DESIGN OF FIR FILTERS WITH SPARSE IMPULSE RESPONSES

*Ryo Matsuoka and Masahiro Okuda**

Univ. of Kitakyushu, Faculty of Environmental Engineering Kitakyushu, Japan

ABSTRACT

In this paper, we present a numerical algorithm for the design of FIR filters with sparse impulse responses. Our method minimizes the number of nonzero entries in the impulse response together with the least squares error of its frequency response. We show that the FIR filters with sparse coefficients can outperform a conventional least squares approach and the Parks-McCllelan method under the condition of the same number of multipliers.

***Index Terms*—** FIR filter design, coefficient sparsity, l_0 approximation,

1. INTRODUCTION

The design of FIR filters is an important issue in digital signal processing. Many design methods have been proposed by a number of authors [1]-[12]. Especially the least-squares (LS) method, which minimizes the mean squared error of frequency responses, is widely used due to its simplicity and flexibility [7]-[12].

Most of conventional filter design approaches aim to minimize a filter order and maximize the filter performance. Even though the FIR filter coefficients designed by the LS method is optimal in the least squares sense, it is not necessarily optimal among the set of filters with the same number of multipliers, that is, less mean squared error can be achieved by a filter that has the same number of multipliers, but has longer impulse response with some zero-valued entries. To minimize the number of multipliers instead of the filter order, some approaches design the filter with zero-valued taps, which is often called sparse filters [13]-[14] (Fig.1). In [14], an efficient design method is proposed. However the method needs to design filters iteratively, and high computational effort is required for high order filters.

In this paper, we present a numerical approach for designing the sparse filters. Our method consists of two steps. In the first step, we find the position of zero-valued coefficients using the l_2-l_0 optimization, and then in the second step the filter coefficients are determined while keeping the coefficients determined by the first step fixed to zero. Our method does not guarantee optimality in the sense of sparsity, but has better performance than the conventional filter design methods.

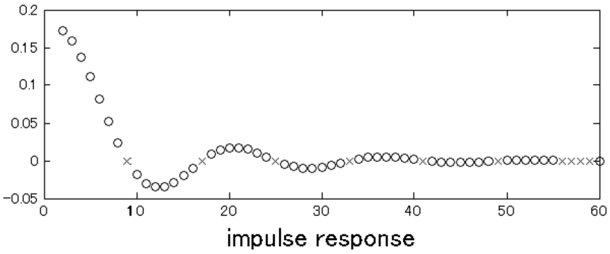

Fig. 1. Impulse response of sparse filters: 'x' indicates zero coefficients.

In Section 2, the basic theory of the least squares method for the FIR filters is briefly described. In Section 3, our design problem is formulated and a design algorithm that considers the sparsity of coefficients is proposed. In Section 4, several examples are illustrated to verify the validity of the proposed algorithm, and a comparison with the conventional methods are shown. In the last section, we make comments on the advantages of the algorithm.

2. CONVENTIONAL WEIGHTED LEAST-SQUARES METHOD

The magnitude response $H(\omega)$ of the linear phase FIR filters and more general non-linear phase FIR filters can be written as a linear combination of trigonometric basis functions

$$H(\omega) = \sum_{n=0}^{N-1} a_n \phi_n(\omega), \tag{1}$$

where, for example, $\phi_n(\omega) = \cos(n\omega)$ for even-order symmetric linear phase filters, and $\phi_n(\omega) = e^{jn\omega}$ for more general formula.

Our design method can handle all types of the FIR filters expressed by (1), however due to the limited space, we show only the case of the even-oder symmetric linear phase FIR filters.

We here repeat $H(\omega)$ for the even-order symmetric filter:

$$H(\omega) = \sum_{n=0}^{N-1} a_n \cos(n\omega), \tag{2}$$

*We are grateful for the support of Japan Society for the Promotion of Science and KDDI Foundation.

where $N = (N_0 - 1)/2 + 1$ and N_0 is its filter length.

