On the Design of Highly Accurate and Efficient IIR and FIR Filters

Gregory Beylkin*, Ryan D. Lewis, and Lucas Monzón

Abstract—**We describe a systematic method for designing highly accurate and efficient infinite impulse response (IIR) and finite impulse response (FIR) filters given their specifications. In our approach, we first meet the specifications by constructing an IIR filter with, possibly, a large number of poles. We then construct, for any given accuracy, an optimal IIR version of such filter (with a minimal number of poles). Finally, also for any given accuracy, we convert the IIR filter to an efficient FIR filter cascade (either serial or parallel). Since in this FIR approximation the non-causal part of the IIR filter only introduces an additional delay (as a function of the desired accuracy), our IIR construction does not have to enforce causality. Thus, we obtain a simple method for constructing linear phase filters if the specifications so require. All of these procedures are accomplished via robust, fast algorithms. We provide several illustrative examples of our method.**

Index Terms—**Approximation algorithms, digital filter design, FIR filters, IIR filters, optimal rational approximations, quadrature mirror filters.**

I. Introduction

IN HIS 2006 paper "The Rise and Fall of Recursive Digital Filters," [1] Rader gives a brief history of filter design methods. He describes how the perceived pros and cons of recursive and non-recursive filters changed over time as new design and implementation techniques were discovered. The goal of our paper is to offer an addendum to this history by providing a new systematic method of designing both types of filters. Our approach is based on a combination of several approximation algorithms and a few observations. We cite algorithms for constructing near-optimal rational approximations [2], [3], a new high accuracy reduction algorithm [4], and a somewhat obscure short note [5]. Our key observation is that it is relatively easy to construct an accurate but sub-optimal (with a large number of poles) rational filter that satisfies the design criteria. We describe an effective approach for the sub-optimal construction well suited for the optimization algorithm. We then rely on robust nonlinear algorithms for optimal rational approximation to minimize the number of poles for a desired accuracy.

We first construct an infinite impulse response (IIR) filter that satisfies the design criteria without attempting to make the design optimal. We next find an equivalent (vis-à-vis the specifications) IIR filter with a near-minimal number of poles. We then convert, for any given accuracy, the IIR filter to an efficient finite impulse response (FIR) filter. It is well known that approximating a rational function with a polynomial for a set accuracy may require a polynomial of high degree. Despite the high degree of our FIR filter, its implementation cost is low and requires only $\mathscr{O}\left(\log\varepsilon^{-1}\right)$ operations, where ε is the desired accuracy. This efficiency is achieved by expressing the FIR filter as a cascade (either serial or parallel) where each factor is computationally inexpensive. Importantly for the many applications that require linear phase filters, we may easily design IIR filters with exact linear phase. In our method, the non-causal part of the IIR filter results in a finite delay in the FIR approximation that does not disturb the phase of the filter.

The combination of these design steps leads to a robust, nearly automatic, method for filter design. We believe that our approach contributes to the state-of-the-art of filter design as summarized in the conclusion of Rader's paper.

This research was partially supported by NSF grants DMS-1009951, DGE-0801680, DMS-0602284, and DOE/ORNL grant 4000038129.

G. Beylkin (corresponding author), R. D. Lewis, and L. Monzón are with the Department of Applied Mathematics, University of Colorado, Boulder, CO 80309-0526, United States (e-mail: beylkin@colorado.edu; lewisrd@colorado.edu; lucas.monzon@colorado.edu).

II. Preliminaries

In this section, we introduce notation and present the algorithms used in our filter design method. Given a filter, we identify its impulse response $h(n)$ with its z-transform,

$$H(z)=\sum_{n=-\infty}^{\infty}h(n)z^{-n}, \tag{II.1}$$

where the sum in (II.1) converges on the unit circle. To recall, if $h(n)$ contains only a finite number of nonzero terms, then $H(z)$ is a FIR filter. Otherwise, $H(z)$ is an IIR filter. If $h(n)=0$ for all $n<0$ then $H(z)$ is causal (and non-causal otherwise).

