j_{M-1}}$. Then elementary algebra shows that the PCFB defined in [37] as above, for the permutation j_i of the decimators n_i, is a solution for $OP1$. Indeed it suffices to prove this assuming $j_i = i$. This amounts to showing that $\sum_i k_i\hat{\sigma}_i^2/n_i \leq \sum_i k_i\sigma_i^2/n_i$ given (5.5) and $k_i < k_{i+1}$. The proof follows from the fact that

$$\sum_{i=0}^{M-1} k_i \frac{\sigma_i^2}{n_i} = \sum_{i=0}^{M-2} (k_i - k_{i+1}) \sum_{l=0}^{i} \frac{\sigma_l^2}{n_l} + k_{M-1} \sum_{i=0}^{M-1} \frac{\sigma_i^2}{n_i}. \tag{5.6}$$

In summary, the earlier definitions result in several PCFBs, one for each permutation of the decimators n_i, and in optimality for the problem $OP1$ for linear objectives (i.e. $h_i(x) = k_i x$ in (5.2)). However, in general optimality for the more interesting problem $OP2$ cannot be claimed unless *all* the PCFBs (for all permutations of n_i) exist (in which case the best one of these would solve $OP2$). Apart from the very special result in [65] (specific to unconstrained dyadic tree–structured FB classes and using a particular ordering of subbands), there are no general existence results known for these PCFBs. Further, since the PCFB definition does not use the concept of majorization, there are no optimality results for general concave h_i (besides $h_i(x) = k_i x$). These observations show that our definitions of Section 5.2.1 more naturally generalize the PCFB concept to nonuniform FBs.

5.3 Existence of nonuniform PCFBs

We begin by discussing existence issues for nonuniform PCFBs defined via equivalent uniform FBs. As seen in Section 5.2.2, this is the preferred nonuniform PCFB definition, leading to broader optimality properties. For these PCFBs, we will state an easily proved existence result (Theorem 5.1) and a more nontrivial nonexistence result (Theorem 5.2). The latter is discussed in Section 5.3.1 and proved in Section 5.3.2. In these sections, by PCFBs we will always mean those defined by equivalent uniform FBs (even if this is not explicitly stated). In Section 5.3.3 however, we deal exclusively with the other PCFB definition, establishing an existence result (Theorem 5.3).

Let $\mathcal{N}$ be any class of nonuniform FBs with a given set of decimators n_i. Let $\mathcal{E}$ be the corresponding class of equivalent uniform FBs derived from FBs in $\mathcal{N}$. By the PCFB definition (using equivalent uniform FBs), PCFBs for $\mathcal{N}$ and $\mathcal{E}$ are equivalent. Now if the PCFB for some FB class $\mathcal{C} \supset \mathcal{E}$ lies in $\mathcal{E}$, then it is also a PCFB for $\mathcal{E}$. For example we could take $\mathcal{C} = \mathcal{C}^u$, the unconstrained class, whose PCFB has been well studied (Section 2.5.2). This gives rise to the following result:

Theorem 5.1. Suppose a FB in a (nonuniform) FB class $\mathcal{N}$ produces subbands that are totally decorrelated and *white*. Then it is a PCFB for $\mathcal{N}$. In other words, it a PCFB for the class of all (unconstrained) orthonormal FBs with the given set of decimators. For example, a FB with nonoverlapping brickwall analysis filters is a PCFB if the input spectrum is constant on the support of each filter (Fig. 5.5a). ∎

The result immediately follows from the fact, evident from Fig. 5.2, that under its premise, the

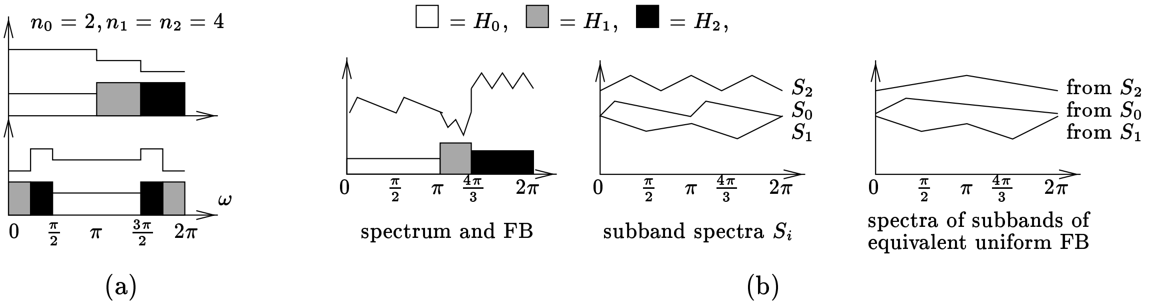


Figure 5.5: Nonuniform PCFBs for certain input spectra. (a) Simple spectra (Theorem 5.1). (b) More complex spectra (for $n_0 = 2, n_1 = 6, n_2 = 3$).

equivalent uniform FB has white and totally decorrelated subbands too, making it the unconstrained class PCFB. (It obeys total decorrelation and spectral majorization, Section 2.5.2.) More generally, suppose the i-th nonuniform subband, with decimator n_i, has a spectrum of form $A_i(e^{j\omega p_i})$ where $p_i = L/n_i$ and L is any common multiple of the n_i. Then, some thought (using Fig. 5.2) shows that the p_i corresponding subbands in the equivalent L channel uniform FB are totally decorrelated with identical spectra $A_i(e^{j\omega})$. Thus, if the nonuniform subbands are totally decorrelated, so are the uniform ones, and further if the $A_i(e^{j\omega})$ obey spectral majorization, then the uniform FB is an unconstrained class PCFB, and hence so is the nonuniform one. This allows creation of more complex input spectra (e.g., as in Fig. 5.5b) for which PCFBs exist (the PCFBs being appropriate brickwall FBs). We now study an important *nonexistence* result for nonuniform PCFBs, namely:

Theorem 5.2. Let $\mathcal{N}^u$ be the class of unconstrained nonuniform FBs with a given set of decimators n_i *not all equal.* Let the input spectrum $S(e^{j\omega})$ be *strictly monotonic* (increasing or decreasing) for $a \leq \omega < a + 2\pi$ for some real a. Then there is no PCFB for the class $\mathcal{N}^u$. ■

5.3.1 Discussion on Theorem 5.2

Existence of nonuniform PCFBs is very delicate. As shown in Fig. 5.6, a slight perturbation of a monotone input spectrum for which a PCFB for $\mathcal{N}^u$ exists (e.g., the flat or white spectrum) can change it to a spectrum that is strictly monotone (on an interval of form $[a, a + 2\pi)$), destroying existence of the PCFB. In contrast, for the *uniform* unconstrained FB class $\mathcal{C}^u$, PCFBs exist for all input spectra, and moreover usually several perturbations on the spectrum do not even change the PCFB. Of course, most objective functions defined on the FB class have their values for each FB perturbed only slightly by small perturbations of the input spectrum. Thus, consider the FB of Fig. 5.6. Being a PCFB for spectrum A, it optimizes many concave objectives as explained in Section 5.2.2. For the perturbed spectrum B, the FB is no longer a PCFB, and so it will not optimize all these objectives. However it will be quite close to optimal if the perturbation is small enough.

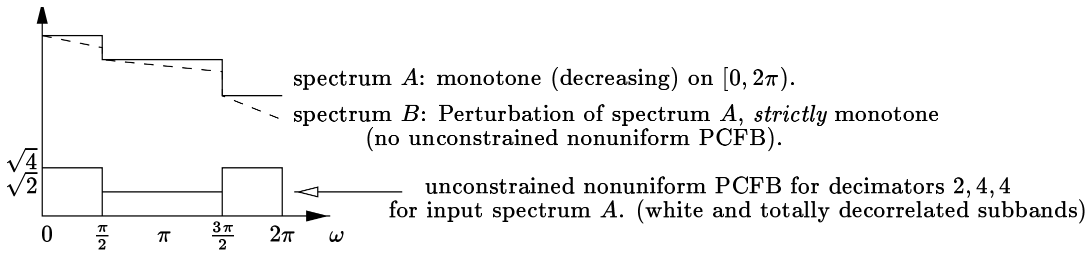


Figure 5.6: Delicateness of existence of nonuniform PCFB for unconstrained class.

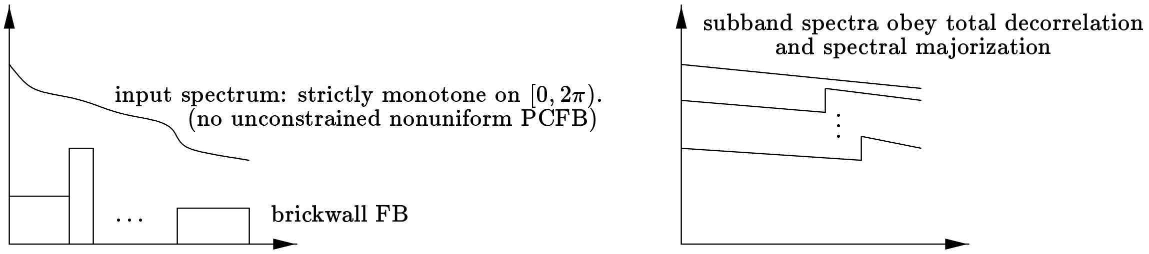


Figure 5.7: Total decorrelation and spectral majorization do not imply nonuniform PCFBs.

Total decorrelation and spectral majorization do not imply nonuniform PCFBs. For the class $\mathcal{C}^u$, we know (Section 2.5.2) that FBs with subbands obeying the properties of spectral majorization and total decorrelation are PCFBs. These properties are evidently well defined for subbands of nonuniform FBs too. Spectral majorization is the property that the subband spectra do not 'cross each other', and total decorrelation is achieved for example if the analysis filters have nonoverlapping supports, just as for uniform FBs. Thus, as shown in Fig. 5.7, if the input spectrum is monotone on $[a, a + 2\pi)$ then any contiguous–stacked brickwall FB with a as a filter band–edge has subbands obeying both these conditions. However the FB is not a PCFB for $\mathcal{N}^u$ if the spectrum is strictly monotone (since there is no PCFB in this case). Thus, these conditions do not imply the principal component property for $\mathcal{N}^u$ (in contrast with the case of $\mathcal{C}^u$).

5.3.2 Proof of Theorem 5.2

Without loss of generality, let $n_0 \geq n_1 \geq \ldots \geq n_{M-1}$. We prove the theorem for $a = 0$ assuming that $S(e^{j\omega})$ is strictly *decreasing* and that $n_0 \neq n_1$ (i.e. $n_0 > n_1$). The proof will show that this does not lose generality. The basic idea of the proof is as follows: Let $\mathcal{E}^u$ be the class of equivalent uniform FBs derived from FBs in $\mathcal{N}^u$.

1. We apply the sequential compaction algorithm of Section 3.1.3 to the uniform FB class $\mathcal{E}^u$, and show that this results uniquely in the FB denoted by FB_A in Fig. 5.7 (more precisely, in the equivalent uniform FB derived from FB_A). From Section 3.1.4, this means that if the class $\mathcal{N}^u$ has a PCFB for the psd $S(e^{j\omega})$, the PCFB must be FB_A.

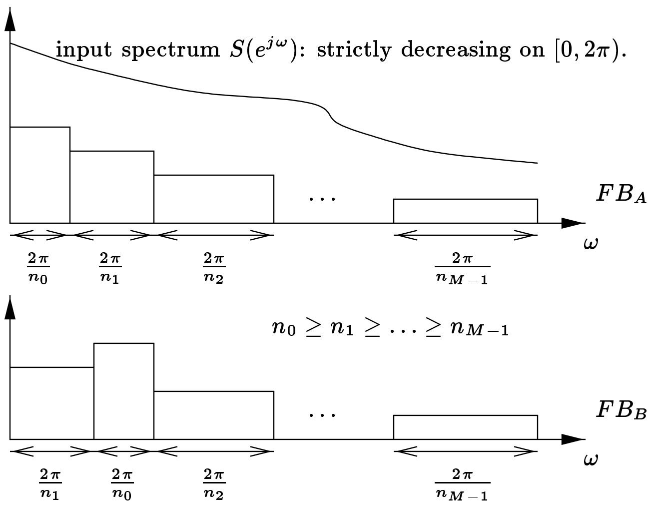


Figure 5.8: Brickwall FBs used in proving Theorem 5.2.

2. We then show that FB_A is not the PCFB by proving that its equivalent uniform subband variance vector does not majorize the variance vector of the distinct FB in $\mathcal{N}^u$ denoted by FB_B in Fig. 5.8. This means that there is no PCFB.

Proof of Step 1: The sequential algorithm involves maximizing the largest variance in the appropriate variance vector, then searching among all variance vectors having this maximum possible largest variance in order to maximize the second–largest variance, and so on. The entries of the equivalent uniform variance vector (5.4) are the nonuniform subband variances repeated appropriately many times. Hence running the algorithm on $\mathcal{E}^u$ is equivalent to sequentially maximizing the nonuniform FB subband variances. The ideal *compaction*(M) *filter* [61] produces the largest possible variance in a subband with decimator M, in any orthonormal FB. For a spectrum that strictly decreases on $[0, 2\pi)$ the compaction(M) filter is unique, supported on $[0, 2\pi/M]$, and has output variance that strictly increases in M. As n_0 is the largest decimator in the FB (i.e., the largest possible M), this proves that among all FBs in $\mathcal{N}^u$, FB_A maximizes the largest subband variance. (This maximum variance is produced by the filter in FB_A corresponding to decimator n_0.) Now all FBs are orthonormal, and so their analysis filters $H_i(z)$ obey $[H_i(e^{j\omega})H_j^*(e^{j\omega})]\downarrow_{\gcd(n_i,n_j)} = 0$ for $i \neq j$ (Section 6.2.2). Hence, if an analysis filter in the FB, corresponding to decimator M, has an aliasfree(M) support, then this support does not overlap with the supports of any of the other analysis filters. Thus we can repeat the same argument for the second–largest variance, and so on. This shows that FB_A is the unique output of the sequential compaction algorithm.

Proof of Step 2: If a filter in an orthonormal FB is ideal, supported on $[c, d]$ and has constant magnitude on its support (like all the filters in FB_A, FB_B), then the corresponding subband variance

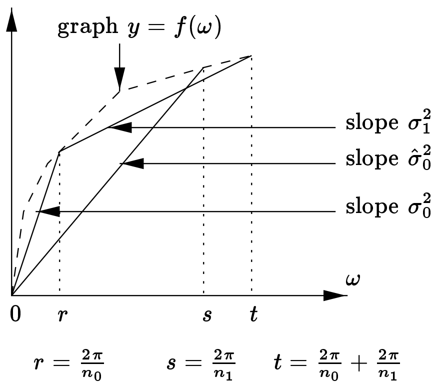


Figure 5.9: Subband variances as chord slopes.

is given by direct calculation as

$$\frac{f(d) - f(c)}{d - c}, \quad \text{where} \quad f(\omega) = \int_0^{\omega} S(e^{j\omega'})\, d\omega'. \tag{5.7}$$

The calculation uses two consequences of FB orthonormality, namely that the support length $d - c = 2\pi/n_i$ where n_i is the corresponding channel decimator, and that the constant passband magnitude of the filter is $\sqrt{n_i}$. Notice that the variance expression (5.7) is the slope of the chord on the graph of $y = f(\omega)$, connecting the end points with abscissae c, d. In our case, f is strictly concave on $[0, 2\pi)$, as its derivative is $S(e^{j\omega})$ which is strictly decreasing. For chords of concave functions, increasing either or both of the abscissae of their endpoints can never cause the slope to increase. As a result, for both the brickwall FBs FB_A and FB_B, the subband variance is nonincreasing as the corresponding filter band–edges increase from 0 to 2π. Thus the largest and next–largest subband variances of FB_A are σ_0^2 and σ_1^2, corresponding to decimators n_0 and n_1 respectively; while the largest subband variance of FB_B is $\hat{\sigma}_0^2$ which corresponds to decimator n_1. All these variances are identifiable as slopes of chords on the graph $y = f(\omega)$ as in Fig. 5.9. Now we define $L = \mathrm{lcm}\{n_i\}$ and $p_i = L/n_i$, and let $\mathbf{v}_A, \mathbf{v}_B$ be the equivalent uniform subband variance vectors of FB_A, FB_B respectively. We use the relation (5.4) between the subband variances of a nonuniform FB and its equivalent uniform FB. This shows that the p_1 largest variances in $\mathbf{v}_B$ have value $\hat{\sigma}_0^2$. On the other hand, of the p_1 largest variances in $\mathbf{v}_A$, there are p_0 variances of value σ_0^2 and $p_1 - p_0$ variances of value σ_1^2. (Recall that $n_0 > n_1$, so $p_1 - p_0 > 0$.) Now referring to the definition (2.7) of majorization, we will have proved that the $\mathbf{v}_A$ does not majorize $\mathbf{v}_B$ if we show that

$$p_0\sigma_0^2 + (p_1 - p_0)\sigma_1^2 < p_1\hat{\sigma}_0^2$$

To do this, we substitute the chord slopes from Fig. 5.9 for the variances. Using $p_k = L/n_k$ and deleting a factor of $L/(2\pi)$, we see that proving the above equation reduces to proving that

$$\frac{n_1}{n_0}f(r) + (1 - \frac{n_1}{n_0})f(t) < f(s), \quad \text{where } r, s, t \text{ are as in Fig. 5.9.} \tag{5.8}$$

This is true by strict concavity of f, since $0 < \frac{n_1}{n_0} < 1$ and $s = \frac{n_1}{n_0}r + (1 - \frac{n_1}{n_0})t$.

Generalizing the proof. If the input spectrum is strictly increasing rather than decreasing on $[0, 2\pi)$ then very similar arguments hold ($f(\omega)$ is now convex). Alternatively we may observe that reflecting the frequency band $\omega \in [0, 2\pi)$ about $\omega = \pi$ gives a decreasing spectrum, and so a separate proof is really unnecessary. Similar comments hold if the interval of strict monotonicity is $[a, a + 2\pi)$ for some $a \neq 0$. If after arranging the decimators as $n_0 \geq n_1 \geq \ldots \geq n_{M-1}$ it happens that $n_0 = n_1$, then we find the smallest i for which $n_i \neq n_{i+1}$, and have n_i, n_{i+1} play the roles played by n_0, n_1 respectively in the above proof. (There is such an i because the FBs are not uniform.)

The result does not hold if the spectrum $S(e^{j\omega})$ is simply monotone as opposed to *strictly* monotone, as shown by the white spectrum or spectrum A of Fig. 5.6. However, examining exactly why the above proof fails for such spectra will reveal that the strictness requirement on the monotonicity can be considerably weakened. In fact it suffices that the spectrum be monotone and *non–constant* over certain intervals that depend on the n_i. The strictness of the monotonicity of $S(e^{j\omega})$, or equivalently, of the concavity of $f(\omega)$ of (5.7), was used in exactly two places: (a) in Step 1, to claim *uniqueness* of FB_A (of Fig. 5.8) as the output of the sequential compaction algorithm, and (b) in Step 2, to prove (5.8). Now (5.8) does not require the concavity of f to be strict on the entire interval $[r, t]$; it suffices that f not be linear, i.e., $S(e^{j\omega})$ not be constant, on this interval. Similarly, the uniqueness of FB_A is not necessary; it suffices that the sequential algorithm necessarily yields a FB containing the 'leftmost' filter of FB_A (corresponding to decimator n_0, see Fig. 5.8). Examining the slopes in Fig. 5.9 shows that this will happen if f is not linear, i.e., $S(e^{j\omega})$ is not constant, on $[0, \frac{2\pi}{n_1}]$. Thus, for the above proof to be valid, $S(e^{j\omega})$ does not have to be *strictly* decreasing on $[0, 2\pi)$; it suffices for it to be decreasing and not constant on either of the intervals $[0, \frac{2\pi}{n_1}]$ and $[\frac{2\pi}{n_0}, \frac{2\pi}{n_0} + \frac{2\pi}{n_1}]$.

The case of real spectra. For a real random process, the spectrum $S(e^{j\omega})$ is even in ω, and hence cannot be monotone on $[a, a + 2\pi)$ (for any a), thus preventing applicability of Theorem 5.2. If such a spectrum is strictly monotone on $[0, \pi)$, can it have a PCFB for the unconstrained orthonormal FB class $\mathcal{N}^u$? We could expect that it cannot (generalizing Theorem 5.2), however, we currently have no proof. The proof of Theorem 5.2 used contiguous–stacked brickwall FBs, which always belong to $\mathcal{N}^u$. A direct generalization to real process spectra would try to use *real–coefficient* brickwall FBs, where all filters $H_i(e^{j\omega})$ are also even in ω. Such FBs however lie in $\mathcal{N}^u$ only if their decimators obey the conditions of the bandpass sampling theorem, as discussed in detail in Section 6.3.2.

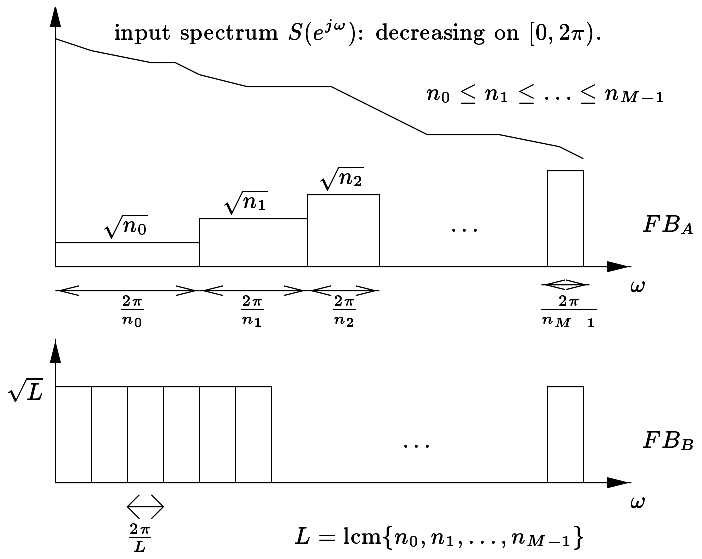


Figure 5.10: Unconstrained PCFB defined by normalized subband variances, for monotone spectra.

5.3.3 A result for PCFBs defined by normalized variances

Theorem 5.3. Let $\mathcal{N}^u$ be the class of unconstrained nonuniform orthonormal FBs with a given set of decimators n_i. A PCFB for $\mathcal{N}^u$ defined using the normalized subband variance vector always exists if the input spectrum $S(e^{j\omega})$ is monotonic (increasing or decreasing) on $[a, a + 2\pi)$ for some real a. Figure 5.10 shows a PCFB, namely FB_A, the contiguous stacked brickwall FB with a as band–edge and filter bandwidths decreasing monotonically with the spectrum. ■

Discussion: As with Theorem 5.2, it suffices to prove the result assuming that $S(e^{j\omega})$ is monotone decreasing on $[0, 2\pi)$, as in Fig. 5.10. The result is well known for the special case of uniform FBs (i.e., when all n_i are equal). In other words, with L as the lcm of the n_i, the filter bank FB_B of Fig. 5.10 is the PCFB for the class $\mathcal{C}^u$ of unconstrained L channel uniform orthonormal FBs. The result for nonuniform FBs essentially follows from this and the notion of the equivalent uniform FB. Notice that the result is strikingly different from that of the earlier section: PCFBs defined by the equivalent uniform FBs do not exist for strictly monotone spectra (Theorem 5.2).

Proof of Theorem 5.3: Consider an arbitrary FB in $\mathcal{N}^u$, with normalized subband variances $\frac{\sigma_i^2}{n_i}$. We must show that the sum $S(P)$ of the P largest of these variances does not exceed the corresponding sum for FB_A of Fig. 5.10. Let $p_i = L/n_i$. Now, the normalized variance $\frac{\sigma_i^2}{n_i}$ multiplied by the constant scale L equals $p_i\sigma_i^2$ which is the sum of the variances in the p_i corresponding subbands of the equivalent uniform FB (all these subbands have the same variance σ_i^2). Let the P largest normalized variances, which sum to $S(P)$, occur in subbands with indices in the (P–element) set I_P. Thus, $S(P) \cdot L$ equals the sum of $\sum_{i \in I_P} p_i$ subband variances in the equivalent L channel uniform

FB. This sum cannot exceed the sum of the largest $\sum_{i\in I_P} p_i$ subband variances in FB_B, because FB_B is a PCFB for the unconstrained uniform FB class $\mathcal{C}^u$. This bound still depends on the chosen FB in $\mathcal{N}^u$ via the set I_P. However we can obtain a FB–independent bound by replacing $\sum_{i\in I_P} p_i$ with an upper bound on it, namely the sum of the P largest of the p_i. The proof will then be completed if we show that FB_A from $\mathcal{N}^u$ actually attains this bound. In other words, assuming that $p_0 \geq p_1 \geq \ldots \geq p_{M-1}$ as in Fig. 5.10, i.e., that $n_i \leq n_{i+1}$, we must show the following claim:

Claim: The P largest normalized subband variances $\frac{\sigma_i^2}{n_i}$ for FB_A occur in its channels corresponding to decimators $n_0, n_1, \ldots, n_{P-1}$, and have a sum that multiplied by L equals the sum of the $\sum_{i=0}^{P-1} p_i$ largest subband variances of FB_B.

