A NEW TECHNIQUE FOR THE DESIGN OF FINITE PRECISION M-D FIR DIGITAL FILTERS

BY

Ahmad A. Masoud

Electrical Engineering Department., Queen's University, Kingston,
Ontario, Canada K7L 3N6

ABSTRACT

A method is proposed for the design of M-D finite precision FIR digital filters. The method operates by predistorting the frequency response prior to quantization. Such a predistortion is designed to counteract the degradation caused by the finite wordlength. The error measure used here is a convex function of the filter coefficients. Such a property enables us to replace the time consuming optimization techniques using direct search with the very fast techniques utilizing the first and second derivatives. Design examples are given for both the 1-D and the 2-D case. The 1-D results are compared to a recently published method [1].

I. INTRODUCTION

Designing filters with finite precision coefficients is desired to reduce the distortion caused by constraining the wordlength to a finite size [2]. The design process requires the finding of a discrete set of filter coefficients that minimizes some error measure. Compared to the continuous case, minimization on a discrete set of coefficients is a very time consuming process [3]. The problem becomes exceedingly difficult when extended to the M-D case. Little work has been done involving the design of finite precision 2-D FIR filters [4]; let alone higher dimensional filters. Here, a method is proposed for the design of finite precision M-D FIR filters. The method operates by predistorting the infinite precision frequency response prior to quantization. Such a process is designed to counteract the degradation introduced by the finite wordlenth. The approximation is used to construct a convex error measure with the aim of isolating design complexity from the number of quantization levels. Such an approach, although not optimum, is found to be fast and efficient. The method can, also, be efficiently used for designing 1-D finite precision FIR digital filters.

II. THE PROPOSED METHOD

The coefficients of the infinite precision design ($h(n_1,.,n_M)$) are assumed to be normalized to the interval [-1,+1]. The approximation is carried out for the uniform quantization nonlinearity ($Q(x)$). The nonlinearity is approximated with the following function :

$$Q(x) \simeq \Gamma(x) = x - \frac{\Delta}{2\pi}\cdot\sin(\frac{2\pi}{\Delta}\, x) \qquad (1)$$

$$x \in [-1,+1]$$

where Δ is the quantization step (Figure 1) :

$$\Delta = \frac{1}{2^{L-1} - 1}$$

and L is the number of bits (sign bit included).

To implement the proposed approach, a new set of infinite precision coefficients ($h'(n_1,.,n_M)$) are computed to replace the original coefficients. The replacement aims at making the distance function (to be constructed later). between $H(\omega_1,.,\omega_M,h)$ and $H(\omega_1,.,\omega_M,Q(h'))$ less than that between $H(\omega_1,.,\omega_M,h)$ and $H(\omega_1,.,\omega_M,Q(h))$. From now on, the quantization nonlinearity ($Q(x)$) will be replaced by its approximate $\Gamma(x)$.

Let $\hat{h}(n_1,.,n_M)$ be the quantized filter coefficients, and $\hat{h}_\alpha(n_1,.,n_M)$ be its approximation obtained using $\Gamma(x)$. Thus :

$$\hat{h}_\alpha(n_1,.,n_M)=\Gamma(h(n_1,.,n_M)) - \frac{1}{c}\cdot\sin(c\cdot h(n_1,.,n_M)) \qquad (2)$$

where $c = 2\pi/\Delta$,and M is the number of dimensions. With no loss of generality only the even symmetric case is

CH2868-8/90/0000-2450$1.00 © 1990 IEEE

considered here. The frequency response corresponding to the approximate quantized filter coefficients is :

$$H(\omega_1,.,\omega_M,\hat{h}_a(n_1,.n_M))$$

$$= \sum_{n_1}..\sum_{n_M} (h(n_1,.,n_M)) - \frac{1}{c}\sin(c\cdot h(n_1,.,n_M)))\cdot \prod_{i=1}^{M}\cos(\omega_i\cdot n_i)$$

$$= \sum_{n_1}..\sum_{n_M} h(n_1,.,n_M)\cdot\prod_{i=1}^{M}\cos(\omega_i\cdot n_i) - \frac{1}{c}\sum_{n_1}..\sum_{n_M} \sin(c\cdot h(n_1,.,n_M))\cdot\prod_{i=1}^{M}\cos(\omega_i\cdot n_i)$$

$$= H(\omega_1,.,\omega_M,h) + \frac{1}{c}\cdot H_r(\omega_1,.,\omega_M,h) \qquad (3)$$

As expected the quantized response consists of two terms, the infinite precision term H, and the distortion term H_r scaled by the constant 1/c. Computing the mean square value of the distortion term we have :

$$\frac{\pi^M}{2c^2}\cdot\prod_{i=1}^{M} N_i - \frac{\pi^M}{2c^2}\sum_{n_1}..\sum_{n_M}\cos(2c\cdot h(n_1,.,n_M)) \qquad (4)$$

where N_i is the length of the filter in the i'th dimension.

