NON-CONVEX OPTIMIZATION FOR THE DESIGN OF SPARSE FIR FILTERS

Dennis Wei

Massachusetts Institute of Technology, Department of Electrical Engineering and Computer Science
77 Massachusetts Avenue, Cambridge, MA 02139, USA

ABSTRACT

This paper presents a method for designing sparse FIR filters by means of a sequence of p-norm minimization problems with p gradually decreasing from 1 toward 0. The lack of convexity for $p < 1$ is partially overcome by appropriately initializing each subproblem. A necessary condition of optimality is derived for the subproblem of p-norm minimization, forming the basis for an efficient local search algorithm. Examples demonstrate that the method is capable of producing filters approaching the optimal level of sparsity for a given set of specifications.

Index Terms— Sparse filters, non-convex optimization, FIR digital filters.

1. INTRODUCTION

Reducing the computational complexity of discrete-time filters has inspired a wide variety of approaches, e.g. [1–4]. This paper focuses on the design of sparse FIR filters, i.e., filters with relatively few non-zero coefficients. Sparse designs allow for the elimination of arithmetic operations corresponding to zero-valued coefficients, and may be incorporated in cascade structures such as those in [1–3] to yield even more efficient implementations. Sparsity is also of interest in the closely related problem of designing linear sensor arrays.

Designing a filter with maximal sparsity subject to a set of specifications is computationally difficult. While the problem can be solved using integer programming methods (e.g. [5]), the associated complexity can be prohibitive for problems with many coefficients. This has motivated research in efficient approximate methods directed at obtaining reasonably sparse but not necessarily optimal designs [6–8]. Of particular relevance to this work is [8], in which the 1-norm of the filter coefficients is minimized as part of the design algorithm. In the current paper, the approach of [8] is extended to the family of functions defined by

$$\|\mathbf{b}\|_p = \left(\sum_{n=1}^{M} |b_n|^p \right)^{1/p} \tag{1}$$

for $0 < p < 1$. It is convenient to refer to $\|\mathbf{b}\|_p$ as a *p-norm* for all $p > 0$ even though (1) defines a valid norm only for $p \geq 1$.

The p-norms have the desirable property of being an asymptotically exact measure of sparsity as p approaches zero. They do however pose their own difficulties for optimization as they are non-convex for $p < 1$. To mitigate the lack of convexity, a sequential optimization procedure is proposed in which p is slowly decreased from 1 toward 0. We present a simplex-like algorithm for solving the individual p-norm minimization problems, based on a vertex condition of optimality to be derived.

The approaches taken in [8] and in this work have parallels in the literature on sparse solutions of underdetermined systems of linear equations. For example, in compressive sensing both the 1-norm [9] and the p-norm for $p < 1$ [10] have been successfully applied, while the ideas of parameterized approximation and sequential optimization in [11] are similar to those in this work. However, as discussed in Section 2, the filter design problem differs significantly from the solution of underdetermined linear equations.

In Section 2, the problem of sparse filter design is formulated. Section 3 discusses a method for designing sparse filters involving a sequence of p-norm minimizations. The problem of p-norm minimization is analyzed in Section 4 and a necessary condition of optimality is given. Section 5 summarizes both our algorithm for p-norm minimization and the overall design algorithm. The performance of the algorithm is demonstrated through examples in Section 6.

2. PROBLEM FORMULATION

We focus on the design of causal, linear-phase FIR filters of length $N + 1$, for which the frequency response takes the form

$$H(e^{j\omega}) = e^{-j\omega N/2} \sum_{n=1}^{M} b_n T(n, \omega),$$

where $M = \lceil (N+1)/2 \rceil$, $T(n, \omega)$ is an appropriate trigonometric function, and the coefficients b_n are simply related to the impulse response (see [12] for details). We regard N as a fixed parameter representing the maximum allowable number of delays, with the understanding that the final design may require fewer than N delays if coefficients at the ends of the impulse response are zero.

