Peak-Error-Constrained Sparse FIR Filter Design Using Iterative L_1 Optimization

Aimin Jiang*, Hon Keung Kwan†, Yanping Zhu‡, Xiaofeng Liu*, Ning Xu* and Xiao Yao*
*College of Internet of Things Engineering
Hohai University, Changzhou, China
Emails: jiangam@hhuc.edu.cn, {xfliubme, xuningdlts}@gmail.com, yaox@hhu.edu.cn
†Department of Electrical and Computer Engineering
University of Windsor, Windsor, Canada
Email: kwan1@uwindsor.ca
‡School of Information Science and Engineering
Changzhou University, Changzhou, China
Email: zhuyanping@cczu.edu.cn

***Abstract*—In this paper, a novel algorithm is presented for the design of sparse linear-phase FIR filters. Compared to traditional l_1-optimization-based methods, the proposed algorithm minimizes l_1 norm of a portion (instead of all) of nonzero coefficients. In this way, some nonzero coefficients at crucial positions are not affected by l_1 norm utilized in the objective function. The proposed algorithm employs an iterative procedure and the index set of these crucial coefficients is updated in each iteration. Simulation results demonstrate that the proposed algorithm can achieve better design results than both greedy methods and traditional l_1-optimization-based methods.**

***Index Terms*—FIR filters, iterative l_1 optimization, l_0 norm, linear program, sparsity.**

I. Introduction

Finite impulse response (FIR) filters are widely employed in various applications of signal processing [1], [2]. Traditional design methods mainly consider how to improve the approximation performance of an FIR filter given design specifications and many design problems can be formulated equivalently as convex optimization problems [3]. However, the implementation complexity of an FIR filter is seldom taken into account by traditional design methods.

Lately, sparse FIR filter designs draw much attention. A sparse FIR filter contains a number of zero coefficients (i.e., impulse responses). Arithmetic operations or circuit components corresponding to these coefficients then can be ignored, leading to a lower hardware cost and lower power consumption. Generally, the sparsity of an FIR filter can be evaluated by l_0 (quasi-)norm of filter coefficients. However, l_0 (quasi-)norm results in a combinatorial optimization problem. Optimal designs of sparse FIR filters could be obtained by traditional combinatorial optimization techniques [4], [5], [6]. However, their computational complexity is too high for large-scale design problems or applications where specifications need to be updated to accommodate the change of environments.

Using l_0 norm, some design problems of sparse FIR filters can be viewed as a special class of sparse coding problems. With the advances of sparse representation and compressive sensing [7], [8], optimization techniques originally developed for sparse coding are adopted in many algorithms to solve sparse FIR filter design problems. For instance, orthogonal matching pursuit (OMP) is utilized by [9]. However, because of essential difference between sparse coding and sparse FIR filter design, it is generally hard to obtain satisfying results by directly applying sparse coding techniques in sparse FIR filter design. Thus, more complicated searching strategies are required. Two heuristic algorithms are proposed in [10]. The first one, called the successive thinning algorithm, is based on two zero-coefficient selection rules, that is, the minimum-increase and the smallest-coefficient. In each iteration, one index is incorporated in the index set of zero coefficients until some approximation error constraints are violated. A regularized objective function, composed by l_1 norm of filter coefficients and the minimax or weighted least-squares (WLS) approximation error, is minimized in [11]. A real sparse design is obtained by hard thresholding after solving the regularized problem. The iterative l_1 optimization is proposed in [12]. Compared to traditional l_1 optimization, a weighting function updated in each iteration is introduced in [12] and larger weights are assigned to coefficients with smaller magnitudes. In [13], the iterative shrinkage/thresholding (IST) techniques are applied to tackle sparse FIR filter design problems. But different to the traditional IST, simpler subproblems constructed in each iteration are solved via their dual problems. It is proven that, under a sufficient condition, optimal solutions to the original subproblems can be obtained by solutions to dual problems. Using the similar design strategy, an efficient design algorithm is also developed in [14] for the design of sparse FIR filters subject to WLS approximation errors. Considering the fact that, given a filter order, minimizing the number of nonzero coefficients generally yields a sparse FIR filter of lower order, an iterative design algorithm is presented in [15], where both filter order and sparsity of an FIR filter are jointly optimized. The iterative-reweighted-least-squares (IRLS) algorithm is utilized to solve this joint optimization problem.

