FIR (Sparse) Filter Design Using IST Algorithm

Niraja Singh
Department of Electronics and Communication Engineering
Radharaman Institute of Technology and Science, Bhopal, Madhya Pradesh, India

Pushpraj Tanwar
Department of Electronics and Communication Engineering
Radharaman Institute of Technology and Science, Bhopal, Madhya Pradesh, India

Abstract-The objective of the sparse FIR filter design problem considered in this paper is to reduce the number of nonzero valued coefficients. The proposed design method is subject to the iterative shrinkage/thresh holding(IST) algorithms, which are used in sparse and redundant representation for signals. To reduce the number of nonzero valued filter coefficient, subject to a least square (weighted) approximation error constraints imposed on frequency domain. The design algorithm is to transfer the original non-convex problems to a series of constrained sub-problems in a simpler form, despite of their non-convexity, solved by a numerical approach developed by in this paper.

Index Term: IST, WLS, Sparse filter, FIR filter and digital filters

I. INTRODUCTION

Sparseness in this filter design means a filter with the majority of the coefficients being zero. In this paper, we concern sparse FIR filter designs, which aim to find FIR filters with as few nonzero-valued coefficients as possible. If a sparse FIR filter is attained, the multipliers corresponding to the zero-valued coefficients can be omitted, which consequently reduce the implementation complexity. Theoretically speaking, we can locate optimal positions of zero-valued impulse responses by exhaustive search. However, this is impossible due to its huge computational complexity when the filter order is large. In practice, it is more realistic to find locally optimal designs. The major difficulty of local search is to determine which filter coefficients could impact the performance of an FIR filter less than some others. In [12], orthogonal matching pursuit (OMP) algorithm is employed to determine sparse solutions to equiripple filter design problems, which are recast as a linear system of equations. This algorithm is belonged to greedy methods. Compared with exhaustive search, the computational complexity of the OMP design algorithm is dramatically reduced. In [13], the branch-and-bound algorithm is applied to the design of sparse FIR filters. Following the depth-first search, filter coefficients to be zeros are gradually incorporated in sub-problems until the given specifications are violated. A half-band FIR filter design algorithm is presented in [14], where by introducing auxiliary variables the sparse design problem is formulated as a mixed integer linear programming (MILP) problem and then solved by branch-and-bound or branch-and-cut technique. Although the design algorithms in [13] and [14] employ intelligent search to locate optimal solutions, the computational complexity is still high especially when designing FIR filters with high orders. Using linear programming (LP) as building blocks, two heuristic approaches are proposed in [15] to design sparse FIR filters. The first approach iteratively thins nonzero-valued coefficients. In each iterative step, one or more coefficients are forced to zeros using either the minimum-increase or smallest-coefficient rule. The second approach is to determine the sparsest filter by solving an approximate problem with an -norm objective function. Based on the similar idea of the second approach in [15], a two-phase design algorithm is proposed in [16]. A convex design problem with an -norm regularization term is first solved. Then, the hard thresh holding operation is applied to force coefficients with small magnitudes equal to 0. Finally, a post-processing is employed to further refine the design result. Similar design strategy is also applied in sparse IIR filter design.

Many of design approaches described above only consider linear-phase FIR filter designs subject to peak error constraints imposed on magnitude response. However, some of them can be readily extended to handle general FIR filter designs. In this paper, we present an iterative shrinkage/thresh holding (IST)-based algorithm for sparse FIR filters designs subject to a WLS approximation error constraint. Both linear- and nonlinear-phase FIR filter designs are handled under a unified framework. The basic idea behind the proposed design algorithm is to successively transform the original non convex design problem to a series of sub-problems in a simpler form. Although these sub-problems are still non-convex, they can be efficiently solved by a numerical approach developed in this paper. Furthermore, by means of the Lagrange dual problems of such sub-problems, it will be demonstrated in Section C that the obtained solutions are optimal to their respective sub-problems.

The paper is organized as follows. The sparse FIR filter design problem is first presented in Section B. A novel iterative design method is then proposed in Section C. Several design examples are presented in. Finally, conclusions are drawn in Section D.