We introduce two filter design algorithms whose origins may be traced to the work of Adamjan, Arov, and Krein (AAK theory) [6], [7], [8]. The algorithm in Section II-A is often adequate but may require extended precision arithmetic for intermediate computations. The reduction algorithm in Section II-B (see [2], [9]) is significantly more efficient and its new version in [4] achieves high accuracy using only the standard double precision arithmetic. Finally, following [5], we describe an algorithm to convert IIR filters to efficient FIR filters while maintaining arbitrary finite accuracy.

A. Designing IIR Filters From a Desired Impulse Response

Our first algorithm constructs an IIR filter $H(z)$ whose impulse response $h(n)$ agrees with some desired impulse response $h_d(n)$, up to some finite but arbitrary accuracy $\varepsilon > 0$ over a certain range of the index $n \in \mathbb{Z}$.

Our solution makes use of an algorithm in [2], [9]. Given a sequence

$$h_d(n), \quad 1 \le n \le 2N+1$$

and a target accuracy $\varepsilon > 0$, we determine the optimal (minimal) number of nodes γ_m and weights w_m such that

$$\left| h_d(n) - \sum_{m=1}^{M} w_m \gamma_m^n \right| < \varepsilon, \quad 1 \le n \le 2N+1. \tag{II.2}$$

We now describe the steps of the algorithm to obtain this approximation.

Algorithm 1:

- Build the $N+1 \times N+1$ Hankel matrix

$$\mathbf{H}_{k\ell} = h_d(k+\ell+1), \quad k, \ell \in [0, N]. \tag{II.3}$$

- Find a vector $\mathbf{u} = (u_0, \ldots, u_N)^T$ satisfying

$$\mathbf{H}\mathbf{u} = \sigma \overline{\mathbf{u}}, \tag{II.4}$$

  with positive σ close to the target accuracy ε, where $\overline{\mathbf{u}} = (\overline{u}_0, \ldots, \overline{u}_N)^T$ denotes the element-wise complex conjugate of the vector $\mathbf{u}$. A problem of this form is known as a con-eigenvalue problem (see, e.g., [10, §4.6]), $\mathbf{u}$ is a con-eigenvector, and σ is a con-eigenvalue. In our case, $\mathbf{H}$ is a Hankel matrix and hence symmetric; the existence of a solution $(\sigma, \mathbf{u})$ follows from Takagi's factorization (see, e.g., [2, pp. 22]), as does the fact that we may take σ to be a singular value of $\mathbf{H}$ and $\mathbf{u}$ to be a specific singular vector.
- Given singular values $\sigma_0 \ge \sigma_1 \ge \ldots \ge \sigma_N$, we select a sufficiently small σ_M, which determines the accuracy of approximation, and the corresponding singular vector $\mathbf{u} = (u_0, \ldots, u_N)^T$.
- Compute the roots γ_m of the con-eigenpolynomial $u(z) = \sum_{n=0}^{N} u_n z^n$ whose coefficients are the entries of the vector $\mathbf{u}$ from the previous step.
- Obtain the weights w_m by solving the least-squares Vandermonde system

$$\sum_{m=1}^{N} w_m \gamma_m^n = h_d(n), \quad 1 \le n \le 2N+1. \tag{II.5}$$

  Typically, only M weights w_m have absolute value larger than the target accuracy ε. We then retain only those nodes γ_m that correspond to the significant weights and solve the corresponding Vandermonde system (II.5) again. For cases of practical interest in digital filtering, the sequence $h_d(n)$ exhibits decay as n becomes large. As a result, the nodes of interest lie inside the unit disk, $|\gamma_m| < 1$.

Remark 1.

