Submitted to IEEE Trans. Signal Processing

FILTER-BANK OPTIMIZATION WITH CONVEX OBJECTIVES, AND THE OPTIMALITY OF PRINCIPAL COMPONENT FORMS[1]

Sony Akkarakaran, Student member, IEEE and P.P.Vaidyanathan, Fellow, IEEE

Department of Electrical Engineering 136-93

California Institute of Technology

Pasadena, CA 91125 USA.

Phone: (626) 395-4681   Fax: (626) 795-8649

Contact Author: P.P.Vaidyanathan, Email: ppvnath@sys.caltech.edu

EDICS number: 2-FILB (Filter Bank Design and Theory)

Abstract. This paper proposes a general framework for the optimization of orthonormal filter banks (FB's) for given input statistics. This includes as special cases, many recent results on filter bank optimization for compression. It also solves problems that have not been considered thus far. FB optimization for coding gain maximization (for compression applications) has been well studied before. The optimum FB has been known to satisfy the *principal component* property, i.e., it minimizes the mean-square error caused by reconstruction after dropping the P weakest (lowest variance) subbands, for any P. In this paper we point out a much stronger connection between this property and the optimality of the FB. The main result is that a principal component filter bank (PCFB) is optimum whenever the minimization objective is a *concave function* of the subband variances produced by the FB. This result has its grounding in majorization and convex function theory, and in particular explains the optimality of PCFB's for compression. We use the result to show various other optimality properties of PCFB's especially for noise suppression applications. Suppose the FB input is a signal corrupted by additive white noise, the desired output is the pure signal, and the subbands of the FB are processed to minimize the output noise. If each subband processor is a zeroth order Wiener filter for its input, we can show that the expected mean square value of the output noise is a concave function of the subband signal variances. Hence a PCFB is optimum in the sense of minimizing this mean square error. The above–mentioned concavity of the error, and hence PCFB optimality, continues to hold even with certain other subband processors such as subband hard thresholds and constant multipliers, though these are not of serious practical interest. We prove certain extensions of this PCFB optimality result to cases where the input noise is *colored*, and the FB optimization is over a larger class that includes *biorthogonal* FB's. We also show that PCFB's do not exist for the classes of DFT and cosine-modulated FB's.

[1] Work supported in part by the National Science Foundation under Grant MIP 0703755.

1 Introduction

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. This paper proposes a general framework for the optimization of 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 thus far.

A generic signal processing scheme using an M-channel uniform perfect reconstruction filter bank (FB) is shown in Fig. 1. The FB is said to be **orthonormal** if the $M \times M$ analysis polyphase matrix $\mathbf{E}(e^{j\omega})$ is unitary for all ω. 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 $\mathbf{S}_{\mathbf{xx}}(e^{j\omega})$. We are also given a class $\mathcal{C}$ of **orthonormal uniform M-channel FB's**. Examples are the class of FB's in which all filters are FIR with a given bound on their order, or the class of unconstrained FB's (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 $\mathbf{S}_{\mathbf{xx}}(e^{j\omega})$, for use in the system of Fig. 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. 1, and the nature of the subband processors P_i. This is done in detail in Section 2 in a general setting.

1.1 Relevant earlier work

Consider in particular the case where the P_i are quantizers for signal compression. We use the model of [13] 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 noise processes are jointly stationary, 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 [13] 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 3.2. PCFB's for certain classes of FB's have been studied earlier. For example, let $\mathcal{C}^t$ denote the class of all M-channel orthogonal transform coders, i.e., FB's as in Fig. 1 where $\mathbf{E}(z)$ is a constant unitary matrix $\mathbf{T}$. The **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 [12] that the KLT is a PCFB for $\mathcal{C}^t$. For the class $\mathcal{C}^u$ of *all* (unconstrained) orthonormal M-channel FB's, construction of the PCFB has been studied by Tsatsanis and Giannakis [23], and independently by Vaidyanathan [25]. The goal of [25] 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 PCFB's [24], 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 [16, Th. 9.8, pg. 398] using a theorem of Hardy et al. In the present paper 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 PCFB's for all these objectives. This covers a wider class of applications, as shown in Section 6. It includes the conjecture of [24] (proved in [16]) as a special case where $h_i \equiv h$ for all i. It also includes the minimization objective of [13] 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 [18],[19]. 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* FB's. In an important preprint [19], Moulin et al study the minimization of this modified objective over the class of *all* (unconstrained) biorthogonal FB's, 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 [18], an algorithm is proposed for PCFB design for a certain class of FIR orthonormal FB's. 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). The present paper studies the geometric structure of the optimization search space, and thereby reveals several new optimality properties of PCFB's, especially those connected with noise reduction. Preliminary results of this work have been presented in [1],[2].

1.2 Main aims of this paper

This paper 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 FB's, 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 noise suppression. Suppose the FB input is a signal buried in noise, and the system of Fig. 1 aims to improve the signal to noise ratio. We consider the case where each subband processor P_i is 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 PCFB's 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, although these are of no serious practical interest. Thus we have a general problem formulation (Section 2) and a unified theory of optimal FB's (Section 3), which simultaneously explains the optimality of PCFB's for progressive transmission (Section 3.2), compression (Section 4.4) and noise suppression (Section 6). To emphasize the fact that PCFB's do not always exist, we also show in Section 5 that the classes of DFT and cosine-modulated FB's do not have PCFB's.

1.3 Notations

Superscripts (*) and (^{T}) denote the complex conjugate and matrix (or vector) transpose respectively, while superscript 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 transforms. $\mathcal{R}^M$ 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 $\text{diag}(\mathbf{A})$ the column vector consisting of the diagonal entries of the square matrix $\mathbf{A}$.

2 Problem formulation

We are given a class $\mathcal{C}$ of **orthonormal uniform M-channel FB's**. 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 Fig. 1). 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. 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 FB's 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. 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.[2] 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.

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

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

$$v_i^{(e)}(n) = v_i^{(d)}(n) - v_i^{(y)}(n). \tag{2}$$

Thus the processor P_i must try to produce an output 'as close to' $v_i^{(d)}(n)$ as possible, in the sense of minimizing $E[|v_i^{(e)}(n)|^2]$. In many situations to be discussed in detail later, the processors P_i are such that

$$E[|v_i^{(e)}(n)|^2] = h_i(\sigma_i^2). \tag{3}$$

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

$$g(\mathbf{v}) = \frac{1}{M}\sum_{i=0}^{M-1} h_i(\sigma_i^2). \tag{4}$$

Hence the minimization **objective** is **purely a function of the subband variance vector**. This function g of (4) 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 FB's 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 3.1). The central result of the present paper, described in detail in Section 3, 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 3.1). Since the set $\mathcal{S}$ plays an important role in the further discussion, we summarize the main definitions and facts pertaining to it:

2.1 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 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] 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.

2. **Computing the subband 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. 1, the vector process $(v_0^{(x)}(n), v_0^{(x)}(n), \ldots, v_0^{(x)}(n))^T$ 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
$$\mathbf{v} = \frac{1}{2\pi}\int_0^{2\pi} \mathrm{diag}\left(\mathbf{E}(e^{j\omega})\mathbf{S_{xx}}(e^{j\omega})\mathbf{E}^\dagger(e^{j\omega})\right)\, d\omega. \tag{5}$$

3. The optimization **search space** is defined as the set $\mathcal{S}$ of all subband variance vectors corresponding to all FB's 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 (5) 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 FB's: 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 FB's 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.