II. PROBLEM FORMULATION

Let Hd(ω)be a given desirable frequency response and ΩI shows all frequencies. A dense grid of frequency points $\omega_i \ (i = 1,2 \ldots\ldots K)$ are sampled over ΩI. The original WLS design problem can be expressed by

$$\min_h \frac{1}{K} \sum_{\omega_i \in \Omega_I} W(\omega_i) |H(e^{j\omega_i}) - \mathrm{Hd}(\omega_i)|^2 \qquad (1)$$

The filter coefficient vector is defined by

$$h = [h_0 \ h_1 \ \ldots\ldots \ h_N]^T \qquad (2)$$

And $W(\omega) \geq 0$ is a given weighting function. In (2), the superscript represents the transpose operation of a matrix or vector. In this paper, we only consider design problems of FIR filters with real-valued coefficients. However, the proposed algorithm can be readily extended to handle design problems of FIR filters with complex-valued coefficients Using h, the frequency response $H(e^{jw})$ of an FIR filter can be computed by

$$H(e^{jw}) = V_N^T(\omega) h \qquad (3)$$

Where

$$V_N(\omega) = [1 \ He^{-jw} \ \ldots\ldots \ e^{-jNw}]^T, \qquad (4)$$

It is known that design problem (1) is convex and can be equivalently cast as

$$\min_h \|\mathrm{F}h - \mathrm{d}\|_2^2 \qquad (5)$$

Were

$$\mathrm{F} = \begin{bmatrix} \sqrt{\frac{W(\omega_1)}{K}} \mathrm{Re}\{V_N^T(\omega_1)\} \\ \sqrt{\frac{W(\omega_1)}{K}} \mathrm{Im}\{V_N^T(\omega_1)\} \\ \vdots \\ \sqrt{\frac{W(\omega_K)}{K}} \mathrm{Re}\{V_N^T(\omega_K)\} \\ \sqrt{\frac{W(\omega_K)}{K}} \mathrm{Im}\{V_N^T(\omega_K)\} \end{bmatrix} \qquad (6)$$

$$\mathrm{d} = \begin{bmatrix} \sqrt{\frac{W(\omega_1)}{K}} \mathrm{Re}\{\mathrm{Hd}(\omega_1)\} \\ \sqrt{\frac{W(\omega_1)}{K}} \mathrm{Im}\{\mathrm{Hd}(\omega_1)\} \\ \vdots \\ \sqrt{\frac{W(\omega_K)}{K}} \mathrm{Re}\{\mathrm{Hd}(\omega_K)\} \\ \sqrt{\frac{W(\omega_K)}{K}} \mathrm{Im}\{\mathrm{Hd}(\omega_K)\} \end{bmatrix} \qquad (7)$$

and $\|X\|_2$ denotes the Euclidean norm of a vector $X \epsilon \mathrm{R}^n$. In (6) and (7), Re{.}and Im{.}are operators used to retrieve real and imaginary parts of a complex value, respectively. The optimal solution $h_{opt}$of (5) can be obtained by solving a set of linear equations or using some convex optimization solver [9] or [10] if additional convex constraints are incorporated. For convenience of latter discussion, we designate δ_{opt} as the optimal objective value of the WLS design problem (5).

Based on (5), the sparse FIR filter design problem under consideration is expressed by

$$\min_{h} \quad \|h\|_0 \qquad (8a)$$

$$s.t. \quad \|Fh - d\|^2_0 \le \delta_d \qquad (8b)$$

Where $\|X\|_0$ denotes the l_0 norm of x, which is equal to the number of nonzero-valued elements of x. It should be mentioned here that the norm is essentially a semi-norm as it does not satisfy the property of positive scalability. In (8), δ_d is a specified upper bound of the WLS approximation error. Without loss of generalization, we assume that

$$\delta_d = \rho\delta_{opt}, \quad \rho \ge 1 \qquad (9)$$

We can achieve a sparser design by a larger ρ, while a lower WLS approximation error by a smaller ρ. The design problem (8) is highly non-convex and, in general, it is difficult to definitely obtain its optimal solutions. Note that if we appropriately choose F, h and d (8) can also be applied to formulate sparse linear-phase FIR filter design problems. For instance, in order to design an Nth-order Class I linear-phase FIR filter, F and h can be chosen, respectively, as