We assume that the amplitude of $H(e^{j\omega})$ is chosen such that the maximum weighted error relative to the ideal frequency response $H_d(e^{j\omega})$ is no greater than a desired tolerance δ_d, i.e.,

$$W(\omega) \left| \sum_{n=1}^{M} b_n T(n, \omega) - H_d(e^{j\omega}) \right| \leq \delta_d \quad \forall\, \omega \in \mathcal{F}, \tag{2}$$

where $W(\omega)$ is a strictly positive weighting function and $\mathcal{F}$ is a closed subset of $[0, \pi]$. We approximate the infinite set of constraints in (2) by a finite subset corresponding to closely spaced frequencies $\omega_1, \omega_2, \ldots, \omega_K$.[1] As a result, (2) can be rewritten as a set of $2K$ linear inequalities in the coefficients b_n, and consequently the set of feasible coefficients is a polyhedron, to be denoted by P.

This work was supported in part by the MIT William Asbjornsen Albert Memorial Fellowship, the Texas Instruments Leadership University Program, and BAE Systems PO 112991.

[1] In our experience, it is sufficient to set $K \sim 10M$ and to distribute the frequencies $\omega_1, \ldots, \omega_K$ uniformly over $\mathcal{F}$ to ensure that (2) is satisfied. This is consistent with guidelines reported in [6, 7].

We use as a measure of complexity the number of non-zero coefficients, which corresponds exactly to

$$\|\mathbf{b}\|_0 \equiv \lim_{p\to 0} \|\mathbf{b}\|_p^p \tag{3}$$

with $\mathbf{b} = (b_1, b_2, \ldots, b_M)$. The function $\|\mathbf{b}\|_0$ is often referred to as the 0-norm for convenience despite not being a true norm. The problem of sparse filter design can be stated as

$$\min_{\mathbf{b}\in P} \quad \|\mathbf{b}\|_0. \tag{4}$$

Problem (4) differs from the problem of obtaining a sparse solution to an underdetermined system of linear equations. The latter has the form

$$\begin{aligned} \min_{\mathbf{x}} \quad & \|\mathbf{x}\|_0 \\ \text{s.t.} \quad & \mathbf{\Phi x} = \mathbf{y}, \end{aligned} \tag{5}$$

with $\dim(\mathbf{y}) < \dim(\mathbf{x})$, i.e., fewer constraints than variables. This contrasts with (4) in which the number of constraints $2K$ must be much larger than the number of variables M in order to yield a good approximation to (2). Moreover, the constraints in (5) are linear equalities as opposed to inequalities.

3. DESIGN USING P-NORM MINIMIZATION

In this section, we outline an approach to designing sparse filters that involves a sequence of p-norm minimizations with $0 < p \leq 1$. Our approach is based on the ability of the p-norms to approximate the 0-norm arbitrarily closely as seen in (3). We are thus led to consider problems of the form

$$\min_{\mathbf{b}\in P} \quad \|\mathbf{b}\|_p^p \tag{6}$$

for values of p approaching zero. We refer to a solution of (6) as a minimum p-norm solution, noting that the minimizer is not affected by replacing $\|\mathbf{b}\|_p^p$ with $\|\mathbf{b}\|_p$.

To further motivate the use of the p-norms, we discuss the two-dimensional example in Fig. 1. Consider first the case $p = 1$ in (6), as was done in [8]. The solution can be determined graphically by constructing the smallest ℓ^1 ball, which has a diamond shape, that intersects the feasible region, in this case at a vertex that does not correspond to a sparse solution. Now consider the same minimization for $p < 1$. As p decreases from 1, the boundaries of the ℓ^p ball curve inward and extend farther along the coordinate axes than they do elsewhere. Consequently, the solutions tend toward the axes and eventually converge to the true sparsest solution.

The behaviour seen in the preceding example is formalized in the following proposition.

Proposition 1. *Let $\{p^{(i)},\ i = 0, 1, \ldots\}$ be a sequence of positive numbers converging to zero, and $\{\mathbf{b}^{(i)}\}$ be a sequence of optimal solutions to the corresponding $p^{(i)}$-norm minimization problems* (6)*. Then every limit point of $\{\mathbf{b}^{(i)}\}$ is a global minimum of the* 0*-norm problem* (4)*.*

The proof is by contradiction and is omitted.