The remainder of the paper is organized as follows. In

978-0-9928-6265-7/16/$31.00

Section II, the proposed algorithm is developed. Simulation results are presented in Section III. Finally, we conclude the paper in Section IV.

II. Proposed Algorithm

For clarity of presentation, we only consider in this paper the design of sparse linear-phase FIR filters of type I with even order (odd length) and even symmetry. But design problems of sparse FIR filters of other types can also be tackled by the proposed algorithm with appropriate modifications. Let N represent the filter order of an FIR filter and $\mathbf{x} = [x_0\, x_1\, \ldots\, x_{\frac{N}{2}}]^T$ be filter coefficients to be optimized. Based on the symmetric structure of an FIR filter of type I, filter coefficients can be recovered by $h_n = h_{N-n} = x_n$ for $n = 0, \ldots, N/2$. To achieve the sparsest filter under given specifications, we formally cast the design problem as

$$\min_{\mathbf{x}} \quad \|\mathbf{x}\|_0 \tag{1a}$$

$$\text{s.t.} \quad \left|\mathbf{c}^T(\omega)\mathbf{x} - D(\omega)\right| \leq \delta(\omega), \quad \forall \omega \in \Omega_I \tag{1b}$$

where l_0 (quasi-)norm counts the number of nonzero coefficients, $\mathbf{c}(\omega)$ is defined by $\mathbf{c}(\omega) = \left[2\cos\frac{N}{2}\omega\ \ldots\ 2\cos\omega\ 1\right]^T$, $D(\omega)$ and $\delta(\omega)$ denote, respectively, the ideal magnitude response and the specified upper bound of magnitude approximation error. If a sparsity pattern of $\mathbf{x}$ is known, a potential solution to (1) can be attained by solving the following linear program

$$\min_{\mathbf{x},s} \quad s \tag{2a}$$

$$\text{s.t.} \quad \left|\mathbf{c}^T(\omega)\mathbf{x} - D(\omega)\right| \leq \delta(\omega) + s, \quad \forall \omega \in \Omega_I, \tag{2b}$$

$$x_n = 0,\ \forall n \in \mathcal{Z} \tag{2c}$$

where $\mathcal{Z}$ is a set of indices corresponding to zero elements of $\mathbf{x}$. If the optimal solution to (2) is nonpositive, this sparsity pattern is feasible to (1).

As l_1 norm is a convex function closest to l_0 norm, a classical technique to solve (1) is to replace $\|\mathbf{x}\|_0$ by its l_1 counterpart. However, the sparsity level of the optimal solution to (1) is highly affected by constraints (1b). If the given specifications are stringent, using l_1 norm generally cannot lead to optimal or even sparse solutions. Some other design methods choose indices of zero coefficients by heuristic criteria, such as, the minimum increase of approximation error in each iterative step [10]. However, there is no guarantee that adopting them could definitely lead to the optimal designs.

It is worth noting that classical l_1-optimization-based design methods aforementioned essentially assume that all the nonzero coefficients could be nullified and, thereby, their magnitudes need to be punished by their l_1 norm in the objective function. This design strategy works well when the sparsity of the optimal solution to (1) is extremely sparse. But if the given specifications are stringent, design results obtained by classical l_1 optimization could be far away from the optimal ones, since nonzero coefficients (especially those with large magnitudes) could be punished excessively by their l_1 norm. Furthermore, to satisfy the constraints imposed on magnitude approximation errors, filter coefficients at some crucial positions have to be nonzero and their magnitudes could be very large. In view of these points, it is actually unnecessary to incorporate all the nonzero coefficients in l_1 norm adopted by the objective function. In practice, however, it is difficult to identify the indices of these crucial coefficients. To overcome this difficulty, a novel greedy strategy based on the l_1 optimization is developed as follows. Let $\mathcal{Z}^{(t-1)}$ denote the index set of zero coefficients obtained in the previous iteration and $\mathcal{N}^{(t-1)}$ be the complement of $\mathcal{Z}^{(t-1)}$. They are initialized by $\mathcal{Z}^{(0)} = \emptyset$ and $\mathcal{N}^{(0)} = \{0, 1, \cdots, M\}$. Then, in each iteration, the following design problem is solved