3 The optimality of PCFB's

We now show that principal component filter banks (PCFB's) 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 3.1. PCFB's are defined and described in Section 3.2. In Section 3.3 we show the connection between PCFB's and special convex sets called polytopes, and thereby prove the main result of the paper.

3.1 Convexity theory [20]

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.

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 $0 \leq \alpha_i \leq 1$ 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 [8],[22].

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,

$$f(\mu\mathbf{x} + (1-\mu)\mathbf{y}) \geq \mu f(\mathbf{x}) + (1-\mu)f(\mathbf{y}) \quad \text{whenever } 0 \leq \mu \leq 1. \tag{6}$$

Graphically, this means that the function f is always above its chord; see Fig. 2c. The domain D of f has to be convex to ensure that the argument of f on the left side of (6) is in D, i.e., to ensure that the above definition makes sense. For a concave function f, we can use (6) to show by induction that for any $\mathbf{x}_i \in D$,

$$f(\sum_{i=1}^{N} \alpha_i \mathbf{x}_i) \geq \sum_{i=1}^{N} \alpha_i f(\mathbf{x}_i) \quad \text{whenever} \quad 0 \leq \alpha_i \leq 1 \text{ and } \sum_{i=1}^{N} \alpha_i = 1. \tag{7}$$

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 (6) is achieved for distinct $\mathbf{x}, \mathbf{y}$ iff μ is either 0 or 1. For such f, equality is achieved in (7) 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, $P \stackrel{\triangle}{=} \text{co}(E)$ 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. 3a,b,c. The following result on extreme points, illustrated by Fig. 3d, is vital in explaining PCFB optimality.

Theorem 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,

$$f(\mathbf{v}) = f(\sum_{i=1}^{N} \alpha_i \mathbf{v}_i) \geq \sum_{i=1}^{N} \alpha_i f(\mathbf{v}_i) \qquad \text{(by (7), i.e., Jensen's inequality)} \tag{8}$$

$$\geq \sum_{i=1}^{N} \alpha_i f(\mathbf{v}_j) = f(\mathbf{v}_j) \qquad \text{(by definition of } \mathbf{v}_j \text{ and using } \sum_{i=1}^{N} \alpha_i = 1). \tag{9}$$

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. $\bigtriangledown\bigtriangledown\bigtriangledown$

Extreme points of general convex sets: A point $\mathbf{v}$ in a convex set D is said to be an extreme point of D if it cannot be expressed as a nontrivial convex combination of points of D, i.e., $\mathbf{v} = \sum_{i=1}^{J} \alpha_i \mathbf{x}_i$ for $\mathbf{x}_i \in D$ and $0 < \alpha_i < 1$, $\sum_{i=1}^{J} \alpha_i = 1$ implies $\mathbf{x}_1 = \ldots = \mathbf{x}_J \ (= \mathbf{v})$. This definition can be verified to be equivalent to the earlier definition of extreme points of polytopes when D is a polytope. Thus a polytope is simply a convex set with finitely many extreme points. We may note (though we do not use) the fact that Theorem 1 holds even if the domain P is a general *compact* convex set. This is proved in a very similar manner, using one additional key result: Every compact convex set is the convex hull of the set of its extreme points (Krein–Milman theorem) [20]. Polytopes are special compact convex sets (i.e., those with finitely many extreme points). Another easily proved fact on extreme points that we will use in Section 3.3.3 is as follows: For any set D, the extreme points of $\mathrm{co}(D)$ always lie in D.

3.2 PCFB's 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

$$\sum_{i=0}^{P} a_i \geq \sum_{i=0}^{P} b_i \qquad \text{for all } P = 0, 1, \ldots, M-1, \text{ with equality holding when } P = M-1. \tag{10}$$

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: PCFB's. Let $\mathcal{C}$ be the given class of orthonormal uniform M-channel FB's, 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. 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 $\mathbf{S}_{\mathbf{xx}}(e^{j\omega})$, if its subband variance vector (defined in Section 2.1) majorizes the subband variance vector of every FB in the class $\mathcal{C}$.

Remarks on the PCFB definition

1. **A simple optimality property**. In Fig. 1, 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 :

$$m_i = \begin{cases} 1 & \text{for } 0 \leq i \leq P-1 \\ 0 & \text{for } P \leq i \leq M-1 \end{cases} \tag{11}$$

for a fixed integer P with $0 \leq P \leq M$. This system keeps the P strongest (largest variance) subbands and discards the others. If the desired output signal $d(n)$ equals the input $x(n)$, then all assumptions of Section 2 are satisfied, and the minimization objective indeed has the form of (4): The optimum FB is the one minimizing $\frac{1}{M}\sum_{i=P}^{M-1}\sigma_i^2$. Now all FB's have the same value of $\sum_{i=0}^{M-1}\sigma_i^2$, so the optimum FB is the one maximizing $\sum_{i=0}^{P-1}\sigma_i^2$. Thus, from the definitions of PCFB's and majorization it follows that a PCFB, if it exists, has the property of being optimum for this problem for *all* values of P. In fact this property is the origin of the concept of a PCFB [23], and is clearly equivalent to its definition. PCFB's are also optimal for many other problems, as Section 3.3 will show.

2. **Existence of PCFB**. Given the class $\mathcal{C}$ of FB's 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), all FB's in $\mathcal{C}$ are PCFB's, no matter what $\mathcal{C}$ is. Section 4 studies certain classes $\mathcal{C}$ for which PCFB's 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 5 studies certain classes $\mathcal{C}$ for which PCFB's do not exist for large families of input spectra.[3]

3. **Nonuniqueness of PCFB**. From the definition of majorization, any permutation of a PCFB is also a PCFB. Further it is possible that two FB's which are not permutations of each other are both PCFB's, i.e., the PCFB need not be unique. However, all PCFB's 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 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.

3.3 Principal components, convex polytopes and PCFB optimality

Let $\mathcal{C}$ be the given class of orthonormal uniform M-channel FB's, and $\mathbf{S_{xx}}(e^{j\omega})$ the psd matrix of the vector input $\mathbf{x}(n)$ of Fig. 1. Let $\mathcal{S}$ be the set of all subband variance vectors of all FB's in $\mathcal{C}$ for input $\mathbf{x}(n)$. We have

Theorem 2: PCFB's 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 $\text{co}(\mathcal{S})$ 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 3: Optimality of PCFB's. 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 $\text{co}(\mathcal{S})$. Further if this function is strictly concave, the optimum FB is necessarily a PCFB.

Theorem 3 follows directly from Theorem 2 (proved in Section 3.3.3) and Theorem 1 of Section 3.1. Note that the FB optimization involves choosing the best vector from $\mathcal{S}$, but Theorem 1 is used here to find the best vector from $\text{co}(\mathcal{S}) \supset \mathcal{S}$. However, Theorem 2 shows that the best vector from $\text{co}(\mathcal{S})$ in fact lies in $\mathcal{S}$

[3] 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.

