PROPERTIES OF APPROXIMATE PARKS-McCLELLAN FILTERS

Li Lee and Alan V. Oppenheim

Research Laboratory of Electronics
MIT, Cambridge, MA 02139

ABSTRACT

It has been observed empirically that each coefficient in a Parks-McClellan filter converges to a steady state value as the filter length increases. This suggests the possibility of obtaining filters that are near optimal while "re-using" filter coefficients from shorter filters in the design of longer filters. In the context of approximate processing this then allows a filtering operation to be done in stages. This paper demonstrates this observation and examines some of its implications.

1. INTRODUCTION

For digital signal processing applications with real-time or low-power constraints, it is often desirable to use algorithms whose output quality can be adjusted depending on the availability of resources such as time or power. For this reason, recently there has been increased interest in approximate signal processing algorithms whose intermediate results represent successively better approximations to the desired solution [1] [2] [3] [4].

When designing FIR filters, successively better solutions become possible as the filter length is allowed to increase. For example, if windowing is used to design a low-pass filter, greater stopband attenuation and sharper transition bands would result from longer windows. However, to design the filter with the optimal stopband attenuation for a given length and transition band, the Parks-McClellan algorithm is used.

In this paper, we explore the possibility of designing FIR filters for approximate signal processing under the minimax criterion used in Parks-McClellan (hereafter written as P-M) filters. The study has been motivated by the observation that for P-M filters designed under fixed transition band constraints, the value of a given coefficient as a function of the filter length N converges as N becomes large. This "settling" behavior suggests the possibility of "reusing" filter coefficients (i.e., those which had already converged to within a given level of error) from shorter filters in the design of longer filters, and vice versa. In the context of approximate signal processing, this would allow a filtering operation to be done in stages, each extending the previous by adding on the results of filtering using only those filter taps which changed or became non-zero. Alternately, it would allow the filter length to be adjusted to adapt to changing resources.

The purpose of this paper is to examine some of the implications of this observation. After describing the notation and assumptions we make in the paper, we demonstrate empirically and mathematically that P-M filter coefficients indeed converge as functions of filter length. Additionally, we use linear programming to explore and evaluate two possible strategies to "reuse" filter coefficients from other filters. We find that one of these approximate FIR filtering schemes appears to offer performance competitive with P-M filters at each stage of computation.

This research was supported in part by the Department of the Navy, Office of the Chief of Naval Research, contract number N00014-93-1-0686, in part by the U.S. Air Force Office of Scientific Research under Grant AFOSR-F49620-96-1-0072, and in part through collaborative participation in the Advanced Sensors Consortium sponsored by the U.S. Army Research Laboratory under Cooperative Agreement DAAL01-96-2-0001. Li Lee is supported by a fellowship from the AT&T Graduate Research Program for Women.

2. NOTATION

This study is limited to type I linear phase low-pass filters, which are symmetric with an odd number of taps. The impulse response of a causal filter with $2N+1$ taps is denoted by $h[n] = h_N[n-N]$. For convenience in notation we refer to $h_N[n]$ as the impulse response, recognizing that the corresponding causal impulse response is easily obtained through a simple time shift. The corresponding frequency response is then the Fourier transform of $h_N[n]$, *i.e.*,

$$H_N(e^{j\omega}) = h_N[0] + 2\sum_{n=1}^{N} h_N[n]\cos(n\omega) \qquad (1)$$

Additionally, $H_d(e^{j\omega})$ denotes the desired frequency response,

$$H_d(e^{j\omega}) = \begin{cases} 1, & 0 \le \omega \le \omega_p, \\ 0, & \omega_s \le \omega \le \pi \end{cases} \qquad (2)$$

where ω_p is the passband edge frequency and ω_s is the stopband edge frequency. In the empirical studies described in this paper, the error tolerance was chosen to be the same for both the passband and the stopband. The error measure, which is the maximum error over the pass- and stopbands, is denoted by E_N:

$$E_N = \max(|H_d(e^{j\omega}) - H_N(e^{j\omega})|). \qquad (3)$$

A P-M filter of length $2N+1$ achieves the minimal E_N possible. In this paper, $h_N[n]$ always denotes a P-M filter, whereas $g_N[n]$ denotes an approximate filter.