Here we briefly review the conventional weighted LS (WLS) method. Generally, the mean squared error of Eq.(2) over the interval $[0, \pi]$ is defined as

$$\Phi_1 = \frac{1}{\pi} \int_0^{\pi} W(\omega)|H(\omega) - D(\omega)|^2 d\omega, \tag{3}$$

where $W(\omega)$ is a weight function which is not identically zero and has positive values, and $D(\omega)$ is a desired frequency response. The optimal filter coefficients $\{a_n\}_{n=0}^{N-1}$ in the LS sense can be uniquely determined by solving the normal equation:

$$\boldsymbol{Q}\boldsymbol{a} = \boldsymbol{p}, \tag{4}$$

where

$$\boldsymbol{a} = [a_0 \ \ a_1 \ \ \cdots a_{N-1}]^T$$

$$\boldsymbol{p}_m = \frac{1}{\pi} \int_0^{\pi} W(\omega)D(\omega)\cos(m\omega)d\omega \tag{5}$$

$$\boldsymbol{Q}_{m,n} = \frac{1}{\pi} \int_0^{\pi} W(\omega)\cos(m\omega)\cos(n\omega)d\omega, \tag{6}$$

where $\boldsymbol{p}_m$ and $\boldsymbol{Q}_{m,n}$ are m-th and (m, n)-th elements of $\boldsymbol{p}$ and $\boldsymbol{Q}$, respectively.

When $D(\omega)$ and $W(\omega)$ are simple functions, Eqs. (5) and (6) can be easily calculated in a closed form. However, commonly, the integrals of Eqs.(5) and (6) are difficult to derive if both $D(\omega)$ and $W(\omega)$ are arbitrary. In particular, when the weight function is given by an error response, as in Lawson's algorithm (discussed later), we can never calculate the integral in Eqs.(5) and (6). Therefore, in practice, it is efficient to define the following cost function which is expressed as the finite sum of the errors on the discretized frequency points.

$$\Phi_2 = \frac{1}{L} \sum_{l=0}^{L-1} W(\omega_l)|H(\omega_l) - D(\omega_l)|^2 \tag{7}$$

We denote it by using the l_2 norm of the error

$$\Phi_2^{1/2} = ||\boldsymbol{W}(\boldsymbol{R}\boldsymbol{a} - \boldsymbol{d})||_2, \tag{8}$$

where the (l, k)-th elements of $\boldsymbol{R}$ is

$$\boldsymbol{R}_{l,k} = \cos(k\omega_l) \tag{9}$$

and

$$\boldsymbol{d} = [D(\omega_0) \ \ D(\omega_2) \cdots \ \ D(\omega_{L-1})]^T \tag{10}$$

$$\boldsymbol{W} = diag\{W(\omega_0) \ \ W(\omega_2) \cdots \ \ W(\omega_{L-1})\} \tag{11}$$

In this case, $\boldsymbol{p}_m$ and $\boldsymbol{Q}_{m,n}$ of Eqs.(5) and (6) in the normal equation are rewritten as follows.

$$\boldsymbol{p}_m = \frac{1}{L} \sum_{l=0}^{L-1} W(\omega_l)D(\omega_l)\cos(m\omega_l) \tag{12}$$

$$\boldsymbol{Q}_{m,n} = \frac{1}{L} \sum_{l=0}^{L-1} W(\omega_l)\cos(m\omega_l)\cos(n\omega_l) \tag{13}$$

3. DESIGN ALGORITHM

The conventional WLS method is optimal in the LS sense under the condition that the filter length is fixed. If one allows a longer filter length with some zero coefficients, it can achieve better approximation than the non-sparse filters with the same number of multipliers. Our goal is to design such sparse filters. The proposed design algorithm consists of two steps as shown below:

1. The positions of filter coefficients to be zero is determined. We accomplish it to solve the sparse approximation problem (discussed in Sec.3.1).
2. The zero-valued coefficients in the previous step are fixed to zero and the filter coefficients are found by using a conventional least squares design (Sec.3.3).