To prove the claim, note that the i-th filter of FB_A is ideal brickwall with a support that equals the union of the supports of p_i filters in the brickwall filter bank FB_B. From this, it is easily verified that the sum of the subband variances corresponding to these p_i filters equals L times the normalized variance $\frac{\sigma_i^2}{n_i}$ of FB_A. The claim is then proved by noticing the order in which the filter supports in FB_A are created by merging those in FB_B: For $i = 0$, we merge the p_0 filters corresponding to the p_0 largest subband variances of FB_B. Then for $i = 1$ we merge the p_1 filters corresponding to the next largest variances in FB_B, and so on. $\bigtriangledown\bigtriangledown\bigtriangledown$

The above proof allows creation of more general (non-monotone) spectra $S(e^{j\omega})$ for which PCFBs (defined by normalized variances) for the class $\mathcal{N}^u$ exist. We replace FB_B by the brickwall PCFB for $S(e^{j\omega})$ for the uniform FB class $\mathcal{C}^u$. We then merge its filter supports in the manner described in proving the above claim, to create a nonuniform brickwall FB with the given decimators n_i. Thus, this FB also obeys the above claim (for the new choice of FB_B), and hence, by the same reasoning, is a PCFB for $\mathcal{N}^u$, *provided it lies in* $\mathcal{N}^u$. However, the process of merging filter supports need not yield a valid FB in $\mathcal{N}^u$. This is because in a brickwall FB in $\mathcal{N}^u$, a filter corresponding to decimator n_i must have an aliasfree(n_i) support, which merging p_i disjoint aliasfree(L) supports does not always yield. Thus the method does not prove PCFB existence for *all* input spectra.

5.4 Conclusion

We have presented two different definitions of nonuniform PCFBs, generalizing the uniform ones. We have studied their optimality, showing that like uniform PCFBs, they minimize many concave objectives although of a somewhat more restricted form. The two PCFB definitions yield strikingly different existence and optimality results. The definition using equivalent uniform FBs was shown to have a broader optimality. With this definition, if the input spectrum is strictly monotonic on $[a, a+2\pi)$, there are no nonuniform PCFBs for the class of unconstrained orthonormal FBs with any given set of decimators (not all equal). Thus, existence of nonuniform PCFBs is much more delicate than that of uniform PCFBs, and can be destroyed by small perturbations of the input spectra.

Chapter 6 Nonuniform FBs: Parameterizations

In Chapter 5 we studied the statistical optimization of nonuniform orthonormal FBs, which are special cases of (nonuniform) FBs with the PR property. An even more fundamental problem is that of parameterizing nonuniform PRFBs, i.e., finding all possible PRFBs with a given set of decimation ratios (obeying maximal decimation). This problem has been extensively studied for uniform FBs (i.e., when the decimators are all equal), and the results can be used to generate a rich family of nonuniform PRFBs too, by cascading uniform ones in tree structures. However, not *all* nonuniform PRFBs can be built in this manner. The parameterization problem in its full generality remains unsolved for several classes of nonuniform FBs, including very basic ones such as the classes of all rational or FIR FBs. In this chapter, we digress from FB optimization (the unifying theme of the earlier chapters) to study this fascinating problem, which is all the more interesting being a basic question that has remained unanswered even after over two decades of filter bank research.

6.1 Introduction

Figure 1.8 shows an M channel nonuniform filter bank (FB). The FB is said to be maximally decimated if the channel decimation rates n_i are integers satisfying

$$\sum_{i=0}^{M-1} \frac{1}{n_i} = 1. \qquad \text{(maximal decimation condition)} \qquad (6.1)$$

Figure 6.1a shows a maximally decimated *uniform* FB, which is a special case of Fig. 1.8 where $n_i = M$ for all i. For this case, as explained in Section 1.1.2, the system can be equivalently redrawn using the analysis and synthesis polyphase matrices $\mathbf{E}(z)$ and $\mathbf{R}(z)$, as shown in Fig. 6.1b. The condition for perfect reconstruction (PR) is then easily expressed as $\mathbf{R}(z) = \mathbf{E}^{-1}(z)$. Due to this, the theory and design of uniform PRFBs is an extremely well developed subject. Numerous parameterization results list all possible M channel uniform PRFBs with various sets of properties such as paraunitariness, FIR filters, linear phase filters, etc.

In contrast, several issues involved in achieving PR in nonuniform FBs remain unresolved. For example, given a general set of positive integers n_i obeying maximal decimation (6.1), how do we determine whether or not there exists a rational PRFB (i.e., one with rational filters) using the n_i as decimators? If the n_i are all equal, clearly such a FB exists (as it is then uniform). Similarly, it also exists if the n_i are arrangeable in a tree so that such a PRFB can be built by cascading uniform PRFBs in a tree structure (Section 6.4.1). This is the most common approach to achieving

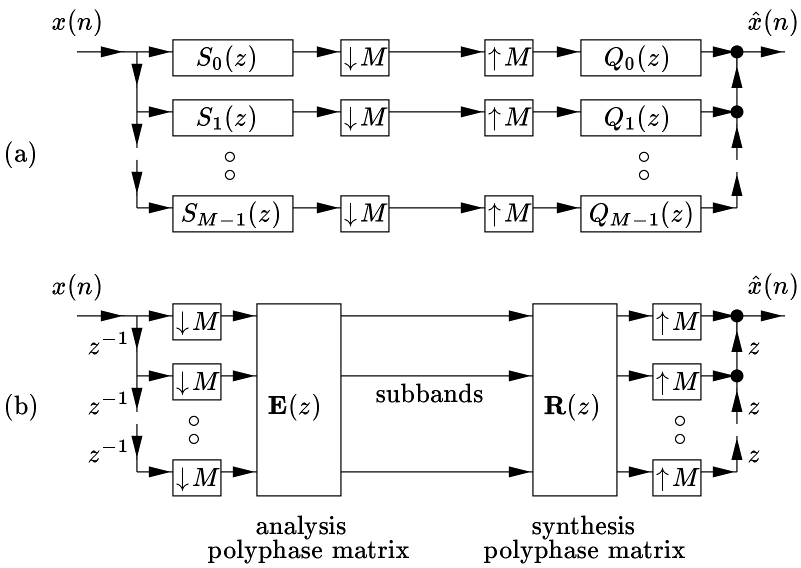


Figure 6.1: Uniform maximally decimated filter bank. (a) Showing analysis and synthesis filters. (b) Polyphase representation.

PR in nonuniform FBs. In particular, it is used to build the FBs that implement the dyadic wavelet transforms [64],[67]: Such a FB has a *dyadic* decimator–set, i.e., one of form $\{2, 2^2, \ldots, 2^{r-1}, 2^r, 2^r\}$ for some integer $r \geq 1$, and is built using a dyadic tree (i.e., one built from a cascade of r two channel FBs). However, there are sets of decimators n_i that cannot be arranged in a tree as described above, and yet permit existence of rational PRFBs in which in fact all filters are *delays*. Further, even if the decimators are arrangeable in a tree, it is possible that there are PRFBs using those decimators that cannot be realized using the tree. These facts will be discussed in detail with examples in Section 6.4.2. Thus *tree structures of uniform PRFBs are far from being a full solution* to the PR problem for nonuniform FBs.

Derivability of decimators from a tree (as described above) is a sufficient condition for existence of rational PRFBs using the decimators. There are certain other conditions that are known to be necessary, e.g., there are no rational PRFBs using the decimator–set $\{2, 3, 6\}$ because no two decimators of such a FB can be coprime (Section 6.6.1). However, a *condition that is both necessary and sufficient remains unknown.* The present chapter studies this and related problems. An important part of our study is to significantly improve upon the known conditions, i.e., to derive new ones, strengthen necessary conditions and weaken sufficient ones. Another contribution is to study the conditions for reducibility of PRFBs to tree structures. For example, it has been shown [19],[55] that all rational PRFBs with dyadic decimator–sets *must* be derivable from dyadic trees. In Section 6.7, we will considerably generalize this result. Although these problems in their full generality remain unresolved, we believe this work to be an important step towards a complete understanding of this subject.

6.1.1 Relevant earlier work

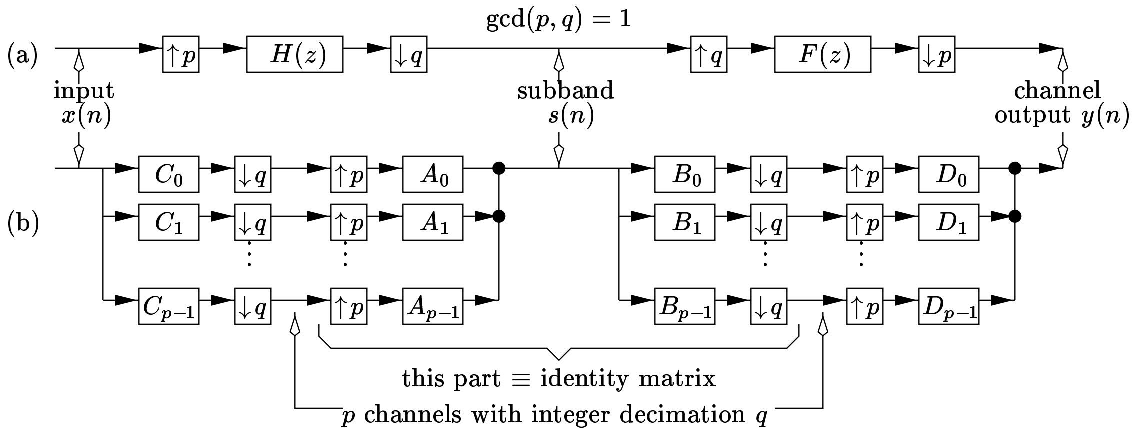
Trees of uniform FBs, and near–PR designs: A very common approach to nonuniform PRFB design is to cascade uniform PRFBs in a tree-structure, e.g., as is done to implement dyadic wavelet transforms [64],[67]. However, as stated earlier, there are nonuniform PRFBs that cannot be built in this manner. Many works deal with *approximate reconstruction* (or 'near–PR') nonuniform FBs, e.g., the frequency domain approaches of Li et al. [40], the time domain methods of Nayebi et al. [47], and other references therein. These are very useful from a practical standpoint, giving FBs with excellent filter responses and low aliasing distortions. However, they do not address the many theoretical issues involved in obtaining *exact* reconstruction.

FBs with fractional decimators: Kovacevic and Vetterli have studied a more general system [39] where each channel of the FB has a decimation rate that is fractional, i.e., of form q/p where p, q are coprime positive integers. Such a channel, shown in Fig. 6.2a, is completely equivalent to the system of Fig. 6.2b. By this we mean that given any one of these systems, we can choose the filters in the other so that the same input $x(n)$ for both systems always produces the same signals $s(n)$ and $y(n)$ as shown. A choice ensuring this is shown in Fig. 6.2c (polyphase components are defined in Section 1.1.2). The equivalence under this choice is provable using the discussion on fractional decimation in [64, Section 4.3.3]. If the $A_i(z)$ differ from the special choice of Fig. 6.2c, we can replace them by this choice and modify the $C_i(z)$ so that the signal $s(n)$ is unaffected. This is done by performing a p-th order polyphase decomposition of the $A_i(z)$, using the fact that p, v are coprime, and moving the resulting polyphase matrix to the left. A similar comment holds for the $B_i(z)$.

From the equivalence shown in Fig. 6.2, we conclude that the PR problems for integer–decimated and rationally decimated FBs are fully equivalent. Another concern besides PR in rationally decimated FBs is the nature of their spectral analysis: Does a subband represent a contiguous portion of the input spectrum, or do the decimators and expanders in Fig. 6.2b cause it to contain separate parts, possibly mirrored and shuffled in order? This issue is studied in [39].[1] However, as far as the PR problem is concerned, it is enough to study FBs with integer decimators, and that is the approach we shall use.

Other more general multirate structures: As we will see in Section 6.2.2, nonuniform PRFBs are hard to design because of certain structural constraints that their associated polyphase matrices must obey. This is the origin of the central problem studied in our work: These structural constraints cannot be obeyed by rational FBs unless their decimators satisfy various conditions, which we aim to characterize. However, the constraints vanish if we use more general systems in the channels of the FB, e.g., if the filters are allowed to be periodically time–varying (Section 6.2.3). Chen and Qiu [15]

[1] It becomes less serious if we allow modulators at appropriate points within the FB.



(c)

$C_i(z) = z^{iu}R_i(z)$ $\quad$ $A_i(z) = z^{-iv}$, $B_i(z) = z^{iv}$ $\quad$ $D_i(z) = z^{-iu}E_i(z)$

$(R_0(z), \ldots, R_{p-1})^T$: p-th order synthesis polyphase vector of $H(z)$ $\quad$ $up - vq = 1$ $\quad$ $(E_0(z), \ldots, E_{p-1}(z))$: p-th order analysis polyphase vector of $F(z)$

Figure 6.2: FB with rational decimators. (a) Single channel with decimator q/p. (b) Equivalent system of p channels with decimator q. (c) A possible set of filter choices ensuring the equivalence.

and Shenoy [54] have studied multirate and FB design using such more general structures. The PR design then allows as much or even more freedom than that in the well–studied PR designs for the traditional *uniform* FB of Fig. 6.1. Our work is restricted to the usual nonuniform FB structure of Fig. 1.8 that does not use such generalized multirate structures.

PR conditions on decimators, and reducibility to tree structures: A necessary condition on the (integer) decimators for PR with rational FBs was first stated in [28]. Called the compatibility test, it was generalized by Djokovic and Vaidyanathan [21], who also pointed out another such condition (pairwise noncoprimeness). We will considerably generalize these conditions. Another related work has involved showing derivability of FBs using dyadic decimator–sets from dyadic trees [55],[19], as explained earlier. These results too will be significantly strengthened. Among various more general situations studied include certain non–dyadic sets, unconstrained FBs, and tree structures whose constituent FBs need not be uniform.

6.1.2 Chapter outline

Section 6.2 reviews the PR conditions on the filters of uniform FBs, and their generalization to nonuniform FBs, derivable using a transformation of nonuniform FBs to equivalent uniform ones. It shows how in spite of this transform, the nonuniform PRFB design does not reduce to a uniform PR design, unless the filters of the nonuniform FB are allowed to be *time varying.* In Section 6.3 we formally state the central problem, and study its solution for classes of unconstrained FBs (where the filters of the FB have no constraints such as rationality). Section 6.4 analyzes the role of tree structures in the study of the main problem. It shows how tree structures of uniform PRFBs do not provide a full solution (Section 6.4.2), and how trees can be used to improve upon known PR conditions on the decimators (Section 6.4.3). Section 6.5 solves the central problem of the chapter for the class of *delay–chains* (FBs in which all filters are delays): It states the necessary and sufficient condition for a set of decimators to be usable to build a PR delay–chain, and presents algorithms to test the condition. Subsequent sections focus mainly on the class of rational FBs. Section 6.6 states the earlier known necessary conditions on decimators of rational PRFBs, and generalizes them in several ways. Section 6.7 generalizes [55],[19] by finding weaker conditions on decimators under which all PRFBs using them can be derived from certain tree structures. Section 6.8 summarizes all known necessary PR conditions on the decimators, and studies their inter-relationships. We conclude by noting many open problems in the area.

6.1.3 Preliminaries

This chapter extensively uses various polyphase concepts [64] reviewed in Section 1.1.2. An easily proved result that we often use is the following:

Lemma 6.1: Polyphase lemma. Let $\mathbf{e}(z), \mathbf{r}(z)$ be the M-th order analysis and synthesis polyphase matrices of the filters $H(z)$ and $F(z)$ respectively. Thus $\mathbf{e}(z)$ is a row vector and $\mathbf{r}(z)$ is a column vector. Then,

$$\mathbf{e}(z)\mathbf{r}(z) = (H(z)F(z)) \downarrow_M \tag{6.2}$$

Maximal decimation: **All FBs** studied in this chapter are **maximally decimated** with **integer decimation** rates, even if this is not explicitly stated. Similarly, references to a 'set of decimators' (or 'decimator–set') always implicitly mean a set of positive integers (not necessarily distinct) obeying (6.1).

6.2 Background: Equivalent uniform FBs and PR equations

The main focus of the chapter is to find conditions on the decimators that permit existence of various types of nonuniform perfect reconstruction (PR) FBs with those decimators. To do this, we must first know what conditions on the filters of the FB guarantee the PR property. This section begins by reviewing the PR conditions for uniform FBs. We then use the transformation of a nonuniform FB to an equivalent uniform FB, already encountered in Section 5.1. This yields the general PR conditions for nonuniform FBs, that will be used in all the later sections. In spite of the possible transformation to uniform FBs, the nonuniform PRFB design problem by no means reduces to the uniform PR design. However, such a reduction *does* occur if the nonuniform FB is allowed to have filters that are LPTV(L) (linear periodically time varying with period L) instead of LTI. With LTI filters, achieving PR is tougher, and is the subject of the later sections.

6.2.1 PR for uniform FBs, and the nonuniform to uniform transformation

For the uniform FB of Fig. 6.1, the problem of achieving PR is very well understood. The following are three equivalent necessary and sufficient conditions on the filters for PR in this case [64]:

1. *Biorthogonality condition.* $(S_i(z)Q_j(z))\downarrow_M = \delta(i-j)$.

2. *AC matrix formulation.* Let $W = e^{-j2\pi/M}$. Then,

$$\begin{bmatrix} A_0(z) \\ A_1(z) \\ \vdots \\ A_{M-1}(z) \end{bmatrix} \triangleq \underbrace{\begin{bmatrix} S_0(z) & S_1(z) & \dots & S_{M-1}(z) \\ S_0(zW) & S_1(zW) & \dots & S_{M-1}(zW) \\ \vdots & \vdots & \ddots & \vdots \\ S_0(zW^{M-1}) & S_1(zW^{M-1}) & \dots & S_{M-1}(zW^{M-1}) \end{bmatrix}}_{\text{alias cancellation (AC) matrix } \mathbf{S}(z)} \begin{bmatrix} Q_0(z) \\ Q_1(z) \\ \vdots \\ Q_{M-1}(z) \end{bmatrix} = \begin{bmatrix} M \\ 0 \\ \vdots \\ 0 \end{bmatrix} \tag{6.3}$$

   For any uniform FB (PR or otherwise), the $A_i(z)$ defined above are called the 'aliasing gains'. The PR condition (6.3) thus specifies all aliasing gains. It arises from the frequency domain relation between the output $\widehat{X}(z)$ and input $X(z)$ of *any* uniform FB (PR or otherwise):

$$\widehat{X}(z) = \frac{1}{M}\sum_{i=0}^{M-1} A_i(z)X(zW^i) \tag{6.4}$$

3. *Polyphase formulation.* If $\mathbf{E}(z), \mathbf{R}(z)$ are respectively the M-th order analysis and synthesis polyphase matrices of the FB (as in Fig. 6.1b), then $\mathbf{R}(z) = \mathbf{E}^{-1}(z)$. That this is equivalent to the biorthogonality condition stated earlier follows from the polyphase lemma (Section 6.1.3), which shows that the ij-th entry of $\mathbf{E}(z)\mathbf{R}(z)$ is precisely the quantity $(S_i(z)Q_j(z))\downarrow_M$ occurring in the biorthogonality condition.

Now, as seen in Section 5.1, any nonuniform FB (as in Fig. 1.8) is transformable into a uniform FB, which we call its equivalent uniform FB [3],[21],[28],[39]. This transform is described by Fig. 5.2, which shows how a single channel with decimator n_k is replaceable by p_k channels with decimators $L = n_k p_k$. Repeating this process on all channels of the nonuniform FB, with L as any common multiple of all its decimators n_i (usually $L = \text{lcm}\{n_i\}$), yields a *uniform* L channel FB. The nonuniform FB has PR if and only if the equivalent uniform FB has PR. The filters in the uniform FB are various delayed versions of those in the nonuniform one. Inserting these relations between the filters into the PR conditions for uniform FBs gives the PR conditions for nonuniform FBs. These conditions, described next, generalize the uniform FB PR conditions, and are heavily used later.

6.2.2 The general PR conditions for nonuniform FBs

Biorthogonality condition. The uniform FB biorthogonality condition, when applied to the uniform FB derived from the nonuniform one of Fig. 1.8, is equivalent to

$$(H_i(z)F_j(z)) \downarrow_{\gcd(n_i,n_j)} = \delta(i-j) \qquad \text{(biorthogonality condition)} \tag{6.5}$$

This has been observed earlier [21],[55]. Appendix A contains a proof for easy reference. The condition gets its name from its time-domain equivalent. To describe this, let $h_i(n), f_i(n)$ respectively be the impulse responses of $H_i(z), F_i(z)$. We define two sets of sequences

$$\{\mu_{ik}(n) = h_i^*(kn_i - n) \ \mid i = 0, 1, \ldots, M-1, \ k = \text{any integer}\} \tag{6.6}$$

$$\{\eta_{jl}(n) = f_j(n - ln_j) \ \mid j = 0, 1, \ldots, M-1, \ l = \text{any integer}\} \tag{6.7}$$

The action of the FB is now elegantly expressed using these sequences: If $x(n)$ is the FB input sequence,

$$c_j(l) = \left\langle x(n), \mu_{jl}(n) \right\rangle = \sum_{n=-\infty}^{\infty} x(n) h_j(ln_j - n) \qquad (j\text{-th subband signal}) \tag{6.8}$$

$$\widehat{x}(n) = \sum_{j=0}^{M-1} \sum_{l=-\infty}^{\infty} c_j(l) \eta_{jl}(n) = \sum_{j=0}^{M-1} \sum_{l=-\infty}^{\infty} c_j(l) f_j(n - ln_j) \qquad \text{(FB output)} \tag{6.9}$$

Here $\langle a(n), b(n) \rangle = \sum_n a(n) b^*(n)$ is the inner product of the sequences $a(n)$ and $b(n)$ (in the space of all sequences $x(n)$ for which $\sum_n |x(n)|^2$ is finite). Thus, the FB output $\widehat{x}(n)$ is a linear combination of the sequences from (6.7), using weights $c_j(l)$ that are inner products of the input $x(n)$ with the sequences from (6.6). Thus PR (i.e., $\widehat{x}(n) = x(n)$) is achieved if the two sets (6.6),(6.7) form a biorthogonal system, i.e., if

$$\langle \mu_{ik}(n), \eta_{jl}(n) \rangle = \delta(i-j)\delta(k-l) \tag{6.10}$$

This can indeed be shown to be the 'time domain' equivalent of (6.5).

AC matrix formulation [21]. In (6.3), we set $M = L, W = e^{-j2\pi/L}$, and the filters as those of the uniform FB derived as in Fig. 5.2, from the nonuniform FB of Fig. 1.8. The i-th row in (6.3) is a sum of filter product terms $S_j(zW^i)Q_j(z)$. We group terms arising from the k-th subband in Fig. 1.8, i.e., those with $S_j(z) = z^{-ln_k} H_k(z)$ and $Q_j(z) = z^{ln_k} F_k(z)$ for $l = 0, 1, \ldots, p_k - 1$ where $n_k p_k = L$ (see Fig. 5.2). This yields a sum of form

$$H_k(zW^i)F_k(z)A_{ik}, \quad \text{where } A_{ik} = \sum_{l=0}^{p_k-1} W^{-iln_k} = \sum_{l=0}^{p_k-1} e^{-j2\pi il/p_k} = \begin{cases} p_k & \text{if } i \text{ is a multiple of } p_k \\ 0 & \text{otherwise} \end{cases}$$

Thus, we can rewrite (6.3) using a new L–row AC matrix $\mathbf{H}(z)$ that has only M columns (one for each analysis filter of the nonuniform FB), as follows:

$$\underbrace{\begin{bmatrix} \mathbf{h}_0(z) & \mathbf{h}_1(z) & \ldots & \mathbf{h}_{M-1}(z) \end{bmatrix}}_{\text{AC matrix } \mathbf{H}(z)} \begin{bmatrix} F_0(z) & F_1(z) & \ldots & F_{M-1}(z) \end{bmatrix}^T = \begin{bmatrix} L & 0 & \ldots & 0 \end{bmatrix}^T, \text{ where} \tag{6.11}$$

$$\mathbf{h}_i(z) = p_i \begin{bmatrix} H_i(z) \underbrace{0 \ \ldots \ 0}_{p_i - 1 \text{ zeros}} & H_i(zW^{p_i}) \underbrace{0 \ \ldots \ 0}_{p_i - 1 \text{ zeros}} & H_i(zW^{2p_i}) & \ldots & H(zW^{(n_i-1)p_i}) \underbrace{0 \ \ldots \ 0}_{p_i - 1 \text{ zeros}} \end{bmatrix}^T \tag{6.12}$$

If $n_i = M$ and $p_i = 1$ for all i (i.e., if the FB is uniform), the form of $\mathbf{H}(z)$ indeed reduces to that of (6.3).