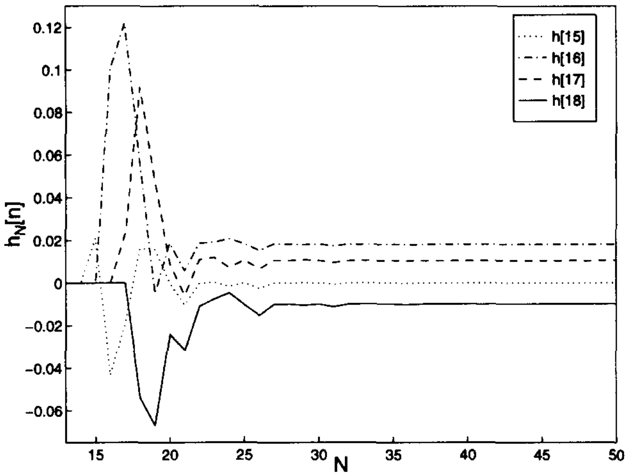

Figure 1: Four coefficients of P-M filters are tracked as the lengths of the filters increase. The filter design parameters are $\omega_p = 0.39\pi$ and $\omega_s = 0.41\pi$.

3. CONVERGENCE OF FILTER COEFFICIENTS

In this section, we discuss the behavior of P-M filter coefficients as the filter length increases. We note in particular that the coefficients appear to converge, and we present a mathematical justification for this convergence behavior.

In Figure 1, the behavior of the values of four P-M filter coefficients is shown as filter length increases. We note that initially there is considerable variation in the value of the coefficients. But as the length increases, the variations decrease and each coefficient appears to settle to a relatively constant value. This behavior has been observed consistently in our empirical studies.

Mathematically, the convergence behavior of the individual coefficients can be explained by the behavior of the functions $H_N(e^{j\omega})$ as N increases. Because the set of all possible filters $h_N[n]$ is a subset of the filters $h_{N+1}[n]$, the maximum error in the pass- and stopbands can only decrease as the filter length increases. Therefore,

$$\lim_{N\to\infty} E_N = 0. \quad (4)$$

Further, since Eqn. (3) represents a distance measure between H_d and H_N in the space of functions over $X = [0,\omega_p] \cup [\omega_s,\pi]$, Eqn. (4) shows that H_N converges uniformly to H_d everywhere except for the transition region.

While the behavior of the functions within the transition region $[\omega_s,\omega_p]$ is not as easily characterized, empirical observations suggest that either the sequence $H_N(e^{j\omega})$ converges uniformly to some continuous function $H(e^{j\omega})$ or it converges uniformly everywhere except at the middle of the transition region.

In the case that $H_N(e^{j\omega})$ does converge uniformly to some function $H(e^{j\omega})$ over $[0,\pi]$, then the convergence of $h_N[n_0]$ in the time domain results from the following:

$$\begin{aligned}
& \lim_{N\to\infty} |h_N[n_0] - h[n_0]| \\
= & \lim_{N\to\infty} \left|\frac{1}{2\pi}\int_{-\pi}^{\pi} (H_N(e^{j\omega}) - H(e^{j\omega}))\cos(n_0\omega)d\omega\right| \\
\le & \lim_{N\to\infty} \frac{1}{\pi}\int_0^{\pi} |H_N(e^{j\omega}) - H(e^{j\omega})|d\omega \\
\le & \frac{1}{\pi}\int_0^{\pi} \lim_{N\to\infty} E_N d\omega \\
= & 0
\end{aligned} \quad (5)$$

Further, because an integral is used, the above equations hold even if isolated points of $H_N(e^{j\omega})$ do not converge uniformly to $H(e^{j\omega})$ on the transition region. Therefore, the coefficients individually converge whether or not uniform convergence takes place over the entire frequency axis.