TABLE I

Major steps of the proposed algorithm

Input:	Filter order N, desired magnitude response $D(\omega)$, and upper bound $\delta(\omega)$ of magnitude approximation error
Output:	Filter coefficients $\mathbf{x}$
Step 1:	Set $t = 0$ and initialize $\mathcal{Z}^{(0)}$ and $\mathcal{N}^{(0)}$. Then, solve (2) with $\mathcal{Z}^{(0)}$ to obtain $\mathbf{x}^{(0)}$.
Step 2:	Update $\mathcal{Z}^{(t)}$ and $\mathcal{N}^{(t)}$ according to (4) and (5).
Step 3:	Set $t = t + 1$. Then, construct $\mathcal{T}^{(t)}$ and solve (3) to obtain $\mathbf{x}^{(t)}$.
Step 4:	If the convergence condition is satisfied, solve (2) using the final $\mathcal{Z}^{(t)}$ and return the obtained design result; Otherwise, go to Step 2.

$$\min_{\mathbf{x}} \quad \|\mathbf{x}_{\mathcal{T}^{(t)}}\|_1 \tag{3a}$$

$$\text{s.t.} \quad \left|\mathbf{c}^T(\omega)\mathbf{x} - D(\omega)\right| \leq \delta(\omega), \quad \forall \omega \in \Omega_I \tag{3b}$$

$$x_n = 0,\ \forall n \in \mathcal{Z}^{(t-1)} \tag{3c}$$

where $\mathbf{x}_{\mathcal{T}^{(t)}} = [x_{j_1}, x_{j_2}, \cdots, x_{j_T}]^T$ is a sub-vector of $\mathbf{x}$ and $\mathcal{T}^{(t)} = \{j_n \in \mathcal{N}^{(t-1)}, n = 1, \cdots, T\}$. Generally, the number of elements of $\mathbf{x}_{\mathcal{T}^{(t)}}$ should be less than $\left|\mathcal{N}^{(t-1)}\right|$. Let $\mathbf{x}^{(t)}$ be the solution achieved by solving (3). If there exist zeros in $\mathbf{x}^{(t)}$, both $\mathcal{Z}^{(t-1)}$ and $\mathcal{N}^{(t-1)}$ are updated by

$$\mathcal{Z}^{(t)} = \mathcal{Z}^{(t-1)} \bigcup \left\{n | x_n^{(t)} = 0, n \in \mathcal{N}^{(t-1)}\right\}, \tag{4}$$

$$\mathcal{N}^{(t)} = \mathcal{N}^{(t-1)} - \left\{n | x_n^{(t)} = 0, n \in \mathcal{N}^{(t-1)}\right\}. \tag{5}$$

It can be observed that, although all the nonzero coefficients $\{x_n\}_{n\in\mathcal{N}^{(t-1)}}$ are optimized in (3), only a portion of nonzero coefficients $\{x_n\}_{n\in\mathcal{T}^{(t)}}$ take part in the construction of the objective function, while filter coefficients $\{x_n\}_{n\in\mathcal{N}^{(t-1)}-\mathcal{T}^{(t)}}$ are optimized to ensure that magnitude approximation error constraints (3b) are satisfied.

Table I summaries the major steps of the proposed algorithm. Two comments regarding the proposed algorithm are given below.

1) Indices in $\mathcal{T}^{(t)}$ has to be appropriately selected in each iteration, as they highly affect design results obtained in the subsequent iterations. Inspired by l_1 norm, $\mathcal{T}^{(t)}$ consists of the indices of filter coefficients with the T

TABLE II
SPECIFICATIONS OF EXAMPLE 1

Passband region	$[0,\ 0.3\pi]$
Stopband region	$[0.5\pi,\ \pi]$
Filter order N	60, 70, and 80
Passband magnitude	Within ± 0.001 dB of unity
Stopband magnitude	Below -60, -65, -70, -75, and -80 dB

smallest magnitudes in our proposed algorithm. This selection rule implies that

$$\left|x_{j_1}^{(t-1)}\right| \leq \cdots \leq \left|x_{j_T}^{(t-1)}\right| \leq \left|x_k^{(t-1)}\right|, \forall k \in \mathcal{N}^{(t-1)} - \mathcal{T}^{(t)}. \quad (6)$$

To determine $\mathcal{T}^{(1)}$, the design problem (2) with $\mathcal{Z}^{(0)} = \emptyset$ has to be solved.

