THE BEST BASIS PROBLEM, COMPACTION PROBLEM AND PCFB DESIGN PROBLEMS

Sony Akkarakaran and P.P. Vaidyanathan

Department of Electrical Engineering 136-93
California Institute of Technology
Pasadena, CA 91125 USA.

ABSTRACT

In a companion paper, we have considered the problem of optimization of filter banks (FB's) for given input statistics. We have pointed out a strong connection between FB optimality and the principal component property. We have shown that *principal component filter banks* (PCFB's) are optimal for various signal processing schemes such as coding and denoising. In the present paper, we examine the nature of the FB optimization problems for these schemes in situations where a PCFB does not exist. We describe an algorithm involving a sequential design of compaction filters, which is known to produce PCFB's if they exist. We then demonstrate in an insightful manner how this algorithm can be suboptimal when PCFB's do not exist. This was earlier shown only by numerical examples.

1. INTRODUCTION

A generic filter-bank based signal processing scheme is shown in Fig. 1. For such a system, a companion paper [1] addresses the following problem: Find the best filter bank (FB) from a class $\mathcal{C}$ of uniform M-channel orthonormal FB's, for given statistics of the input $x(n)$. For various schemes described by Fig. 1, the subband processors P_i are such that the FB optimization objective is a concave function of the subband variances produced by the FB. Whenever this happens, we have shown that a *principal component filter bank* [1, 2] for the class $\mathcal{C}$ is always optimum within $\mathcal{C}$.

The existence of a PCFB for the class $\mathcal{C}$ is thus seen to make these FB optimization problems analytically tractable. In particular, when $\mathcal{C}$ is the class of all (unconstrained) M-channel orthonormal FB's, the PCFB can be obtained by an algorithm involving design of a sequence of compaction filters, as observed in [3]. In the present paper we state this algorithm formally in a way that makes it well-defined for arbitrary classes $\mathcal{C}$. We then use the geometric approach described in [1] to show that this algorithm always produces a PCFB if one exists. Further we show that in absence of a PCFB, the algorithm will be suboptimal for a large number of objectives. Since this was previously shown only by numerical examples [4], it provides a new insight into the problem.

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

2. REVIEW OF PCFB OPTIMALITY

Following [1], we define the **subband variance vector** associated with the system of Fig. 1, as the vector $\mathbf{v} = (\sigma_0^2, \sigma_1^2, \ldots, \sigma_{M-1}^2)^T$ whose i-th entry is the variance of the subband signal input to the processor P_i, for $i = 0, 1, \ldots, M-1$. The optimization **search space** is defined as the set S consisting of all subband variance vectors associated with all FB's in the given class $\mathcal{C}$. Since there is no constraint on which processor to use in which subband in Fig. 1, the set S is 'permutation-symmetric': If $\mathbf{v}$ is in S then all vectors obtained by permuting the entries of $\mathbf{v}$ are also in S, and correspond to the same FB in $\mathcal{C}$. The optimization objective is a real valued function g defined on the set S. We assume that g has an extension to the convex hull of S (denoted by $\text{co}(S)$) [1] on which it is concave.

As noted in [1], existence of a PCFB implies that the set $\text{co}(S)$ is a *polytope*. By definition this means that $\text{co}(S) = \text{co}(E)$ where E is a *finite* set. Assuming that E is chosen to have as few elements as possible, the vectors in E are known as *corners* of the set $\text{co}(S)$. When a PCFB exists, in fact these corners are permutations of each other, and correspond to the PCFB. Thus, since g is assumed to be concave over the polytope $\text{co}(S)$, at least one corner of this polytope is a minimum of g over $\text{co}(S)$ (and hence over S), as illustrated in Fig. 2. Thus a PCFB is always optimal for such objectives g. For various FB based signal processing schemes, the processors P_i of Fig. 1 are such that g is indeed concave on $\text{co}(S)$. g usually takes the form