Polyphase formulation. The PR condition is $\mathbf{R}(z) = \mathbf{E}^{-1}(z)$, just as for uniform FBs. However, as the equivalent uniform FB has interdependencies between the filters, its analysis polyphase matrix $\mathbf{E}(z)$ has a special structure [3]: Its rows can be partitioned into groups, where the k-th group corresponds to the k-th subband analysis filter $H_k(z)$ in Fig. 1.8. This group has $p_k = L/n_k$ rows as follows:

$$\left[\begin{array}{ccc|ccc|ccc} E_0^k & \cdots & E_{n_k-1}^k & E_{n_k}^k & \cdots & E_{(p_k-1)n_k-1}^k & E_{(p_k-1)n_k}^k & \cdots & E_{L-1}^k \\ z^{-1}E_{(p_k-1)n_k}^k & \cdots & z^{-1}E_{L-1}^k & E_0^k & \cdots & E_{(p_k-2)n_k-1}^k & E_{(p_k-2)n_k}^k & \cdots & E_{(p_k-1)n_k-1}^k \\ z^{-1}E_{(p_k-2)n_k}^k & \cdots & z^{-1}E_{(p_k-1)n_k-1}^k & z^{-1}E_{(p_k-1)n_k}^k & \cdots & E_{(p_k-3)n_k-1}^k & E_{(p_k-3)n_k}^k & \cdots & E_{(p_k-2)n_k-1}^k \\ \vdots & \cdots & \vdots & \vdots & \cdots & \vdots & \vdots & \cdots & \vdots \\ z^{-1}E_{n_k}^k & \cdots & z^{-1}E_{2n_k-1}^k & z^{-1}E_{2n_k}^k & \cdots & z^{-1}E_{L-1}^k & E_0^k & \cdots & E_{n_k-1}^k \end{array}\right] \tag{6.13}$$

The first row is the L-th order analysis polyphase matrix (vector) of $H_k(z)$. Each subsequent row is formed by shifting length - n_k blocks of the previous row to the right, with the last block multiplied by z^{-1} and circulated back to the left end.[2] These rows are the polyphase vectors of filters $z^{-an_k}H_k(z)$ for $a = 1, 2, \ldots, p_k - 1$. Similarly, the synthesis polyphase matrix $\mathbf{R}(z)$ of the equivalent uniform FB has *columns* arrangeable into groups. The k-th group has a form like the transpose of (6.13), with the $E_l^k(z)$ replaced by the entries $R_l^k(z)$ of the L-th order synthesis polyphase vector of the synthesis filter $F_k(z)$, and the z^{-1} factors replaced by z elements.

The paraunitary case. The uniform FB of Fig. 6.1 is said to be paraunitary (or orthonormal) if $\mathbf{E}^{-1}(z) = \widetilde{\mathbf{E}}(z)$; or in other words, if PR is obtained with $\mathbf{R}(z) = \widetilde{\mathbf{E}}(z)$, or equivalently with $Q_i(z) = \widetilde{S}_i(z)$. By generalization, the nonuniform FB of Fig. 1.8 is said to be orthonormal if PR is obtained (i.e., (6.5) is obeyed) with $F_i(z) = \widetilde{H_i}(z)$. From the relations between the filters of the

[2] The submatrix (6.13) of $\mathbf{E}(z)$ is *block pseudocirculant* with block size $1 \times n_k$ (generalizing the notion of pseudo-circulants [64]).

nonuniform and the equivalent uniform FB, we see that each of these is paraunitary if and only if the other is. Notice that the two sets of (6.6),(6.7) which form a biorthogonal system in any PRFB, will *coincide*, hence forming an *orthonormal* system, if and only if the FB is paraunitary. This is because $F_i(z) = \widetilde{H}_i(z)$ is equivalent to $\eta_{jl}(n) = \mu_{jl}(n)$ in (6.6),(6.7). A general PRFB that is not necessarily orthonormal is often called a biorthogonal FB, due to the condition (6.5). Two additional properties of orthonormal FBs, proved for uniform FBs in [64], are as follows:

$$\frac{1}{2\pi}\int_0^{2\pi} \left|H_i(e^{j\omega})\right|^2 d\omega = 1 \ \ \text{(unit energy)}, \ \ \text{and} \ \ \sum_{i=0}^{M-1} \frac{H_i(z)\widetilde{H}_i(z)}{n_i} = 1 \ \ \text{(power complementarity)}. \tag{6.14}$$

We can prove these for nonuniform FBs using the result for uniform ones and the transformation of Fig. 5.2.

6.2.3 Relation between the nonuniform and uniform PR designs

Transforming a nonuniform FB to an equivalent uniform one helps to find the PR conditions on its filters. These two FBs also share several properties (i.e., each has the property iff the other does). Examples are PR and paraunitariness; and rationality, stability, and FIR nature of filters. However, the equivalent uniform FB does not help in designing nonuniform PRFBs. This is due to its special structure: It has groups of filters that are delayed versions of each other. There are no known uniform PRFB design methods that allow imposition of this structure. Notice that the delayed versions of a filter have the *same* magnitude response, while uniform PRFB designs usually approximate ideal *nonoverlapping* analysis filter responses.

Most choices of the analysis filters $H_i(z)$ of Fig. 1.8 yield an equivalent uniform FB with an *invertible* analysis polyphase matrix $\mathbf{E}(z)$. However, this is not sufficient for existence of LTI synthesis filters ($F_i(z)$ of Fig. 1.8) resulting in PR: For this we further require that the inverse $\mathbf{R}(z) = \mathbf{E}^{-1}(z)$ have the special structure described in Section 6.2.2. This added constraint is not always easy to satisfy. If $\mathbf{E}(z)$ is paraunitary, then $\mathbf{R}(z)$, being equal to $\widetilde{\mathbf{E}}(z)$, automatically has the desired structure, and a nonuniform (paraunitary) PRFB is possible. However, again none of the many known parameterizations of uniform paraunitary FBs [64] allow imposition of the special structure of Section 6.2.2 that $\mathbf{E}(z)$ must have in order to represent a nonuniform FB.

The structural constraints on $\mathbf{E}(z)$ and $\mathbf{R}(z)$ can however be completely given up if the filters in the nonuniform FB are allowed to be LPTV(L) instead of LTI [3]. This is shown by Fig. 6.3, in which $p_k = L/n_k$ channels of a uniform L channel (maximally decimated) FB are converted into a single channel with decimator n_k. The analysis and synthesis filters in this channel are LPTV(L). The procedure is repeated for each k using disjoint subsets of channels of the uniform FB. Clearly the nonuniform FB has the PR property if and only if the uniform one does. In the rest of this

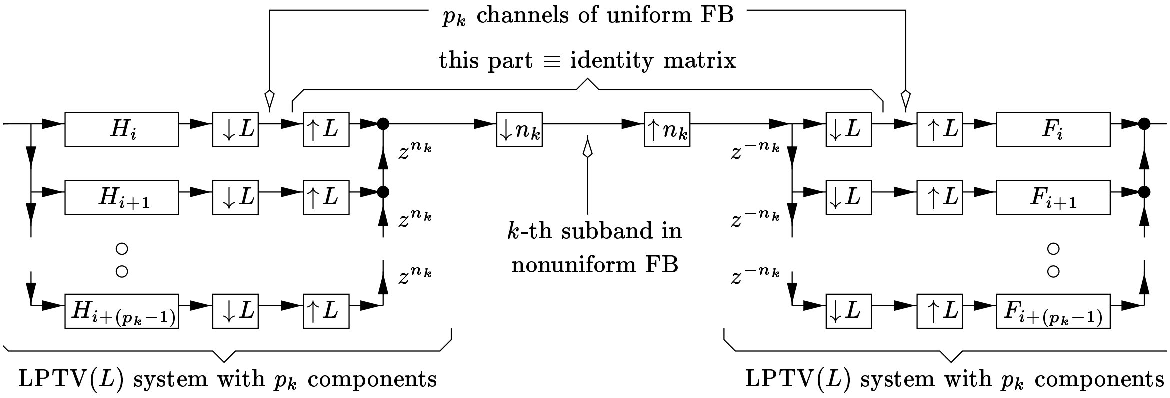


Figure 6.3: Equivalence between uniform FBs and nonuniform FBs with LPTV filters.

chapter, we assume all analysis and synthesis filters of all FBs to be LTI. The nonuniform PR design is then significantly harder.

6.3 Problem statement, and unconstrained FBs

6.3.1 Problem statement

The nonuniform perfect reconstruction (PR) FB design problem in its full generality can be stated as follows:

1. **Conditions on decimators for PR.** Given a set of positive integers n_i satisfying the maximal decimation condition (6.1), find necessary and sufficient conditions on the n_i for existence of a PRFB in some specified class $\mathcal{C}$ of FBs, having the n_i as decimators.

2. **Parameterization of the PRFBs.** When the n_i satisfy such a condition, find *all* possible PRFBs in $\mathcal{C}$ having n_i as decimators.

The FB class $\mathcal{C}$ here is defined by some constraint on the filters of its constituent FBs. Important examples that we will consider are delay–chains (FBs in which all filters are delays), rational FBs and FIR FBs. Other constraints that the class $\mathcal{C}$ can impose are realness of filter coefficients, stability of filters, and paraunitariness (or orthonormality). Note that in general the class definition does not directly by itself impose any constraint on either the number of channels or the nature of the decimators in the FB. However, the requirement that a FB in the class be maximally decimated and have PR could impose various constraints on these parameters. The statement of the problem is to characterize (a) the nature of these constraints, and (b) all PRFBs in $\mathcal{C}$ having a general decimator–set that obeys these constraints.

The solution to the problem of course depends on the FB class $\mathcal{C}$. It is completely known for delay–chains, but unknown for rational FBs. Notice that the parameterization problem depends on

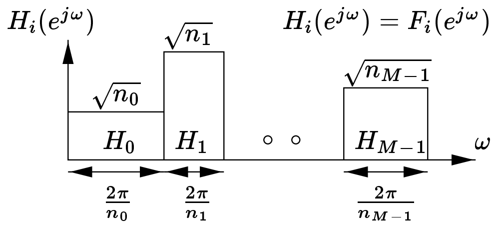


Figure 6.4: Ideal contiguous–stacked complex coefficient brickwall FB.

first finding conditions on the decimators for PR, which can be quite tough in itself. So we will mainly focus on finding conditions for PR. Our aim will be to weaken the sufficient conditions and strengthen necessary ones until we obtain a set of necessary *and* sufficient conditions (the final goal, which we do not always achieve). We will also derive some results on the parameterization problem, especially in connection with tree structures.

6.3.2 FBs with unconstrained complex and real coefficient filters

Let the class $\mathcal{C}$ in the above formulation be simply the class of all FBs, with no filter constraints (i.e., allowing ideal brickwall filters etc.). Then a PRFB in $\mathcal{C}$ always exists, no matter what the decimators n_i are (of course, provided they obey (6.1)). This is because the FB in Fig. 6.4, with ideal contiguous–stacked brickwall filters, always has PR. In fact it is a paraunitary FB. We will hence exclude this class $\mathcal{C}$ from all further discussion.

Note that the filters of Fig. 6.4 always have complex coefficients. Now let $\mathcal{C}$ be the class of all *real coefficient* FBs (i.e., FBs in which all filters have real coefficients). No other constraint is imposed, so the filters could still be ideal. However, it is now more difficult to find conditions on the decimators for existence of PRFBs in $\mathcal{C}$. Taking a cue from Fig. 6.4, we can examine *brickwall* FBs, i.e., FBs as in Fig. 1.8 where the filters $H_i(e^{j\omega})$ have nonoverlapping supports, are constant on their supports and $H_i(z) = F_i(z)$. Since the H_i partition the input spectrum, PR is possible if and only if for each i, the i-th channel perfectly reconstructs all inputs that are bandlimited to the passband of $H_i(e^{j\omega})$. (In fact we then get a paraunitary PRFB, by suitable scaling of the filters.) This equivalently means that $H_i(e^{j\omega})$ has an aliasfree(n_i) support. For the (real coefficient) FB of Fig. 6.5, the bandpass sampling theorem states that this happens iff the band–edges of H_i are at integer multiples of π/n_i [39]. Thus, the FB of Fig. 6.5 has PR if and only if

$$\sum_{i=0}^{k} \frac{1}{n_i} \text{ is an integer multiple of } \frac{1}{n_{k+1}} \text{ for all } k = 0, 1, \ldots, M-2. \tag{6.15}$$

Thus, a given set of decimators n_i can be used to build a real coefficient PRFB of the form of Fig. 6.5 if and only if (6.15) holds for some ordering of the n_i. For example, the set $\{2, 3, 6\}$ obeys

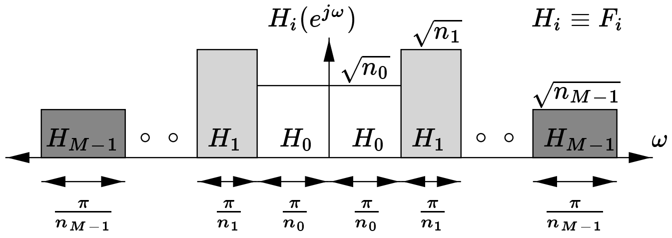


Figure 6.5: Ideal contiguous–stacked real coefficient brickwall FB.

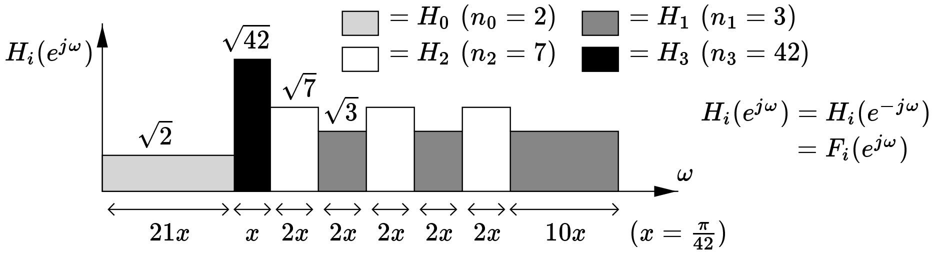


Figure 6.6: Non–contiguous stacked ideal real coefficient brickwall FB.

this condition (with ordering (2,6,3) or (3,6,2)). The set $\{2, 3, 7, 42\}$ violates the condition (it is the only such set with ≤ 4 decimators). However, this does not preclude existence of PRFBs with more complicated stackings of nonoverlapping real coefficient brickwall filters, e.g., as in Fig. 6.6. Given a set S of decimators, does such a PRFB using the set S *always* exist? Does its nonexistence imply that there is no PRFB using S with real coefficient filters (ideal or otherwise) at all? Currently we do not know the answers to these questions.

An important class of FBs studied in the later sections is that of all *rational* FBs, i.e., those in which all filters have rational transfer functions. As Section 6.6.1 will show, neither of the above decimator–sets $\{2, 3, 6\}$ and $\{2, 3, 7, 42\}$ permit existence of a rational PRFB (since they have pairs of coprime decimators). Thus it is tempting to conclude that the decimators of rational PRFBs are more restricted than those of real coefficient PRFBs. Indeed, intuition suggests that for any decimator–set S, existence of rational PRFBs using S implies that of real coefficient rational PRFBs using S. This is in fact true for all sets S for which rational PRFBs are currently known to exist. However, as we will see later, there are many sets for which it is not known whether either rational PRFBs or PRFBs with real coefficient filters (rational or otherwise) exist. Thus, in general we do not know whether existence of the former implies that of the latter. The constraint of realness of filter coefficients will not be applied or studied any further.

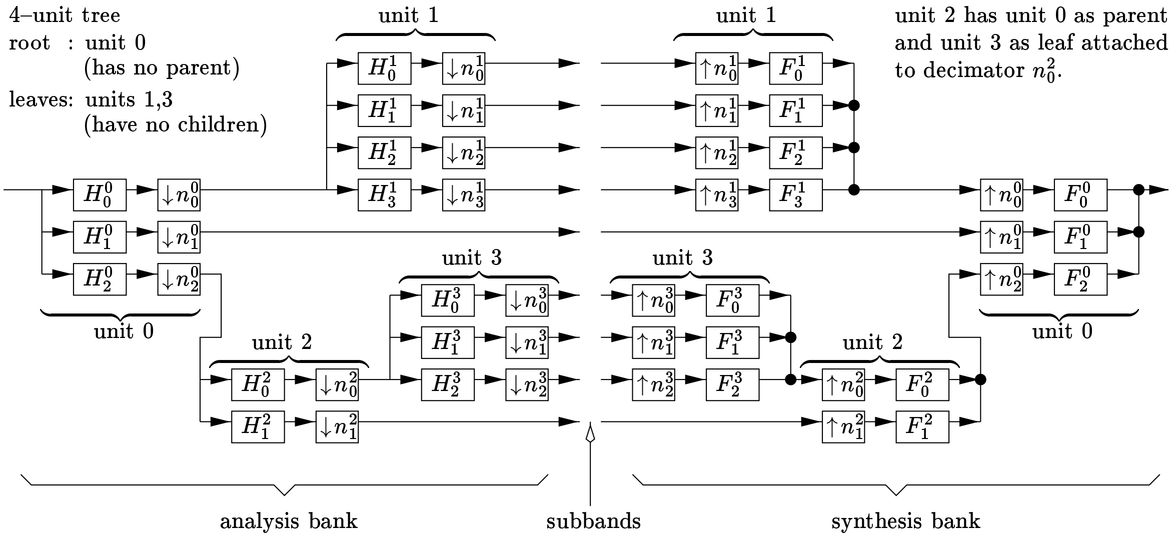


Figure 6.7: Tree structure of filter banks.

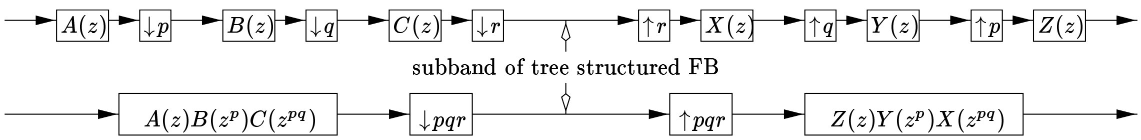


Figure 6.8: Structure of a channel of a tree structured FB.

6.4 Tree Structures

Cascading uniform PRFBs in a tree structure is the most common method of designing nonuniform PRFBs. As pointed out in Section 6.1, this method, though useful, is far from providing a complete PR theory of nonuniform FBs, i.e., a full solution to either of the two basic problems posed in Section 6.3.1. However, tree structures do provide very useful tools in the study of these problems. This section aims at analyzing their role in this study. Section 6.4.1 defines some basic terminology we will often use later in describing and studying tree structures. Section 6.4.2 analyzes the method of cascading uniform PRFBs in tree structures, and shows with examples how it falls short of a full PR theory of nonuniform FBs. Section 6.4.3 presents general methods that use trees to improve upon known PR conditions on the decimators of nonuniform FBs belonging to various FB classes. By 'improving a PR condition' we mean strengthening a necessary condition, or weakening a sufficient one. These methods will be applied to specific conditions later on.

6.4.1 Basics and terminology

A **tree structured FB** is one of the form shown in Fig. 6.7, built by repeated insertion of FBs into the subbands of other FBs. These constituent FBs of the tree structure will be called its **units**. They could be either uniform or nonuniform FBs, and may themselves be tree structured FBs. The

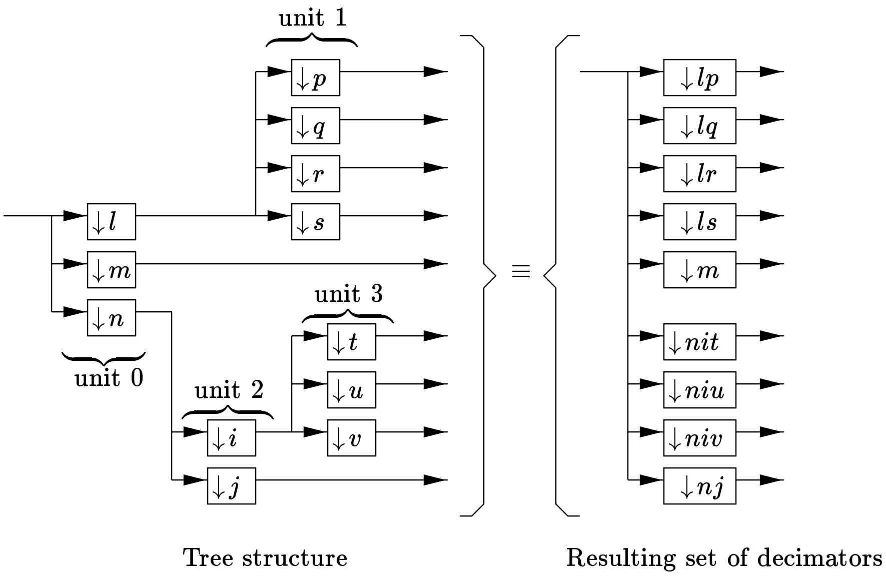


Figure 6.9: Tree structure of sets of decimators.

terms **parent, child, root and leaf units** will often be used to describe the relative positions of the units in the tree; their meanings are presumed to be self–evident or clear from the examples shown in Fig. 6.7. We also use the term **descendant**, an obvious extension of 'child'.

Figure 6.8 shows how the decimators and filters of a tree structured FB are related to those of its units. The same FB may be derivable from many trees, differing in the choice of filters in the units (e.g., in Fig. 6.8, replace filter $A(z)$ by $A(z)/T(z^p)$ and $B(z)$ by $B(z)T(z)$ for arbitrary $T(z)$) or even in the sets of decimators in the units and the number of units (e.g., combine units 2,3 of Fig. 6.7 into a single FB). Every FB is derivable from a **trivial tree**, which by definition is one with only one unit, i.e., one whose root is also a leaf. We will also use the notion of a **tree structured set of decimators**. Shown in Fig. 6.9, this is defined exactly as a tree structured FB, except that the units of the tree are now just sets of decimators rather than FBs. The distinction is made because while a decimator–set S may sometimes be derivable from many tree structures, a FB using S need not always be derivable from all of these. In fact derivability from *all* these trees usually occurs only in very special cases (e.g., when the set S is *dyadic*, Section 6.7). Often we have the other extreme where the trivial tree is the only one that the FB is derivable from. Two other useful notions are as follows:

Uniform–trees. A uniform–tree structure of FBs or decimator–sets is a tree structure in which all units are uniform. (A uniform decimator–set, like a uniform FB, is one in which all decimators are equal.) Its importance, elaborated in Section 6.4.2, stems from the fact that design of uniform FBs is so well understood.

Properties preserved by trees. It is fairly clear that a tree structure whose units are PRFBs generates a (tree structured) PRFB. Similarly a tree of rational FBs generates a rational FB. In general we say that a property of FBs is **preserved by tree structures** if it is true that whenever all units of a tree of FBs obey the property, so does the resulting tree structured FB. A similar statement holds for properties of decimator–sets. Two obvious but important properties of FBs preserved by trees are PR and maximal decimation. From Fig. 6.8, we can infer that the property of having filters that are rational, stable, real coefficient, FIR, or delays, and also the paraunitariness property, are all preserved by tree structures. The property of being a uniform FB is clearly *not* preserved by trees. Other useful nontrivial examples will be presented later (Section 6.6.3). As Section 6.4.3 will show, the ability of trees to strengthen known PR conditions on the decimator–sets depends crucially on whether or not certain properties are preserved by trees.

6.4.2 Uniform–trees: An incomplete PR theory for nonuniform FBs

A uniform–tree of FBs or decimator–sets is one in which each unit is uniform (i.e., all its decimators are equal). Its role in the central problem of Section 6.3.1 can be summarized as follows:

Role of uniform–trees. Derivability of a decimator–set S from a uniform–tree is a *sufficient* condition on S for existence of PRFBs using S and belonging to the specified FB class $\mathcal{C}$, for all $\mathcal{C}$ of interest in this work.

This statement follows from the simple fact that a uniform–tree whose units are rational PRFBs generates a rational PRFB, and so on. More generally, the statement holds for every FB class $\mathcal{C}$ having two features, namely (a) $\mathcal{C}$ contains uniform PRFBs with all decimation rates, and (b) the property of being in $\mathcal{C}$ is preserved by tree structures. All $\mathcal{C}$ of interest here, e.g., the rational and FIR FB classes, have these features. Thus it is important to have an algorithm to test whether or not a given decimator–set S is derivable from a uniform–tree. Such derivability is assured, for instance, if S has no more than two *distinct* decimators, or if each decimator divides every decimator greater than itself (e.g., when they are all powers of the same integer). Appendix B proves this, and gives complete algorithms to test for derivability from uniform–trees.

Due to the common use of uniform–trees to design nonuniform PRFBs, the term 'tree structure' in the literature sometimes implicitly refers only to uniform–trees. In this work however, trees are always more general, i.e., unless explicitly referred to as uniform–trees, they could have nonuniform units too. This is necessary, for as we now show, uniform–trees do not provide a complete PR theory for nonuniform FBs.