(and corresponds to the PCFB). Hence it must be optimum over $\mathcal{S}$. Also note that all permutations of a PCFB are PCFB's, and the above theorems do not specify which of these is the optimum. All of them need not be equally good in general. However the optimum can be found by a finite search over these PCFB's.

Theorem 3 shows optimality of PCFB's for a number of signal processing problems. In Section 2, 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 (4). If the functions h_i in (4) are all concave on $\mathcal{R}_+$, then g is concave on the domain $\mathrm{co}(\mathcal{S})$ [22]. This happens in several problems as we will see in later sections. Thus, Theorem 3 shows PCFB optimality for all these problems. To prove Theorem 2 and hence Theorem 3, we first review some results on majorization theory [11].

3.3.1 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).

3.3.2 Relevant results from majorization theory

1. **Majorization theorem** [10],[11]: 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** [11]: 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

$$\sum_{i=1}^{N} \alpha_i \mathbf{P}_i = \mathbf{Q}, \quad \text{where} \quad 0 \leq \alpha_i \leq 1, \quad \text{and} \quad \sum_{i=1}^{N} \alpha_i = 1. \tag{12}$$

3. **Orthostochastic majorization theorem** [11]: 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 $\mathbf{H}$ with entries of $\mathbf{a}$ as its eigenvalues and entries of $\mathbf{b}$ on its diagonal.

On the proofs: The majorization theorem actually follows from the orthostochastic majorization theorem. See [10] or [27] for an independent proof. Regarding Birkhoff's theorem, as all permutation matrices are doubly stochastic, so is their convex combination $\mathbf{Q}$ of (12) (Appendix A). The converse proof is more elaborate [11]. 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 4.2. Its proof is more involved [11]. 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 4.2.

3.3.3 Proof of Theorem 2

Let a PCFB for the class $\mathcal{C}$ exist for the given input psd $\mathbf{S_{xx}}(e^{j\omega})$. Let $\mathbf{v}_*$ be the PCFB subband variance vector (unique up to permutation; see Section 3.2). Let $\mathbf{P}_j$ be the $M \times M$ permutation matrices, for $j = 1, 2, \ldots, J$ (where $J = M!$), and let $\mathbf{v}_j \stackrel{\triangle}{=} \mathbf{P}_j \mathbf{v}_*$. Thus $E \stackrel{\triangle}{=} \{\mathbf{v}_j : j = 1, 2, \ldots, J\}$ 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 PCFB's, $\mathbf{v}_*$ majorizes $\mathbf{v}$. So by the majorization theorem (Section 3.3.1), $\mathbf{v} = \mathbf{Q}\mathbf{v}_*$ for some doubly stochastic matrix $\mathbf{Q}$. By Birkhoff's theorem (Section 3.3.1), $\mathbf{Q}$ is some convex combination of the $\mathbf{P}_j$. Thus,

$$\mathbf{v} = \mathbf{Q}\mathbf{v}_* = \sum_{j=1}^{J} \alpha_j \mathbf{P}_j \mathbf{v}_* = \sum_{j=1}^{J} \alpha_j \mathbf{v}_j \quad \text{for some } \alpha_j \text{ such that } \; 0 \leq \alpha_j \leq 1, \sum_{j=1}^{J} \alpha_j = 1. \tag{13}$$

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 2.1), $E \subset \mathcal{S}$, 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 $E \stackrel{\triangle}{=} \{\mathbf{v}_j : j = 1, 2, \ldots, J\}$, 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

$$0 \leq \alpha_j \leq 1, \; \sum_{j=1}^{J} \alpha_j = 1 \quad \text{and} \quad \mathbf{v} = \sum_{j=1}^{J} \alpha_j \mathbf{v}_j = \sum_{j=1}^{J} \alpha_j \mathbf{P}_j \mathbf{v}_* = \mathbf{Q}\mathbf{v}_* \; \text{ where } \; \mathbf{Q} \stackrel{\triangle}{=} \sum_{j=1}^{J} \alpha_j \mathbf{P}_j. \tag{14}$$

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. Indeed there is such a FB in $\mathcal{C}$: As $\text{co}(\mathcal{S}) = \text{co}(E)$ and E is a finite set, the extreme points of the polytope $\text{co}(\mathcal{S})$ lie within E; and they also lie in $\mathcal{S}$ (Section 3.1). 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 2.1). $\triangledown \triangledown \triangledown$

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 expressible as 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 (4) where the functions h_i are concave on $\mathcal{R}_+$. Theorem 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 (4)). In fact even this is not the complete family of objectives minimized by PCFB's. 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 PCFB's. 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** [17]. 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 PCFB's minimize all Schur–concave objectives. Full characterizations and several interesting examples of such functions can be found in [17].

4 PCFB's for standard classes, and optimality for compression

This section first shows existence of PCFB's for three special classes of FB's, namely classes with $M = 2$ channels, the class of M-channel orthogonal transform coders, and that of all M-channel orthonormal FB's. 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 FB's. 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* FB's, e.g. that of FIR or IIR FB's 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 2.1). Thus $\mathcal{S}$ lies wholly on a line of slope -1 in $\mathcal{R}^2$. So $\text{co}(\mathcal{S})$ is an interval on this line; see Fig. 4. Thus $\text{co}(\mathcal{S})$ 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 $\text{co}(\mathcal{S})$, irrespective of the input psd.[4]

4.1 The transform coder class

The transform coder class $\mathcal{C}^t$ is defined as the class of uniform M-channel orthonormal FB's whose polyphase matrix ($\mathbf{E}(z)$ in Fig. 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. 1. We have

Theorem 4: Transform coders – KLT, PCFB's 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 $\text{co}(\mathcal{S})$ that is a polytope, as stated by Theorem 2, Section 3.3.
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 (5) becomes

$$\mathbf{v} = \text{diag}(\mathbf{R_{vv}}), \quad \text{where} \quad \mathbf{R_{vv}} = \mathbf{T R_{xx} T}^\dagger, \quad \text{where} \quad \mathbf{R_{xx}} = \frac{1}{2\pi} \int_0^{2\pi} \mathbf{S_{xx}}(e^{j\omega}) \, d\omega. \tag{15}$$

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. 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 $\mathbf{v}_* \stackrel{\triangle}{=} \text{diag}(\mathbf{\Lambda})$ 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 3.3.1 (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

[4]If $\text{co}(\mathcal{S})$ 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 psd's.

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 3.3.1. 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 $\mathcal{S} = \text{co}(\mathcal{S})$.