The problem of p-norm minimization (6) can be difficult when $p < 1$ since the objective function is non-convex. To mitigate the lack of convexity, we propose solving a sequence of p-norm minimizations as opposed to a single minimization, beginning with the case $p = 1$ and decreasing p gradually thereafter toward zero, e.g. according to

$$p^{(i)} = \alpha^i, \quad i = 0, 1, \ldots, \tag{7}$$

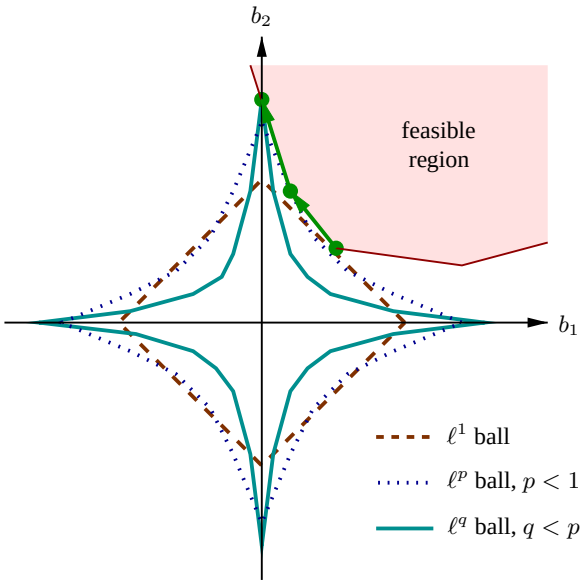

Fig. 1. Graphical representation of the minimization of various p-norms. The circles and arrows indicate the path traced by the optimal solutions.

where i is an index for the p-norm subproblems and α is slightly less than 1. For $p^{(0)} = 1$, (6) is a convex problem and can be solved efficiently to yield a global minimum. For $p^{(1)} = \alpha$, one might expect that an α-norm minimizer should be close in some sense to a 1-norm minimizer, and therefore the latter could be a promising initialization for obtaining the former. Generalizing this idea, a solution to subproblem i can be used to initialize subproblem $i + 1$. We note that a similar process of sequential optimization appears in [11]. It is conjectured that if α is close enough to 1, the sequence of solutions obtained using this initialization strategy will remain globally optimal for a significant range of p values.

4. ANALYSIS OF P-NORM MINIMIZATION

In this section, we present a more detailed analysis of the optimization problem in (6) with $p \in (0, 1]$, which is a recurring subproblem in the method outlined in Section 3. We first recast problem (6) into an equivalent form by expressing each coefficient b_n as the difference between two non-negative variables x_{2n-1} and x_{2n},

$$b_n = x_{2n-1} - x_{2n}, \quad x_{2n-1} \geq 0,\ x_{2n} \geq 0, \quad n = 1, \ldots, M. \tag{8}$$

Under the condition that $x_{2n-1}x_{2n} = 0$ for all n, the representation in (8) is unique and we also have

$$|b_n| = x_{2n-1} + x_{2n}, \quad n = 1, \ldots, M. \tag{9}$$

Using (8), (9) and (1), problem (6) can be transformed into

$$\begin{aligned} \min_{\mathbf{x}} \quad & F(\mathbf{x}) \\ \text{s.t.} \quad & W(\omega_k)\left|\sum_{n=1}^{M}(x_{2n-1} - x_{2n})T(n,\omega_k) - H_d(e^{j\omega_k})\right| \leq \delta_d, \\ & \qquad\qquad k = 1, \ldots, K, \\ & x_{2n-1} \geq 0,\ x_{2n} \geq 0, \quad n = 1, \ldots, M, \end{aligned} \tag{10}$$

where

$$F(\mathbf{x}) = \sum_{n=1}^{M} (x_{2n-1} + x_{2n})^p \,. \tag{11}$$

Problems (6) and (10) are equivalent in the sense that there is a one-to-one correspondence between their respective optimal solutions. The nonlinear constraints $x_{2n-1}x_{2n} = 0$, $n = 1, \ldots, M$, are not needed because it can be shown that they are automatically satisfied by all optimal solutions of (10). Hence the feasible set for problem (10) is also a polyhedron, which we denote by $\tilde{P}$ for convenience.