- Typically, singular values decay rapidly so the number of terms M in the approximation (II.2) satisfies $M = \mathcal{O}(\log \varepsilon^{-1})$.
- To approximate a sequence

$$h_d(n), \quad -2N-1 \le n \le -1,$$

  by a sum

$$h_d(n) \approx \sum_{m=1}^{M} w_m \gamma_{m,}^n \quad -2N-1 \le n \le -1,$$

  we simply reindex $n \longmapsto -n$ and use Algorithm 1. In this case the nodes γ_m lie outside the unit disk, $|\gamma_m| > 1$, provided $h_d(n)$ decays as n becomes large and negative.
- We note that we may formulate this algorithm in terms of the singular value decomposition (SVD) without invoking the con-eigenvalue problem. However, Algorithm 3, which may be derived from Algorithm 1, requires this formulation. For a detailed analysis we refer to [2], [9].
- The nodes γ_m turn out to be the poles of the transfer function $H(z)$ that we construct via the next algorithm.

Let us now describe how to use Algorithm 1 to solve a filter design problem. Given a desired impulse response $h_d(n)$ for $n \in [-N_2, N_1]$ and target accuracy $\varepsilon > 0$, construct an IIR filter $H(z)$ with a (nearly) minimal number of poles whose impulse response $h(n)$ satisfies

$$|h_d(n) - h(n)| < \varepsilon, \quad n \in [-N_2, N_1]. \tag{II.6}$$

In the special case that $N_2 = 0$, $H(z)$ is a causal filter.

Algorithm 2:

- Determine poles γ_m^{in} and weights w_m^{in} such that

$$\left| h_d(n) - \sum_{m=1}^{M^{\text{in}}} w_m^{\text{in}} \left(\gamma_m^{\text{in}}\right)^n \right| < \varepsilon, \quad 1 \le n \le N_1,$$

  where $\left|\gamma_m^{\text{in}}\right| < 1$ using Algorithm 1.
- Determine poles γ_m^{out} and weights w_m^{out} such that

$$\left| h_d(n) - \sum_{m=1}^{M^{\text{out}}} w_m^{\text{out}} \left(\gamma_m^{\text{out}}\right)^n \right| < \varepsilon, \quad -N_2 \le n \le -1,$$

  where $|\gamma_m^{\text{out}}| > 1$ again using Algorithm 1.
- Compute the constant w_0 as

$$w_0 = h_d(0) - \sum_{m=1}^{M^{\text{in}}} w_m^{\text{in}} - \sum_{m=1}^{M^{\text{out}}} w_m^{\text{out}}.$$

- The resulting IIR filter $H(z)$, with impulse response $h(n)$, has $M^{\text{in}} + M^{\text{out}}$ poles and is given by

$$H(z) = w_0 + \sum_{m=1}^{M^{\text{in}}} \frac{w_m^{\text{in}}}{1 - \gamma_m^{\text{in}}/z} + \sum_{m=1}^{M^{\text{out}}} \frac{w_m^{\text{out}}}{1 - z/\gamma_m^{\text{out}}}. \tag{II.7}$$

It may not be immediately obvious why this algorithm should work. Indeed, it is rather surprising that the poles of an optimal IIR filter are related to the roots of a con-eigenpolynomial of a Hankel matrix constructed from the filter's impulse response. The theory underlying our method may be found in [3] and traced back to the work of Adamjan, Arov, and Krein (AAK theory) [6], [7], [8]. In this sense, our algorithm is related to algorithms in [11] and [12]. But while those algorithms suggest that the input sequence $h_d(n)$ be windowed in some fashion—thereby modifying (perhaps substantially) the desired frequency response $H_d(e^{j\omega})$—ours does not. Also, our algorithm leads to a way to reduce the number of poles in a sub-optimal IIR filter, which we describe below. First, let us make a few remarks about typical filter design problems.