$$g(\sigma_0^2, \sigma_1^2, \ldots, \sigma_{M-1}^2) = \frac{1}{M} \sum_{i=0}^{M-1} h_i(\sigma_i^2) \qquad (1)$$

which is concave on $\text{co}(S)$ whenever the h_i are concave on $[0, \infty)$. For example, in the high bitrate coding problem of [3], $h_i(x) = \log(x)$. In the noise suppression problems considered in [1], $h_i(x) = \frac{x\eta^2}{x+\eta^2}$ when P_i is a zeroth order Wiener filter, and $h_i(x) = \min(x, \eta^2)$ when P_i is a hard-threshold operation (where η^2 is the noise variance).

Thus whenever $\text{co}(S)$ is a polytope, the optimization can be reduced to a search over the *finite* set of FB's that correspond to the corners of the polytope. When a PCFB exists, this set has exactly one element, namely the PCFB. If there is no PCFB, one could hope that if $\text{co}(S)$ is indeed still a polytope, then it would not be very difficult to identify this finite set of FB's that corresponds to its corners (and thereby solve the optimization problem). However, a polytope is a fairly structured object. Given an input power spectrum and a class $\mathcal{C}$ of FB's, say the class of FIR FB's with a given bound on the filter orders, there is no apriori reason to believe that the corresponding set co(S) is a polytope. In general, co(S) would thus be a bounded convex set that is not necessarily a polytope. We shall assume that co(S) is closed (or compact), which will be true for most 'natural' classes $\mathcal{C}$ of FB's. We will next observe that corners can be defined for arbitrary convex sets (not necessarily polytopes), and note that they have optimality properties similar to those discussed above.

3. ARBITRARY CONVEX SETS: CORNERS AND THEIR OPTIMALITY

Definition [5]. Let B be a convex subset of $\mathcal{R}^M$. A point $\mathbf{z} \in B$ is said to be an *extreme point*, or a *corner* of B if

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

$$\text{implies} \quad \mathbf{x} = \mathbf{y}\ (= \mathbf{z}).$$

Geometrically, we cannot draw a line-segment that contains $\mathbf{z}$ in its interior (i.e. not as an endpoint) and yet lies wholly within the set B. The interior of B cannot have any corners, because around each point in the interior we can draw a ball that lies wholly in B. So all the corners lie on the boundary. However, not all boundary points are necessarily corners. If B is a polytope, the above definition can be verified to coincide with the earlier definition of corners of a polytope. These points are illustrated in Fig. 3, which shows the corners of some closed and bounded (or compact) convex sets.

It can be shown without much effort, that every compact convex set is the convex hull of its boundary, and that it has at least one corner. The proof of the following result however is less obvious:

Krein-Milman theorem / Internal representation of convex sets [5, 6]: Every compact convex set is the convex hull of its corners.

This result is evidently true for polytopes, and can be verified to be true in the examples shown in Fig. 3. The result is thus intuitively clear (although its formal proof might not be obvious). Its importance lies in the fact that it can be used to immediately prove

Optimality of corners: Given any function g that is concave on a compact convex set D, at least one of the corners of D is a minimum of g. Further if g is strictly concave then its minimum is necessarily at a corner of the set.

For the special case when the compact convex set is a polytope, this result was discussed in [1] and is illustrated in Fig. 2. Fig. 4 illustrates the result for a compact convex set that is *not* a polytope. In Fig. 4, all corners are 'equally good', i.e. all are minima, but this of course need not be true in general.