4. APPROXIMATE FILTERS

One consequence of the convergence behavior of the coefficients is that in general, $h_N[n] - h_{N+M}[n]$ is small for the center coefficients $|n| \le N_0$, which have sufficiently "settled". This suggests the possibility of using the same coefficients for both $h_N[n]$ and $h_{N+M}[n]$ for $|n| \le N_0$. A filter of length $2N+1$ which constrained a subsets of its coefficients to be identical to those coefficients from $h_{N+M}[n]$ would obviously be only an approximation to the optimal filter $h_N[n]$. However, if the assumed coefficients are close to those of the actual optimal filter, the approximation may result in only a slight degradation in performance.

Approximate signal processing algorithms are typically structured such that successively refined answers are yielded at intermediate points of computation. In the case of FIR filtering, approximate P-M filters can be incorporated into an overall filtering algorithm under two possible strategies. Both of them progressively improve the filter output quality at each refinement stage by augmenting the filter. Specifically, at each stage of refinement a portion of the filter coefficients are retained from the previous stage and the remaining coefficients are updated. The input is processed through only those filter taps which changed, and the result is added to the output of the previous filtering stage.

The difference between the two methods is that while one method (which we refer to as Method I) ends with an optimal longest filter at the last filtering stage, the other (Method II) starts with an optimal shortest filter at the first step. More specifically, Method I starts by designing a P-M filter to be used at the last filter stage. The shorter filters used in the intermediate stages of the algorithm then constrain their center filter coefficients ($|n| \le N_0$) to the values of this final filter. In contrast, Method II starts with the design of a P-M filter for the first filter stage. Then at each stage, the center taps ($|n| \le N_0$) of the existing filter are kept fixed, the rest are modified, and new taps are added.

In both methods, the filters used in the intermediate stages are suboptimal, and the remainder of the paper attempts to characterize the cost of the approximations by comparing the stopband attenuation of the approximate filters to those of P-M filters. To investigate the performance of the approximate filters, an interior-point method for linear programming was used to solve the following problem:

> Design an FIR filter $g_L[n]$ of $2L+1$ taps such that $g_L[n] = h_M[n]$ for $|n| \le N_0$, and $\max |G(e^{j\omega}) - H_d(e^{j\omega})|$ is minimized over the pass- and stop-bands.

Letting

$$A(e^{j\omega}) = h_M[0] + 2\sum_{n=1}^{N_0} h_M[n]\cos(n\omega), \qquad (6)$$

the algorithm solves the linear programming problem which minimizes the maximum error δ, under the constraints

$$|2\sum_{N_0+1}^{L} g_L[n]\cos(n\omega) + A(e^{j\omega}) - H_d(e^{j\omega})| \le \delta, \qquad (7)$$

for $0 \le \omega \le \omega_p$ and $\omega_s \le \omega \le \pi$. While the efficiency of the algorithm is comparable to that of other linear programming algorithms, it is more computationally intensive than the Remez algorithm. Therefore, it is not the intent of this paper to propose this specific algorithm as an efficient method for obtaining the filter coefficients in the incremental refinement procedure. Rather, we hope to use its results to compare the error of the filter $g_L[n]$ with that achieved by the P-M filter $h_L[n]$ under different ways of choosing N_0. The results would give a good indication the feasibility of using Method I or II to reuse filter coefficients in the process of lengthening the FIR filter. It then remains to find a suitably efficient algorithm for obtaining the remaining coefficients.

5. EXPERIMENTAL RESULTS

5.1. Method I

In this section, we present by example some observations related to the approximate filters which would be used in Method I. Based on these observations, we discuss the feasibility of performing approximate FIR filtering using this strategy and comment on the effect of the choice of parameters such as N_0 on the performance of the approximations.