To construct the error measure let us consider the approximate quantized predistorted response corresponding to the coefficient set $\{\hat{h}'_a(n_1,.,n_M) = \Gamma(h'(n_1,.,n_M))\}$. By choosing h' different from h the distortion term in (3) will change, at the same time the infinite precision component will deviate from the original infinite precision design; this deviation will be referred to as the deviation error. The set {h'} is chosen such that the reduction in the mean square value of the distortion component is greater than the increase in the mean square value of the deviation component. This is expected to lead to a decrease in the whole error. The distortion error $E_d(h')$ is equal to :

$$\frac{\pi^M}{2c^2}\prod_{i=1}^{M} N_i - \frac{\pi^M}{2c^2}\sum_{n_1}..\sum_{n_M}\cos(\ 2c\cdot h'(n_1,.,n_M))$$

The deviation error $E_\Delta(h')$ is equal to :

$$\pi^M\sum_{n_1}..\sum_{n_M}(h'(nl,.,n_M) - h(n_1,.,n_M))^2$$

The adopted error measure is taken to be the sum of both error components, that is :

$$\pi^M\sum_{n_1}..\sum_{n_M}[\ (h'(n_1,.,n_M) - h(n_1,.,n_M))^2 - \frac{1}{2c^2}\cos(2c\cdot h'(n_1,.,n_M))] + \frac{\pi^M}{2c^2}\cdot\prod_{i=1}^{M} N_i \qquad (5)$$

By taking the first and second derivative one can easily show that the error measure is a convex function of its arguments.

One important consideration when optimizing (5) is the constant term. Although, this term may give the false notion that it can be neglected in the design, its presence in a time domain error measure reflects itself as a frequency domain error concentrated near the origin of the frequency axis. One way to deal with this is to offset the coefficients obtained by minimizing (5) with a constant β prior to quantization. β can be computed by minimizing the following error measure:

$$\min_{\beta}\int_0^{2\pi}.\int_0^{2\pi}(\ H_d(\omega_1,.,\omega_M) - \sum_{n_1}..\sum_{n_M}[h'_o(n_1,.n_M) - \beta]\cdot\prod_{i=1}^{M}\cos(\omega_i\cdot n_i))^2 d\omega_1.d\omega_M \qquad (6)$$

where the square brackets represents the quantization nonlinearity. It is obvious that using direct search to find β can not be considered as a computational burden. β is observed to be a small constant near the origin.

III. EXAMPLES

To demonstrate the ability of the technique in designing finite precision filters two examples are provided for the 1-D and the 2-D case.

1- 1-D lowpass filter.

Here, the method is compared to another method for designing 1-D finite precision filters. The technique is based on error spectrum shaping (ESS) [1]. A low pass filter having the following desired characteristics is to be designed :

$$H_d(\omega) = \begin{cases} 1 & |\omega| \le .4\pi \\ 0 & \text{else} \end{cases}$$

The infinite precision design obtained using the McClellan algorithm is shown below :

Table 1. 1-D LP filter, infinite precision design.

ω_p	ω_s	N	δ_p dB	δ_s dB
.4π	.6π	65	.0005	-132.1

the quantized response :

Table 2. 1-D LP filter, finite precision design.

L	Rounding		ESS		Prop. method	
	δ_p	dB δ_s	δ_p	dB δ_s	δ_p	dB δ_s
16	.0011	-79.7	.0035	-97.7	.00065	-92.4

2. 2-D lowpass filter.

The desired characteristics are :

$$H_d(\omega_1, \omega_M) = \begin{cases} 1 & |\omega_1|, |\omega_2| \le .4\pi \\ 0 & \text{else where} \end{cases}$$

The infinite precision design is shown in the table below :

Table 3. 2-D LP filter, infinite precision design.

length	δ_p dB	δ_s dB
11x11	.5876	-26.0

and the finite precision design is :

Tabel 4. 2-D LP filter, finite precision design.

L	Rounding		Proposed method	
	δ_p	dB δ_s	δ_p	dB δ_s
4	.5307	-14.8	.013	-16.5

IV CONCLUSIONS

A method is proposed for the design of finite precision M-D FIR digital filters. The method functions by predistorting the infinite precision frequency design in order to alleviate the distortion introduced by quantization. The quantization nonlinearity is approximated with a mathematically manageable function. Such an approximation enable us to construct a time domain error measure that is a convex function of the filter coefficients. With such a property it is possible to replace the time consuming optimization techniques based on Integer Programming with the much faster techniques based on the first and second derivatives. It ought to be noticed that the accuracy in approximating the nonlinearity can be improved to an arbitrary degree. Nevertheless, such an increase can be at the expense of adding more complexity, and the danger of losing some of the desirable properties such as convexity. Although the method does not produce optimum results, it is capable of yielding a satisfactory design.