2) The proposed algorithm terminates the iterative procedure when

$$\frac{\left\|\mathbf{x}^{(t)} - \mathbf{x}^{(t-1)}\right\|_2}{\left\|\mathbf{x}^{(t-1)}\right\|_2} \leq \varepsilon \quad (7)$$

where $\|\cdot\|_2$ represents l_2 norm of a vector and ε is a prescribed threshold. In general, if the sparsity level of filter coefficients no longer increases, the iterative procedure would converge in several iterations, as the index set $\mathcal{T}^{(t)}$ does not vary with t. After the convergence of the iterative procedure, the equiripple design can be further obtained by tackling (2) using the final $\mathcal{Z}^{(t)}$.

III. SIMULATIONS

In this section, two design examples are presented to demonstrate the effectiveness of the proposed algorithm. In our designs, parameter ε is chosen as 10^{-6}. Constraints (2b) and (3b) are imposed on 501 discrete frequency points within $[0, \pi]$. Both (2) and (3) are linear programs, which are solved by SeDuMi [16].

In the first example, a set of lowpass filters are designed using specifications given in Table II. In all the designs of this example, parameter T is always chosen equal to 2. Numbers of zero coefficients for each pair of specific filter order N and the minimum stopband magnitude attenuation are summarized in Table III. To evaluate its performance, the proposed algorithm is compared to the successive thinning algorithm [10] in which two zero-coefficient selection rules are suggested, that is, the smallest-coefficient and the minimum-increase. It can be observed that the proposed algorithm can obtain FIR filters with more zero coefficients in most designs. Magnitude responses of FIR filters obtained by three design methods are presented in Fig. 1, where filter order and the minimum stopband magnitude attenuation are chosen, respectively, equal to 60 and -70 dB. Impulse responses of each FIR filter are depicted in Fig. 2. Due to their symmetric structure, only half of impulse responses are shown in Fig. 2. Note that, most of zero coefficients are located towards both ends of impulse responses of a linear-phase FIR filter. Thus, the resulting filters could be considered as lower-order filters.

In the second example, we employ the proposed algorithm to design bandpass filters using various N and $\delta(\omega)$. The

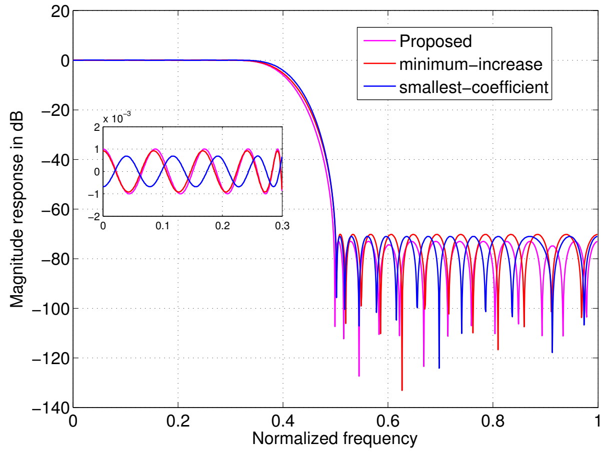

Fig. 1. Magnitude responses of FIR filters obtained in Example 1.

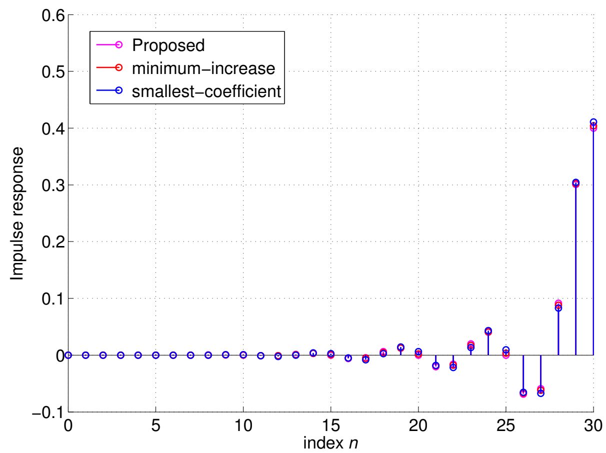

Fig. 2. Impulse responses of FIR filters obtained in Example 1.