Proof of optimality of corners: Let $\mathbf{v}_{opt}$ be the minimum of g over D. (Its existence is either assumed or follows if g is assumed to be continuous.) By the Krein-Milman theorem, $\mathbf{v}_{opt}$ is a convex combination of some set of corners of D, i.e.

$$\mathbf{v}_{opt} = \sum_{j=1}^{J} \beta_j \mathbf{z}_j \quad \text{where} \quad \beta_j \in [0,1],\ \sum_{j=1}^{J} \beta_j = 1 \qquad (2)$$

for some distinct corners $\mathbf{z}_j$ of D. Now at least one of the $\mathbf{z}_j$ has to be a minimum of g over D. If not, then $g(\mathbf{z}_j) > g(\mathbf{v}_{opt})$ for all $j = 1, 2, \ldots, J$, and hence

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

i.e. $g(\mathbf{v}_{opt}) > g(\mathbf{v}_{opt})$ which is a contradiction. Hence at least one corner of D is a minimum of g over D. The first inequality above is the Jensen's inequality for concave functions. If g is strictly concave, then this inequality is strict unless one of the β_j is unity. Hence in this case $\mathbf{v}_{opt}$ equals the corresponding $\mathbf{z}_j$, i.e. the minimum is necessarily at a corner of D.

$\nabla \nabla \nabla$

In our FB optimization problem, $D = \text{co}(S)$ where S is the search space. Let E be the set of corners of D, so $E \subset S$ and $\text{co}(S) = \text{co}(E)$. From the above result, the optimization over co(S) can be reduced to one over E. Thus the analytical tractability of the problems of [1] can be traced to the fact that co(S) is a polytope, i.e. that E is finite. In general, 'almost every' corner in E is associated with a concave (in fact linear) objective for which the corner is the unique minimum.[1] Thus, if a PCFB does not exist, or more precisely if E is not finite, it is not possible to make a general statement about the optimality of any single FB for a large class of objectives. It might be possible to avoid a suboptimum numerical search for a specific objective. However the analytical solution will have to exploit the specific structures of both the objective function and the set co(S). Thus we see that in absence of a PCFB, the problem of finding the optimum FB for a given concave objective usually becomes *analytically intractable.* So in such cases, a numerical procedure (that in general gives a suboptimum solution) such as a gradient-descent based algorithm is usually needed. It is enough to search for the minima over the set E (as opposed to S or co(S)); however it is not known to the authors at this time whether there are numerical search procedures that can exploit this fact.

4. THE SEQUENCE - OF - COMPACTION-FILTERS ALGORITHM

This is an algorithm that has sometimes been proposed [3, 4] to find a 'good' FB in classes $\mathcal{C}$ that need not necessarily have PCFB's. We first state the algorithm in a precise way that holds for any general class $\mathcal{C}$. This will show that it produces FB's for which the corresponding subband variance vector *is a corner of* co(S). The optimality of the algorithm is then examined in this light.

[1]This can be proved using the concept of 'exposed points' or 'tangent hyperplanes' to convex sets [6].

4.1. Algorithm statement

Let us be given the class $\mathcal{C}$ of FB's, and the corresponding optimization search space S. For each $\mathbf{v} \in S$, define $\mathbf{w}$ as the vector obtained by arranging the entries of $\mathbf{v}$ in increasing order. Let $T_0 \subset S$ be the set of all such $\mathbf{w}$ for all $\mathbf{v} \in S$. The i-th coordinate of a vector $\mathbf{w} \in \mathcal{R}^M$ is denoted by w_i, for $i = 0, 1, \ldots, M-1$. Then,

1. Find $\alpha_0 \triangleq \max\{w_0 : \mathbf{w} \in T_0\}$. Let $T_1 \subseteq T_0$ be the set of all $\mathbf{w} \in T_0$ such that $w_0 = \alpha_0$.
2. Find $\alpha_1 \triangleq \max\{w_1 : \mathbf{w} \in T_1\}$. Let $T_2 \subseteq T_1$ be the set of all $\mathbf{w} \in T_1$ such that $w_1 = \alpha_1$.
3. Continue the process, i.e. find $\alpha_i \triangleq \max\{w_i : \mathbf{w} \in T_i\}$. Let $T_{i+1} \subseteq T_i$ be the set of all $\mathbf{w} \in T_i$ such that $w_i = \alpha_i$. Do this for $i = 0, 1, \ldots, M-1$.
4. This procedure uniquely defines a vector
$$\mathbf{v}_\alpha \triangleq (\alpha_0, \alpha_1, \ldots, \alpha_{M-1})^T \in S \tag{3}$$
with $\alpha_0 \geq \alpha_1 \geq \ldots \geq \alpha_{M-1}$. In fact the set T_M consists of the single element $\mathbf{v}_\alpha$. The output of the algorithm is any FB in $\mathcal{C}$ with subband variance vector $\mathbf{v}_\alpha$ (or any of its permutations).