3.1. Sparse Approximation

Here we define $||\boldsymbol{a}||_0$ as the number of non-zero elements in $\boldsymbol{a}$ (It is often called zero-norm or l_0-norm, despite not satisfying the properties of the norm). Our aim is to approximate the desired frequency response by the linear phase FIR filter with minimum number of coefficients. To fulfill it, we define the filter design problem as a process to minimize the cost function:

$$\min_{\boldsymbol{a}} \frac{1}{2}||\boldsymbol{W}(\boldsymbol{R}\boldsymbol{a} - \boldsymbol{d})||_2^2 + \beta||\boldsymbol{a}||_0, \tag{14}$$

where $\boldsymbol{d}$ is a desired frequency response and the parameter β is introduced to control the balance between the error of the filter and the number of the coefficients. Since the second term of the cost function (14) is a discrete metric, the problem is inherently difficult to solve by conventional methods such as gradient decent based optimization.

We adopt a method based on the variable splitting and quadratic penalty, which is recently applied to image restoration problems [15], [16], [17]. Introducing an auxiliary parameter $\boldsymbol{s}$ that corresponds the coefficients $\boldsymbol{a}$, the problem (14) is equivalent to the following minimization problem:

$$\min_{\boldsymbol{a},\boldsymbol{s}} \quad \frac{1}{2}||\boldsymbol{W}(\boldsymbol{R}\boldsymbol{a} - \boldsymbol{d})||_2^2 + \beta||\boldsymbol{s}||_0, \quad \text{subject to} \quad ||\boldsymbol{a} - \boldsymbol{s}||_2^2 = 0 \tag{15}$$

Instead of solving (15), we add the penalty term $||\boldsymbol{a} - \boldsymbol{s}||_2^2$ to the cost. In the end, our design problem is stated as

$$\min_{\boldsymbol{a},\boldsymbol{s}} \frac{1}{2}||\boldsymbol{W}(\boldsymbol{R}\boldsymbol{a} - \boldsymbol{d})||_2^2 + \beta||\boldsymbol{s}||_0 + \frac{\gamma}{2}||\boldsymbol{a} - \boldsymbol{s}||_2^2, \tag{16}$$

where γ is a weighting parameter that controls similarity between $\boldsymbol{a}$ and $\boldsymbol{s}$. The strategy of the variable splitting approach is that, we start with initial values of $\boldsymbol{s}^0$, and then repeatedly solve two sub-problems:

1.

$$\boldsymbol{a}^k = \arg\min_{\boldsymbol{a}} \frac{1}{2}\|\boldsymbol{W}(\boldsymbol{R}\boldsymbol{a} - \boldsymbol{d})\|_2^2 + \frac{\gamma}{2}\|\boldsymbol{a} - \boldsymbol{s}^{k-1}\|_2^2 \tag{17}$$

2.

$$\boldsymbol{s}^k = \arg\min_{\boldsymbol{s}} \beta\|\boldsymbol{s}\|_0 + \frac{\gamma}{2}\|\boldsymbol{a}^k - \boldsymbol{s}\|_2^2 \tag{18}$$

3. $\gamma^{k+1} = \mu \cdot \gamma^k, \ \ (\mu > 1)$

4. $k = k + 1;$

The first sub-problem (17) has a simple quadratic form. The solution is determined by solving

$$(\boldsymbol{R}^T\boldsymbol{W}^2\boldsymbol{R} + \gamma^k\boldsymbol{I})\boldsymbol{a}^k = \boldsymbol{R}^T\boldsymbol{W}^2\boldsymbol{d} + \gamma^k\boldsymbol{s}^{k-1} \tag{19}$$

The optimal solution of the second sub-problem (18) is found for each coefficient individually, that is, the problem is equivalent to the minimization of the function $E(s_n)$ for $n = 0, 1, \cdots, N-1$ as follows

$$\min_{s_n} \quad E(s_n) = \beta C(s_n) + \frac{\gamma}{2}(a_n - s_n)^2, \quad (n = 0, 1, \cdots, N-1) \tag{20}$$

where s_n is an element in $\boldsymbol{s}$, and the function $C(s_n)$ has 1 if $s_n \neq 0$ and 0 otherwise (we omit the superscript k). In the end, (18) is minimized when

$$s_n^* = \begin{cases} 0, & a_n^2 < 2\beta/\gamma \\ a_n, & \text{otherwise} \end{cases} \tag{21}$$