In Method I, the same filter $h_M[n]$ is used to constrain the taps of the approximate filters $g_L[n]$, where $L < M$. One design issue deals with the choice of N_0 for each possible value of L. If N_0 is small, the approximation $g_L[n]$ will be closer to the optimal P-M filter. However, a large N_0 makes the approximate filtering scheme more efficient, since only a few filter coefficients would have to be updated for the next computational stage. Along with this design tradeoff is a further question of how to choose N_0 as a function of L.

Figures 2 and 3 show the results of two possible approaches to choosing N_0 as a function of L. In Figure 2, we choose $N_0 = L - c$, for $c = 10, 20, 30$. In Figure 3, we choose $N_0 = \alpha L$, for $\alpha = 0.1, 0.3, 0.5, 0.7$. Both figures show the stopband attenuations achieved by approximate P-M filters $G_L(e^{j\omega})$ for each of those settings, as well as the lower bounds set by the attenuation of the P-M filters. The optimal filter $h_{75}[n]$ used to constrain the taps of $G_L(e^{j\omega})$ is a P-M filter of 151 taps. Additionally, the design parameters were $\omega_p = 0.38\pi$ and $\omega_s = 0.42\pi$ for all of the filters. For example, in Figure 2, the filter $g_{50}[n]$, represented by the circle('o') for $L = 50$, was generated using linear programming under the constraint that $g_{50}[n] = h_{75}[n]$ for $|n| < L - 10 = 40$. Similarly, in Figure 3, the filter $g_{60}[n]$, represented by the '+' for $L = 60$, was generated using linear programming under the constraint that $g_{60}[n] = h_{75}[n]$ for $|n| < 0.3L = 18$.

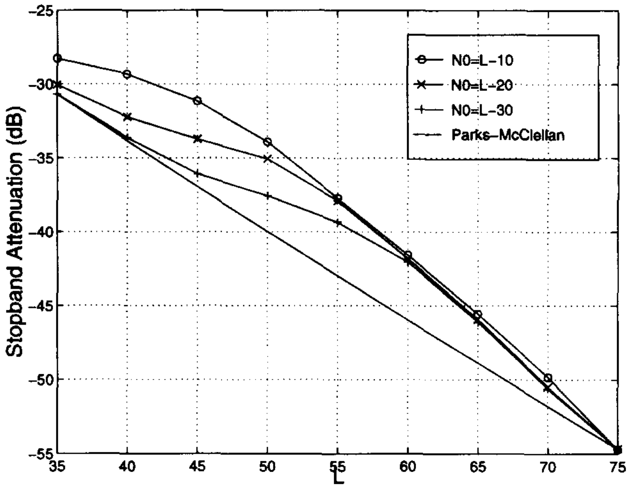

Figure 2: Stopband attenuation of approximate filters $g_L[n]$ as compared with the corresponding P-M filters $h_L[n]$. $g_L[n]$ is a length $2L + 1$ FIR filter which achieves the greatest stopband attenuation under the constraint that $g_L[n] = h_{75}[n]$ for $|n| \le N_0$, and $N_0 = L - c$, as labelled.