$$F = \begin{bmatrix} 2\sqrt{\frac{W(\omega_1)}{K}}\cos\frac{N}{2}\omega_1 & \cdots & 2\sqrt{\frac{W(\omega_1)}{K}}\cos\omega_1 & 1 \\ \vdots & \ddots & \vdots & \vdots \\ 2\sqrt{\frac{W(\omega_K)}{K}}\cos\frac{N}{2}\omega_K & \cdots & 2\sqrt{\frac{W(\omega_K)}{K}}\cos\omega_K & 1 \end{bmatrix}$$

$$h = [h_0 \;\; h_1 \;\dots\; h_{\frac{N}{2}}]^T$$

Correspondingly, d is composed by the desired magnitude responses sampled on the frequencies $\omega_i, \; i = 1, 2, \dots, K$, . For linear-phase FIR filter designs, the WLS approximation error constraint (8b) only needs to be imposed on the magnitude response instead of the frequency response, as the group delay of a linear-phase FIR filter is constant over the whole frequency band.

The non-convexity of (8) can be overcome by replacing the l_0 norm of h by its l_1 norm in the objective function. Although it is a good approximation of (8), the l_1 norm design problem does not directly lead to a real sparse solution. In the next section, we shall introduce an efficient design algorithm to solve (8), which can yield sparse filters directly.

III. ITERATIVE DESIGN METHOD

One promising way to efficiently solve (8) is to mix its l_0 norm objective function and the l_2 norm constraint in the form

$$f(h) = \|Fh - d\|_2^2 + \mu\|h\|_0 \qquad (10)$$

where the positive regularization parameter μ controls the tradeoff between the approximation error and the sparsity of filter coefficients. Many numerical approaches have been proposed to tackle this unconstrained optimization problem. Among them, the IST algorithms have drawn much attention due to their computational efficiency. The IST algorithms were proposed independently by several authors. The major advantage of the IST algorithms is that in each iterative step of the IST algorithms the optimization can be decomposed into a set of independent scalar optimization problems, which generally have closed-form solutions, such that large-scale problems can be efficiently resolved. Readers are referred to [7] and [8] for a more thorough review of the IST algorithms and their variants. The proposed algorithm is based on a design strategy similar to that of the IST algorithms. However, instead of minimizing the regularized objective function $f(h)$, the proposed algorithm directly deals with (8) so as to avoid the dilemma of choosing the regularization parameter μ. The proposed algorithm contains an iterative design procedure, which is developed in Section C. In each iteration, a sub-problem is constructed using the solution obtained in the previous iterative step. It will be shown that although this sub-problem is still non-convex, its globally optimal solution can be efficiently obtained in each iterative step by a numerical method presented in Section-C.

IV. Iterative Design Procedure

The proposed design algorithm adopts an iterative procedure to design sparse FIR filters. The initial point $h^{(0)}$ can be obtained by solving the original WLS design problem (5). In the $(k+1)$-th iterative step, we construct a constrained sub-problem other than the regularized one. Its objective function is in the same form as in (8). Constraint (8b) is modified as

$$\|Fh - d\|_2^2 + s\left(h, h^{(k)}\right) \leq \delta_d \tag{11}$$

Where $h^{(k)}$ the design result is in the last iterative step and the additional term $s\left(h, h^{(k)}\right)$ is defined by

$$s\left(h, h^{(k)}\right) = c\left\|h - h^{(k)}\right\|_2^2 - \left\|Fh - Fh^{(k)}\right\|_2^2$$

$$(h - h^{(k)})^T \ (cI - F^TF) \ (h - h^{(k)}) \tag{12}$$

In (12), is chosen such that $s\left(h, h^{(k)}\right)$ is always convex, which further implies

$$c \geq \lambda_{max}(F^TF) \tag{13}$$

Where $\lambda_{max}(A)$ represents the maximal eigenvalue of a symmetric matrix. It should be mentioned that under (13) the term $s\left(h, h^{(k)}\right)$ is always nonnegative. Thus, the feasibility domain defined by (11) is contained within the one defined by (8b). In order to make the restricted feasibility domain as large as possible, c should be chosen as $\lambda_{max}(F^TF)$.