(for derivation, see Appendix)

3.2. Balance control

In the algorithm, the desired number of non-zero coefficients (denoted by N_d) is specified by a user. The parameter β in the minimization problem (16) determines the number of non-zero coefficients. However it is difficult to explicitly formulate the relationship with β and N_d. We adopt a heuristic approach, in which β adaptively changes in the iterations according to the number of non-zero coefficients. If the actual number of non-zero coefficients in an iteration is larger than N_d, β is increased to $r_u \cdot \beta$ $(r_u > 1)$, otherwise β is decreased to $r_l \cdot \beta$ $(r_l < 1)$, where r_u, and r_l are newly introduced scaling parameters. When the number of the coefficients reaches N_d, then β is fixed. The algorithm starts with large values of r_u and r_l, then those are gradually decreased.

3.3. Constrained Filter Design

The aim of the algorithm in Sec.3.1 is to find the positions for the filter coefficients to be zero. Once the positions are determined, we re-design the filter by the constrained least squares approach to achieve an optimal filter:

$$\boldsymbol{a}^* = \arg\min_{\boldsymbol{a}} \Phi_2 \quad s.t.\ a_n = 0 \ \ (\forall n \in \mathcal{S}),$$

where $\mathcal{S}$ is the set of the positions of zero coefficients. The solution for the problem is given by solving the normal equation.

4. EXAMPLES AND COMPARISON

In this section, several numerical examples are shown to verify the advantage of the proposed algorithm. All examples were designed in MATLAB. All frequencies are normalized by π and frequency points are equally spaced.

Example 1: Least Squares Filter

A low-pass filter with a narrow transition band was designed. The filter length N_0 is 1059 ($N = 530$) and the number of non-zero coefficients is $N_d = 859$. The passband and stopband edges $\omega_p = 0.05$ and $\omega_s = 0.055$. The result is compared to two conventional least squares approaches:

1. (LS1) Design a filter of length N_d by the conventional LS method.
2. (LS2) Design a filter of length N_0 by the conventional LS method and then force the $N_0 - N_d$ smallest coefficients zero.

Fig.2 illustrates the log-magnitude response of filters designed the (a) LS1 and (b) the proposed method. The mean squared errors of LS1, LS2 and the proposed method were $4.48 \cdot 10^{-3}$, $4.34 \cdot 10^{-3}$, $3.08 \cdot 10^{-3}$, respectively.

We tested hundreds of design examples, and all examples converge to filters with smaller errors than ones of the conventional LS methods (LS1 and LS2). Some of them are listed in Table 1.

Table 1. Squared error(ω_p: passband edge, ω_s: stopband edge, N: filter length, N_d: # of non-zero coeffs., all the three filters listed have same number of non-zero coefficients)

$(\omega_p, \omega_s, N_0, N_0 - N_d)$,	LS1	LS2	Proposed method
(0.2, 0.26, 99, 40)	2.00e-2	3.42e-2	1.24e-2
(0.1, 0.14, 199, 40)	2.23e-4	7.13e-4	1.46e-4
(0.1, 0.14, 199, 80)	3.60e-3	1.35e-2	2.00e-3
(0.1, 0.11, 459, 100)	1.47e-2	1.24e-2	9.69e-3
(0.03, 0.035, 1199, 400)	7.38e-3	1.16e-2	4.81e-3
(0.05, 0.052, 2059, 1000)	0.183	0.142	0.111

Example 2: Chebyshev approximation

In this example, we apply the proposed method to the Chebyshev approximation, and compare it with the Parks-McClellan (PM) algorithm [3].

In the algorithm modified Lawson's algorithm [12] is used for the Chebyshev approximation, in which the WLS problems are solved iteratively. In each iteration the weighting function is updated by

$$W_{k+1}(\omega_l) = W_k(\omega_l)\frac{W_0(\omega_l)E_k^{env}(\omega_l)}{\sum_i W_0(\omega_l)E_k^{env}(\omega_l)}, \quad (22)$$

where $E_k^{env}(\omega)$ is the piecewise-linear envelope function of the error (for detail, see [12]).