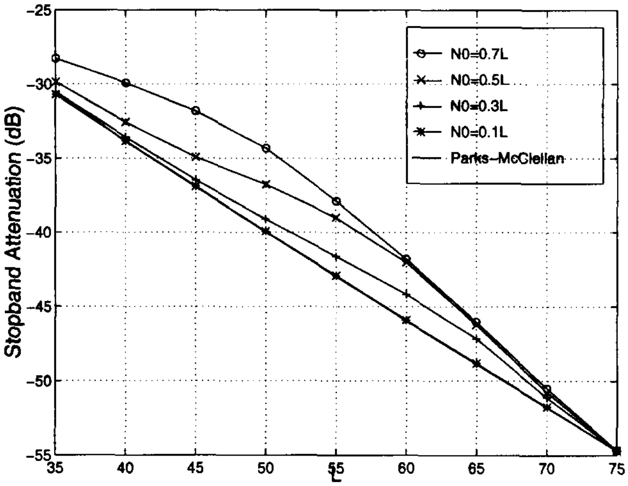

Figure 3: Stopband attenuation of approximate filters $g_L[n]$ as compared with the corresponding P-M filters $h_L[n]$. $g_L[n]$ is a length $2L + 1$ FIR filter which achieves the greatest stopband attenuation under the constraint that $g_L[n] = h_{75}[n]$ for $|n| \le N_0$, and $N_0 = \alpha L$, as labelled.

It is clear from the figures that the performance of the approximate filters improves with decreased N_0, as expected. In Figure 2, where $N_0 = L - c$, the constant c significantly varies the stopband attenuation achieved by the approximation $G_L(e^{j\omega})$ for small L, but its effect decreases as L increases. In contrast, in Figure 3, where $N_0 = \alpha L$, the deviation between the performance of the approximations $G_L(e^{j\omega})$ and that of the P-M filters $H_L(e^{j\omega})$ appears to be sensitive to the value of α for every value of L except $L = 75$, where $g_L[n]$ is fixed to be the optimal P-M filter. This comparison suggests that choosing N_0 as a fraction of L is a better choice. These plots also suggest that Method I is feasible way to structure approximate FIR filtering. For $\alpha = 0.5$, in the worst case ($L = 55$), the stopband attenuation for $G_L(e^{j\omega})$ was only 4 dBs higher than that achieved by the P-M filter.

5.2. Method II

In this section, we examine the performance of approximate filters which would be used in Method II. Under Method II, one starts the approximate processing with an optimal P-M filter. At the end of each processing stage, the taps of the current filter form the constraints for the design of the approximate filter for the next stage. The plots in this section will indicate the limitations of structuring approximate FIR filtering this way. As in the last section, we also comment on the effect of the choice of N_0 on the performance of the approximate filters.

Figures 4 and 5 show the stopband attenuation achieved by each stage of the approximate filtering when we structure FIR filtering using Method II under two schemes of choosing N_0. In Figure 4 we choose $N_0 = L - c$, for $c = 10, 20, 30, 40$, and in Figure 5, we choose $N_0 = \alpha L$, for $\alpha = 0.1, 0.2, 0.3, 0.5$. For example, in Figure 4, the curve connecting the asterisks ('*') shows the stopband attenuation achieved when each $g_L[n]$ is generated under the constraint $g_L[n] = g_{L-5}[n]$ for $|n| < L - 40$. Similarly in Figure 5, the curve connecting the asterisks ('*') shows the stopband attenuation achieved when each $g_L[n]$ is generated under the constraint $g_L[n] = g_{L-5}[n]$ for $|n| < 0.1L$. In both plots, $g_{45}[n]$ is set to be the P-M filter of length 91, and the design parameters are $\omega_p = .38\pi$ and $\omega_s = 0.42\pi$.

The curves show that with Method II, the stopband attenuations achieved by the approximate filters reach a breakpoint in L. After that point, increasing L no longer improves the performance of the filter, and the performance difference between the approximate filters and the P-M filters increases rapidly. The parameters c in Figure 4 and α in Figure 5 influence the position of the breakpoint, but not the behavior of the approximate filters after it. Additionally, the figures show that N_0 has to be chosen to be quite small (large c or small α) for the breakpoint to take place for a large value of L. A comparison of these figures with Figures 2 and 3 clearly indicates that Method I offers much better performance over a wider range of L than Method II.