detailed specifications are given in Table IV. In this example, we set T equal to 6. The proposed algorithm is compared to the iterative reweighted l_1 (IRL1) design algorithm [12], which iteratively solves the following problem

$$\min_{\mathbf{x},\gamma} \quad \gamma + \mu \left\|\mathbf{W}^{(l)}\mathbf{x}\right\|_1 \quad (8a)$$

$$\text{s.t.} \quad \left|\mathbf{v}_N^T(\omega)\mathbf{x} - D(\omega)\right| \leq \gamma, \quad \forall \omega \in \Omega_I, \quad (8b)$$

$$\gamma \leq \delta, \quad (8c)$$

$$x_i = 0, \quad \forall i \in \mathcal{Z}^{(l)} \quad (8d)$$

where μ is a regularization parameter always set equal to 1, $\mathbf{W}^{(l)}$ is a diagonal matrix whose diagonal elements are updated in each iteration by $W_i^{(l)} = \left[\left|x_i^{(l-1)}\right| + \sigma\right]^{-1}$, and σ takes a small value to avoid the division by zero. Problem formulation (8) shows that the IRL1 algorithm also works with the l_1 optimization. But it aims to minimize the weighted l_1 norm of all the nonzero coefficients. Design results of both algorithms are summarized in Table V, which shows that the proposed algorithm achieves much better results when only considering in each iteration the sparsity of a portion of

TABLE III
NUMBER OF ZERO COEFFICIENTS OF SPARSE FIR FILTERS OBTAINED IN EXAMPLE 1

Stopband magnitude level (dB)	Proposed			Minimum-increase [10]			Smallest-coefficient [10]		
	N=60	N=70	N=80	N=60	N=70	N=80	N=60	N=70	N=80
−60	24	34	44	22	32	42	22	32	42
−65	22	32	42	18	28	38	16	26	36
−70	22	32	42	18	28	38	16	20	30
−75	20	30	40	16	26	36	16	26	36
−80	20	30	40	20	30	40	20	30	40

TABLE IV
SPECIFICATIONS OF EXAMPLE 2

Maximum filter order N	160, 180, 200, 220 and 240
Passband region	$[0.3\pi, 0.4\pi]$
Stopband region	$[0, 0.25\pi] \cup [0.5\pi, \pi]$
Upper bound $\delta(\omega)$ of magnitude approximation error	Below −60, −70, −80, −90 and −100 dB

TABLE V
NUMBER OF ZERO COEFFICIENTS OF SPARSE FIR FILTERS OBTAINED IN EXAMPLE 2

N	160	180	200	220	240
$\delta(\omega)$	−60 dB	−70 dB	−80 dB	−90 dB	−100 dB
Proposed	48	40	30	28	20
IRL1 [12]	42	32	20	16	12

nonzero coefficients. Impulse responses of FIR filters obtained using $N = 160$ and $\delta(\omega) = -60$ dB are given in Fig. 3. As already observed in Example 1, sparse FIR filters generally have lower filter orders.

IV. CONCLUSIONS

In this paper, we develop a novel algorithm for the design of sparse linear-phase FIR filters based on the iterative l_1 optimization. Different to classical l_1 optimization techniques, the proposed algorithm aims to minimize in each iteration l_1 norm of a portion of nonzero coefficients. This design strategy is inspired by the fact that some nonzero coefficients located at crucial positions cannot be nullified and minimizing their magnitudes could drive design results away from the optimal solutions. In practice, a small number of filter coefficients with the smallest magnitudes are considered as ones that tend to be zeros. Simulation results demonstrate that the proposed algorithm can achieve better designs than both classical l_1-optimization-based algorithms and greedy algorithms using more complicated zero-coefficient selection rules.