When $p = 1$, (10) is a linear programming problem and can be solved using standard techniques [13]. When $p < 1$, it can be verified that $F(\mathbf{x})$ in (11) is a concave function, and therefore $F(\mathbf{x})$ attains a minimum at a vertex of $\tilde{P}$ (see Prop. B.20 in [14]). The vertex condition can be strengthened somewhat as stated in Theorem 1. Theorem 1 also generalizes the usual condition that holds at a local minimum $\mathbf{x}^*$ (see [14]),

$$\nabla F(\mathbf{x}^*)'(\mathbf{x} - \mathbf{x}^*) \geq 0 \quad \forall\, \mathbf{x} \in \tilde{P}, \tag{12}$$

which may not apply in the case of problem (10) because the gradient $\nabla F(\mathbf{x})$ is not defined everywhere. The generalization of (12) can be stated in terms of the following definitions: Given a local minimum $\mathbf{x}^*$ of problem (10), define $\mathcal{N}$ and $\mathcal{Z}$ to be the sets of indices n for which $x^*_{2n-1} + x^*_{2n} > 0$ and $x^*_{2n-1} + x^*_{2n} = 0$ respectively. For an arbitrary vector $\mathbf{x}$, denote by $\mathbf{x}_\mathcal{N}$ the $2|\mathcal{N}|$-dimensional vector obtained by extracting from $\mathbf{x}$ the components x_{2n-1}, x_{2n} for all $n \in \mathcal{N}$. Let

$$F_\mathcal{N}(\mathbf{x}_\mathcal{N}) = \sum_{n \in \mathcal{N}} (x_{2n-1} + x_{2n})^p \,.$$

Also define $\tilde{P}_\mathcal{N}$ to be the restriction of $\tilde{P}$ to the hyperplane defined by $x_{2n-1} = x_{2n} = 0$, $n \in \mathcal{Z}$, i.e.,

$$\tilde{P}_\mathcal{N} = \left\{ \mathbf{x}_\mathcal{N} \,|\, \mathbf{x} \in \tilde{P};\ x_{2n-1} = x_{2n} = 0,\ n \in \mathcal{Z} \right\}.$$

Theorem 1 can now be stated as follows:

Theorem 1. *Let $\mathbf{x}^*$ be a local minimum of the problem in* (10) *with $0 < p < 1$. Then the following conditions hold:*

(a)

$$\nabla F_\mathcal{N}(\mathbf{x}^*_\mathcal{N})' (\mathbf{x}_\mathcal{N} - \mathbf{x}^*_\mathcal{N}) > 0 \quad \forall\, \mathbf{x}_\mathcal{N} \in \tilde{P}_\mathcal{N},\ \mathbf{x}_\mathcal{N} \neq \mathbf{x}^*_\mathcal{N}.$$

(b) $\mathbf{x}^*$ is a vertex of $\tilde{P}$.

The proof is omitted.

The vertex condition of optimality forms the basis for a simplex-like algorithm, described in the next section, that is directed at solving the problem of p-norm minimization in the case $p < 1$. We remark that Theorem 1 has the potential to be applied more broadly: statement (a) holds as long as the feasible set is convex, and both statements hold for any polyhedral feasible set. These results could be of use, for instance, in compressive sensing problems in which the measurement uncertainties are represented by linear inequalities.

5. DESCRIPTION OF ALGORITHM

Our overall algorithm for designing sparse FIR filters combines the sequential procedure of Section 3 with an algorithm for p-norm minimization to be described. For concreteness, we assume that p decreases according to (7). The sequential process terminates when p has decreased to an acceptably small value $p_{\min}$, or when the solution is deemed to have converged.

The case $p = p^0 = 1$ corresponds to a linear programming problem and hence any standard solver may be used.[2] For $p < 1$, we propose a local search algorithm in which the search is restricted to the vertices of the feasible polyhedron $\tilde{P}$, based on the optimality condition in Theorem 1. In each iteration, all vertices adjacent to the current vertex solution are searched for lower values of the objective function $F(\mathbf{x})$ in (11). If some of the adjacent vertices have lower objective values, the algorithm moves to the vertex with the lowest value and the search continues. Otherwise the algorithm terminates.

The local search algorithm is similar to the simplex method for linear programming in that it searches for lower function values by moving from one vertex to another along edges of the feasible polyhedron. As a consequence, the algebraic characterization of vertices and the procedure for moving between them are the same as in the simplex method, and are omitted for this reason. The interested reader is referred to linear programming texts (e.g. [13]).