Remark 2. In many cases of practical interest, some type of symmetry exists between $h_d(n)$ and $h_d(-n)$. In such cases a corresponding symmetry is induced between poles inside and outside the unit disk and their corresponding weights. For example, it is quite common for the impulse response to be real and symmetric,

$$h_d(n) \in \mathbb{R} \quad \text{and} \quad h_d(-n) = h_d(n),$$

in which case it is not difficult to show that poles appear at conjugate-reciprocal locations and the corresponding weights are complex conjugates, so that with a suitable reordering

$$M^{\text{in}} = M^{\text{out}}, \; w_m^{\text{in}} = \overline{w}_m^{\text{out}}, \; \text{and} \; \gamma_m^{\text{in}} = 1/\overline{\gamma}_m^{\text{out}}.$$

Additionally, poles inside the unit disk appear in conjugate pairs, so that for each $m \in [1, M^{\text{in}}]$, either both w_m^{in} and γ_m^{in} are real, or there exists a $m' \in [1, M^{\text{in}}]$ such that

$$w_m^{\text{in}} = \overline{w}_{m'}^{\text{in}} \quad \text{and} \quad \gamma_m^{\text{in}} = \overline{\gamma}_{m'}^{\text{in}}.$$

If such symmetries are present, then it is not necessary to approximate the negative half of the sequence $h_d(n)$. Instead, we approximate only the positive half, which gives us the poles and corresponding weights inside the unit disk, and then use the appropriate symmetry relations to obtain the poles and weights outside the unit disk.

Remark 3. If we are given a frequency response $H_d(e^{j\omega})$, we may use or design an appropriate quadrature rule to compute the impulse response $h_d(n)$. We need to compute a sufficient number of terms so that h_d has decayed to a level substantially smaller than ε for both negative and positive indices. This may lead to a rather large matrix $\mathbf{H}$; the algorithm we describe next in combination with the construction in Section III allows us to avoid computing with large matrices.

B. Reduction of the Number of Poles

The filter design algorithm in Section II-A is simple to implement and produces excellent filters. As input, it requires a portion of the desired impulse response, $h_d(n)$. For the output filter $H(z)$ to be satisfactory—i.e., for $\left|H_d(e^{j\omega}) - H(e^{j\omega})\right|$ to be less than the target accuracy ε—the portion of $h_d(n)$ provided as input should have decayed to a level smaller than ε. If $H_d(e^{j\omega})$ contains sharp transitions or is highly peaked, then the sequence $h_d(n)$ decays slowly, resulting in a large Hankel matrix $\mathbf{H}$ in (II.3). Computing the SVD of this matrix can be time consuming and may require extended precision arithmetic. In this section we present an alternative approach: by reducing the number of poles in a sub-optimal (but easy to obtain) IIR filter, we bypass a costly SVD.

In Section III we demonstrate how to obtain a sub-optimal (with a large number of poles) IIR filter satisfying a particular set of filter design requirements. We now describe an algorithm that takes such a sub-optimal filter as input and produces a near-optimal filter as output. We write the sub-optimal filter as

$$T_0(z) + H_0(z) = T_0(z) + \sum_{m=1}^{M_0^{\text{in}}} \frac{s_m^{\text{in}}}{1 - p_m^{\text{in}}/z} + \sum_{m=1}^{M_0^{\text{out}}} \frac{s_m^{\text{out}}}{1 - z/p_m^{\text{out}}},$$

where $|p_m^{\text{in}}| < 1$ and $|p_m^{\text{out}}| > 1$. We separate the Laurent polynomial $T_0(z)$ in the filter description to make $H_0(z)$ a proper rational function. A Laurent polynomial $T_0(z)$ is a finite linear combination of positive and negative integer powers of z. The important property is that the poles of $T_0(z)$ (if any) be located at the origin. In many cases $T_0(z)$ is simply a constant; for example, in (II.7) $T_0(z) = w_0$. We also assume that the poles of $H_0(z)$ are simple.

Given a target accuracy ε, we find a filter $H(z)$ of the form

$$H(z) = \sum_{m=1}^{M^{\text{in}}} \frac{w_m^{\text{in}}}{1 - \gamma_m^{\text{in}}/z} + \sum_{m=1}^{M^{\text{out}}} \frac{w_m^{\text{out}}}{1 - z/\gamma_m^{\text{out}}},$$

such that

$$\left|H_0(e^{j\omega}) - H(e^{j\omega})\right| < \varepsilon,$$

with $M^{\text{in}} < M_0^{\text{in}}$ and $M^{\text{out}} < M_0^{\text{out}}$. This process, which we call *reduction*, is performed separately on the poles inside and outside the unit disk. Let us describe the procedure for reducing the interior poles; the procedure for reducing exterior poles is completely analogous. For simplicity of notation, we drop superscripts and let $s_m = s_m^{\text{in}}$, $p_m = p_m^{\text{in}}$, and $M_0 = M_0^{\text{in}}$.