The algorithm for the Chebyshev approximation is stated as follows.

1. The initial weight $W^0(\omega)$ is given and start with $W^1(\omega) = W^0(\omega)$.
2. Solve the sparse approximation in Sec.3.1.
3. The weight is updated by (22).
4. If it converges, then go to Step 5, otherwise go back to Step 2.
5. Fix the zero-valued coefficients, the conventional Lawson's algorithm [12] is performed to re-design the filter.

We design the filter, whose passband and stopband edges are $\omega_p = 0.1$ and $\omega_s = 0.13$, respectively.

We compare our results with the PM method that guarantees its optimality for the non-sparse filters[1]. We adjust the initial weighting function $W^0(\omega)$ to obtain the same amount of passband ripples, and then compare the stopband ripple with the PM method. Fig.3 shows the results of the designed filter with $N = 259$ and $N_d = 139$. Table 2 gives some of numerical design examples. Our method outperformes PM method by 2-8 dB in the attenuation. According to the paper [14], the sparse filter [14] also increase the level of attenuation by 2-8 dB over the PM method. However the method [14] needs iterative design of the optimal filter, and if the length is increased, the number of iteration will become much larger. On the other hand, in our algorithm the design of 300-tap equiripple filter needs only a few seconds to converge with Intel Core i7 2.93GHz CPU.

Appendix

This step is equivalent to hard thresholding in the shrinkage algorithm. When $s_n \neq 0$, we have

$$E(s_n) = \beta + \frac{\gamma}{2}(a_n - s_n)^2$$

and then $E(s_n)$ has the minimum value

$$E(a_n) = \beta$$

[1]executed in MATLAB using 'firpm.m'

Table 2. Chebyshev Approximation ($\omega_p = 0.1$, $\omega_s = 0.13$, N: filter length, N_d: # of non-zero coeffs., the length of the PM filter is N_d, p.r.: maximum passband ripple[dB], s.a.: minimum stopband attenuation[dB])

	PM method		Proposed Method	
(N, N_d)	p.r.	s.a.	p.r.	s.a.
(159, 79)	3.12e-2	25.2	3.12e-2	30.1
(199, 99)	1.60e-2	27.9	1.60e-2	35.9
(239, 119)	8.75e-3	33.1	8.75e-3	41.2
(259, 139)	5.53e-3	37.6	5.53e-3	45.1
(319, 179)	2.33e-3	48.9	2.33e-3	52.6

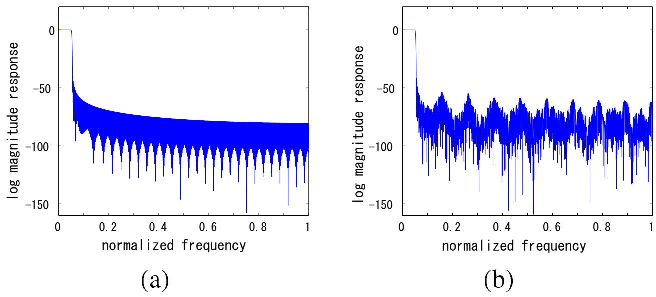

Fig. 2. Example 1 (a) Conventional LS method, (b) Sparse Filter

at $s_n = a_n$. In the case of $s_n = 0$,

$$E(0) = \frac{\gamma}{2}a_n^2$$

holds.

Thus if

$$\beta > \frac{\gamma}{2}a_n^2$$

is satisfied, $E(s_n)$ has the minimum value $\frac{\gamma}{2}a^2$ at $s_n = 0$. Otherwise $E(s_n)$ has minimum value β at $s_n = a_n$. In the end, (21) holds.