4.2 The unconstrained class

The class $\mathcal{C}^u$ is defined to contain *all* uniform M-channel orthonormal FB's, with no constraints on the filters besides those imposed by orthonormality. So FB's in $\mathcal{C}^u$ could have ideal unrealizable filters. We could in effect think of $\mathcal{C}^u$ as the set of all $M \times M$ matrices $\mathbf{E}(e^{j\omega})$ that are unitary for all ω. ($\mathbf{E}(e^{j\omega})$ represents the analysis polyphase matrix.) An exact analog of Theorem 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 [24],[25]. 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 $\mathbf{S_{xx}}(e^{j\omega})$ for each ω, i.e., $\mathbf{K}(e^{j\omega})\mathbf{S_{xx}}(e^{j\omega})\mathbf{K}^\dagger(e^{j\omega}) = \boldsymbol{\Lambda}(e^{j\omega})$ where $\boldsymbol{\Lambda}(e^{j\omega})$ is diagonal (for all ω) and $\text{diag}(\boldsymbol{\Lambda}(e^{j\omega})) = (\lambda_0(e^{j\omega}), \lambda_1(e^{j\omega}), \ldots, \lambda_{M-1}(e^{j\omega}))^T$. Using (5), the subband variance vector $\mathbf{v}$ of an arbitrary FB $\mathbf{E}(e^{j\omega}) \in \mathcal{C}^u$ is given by

$$2\pi\mathbf{v} = \int_0^{2\pi} \text{diag}(\mathbf{E}(e^{j\omega})\mathbf{S_{xx}}(e^{j\omega})\mathbf{E}^\dagger(e^{j\omega}))\, d\omega = \int_0^{2\pi} \mathbf{Q}(e^{j\omega})\, (\lambda_0(e^{j\omega}), \lambda_1(e^{j\omega}), \ldots, \lambda_{M-1}(e^{j\omega}))^T\, d\omega. \quad (16)$$

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 (16) 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 (16). The integration process 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 $\mathbf{S_{xx}}(e^{j\omega})$, and (2) causing **spectral majorization** [25] – the said ordering of eigenvalues of $\mathbf{S_{xx}}(e^{j\omega})$. Note that the PCFB for $\mathcal{C}^u$ has uncorrelated subband *processes*, unlike the *instantaneous* decorrelation produced by the KLT (Theorem 4).

Proving $\mathcal{S} = \text{co}(\mathcal{S})$: 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 3.3.1), $\mathbf{v} = \mathbf{Q}\mathbf{v}_*$ 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 $\mathcal{C}^u$, (16) 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 $\mathcal{S} = \text{co}(\mathcal{S})$.

4.3 PCFB optimality for coding/compression

Here we consider the problems of [13],[25], where the processors P_i of Fig. 1 are quantizers, and the desired output $d(n)$ equals the input $x(n)$. This situation fits the general problem formulation of Section 2 under appropriate quantizer models. The subband error signal $v_i^{(e)}(n)$ of Section 2 here represents the i-th subband quantization noise. Under the quantizer model we assume, this noise is zero mean with variance

$$E[|v_i^{(e)}(n)|^2] = f_i(b_i)\sigma_i^2. \tag{17}$$

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 [13]. 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. Comparing (17) with (3) reveals the minimization objective g to be as in (4), i.e.,

$$g(\sigma_0^2, \sigma_1^2, \ldots, \sigma_{M-1}^2) = \frac{1}{M}\sum_{i=0}^{M-1} h_i(\sigma_i^2) \qquad \text{with} \qquad h_i(x) = f_i(b_i)x. \tag{18}$$

Thus the h_i are linear (and hence concave), so g is indeed concave. So by Theorem 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 (and hence for PCFB's 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 [25], i.e., uniform quantization under the high bit rate 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 [9], 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 [13]. 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 (18), 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 [13] without using any result on majorization: As c_i are constants, the optimization is unaffected by taking $c_i = 0$. With $m_0 \leq m_1 \leq \ldots \leq m_{M-1}$ and $\sigma_0^2 \geq \sigma_1^2 \geq \ldots \geq \sigma_{M-1}^2$,

$$Mg(\sigma_0^2, \sigma_1^2, \ldots, \sigma_{M-1}^2) = \sum_{i=0}^{M-1} m_i \sigma_i^2 = \sum_{i=0}^{M-2} (m_i - m_{i+1}) \left(\sum_{j=0}^{i} \sigma_j^2 \right) + m_{M-1} \left(\sum_{j=0}^{M-1} \sigma_j^2 \right). \qquad (19)$$

As the last term is constant for all FB's, and since $m_i - m_{i+1} \leq 0$, 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 6 that result in nonlinear concave h_i.

5 Filter Bank classes having no PCFB

Existence of a PCFB for a class $\mathcal{C}$ of orthonormal FB's implies a very strong condition on the subband variance vectors of the FB's in $\mathcal{C}$. There are many classes $\mathcal{C}$ which do not have PCFB's. Indeed it seems quite plausible that the classes of Section 4 are the only ones having PCFB's for all input power spectra. This section reviews some known results on nonexistence of PCFB's, and shows that the classes of ideal DFT and cosine-modulated FB's do not have PCFB's for several input spectra.

If a PCFB for the given class $\mathcal{C}$ of FB's exists, it simultaneously optimizes over $\mathcal{C}$, several functions of the subband variances (Section 3). So we can show nonexistence of PCFB's for $\mathcal{C}$ by proving that no single FB in $\mathcal{C}$ can optimize two of such functions. This method is used in [14],[18] for certain classes of FIR FB's, 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 PCFB's has not yet been *proved* for any reasonably general FIR class, say the class of all M-channel ($M > 2$) FIR orthonormal FB's with polyphase matrix of McMillan degree $\mu > 0$ (though it seems very likely that such classes do not have PCFB's). We now prove nonexistence of PCFB's for the classes of DFT and cosine-modulated FB's.

Definition. The class $\mathcal{C}^{dft}$ of M-channel orthonormal **DFT FB's** is the one containing all FB's as in Fig. 1 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** FB's (CMFB's) is the one containing all FB's as in Fig. 1 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 5: PCFB nonexistence for DFT, cosine-modulated FB classes. There are families of input psd's 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. 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 FB's produced by the prototypes. 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* [25] 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. 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 [4] a CMFB having $H_0^{(1)}(e^{j\omega})$ of Fig. 5b as one of its filters is necessarily the CMFB produced by $P^{(1)}(e^{j\omega})$ of Fig. 5b as prototype.

6 Optimal noise reduction with filter-banks

Suppose the FB input $x(n)$ of Fig. 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. 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. 6. This problem fits the formulation of Section 2 if we assume that $s(n)$ and $\mu(n)$ uncorrelated, and that $\mu(n)$ is white with a fixed known variance $\eta^2 > 0$. Indeed, using the notation of Section 2, 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 again white with variance η^2, and are uncorrelated to the signal components. The subband error process is

$$v_i^{(e)}(n) = v_i^{(d)}(n) - v_i^{(y)}(n) = v_i^{(s)}(n) - k_i v_i^{(x)}(n) = (1 - k_i) v_i^{(s)}(n) - k_i v_i^{(\mu)}(n). \tag{20}$$

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

$$E[|v_i^{(e)}(n)|^2] = |1-k_i|^2\,\sigma_i^2 + |k_i|^2\,\eta^2, \tag{21}$$

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 (21)) 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, (21) becomes $E[|v_i^{(e)}(n)|^2] = \frac{\sigma_i^2\eta^2}{\sigma_i^2+\eta^2}$, which is as in (3) with

$$h_i(x) = \frac{x\eta^2}{x+\eta^2}. \tag{22}$$

This function h_i is plotted in Fig. 7, and is easily verified to be concave on $[0,\infty)$. So by Theorem 3, PCFB's are optimal if the subband multipliers k_i are zeroth order Wiener filters.