After some manipulations, constraint (11) can be rewritten by

$$\left\|h - b^{(k)}\right\|_2^2 \leq u^{(k)} \tag{14}$$

Where

$$b^{(k)} - \frac{1}{c}F^T\left(d - Fh^{(k)}\right) + h^{(k)} \tag{15}$$

$$u^{(k)} = \frac{1}{c}\delta_d - \frac{1}{c^2}\left(Fh^{(k)} - d\right)^T (cI - FF^T) \left(Fh^{(k)} - d\right) \tag{16}$$

In (16), represents an identity matrix whose size can be determined in context. Using (11) or, equivalently, (14), we can reformulate the sub-problem in each iterative step as

$$\min_h \qquad \|h\|_0 \tag{17}$$

$$\text{s.t.} \quad \left\|h - b^{(k)}\right\|_2^2 \leq u^{(k)} \tag{17b}$$

Note that (17) cannot be further decomposed to a set of scalar optimization problems. We shall introduce an efficient design method in the next subsection to solve (17).

Let $h^{(k+1)}$ denote an optimal solution to (17). The following proposition shows that by appropriately selecting an initial point for the iterative design procedure, the feasibility domain of (17) is nonempty. Then, if in each iteration a global solution to (17) can be always achieved, we can find a local solution of the original design problem (8).

Proposition 1: If

$$\|Fh^{(0)} - d\|_2^2 \le \delta_d \qquad (18)$$

the feasibility domain defined by (14) is nonempty for any $k \ge 0$.

Proof: First of all, we show that for some specific k, if $\|Fh^{(k)} - d\|_2^2 \le \delta_d$ is satisfied, is nonnegative.

Since

$$\|Fh^{(k)} - d\|_2^2 \le \delta_d \qquad (19)$$

And

$$\lambda_{max}(cI - FF^T) = c - \lambda_{max}(FF^T) \le c,$$

We have

$$(Fh^{(k)} - d)^T (cI - FF^T) (Fh^{(k)} - d) \le c\delta_d$$

And consequently

$$u^{(k)} \ge \frac{1}{c}\delta_d - \frac{1}{c^2}c\delta_d = 0$$

If $u^{(k)} \ge 0$, there exists at least one feasible point (e.g. $b^{(k)}$,) for the design problem (17). Combined with the fact that the optimal solution $h^{(k+1)}$ of (17) also satisfies (19), in all the successive iterative steps, the feasibility domain defined by (14) is always nonempty. If the iterative procedure starts from a given $h^{(0)}$ which satisfies (18), then the feasibility domain defined by (14) is nonempty for any $k \ge 0$.

In our designs, the iterative design procedure continues until the following condition is satisfied

$$\frac{\|h^{(k+1)} - h^{(k)}\|_2}{\|h^{(k)}\|_2} \le \epsilon \qquad (20)$$

or the number of iterations is larger than a predefined number MaxIter. Although the convergence of the proposed design algorithm has not been strictly guaranteed, in all the designs we have conducted, the iterative procedure can always converge to final solutions.

After the iterative procedure described above, we may further reduce the WLS approximation error by solving a WLS design problem similar to (5)

$$\min_h \|Fh - d\|_2^2 \qquad (21a)$$

$$s.t. \quad h_n = 0, \ _n\in \qquad (21b)$$

In (21b),

is a subset of indices at which h_n is identified to be zero by the iterative procedure described before. The nonzero valued coefficients of the optimal solution h^* to (21) can be computed by

$$h^* = (F^T F)^{-1} F^T d \qquad (22)$$

where F is obtained by extracting from F the columns corresponding to the complement set of

Note that if

is an empty set, (22) is also the optimal solution to (5).