Algorithm 3:

- Write each weight s_m in polar form, $s_m = \rho_m e^{j\theta_m}$, and compute the square roots $c_m = \rho_m^{\frac{1}{2}} e^{j\frac{\theta_m}{2}}$.
- Construct the $M_0 \times M_0$ positive definite matrix $\mathbf{A}$, where
$$\mathbf{A}_{mn} = \frac{c_m \overline{c}_n}{1 - p_m \overline{p}_n}.$$
- Find a vector $\mathbf{u} = (u_1, \ldots, u_{M_0})^T$ satisfying the con-eigenproblem
$$\mathbf{A}\mathbf{u} = \sigma \overline{\mathbf{u}}, \tag{II.8}$$
with positive $\sigma = \sigma_M$ close to the target accuracy ε, where the con-eigenvalues are ordered, $\sigma_0 \geq \sigma_1 \geq \ldots \geq \sigma_{M_0-1}$.The matrix $\mathbf{A}$ is not necessarily symmetric, so the con-eigenvalue σ need not be a singular value of $\mathbf{A}$, but it may be shown that σ^2 is an eigenvalue of $\overline{\mathbf{A}}\mathbf{A}$ [10, §4.6].
- Use the elements of the con-eigenvector $\mathbf{u} = \mathbf{u}_M$ from the previous step to build the con-eigenfunction $u(z)$,
$$u(z) = \frac{1}{\sigma} \sum_{m=1}^{M_0} \frac{\overline{s}_m u_m}{1 - \overline{p}_m z}.$$
AAK theory guarantees that $u(z)$ has exactly M roots $\gamma_1, \gamma_2, \ldots, \gamma_M$ inside the unit disk.
- Obtain the weights $w_1, w_2, \ldots, w_M$ as the unique solution of the $M \times M$ linear system
$$\sum_{m=1}^{M} \frac{1}{1 - \gamma_m \overline{\gamma}_n} w_m = \sum_{m=1}^{M_0} \frac{1}{1 - p_m \overline{\gamma}_n} s_m.$$

The resulting IIR filter

$$H(z) = \sum_{m=1}^{M} \frac{w_m}{1 - \gamma_m/z}$$

is near-optimal and satisfies

$$\left| \sum_{m=1}^{M_0} \frac{s_m}{1 - p_m/z} - \sum_{m=1}^{M} \frac{w_m}{1 - \gamma_m/z} \right| < k\varepsilon, \quad |z| = 1,$$

for $k \approx 1$.

We do not derive this algorithm here (see [4] for details) and note that it can be obtained from the discussion in [2, §6] or justified using results from AAK theory [6], [7], [8]. We note that several significant improvements to this algorithm which use the Cauchy structure of $\mathbf{A}$ appear in [4]. The key improvements in [4] are the speed of the algorithm and a *relative* accuracy of computed con-eigenvalues resulting in accurate computations using the standard double precision arithmetic.

C. Efficient FIR Approximation of IIR Filters

In many situations the straightforward recursive realization of an IIR filter may be inconvenient. For example, an IIR filter with linear phase requires poles both inside and outside the unit disk. The data must then be accessed in reverse-time order to obtain a stable recursive realization. Also, recursive realizations implemented using fixed-point arithmetic may allow errors to accumulate, potentially reducing the filter accuracy to an unacceptable level. For these reasons, it is desirable to find FIR approximations of IIR filters.

The traditional approach to this problem uses some optimization criterion to find a fixed-length FIR filter (see, e.g., [13]); efficiency is obtained by requesting a short filter. Instead, we use the approach in [5]: we specify the target accuracy ε, but do not fix the order of the FIR filter. We obtain a factored FIR filter where each factor is particularly simple, resulting in an efficient cascade realization. We briefly present this approximation method and refer to [5] for the details.