Deficiencies of uniform–trees.

1. *Uniform–tree condition is not necessary for PR:* There are decimator–sets S that are not derivable from uniform–trees, but can be used to build PR *delay–chain* FBs, i.e., FBs in which

all filters have the form z^{-k} for integer k. An example [21] is the set $S = \{6, 10, 15, 30, \ldots, 30\}$ (30 occurring 20 times), discussed in detail and generalized in Section 6.5.3. A delay–chain belongs to every FB class $\mathcal{C}$ of interest here (e.g., the FIR class). Thus, the uniform–tree condition is not necessary for any of these classes.

2. *Uniform–tree FBs are not a full parameterization:* Even if a decimator–set S is derivable from a uniform–tree, there may be PRFBs using S which are not derivable from any uniform–tree of FBs. We will now illustrate two examples of such a situation.

Example 6.1: Based on modifying filters of tree structured FBs. We take the analysis bank of a tree structured PRFB, find all subbands with a fixed decimation rate N, and transform them using an invertible square matrix $\mathbf{E}(z)$. If $\mathbf{h}(z)$ is the vector of analysis filters in the channels being transformed, the transform is equivalent to replacing $\mathbf{h}(z)$ by $\mathbf{h}'(z) = \mathbf{E}(z^N)\mathbf{h}(z)$. We preserve PR by effecting a corresponding change of synthesis filters using the inverse transform $\mathbf{E}^{-1}(z)$. Now if all the subbands being transformed come from the same unit FB in the tree, the transform can be effected by modifying only the filters of this unit, and the tree structure is preserved. More generally, if $\mathbf{E}(z)$ is block–diagonal with each block acting on subbands coming from the same unit of the tree, then again the tree structure can be preserved. However, this is no longer possible in general once we choose $\mathbf{E}(z)$ to avoid this degeneracy. In fact it is then fairly easy to ensure that the special structure of $\mathbf{h}(z)$ imposed by the tree is absent in the new filter vector $\mathbf{h}'(z)$. Thus, the new FB has the same decimator–set S but cannot be derived from the same tree. In particular if we choose a set S with a unique uniform–tree representation which is chosen as the starting tree in the above construction, the new FB is not derivable from a uniform–tree though its decimators are. An example of this kind is shown in [55], using FIR orthonormal FBs having the decimator–set $\{6, 6, 6, 6, 9, 9, 9\}$.

Example 6.2: Based on PR delay–chains. Consider the set with decimators 6,10,15,30 occurring 2,4,1,6 times respectively. From Section 6.5.2 we can show that this set can be used to build a PR delay–chain FB, which is clearly not derivable from a uniform–tree as the gcd of its decimators is unity. Now we build a tree in which the root is a uniform two channel PR delay–chain and the leaves are two such (identical) FBs, both children of the root. This yields a new (tree structured) PR delay–chain FB in which the gcd of the decimators is 2. Thus if this FB is to be derived from a *uniform–tree*, the root of the tree must be uniform with decimator 2. From its construction, this implies that in fact the new FB is *not* derivable from a uniform–tree. However, its set of decimators *is* derivable from a uniform–tree (in fact, in multiple ways). Note that this example cannot be produced starting from a uniform–tree of FBs in the manner used to create Example 6.1 above. Thus it shows a *deeper reason* for the incompleteness of FB parameterizations using uniform–trees.

6.4.3 Using trees to improve PR conditions on the decimators

Weakening sufficient conditions

Let S be a general decimator–set (obeying (6.1)). We seek conditions on S that permit existence of a PRFB that uses S as its decimator–set and belongs to some specified FB class $\mathcal{C}$. For all $\mathcal{C}$ of interest here, the most elementary but very strong *sufficient* condition on S for this purpose is that S be uniform (i.e., all its decimators be equal), as uniform FBs can always be built. However, using the fact that the FB class definition (i.e., property of being in the class $\mathcal{C}$) is preserved by tree structures, Section 6.4.2 has obtained a much weaker (and hence improved) sufficient condition, namely that S be derivable from a *uniform–tree*.

The process just described can be easily generalized to improve (i.e., weaken) *any* sufficient condition $\mathcal{P}$ on the decimator–set S (rather than merely the condition that S be uniform). The only requirement for this process to work is that the FB class definition be preserved by trees (which holds for all classes of interest here). The improved sufficient condition, denoted by $\mathcal{P}'$, states that S be derivable from a tree in which all units obey the original sufficient condition $\mathcal{P}$. Because S is always derivable from the trivial tree, the new condition is indeed weaker, i.e., $\mathcal{P}$ implies $\mathcal{P}'$. It is also easy to test for $\mathcal{P}'$ once we have a test for the original condition $\mathcal{P}$: We simply list all possible tree representations of S and run the test for $\mathcal{P}$ on all units of each of them. (Of course the specific nature of $\mathcal{P}$ could make even faster tests possible.)

It may happen that $\mathcal{P}' \equiv \mathcal{P}$, i.e., the 'weaker' condition is *not strictly weaker*. For example, suppose $\mathcal{P}$ itself is preserved by trees. Then if S is derivable from a tree in which each unit obeys $\mathcal{P}$ (i.e., $\mathcal{P}'$ holds), it implies that S itself obeys $\mathcal{P}$. Thus $\mathcal{P}' \equiv \mathcal{P}$ here. In fact a little further thought shows that $\mathcal{P}' \equiv \mathcal{P}$ if and only if $\mathcal{P}$ is preserved by trees. Note that by its definition, $\mathcal{P}'$ itself is preserved by trees. Thus repeated application of the above method cannot weaken the sufficient condition $\mathcal{P}$ any more than the first one does.

The only currently known instance where the above method strictly weakens a sufficient PR condition on decimator–sets is the one mentioned at the beginning of this section, leading to the uniform–tree sufficient condition. (Of course we can create other artificial instances, which lead to sufficient conditions that are stronger, and hence not as useful.) We will now show a method to improve *necessary* conditions, and see that there are more nontrivial examples where this method causes a strict improvement.

Strengthening necessary conditions

We begin by illustrating the general method using a specific necessary condition that follows from Theorem 6.4 of Section 6.6.2. The condition states that the decimator–set of a rational PRFB cannot have a subset of $g+1$ decimators within which the gcd of any pair is g. The set $S = \{2, 4, 8, 12, 24\}$

can be seen to obey this condition. Suppose there is a rational PRFB using decimator–set S. We can create tree structures whose units are this and other rational PRFBs. The resulting FB also is a (tree structured) rational PRFB. Hence its decimator–set must obey the above necessary condition too. Thus we can obtain a new and stronger necessary condition on S by applying the original one to all the tree structured decimator–sets created from S as just described. In the present case, this new condition is *strictly stronger:* Using a two–unit tree in which the leaf is uniform with decimator 2 and is attached to the decimator 2 in the root S, we obtain the decimator–set $\{4,4,4,8,12,24\}$ which violates the original condition. (Its subset $\{4,4,4,8,12\}$ has 5 decimators within which the gcd of any pair is 4.)

The new necessary condition $\mathcal{P}''$ created as above from the original condition $\mathcal{P}$ will be referred to as the **tree version** of the necessary condition $\mathcal{P}$. It is stronger, i.e., $\mathcal{P}''$ implies $\mathcal{P}$, because in particular the tree chosen in the above construction can be taken as the trivial one with S as its only unit. Generalizing the above example, we summarize the method of strengthening necessary conditions as follows:

Theorem 6.1: Tree versions of necessary conditions. Let $\mathcal{C}$ be a FB class such that the property of being in $\mathcal{C}$ is preserved by trees. Let $\mathcal{P}$ be a *necessary* condition on a general decimator–set S for existence of a PRFB in $\mathcal{C}$ with S as its set of decimators. Consider any decimator–set S'' derivable from a tree structure in which each unit is either identical to S or allows building of PRFBs in $\mathcal{C}$ (i.e., obeys some sufficient condition). Let $\mathcal{P}''$ be the condition that all such sets S'' satisfy $\mathcal{P}$. Then $\mathcal{P}''$ is also a necessary condition on S, called the *tree version* of the necessary condition $\mathcal{P}$. ■

Remarks:

1. We have just defined tree versions of necessary conditions, which are stronger necessary conditions. Earlier we had defined tree versions of sufficient conditions, which are weaker sufficient conditions. Some basic differences exist between these two methods of using trees to improve known conditions. For example, the above definition of the necessary condition $\mathcal{P}''$ involves a known sufficient condition. The weaker this sufficient condition, the stronger $\mathcal{P}''$ becomes. This is notably different from the earlier situation for tree versions of sufficient conditions.

2. *Algorithm to test $\mathcal{P}''$.* The condition $\mathcal{P}''$ on S demands that $\mathcal{P}$ hold for several sets S'' derived from S (including S itself) as described above. As there are infinitely many of the S'', we cannot state a general algorithm that tests for $\mathcal{P}''$. One needs specific tests designed using the features of $\mathcal{P}$ and the sufficient condition used to define $\mathcal{P}''$. This is again unlike the situation for tree versions of sufficient conditions.

3. *When are tree versions not* strictly *stronger?* Suppose $\mathcal{P}$ is preserved by tree structures. Then if S obeys $\mathcal{P}$, all units in the tree generating S'' obey $\mathcal{P}$, and hence so does S''. Thus, S obeys

$\mathcal{P}''$ too, i.e., $\mathcal{P}'' \equiv \mathcal{P}$. Here too, as with tree versions of sufficient conditions, $\mathcal{P}''$ is preserved by trees, and is hence unchanged by forming its tree version. The only difference is that now we cannot in general claim that $\mathcal{P}'' \equiv \mathcal{P}$ implies that $\mathcal{P}$ is preserved by tree structures.

Tree versions of necessary conditions have not been observed earlier. A possible reason for this is that the simplest known necessary conditions for the rational FB class (compatibility and pairwise noncoprimeness, Section 6.6.1) are preserved by trees, and are hence identical to their tree versions. Section 6.6.3 shows another necessary condition made strictly stronger by forming its tree version.

6.5 Delay–chains

A delay–chain FB is one in which all filters are delays, i.e., of form z^{-k} for integer k. (We call z^{-k} a delay even though it is actually an 'advance' for $k < 0$.) Such a FB, while quite useless from a practical standpoint, is a useful tool in solving the problems of Section 6.3.1. This section presents a complete solution to these problems when the class $\mathcal{C}$ of FBs under study is that of delay–chain FBs. Because delays trivially obey various filter properties, delay–chain PRFBs belong to every class $\mathcal{C}$ studied in this chapter: They are FBs with FIR, rational, stable, real coefficient filters, and we will see that they are also paraunitary. Thus, solving the problem of Section 6.3.1 for the class of delay–chains yields a *sufficient* condition on the decimators for existence of PRFBs in any of these classes. We will see that this condition is strictly weaker than the other sufficient condition studied earlier in Section 6.4.2, namely derivability from uniform–trees. In fact existence of a delay–chain is the weakest known sufficient condition for all of the earlier mentioned rational FB classes.

6.5.1 PR condition on the set of decimators

In Fig. 1.8, if $H_k(z) = z^{-l_k}$ for all k, the k-th subband signal is $c_k(n) = x(n_k n - l_k)$, i.e., it contains a certain subset of the input samples $x(n)$. Let $L = \mathrm{lcm}\{n_i\}$, and consider any L consecutive samples of $x(n)$. We see that the k-th subband contains exactly L/n_k of these samples. Due to maximal decimation, we have

$$\sum_k (L/n_k) = L \tag{6.16}$$

Thus, if any of the L chosen input samples occurs in more than one subband, there must be a sample that does not occur in any subband. In this case, PR is clearly impossible no matter what the choice of synthesis filters. On the other hand, if none of the input samples occurs in more than one subband, then (6.16) implies that each of them occurs in exactly one subband. We can then achieve PR by appropriately interleaving the subband samples, which is done by the choice of synthesis filters as $F_k(z) = z^{l_k}$. Thus, PR is possible iff no input sample occurs in more than one subband. This condition means that if $i \neq j$, then $n_i n - l_i \neq n_j m - l_j$, i.e., $l_i - l_j \neq n_i n - n_j m$, for any integers n, m. As n, m range over all integers, the right side here ranges over all multiples of $\gcd(n_i, n_j)$. Thus the PR condition may be summarized as follows:

Theorem 6.2: Delay–chain PRFBs. In Fig. 1.8, if $H_k(z) = z^{-l_k}$ for integers l_k, PR is possible iff no input sample occurs in more than one subband. Under this condition, PR is obtained with the unique choice $F_k(z) = \widetilde{H}_k(z) = z^{l_k}$, yielding a PR delay–chain FB, which is thus always paraunitary. The necessary and sufficient condition on the decimators n_i for existence of such a FB is that there exist integers l_i satisfying

$$(l_i - l_j) \not\equiv 0 \pmod{\gcd(n_i, n_j)} \quad \text{if } i \neq j \tag{6.17}$$

6.5.2 Testing the PR condition

Given the decimators n_k, it is required to test for existence of integers $l_0, l_1, \ldots, l_{M-1}$ obeying (6.17). Now if (6.17) holds for some integers l_k, then it also holds if each l_k is replaced by $l_k + m_k n_k + C$ for any integers m_k and any fixed integer C. Hence, without loss of generality we can assume that $0 \leq l_k < n_k$ and $l_0 = 0$. This makes the number of possible sets of l_k *finite*, so clearly there is an algorithm for our purpose.

For example, we can try to assign the l_k sequentially, as follows: Suppose we have $l_0, l_1, \ldots, l_{N-1}$ obeying (6.17) for some $N < M$. We assign to l_N all possible values obeying $0 \leq l_N < n_N$ and $(l_N - l_j) \not\equiv 0 \pmod{\gcd(n_N, n_j)}$ for $j = 0, 1, \ldots, N-1$. Each value yields a larger set $l_0, l_1, \ldots, l_N$ obeying (6.17), and we can now repeat the process on this set. If there is no valid choice for l_N, we must restart with another valid set of choices for $l_0, l_1, \ldots, l_{N-1}$. Initializing this recursive scheme using $l_0 = 0$, we can thus list *all* sets $\{l_0, l_1, \ldots, l_{M-1}\}$ obeying (6.17). In particular this finds whether or not there exist such sets. This solves both problems of Section 6.3.1 for the class $\mathcal{C}$ of delay–chain FBs. To determine only the existence of a PR delay–chain, the above algorithm can often be accelerated using the following result:

Fact 1. Let integers $l_0, \ldots, l_{N-1}$ obey (6.17) for some $N < M$, and let n_N be a common multiple of $n_0, n_1, \ldots, n_{N-1}$. Then there is an integer l_N such that $0 \leq l_N < n_N$ and $l_0, \ldots, l_{N-1}, l_N$ obey (6.17) too.

Proof: From the premise, $l_0, \ldots, l_{N-1}, l_N$ will satisfy (6.17) if and only if

$$(l_N - l_j) \not\equiv 0 \pmod{\gcd(n_N, n_j)} \quad \text{for } j = 0, 1, \ldots, N-1. \tag{6.18}$$

Also $\gcd(n_N, n_j) = n_j$. Thus (6.18) is equivalent to $l_N \neq l_j + nn_j$ for all integers n, for $j = 0, 1, \ldots, N-1$. For each j there are $\frac{n_N}{n_j}$ integers of the form $l_j + nn_j$ in the range $[0, n_N)$. Thus, of the n_N integers in $[0, n_N)$, at most $B = \sum_{k=0}^{N-1} \frac{n_N}{n_k}$ of them are excluded as possible choices of l_N (in fact we can even show that exactly B choices are excluded). As $N < M$, maximal decimation (6.1) means that $B < n_N$, so there are still valid choices of l_N in the interval $[0, n_N)$. $\triangledown\triangledown\triangledown$

Thus, suppose there is a decimator n_N such that each $n_j \geq n_N$ is a multiple of all $n_i < n_j$. It then suffices to verify existence of valid delays l_k obeying (6.17) for all $n_k < n_N$. As an extreme case, if every n_j is a multiple of all $n_i < n_j$ (i.e., every n_j divides all $n_i > n_j$), then a delay–chain PRFB always exists. In fact the decimator–set is then derivable from a uniform–tree (Appendix B). Fact 1 is also useful in proving Theorem 6.3 which follows soon.

Nonuniqueness of delay–chains: When a decimator–set allows building of a PR delay–chain FB, in general this delay–chain is not unique. The non-uniqueness can be much deeper than that caused simply by adding a constant delay to all the filters. For example, when several delay–chains are possible, it could happen that some of them are also derivable from uniform–trees, while some others are not, as seen in Section 6.4.2.

6.5.3 Delay–chains vs. uniform–trees

Our study of tree structures showed that (a) known PR conditions on decimators can sometimes be strengthened using trees (Section 6.4.3), and (b) derivability of the decimators from a uniform–tree is a *sufficient* PR condition for all FB classes that we study (Section 6.4.2). Does this teach us more about delay–chains? Firstly, the condition (6.17) is both necessary and sufficient for existence of PR delay–chains. Hence it remains unaltered by the procedures of Section 6.4.3. Next, the uniform–tree condition is *not* necessary, as we now show:

Theorem 6.3: PR delay–chains without uniform–trees. There are infinitely many PR delay–chain FBs that cannot be derived from uniform–trees. Such FBs can be built using every set of decimators of the form

$$S = \{n_0, n_1, n_2, L, L, \ldots, L\} \text{ where } L = \mathrm{lcm}(n_0, n_1, n_2) \text{ occurs } L\left(1 - \sum_{i=0}^{2} \frac{1}{n_i}\right) \text{ times,} \tag{6.19}$$

and $(n_0, n_1, n_2) = (m_1 m_2, m_2 m_0, m_0 m_1)$ for some pairwise coprime integers $m_i > 1$. $\blacksquare$

Proof: By Fact 1, decimators of S allow building of a PR delay–chain FB iff there are integers l_0, l_1, l_2 obeying (6.17) for $i, j \in \{0, 1, 2\}$. This condition is easily ensured, in fact it holds iff $\gcd(n_i, n_j) \geq 2$ for $i, j \in \{0, 1, 2\}$ with strict inequality for at least one $i \neq j$. (We can then make a valid choice of the l_i from the numbers 0,1,2.) Further if $\gcd(n_0, n_1, n_2) = 1$, the set cannot be derived from a uniform–tree (Appendix B). Both these requirements are satisfied by the choice of n_i stated by the theorem. $\triangledown\triangledown\triangledown$

An example of a delay–chain PRFB not derivable from a uniform–tree was first shown in [21]. Its set of decimators $\{6, 10, 15, 30, 30, \ldots, 30\}$ (30 occurring 20 times) is a special case of the construction of Theorem 6.3 with $(m_0, m_1, m_2) = (5, 3, 2)$. This is not the only way to produce such examples: Delay–chain PRFBs can also be built with the decimator values $6, 10, 15, 30$ when the number of their respective occurrences are $2, 4, 1, 6$ or $2, 3, 2, 7$. The former set of decimators is the *smallest* such example.[3] It can be used as the root of a tree to derive the example of [21], but not the latter example. In all these cases, the decimators have no common factor, ensuring that they are not derivable from uniform–trees. In fact if the decimators of a delay–chain PRFB do have a common factor, the FB can be built from smaller PR delay–chains as follows:

Fact 2. Let all decimators in a PR delay–chain FB have common factor $K > 1$. Then the FB can be derived from a tree structure in which each unit is a PR delay–chain FB and the root is uniform with decimator K.

Proof: Let $x(n)$ be the FB input. For $0 \leq k < K$, let $f_k(n) = x(Kn - k)$, which is the k-th subband signal in a uniform K channel delay–chain PRFB. Now consider the i-th channel of the given PR delay–chain, with decimator n_i, analysis filter z^{-l_i}, and hence, subband signal $x(n_i n - l_i)$. Since n_i is a multiple of K, either *all* its samples lie in the sequence $f_k(n)$, or none of them do (depending on whether or not $l_i \equiv k \pmod{K}$). We now collect all subbands whose samples do lie (entirely) in $f_k(n)$. Due to the PR condition for delay–chains (Theorem 6.2), these subbands collectively contain all samples of $f_k(n)$ (as none of the other subbands have any of them), and each of these samples occurs in exactly one of these subbands. Further the delays in all these subbands are equal (to k) modulo K. Thus these subbands can be generated by inserting a suitable delay–chain PRFB as a child (in a tree) in the k-th subband signal $f_k(n)$ of a uniform K channel delay–chain PRFB. Repeating this process for $k = 0, 1, \ldots, K - 1$ yields the desired tree structure. $\bigtriangledown\bigtriangledown\bigtriangledown$

Remarks:

1. The above result does not generalize easily to other classes of FBs (besides delay–chains). For example, consider the decimators $\{4, 4, 4, 4\}$, having common factor $K = 2$. These decimators can be used to build *rational* and *FIR* PRFBs that are not derivable from *any* tree structure (besides the trivial one).

2. A common factor $K > 1$ among all decimators does not by itself ensure their derivability from a tree whose root is uniform with K as decimator.[4] However, if the decimators also allow building of a delay–chain PRFB, then by Fact 2, there is at least one such tree, as the FB itself is derivable from such a tree.

3. All decimators of a PR delay-chain FB need not have a common factor $K > 1$ (see the example in Theorem 6.3). However, further conditions on the decimators can force such a

[3] This is true when size is measured by either the number of decimators, or their lcm, or the largest one. In fact there is no other example with ≤ 13 decimators, as an exhaustive search aided by a computer and Fact 2 will show.

[4] The set of decimators $\{4, 6, 6, 10, 10, 10, 10, 60\}$ shows this for $K = 2$. The choice of root prevents the leaves from obeying (6.1).

common factor to exist, thus making Fact 2 apply. For example, suppose the PR delay–chain has a decimator of value m occurring $m-1$ times (m is thus the smallest decimator). Then all decimators must have m as a factor. This is provable by a slight extension of the proof of Fact 2. In fact it even generalizes to rational FBs in place of delay–chains (Theorem 6.5, Section 6.7), although this is harder to prove.

6.6 The class of rational FBs

In this section and most of Section 6.7, the FB class $\mathcal{C}$ of interest is that of rational FBs, i.e., FBs in which all filters are rational. We seek necessary and sufficient conditions on a decimator–set S for existence of a rational PRFB using S. The *weakest known sufficient condition* is that of existence of a PR *delay–chain* (Section 6.5). This is clearly sufficient since delay–chains are rational FBs, but *is it also necessary*? Or is there a decimator–set which does not permit existence of PR delay–chains, but allows building of rational PRFBs (whose filters are not all delays)? This is a major open question in the PR theory of nonuniform FBs.

A possible approach to answer the above question is to try to build a rational PRFB with decimators that do not allow building of PR delay–chains. However, starting with an arbitrary decimator–set such as $S = \{2, 3, 6\}$ does not help, as S violates a known *necessary* condition (called 'compatibility', Section 6.6.1) on the decimators of a rational PRFB. Such sets must be excluded, and to this end it helps to derive more necessary conditions. This is our main contribution in this section. The previously known necessary conditions for PR are described in Section 6.6.1. Each subsequent subsection develops a new necessary condition that is strictly stronger than a previously known one. Table 1 (Section 6.8) presents a comprehensive summary of all known conditions, many of which are new results of the present work. The table studies the interrelationship between the conditions, and lists example decimator–sets illustrating their use.

All the new necessary conditions we develop still collectively remain insufficient for existence of *delay–chain* PRFBs, and thus it is *still not known* whether they are sufficient for existence of rational PRFBs. Our work reduces the 'gap' between the necessary conditions and the sufficient one. Proving that the sufficient condition is in fact necessary would in some sense render obsolete most of the present section. However this appears tough to do, in fact the statement may not even be true. Our work is a step towards the truth.

6.6.1 Previously known necessary conditions on decimators

1. *Pairwise noncoprimeness.* No two decimators of a rational PRFB can be coprime [21]. If $\gcd(n_i, n_j) = 1$ for two decimators n_i, n_j in Fig. 1.8, the biorthogonality condition (6.5) for PR implies $H_iF_j = 0$ and $(H_iF_i)\downarrow_{n_i} = (H_jF_j)\downarrow_{n_j} = 1$. This is impossible for a rational FB,

as $H_i F_j = 0$ forces $H_i \equiv 0$ or $F_j \equiv 0$.

2. *Compatibility.* Every decimator occurring only once must divide some other decimator [3, 28, 21]. In particular, the largest decimator must occur at least twice. As Section 6.6.4 will show (see Example A), without this condition the rational FB cannot even be a nonzero LTI system, let alone have PR.

3. *Strong compatibility.* This condition, developed in [21], places a lower bound $b_j \geq 1$ on the number of occurrences N_j of each decimator n_j. The condition is stated as follows:

$$N_j \geq b_j \triangleq \frac{1}{p_j}\left(\min_{p_i \neq p_j} \operatorname{lcm}(p_i, p_j)\right), \quad \text{where } p_i = \frac{L}{n_i}, \quad L = K(\operatorname{lcm}\{n_i\}) \text{ for any positive integer } K \tag{6.20}$$

   This will be shown in Section 6.6.4, which in fact proves a new condition strictly stronger than the above.