In our experience with the sequential algorithm, the number of non-zero coefficients decreases more rapidly for p near 1 and less rapidly as p decreases. Around $p = 0.1$, the algorithm often converges to a solution that appears to be locally minimal for all smaller values of p. To determine if additional coefficients can be set to zero after convergence, we employ a re-optimization strategy loosely similar to the one in [8].

Denote by $\mathcal{Z}$ the set of indices n such that $b_n = 0$ in the final solution given by the sequential algorithm. The first step in the re-optimization is to minimize the maximum weighted error relative to $H_d(e^{j\omega})$ while constraining all coefficients with indices in $\mathcal{Z}$ to zero. This constrained minimax optimization can be formulated as

$$\begin{aligned} \min_{\delta, \mathbf{b}} \quad & \delta \\ \text{s.t.} \quad & W(\omega_k) \left| \sum_{n=1}^{M} b_n T(n, \omega_k) - H_d(e^{j\omega_k}) \right| \leq \delta, \quad k = 1, \ldots, K, \\ & b_n = 0, \quad n \in \mathcal{Z}. \end{aligned} \tag{13}$$

Once an optimal solution $(\delta^*, \mathbf{b}^*)$ is obtained, the index m corresponding to the coefficient b^*_m with the smallest magnitude is added to $\mathcal{Z}$, thus decreasing the number of non-zero coefficients, and (13) is re-solved. The process of zeroing the smallest coefficient and re-solving (13) continues until the maximum error exceeds δ_d, at which point the last feasible solution is taken to be the final design.

The re-optimization is occasionally able to generate one or two additional zero-valued coefficients after the p-norm algorithm converges. In addition, the re-optimized design almost always meets the frequency response constraints with a non-zero margin, i.e., the maximum error is strictly less than δ_d. Thus the final design usually satisfies the constraints in (2) at all frequencies and not just at the constrained frequencies $\omega_1, \ldots, \omega_K$.

The complexity of the overall algorithm is equivalent to a small number of M-dimensional linear programs. The equivalent number of linear programs depends on the number of subproblems (e.g. 5–10 with $\alpha = 0.98$ and $p_{\min} = 0.01$), but does not grow with M. The efficiency can be improved by exploiting the structure of the constraints in (2) in performing the required matrix inversions.

[2] To facilitate the initialization of the next subproblem, the solver should return a vertex solution, which is guaranteed to exist.

6. DESIGN EXAMPLES

In this section, we present a number of examples to illustrate the potential of our algorithm. For all examples, we use $\alpha = 0.98$ and $p_{\min} = 0.01$. The parameter N ranges from the number of delays required by the minimum-length Parks-McClellan design to 1.25 times that number.

In Example 1, we compare the algorithm in this work to the 1-norm algorithm using the example presented in [8]. The specifications are as follows: passband edge of 0.20π, stopband edge of 0.25π, passband ripple of 0.01 (linear) and stopband ripple of 0.1 (linear). The minimum-length Parks-McClellan design has 52 non-zero coefficients in its impulse response and requires 51 delays. The sparse design in [8] requires 41 non-zeros and 64 delays. Using our p-norm algorithm, the number of non-zeros is further reduced to 32, with 63 delays. Fig. 2 compares the impulse responses corresponding to the Parks-McClellan design and the design produced by our algorithm. The zero-valued coefficients in the sparse design tend to occur at locations corresponding to small coefficients in the Parks-McClellan design.

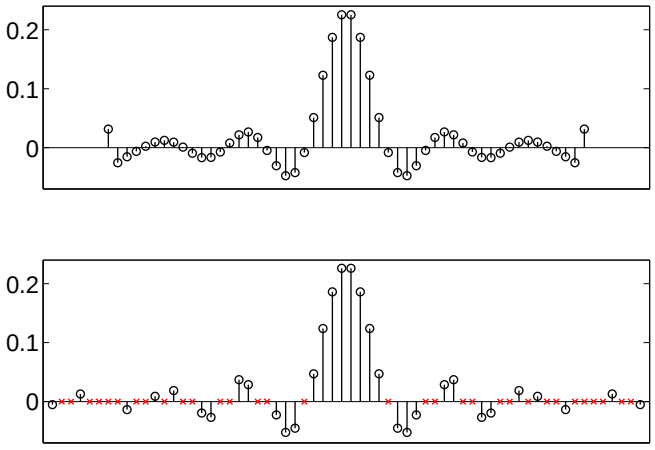

Fig. 2. Impulse responses corresponding to the Parks-McClellan design (top) and the p-norm design (bottom) for Example 1. Zero-valued coefficients are marked by x's.