In our method, the problem of finding a FIR filter amounts to approximating a rational function with a polynomial for some prescribed accuracy ε. The construction is based on the simple identity

$$\frac{1}{1-z} = \prod_{n=0}^{\infty}\left(1+z^{2^n}\right), \quad |z|<1. \tag{II.9}$$

We adapt the approach in [5] to IIR filters expressed as partial fractions as constructed by Algorithms 2 and 3. The following Lemma shows how to approximate a single term in the partial fraction expansion (II.7) by a FIR filter.

Lemma 4. *Let γ, w be complex-valued with $|\gamma|<1$, and let N be a positive integer. Then, for both causal and anti-causal partial fractions, we have the bound*

$$\left|\frac{w}{1-\gamma/z} - w\prod_{n=0}^{N}\left[1+\left(\frac{\gamma}{z}\right)^{2^n}\right]\right| \le |w|\,\frac{|\gamma|^{2^{N+1}}}{1-|\gamma|} \tag{II.10}$$

and

$$\left|\frac{w}{1-\gamma z} - w\prod_{n=0}^{N}\left[1+(\gamma z)^{2^n}\right]\right| \le |w|\,\frac{|\gamma|^{2^{N+1}}}{1-|\gamma|} \tag{II.11}$$

for all $|z|=1$.

Proof: From (II.9) it follows that

$$\prod_{n=0}^{N}\left[1+(\gamma z)^{2^n}\right] = \frac{1-(\gamma z)^{2^{N+1}}}{1-\gamma z} = \sum_{k=0}^{2^{N+1}-1}(\gamma z)^k, \tag{II.12}$$

for $|z|=1$. Apply (II.12) to the identity

$$\frac{w}{1-\gamma z} = w\sum_{k=0}^{\infty}(\gamma z)^k$$

to obtain (II.11). The proof of (II.10) is identical. ■

Even though the sum on the right hand side of (II.12) contains 2^{N+1} terms, the sum is represented by only $N+1$ factors in the product on the left. Given the desired accuracy ε, inequalities (II.10) and (II.11) show that the number of factors $N+1$ depends only sub-logarithmically on ε^{-1}. The next proposition shows how to approximate the entire IIR filter (II.7) by a FIR filter with a bounded absolute error. We omit the proof since it is an immediate consequence of Lemma 4.

Proposition 5. *Given an IIR filter in the form* (II.7)*, define the FIR filter*

$$\widetilde{H}(z) = w_0 + \sum_{m=1}^{M^{in}} w_m^{in}\prod_{n=0}^{N_m^{in}}\left[1+\left(\frac{\gamma_m^{in}}{z}\right)^{2^n}\right] + \sum_{m=1}^{M^{out}} w_m^{out}\prod_{n=0}^{N_m^{out}}\left[1+\left(\frac{z}{\gamma_m^{out}}\right)^{2^n}\right], \tag{II.13}$$

where N_m^{in}, $m=1,2,\ldots,M^{in}$ and N_m^{out}, $m=1,2,\ldots,M^{out}$ satisfy

$$\sum_{m=1}^{M^{in}}\left|w_m^{in}\right|\frac{\left|\gamma_m^{in}\right|^{2^{N_m^{in}+1}}}{1-\left|\gamma_m^{in}\right|} + \sum_{m=1}^{M^{out}}\left|w_m^{out}\right|\frac{\left|\gamma_m^{out}\right|^{-2^{N_m^{out}+1}}}{1-\left|\gamma_m^{out}\right|^{-1}} < \varepsilon. \tag{II.14}$$