6.1 Remarks on PCFB optimality for noise reduction

PCFB's for the pure or the noisy signal? Notice a difference between the argument σ_i^2 of h_i here and in (3): In (3), 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 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 FB's 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. 6. However, we may note that there are other choices that also result in an error function (21) 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 (21) 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:

$$k_i = \begin{cases} 1 & \text{if } \sigma_i^2 \geq T \\ 0 & \text{otherwise} \end{cases} \tag{23}$$

The resulting subband error functions h_i are plotted in Fig. 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) [22]. Unlike the Wiener filter however, these choices of multiplier k_i are of no serious practical interest, and are mentioned here only to demonstrate an academic implication of PCFB optimality. More practical hard thresholding schemes for noise suppression [7] 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 arguments of Section 3 involving majorization theory or the concavity of the function (22). 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 4.3 (see Equation (19)) without using convexity theory. This is possible since the $h_i(\sigma_i^2)$ in this case are as in (21) 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 PCFB's 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 6: Optimum FB based white noise suppression. In Fig. 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.

6.2 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 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 FB's are *biorthogonal* as opposed to orthonormal.

Consider the system of Fig. 8 where the bold-face vectors $\mathbf{s}(n), \boldsymbol{\mu}(n), \mathbf{x}(n)$ and $\mathbf{y}(n)$ are all M-fold blocked versions of the corresponding scalar 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 $\boldsymbol{\mu}(n)$ are uncorrelated WSS vector processes with psd matrices $\mathbf{S_{ss}}(e^{j\omega})$ and $\mathbf{S}_{\boldsymbol{\mu\mu}}(e^{j\omega})$ respectively. The blocked version of the error is then $\mathbf{e}(n) = \mathbf{y}(n) - \mathbf{s}(n)$ which is WSS with psd matrix $\mathbf{S_{ee}}(e^{j\omega})$ as follows: ($\mathbf{I}$ denotes the identity matrix)

$$\mathbf{S_{ee}}(e^{j\omega}) = \left[\mathbf{A}(e^{j\omega}) - \mathbf{I}\right] \mathbf{S_{ss}}(e^{j\omega}) \left[\mathbf{A}(e^{j\omega}) - \mathbf{I}\right]^{\dagger} + \mathbf{A}(e^{j\omega})\mathbf{S}_{\boldsymbol{\mu\mu}}(e^{j\omega})\mathbf{A}^{\dagger}(e^{j\omega}), \tag{24}$$

$$\text{where} \quad \mathbf{A}(e^{j\omega}) = \mathbf{R}(e^{j\omega})\mathbf{D}(e^{j\omega})\mathbf{E}(e^{j\omega}), \quad \text{and of course} \quad \mathbf{R}(e^{j\omega}) = \mathbf{E}^{-1}(e^{j\omega}). \tag{25}$$

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 $\mathbf{s}(n), \boldsymbol{\mu}(n)$ through transfer matrices $\left[\mathbf{A}(e^{j\omega}) - \mathbf{I}\right]$ and $\mathbf{A}(e^{j\omega})$ respectively. Since $\mathbf{s}(n), \boldsymbol{\mu}(n)$ 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 psd's, each easy to compute. Note that (24),(25) 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 $\mathbf{S}_{\boldsymbol{\mu\mu}}(e^{j\omega})$ is the identity matrix). The average mean-square value of the error $e(n)$ is

$$\varepsilon = \frac{1}{M}\text{trace}\left(\boldsymbol{\Sigma}_{\mathbf{ee}}\right), \text{ where } \boldsymbol{\Sigma}_{\mathbf{ee}} = \frac{1}{2\pi}\int_0^{2\pi} \mathbf{S_{ee}}(e^{j\omega})\, d\omega = \text{ autocorrelation matrix of } \mathbf{e}(n). \tag{26}$$

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

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

$$\boldsymbol{\Sigma}_{\mathbf{ee}} = \left[\mathbf{A} - \mathbf{I}\right] \boldsymbol{\Sigma}_{\mathbf{ss}} \left[\mathbf{A} - \mathbf{I}\right]^{\dagger} + \mathbf{A}\boldsymbol{\Sigma}_{\boldsymbol{\mu\mu}}\mathbf{A}^{\dagger}, \tag{27}$$

where $\boldsymbol{\Sigma}_{\mathbf{ss}}, \boldsymbol{\Sigma}_{\boldsymbol{\mu\mu}}$ are autocorrelation matrices of $\mathbf{s}(n)$ and $\boldsymbol{\mu}(n)$ 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.,

$$\mathbf{A} = \mathbf{RDE} = \boldsymbol{\Sigma}_{\mathbf{ss}} \left[\boldsymbol{\Sigma}_{\mathbf{ss}} + \boldsymbol{\Sigma}_{\boldsymbol{\mu\mu}}\right]^{-1}. \tag{28}$$

Suppose the signal and noise have a *common* KLT; i.e., for some unitary $\mathbf{T}$ both $\mathbf{T}\boldsymbol{\Sigma}_{\mathbf{ss}}\mathbf{T}^{\dagger} = \boldsymbol{\Lambda}_{\mathbf{ss}}$ and $\mathbf{T}\boldsymbol{\Sigma}_{\boldsymbol{\mu\mu}}\mathbf{T}^{\dagger} = \boldsymbol{\Lambda}_{\boldsymbol{\mu\mu}}$ are diagonal matrices. Then substitution in (28) shows that $\mathbf{A} = \mathbf{RDE} = \mathbf{T}^{\dagger}\mathbf{WT}$ where $\mathbf{W} = \boldsymbol{\Lambda}_{\mathbf{ss}} \left[\boldsymbol{\Lambda}_{\mathbf{ss}} + \boldsymbol{\Lambda}_{\boldsymbol{\mu\mu}}\right]^{-1}$ 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 7: Optimum memoryless transform for subband Wiener filtering. In Fig. 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. 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 6 and 7. Theorem 6 proves optimality of a PCFB for a general class $\mathcal{C}$ of orthonormal FB's, 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 7 focuses on the case when *all* subband multipliers are *Wiener filters*, and on a *special class* of FB's, namely the class $\mathcal{C}^b$ of all FB's with a *constant* (memoryless) polyphase matrix. Notice that $\mathcal{C}^b$ includes the orthogonal transform coder class $\mathcal{C}^t$. All Theorem 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 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 6 and 7 have a common element, which they generalize in different directions.

Further generalizations. Attempts to combine Theorems 6 and 7 yield many interesting further generalizations and open problems. For example, let us restrict attention to *orthogonal* transforms in Theorem 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 7). This result is shown in [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 6 needs some modifications as shown in [3]. It also appears plausible that the above optimality of the common KLT extends to the class of *all memoryless biorthogonal* transforms.[5] Verifying this is however currently an open problem.