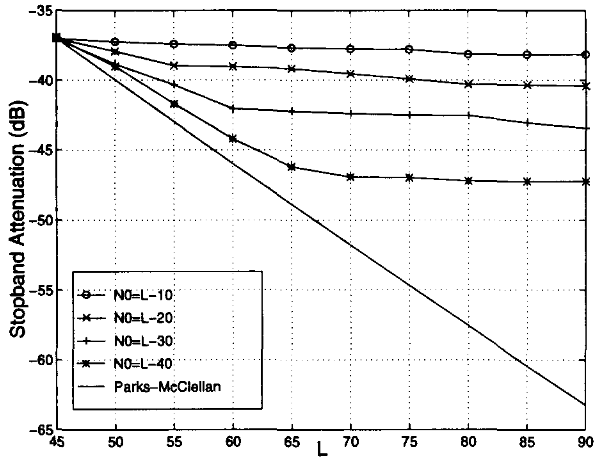

Figure 4: Stopband attenuation of approximate filters $g_L[n]$ as compared with the corresponding P-M filters $h_L[n]$. $g_L[n]$ is a length $2L + 1$ FIR filter which achieves the greatest stopband attenuation under the constraint that $g_L[n] = g_{L-5}[n]$ for $|n| \leq N_0$, and $N_0 = L - c$, as labelled. $g_{45}[n]$ is the P-M filter of length 91.

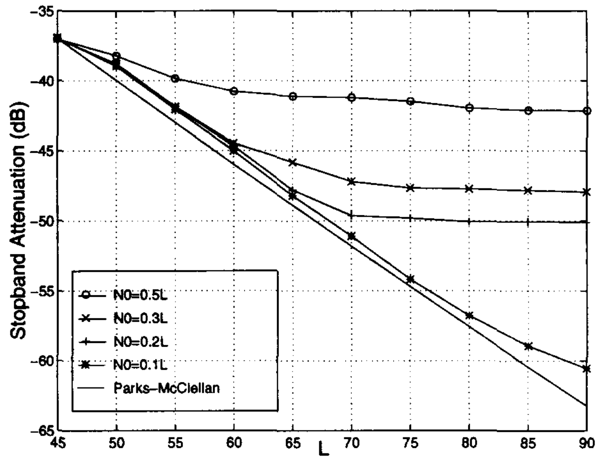

Figure 5: Stopband attenuation of approximate filters $g_L[n]$ as compared with the corresponding P-M filters $h_L[n]$. $g_L[n]$ is a length $2L + 1$ FIR filter which achieves the greatest stopband attenuation under the constraint that $g_L[n] = g_{L-5}[n]$ for $|n| \leq N_0$, and $N_0 = \alpha L$, as labelled. $g_{45}[n]$ is the P-M filter of length 91.

6. SUMMARY

This paper presented an empirical study which showed the feasibility of using approximate P-M filters for approximate signal processing. By exploiting the observation that filter coefficients of P-M filters converge, we found that placing appropriate constraints on the center coefficients of the approximate filters only slightly degraded their performance. Future research directions include developing a firm theoretical understanding of our observations and finding an efficient algorithm for designing the approximate filters.

7. REFERENCES

[1] J.T. Ludwig, S.H. Nawab and A. Chandrakasan, "Low-Power Digital Filtering using Approximation Processing", *IEEE Journal on Solid State Circuits*, vol. 31, no. 3, pp.395-400, March 1996.

[2] S.H. Nawab, A.V. Oppenheim, A.P. Chandrakasan, J.M. Winograd, and J.T.Ludwig, "Approximate Signal Processing," *VLSI Signal Processing*, Fall 1996

[3] J.M. Winograd and S.H. Nawab, "Incremental Refinement of DFT and STFT Approximations", *Signal Processing Letters*, vol. 2,no. 2, pp. 25-27, February 1995.

[4] Joseph M. Winograd, S. Hamid Nawab, and Alan V. Oppenheim, "FFT-Based Incremental Refinement of Suboptimal Detection", *Proc. ICASSP 1996*, vol. 5, pp. 2479-2482.