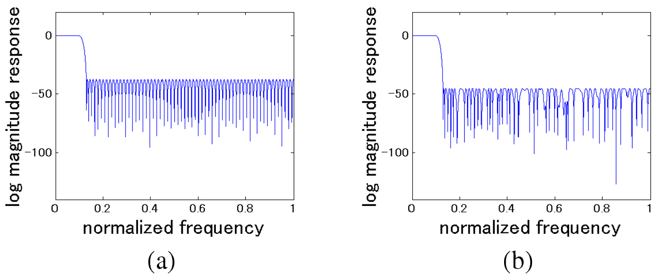

Fig. 3. Example 2 (a) PM method, (b) Sparse Filter

5. REFERENCES

[1] L.R.Rabiner and B.Gold, Theory and Application of Digital Signal Processing, Englewood Cliffs , NJ; Prentice-Hall.

[2] A.V. Oppenheim and R.W. Schafer, Digital Signal Processing ,Englewood Cliffs , NJ; Prentice-Hall.

[3] T.W.Parks and J.H.McClellan, "Chebyshev Approximation For Nonrecursive Digital Filters with Linear Phase", *IEEE Trans. Circuit Theory*,vol.**CT-19**, pp.189-194, 1972.

[4] J.H.McClellan,T.H.Parks and L.R.Rabiner, "A Computer Program for Designing Optimum FIR Linear Phase Digital Filters", *IEEE Trans. Audio. Electroacoustics*, vol.**AU-21**, No.6, pp.506-526, 1973.

[5] J.H.McClellan and T.H.Parks, "A Unified Approach to the Design of Optimum FIR Linear-Phase Digital Filters ", *IEEE Trans. Circuit Theory*, vol.**CT-20**, No.6, pp.506-526, 1973.

[6] P.P. Vaidyanathan and T.Q. Nguyen, "Eigenfilters: A new approach to least squares FIR filter design and applications," *IEEE Trans Circuits Syst.*, vol.**CAS-22** pp.943-953 ,1975.

[7] V.R. Algazi and M. Suk, " On the frequency weighted least squares design of finite duration filters", *IEEE Trans Circuits Syst.*, vol.**CAS-34** pp.80-95 ,1987.

[8] V.R. Algazi, M. Suk and C.S. Rim, "Design of almost minimax FIR filters in one and two dimensions by WLS techniques," *IEEE Trans Circuits Syst.*, vol.**CAS-33** pp.590-596 ,1986.

[9] Masahiro Okuda, Masaaki Ikehara, Shin-ichi Takahashi,"Fast and Stable Least Squares Approach for the Design of Linear Phase FIR Filters", IEEE Trans. on Signal Processing, vol.46, (no.6), p.1485-93, June 1998

[10] C.S. Burrus, A.W. Soewito and R.A. Gopinath, "Least squares error FIR filter design with transition bands," *IEEE Sigmal Processing.*, vol.**SP-40** pp.1327-1340 ,1992.

[11] C.S. Burrus, "Multiband least squares FIR filter design," *IEEE Sigmal Processing.*, vol.**SP-43** pp.412-421 ,1995.

[12] Y.C.Lim, J.H.Lee, C.K.Chen and R.H.Yang, "A Weighted Least Squares Algorithm for Quasi-Equiripple FIR and IIR Digital Filter Design", *IEEE Sigmal Processing.*, vol.**SP-40** pp.551-558 ,1992.

[13] D.Wei, "Non-convex optimization for the design of sparse FIR filters, " IEEESP 15th Workshop on Statistical Signal Processing, 2009

[14] T.Baran, D.Wei, and A.V. Oppenheim, "Linear Programming Algorithms for Sparse Filter Design," IEEE Transactions on Signal Processing, Vol.58, No.3, Page(s): 1605 - 1617 , 2010

[15] Yilun Wang and Junfeng Yang and Wotao Yin and Yin Zhang, " A new alternating minimization algorithm for total variation image reconstruction," SIAM J. IMAGING SCI, pp.248-272, 2008

[16] Jose M. Bioucas-Dias, Mario A. T. Figueiredo: "An iterative algorithm for linear inverse problems with compound regularizers," ICIP 2008: 685-688

[17] Li Xu, Cewu Lu, Yi Xu, Jiaya Jia, "Image smoothing via L_0 gradient minimization," Proceedings of the 2011 SIGGRAPH Asia Conference Article No. 174, 2011