Note that the bound b_j of (6.20) is independent of the integer K. Also, it only needs verification for *distinct* decimator values, because if $n_i = n_j$ then $N_i = N_j$, $b_i = b_j$. For a *uniform* set of decimators, $p_i = p_j$ for all i, j, so we define $b_j = 1$ here (so that the bound holds). Excluding this case, $b_j = 1$ iff p_j is a multiple of some $p_i \neq p_j$, i.e., iff n_j divides some distinct decimator n_i. So the bound need not be checked for such decimators. Also, strong compatibility implies compatibility, because it demands that any n_j occurring only once (i.e., with $N_j = 1$) must have $b_j = 1$, i.e., must divide some other decimator. In fact strong compatibility is a *strictly* stronger necessary condition than compatibility, as shown by the set of decimators $\{2, 4, 6, 24, 24\}$. However it does not imply pairwise noncoprimeness [3] (shown by $\{2, 5, 10, 10, 10\}$). Likewise, a set could satisfy pairwise noncoprimeness but violate compatibility (and hence strong compatibility), e.g., $\{2, 4, 6, 12\}$.

6.6.2 The pairwise gcd test

Theorem 6.4: Pairwise gcd test. Among the decimators of a rational PRFB, there cannot be a subset of $g+1$ decimators such that the gcd of any two elements from the subset is a factor of g. In particular (for $g = 1$), this implies the pairwise noncoprimeness condition (Section 6.6.1). ■

Proof: As with pairwise noncoprimeness, the proof uses the biorthogonality condition (6.5) for PR. Let $g+1$ decimators $n_0, n_1, \ldots, n_g$ be such that the gcd of any pair divides g. From (6.5), $(H_i(z)F_j(z)) \downarrow_{\gcd(n_i, n_j)} = 0$ if $i, j \in \{0, 1, \ldots, g\}$, $i \neq j$. In this case $g/\gcd(n_i, n_j)$ is given to be an integer, so decimating both sides by it,

$$(H_i(z)F_j(z)) \downarrow_g = 0, \quad \text{for } i, j \in \{0, 1, \ldots, g\},\ i \neq j \tag{6.21}$$

Form the g-th order analysis polyphase matrix $\mathbf{E}(z)$ (of size $(g+1) \times g$) of the filters $H_i(z)$, and the

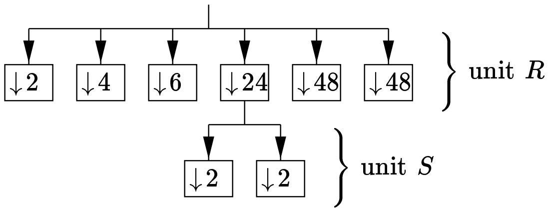


Figure 6.10: Showing that strong compatibility is not preserved by trees.

g-th order synthesis polyphase matrix $\mathbf{R}(z)$ (of size $g \times (g+1)$) of the F_i. Thus, from the polyphase lemma (Section 6.1.3), for $i, j \in \{0, 1, \ldots, g\}$, $(H_i(z)F_j(z)) \downarrow_g$ is the ij-th entry in the product $\mathbf{P}(z) = \mathbf{E}(z)\mathbf{R}(z)$. Hence by (6.21), $\mathbf{P}(z)$ is a $((g+1) \times (g+1))$ *diagonal* matrix. Its i-th diagonal entry is the filter $(H_i(z)F_i(z)) \downarrow_g$, with impulse response $c_i(gn)$, where $c_i(n)$ is the impulse response of H_iF_i. From (6.5), $(H_i(z)F_i(z)) \downarrow_{n_i} = 1$, so $c_i(n_in) = \delta(n)$ i.e., $c_i(0) = 1$. Hence $c_i(gn) \not\equiv 0$, i.e., no diagonal element of $\mathbf{P}(z)$ is identically zero. Thus, as these elements are rational filters, there is a z such that $\mathbf{P}(z)$ has full rank $g+1$. However, this is impossible from the sizes of $\mathbf{E}(z), \mathbf{R}(z)$. $\triangledown\triangledown\triangledown$

6.6.3 Tree version of strong compatibility

In Section 6.4.3, we saw how given a necessary condition $\mathcal{P}$ on the decimators for PR, we could form its 'tree version' $\mathcal{P}''$, which is a stronger (though not necessarily *strictly* stronger) necessary condition. We can apply this process to the conditions of Section 6.6.1. Some thought shows that both the pairwise noncoprimeness and the compatibility conditions are preserved by tree structures. They are hence identical to their tree versions (as seen in Section 6.4.3). However, the same is not true with strong compatibility: Its tree version is strictly stronger than itself. This is shown by the two–unit tree in Fig. 6.10. Both units R and S are strong compatible, and S allows building of rational PRFBs (as it is uniform). However the resulting set of decimators is not strong compatible. Hence, though R obeys strong compatibility, it violates its tree version.

A complete algorithm to test this new necessary condition is described in Appendix E. Its derivation involves characterizing trees similar to that in Fig. 6.10. This is done by the following results:

Fact 3. Consider a set T of decimators derived from a 2-unit tree structure having root R and leaf S attached to decimator m_0 of R. Suppose R, S are strong compatible but T is not. Then S is a uniform unit, i.e., all its decimators have equal value K. The decimator m_0 of R does not occur in T, i.e., it occurs only once in R. The decimator m_0K of T obtained at the leaf S also occurs in R. Decimators of this value m_0K are the only ones violating the strong compatibility lower bound on the number of their occurrences in T.

Fact 4. Let a set D of decimators satisfy strong compatibility but violate its tree version. Then there is a tree $\mathcal{T}$ generating a set T of decimators, such that $\mathcal{T}$ and T have the following properties:

1. The tree $\mathcal{T}$ has root D. All leaves of the tree are uniform and are children of its root. All decimators obtained at the leaves have equal value d.

2. If d_i are the decimators of D to which leaves are attached in $\mathcal{T}$, then no decimator in T has value d_i.

3. If $d \notin D$, then $d = \operatorname{lcm}\{d_i\}$. Hence, if $d \notin D$, the d_i are not all equal (for otherwise, $d = d_i \in D$).

4. Decimator $d \in T$ violates the strong compatibility lower bound on the number of its occurrences in T.

Fact 3 is proved in Appendix C, and is used to prove Fact 4 in Appendix D. Fact 4 gives an algorithm to test whether the set D obeys the tree version of the strong compatibility condition: We find all trees with root D and properties 1,2 and 3 listed in Fact 4. It can be seen that there are only finitely many such trees, and from Fact 4, D violates the condition if and only if one of these trees also obeys property 4. This idea is the basis of the detailed algorithm of Appendix E.

6.6.4 The AC-matrix test

The necessary condition derived here relies heavily on the AC matrix formulation (6.11),(6.12) of the PR condition on the filters of the FB. The algorithm to test the condition is described in Appendix F, and may be taken as the statement of the condition (i.e., this condition, unlike the earlier ones, does not have a short / simple statement). Like the test of Section 6.6.3, this test also strictly strengthens strong compatibility, but in an independent direction. In this section, we derive two lemmas and use them to explain the operation of the test, illustrate it with examples, and thus justify the algorithm of Appendix F. Developing the new test will also *prove* that strong compatibility is necessary for PR; a result *assumed* in deriving the test of Section 6.6.3. We will further show that (simple) compatibility is necessary even if we allow the rational FB to violate PR but merely insist that it be an *LTI* system (i.e., an *aliasfree* FB) that is not identically zero.

Two key results used by the test

Lemma 6.2. In Fig. 1.8, if $\sum_k H_{i_k}(z)F_{i_k}(z) = 0$ for any set of i_k, $0 \leq i_k < M$, then the FB cannot have PR.

Proof: If the FB of Fig. 1.8 has PR, $(H_{i_k}(z)F_{i_k}(z)) \downarrow_{n_{i_k}} = 1$ by biorthogonality (6.5). Let $L = \operatorname{lcm}\{n_i\}$, thus $(H_{i_k}(z)F_{i_k}(z)) \downarrow_L = (1) \downarrow_{(L/n_{i_k})} = 1$. So $(\sum_k H_{i_k}(z)F_{i_k}(z)) \downarrow_L \neq 0$, violating $\sum_k H_{i_k}(z)F_{i_k}(z) = 0$. $\bigtriangledown\bigtriangledown\bigtriangledown$

Lemma 6.3. Given *rational* filters $B_i(z), C_i(z)$, $0 \leq i < N$, let $W = e^{-j2\pi/M}$ and $G_l(z) = \sum_{i=0}^{N-1} B_i(zW^l)C_i(z)$. If $G_l(z) = 0$ for N values of l occurring consecutively in an arithmetic progression, then $G_l(z) = 0$ *for all* values of l in this progression. (The lemma in fact holds for any nonzero complex W.)

Proof: For $N = 1$, the lemma is to be interpreted as follows: If $B_0(zW^l)C_0(z) = 0$ for some l, then it holds for all l. This is clearly true: Rational filters $B_0(z), C_0(z)$ obey $B_0(zW^l)C_0(z) = 0$ iff $B_0 \equiv 0$ or $C_0 \equiv 0$ or both. (Note however that this is in general false if we remove the rationality constraint.) Hence, let $N > 1$. Let the N given consecutive values of l in arithmetic progression be $s, s+d, s+2d, \ldots, s+(N-1)d$. Define

$$\mathbf{b}(z) = \begin{bmatrix} B_0(zW^s) & B_1(zW^s) & \ldots & B_{N-1}(zW^s) \end{bmatrix}, \mathbf{c}(z) = \begin{bmatrix} C_0(z) & C_1(z) & \ldots & C_{N-1}(z) \end{bmatrix}^T \tag{6.22}$$

The lemma can then be stated as follows: If $\mathbf{b}(zW^{nd})\mathbf{c}(z) = 0$ for $n = 0, 1, \ldots, N-1$ then it is true for all integers n. To show this, form the square matrix $\mathbf{B}(z)$ with rows $\mathbf{b}(zW^{nd})$, $0 \leq n < N$. By the premise of the lemma, $\mathbf{B}(z)\mathbf{c}(z)$ is the zero vector. This implies linear dependence of the columns of $\mathbf{B}(z)$, and hence of its rows, as it is square. So $\sum_{n=0}^{N-1} \alpha_n(z)\mathbf{b}(zW^{nd}) = 0$ for some rational filters $\alpha_n(z)$ not all identically zero. Let r be the maximum n for which $\alpha_n(z) \not\equiv 0$. Divide the above relation by $\alpha_r(z)$. (This is allowed solely due to the rationality assumption: Otherwise $\alpha_r(e^{j\omega})$ could for instance be zero in an interval.) This yields

$$\mathbf{b}(zW^{rd}) = \sum_{n=0}^{r-1} \beta_n(z)\mathbf{b}(zW^{nd}) \qquad \text{for some rational filters } \beta_n(z). \tag{6.23}$$

Replacing z by zW^d and postmultiplying by $\mathbf{c}(z)$ shows $\mathbf{b}(zW^{nd})\mathbf{c}(z) = 0$ for $n = r+1$. Using this and repeating the process shows the same for $n = r+2$. Carrying on this way, the result is shown for all $n \geq 0$. For $n < 0$ we use a similar procedure, now choosing r as the *minimum* n for which $\alpha_n(z) \not\equiv 0$. $\nabla\nabla\nabla$

Deriving the test from the AC matrix and Lemmas 6.2,6.3

Examine closely Equations (6.11),(6.12) which are equivalent to PR. Number the rows of the AC matrix from 0 to $L-1$, and the columns from 0 to $M-1$, and let $W = e^{-j2\pi/L}$. The l-th row equation in (6.11) has form

$$\sum_k p_{i_k} H_{i_k}(zW^l)F_{i_k}(z) = 0 \tag{6.24}$$

The summation here ranges over all indices i_k for which the i_k-th column in the AC matrix has a nonzero entry in the l-th row. This happens if and only if l is a multiple of $p_{i_k} = L/n_{i_k}$. Thus, suppose that for all integers m in some set S, the number $l = mp_0$ is not a multiple of any $p_i \neq p_0$.

For all these $l = mp_0$, (6.24) holds with the summation being over the same set of filters, i.e., those corresponding to decimator value n_0. Thus, if N_0 is the number of occurrences of decimator n_0, (6.24) takes the form

$$\sum_{i=0}^{N_0-1} p_0\, H_i(zW^l)F_i(z) = 0, \quad \text{for } l = mp_0, \quad \text{for } m \in S \tag{6.25}$$

This is very similar to the system $\sum_{i=0}^{N-1} B_i(zW^l)C_i(z) = 0$ of Lemma 6.3, with $N = N_0$. The only difference is that here the premise of the lemma may or may not hold, i.e., (6.25) may or may not hold for N_0 values of l occurring consecutively in an arithmetic progression.

The main idea of the test we are developing is to find all progressions for which the premise of Lemma 6.3 actually holds, and then use the lemma. This may sometimes allow us to deduce that (6.25) actually holds for other values of l too, besides those stated in (6.25) itself. If $l = 0$ turns out to be one of these, then by Lemma 6.2 we can conclude that PR is impossible, i.e., the given set of decimators fails the test. To perform such a test, we must use the known values of l from (6.25) to find progressions obeying the premise of Lemma 6.3. From (6.25) it clearly suffices to examine progressions of integers whose common difference d is an integer multiple of p_0. There are infinitely many such progressions, each an infinite sequence. However, since $W^L = 1$, it suffices to consider the progressions modulo L, and to restrict their common difference d as $d < L$. In fact even $d \leq \lfloor L/2 \rfloor$ suffices, as any progression with common difference $L - d$ can be generated in reverse order by one with difference d. We will now show examples of the working of the above test.

Example A: Compatibility is a special case, and is necessary even for aliasfree FBs. Suppose p_0 occurs only once and is not a multiple of any $p_i \neq p_0$. In other words, n_0 occurs only once and does not divide any other n_i, i.e., compatibility (Section 6.6.1) is violated. Then (6.25) holds with $N_0 = m = 1$, and a trivial use of Lemma 6.3 shows that indeed (6.25) also holds for $l = 0$ (i.e., $H_0F_0 \equiv 0$). Thus by Lemma 6.2, PR is impossible. Hence, passing our new test implies compatibility of the decimators. In fact, even if the rational FB does not have PR but is *LTI* (i.e., aliasfree) with transfer function $T(z)$, (6.11),(6.12) still hold with L replaced by $LT(z)$ in (6.11). Hence, if n_0 violates compatibility as described above, the conclusion $H_0F_0 \equiv 0$ still holds. Thus the FB input–output relation is preserved even if we drop the 0-th channel, making the FB *overdecimated.* As the input of such a FB cannot be recovered from its output using any LTI system, we must have $T(z) \equiv 0$ (else we could use the LTI inverse $1/T(z)$). Thus, compatibility is necessary even if all we demand of the rational FB is that it be aliasfree and not identically zero (as opposed to having PR).

Example B: A specific set of decimators. Consider the set $\{4, 6, 6, 6, 10, 20, 20, 20\}$. This has 8 decimators with lcm $L = 60$, and $p_5 = p_6 = p_7 = 60/20 = 3$. For $m = 7, 9, 11$, the numbers $l = 3m$

are not multiples of any $p_i \neq 3$. Thus for these l, the l-th row equation in (6.11) reads as

$$\sum_{i=5}^{7} 3H_i(zW^l)F_i(z) = 0 \quad \text{for } l = 3m,\ m = 7, 9, 11 \tag{6.26}$$

The sum has three terms, and the three values of l occur consecutively in an arithmetic progression. Thus by Lemma 6.3, (6.26) holds for all l in this progression, specifically for $l = 3 \times 5 = 15$, which is not a multiple of any p_i besides $p_5 = 3$ and $p_0 = 15$. The 15-th row equation in (6.11) initially reads as $15H_0(zW^l)F_0(z) + 3\sum_{i=5}^{7} H_i(zW^l)F_i(z) = 0$ (for $l = 15$), but in the light of the above conclusion it now further says that $H_0(zW^l)F_0(z) = 0$. Now another application of Lemma 6.3 (for the trivial case of $N = 1$) shows that $H_0(z)F_0(z) = 0$ and thus PR is impossible by Lemma 6.2.

Based on the above discussion and examples, Appendix F shows a complete general algorithm to test the necessary condition derived above using Lemmas 6.2 and 6.3. This test, called the AC matrix test, implies not merely compatibility (Example A) but strong compatibility too (Appendix F), and is in fact strictly stronger than strong compatibility: The decimators of Example B are strong compatible and yet fail the test.

6.7 Conditions based on reductions to tree structures

As seen in Section 6.4.2, a decimator–set may be derivable from trees in many ways, but a FB using the decimator–set may not be derivable from all of these trees. However, all PRFBs using decimators obeying certain conditions must be derivable from certain nontrivial trees associated with the conditions. Fact 2 is a result of this type for delay–chains. Another example is as follows: If the decimator–set comes from a dyadic or 'wavelet' tree (i.e., has form $\{2, 2^2, \ldots, 2^{r-1}, 2^r, 2^r\}$ for some integer $r > 0$), then *all* PRFBs using those decimators are derivable from this tree. This was proved in [55],[19] for rational orthonormal and biorthogonal FBs respectively. It parameterizes all FBs with dyadic decimator–sets, i.e., solves problem 2 of Section 6.3.1 for such sets. However, it does not reveal any new conditions on decimators for existence of rational PRFBs (problem 1 of Section 6.3.1). This is because it concerns only dyadic decimator–sets, which, being derivable from uniform–trees, are already known to allow building of PRFBs in every FB class of interest here.

Suppose on the other hand that we have a condition on a more general decimator–set S that allows us to conclude that every rational PRFB using S is derivable from some (nontrivial) tree. Such a condition provides a parameterization result for FBs using such decimator–sets S. Further, it reduces the problem of existence of rational PRFBs using S to that of existence of rational PRFBs using the smaller decimator–sets in the units of the tree. We can obtain a new necessary condition on S for existence of such a FB, by applying all the known conditions on these smaller sets. In this section, we derive three such conditions (Theorems 6.5,6.6,6.7) all of which yield as a special case,

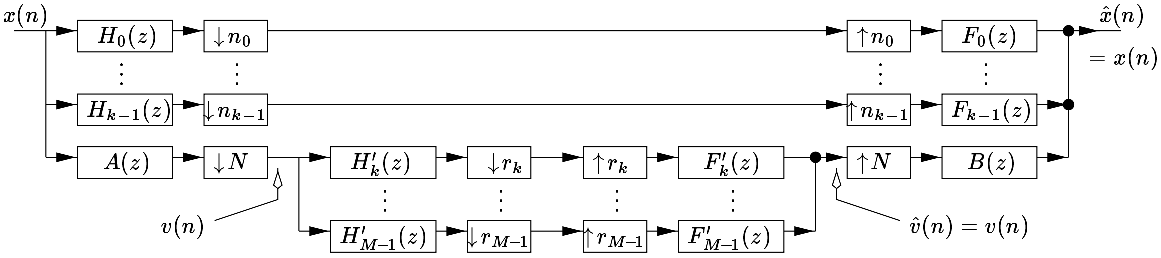


Figure 6.11: Root extraction test (Theorem 6.5): Showing equivalent tree structure for any rational PRFB with decimators obeying the premise (6.27) of the test.

the result on dyadic FBs mentioned earlier. We refer the reader to Table 1 (Section 6.8) for example decimator–sets showing the use of the new necessary conditions generated by these results. Finally, we present two other results (Theorems 6.8,6.9) that also pertain to other filter constraints besides rationality such as orthonormality, stability and the FIR property.

Theorem 6.5: Root extraction test. Given a set of decimators $n_0, n_1, \ldots, n_{M-1}$ such that

$$\sum_{i=k}^{M-1} \frac{1}{n_i} = \frac{1}{N}, \quad \text{where } N = \text{integer multiple of } n_0, n_1, \ldots, n_{k-1}. \tag{6.27}$$

Let there be a rational PRFB using these decimators. Then all n_i for $i \geq k$ are multiples of N, and the FB is always derivable from a two unit tree structure of rational PRFBs in which the root has decimators $n_0, n_1, \ldots, n_{k-1}, N$, as shown in Fig. 6.11. ■

This result is a special case of Theorem 6.6 (which is proved in Appendix G). A corollary obtained with $N = 2$ is that any rational PRFB having a decimator of value 2 must be derivable from a two unit tree of rational PRFBs in which the root is a uniform two band FB. Repeated use of this corollary shows the result of [19] on derivability of rational biorthogonal FBs with dyadic decimators from trees. The corresponding result of [55] for orthonormal FBs does not directly follow: Theorems 6.5,6.6 do not themselves show how to ensure orthonormality of all the units of the tree, given that of the overall FB. This can be done using Theorem 6.8 which follows later. Finally, note that even nonrational PRFBs with decimators n_i obeying (6.27) must be derivable from trees as in Fig. 6.11, *provided* the n_i have the property that all n_i for $i \geq k$ are multiples of N. In other words, this property no longer follows from (6.27) (as is clear from Section 6.3.2), but the derivability from a tree follows if the property is made an additional premise (Appendix G).

Theorem 6.6: Generalized root extraction test. Suppose the given decimator–set $D =$

$\{n_0, n_1, \ldots, n_{M-1}\}$ has disjoint[5] subsets S, T_1, T_2 such that S is nonempty and

$$\sum_{n_i \in S} \frac{1}{n_i} = \frac{1}{N} \text{ for some integer } N, \tag{6.28}$$

$$T_1 = \{n_i : n_i \in D, n_i = \text{factor of } N\}, \tag{6.29}$$

$$\gcd(n_i, n_j) = \text{factor of } N \quad \text{whenever } n_i \in S \cup T_2, n_j \in T_2, i \neq j, \text{ and} \tag{6.30}$$

$$\sum_{n_i \in T_1} \frac{N}{n_i} + |T_2| = N - 1 \ \ (\text{where } |T_2| = \text{number of elements in } T_2). \tag{6.31}$$

Then, if a rational PRFB exists with these decimators, all $n_i \in S$ are multiples of N, and the FB is derivable from a 2-unit tree of rational PRFBs. This tree has root decimator–set obtained from D by replacing $S \subset D$ by a single decimator of value N. The leaf–decimator set is derived from S by dividing all its elements by N. ■

Comments on Theorem 6.6. This result, proved in Appendix G, is more complicated to state but also more general than Theorem 6.5. Theorem 6.5 represents the special case when T_2 is empty and $S \cup T_1 = D$ (in which case (6.28) and (6.29) imply (6.31), due to (6.1)). Note that while one of the sets T_1, T_2 can be empty, (6.31) shows that they cannot *both* be empty except in the trivial case where $N = 1$ and $S = D$. With S, T_1 defined as in (6.28),(6.29), their disjointedness is equivalent to $|S| > 1$, which ensures that each entry of T_1 is less than all entries of S. Disjointedness of T_2 from S, T_1 is a separate requirement that does not follow from (6.28)–(6.31). Both S and T_1 can have multiple occurrences of a given decimator value; in fact from (6.29), every $n_i \in T_1$ occurs as many times in T_1 as it does in D. However, entries of T_2 are all distinct from each other, for else by (6.30), T_2 would have some elements that are factors of N and are hence in T_1 too, violating their disjointedness. Unlike Theorem 6.5, Theorem 6.6 is not obeyed by nonrational FBs even if the condition that all $n_i \in S$ be multiples of N is made an additional premise (a counterexample can be created with brickwall FBs).

Theorem 6.7: AC matrix based leaf extraction test. Consider Fig. 6.12a, showing a subset of the channels of some maximally decimated FB. Suppose the system in Fig. 6.12a is not identically zero, and all its filters are rational. Then the following statements are equivalent:

(a) Let $W = \exp(\frac{-j2\pi}{KM})$, $G_l(z) \stackrel{\triangle}{=} \sum_{i=0}^{K-1} H_i(zW^l)F_i(z)$. Then $G_l(z) = 0$ for all $l \in \{0, 1, \ldots, KM - 1\}$ that are not integer multiples of K (or equivalently, as $W^{KM} = 1$, for all such integers l).

(b) There are rational filters A, B, H_i', F_i' such that the systems of Figs. 6.12a, 6.12b are equivalent (i.e., for $i = 0, 1, \ldots, K-1$, $H_i(z) = A(z)H_i'(z^M)$, $F_i(z) = B(z)F_i'(z^M)$) and the H_i', F_i' form a K band PRFB. ■

[5]Most 'sets' in our work, including D, S, T_1 here, are really 'multisets', i.e., can contain multiple occurrences of the same decimator value. However, *disjointedness* here has its usual set–theoretic meaning. Thus, if one of the sets S, T_1, T_2 has a decimator of value n, the others can have no decimator of value n even if D has several such decimators.