Step 3 for each i represents a maximization of a subband variance followed by a narrowing down of the search. This gives the algorithm its name. By construction, the vector $\mathbf{v}_\alpha$ has the special property that it is the *greatest* of all vectors in S in the 'dictionary ordering' on $\mathcal{R}^M$. In fact, a concise restatement of the above algorithm is that it finds the vector $\mathbf{v}_\alpha$ with this property, and finds all FB's in $\mathcal{C}$ that have subband variance vector $\mathbf{v}_\alpha$ (or its permutations).

4.2. Connection to compaction filters

Given the power spectrum of a wide sense stationary (WSS) input, the notion of an optimum *compaction filter* for the input has been defined in [3]: It is the filter that maximizes its output variance among all filters in the class $\mathcal{F}_0$ of Nyquist(M) filters. The motivation for the Nyquist(M) constraint is the fact that these filters are used to construct an orthonormal FB, and any filter in such a FB is Nyquist(M).

The compaction filter can be constructed by a procedure outlined in [3]. This procedure always results in an ideal filter, i.e. one supported on an aliasfree(M) zone and having constant magnitude on this zone. If such a filter is to be part of an orthonormal FB, its support cannot overlap with the supports of the remaining filters in the FB. Thus given the compaction filter H_0 for the input, we can complete the FB by a sequential design of compaction filters [3]: For $i = 1, 2, \ldots, M-1$, the i-th filter H_i in the FB is one that maximizes its output variance among all filters in the class $\mathcal{F}_i \subset \mathcal{F}_{i-1}$. Here the class $\mathcal{F}_i$ is defined to consist of filters whose supports do not overlap those of the previously designed (ideal) filters $H_0, H_1, \ldots, H_{i-1}$. An equivalent definition of H_i is as follows: Consider the power spectrum obtained by setting to zero the bands of the original input spectrum that fall within the supports of the previously defined filters $H_0, H_1, \ldots, H_{i-1}$. Then H_i is a compaction filter for this modified power spectrum.

The above procedure from [3] can now be seen to be exactly the algorithm of Section 4.1 applied to the class $\mathcal{C}$ of all (unconstrained) orthonormal FB's. However, stating this algorithm using the notion of a compaction filter becomes difficult for a *general* class $\mathcal{C}$. To illustrate this, let $\mathcal{C}$ be the class of all M-channel FIR orthonormal FB's having filters with orders bounded by N. With this class in mind, it might appear reasonable to make the following definition: A *FIR compaction filter* is one that maximizes its output variance among all filters in the class $\mathcal{G}_0$ of all Nyquist(M) filters *with order bounded by* N. Indeed, design of such filters has been studied [7]. However, completing the FB poses difficulties: Letting H_0 be the FIR compaction filter, it is not even clear whether there is any FB in $\mathcal{C}$ that contains H_0 as one of its filters. The statement of Section 4.1 circumvents this difficulty.