6.2.2 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. 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 error ε of (26) where $\mathbf{S}_{\mathbf{ee}}(e^{j\omega})$ is as in (24),(25), 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 no analytical results are known to the authors at this time. N-th order Wiener filters ($N > 0$) cannot be handled like zeroth order ones as in Section 6.1. This is because the minimization objective now depends on not just the subband variances, but on N coefficients in the autocorrelation sequences of the subband random processes.

If *ideal* Wiener filters are used in each subband, an analog of Theorem 7 can be stated: In this case, any orthonormal FB whose polyphase matrix $\mathbf{E}(e^{j\omega})$ diagonalizes both the signal and noise psd matrices is

[5] An analogous result is true for the high bitrate coding problem with optimal bit-allocation (Section 4.4), i.e., the signal KLT is optimal over all memoryless biorthogonal transforms. This is proved using the Hadamard inequality for determinants.

optimal over the class of all unconstrained *biorthogonal* FB's. This result is obtained by repeating the methods used to prove Theorem 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: Diagonalization of the psd matrices is sufficient, and there is no constraint on the ordering of the subband spectra. However, this result is not very interesting when the scalar signal and noise input to the FB are WSS (as opposed to CWSS(M)): In this case, diagonalization of psd matrices is trivial using any orthonormal FB with non-overlapping analysis filters. The resulting system is then equivalent to an ideal scalar Wiener filter acting directly on the scalar input without use of any FB.

7 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 PCFB's for compression can be explained in this manner. PCFB's 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 FB's, and to the case when the noise is colored. We have also shown that the classes of ideal DFT and cosine-modulated FB's do not have PCFB's.

A companion paper [3] contains further results on colored noise suppression. It proves optimality of the common signal and noise KLT among the orthogonal transform coder class, for noise suppression using any combination of Wiener filters, hard thresholds and constants as subband multipliers. This further generalizes some parts of Theorem 7 of the present work. It is also shown in [3] that an analogous result on optimality of a common signal and noise PCFB for the class $\mathcal{C}^u$ of unconstrained FB's is false. We study the effect of absence of a PCFB on the FB optimization, and show that in general the problem becomes analytically intractable. We examine the connection between compaction filters, PCFB's and FB optimization. Extensions to nonuniform FB's have been studied in [5].

Acknowledgment. The authors gratefully thank the reviewers for their many useful suggestions that improved the paper significantly. We are also thankful to Prof. Soura Dasgupta for bringing reference [17] to our attention.

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,

$$\begin{array}{lccc} & \mathbf{Ck} = \alpha\mathbf{Ak} + (1-\alpha)\mathbf{Bk} = \alpha\mathbf{k} + (1-\alpha)\mathbf{k} = \mathbf{k}, & \text{and similarly} & \mathbf{k}^T\mathbf{C} = \mathbf{C}. \\ \text{Likewise,} & \mathbf{Dk} = \mathbf{ABk} = \mathbf{Ak} = \mathbf{k}, & \text{and similarly} & \mathbf{k}^T\mathbf{D} = \mathbf{D}. \end{array} \tag{29}$$

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 matrix $\mathbf{A} = \frac{1}{2}\begin{bmatrix} 1 & 1 & 0 \\ 1 & 0 & 1 \\ 0 & 1 & 1 \end{bmatrix}$. 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 $M > 2$.

References

[1] S.Akkarakaran and P.P.Vaidyanathan, "On optimization of filter banks with denoising applications," in *Proc. IEEE ISCAS,* Orlando, FL, June 1999.

[2] S.Akkarakaran and P.P.Vaidyanathan, "Optimized Orthonormal Transforms for SNR Improvement by Subband Processing," in *Proc. IEEE Wkshop SP Adv. Wireless Comm.,* Annapolis, MD, May 1999.

[3] S.Akkarakaran and P.P.Vaidyanathan, "Results on Principal Component Filter Banks: Colored Noise Suppression and Existence Issues," *submitted to IEEE Trans. Inform. Theory.*

[4] S.Akkarakaran and P.P.Vaidyanathan, "Principal Component Filter Banks: Existence Issues, and Application to Modulated Filter Banks," to appear in *Proc. IEEE ISCAS,* Geneva, May 2000.

[5] S.Akkarakaran and P.P.Vaidyanathan, "On Nonuniform Principal Component Filter Banks: Definitions, Existence and Optimality," in *Proc. SPIE,* San Diego, CA, 2000.

[6] J.B.Conway, *A Course in Functional Analysis.* Springer–Verlag, New York, 1985.

[7] D.L.Donoho and I.M.Johnstone, "Ideal spatial adaptation by wavelet shrinkage," *Biometrika,* vol. 81, no. 3, pp. 425-455, 1994.

[8] N.Dunford and J.T.Schwartz, *Linear Operators, vols. I and II.* Interscience, New York, 1964.

[9] H.Feng and M.Effros, "On the Rate-Distortion Optimality and Computational Efficiency of the Karhunen Loeve Transform for Lossy Data Compression," *preprint.*

[10] G.H.Hardy, J.E.Littlewood, and G.Polya, *Inequalities.* Cambridge University Press, 1934.

[11] R.A.Horn and C.R.Johnson, *Matrix Analysis.* Cambridge University Press, 1985.

[12] 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.

[13] A.Kirac and P.P.Vaidyanathan, "Optimality of Orthonormal Transforms for Subband Coding," in *Proc. IEEE DSP Workshop,* Utah, August 1998.

[14] A.Kirac and P.P.Vaidyanathan, "On Existence of FIR Principal Component Filter Banks," in *Proc. ICASSP,* Seattle, May 1998.

[15] S.Lang, *Real Analysis.* Addison–Wesley, 1983.

[16] S.Mallat, *A Wavelet Tour of Signal Processing.* Academic Press, 1998.

[17] A.W.Marshall and I.Olkin, *Inequalities: Theory of Majorization and its Applications.* Academic Press, 1979.

[18] P.Moulin and M.K.Mihcak, "Theory and Design of Signal-Adapted FIR Paraunitary Filter Banks," *IEEE Trans. Signal Processing,* vol. 46, no. 4, pp. 920-929, April 1998.

[19] P.Moulin, M.Anitescu, and K.Ramchandran, "Theory of Rate-Distortion Optimal, Constrained Filter Banks – Application to IIR and FIR Biorthogonal Designs," *preprint,* June 1998.

[20] R.T.Rockafellar, *Convex Analysis.* Princeton University Press, 1970.

[21] W.Rudin, *Functional Analysis.* McGraw–Hill, 1991.

[22] D.L.Russell, *Optimization Theory.* W.A.Benjamin, New York, 1970.

[23] M.K.Tsatsanis and G.B.Giannakis, “Principal Component Filter Banks for Optimal Multiresolution Analysis,” *IEEE Trans. Signal Processing,* vol. 43, no. 8, pp. 1766-1777, Aug. 1995.

[24] M.Unser, “An extension of the KLT for wavelets and perfect reconstruction filter banks,” in *Proc. SPIE no. 2034, Wavelet Appl. Signal Image Processing,* San Diego, CA, 1993, pp.45-56.

[25] P.P.Vaidyanathan, “Theory of Optimal Orthonormal Subband Coders,” *IEEE Trans. Signal Processing,* vol. 46, no. 6, pp. 1528-1543, June 1998.