TABLE I
SPECIFICATIONS OF DESIGN

Passband region	0.15705
Stopband region	0.27406
Filter order N	90
Parameter ρ	2~6

of a linear-phase FIR filter is constant over the whole frequency band, we only measure the (weighted) L_2 and L_∞ errors on magnitude response of linear-phase FIR filters. All designs presented in this section are conducted on a desktop computer with an AMD Ph II quad core processor of 3.00 GHz.

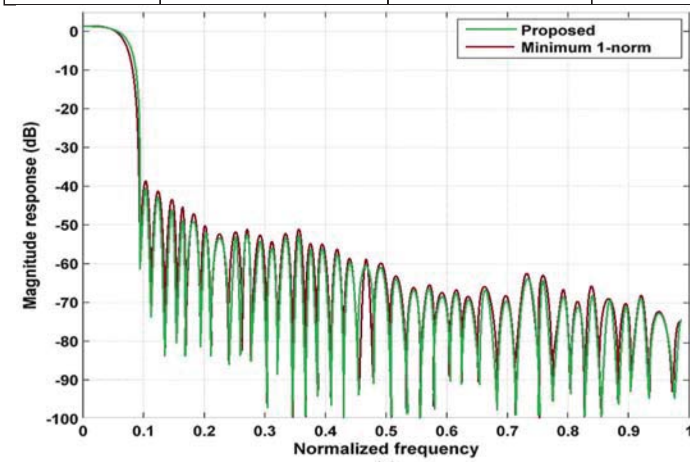
Example 1

The first example is taken from [4], where the objective is to determine optimal weights for a uniform linear beam former. In essence, this design problem can be equivalently translated to a linear-phase FIR filter design problem, whose specifications are given in Table I. For comparison, we also employ the successive thinning and minimum 1-norm algorithms of 4] to design FIR filters. Although these two algorithms are originally proposed for sparse FIR filter designs with peak error constraints, they can be readily extended to deal with WLS designs. Designs with various within [2, 6] are conducted in this example. Note that when $\rho = 2$ the obtained solution is optimal to the WLS design problem (5). All the design results are summarized in Table II. For $\rho = 2.5$, magnitude and impulse responses of the linear phase FIR filters designed, respectively, by the proposed algorithm, and the successive thinning and minimum 1-norm algorithms of [4] are depicted in Fig. 2. Due to the symmetric structure of a linear-phase FIR filter, only half of impulse responses are shown in Fig. 2(c). In each design, the minimum-increase and smallest-coefficient selection rules of the successive thinning algorithm yield the same design result. Thereby, we only plot in Fig. 2 the magnitude and impulse responses of the
FIR filter obtained by the minimum-increase rule of the successive thinning algorithm.
It can be observed that the proposed algorithm can achieve design results comparable to or better than those of the successive thinning algorithm in most of designs. The proposed algorithm attains sparser designs than the successive thinning algorithm when $\rho = 2.5$, 5.0, and 6.0 at a cost of higher magnitude approximation errors. For $\rho = 4.0 \text{ and } 4.5,$ FIR filters designed by these two algorithms have the same number of zero-valued coefficients. However, the design results obtained by the proposed algorithm have lower peak errors but higher L_2 approximation errors. In the designs of $\rho = 3.0,\ 3.5 \text{ and } 4.5,$ both the proposed algorithm and the successive thinning algorithm can achieve the same results. Compared with the minimum 1-norm algorithm, the proposed algorithm yields much better designs in terms of the sparsity of obtained FIR filters.