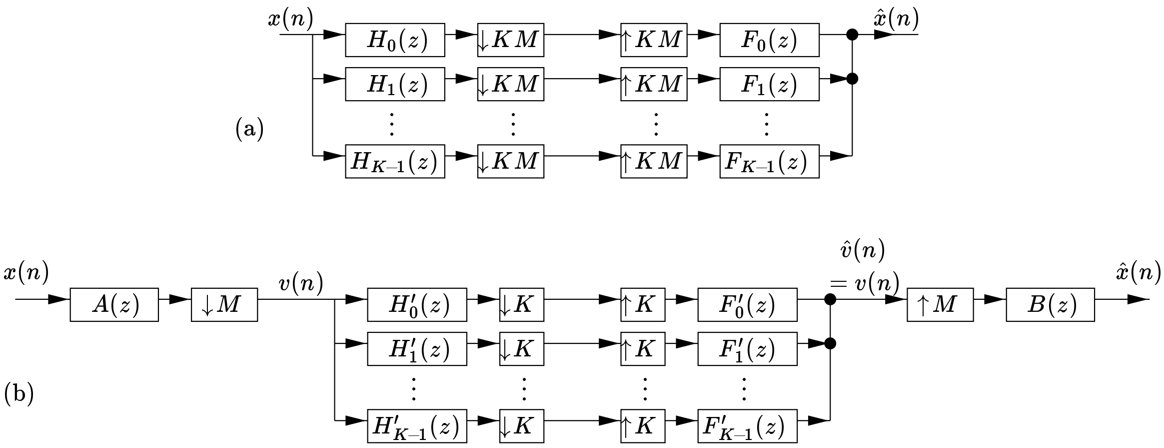


Figure 6.12: Leaf extraction test (Theorem 6.7). (a) K channels with decimator KM. (b) Equivalent structure under the premise of the test.

Application of Theorem 6.7. The result is proved in Appendix G, we mainly use the fact that (a) implies (b) in its statement. Suppose precisely K decimators of a rational PRFB have value KM. Examine the k-th row on the left side of the L–row AC matrix equation (6.11) of the FB (where L is a multiple of all decimators in the FB). For $k = l(L/(KM))$ (l a positive integer), this evaluates to the sum of $G_l(z)$ (defined in Theorem 6.7) and other terms coming from channels whose decimators n_i are such that k is a multiple of L/n_i. If there are no such terms, $G_l(z) = 0$. Even if there are such terms, we have seen in deriving the AC matrix test (Section 6.6.4) how one can sometimes deduce that they sum to zero (and hence that $G_l(z) = 0$) using filter rationality and the other rows in (6.11). Suppose the decimators are such that such a deduction of $G_l(z) = 0$ is possible for all l that are not integer multiples of K. The condition (a) of Theorem 6.7 is then obeyed by the K channels with decimator KM for all rational PRFBs with this set of decimators. Thus, Theorem 6.7 implies that all these PRFBs are derivable from a two unit tree of rational PRFBs, in which the leaf is uniform with decimator K and generates the K channels with decimator KM. Thus for such a decimator–set, existence of rational PRFBs is equivalent to existence of rational PRFBs using the smaller decimator–set in the root of the above–mentioned tree. This technique can be applied to dyadic decimator sets to deduce the result of [19]. However this result follows more easily from Theorem 6.5. Finally, note that even if we remove all rationality restrictions in the statement of Theorem 6.7, (b) still implies (a) (Appendix G). The converse (which is more useful) is however no longer true (a counterexample can be created using brickwall FBs).

Theorems 6.5—6.7 involve a decimator–set D having a subset S whose entries have reciprocals summing to $1/N$ for some integer N. Given a rational PRFB using D, the goal is to derive the subset of its channels corresponding to S from a single channel with decimator N, by attaching a leaf FB using the decimator–set S/N. (S/N is obtained from S by dividing each of its entries by N.) The theorems give various conditions on D under which this can be done for all rational PRFBs using decimator–set D. That S/N is a set of integers is either an assumption or a conclusion. Note that certain *rational* PRFBs are shown to be derivable from trees whose units are all rational PRFBs. If the original PRFB obeys a constraint other than (or besides) rationality, such as orthonormality or the FIR constraint, then can all units of the tree also be chosen to obey this constraint? A partial answer (for certain constraints) lies in the following result:

Theorem 6.8. Consider any of the following properties of FBs: (a) PR, (b) orthonormality, (c) stable filters, and (d) FIR filters. Suppose a tree structure of rational FBs yield a (necessarily rational) FB which obeys one particular property from this list. Then the filters in each unit of the tree can be modified without changing the overall FB, in such a way that each unit also becomes a FB that satisfies that property. ■

This result is proved in Appendix H. There is an important point to note about the list of properties in its statement. One could consider adding to the list, *combinations* of the listed properties, i.e., (e) PR and stable filters, (f) orthonormality and stable filters, (g) PR and FIR filters, (h) orthonormality and FIR filters. However, these *have not been listed.* Thus, for instance if the overall FB has PR with stable filters, all Theorem 6.8 assures is that the individual FBs can be altered to have either PR or stable filters — whether they can have *both* is left undecided. Indeed, it is an *open problem* as to whether or not Theorem 6.8 holds with any of the properties (e)–(h) added to its list of properties (though we believe that it probably does hold even in this case). That it holds for property (h) has been proved *for dyadic trees* in [55, Th. 2]. This proof can be extended to cover both properties (f) and (h) for all uniform–trees in which no unit has more than one child (dyadic trees being a special case). Further extensions (to arbitrary trees) are unknown.[6]

Our last result is one that, given a PRFB with decimator–set S, deduces existence of another PRFB, which has a possibly different decimator–set S_1 and preserves certain properties of the original FB such as rationality and orthonormality. Thus, given a necessary condition $\mathcal{P}$ on S for existence of a PRFB using S with such a property, we can get a stronger necessary condition $\mathcal{P}_1$ by applying $\mathcal{P}$ not merely to S but also to S_1. It may turn out that $S_1 \equiv S$, or that S_1 is derivable from a tree using S as root, in which case $\mathcal{P}_1$ is automatically tested once we test for the tree version $\mathcal{P}''$ of $\mathcal{P}$ (i.e., $\mathcal{P}''$ is even stronger than $\mathcal{P}_1$). However, this does not always happen, i.e., sometimes we indeed get a new condition. The result is as follows:

Theorem 6.9: Subset extension test. Consider any subset of channels of any PRFB. Let the decimators in this subset have lcm L and reciprocals that sum to p/L. Then there exist $L-p$ channels with decimation L which when augmented to the chosen subset, extend it into a PRFB. If the original PRFB has any *one* of the following properties: (a) rational filters, (b) orthonormality,

[6] [55, Th. 4] appears to show Theorem 6.8 with property (h) for all uniform–trees, but in fact it does not: In its proof, $T_i'(z) = T_i(z)/B(z^b)$ has not been shown to be FIR. Similarly, [19, Sec. 6] seems to account for property (e), but in fact it only covers stability (property (c)).

Test type	Test # (i)	Test	Section	Tests implied	Example decimator–set D_i violating test $\#i$
Pairwise gcd based	1	Pairwise noncoprimeness	6.6.1	—	3,4,12,12,12,12,12
	2	Pairwise gcd test	6.6.2, Th. 6.4	1	6,10,14,210,210,...,210
Generalizing compatibility	3	Compatibility	6.6.1	—	6,6,6,9,9,9,9,18
	4	Strong compatibility	6.6.1	3	6,6,6,6,9,9,27,27,45,270,...,270
	5	Tree version of #4	6.6.3, App. E	3,4	4,6,6,12,16,16,16,18,18,72,144,144,144
	6	AC matrix test	6.6.4, App. F	3,4	6,6,6,6,9,9,24,72,72,72,72,72
Tree reduction based	7	Root extraction test	6.7, Th. 6.5	—	4,4,4,10,20,20,20
	8	Generalized root extraction	6.7, Th. 6.6	7	4,4,6,8,24,24,24,24,24
	9	AC–based leaf extraction	6.7, Th. 6.7	—	4,6,6,12,12,12,20,20,30,60,60
	10	Subset extension test	6.7, Th. 6.9	—	4,6,6,8,12,12,36,48,48,72,72,72,72
Tree versions (Section 6.4.3)	11	Tree version of #2	?	1,2	3,6,6,12,18,30,180,...,180
	12	Tree version of #6	?	3,4,6	??
	13	Tree version of #7–10	?	7–10	??

Table 6.1: Necessary conditions on decimators of rational PRFBs.

(c) orthonormality and rational filters, (d) orthonormality and FIR filters, then the new 'extended' FB can also be chosen to have that property. ■

Proof: We use the equivalence between the biorthogonality condition (6.5) and the polyphase formulation of the PR condition on the filters. From Section 6.2.2, Fig. 5.2 and Appendix A, we can see that for a specific i, j, (6.5) is equivalent to $p_i \times p_j$ equations of the form $(S_i(z)Q_j(z))\downarrow_L = \delta(i-j)$. Here $L = n_i p_i = n_j p_j$, and these equations come from choosing S_i, Q_j respectively as the delayed versions of H_i, F_j in Fig. 5.2. The left side of each such equation can be written as an 'inner product' of length L vectors using the polyphase lemma (Section 6.1.3). In order to arrange all these equations into a single polyphase matrix equation, L was chosen in Section 6.2.2 as a multiple of *all* decimators n_i. However, if we restrict attention to a subset of channels of the nonuniform FB (as in Theorem 6.9), it suffices to let L be a multiple of the decimators in this subset. Thus, the subset chosen in the theorem statement corresponds to a matrix equation $\mathbf{E}(z)\mathbf{R}(z) = \mathbf{I}$ where $\mathbf{E}(z), \mathbf{R}(z)$ are of sizes $p \times L$ and $L \times p$ respectively. The theorem then follows by augmenting these matrices into $L \times L$ ones whose product is still the identity. The augmented matrices are the polyphase matrices of the new FB, and the added rows and columns are the L-th order polyphase vectors of the filters in the added channels. Clearly if $\mathbf{E}(z), \mathbf{R}(z)$ are rational, these vectors can also be chosen to be rational. If the original FB is orthonormal, $\mathbf{E}(e^{j\omega})\mathbf{E}^{*T}(e^{j\omega}) = \mathbf{I}$, so we can extend $\mathbf{E}(e^{j\omega})$ into a unitary matrix (for each ω). Further if the original FB is rational or FIR, the extension can be forced to preserve these properties by using the unitary statespace realizations [64, Chap. 14] of $L \times p$ paraunitary systems.[7] ▽ ▽ ▽

6.8 Summary and comparison of necessary conditions

Table 1 lists all currently known necessary conditions on the decimators of rational PRFBs, many of which have been developed in this chapter. The following remarks are in order:

1. For each of the tests numbered $\#i = 1, 2, \ldots, 11$, we have an example decimator–set D_i violating the test. We have chosen D_i so that the only other listed tests it fails are (a) any tests that imply test $\#i$, as shown in the second-last column of the table, and (b) *possibly* one or more of the tests #11–13, which we have not designed algorithms to perform, and hence, cannot currently decide whether or not they are violated. This shows that except for these last three tests, the interdependencies between the tests are exactly as described in the table (in its second-last column). For example, passing the AC matrix test implies nothing about passing the tree version of the strong compatibility test, and vice versa.[8]

2. The above remark applies in particular to the example set D_{11}, which passes all tests #1–10. It fails test #11 because attaching a uniform leaf with decimator 2 to its decimator of value 3 yields a set with a subset of 7 decimators within which the gcd of any pair is 6. Currently we do not have such examples for tests #12,13. Though we have not devised an algorithm to test for the tree version of the pairwise gcd test, the set D_{11} shows that the tree version is strictly stronger than the original test.

3. The AC matrix test (#6) is also strictly strengthened by forming its tree version (#12), as shown by the set $\{3, 4, 8, 12, 12, 24, 24, 24\}$. This set passes test #6, but fails its tree version because attaching uniform leaves with decimator 2 to its decimators of value 12 yields a set that fails test #6. However, this example also fails another test from Table 1, i.e., pairwise noncoprimeness.

4. Each test $\mathcal{P}$ of rows #11–13 is the tree version of some test $\mathcal{P}_1$. As seen in Section 6.4.3, the test $\mathcal{P}$ is well defined, but involves applying $\mathcal{P}_1$ to an infinite family of decimator–sets. Devising a finite algorithm for this process can take ingenuity or hard work, as seen in Section 6.6.3 (and Appendix E) for the tree version of strong compatibility. This is especially true given the complex nature of some of the $\mathcal{P}_1$ (e.g., the AC matrix test). Design of algorithms for these tests is left for future work.

5. There are decimator–sets that obey all the known necessary conditions #1–10 which we have algorithms to verify, and yet do not permit building of delay–chain PRFBs. Examples are the sets $\{4, 6, 6, 12, 12, 12, 16, 24, 48, 48, 48\}$ and $\{6, 6, 6, 6, 9, 12, 36, 36, 36, 36, 36\}$ (all examples have

[7] The extension is made by extending their 'rectangular unitary' *realization matrices* into (square) unitary ones.

[8] Example D_5 actually also fails test #10, but in a manner that makes the use of test #10 equivalent to using the tree version of another test (see discussion on Theorem 6.9).

≥ 11 decimators). Thus, these necessary conditions, taken together, are still not equivalent to the most general known sufficient condition for existence of rational PRFBs, namely, existence of PR delay–chains. We currently do not know whether or not sets of the kind listed above allow building of rational PRFBs. Thus, the main problem of this chapter (Section 6.3.1) remains unsolved for the rational FB class.

6.9 Concluding remarks

We have presented several new conditions on the decimators of rational PRFBs, considerably generalizing many earlier known ones. Our work still leaves necessary and sufficient conditions unknown. The weakest known sufficient condition, when obeyed, allows the decimators to be used to build PRFBs as specialized as delay–chains. Thus, if we impose various less restrictive conditions on the filters of a rational FB (e.g., FIR filters, linear phase filters, orthonormality, etc.), we get many more FB classes for which we do not know the necessary and sufficient conditions on the decimators of PRFBs in the class. It has been shown [21] that existence of rational PRFBs implies that of rational *orthonormal* FBs with stable filters (i.e., all analysis filters have all poles inside the unit circle). However, whether this implies existence of *FIR orthonormal* FBs is not known. Even when a decimator–set is known to allow building of rational PRFBs, complete parameterizations of the possible PRFBs are not known, except in the restricted cases of uniform and dyadic decimator–sets. Partial parameterizations using trees have been presented in Section 6.7. Other specific open problems encountered in our study are listed below:

1. *Non–tree–structured PRFBs with tree structured decimator–sets:* Section 6.4.2 has shown two different constructions leading to such FBs. Are there any more?

2. *Forcing properties of a tree structured FB on all the tree units:* Theorem 6.8 (Section 6.7) shows that this is possible for rational FBs with certain properties (e.g., PR, FIR filters), but it is not known whether it is possible for certain others (e.g., PR *and* FIR filters).

3. *Real coefficient FBs (Section 6.3.2):* Do they always exist? Does existence of rational PRFBs with decimator–set S imply that of real coefficient PRFBs (rational or otherwise) using S?

4. *Algorithms for tree versions of necessary conditions:* These have not been designed for certain necessary conditions due to their complicated nature (e.g., AC matrix test), see Table 1 (Section 6.8).

6.10 Appendices

Appendix A: Proof of Nonuniform Biorthogonality Condition (6.5)

The (uniform) biorthogonality condition on the uniform FB derived from a nonuniform one is equivalent to

$$\begin{aligned}\left(z^{-n_i c} H_i(z) z^{n_i d} F_i(z)\right) \downarrow_L &= \delta(c-d) && (6.32)\\ \left(z^{-n_i a} H_i(z) z^{n_j b} F_j(z)\right) \downarrow_L &= 0 \quad \text{if} \quad i \neq j && (6.33)\end{aligned}$$

Here $c, d, a \in \{0, 1, \ldots, p_i - 1\}$ and $b \in \{0, 1, \ldots, p_j - 1\}$, where $L = n_i p_i = n_j p_j$. We now use the noble identity $\left(X(z^M)Y(z)\right) \downarrow_M = X(z)\,(Y(z)) \downarrow_M$. This shows that (6.32) is equivalent to $\left((H_i(z)F_i(z)) \downarrow_{n_i} z^{d-c}\right) \downarrow_{p_i} = \delta(c-d)$. If $c = 0$, the left side here for $d = 0, 1, \ldots, p_i - 1$ is the d-th entry in the p_i-th order analysis polyphase vector of $(H_i(z)F_i(z)) \downarrow_{n_i}$. So the equation is clearly equivalent to

$$(H_i(z)F_i(z)) \downarrow_{n_i} = 1. \tag{6.34}$$

Next, having (6.33) hold for the said values of a, b is equivalent to having it hold for all integers a, b. This is because $L = n_i p_i = n_j p_j$, and $(A(z)) \downarrow_L = 0$ is equivalent to $\left(z^{qL} A(z)\right) \downarrow_L = 0$ for any integer q and transfer function $A(z)$ (by noble identity). As a, b take all integer values, $-n_i a + n_j b$ takes values $k \gcd(n_i, n_j)$ for all integers k. Thus, using the noble identity, with $A_{ij}(z) \triangleq (H_i(z)F_j(z)) \downarrow_{\gcd(n_i,n_j)}$, (6.33) is equivalent to

$$\left(A_{ij}(z)\ z^k\right) \downarrow_{L/\gcd(n_i,n_j)} = 0 \quad \text{for all integers } k \quad \text{if} \quad i \neq j \tag{6.35}$$

The left side here includes all entries of the order $L/\gcd(n_i, n_j)$ polyphase vector of $A_{ij}(z)$, so (6.35) is equivalent to

$$A_{ij}(z) = (H_i(z)F_j(z)) \downarrow_{\gcd(n_i,n_j)} = 0 \quad \text{if} \quad i \neq j. \tag{6.36}$$

Thus (6.32),(6.33) are equivalent to (6.34),(6.36) respectively; proving the nonuniform biorthogonality equation (6.5).

Appendix B: Derivability of Decimator–sets from a Uniform-tree

Claim: If a set S of decimators satisfying (6.1) is derivable from a uniform–tree, then it is derivable from a uniform–tree in which the root has decimator g, where g is the gcd of all elements of S.

Proof: (Can be skipped without losing continuity.) Use induction on the number N of units in the tree. Let r be the root decimator in the given uniform–tree. Clearly r divides all decimators in S, so r divides g. Now the root can have at most r children. If it has less than r children, then

r is a decimator in S. Hence $g <= r$, implying $g = r$, i.e., the root already has decimator g. This proves the claim for $N = 2$, as the root of a 2-unit tree has 1 child and $1 < r$. If the root has all r children, consider any child along with all its descendants. These units form another uniform–tree. All decimators generated by this new tree are multiples of g/r. Let g' be their gcd, thus $g' = k(g/r)$ for some integer k. The new tree has $\leq N - 1$ units. So by the induction hypothesis, the decimators it generates can be rederived from a uniform–tree with root decimator g'. Since $g' = k(g/r)$, this root unit can then be rederived from a uniform–tree having decimator g/r for root and k for all the g/r leaves, each of which is a child of the root. After making all these replacements on the starting tree, all children of its root now have decimator g/r, hence the root and its children can be replaced by a single uniform unit with decimator g. This proves the claim. $\bigtriangledown\bigtriangledown\bigtriangledown$

The above result suggests an algorithm [39] that tries to build a uniform–tree starting from its root:

Root–to–leaves Algorithm. (Tests derivability of a given set S of decimators from a uniform–tree)

1. Find gcd g of all elements of S. If $g = 1$, then S is not derivable from a uniform–tree.
2. Divide all entries of S by g (represents choosing root decimator g). Find *all* possible partitions of the resulting set into g groups each of which is a valid set of decimators obeying maximal decimation (6.1).
3. We can derive S from a uniform–tree if and only if among these partitions, there is at least one in which each group is derivable from a uniform–tree.

The algorithm is recursive. At Step 2, dividing the entries of S by g yields a set S' lower–bounded by unity. In the ensuing partition of S', any unity element in S' is all by itself a valid group viewed as derivable from a uniform–tree for purposes of Step 3. Such a group denotes absence of a child of the root, just as groups with more than one element represent children of the root. There may possibly be *no* valid partition at Step 2, e.g., when $S = \{4, 6, 6, 10, 10, 10, 10, 120, 120\}$. This of course means that there is no uniform–tree.

Note that though Step 2 can always be implemented in principle, doing it with a simple and efficient algorithm can be tricky. An alternative method builds the tree starting from a leaf and avoids this problem. Its basic idea is in identifying a leaf: Given an arbitrary decimator d in a set S derivable from a uniform–tree, it is not clear whether d is obtained at a leaf unit of the tree. However, this *must* be the case if d is the *maximum* element in S, and further the leaf decimator must of course then divide d. Based on this, we have:

Leaf–to–root Algorithm. (Tests derivability of a given set S of decimators from a uniform–tree)

1. If S has no more than two *distinct* decimators, it is derivable from a tree.

2. Find the maximum entry m in S, and the number of times N that it occurs. For each factor k of m such that $1 < k \leq N$, form a smaller set S_k by setting $S_k = S$ and then replacing k of the elements of value m in S_k by one element of value m/k (i.e., form a leaf unit that is uniform with decimator k).

3. The set S is derivable from a uniform–tree if and only if at least one of the S_k above is.

This is another recursive algorithm, more elegant and simpler to implement, though it may be somewhat unclear whether or not it is faster. Its only step still requiring justification is Step 1. This is easily done: Suppose $S = \{m_0, \ldots, m_0, m_1, \ldots, m_1\}$ with m_i occurring N_i times $(i = 0, 1)$. Let $m_i = gd_i$ where $g = \gcd(m_0, m_1)$. Maximal decimation implies that $N_0 d_1 + N_1 d_0 = g d_0 d_1$. As d_0, d_1 are coprime, this means that $N_i = D_i d_i$ for integers D_i, $i = 0, 1$, where $D_0 + D_1 = g$. Thus S is derivable from a uniform–tree in which the root has decimator g, all its children are leaves, and D_i leaves have decimator d_i $(i = 0, 1)$.

Only necessary and only sufficient conditions. Presence of no more than two distinct decimators, as shown above, is an example of a *sufficient* condition for derivability from a uniform–tree. It is by no means necessary. Another such example is the condition that each decimator divides every decimator larger than itself (a special case is when all of them are powers of the same number). This condition neither implies nor is implied by the earlier one, and neither condition is necessary, as exemplified by the set $\{4, 4, 6, 6, 12, 12\}$. Sufficiency of the new condition is proved using the root–to–leaves algorithm: Clearly $g > 1$ at Step 1, as g is the smallest decimator. At Step 2 in formation of the partition, if we sequentially select elements from the smallest upwards, the condition ensures that at some stage the reciprocals of the selected elements will sum to unity. Repeating this process results in a valid partition, and further each of its groups also satisfies the condition. Thus the proof is completed by induction on the number of decimators.

Derivability of a set of decimators from uniform–trees implies existence of various types of PRFBs (including PR delay–chains) using those decimators. Thus, any conditions necessary for such existence are also necessary for derivability from uniform–trees. Their necessity is often provable directly from the above algorithms. For example, without pairwise noncoprimeness (Section 6.6.1), $g = 1$ at Step 1 of the root–to–leaves algorithm. If compatibility (Section 6.6.1) is violated, i.e., if a decimator d does not divide any other decimator, then eventually $m = d$ and $N = 1$ at Step 2 of the leaf–to–root algorithm, i.e., there are no sets S_k. As tests for such necessary conditions are inconclusive whenever they are satisfied, they cannot replace the earlier complete algorithms, though they can potentially increase their efficiency.