[26] P.P.Vaidyanathan and A.Kirac, “Results on Optimal Biorthogonal Filter Banks,” *IEEE Trans. Circuits and Systems II,* vol. 45, no. 8, pp. 932-947, August 1998.

[27] `http://www.systems.caltech.edu/dsp/students/sony/journ/majproof.ps`

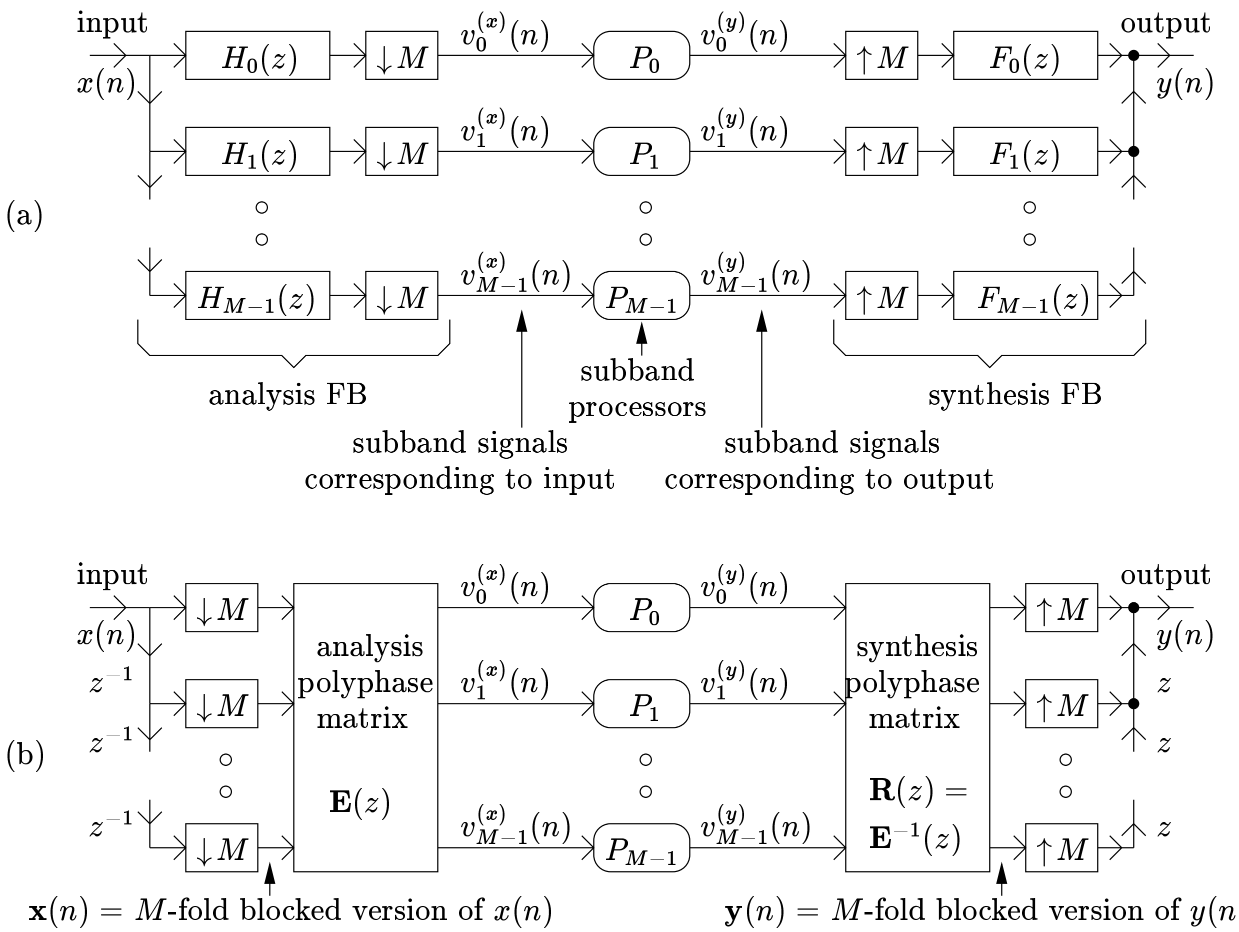

Figure 1: **Generic FB based signal processing scheme.**
(a) Showing analysis and synthesis filters, (b) Polyphase representation.

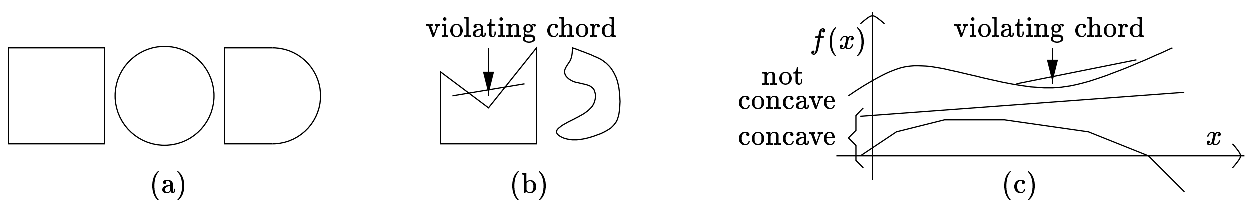

Figure 2: **Convex sets and concave functions.**
(a) Convex sets, (b) Nonconvex sets, (c) Concave functions of one variable.

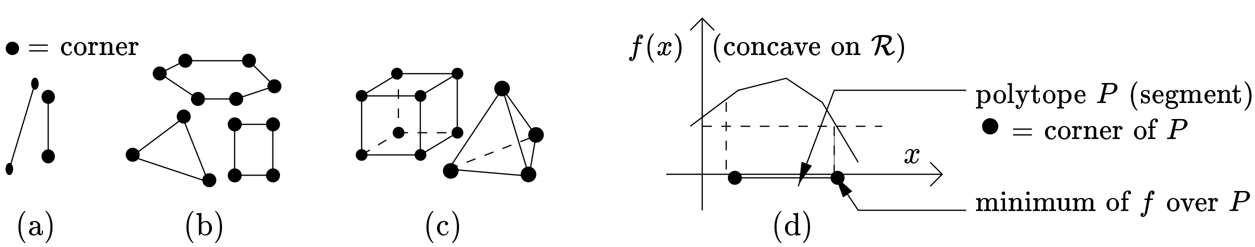

Figure 3: M**-dimensional polytopes, their extreme points and their optimality.**
(a) $M = 1$, (b) $M = 2$, (c) $M = 3$, (d) Optimality of extreme points.