Then the FIR approximation $\widetilde{H}(z)$ in (II.13) *satisfies*

$$\left|H(z)-\widetilde{H}(z)\right| < \varepsilon, \quad |z|=1.$$

Remark 6. If the IIR filter $H(z)$ is non-causal, then the FIR filter (II.13) is also non-causal (viz., $\widetilde{H}(z)$ contains positive powers of z). The highest positive power of z that appears in $\widetilde{H}(z)$ depends on the desired accuracy and determines the non-causal delay associated with the FIR filter. By introducing a pure delay, $z^{-2^{N_{max}+1}+1}\widetilde{H}(z)$, where $N_{\max} = \max\left\{N_1^{\text{out}}, N_2^{\text{out}}, \ldots, N_{M_{\text{out}}}^{\text{out}}\right\}$, we obtain a causal FIR filter. Hence, both causal and non-causal IIR filters yield efficient causal FIR approximations.

III. Filling the Gaps

We now combine the algorithms of Section II to produce a systematic method for designing near-optimal filters with which we create remarkable filters not obtainable (as far as we know) by other techniques. We describe lowpass filter design as a model problem. Although in this case it may be possible to obtain an equivalent design by other means, this example allows us to compare with filters designed using alternative methods. However, for the more complicated filter design problems addressed in Section V, we are not aware of alternative constructions with comparable efficiency.

Our method comprises three steps. We first create a sub-optimal IIR filter to satisfy the design criteria. Next, we use the reduction algorithm of Section II-B to find an equivalent (vis-à-vis the filter specifications) near-optimal IIR approximation of this filter. Finally, we use the FIR approximation algorithm of Section II-C to obtain an efficient FIR filter. Each step in this process introduces some approximation error, so we will allocate a portion of the total allowable error, as given in the filter specifications, to each of the three steps.

Consider the following lowpass filter specification:

$$\begin{aligned} \left|H(e^{j\omega})-1\right| &< 10^{-4}, \quad |\omega| < \frac{80}{140} \\ \left|H(e^{j\omega})\right| &< 10^{-4}, \quad |\omega| > \frac{81}{140}, \end{aligned} \tag{III.1}$$

where $\omega \in (-\pi, \pi)$. The combination of a relatively wide passband and a narrow transition region make this a challenging problem. For example, a multirate approach utilizing decimate-by-two stages would offer only marginal improvement over a single stage approach since decimation could only be performed twice. Furthermore, the passband error specification requires the phase $\arg H(e^{j\omega})$ to be nearly zero throughout the passband. Such a requirement, equivalent to requesting approximately linear phase, is challenging for many IIR filter design techniques.

A straightforward method of using the algorithms of Section II to obtain an IIR filter is to begin with the piecewise linear function $H_p(e^{j\omega})$, where

$$H_p(e^{j\omega}) = \begin{cases} 1, & \text{if } |\omega| < \frac{80}{140} \\ 81 - 140\,|\omega|, & \text{if } |\omega| \in [\frac{80}{140}, \frac{81}{140}] \\ 0, & \text{if } |\omega| > \frac{81}{140}. \end{cases}$$

However, approximating this function allocates too many poles to the sharp corners of the transition region. Instead, we will follow an approach inspired by Butterworth digital filter design (see, e.g., [14, §7.2]) and begin with an an infinitely differentiable rational function that is optimally flat in the passband and the stopband. We define the function $F(w)$ by

$$F(w) = F(w;\delta,N) = \frac{1}{1+\left(\frac{w}{\delta}\right)^{4N}}, \tag{III.2}$$

where $\delta > 0$ and N is a positive integer parameter to be specified later. $F(w)$ is infinitely differentiable on the real axis of the w-plane, and we associate the real axis with analog frequency.