4.3. Is the algorithm optimal?

Fact. The vector $\mathbf{v}_\alpha \in S$ produced by the algorithm of Section 4.1 **is a corner of** co(S).
Proof. Let $\mathbf{v}_\alpha = \gamma\mathbf{x} + (1-\gamma)\mathbf{y}$ for $\gamma \in (0,1)$ and $\mathbf{x}, \mathbf{y} \in$ co(S). Then by definition of a corner (Section 3), the proof will be completed if we show that $\mathbf{x} = \mathbf{y} = \mathbf{v}_\alpha$. Now by definition of the convex hull co(S), $\mathbf{x}, \mathbf{y}$ and hence $\mathbf{v}_\alpha$ can be written as convex combinations of elements of S, i.e. $\mathbf{v}_\alpha = \sum_{j=1}^{J} \beta_j \mathbf{v}^j$ for some $\mathbf{v}^j \in S$ 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 $\alpha_0 \geq v_0^j$, we have $\alpha_0 = v_0^j$. Hence $\mathbf{v}^j \in T_1$. 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 corners of co(S) correspond to the PCFB. Hence the algorithm of Section 4.1 always produces the PCFB, and is thus optimal for many problems [1]. The vector $\mathbf{v}_\alpha$ of (3) here has an additional property: If its entries are arranged in *increasing order*, then in fact it becomes the *least* vector in S in the dictionary ordering on $\mathcal{R}^M$.[2] On the other hand, if a PCFB does not exist, then there will be at least two corners that are not equivalent, i.e. whose coordinates are not permutations of each other. The algorithm of Section 4.1 produces one corner, but the minima could easily be at other non-equivalent corners. Thus the algorithm could be suboptimum.

To illustrate this point, consider the following hypothetical example with $M = 3$ channels: Let co(S) = co(E) where the set E consists of vectors $\mathbf{v}_1 = (3,2,1)^T, \mathbf{v}_2 = (2.9, 2.2, 0.9)^T$ and their permutations. Since E is finite, co(S) is a polytope whose corners lie in E. Since neither of $\mathbf{v}_1, \mathbf{v}_2$ *majorizes* [1] the other, in fact all elements of E

[2] However, the fact that $\mathbf{v}_\alpha$ of (3) has this additional property does not imply that a PCFB exists, unless the number of channels is $M \leq 3$. *Majorization* [1] is an even stronger requirement, as illustrated by the vectors $(10,6,5,1)$ and $(9,8,3,2)$: The former is both greater and lesser than the latter and its permutations, depending on whether the entries of the former are in decreasing or increasing order respectively. However, neither of these two vectors majorizes the other.

are corners of $\text{co}(S)$. A PCFB does not exist because $\mathbf{v}_1, \mathbf{v}_2$ are not permutations of each other. Now consider the high bit-rate coding problem of [3]. Here the objective to be minimized over S is $\pi(\mathbf{v})$, the product of the coordinates of $\mathbf{v} \in S$. (As noted in [1], this is equivalent to minimizing an objective that is concave on $\text{co}(S)$.) Since $\pi(\mathbf{v}_1) = 6$ and $\pi(\mathbf{v}_2) = 5.742$, $\mathbf{v}_2$ is the minimum. However, the algorithm of Section 4.1 produces $\mathbf{v}_\alpha = \mathbf{v}_1$, and is thus suboptimum.

More generally, let $P \subseteq \text{co}(S)$ be the polytope whose corners are permutations of the vector $\mathbf{v}_\alpha$ of (3). Then $P = \text{co}(S)$ iff a PCFB exists. Now consider the function $f(\mathbf{v}) = -d(\mathbf{v}, P)$, where $d(\mathbf{v}, P) = \min\{\|\mathbf{v} - \mathbf{x}\| : \mathbf{x} \in P\}$ is the minimum distance from $\mathbf{v}$ to P using any valid norm $\|.\|$ on $\mathcal{R}^M$. It can be shown that f is a well-defined continuous concave function on $\mathcal{R}^M$. From the definition it is clear that (1) f has a constant value (zero) on P, and (2) if a PCFB does not exist, then P is actually the set of *maxima* of f over $\text{co}(S)$. Since the algorithm of Section 4.1 produces subband variance vector $\mathbf{v}_\alpha \in P$, it leads to the worst possible choice of FB's for an infinite family of such concave objectives f.