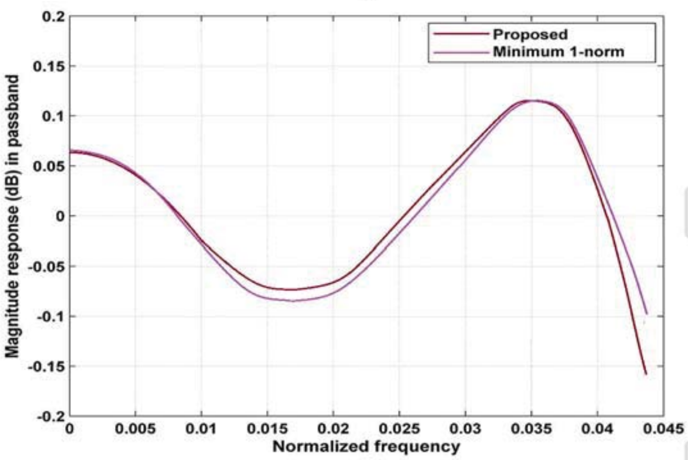
For the design with $\rho = 2.5$, we also plot in Fig. 3 the variation of $\|h^{(k+1)} - h^{(k)}\|_2 / \|h^{(k)}\|_2$ with respect to the iteration number . It can be observed that the proposed algorithm depicted in Fig. 1 is repeated for two times. During the second stage, the curve decreases monotonically with respect to k. Nevertheless, several leaps appear in the curve of the first stage. A large number of experiments reveal that these leaps generally happen with the change of the number or indices of zero-valued coefficients.

TABLE II NUMBER OF ZERO-VALUED COEFFICIENTS AND APPROXIMATION ERRORS FOR DESIGN EXAMPLE

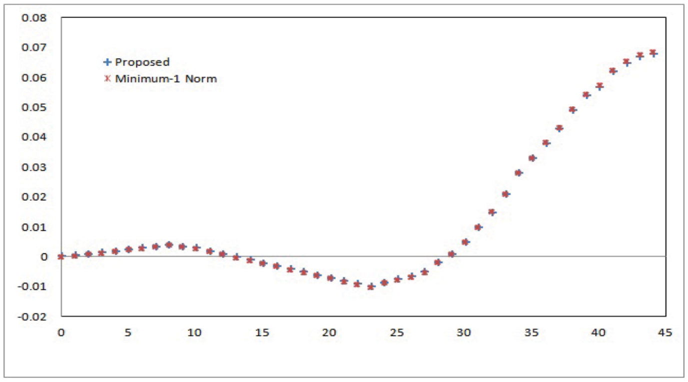
ρ	Proposed algorithm			Minimum 1-norm algorithm [21]		
	Sparsity	$e_{L2}$of MR (X10-4)	$e_{L\infty}$of MR (X10-2)	Sparsity	$e_{L2}$of MR (X10-4)	$e_{L\infty}$of MR (X10-2)
2.0	0	0.643	1.970	0	0.643	1.970
2.5	10	0.962	2.922	6	0.739	2.031
3.0	12	1.079	3.020	10	1.072	2.580
3.5	14	1.310	3.223	12	1.173	2.408
4.0	16	1.591	3.270	14	1.471	2.745
4.5	18	2.011	3.638	16	1.920	3.199
5.0	20	2.451	4.358	16	1.920	3.202
5.5	20	2.451	4.369	18	2.435	3.699
6.0	22	2.901	4.454	18	2.435	3.699

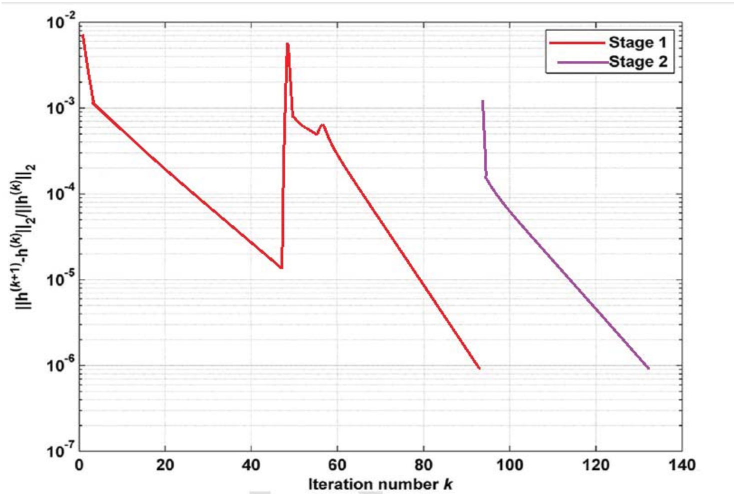


(a)