Appendix C: Proof of Fact 3

Let $R = \{m_0, m_1, \ldots, m_{M-1}\}$, $S = \{k_0, k_1, \ldots, k_{K-1}\}$. So $T = \{n_0, \ldots, n_{K-1}, n_K, n_{K+1}, \ldots, n_{K+M-2}\}$ with $n_i = m_0 k_i$ for $i = 0, 1, \ldots, K-1$ and $n_{K-1+i} = m_i$ for $i = 1, 2, \ldots, M-1$. Let $L = \text{lcm}\{n_i\}$, $p_i = L/n_i$. Let n_i occur N_i times in T. Let b_i be the strong compatibility lower bound on N_i. The proof is in two parts:

Part 1: Uniformity of S. Suppose S is not a uniform unit, we will then show that $b_j \leq N_j$ for all j, i.e., T is strong compatible. Indeed for $j = 0, 1, \ldots, K-1$ we have from (6.20),

$$\begin{aligned} p_j b_j = \min_{p_i \neq p_j} \text{lcm}(p_i, p_j) &\leq \min_{p_i \neq p_j,\, 0 \leq i < K} \text{lcm}(p_i, p_j) \quad \text{(as } S \text{ is not uniform)} && (6.37) \\ &\leq p_j N_j \qquad \text{(as } S \text{ is strong compatible)} && (6.38) \end{aligned}$$

The minimization on the right side of (6.37) is not over an empty set because S is nonuniform, i.e., $p_i \neq p_j$ for at least one i such that $0 \leq i < K$. The right side of (6.37) thus equals $p_j b_j^S$ where b_j^S is the strong compatibility (lower) bound on the number N_j^S of occurrences of k_j in S. This bound holds by strong compatibility of S, and $N_j^S \leq N_j$. This justifies (6.38), and thus $b_j \leq N_j$ for $0 \leq j < K$. For $j \geq K$, if $n_j = m_0$ then $b_j = 1 \leq N_j$, because n_j divides a distinct decimator $n_0 = n_j k_0$. If $n_j \neq m_0$, then $N_j \geq N_j^R$, the number of occurrences of $n_j = m_{j-(K-1)}$ in R. Let b_j^R be the strong compatibility lower bound on N_j^R. Thus $b_j^R \leq N_j^R \leq N_j$, and

$$p_j b_j^R = \min(A, B) \quad \text{where} \quad A = \min_{p_{K-1+i} \neq p_j,\, i \geq 1} \text{lcm}(p_{K-1+i}, p_j), \quad \text{and} \quad B = \text{lcm}(L/m_0, p_j), \quad \text{while}$$

$$p_j b_j = \min_{p_i \neq p_j} \text{lcm}(p_i, p_j)$$

Thus if $A \leq B$ (e.g., this holds if $m_i = m_0$ for some $i > 0$), then $b_j \leq b_j^R \leq N_j$. Even if $A > B$,

$$p_j b_j^R = \text{lcm}(L/m_0, p_j) \geq \min_{p_i \neq p_j,\, 0 \leq i < K} \text{lcm}(p_i, p_j) \geq p_j b_j, \tag{6.39}$$

as $p_i = L/(m_0 k_i)$ for $i < K$, and nonuniformity of S again ensures that $\text{lcm}(p_i, p_j)$ is not being minimized over an empty set. (Nonuniformity of S is not needed here if $L/(m_0 K) \neq p_j$.) So again $b_j \leq b_j^R \leq N_j$. Thus, $b_j \leq N_j$ for all j, i.e., T is strong compatible, contradicting the premise of Fact 3. Hence S must be a uniform unit, i.e., $k_0 = k_1 = \ldots = k_{K-1} = K$. $\bigtriangledown\bigtriangledown\bigtriangledown$

Part 2: Necessary conditions for $b_j > N_j$. We have already shown in Part 1 that if $j \geq K$, then $b_j > N_j$ is possible only if m_0 occurs only once in R and $m_0 K = n_j$. The proof of Fact 3 will be completed if we show a similar statement for $j < K$, i.e., that $b_j > N_j$ is possible only if m_0 occurs only once in R and $m_0 K = n_i$ for some $i \geq K$. To show this, note that for all $j < K$, all the n_j are identical (shown by Part 1), and hence the same holds for the N_j and the b_j. Also $N_j \geq K$. Thus it suffices to show that $b_0 \leq K$ if either $m_0 = m_l = n_{K-1+l}$ for some $l > 0$, or $m_0 K \neq m_i$ for all $i > 0$. If $m_0 = m_l = n_{K-1+l}$ for some $l > 0$, then

$$p_0 b_0 = \min_{p_i \neq p_0} \operatorname{lcm}(p_i, p_0) \leq \operatorname{lcm}(p_{K-1+l}, p_0) = \operatorname{lcm}\left(\frac{L}{m_0}, \frac{L}{m_0 K}\right) = \frac{L}{m_0} = p_0 K,$$

hence $b_0 \leq K$. If on the other hand m_0 occurs exactly once in R, then $m_0 F = m_l = n_{K-1+l}$ for some $F > 1$, $l > 0$ since R is compatible. Thus if $m_0 K \neq m_i$ for all $i > 0$, then

$$p_0 b_0 = \min_{p_i \neq p_0} \operatorname{lcm}(p_i, p_0) \leq \operatorname{lcm}(p_{K-1+l}, p_0) = \operatorname{lcm}\left(\frac{L}{m_0 F}, \frac{L}{m_0 K}\right) \leq \frac{L}{m_0} = p_0 K,$$

hence $b_0 \leq K$ again. This establishes the claim, hence proving Fact 3. $\bigtriangledown \bigtriangledown \bigtriangledown$

Appendix D: Proof of Fact 4

From the premise of Fact 4, there is a tree $\mathcal{T}'$ in which each unit is either D or allows building of rational FBs (e.g., uniform units), such that $\mathcal{T}'$ generates a set of decimators that is not strong compatible. Note that every unit in $\mathcal{T}'$ is strong compatible. We now perform a series of operations on $\mathcal{T}'$, each yielding a new tree with all the properties of $\mathcal{T}'$, until finally we get the tree $\mathcal{T}$ with the desired properties as in Fact 4.

If the root of $\mathcal{T}'$ has a child that is not a leaf, then this child, along with all its descendants, forms a tree with fewer units than $\mathcal{T}'$. We can assume that this tree generates a strong compatible decimator–set (else we can replace $\mathcal{T}'$ by this tree and repeat the process). We then view this tree as a *single* unit. This makes every child of the root of $\mathcal{T}'$ a strong compatible leaf. Next, we delete any leaf such that the residual tree generates a decimator–set that is not strong compatible. This yields the desired tree $\mathcal{T}$ having all properties of $\mathcal{T}'$. We now show that $\mathcal{T}$ and the decimator–set T it generates have all the properties listed in Fact 4.

Properties 1,2,4: For any leaf S of $\mathcal{T}$, we see that $\mathcal{T}$ can be redrawn as a 2-unit tree with strong compatible units R and S. However $\mathcal{T}$ itself generates the set T that is not strong compatible. Thus we can use Fact 3 to conclude the following: (a) All leaves of $\mathcal{T}$ are uniform. (b) For any decimator value obtained at a leaf of $\mathcal{T}$, decimators of T with that value are the only ones in T that violate the strong compatibility lower bound on the number of their occurrences in T. (c) Property 2 of Fact 4 holds. Now (b) implies that all decimators obtained at the leaves have the same value d. Also, (a) implies that $\mathcal{T}$ has root D: Otherwise the root allows building of rational PRFBs, and hence, so does $\mathcal{T}$ (as all children of its root are uniform leaves); violating the fact that T is not strong compatible. This completes the proof of property 1. Property 4 follows from this and conclusion (b) listed above. Thus we have shown properties 1,2,4 of Fact 4. $\bigtriangledown \bigtriangledown \bigtriangledown$

Property 3: Let k_i be the decimator value of the leaf attached to $d_i \in D$ to form $\mathcal{T}$. As $d_i k_i = d$, we have $d = C\,\mathrm{lcm}\{d_i\}$ where $C = \gcd\{k_i\}$. We must show that if $d \notin D$, then $C = 1$. In fact, this may be *false*. Our approach is to assume that $d \notin D$, and then create a new tree $\mathcal{T}^*$ generating a decimator–set T^* with all the properties of Fact 4. This is done by replacing every leaf decimator k_i with k_i/C. (If $k_i = C$ this means deleting the leaf.) Clearly property 1 of Fact 4 continues to hold, with the decimators obtained at the leaves now having value $d^* = d/C = \mathrm{lcm}\{d_i\}$. To prove property 2, let decimator d_i of D have a leaf attached to it in $\mathcal{T}^*$. Then it also has a leaf (uniform with decimator k_i) attached in $\mathcal{T}$. As $d_i \notin T$ (by property 2 for $\mathcal{T}$), the only way to have $d_i \in T^*$ is that d_i be the newly formed decimator d/C. This however means that $k_i = C$ (as $d = d_i k_i$), i.e., the leaf attached to d_i in $\mathcal{T}$ has been deleted in $\mathcal{T}^*$, contradicting the assumption on d_i. Thus $d_i \notin T^*$, i.e., $\mathcal{T}^*$ obeys property 2. Next we prove property 3. As already seen, if $k_j = C$ for some j, then $d^* = d/C = d_j \in D$. Thus, if $d^* \notin D$, then $k_j > C$ for all j, i.e., decimators d_i with leaves attached in $\mathcal{T}$ are the same as those with leaves attached in $\mathcal{T}^*$. So property 3 holds for $\mathcal{T}^*$ from $d^* = d/C = \mathrm{lcm}\{d_i\}$. Lastly, we show property 4, i.e., that d^* violates the strong compatibility lower bound b^* on the number N^* of its occurrences in T^*. Let N be the number of occurrences of d in T, and let b be the strong compatibility lower bound on N. Let L be any common multiple of the decimators of T. We must show that $b^* > N^*$. Since $\mathcal{T}$ obeys property 4, we have $b > N$. Also, by construction of $\mathcal{T}^*$ and the hypothesis $d \notin D$, we have $N^* \geq N/C$. The inequality is strict only if $d/C \in T$, but this would imply (by definition (6.20) of b) that $b \leq \left(\frac{d}{L}\right)\mathrm{lcm}(\frac{L}{d}, \frac{LC}{d}) = C$. Since $N \geq k_i \geq C$, we get $b \leq N$, a contradiction. Thus $d/C \notin T$, and hence $N^* = N/C$. Lastly, $b^* = \left(\frac{d}{LC}\right)\mathrm{lcm}(\frac{LC}{d}, \frac{L}{m})$ for some $m \in T^*$, $m \neq d/C$. Thus $m \in D$ and $m \in T$ too, and $m \neq d$ by the hypothesis $d \notin D$. So $b \leq \left(\frac{d}{L}\right)\mathrm{lcm}(\frac{L}{d}, \frac{L}{m}) \leq C\left(\frac{d}{LC}\right)\mathrm{lcm}(\frac{LC}{d}, \frac{L}{m}) = Cb^*$. Hence, $b^* \geq b/C > N/C = N^*$ (using $b > N$). Thus $b^* > N^*$ as required. $\triangledown\triangledown\triangledown$

Appendix E: Testing Tree Version of Strong Compatibility.

Given a decimator–set D, let $V = \{v_0, v_1, \ldots, v_{K-1}\}$ be the set of *distinct* decimator values in D, with v_i occurring N_i times in D. Let L be any multiple of all the v_i, i.e., of $\mathrm{lcm}\{v_i\}$, and let $p_i = L/v_i$. Then D satisfies the tree version of strong compatibility if and only if Routine 1 below returns the value 'TRUE' for all $v_i \in V$ *and* Routine 2 returns value 'TRUE'.

Routine 1: (To be performed for all $v_i \in V$)

1. Initialization: Set $M = N_i$, $A = V$ and delete v_i from A.

2. If A is empty, return(TRUE). Else, let $j = l$ minimize $\mathrm{lcm}(p_i, p_j)$ over all j such that $v_j \in A$. If $M < \mathrm{lcm}(p_i, p_l)/p_i$, return(FALSE).

3. If v_l does not divide v_i, return(TRUE). Else, add $N_l(v_i/v_l)$ to M and delete v_l from A. This represents attaching to every decimator of value v_l, a leaf that is uniform with decimator v_i/v_l.

Then go to Step 2.

Routine 2:

1. Find all subsets S of V having at least two but less than $K-1$ elements, such that the lcm $l(S)$ of all elements of S does not divide any $v_j \in V$.

2. For each S of Step 1, let $\sigma(S)$ be the sum of all the numbers $N_i(l(S)/v_i)$ for all $v_i \in S$. Let $b(S)$ be the minimum of $\left(\frac{l(S)}{L}\right) \mathrm{lcm}(\frac{L}{l(S)}, \frac{L}{v_i})$ over all $v_i \notin S$. This step represents attaching to every decimator whose value v_i lies in S, a leaf unit that is uniform with decimator $l(S)/v_i$, so that all decimators thus obtained at the leaves have value $l(S)$. In the resulting tree structured set of decimators, $\sigma(S)$ is the number of occurrences of decimator $l(S)$ and $b(S)$ is the strong compatibility lower bound on $\sigma(S)$.

3. If $\sigma(S) \geq b(S)$ *for all* S above, return(TRUE). Else return(FALSE).

The action of the routines is independent of which multiple of lcm$\{v_i\}$ we choose L to be. To explain how the above test works, refer to the statement of Fact 4. Routine 2 lists all trees $\mathcal{T}$ obeying properties 1,2,3 of Fact 4 such that $d \notin D$ (see property 3), and returns a 'FALSE' value if any of these obey property 4. The set S of Step 1 represents choice of the d_i of property 2. We demand that S must have at least two elements, and that $l(S) \neq v_j$ for all $v_j \in V$, to ensure that property 3 holds with $d \notin D$. In fact we further demand that $l(S)$ must not divide any $v_j \in V$, for if it does, $b(S) = 1$ at Step 2. We also exclude sets S with $\geq K-1$ elements, for then $\mathcal{T}$ generates a set with at most two *distinct* decimators. Such a set, being derivable from a uniform–tree (Appendix B), is always strong compatible, i.e., $\sigma(S) \geq b(S)$ will hold at Step 3.

Routine 1 becomes a test for strong compatibility if we delete Step 3 in it. Hence we can assume strong compatibility of the given set of decimators. Thus the only task remaining is to examine whether there is a tree $\mathcal{T}$ obeying all properties of Fact 4 with $d \in D$ in property 3. This is achieved by the addition of Step 3. To see this, let there be such a tree $\mathcal{T}$, with $d = v_i$, producing a set T of decimators. The quantity $b = \mathrm{lcm}(p_i, p_l)/p_i$ of Step 2 is the lower bound on N_i, which holds by assumption of strong compatibility. Now the number N_T of occurrences of v_i in T is at least N_i. Further if $v_l \in T$, then the strong compatibility lower bound on N_T does not exceed b, and hence cannot be violated. Thus $v_l \notin T$, i.e., all decimators of value v_l must have leaves attached to them to convert them into decimators of value v_i. This justifies Step 3.

In the special case when $L \triangleq \mathrm{lcm}\{v_j\} \in V$, Routine 2 can be skipped (it always returns 'TRUE'), and Routine 1 needs execution only for $v_j = L$ (it returns 'TRUE' for all other v_j). This is provable from the fact that for $v_j = L$, $p_j = 1$. In general, Routine 1 appears to be the important part of the test: There are relatively fewer decimator–sets for which violation of the test is detected by Routine 2 but not by Routine 1 (examples of such sets being $\{2,3,24,24,36,36,36\}$ and $\{2,4,6,48,48,72,72,72\}$).

Appendix F: Algorithm for the AC Matrix Test

In the given set of decimators, let $v_0, v_1, \ldots, v_{K-1}$ be the *distinct* decimator values, with v_j occurring N_j times. Let L be any common multiple of the v_j, and let $p_j = L/v_j$. The algorithm is as follows:

1. *Initialization.* Create a matrix $\mathbf{U}$ with rows numbered 0 to $L-1$ and columns 0 to $K-1$, where the lj-th entry u_{lj} is 1 if l is a multiple of p_j, and zero otherwise. Thus $\mathbf{U}$ is initialized to describe the positions of the zero and nonzero entries in the AC matrix (6.11),(6.12). In particular, $u_{0j} = 1$ for all j.

2. Set $\mathbf{U}' = \mathbf{U}$ (saving the current value of $\mathbf{U}$ in $\mathbf{U}'$). For all l, j such that u_{lj} is the only entry in the l-th row having value unity, set $u_{lj} = 2$. This identifies sets of filters having the same decimator value v_j, and satisfying an equation of the form $\sum_i B_i(zW^l)C_i(z) = 0$.

3. For each $d = kp_j$ for integer k satisfying $1 \leq kp_j \leq \lfloor L/2 \rfloor$, let $c_s^d(n) = s + nd$ for $s = 0, p_j, 2p_j, \ldots, d - p_j$. If $u_{lj} = 2$ for $l \equiv c_s^d(n) \pmod{L}$ for N_j consecutive integers n, set $u_{lj} = 2$ for $l \equiv c_s^d(n) \pmod{L}$ for *all* integers n. Do this for each $j = 0, 1, \ldots, K-1$. (This corresponds to use of Lemma 6.3.)

4. If $u_{0j} = 2$ for any j, the given set of decimators fails the AC matrix test. (This is where we use Lemma 6.2.) If $\mathbf{U}' = \mathbf{U}$, the set passes the test. If neither of these happens, go to Step 2.

Passing the above test is a necessary condition on the decimators of any rational PRFB, as the discussion of Section 6.6.4 proves. The test outcome is independent of which common multiple of the v_j we choose L to be. The above algorithm may be made more efficient in many ways (e.g., we can declare the test as passed if $\mathbf{U}' = \mathbf{U}$ after Step 2); our main purpose here is to state a correct (rather than highly efficient) algorithm.

Lastly, we prove that the above test implies strong compatibility. Consider any fixed $j \in \{0, 1, \ldots, K-1\}$, and find the smallest $l > 0$ such that u_{lj} is not set to value 2 at Step 2. This is the smallest nonzero multiple of p_j that is also a multiple of some $p_i \neq p_j$, i.e., it is $\min_{p_i \neq p_j} \operatorname{lcm}(p_i, p_j) = p_j b_j$ where b_j is as in (6.20). Thus, after Step 2, $u_{lj} = 2$ for $l = kp_j$ for $k = 1, 2, \ldots, b_j - 1$. So if $N_j < b_j$, Step 3 will use the sequence $c_0^{p_j}(n)$ to set $u_{lj} = 2$ for all $l = np_j$. In particular it sets $u_{0j} = 2$, which means that the test is failed (see Step 4). Hence if the test is passed, we have $N_j \geq b_j$ for all j, which is the strong compatibility condition (6.20).

Appendix G: Proofs of Theorems 6.6,6.7

Proof of Theorem 6.6: We will prove the claim of the theorem after replacing its premises (6.28)–(6.31) about the decimator–set D by the following premise: The set D has two nonempty disjoint subsets S, T such that

$$\sum_{n_i \in S} \frac{1}{n_i} = \frac{1}{N} \text{ for some integer } N, \tag{6.40}$$

$$|T| = N-1, \text{ and} \tag{6.41}$$

$$\gcd(n_i, n_j) = \text{factor of } N \quad \text{whenever} \quad n_i \in S \cup T,\, n_j \in T,\, i \neq j \tag{6.42}$$

This suffices because from a rational PRFB obeying (6.28)–(6.31), we can create one obeying (6.40)–(6.42) by inserting in each of its channels with decimator $n_i \in T_1$, a uniform rational PRFB with decimator N/n_i. This process preserves the channels corresponding to the decimator subset S, and creates $\left(\sum_{n_i \in T_2} (\frac{N}{n_i})\right)$ new decimators each of value N. The set T consists of T_2 and these new decimators; thus (6.41) follows from (6.31), and (6.42) from (6.30) and the fact that the new decimators have value N. Having proved the claim using (6.40)–(6.42), we remove the inserted uniform leaf FBs to prove it under the original premise (6.28)–(6.31).

Part 1: Proof under additional assumption that all $n_i \in S$ are multiples of N. Let us be given a rational PRFB with decimator–set D and filters as in Fig. 1.8, such that D has disjoint subsets S, T obeying (6.40)–(6.42). Let $\mathbf{E}(z), \mathbf{R}(z)$ respectively be the N-th order analysis and synthesis polyphase matrices of the analysis and synthesis filters corresponding to channels with decimators $n_i \in T$. Let $\mathbf{e}_i(z)$ be the N-th order analysis polyphase vector of $H_i(z)$ where $n_i \in S$. From (6.41), $\mathbf{E}(z), \mathbf{R}(z)$ have sizes $(N-1) \times N$ and $N \times (N-1)$ respectively. We use (6.42) with the PR condition (6.5) and the polyphase lemma, as in Section 6.6.2. This shows that $\mathbf{e}_i(z)\mathbf{R}(z) = \mathbf{0}$, and that $\mathbf{E}(z)\mathbf{R}(z)$ is a $(N-1) \times (N-1)$ diagonal matrix, none of whose diagonal entries is identically zero. This implies (using rationality of the filters) that $\mathbf{R}(z)$ has $N-1$ linearly independent columns. All the $\mathbf{e}_i(z)$, being 'orthogonal' to all these columns, must be 'proportional', i.e., $\mathbf{e}_i(z) = H_i'(z)\mathbf{a}(z)$ for some rational filters $H_i'(z)$ and vector $\mathbf{a}(z)$. Let $A(z)$ be the filter with $\mathbf{a}(z)$ as its N-th order analysis polyphase vector. Computing $H_i(z)$ from $\mathbf{e}_i(z)$ shows that $H_i(z) = A(z)H_i'(z^N)$. A similar argument shows that for all i such that $n_i \in S$, $F_i(z) = B(z)F_i'(z^N)$ for some rational $B(z), F_i'(z)$. Thus, under the additional assumption that all decimators in S are multiples of N, we see that the given rational PRFB is derivable from a two unit tree of rational FBs. The units of the tree have decimator–sets exactly as desired, and using Theorem 6.8, their filters can further be modified so that they also have PR. This completes Part 1 of the proof.

Part 2: Extending Part 1 to nonrational FBs in the setting of Theorem 6.5. When the original premises (6.28)–(6.31) of Theorem 6.6 are obeyed in the special manner that results in the premise of Theorem 6.5, the effect on (6.40)–(6.42) is to cause $D = S \cup T$ and $n_j = N$ for all $n_j \in T$. Now in Part 1, the diagonal elements of $\mathbf{E}(z)\mathbf{R}(z)$ are $(H_j(z)F_j(z)) \downarrow_N$ where $n_j \in T$ (by polyphase lemma). Thus, in the above special case, by (6.5), in fact $\mathbf{E}(z)\mathbf{R}(z)$ is the identity. Hence we can choose the $A(z), B(z)$ of Part 1 to have N-th order analysis and synthesis polyphase vectors $\mathbf{a}(z), \mathbf{b}(z)$ respectively, such that the $N \times N$ matrices $\begin{bmatrix} \mathbf{E}(z) \\ \mathbf{a}(z) \end{bmatrix}$ and $\begin{bmatrix} \mathbf{R}(z) & \mathbf{b}(z) \end{bmatrix}$ have product equal to identity. This possible even without any rationality restriction on the filters (of course A, B are then nonrational in general). These matrices now become the polyphase matrices of the root FB. Thus, the root automatically has PR, and hence so does the leaf (since the overall FB has PR), without the need to use Theorem 6.8 (which requires filter rationality). Thus, for the special case of Theorem 6.5 (as distinct from the general setting of Theorem 6.6), we have extended Part 1 to nonrational FBs.