REFERENCES

[1] J. J. Nielsen," Design of Linear-phase Direct-form FIR Digital Filters with Quantized Coefficients Using Error Spectrum Shaping Techniques", IEEE Trans. on ASSP, Vol.37, July 89, P.P. 1020-1026.

[2] D. S. Chan, L.R. Rabiner," Analysis of Quantization Errors in the Direct Form of Finite Impulse Response Digital Filters", IEEE Trans. on AU, Vol. AU-21, No.4, Aug. 73, P.P. 354-366.

[3] D. M. Kodek," Design of Optimal Finite Word-length FIR Digital Filters Using Integer Programming Techniques", IEEE Trans. on ASSP, Vol. ASSP-28, No.3, June 80, p.p. 304-308.

[4] P. Siohan, A. Benslimane," Finite Precision Design of Optimal Linear-phase 2-D FIR Digital Filters", IEEE Trans. on Circuits and Systems, Vol. CAS-36, No.1, Jan. 89, P.P. 11-21.

ACKNOWLEDGEMENT

The author would like to thank Dr. M.M.Fahmy, elec. eng. dept., Queen's University for his useful comments in writing the manuscript.

APPENDIX

The filter quantized filter coefficients of example 1 obtained using the proposed method are:

#		#	
1	+1.00000	18	+0.00656
2	+0.65292	19	+0.00659
3	+0.03031	20	-0.00345
4	-0.20836	21	-0.00443
5	-0.02878	22	+0.00153
6	+0.11451	23	+0.00268
7	+0.02631	24	-0.00058
8	-0.07159	25	-0.00156
9	-0.02319	26	+0.00009
10	+0.04645	27	+0.00079
11	+0.01965	28	+0.00009
12	-0.03015	29	+0.00031
13	-0.01608	30	-0.00003
14	+0.01904	31	+0.00018
15	+0.01254	32	+0.00012
16	-0.01157	33	+0.00003
17	-0.00937	34	+0.00000

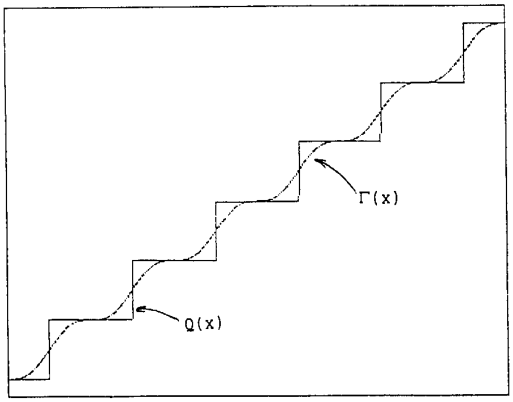

Figure 1. The Uniform quantization nonlinearity Q(x) and it approximation Γ(x).

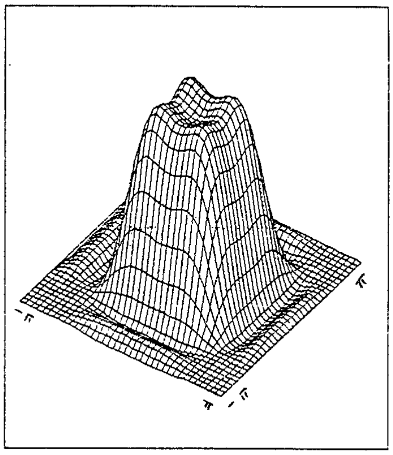

Figure 2.a The infinite precision response.

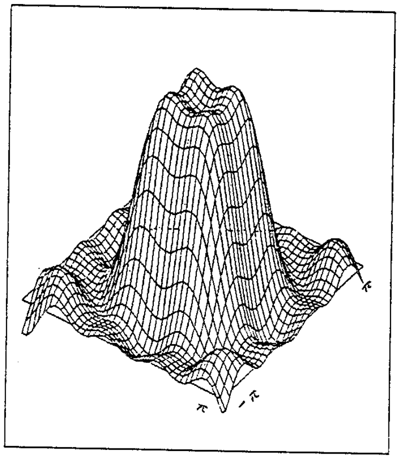

Figure 2.b Direct quantization.

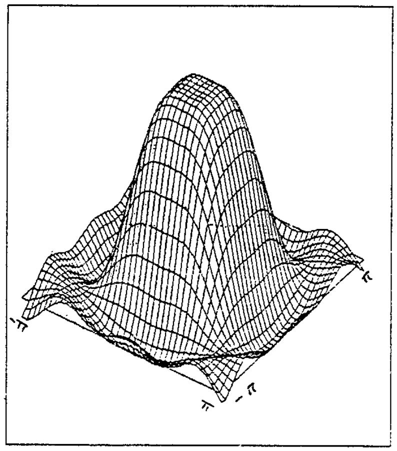

Figure 2.c Proposed method.