(b)





V. CONCLUSION

A novel sparse FIR filter design algorithm in the WLS sense has been presented in this paper. In order to tackle the nonconvexity of the original design problem incurred by the l 0–norm objective function, an iterative design procedure is developed in Section C. In each iterative step, by use of the design result attained in the last iteration, a constrained sub-problem is first constructed. Furthermore, it is demonstrated that the design result obtained is optimal to each sub-problem. Design results obtained by the iterative procedure is finally refined bysolving (22). The design procedure can be repeated for several times to further improve the design performance. It is worth noting that the proposed algorithm is essentially independent on the problem formulation. It can be applied to solve any design problem, which can be formulated in the form of (8). Simulation results reveal that the proposed algorithm can achieve design results comparable to or better than those of some other popular sparse filter design approaches.

REFERENCES

[1] D. Mattera, F. Palmieri, and S. Haykin, “Efficient sparse FIR filter de- sign,” in Proc. IEEE Int. Conf. Acoust., Speech, Signal Process.,2002, vol. 2, pp. 1537–1540.
[2] Y.-S. Song and Y. H. Lee, “Design of sparse FIR filters based on branch-and-bound algorithm,” in Proc. 40th Midwest Symp. Circuits Syst., Aug. 1997, vol. 2, pp. 1445–1448.
[3] O.Gustafsson,L.S.DeBrunner,V.DeBrunner,andH.Johansson,“On the design of sparse half-band like FIR filters,” in Proc. 41st Asilomar Conf. Signal, Syst., Comp., Nov. 2007, pp. 1098–1102
[4] T. Baran, D. Wei, and A. V. Oppenheim, “Linear programming algo- rithms for sparse filter design,” IEEE Trans. Signal Process., vol. 58, pp. 1605–1617, Mar. 2010.
[5] W.-S. Lu and T. Hinamto, “Digital filters with sparse coefficients,” in Proc. IEEE Int. Symp. Circuits Syst., Paris, France, May 2010, pp.
[6] W.-S. Lu and T. Hinamto, “Minimax design of stable IIR filters with sparse coefficients,” in Proc. IEEE Int. Symp. Circuits Syst., Rio de Janeiro, Brazil, May 2011, pp. 398–401.
[7] M. Zibulevsky and M. Elad, “L1–L2 optimization in signal and image processing,”IEEESignalProcess.Mag.,vol.27,pp.76–88,May2010.

[8] M. Elad, M. A. T. Figueiredo, and Y. Ma, “On the role of sparse and redundant representations in image processing,” Pro. IEEE, vol. 98, pp. 972–982, Jun. 2010.

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

[10] K. C. Toh, M. J. Todd, and R. H. Tutuncu, “SDPT3-a MATLAB soft- ware package for semidefinite programming,” Optim. Methods Softw., vol. 11, pp. 545–581, 1999.

[11] W.-S.LuandT.-B.Deng,“Animprovedweightedleast-squaresdesign for variable fractional delay FIR filters,” IEEE Trans. Circuits Syst. II, Analog Digit. Signal Process., vol. 46, no. 8, pp. 1035–1040, Aug. 1999.

[12] W.J.OhandY.H.Lee,“Design of efficient FIRfilters with cyclotomic polynomial prefilters using mixed integer linear programming,” IEEQ21`ZSaE Signal Process. Lett., vol. 3, pp. 239–241, Aug. 1996.

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

[14] J. Adams and A. Willson, Jr., “A new approach to FIR digital filters with fewer multipliers and reduced sensitivity,” IEEE Trans. Circuits Syst., vol. CAS-30, pp. 277–283, May 1983.

[15] J. Adams and A. Willson, Jr., “Some efficient digital prefilter structures,” IEEE Trans. Circuits Syst., vol. CAS-31, pp. 260–266, Mar. 1984.

[16] Y. C. Lim and S. R. Parker, “FIRfilter design over a discrete power-of-two coefficient space,” IEEE Trans. Acoust., Speech, Signal Process., vol. ASSP-31, pp. 583–591, Jun. 1983.