5. CONCLUDING REMARKS

We have described a FB design algorithm involving a sequential maximization of subband variances, that produces PCFB's *if they exist.* We have shown in an insightful way how the algorithm can be suboptimal in the absence of a PCFB. We have also shown how many FB optimization problems often become analytically intractable in absence of a PCFB. The methods used here suggest ways of *analytically proving nonexistence* of PCFB's for certain classes of FB's; as discussed in [8] in the context of nonuniform FB's.

6. REFERENCES

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

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

[3] P.P.Vaidyanathan, "Theory of Optimal Orthonormal Subband Coders," *IEEE Trans. SP,* vol. 46, no. 6, pp. 1528-1543, June 1998.

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

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

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

[7] A.Kirac and P.P.Vaidyanathan, "Theory and Design of Optimum FIR Compaction Filters," *IEEE Trans. SP,* vol. 46, no. 4, pp. 903-919, April 1998.

[8] S.Akkarakaran and P.P.Vaidyanathan, "General Results on Filter Bank Optimization with Convex Objectives, and the Optimality of Principal Component Filter Banks," *in preparation.*

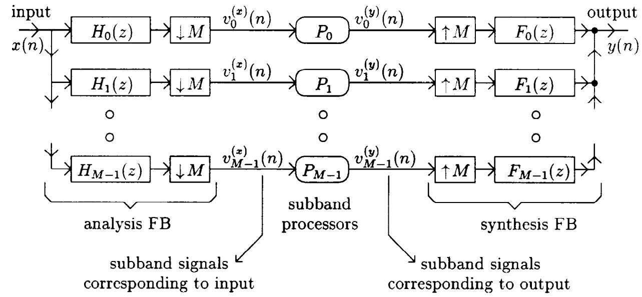

Figure 1: Generic FB based signal processing scheme.

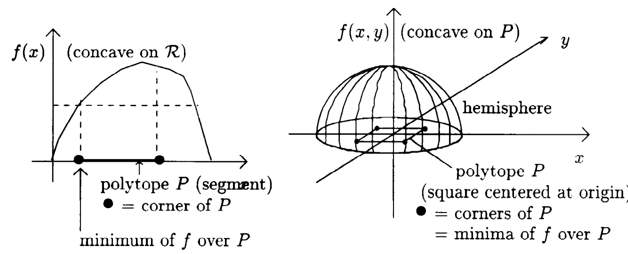

Figure 2: Optimality of corners of polytopes.

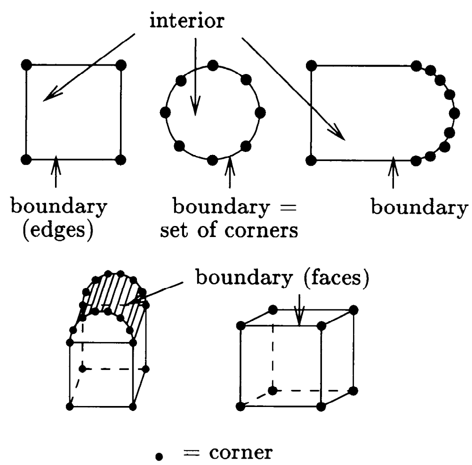

Figure 3: Corners and boundaries of compact convex sets.

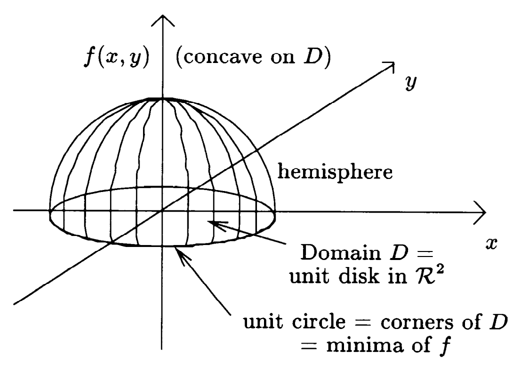

Figure 4: Optimality of corners of compact convex sets.