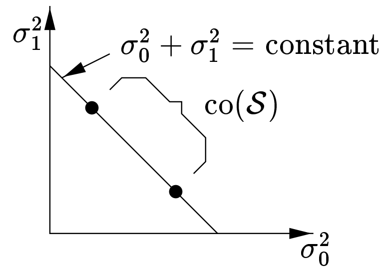

Figure 4: **Search space $\mathcal{S}$ for a class of 2-channel FB's.**

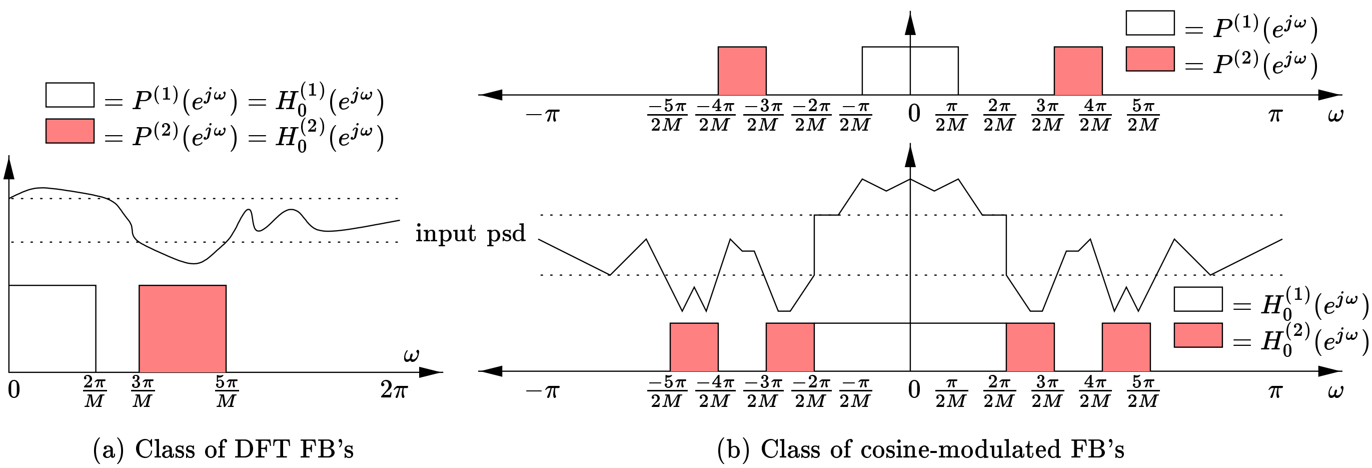

Figure 5: **Nonexistence of PCFB's.**

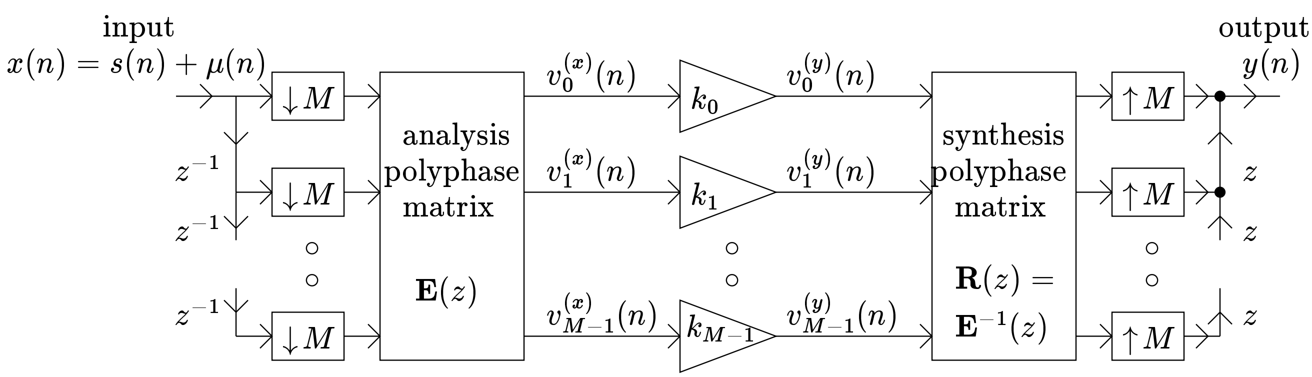

Figure 6: **FB based noise reduction.**

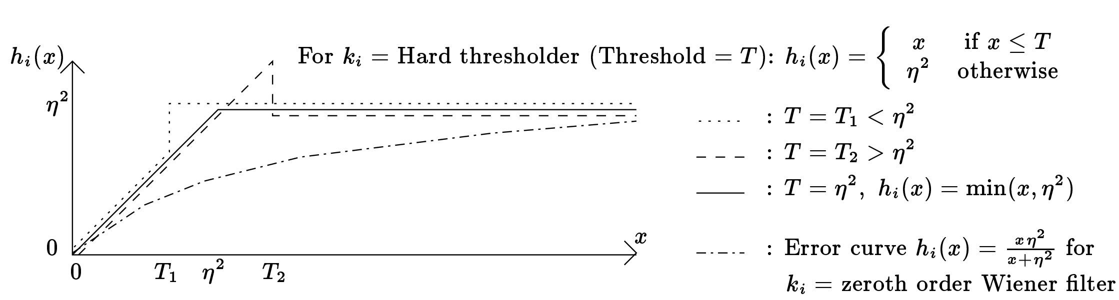

Figure 7: **Subband error functions in noise reduction.**

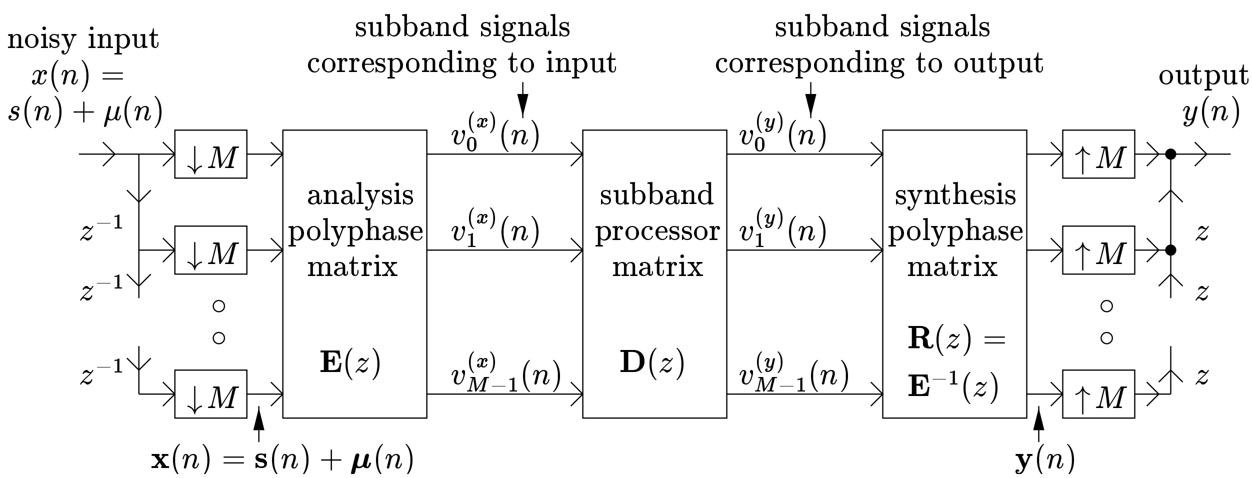

Figure 8: **Subband noise reduction: System of Section 6.3.**

LIST OF FIGURES

FOOTNOTES

1. Manuscript received ______________________________

2. Work supported in part by the National Science Foundation under Grant MIP 0703755.

3. The authors are with the Department of Electrical Engineering, California Institute of Technology, Pasadena, CA 91125.