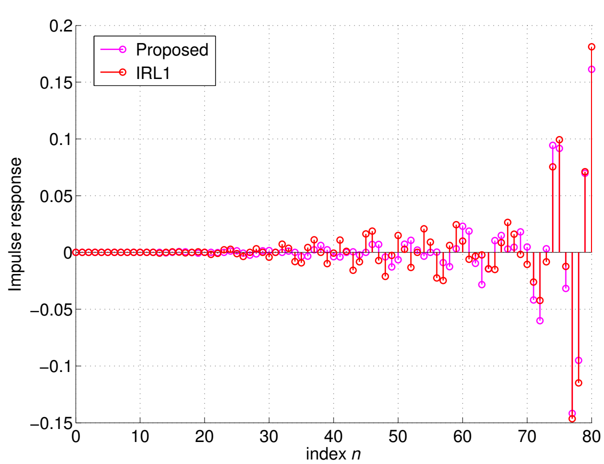

Fig. 3. Impulse responses of FIR filters obtained in Example 2.

ACKNOWLEDGMENT

This work was supported in part by the Fundamental Research Funds for the Central Universities of China under grants 2011B11214, 2011B11114, 2012B07314, and 2012B04014, the National Nature Science Foundation of China under grants 61101158, 61471157, and 61401148, and the Natural Science Foundation of Jiangsu Province, China under grants BK20130238, BK20141159 and BK20141157.

REFERENCES

[1] A. Antoniou, *Digital Filters: Analysis, Design, and Applications*, 2nd Ed. New York: McGraw-Hill, 2000.
[2] T. W. Parks and C. S. Burrus, *Digital Filter Design*. New York: John Wiley & Sons, 1987.
[3] T. N. Davidson, "Enriching the art of FIR filter design via convex optimization," *IEEE Signal Processing Magazine*, vol. 27, pp. 89-101, May 2010.
[4] 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 Process.*, vol. 44, pp. 123-126, Jan. 1996.
[5] O. Gustafsson, L. S. DeBrunner, V. DeBrunner, and H. Johansson, "On the design of sparse half-band like FIR filters," in *41st Asilomar Cons. Signal, Cyst., Comp.*, Pacific Grove, USA, Nov. 2007, pp. 1098-1102.
[6] Y.-S. Song and Y. H. Lee, "Design of sparse FIR filters based on branch-and-bound algorithm," in *Proc. 40th Midwest Simp. Circuits Cyst.*, vol. 2, Sacramento, USA, Aug. 1997, pp. 1445-1448.
[7] M. Glad, *Sparse and Redundant Representations: From Theory to Applications in Signal and Image Processing*. New York, USA: Springer, 2010.
[8] Y. C. Elder and G. Kutyniok, *Compressive Sensing: Theory and Applications*. New York, USA: Cambridge University Press, 2012.
[9] D. Matters, F. Palmer, and S. Hyaline, "Efficient sparse FIR filter design," in *Proc. IEEE Int. Cons. Accost., Speech, Signal Process.*, 2002, vol. 2, pp. 1537–1540.
[10] T. Baron, D. Wei, and A. V. Oppenheim, "Linear programming algorithms for sparse filter design," *IEEE Trans. Signal Process.*, vol. 58, pp. 1605-1617, Mar. 2010.
[11] W.-S. Lu and T. Hinamoto, "Digital filters with sparse coefficients," in *Proc. IEEE Int. Symp. Circuits Syst.*, Paris, France, May 2010, pp. 169-172.
[12] C. Rusu and B. Dumitrescu, "Iterative reweighted l_1 design of sparse FIR filters," Signal Process., vol. 92, pp. 905-911, Apr. 2012.
[13] A. Jiang, H. K. Kwan, and Y. Zhu, "Peak-error-constrained sparse FIR filter design using iterative SOCP," *IEEE Trans. Signal Process.*, vol. 60, pp. 4035-4044, Aug. 2012.
[14] A. Jiang and H. K. Kwan, "WLS design of sparse FIR digital filters," *IEEE Trans. Circuits Syst. I*, vol. 60, pp. 125-135, Jan. 2013.

[15] A. Jiang, H. K. Kwan, Y. Zhu, X. Liu, N. Xu, and Y. Tang, “Design of sparse FIR filters with joint optimization of sparsity and filter order,” *IEEE Trans. Circuits Syst. I*, vol. 62, pp. 195-204, Jan. 2015.

[16] J. F. Sturm, “Using SeDuMi 1.02, a MATLAB toolbox for optimization over symmetric cones,” *Optim. Methods Softw.*, vol. 11/12, pp. 625–653, 1999.