Part 3: Proving the additional premise used in Part 1, using filter rationality. For each i such that $n_i \in S$ we insert a q_i channel uniform rational PRFB within the i-th channel of the given PRFB, where $q_i = \mathrm{lcm}(N, n_i)/n_i$. This forms q_i new decimators of value $n_i q_i$. Let S' be the set of these decimators. Then, the newly formed tree–structured rational PRFB also has a decimator–set satisfying the premises (6.40)–(6.42), with S replaced by S' and T unchanged. Indeed, (6.40),(6.41) obviously hold, while (6.42) follows from the observation that if $\gcd(n_i, n_j)$ is a factor of N and q_i contains precisely the factors of N that are not present in n_i (i.e., $q_i = \mathrm{lcm}(N, n_i)/n_i$) then $\gcd(n_i q_i, n_j)$ is also a factor of N. Further S' also obeys the additional assumption that its elements are multiples of N, by the choice of the q_i. Let $q_i > 1$ and consider two analysis filters $C_i^l(z)$, $l = 0, 1$ of the q_i band leaf FB inserted in the channel with decimator $n_i \in S$. The corresponding analysis filters of the new tree–structured FB are $H_i(z)C_i^l(z^{n_i})$. However, using Theorem 6.6 (which Part 1 has proved for the new FB), these filters have the form $A(z)D_i^l(z^N)$ for some rational $D_i^l(z), A(z)$ where $A(z)$ is independent of l, i. Taking ratios of these filters (a crucial step that requires filter rationality) shows that $\frac{C_i^0(z^{n_i})}{C_i^1(z^{n_i})} = \frac{D_i^0(z^N)}{D_i^1(z^N)}$, which implies that each equals $X_i(z^{\mathrm{lcm}(N,n_i)})$ for some rational $X_i(z)$. Replacing z by z^{1/n_i} and using the definition of q_i, we have $\frac{C_i^0(z)}{C_i^1(z)} = X_i(z^{q_i})$. This means that the q_i-th order analysis polyphase vectors $\mathbf{e}_i^l(z)$ of $C_i^l(z)$, $l = 0, 1$, are linearly dependent, as $\mathbf{e}_i^0(z) = \mathbf{e}_i^1(z)X_i(z)$. Thus, the inserted q_i band uniform leaf FB with the filters $C_i^l(z)$, while assumed to have PR, has an analysis polyphase matrix that is not invertible (since it contains the rows $\mathbf{e}_i^l(z), l = 0, 1$). This contradiction disproves the assumption that $q_i > 1$. Hence $q_i = 1$, or in other words, n_i is a multiple of N. $\bigtriangledown\bigtriangledown\bigtriangledown$

Proof of Theorem 6.7: We first write the input–output relations, analogous to (6.4), of the systems of Fig. 6.12:

$$\hat{X}(z) = \frac{1}{KM}\sum_{l=0}^{KM-1} X(zW^l)G_l(z) \qquad \text{for Fig. 6.12a} \qquad (6.43)$$

$$\hat{X}(z) = \frac{1}{M}\sum_{l=0}^{M-1} A(zW^{Kl})B(z)X(zW^{Kl}) \qquad \text{for Fig. 6.12b} \qquad (6.44)$$

Here G_l are as defined in statement (a) of Theorem 6.7, and (6.44) uses the PR property of the FB formed by the H'_i, F'_i. Comparing (6.43) and (6.44) directly shows that (b) implies (a) in Theorem 6.7, whether or not the filters are rational. We now prove that (a) implies (b) (for which the filters must be rational). Form the M-th order AC matrix $\mathbf{H}(z)$ (of size $M \times K$) using analysis filters $H_i(z)$, i.e., let the q-th row of $\mathbf{H}(z)$ be $(H_0(zW^{Kq}), H_1(zW^{Kq}), \ldots, H_{K-1}(zW^{Kq}))$ for $q = 0, 1, \ldots, M-1$. Thus, the condition (a) is equivalent to

$$\mathbf{H}(zW^l)\mathbf{f}(z) = \mathbf{0} \quad \text{for} \quad l = 1, 2, \ldots, K-1, \quad \text{where} \quad \mathbf{f}(z) = (F_0(z), F_1(z), \ldots, F_{K-1}(z))^T$$

Replacing z by zW^{-l}, $\mathbf{H}(z)\mathbf{f}(zW^{-l}) = \mathbf{0}$. Now the columns $\mathbf{f}(zW^{-l})$, $l = 1, 2, \ldots, K-1$ are linearly independent. For otherwise, there are rational filters $\alpha_l(z)$ such that $\sum_{l=j}^{K-1} \alpha_l(z)\mathbf{f}(zW^{-l}) \equiv 0$, where $1 \leq j < K$ and $\alpha_j(z) \not\equiv 0$. Dividing this by $\alpha_j(z)$ and replacing z with zW^j shows that $\mathbf{H}(z)\mathbf{f}(z) = \mathbf{0}$ too. This would mean that $G_l(z) = 0$ *for all* integers l, i.e., by (6.43), that the system of Fig. 6.12a is identically zero, which contradicts the premise of the theorem. Thus, the $K-1$ columns $\mathbf{f}(zW^{-l})$, $l = 1, 2, \ldots, K-1$ are linearly independent, and each row of $\mathbf{H}(z)$ is 'orthogonal' to all these columns (i.e., their product is identically zero). Hence all these rows must be 'proportional' to each other, i.e., $\mathbf{h}_1(z) = C(z)\mathbf{h}_0(z)$ for some scalar filter $C(z)$, where $\mathbf{h}_q(z)$ is the q-th row of $\mathbf{H}(z)$. This implies that for all $i = 0, 1, \ldots, K-1$,

$$\frac{H_i(zW^K)}{H_0(zW^K)} = \frac{H_i(z)}{H_0(z)} \triangleq D_i(z), \quad \text{thus, } D_i(e^{j\omega}) = D_i(e^{j(\omega + \frac{2\pi}{M})}), \quad \text{i.e., } D_i(e^{j\omega}) \text{ is } \left(\tfrac{2\pi}{M}\right) \text{–periodic.}$$

Hence, $D_i(e^{j\omega}) = P_i(e^{j\omega M})$, i.e., by rationality, $D_i(z) = P_i(z^M)$. Thus, $H_i(z) = A(z)H'_i(z^M)$ where $A(z) = H_0(z)$ and $H'_i(z) = P_i(z)$, showing that the analysis banks of Figs. 6.12a,b can be made equivalent. Next, replacing z with zW^{-l} in condition (a) of the theorem shows that the condition holds even if each H_i is interchanged with F_i. Hence the same process can be repeated for the synthesis banks.

The above process may not ensure the PR property for the K band FB formed by the H'_i, F'_i (which we will refer to as the leaf FB). However, G_l now takes the form

$$\begin{aligned} G_l(z) &= A(zW^l)B(z)\sum_{i=0}^{K-1} H'_i(z^M W^{Ml})F'_i(z^M) = A(zW^l)B(z)G'_l(z^M), \quad \text{where} \\ G'_l(z) &= \sum_{i=0}^{K-1} H'_i(zW_K^l)F'_i(z) \quad (\text{where } W_K = W^M = \exp(\tfrac{-j2\pi}{K})). \end{aligned}$$

Thus, condition (a) implies that $G'_l(z) = 0$ for $l = 1, 2, \ldots, K-1$. (The alternative $A(zW^l)B(z) = 0$ is infeasible as it makes the systems identically zero.) Now the input–output relation of the leaf FB is $\widehat{V}(z) = \frac{1}{K}\sum_{i=0}^{K-1} V(zW_K^l)G'_l(z)$ (analogous to (6.4)). Thus the leaf FB is LTI with (rational) transfer function $U(z) = G_0'(z)/K$. Hence, dividing all the $H_i'(z)$ by $U(z)$ and multiplying $A(z)$ by $U(z^M)$ gives a new system with all the properties desired in condition (b). This proves that (a) implies (b). $\bigtriangledown\bigtriangledown\bigtriangledown$

Appendix H: Proof of Theorem 6.8

It suffices to prove the result for 2–unit trees, as we can continue by induction. A general 2–unit tree is specifiable as follows: The triples of (analysis filter, synthesis filter, decimator) are $(H_i(z), F_i(z), m_i), i = 0, 1, \dots, M-1$ for the root and $(A_i(z), B_i(z), k_i), i = 0, 1, \dots, K-1$ for the leaf, which is attached to decimator m_0 of the root. Thus the filters allowing and requiring modification are H_0, F_0 and the leaf filters A_i, B_i. The overall FB is unaffected iff the modifications preserve all the products $H_0(z)A_i(z^{m_0})$ and $F_0(z)B_i(z^{m_0})$.

Realizing stability, FIR filters: Let all the $H_0(z)A_i(z^{m_0})$ be stable. Then for every unstable pole $z = p$ of $A_j(z)$, there are m_0 unstable poles in $A_j(z^{m_0})$, one at each m_0-th root of p. To cancel these, we must have $H_0(z) = H_0'(z)C(z^{m_0})$ where H_0, H_0' have the same set of poles and $C(z) = (1 - z^{-1}p)$, so that $C(z^{m_0})$ is FIR with m_0 zeroes at the right places. Hence, replacing H_0 by H_0' and the A_i by A_iC removes the unstable pole of A_j and preserves the analysis filters of the overall FB. Thus all A_i can be made stable. Similarly if H_0 has an unstable pole p, each $A_i(z^{m_0})$ must have a zero at p, and hence for each i, $A_i(z) = A_i'(z)(1 - p^{m_0}z^{-1})$ where A_i, A_i' have the same set of poles. Thus, replacing A_i by A_i' and $H_0(z)$ by $H_0(z)(1 - p^{m_0}z^{-m_0})$ removes the unstable pole of H_0. Thus all filters can be made stable while preserving the overall FB. Similarly, if all the $H_0(z)A_i(z^{m_0})$ are FIR, the above argument can be repeated for *all* poles (rather than just the unstable ones), and all analysis filters can be made FIR.

Realizing PR, orthonormality: If the overall FB has PR, from (6.5) we get

$$(H_0(z)A_i(z^{m_0})F_0(z)B_j(z^{m_0}))\downarrow_{\gcd(m_0k_i, m_0k_j)} = ((H_0(z)F_0(z))\downarrow_{m_0} A_i(z)B_j(z))\downarrow_{\gcd(k_i,k_j)} = \delta(i-j) \tag{6.45}$$

With rational filters $X(z), Y(z)$ defined such that $XY = (H_0F_0)\downarrow_{m_0}$, let $A_i' = A_iX$, $B_i' = B_iY$ for all i. Thus from (6.45), $\left(A_i'(z)B_j'(z)\right)\downarrow_{\gcd(k_i,k_j)} = \delta(i-j)$, i.e., replacing each A_i by A_i' and B_i by B_i' causes the leaf FB to obey (6.5) and hence to have PR. The overall FB is preserved on replacing $H_0(z)$ by $H_0'(z) = H_0(z)/X(z^{m_0})$ and $F_0(z)$ by $F_0'(z) = F_0(z)/Y(z^{m_0})$. Since now both the leaf and the overall FB have PR, the root must have PR too. Thus the root and leaf have been modified as desired. Further if the overall FB is orthonormal, then it has PR with $F_0(z)B_i(z^{m_0}) = \widetilde{T}_i(z)$ where $T_i(z) = H_0(z)A_i(z^{m_0})$ (and of course, $F_i = \widetilde{H}_i$ for $i > 0$). Using $\widetilde{PQ} = \widetilde{P}\widetilde{Q}$, this means that (6.45) holds with F_0, B_i replaced by $\widetilde{H}_0, \widetilde{A}_i$ respectively. So we repeat with these substitutions, the earlier arguments used to make the root and leaf PR, and choose X such that $Y = \widetilde{X}$, i.e., such that

$X\widetilde{X} = \left(H_0\widetilde{H_0}\right)\downarrow_{m_0} \stackrel{\triangle}{=} W(z)$. (This is possible by spectral factorization, as $W(z)$ is rational and $W(e^{j\omega}) \geq 0$.) This ensures that the root and leaf are modified to be PR with $F_0' = \widetilde{H_0'}$ and $B_i' = \widetilde{A_i'}$. In other words, for all FBs, PR is obeyed and the synthesis filter corresponding to a given analysis filter D is $\widetilde{D}$. Thus both the root and leaf have been modified to be orthonormal rational FBs.

Chapter 7 Conclusion

In this thesis, we have made an extensive study of orthonormal filter bank optimization and the optimality of principal component filter banks. We have shown that PCFBs minimize all concave functions of the subband variances. This basic result leads to a unified explanation of PCFB optimality for several signal processing schemes, including compression, white noise suppression, and DMT communications. Various extensions were made to colored noise suppression, nonuniform FB optimization, and in some cases, biorthogonal FB optimization. We also saw the limitations of the PCFB concept. PCFBs do not exist for many practical FB classes, and this usually results in an analytically intractable optimization. The unconstrained class PCFBs are usually brickwall and unrealizable, and further fixed transforms such as the DFT achieve comparable performance for sufficiently large number of channels. The importance of the PCFB is that it provides a unified theory for FB optimization, supplies theoretical upper bounds on achievable performance, and indicates the direction in which to proceed when it is desired to adapt the FB to its input to improve performance.

Many issues remain open on both the theoretical and practical fronts in FB optimization. The theoretical issues involve PCFB existence, biorthogonal FBs, and nonuniform FBs. We have seen that PCFBs do not exist for many classes. However, formal analytical proofs of PCFB nonexistence for classes such as the FIR class are not yet available. The characterization of the spectra for which there is no PCFB for such a class is an open problem. A similar comment applies for nonuniform FBs, where such a result is not available even for the unconstrained class. For this class, at least for some special decimator–sets (e.g., the dyadic ones), it may be possible to obtain the optimum FB by a finite search, i.e., the search space may be a polytope, even if a PCFB does not exist. Characterization of the search space has also not been performed. Regarding biorthogonal FB optimization, some results have been obtained for the compression problem: The optimum FB has the form of a PCFB with pre and post filters. We can ask the question whether the same holds for other problems such as noise suppression with various subband operations such as Wiener filtering or hard thresholding. These problems are complicated by the dependence of the objective on all filters in the FB, rather than on merely the subband variances.

The practical issues involve developing efficient algorithms to adapt the FB to the input spectrum to achieve performance comparable to the PCFB. The ideal unrealizable filters of the unconstrained PCFB have to be approximated in a computationally feasible manner. In the DMT system, the channel and noise spectrum have to be learnt and the FB has to be adapted accordingly. Doing this in conjunction with the time and frequency domain equalization poses many challenging problems.

We have also studied the problem of parameterizing nonuniform FBs having perfect reconstruction (PR). Many new results have been derived — new necessary conditions on the decimators of rational PRFBs, and conditions on decimators under which PRFBs using them are necessarily derivable from an appropriate tree structure. Solving this problem in full is not very likely to produce new efficient FB designs, since a rich family of nonuniform FBs built from tree structures is already available. However, from a theoretical standpoint it is a basic question which has surprisingly remained unsolved. We have made a considerable improvement on the existing knowledge about the problem. However, as we have seen, the problem in its full generality is far from solved, and there are several open issues. Indeed, our work is possibly more of a beginning than an end to work in this area.

Bibliography

[1] A.N.Akansu, P.Duhamel, X.Lin, and M.de Courville, "Orthogonal transmultiplexers in communication: A review," IEEE Trans. Sig. Proc., vol. 46, no. 4, pp. 979-995, April 1998.

[2] S.Akkarakaran and P.P.Vaidyanathan, "Filter bank optimization with convex objectives, and the optimality of principal component forms," *IEEE Trans. Sig. Proc.,* vol. 49, pp. 100-114, Jan. 2001.

[3] S.Akkarakaran and P.P.Vaidyanathan, "New results and open problems on nonuniform filter banks," in *Proc. IEEE ICASSP,* Phoenix, AZ, Mar. 1999.

[4] S.Akkarakaran and P.P.Vaidyanathan, "On optimization of filter banks with denoising applications," in *Proc. IEEE ISCAS,* Orlando, FL, June 1999.

[5] S.Akkarakaran and P.P.Vaidyanathan, "The best basis problem, compaction problem and PCFB design problems," in *Proc. IEEE ISCAS,* Orlando, FL, June 1999.

[6] S.Akkarakaran and P.P.Vaidyanathan, "Optimized orthonormal transforms for SNR improvement by subband processing," in *Proc. IEEE Wkshop SP Adv. Wireless Comm.,* pp. 313-316, Annapolis, MD, May 1999.

[7] S.Akkarakaran and P.P.Vaidyanathan, "The Role of Principal Component Filter Banks in Noise Reduction," in *Proc. SPIE,* Denver, CO, July 1999.

[8] S.Akkarakaran and P.P.Vaidyanathan, "Results on principal component filter banks: Colored noise suppression and existence issues," *IEEE Trans. Inform. Theory,* vol. 47, no. 3, pp. 1003-1020, Mar. 2001.

[9] S.Akkarakaran and P.P.Vaidyanathan, "Principal component filter banks: existence issues, and application to modulated filter banks," in *Proc. IEEE ISCAS,* Geneva, May 2000.

[10] S.Akkarakaran and P.P.Vaidyanathan, "On nonuniform principal component filter banks: Definitions, existence and optimality," in *Proc. SPIE,* San Diego, CA, 2000.

[11] S.Akkarakaran and P.P.Vaidyanathan, "Are nonuniform principal component filter banks optimal?," *in Proc. EUSIPCO,* Tampere, Finland, September 2000.

[12] D.P.Bertsekas, "The auction algorithm for assignment and other network flow problems: A tutorial," *Interfaces,* vol. 20, no. 4, pp. 133-149, July–August 1990.

[13] J.A.C.Bingham, "Multicarrier modulation for data transmission: An idea whose time has come," *IEEE Comm. Mag.,* pp. 5-14, May 1990.

[14] F.Bruekers and A.van den Enden, "New networks for perfect inversion and perfect reconstruction," *IEEE J. Sel. Areas Comm.,* vol. 10, no. 1, pp. 130-137, Jan. 1992.

[15] T.Chen and L.Qiu, "General multirate building structures with application to nonuniform filter banks," *IEEE Trans. Circ. Syst.–II,* vol. 45, pp. 948-958, Aug. 1998.

[16] G.Cherubini, E.Eleftheriou, S.Olcer, and J.M.Cioffi, "Filter bank modulation techniques for very high speed digital subscriber lines," *IEEE Comm. Mag.,* pp. 98-104, May 2000.

[17] J.S.Chow, J.C.Tu, and J.M.Cioffi, "A discrete multitone transreceiver system for HDSL applications," *IEEE J. Sel. Areas Comm.,* vol. 9, no. 6, pp. 895-908, Aug. 1991.

[18] T.M.Cover and J.A.Thomas, *Elements of Information Theory.* John Wiley & Sons, Inc., 1991.

[19] S.Dasgupta and A.Pandharipande, "On biorthogonal nonuniform filter banks," *preprint.*

[20] I.Daubechies and W.Sweldens, "Factoring wavelet transforms into lifting steps," *J. Fourier Anal. Appl.,* vol. 4, no. 3, pp. 247-269, 1998.

[21] I.Djokovic and P.P.Vaidyanathan, "Results on biorthogonal filter banks," *ACHA,* pp. 329-343, 1994.

[22] D.L.Donoho and I.M.Johnstone, "Ideal spatial adaptation by wavelet shrinkage," *Biometrika,* vol. 81, no. 3, pp. 425-455, 1994.

[23] H.Feng and M.Effros, "On the rate-distortion optimality and computational efficiency of the Karhunen Loeve Transform for lossy data compression," *preprint.*

[24] N.J.Fliege, *Multirate Digital Signal Processing.* Chichester: John Wiley and Sons, 1994.

[25] R.M.Gray, "On the asymptotic eigenvalue distribution of Toeplitz matrices," *IEEE Trans. Inform. Theory,* vol. IT-18, no. 6, pp. 725-730, Nov. 1972.

[26] P.Gritzmann and B.Sturmfels, "Minkowski addition of polytopes: Computational complexity and applications to grobner bases," *SIAM J. Discrete Math,* vol. 6, no. 2, pp. 246-269, May 1993.

[27] G.H.Hardy, J.E.Littlewood, and G.Polya, *Inequalities.* Cambridge University Press, 1934.

[28] P.-Q.Hoang and P.P.Vaidyanathan, "Non-uniform multirate filter banks: Theory and design," in *Proc. IEEE ISCAS,* Portland, Oregon, May 1989, pp.371-374.

[29] R.A.Horn and C.R.Johnson, *Matrix Analysis.* Cambridge University Press, 1985.

[30] Y.Huang and P.M.Schultheiss, "Block quantization of correlated Gaussian random variables," *IEEE Trans. Commun. Syst.,* vol. C-10, pp. 289-296, Sept. 1963.

[31] I.Kalet, "The multitone channel," *IEEE Trans. Comm.,* vol. 37, no. 2, pp. 119-124, Feb. 1989.

[32] I.Kalet and S.Shamai, "On the capacity of a twisted-wire pair: Gaussian model," *IEEE Trans. Comm.,* vol. 38, no. 3, pp. 379-383, Mar. 1990.

[33] K.J.Kerpez, "Near-end crosstalk is almost Gaussian," *IEEE Trans. Comm.,* vol. 41, no. 5, pp. 670-672, May 1993.

[34] M.R.K.Khansari and E.Dubois, "Pade table, continued fraction expansion, and perfect reconstruction filter banks," *IEEE Trans. Sig. Proc.,* vol. 44, no. 8, pp. 1955-1963, Aug. 1996.

[35] A.Kirac and P.P.Vaidyanathan, "Optimality of orthonormal transforms for subband coding," in *Proc. IEEE DSP Workshop,* Utah, August 1998.

[36] A.Kirac and P.P.Vaidyanathan, "On existence of FIR principal component filter banks," in *Proc. IEEE ICASSP,* Seattle, May 1998.

[37] A.Kirac and P.P.Vaidyanathan, "Optimal nonuniform orthonormal filter banks for subband coding and signal representation," in *Proc. ICIP,* Chicago, 1998.

[38] A.Kirac and P.P.Vaidyanathan, "Theory and Design of Optimum FIR Compaction Filters," *IEEE Trans. Sig. Proc.,* vol. 46, no. 4, pp. 903-919, April 1998.

[39] J.Kovacevic and M.Vetterli, "Perfect Reconstruction Filter Bankswith Rational Sampling Factors," *IEEE Trans. Sig. Proc.,* vol. 41, pp. 2047-2066, June 1993.

[40] J.Li, T.Q.Nguyen, and S.Tantaratana, "A simple design method for near-perfect-reconstruction nonuniform filter banks," *IEEE Trans. Signal Processing,* vol. 45, pp. 2105-2109, Aug. 1997.

[41] Y-P.Lin and S-M.Phoong, "Perfect discrete multitone modulation with optimal transceivers," IEEE Trans. Sig. Proc., vol. 48, no. 6, pp. 1702-1712, June 2000.

[42] S.Mallat, *A Wavelet Tour of Signal Processing.* Academic Press, 1998.

[43] H.S.Malvar, *Signal Processing with Lapped Orthogonal Transforms.* Artech House, 1992.

[44] A.W.Marshall and I.Olkin, *Inequalities: Theory of Majorization and its Applications.* Academic Press, 1979.

[45] P.Moulin and M.K.Mihcak, "Theory and design of signal-adapted FIR paraunitary filter banks," *IEEE Trans. Sig. Proc.,* vol. 46, no. 4, pp. 920-929, April 1998.

[46] P.Moulin, M.Anitescu, and K.Ramchandran, "Theory of rate-distortion optimal, constrained filter banks – Application to IIR and FIR biorthogonal designs," *IEEE Trans. Sig. Proc.,* vol. 48, no. 4, pp. 1120-1132, April 2000.

[47] K.Nayebi, T.P.Barnwell,III, and M.J.T.Smith, "Nonuniform filter banks: A reconstruction and design theory," *IEEE Trans. Sig. Proc.,* pp. 1114-1127, Mar. 1993.

[48] J.G.Proakis, *Digital Communications.* McGraw Hill, 1995.

[49] J.O'Rourke, *Computational Geometry in C.* Cambridge University Press, 1998.

[50] R.T.Rockafellar, *Convex Analysis.* Princeton University Press, 1970.

[51] A.Said and W.A.Pearlman, "A new, fast, and efficient image codec based on set partitioning in hierarchical trees," *IEEE Trans. Circ. Syst. for Video Tech.,* vol. 6, no. 3, pp. 243-250, June 1996.

[52] V.P.Sathe and P.P.Vaidyanathan, "Effects of multirate systems on the statistical properties of random signals," *IEEE Trans. Sig. Proc.,* vol. 41, no. 1, pp. 131-146, Jan. 1993.

[53] J.M.Shapiro, "Embedded image–coding using zerotrees of wavelet coefficients," *IEEE Trans. Sig. Proc.,* vol. 41, no. 12, pp. 3445-3462, Dec. 1993.

[54] R.G.Shenoy, "Multirate specifications via alias–component matrices," *IEEE Trans. Circ. Syst.–II,* vol. 45, pp. 314–320, Mar. 1998.

[55] A.K.Soman and P.P.Vaidyanathan, "On orthonormal wavelets and paraunitary filter banks," *IEEE Trans. Sig. Proc.,* vol. 41, pp. 1170-1183, Mar. 1993.

[56] T.Starr, J.M.Cioffi, and P.J.Silverman, *Understanding Digital Subscriber Line Technology.* Prentice-Hall, NJ, 1999.

[57] G.Strang and T.Q.Nguyen, *Wavelets and Filter Banks.* Wellesley-Cambridge Press, 1996.

[58] G.L.Stuber, *Principles of Mobile Communication.* Kluwer Academic Publishers, 1996.

[59] M.K.Tsatsanis and G.B.Giannakis, "Principal component filter banks for optimal multiresolution analysis," *IEEE Trans. Sig. Proc.,* vol. 43, no. 8, pp. 1766-1777, Aug. 1995.

[60] M.Unser, "An extension of the KLT for wavelets and perfect reconstruction filter banks," in *Proc. SPIE no. 2034, Wavelet Appl. Sig. Image Proc.,* San Diego, CA, 1993, pp.45-56.

[61] P.P.Vaidyanathan, "Theory of optimal orthonormal subband coders," *IEEE Trans. Sig. Proc.,* vol. 46, no. 6, pp. 1528-1543, June 1998.

[62] P.P.Vaidyanathan and A.Kirac, "Results on optimal biorthogonal filter banks," *IEEE Trans. Circ. Syst.–II,* vol. 45, no. 8, pp. 932-947, August 1998.

[63] P.P.Vaidyanathan, Y-P.Lin, S.Akkarakaran, and S-M.Phoong, "Optimality of principal component filter banks for discrete multitone communication systems," in *Proc. IEEE ISCAS,* Geneva, May 2000.

[64] P.P.Vaidyanathan, *Multirate Systems and Filter Banks.* Englewood Cliffs, NJ: Prentice-Hall, 1993.

[65] P.P.Vaidyanathan, "Review of recent results on optimal orthonormal subband coders," in *Proc. SPIE,* San Diego, July 1997.

[66] P.P.Vaidyanathan, Y-P.Lin, S.Akkarakaran, and S-M.Phoong, "Discrete multitone modulation with principal component filter banks," *IEEE Trans. Inform. Theory,* under review.

[67] M.Vetterli and J.Kovacevic, *Wavelets and Subband Coding.* Englewood Cliffs, NJ: Prentice-Hall, 1995.

[68] J-J.Werner, "The HDSL environment," *IEEE J. Sel. Areas Comm.,* vol. 9, no. 6, pp. 785-800, Aug. 1991.

[69] `http://www.systems.caltech.edu/dsp/students/sony/journ/majproof.ps`