We also compare our algorithm to an integer programming method, which is guaranteed to produce maximally sparse designs, using two of the examples in [5]. Table 1 lists the specifications for both filters. Table 2 summarizes the number of non-zero impulse response values and the number of delays resulting from the Parks-McClellan algorithm, the integer programming algorithm of [5], and our p-norm algorithm. The results indicate that our algorithm is capable of yielding reasonably sparse designs with significantly less complexity compared to integer programming. Note also that our design in Example 3 does not require any extra delays relative to the Parks-McClellan design.

	example 2	example 3
passband edge	0.4π	0.1616π
stopband edge	0.5π	0.2224π
passband ripple	0.2 dB	0.1612 dB
stopband attenuation	60 dB	34.548 dB

Table 1. Specifications for Examples 2 and 3.

example	algorithm	non-zeros	delays
2	Parks-McClellan	48	47
	integer programming	40	49
	p-norm	43	50
3	Parks-McClellan	56	55
	integer programming	44	57
	p-norm	46	55

Table 2. Results for Examples 2 and 3.

7. REFERENCES

[1] Y. Neuvo, C.-Y. Dong, and S. Mitra, "Interpolated finite impulse response filters," *IEEE Trans. Acoust., Speech, Signal Processing*, vol. 32, pp. 563–570, June 1984.

[2] Y. C. Lim, "Frequency-response masking approach for the synthesis of sharp linear phase digital filters," *IEEE Trans. Circuits Syst.*, vol. 33, pp. 357–364, Apr. 1986.

[3] R. J. Hartnett and G. F. Boudreaux-Bartels, "On the use of cyclotomic polynomial prefilters for efficient FIR filter design," *IEEE Trans. Signal Processing*, vol. 41, pp. 1766–1779, May 1993.

[4] M. Aktan, A. Yurdakul, and G. Dundar, "An algorithm for the design of low-power hardware-efficient FIR filters," *IEEE Trans. Circuits Syst. I*, vol. 55, pp. 1536–1545, July 2008.

[5] J. T. Kim, W. J. Oh, and Y. H. Lee, "Design of nonuniformly spaced linear-phase FIR filters using mixed integer linear programming," *IEEE Trans. Signal Processing*, vol. 44, pp. 123–126, Jan. 1996.

[6] J. L. H. Webb and D. C. Munson, "Chebyshev optimization of sparse FIR filters using linear programming with an application to beamforming," *IEEE Trans. Signal Processing*, vol. 44, pp. 1912–1922, Aug. 1996.

[7] D. Mattera, F. Palmieri, and S. Haykin, "Efficient sparse FIR filter design," in *Proc. ICASSP*, May 2002, vol. 2, pp. 1537–1540.

[8] T. A. Baran and A. V. Oppenheim, "Design and implementation of discrete-time filters for efficient rate-conversion systems," in *Proc. Asilomar Conf. Signals Syst. Comp.*, Nov. 2007.

[9] D. Donoho, "Compressed sensing," *IEEE Trans. Inform. Theory*, vol. 52, pp. 1289–1306, Apr. 2006.

[10] R. Chartrand, "Exact reconstruction of sparse signals via nonconvex minimization," *IEEE Signal Processing Lett.*, vol. 14, pp. 707–710, Oct. 2007.

[11] G. H. Mohimani, M. Babaie-Zadeh, and C. Jutten, "A fast approach for overcomplete sparse decomposition based on smoothed ℓ^0 norm," *IEEE Trans. Signal Processing*, vol. 57, pp. 289–301, Jan. 2009.

[12] A. V. Oppenheim, R. W. Schafer, and J. R. Buck, *Discrete-Time Signal Processing*, Prentice-Hall, Inc., Upper Saddle River, NJ, 1999.

[13] D. Bertsimas and J. N. Tsitsiklis, *Introduction to Linear Optimization*, Athena Scientific, Nashua, NH, 1997.

[14] D. P. Bertsekas, *Nonlinear Programming*, Athena Scientific, Belmont, MA, 1999.