For real w, the function $F(w)$ has the partial fraction expansion

$$F(w) = 2\,\mathrm{Re}\sum_{n=0}^{2N-1} \frac{r}{1-\gamma_n w},$$

where $r = (4N)^{-1}$ and

$$\gamma_n = \delta^{-1} e^{j\pi\frac{2n+1}{4N}}, \quad n = 0,1,\ldots,2N-1.$$

Applying the Möbius transform

$$w = \alpha(z) = j\frac{1-z}{1+z}, \tag{III.3}$$

we map the unit disk $|z| < 1$ onto the upper half plane $\mathrm{Im}\, w > 0$, and obtain the IIR filter

$$H_d(z) = F(\alpha(z)) = c + 2\,\mathrm{Re}\sum_{n=0}^{2N-1} \frac{\overline{s}_n}{1-\overline{p}_n z},$$

where

$$p_n = \overline{\left(\frac{\gamma_n - j}{\gamma_n + j}\right)} \quad \text{and} \quad s_n = \overline{\left(\frac{2jr\gamma_n}{\gamma_n^2+1}\right)}$$

for $n = 0,1,\ldots,2N-1$, and

$$c = 2\,\mathrm{Re}\sum_{n=0}^{2N-1} \frac{jr}{j-\gamma_n}.$$

For $|z| = 1$ on the unit circle, we write $H_d(z)$ as

$$H_d(z) = c + \sum_{n=0}^{2N-1} \frac{s_n}{1-p_n/z} + \frac{\overline{s}_n}{1-\overline{p}_n z} \tag{III.4}$$

describing a lowpass non-causal IIR filter with linear phase (in fact, $H_d(e^{j\omega})$ is real and nonnegative). The filter consists of $4N$ poles appearing as points with conjugate-reciprocal symmetry. The factor of 4 in the denominator of (III.2) was chosen to produce this 4-fold symmetry. Observe that (III.4) is in the proper form for the reduction algorithm of Section II-B, a fact we will use momentarily.

We now choose δ and N to obtain our preliminary sub-optimal IIR filter $H_d(z)$. In choosing these parameters we only concern ourselves with the accuracy of the approximation. Setting $\delta = 0.295686$ and $N = 393$ produces a filter with a maximum error of 3.3×10^{-5} in both the passband and stopband. This filter—which has 1572 poles—is obviously far from optimal. We now apply to $H_d(z)$ the reduction algorithm from Section II-B to obtain a near-optimal IIR filter $H(z)$. For the con-eigenvalue controlling the approximation error in (II.8), we select $\sigma \approx 3.7\times10^{-5}$. After applying the algorithm, the resulting IIR filter has only 30 poles, 15 inside the unit disk and 15 (conjugate-reciprocal poles) outside the unit disk. Like $H_d(z)$, the frequency response of $H(z)$ is real-valued. It has a maximum error of 8.5×10^{-5} in the passband and stopband.

As a final step, we use the approach in Section II-C to obtain a FIR approximation $\widetilde{H}(z)$ of $H(z)$. We construct the filter $\widetilde{H}(z)$ in the form (II.13), where we expand each pole so that the error of approximation in (II.14) does not exceed 1.5×10^{-5}. The resulting FIR filter satisfies the filter specifications (III.1), has linear phase, and its implementation requires 312 real additions and 161 real multiplications per output sample (we discuss the operation count in Section IV). For comparison, the FIR filter that satisfies (III.1) designed by the Parks-McClellan-Rabiner (PMR) algorithm [15, §5.1] requires 4057 taps and needs 4056 real additions and 2029 real multiplications per output sample. Alternatively, if we were to use the PMR algorithm to produce a filter with the same passband and stopband requiring 161 multiplications per sample, the resulting filter would achieve a stopband attenuation of only 0.21, compared with 10^{-4} for our filter.

The frequency responses of $H_d(z)$, $H(z)$, and $\widetilde{H}(z)$ are shown in Fig. III.1. The poles of the sub-optimal filter $H_d(z)$ and the poles of the near-optimal filter $H(z)$ are displayed in Fig. III.2, where only poles inside the unit disk are shown. The poles of $H(z)$ inside the upper half of the unit disk are listed in Table I. The table also lists how many factors each pole requires in its FIR approximation $\widetilde{H}(z)$.

A different approach to efficient FIR filter design is to decompose the frequency range $\omega \in (-\pi, \pi)$ into subbands and design an efficient FIR filter for each subband [16]. Efficiency is generally obtained by designing sparse FIR filters. Although our method is entirely different, the structure of the resulting

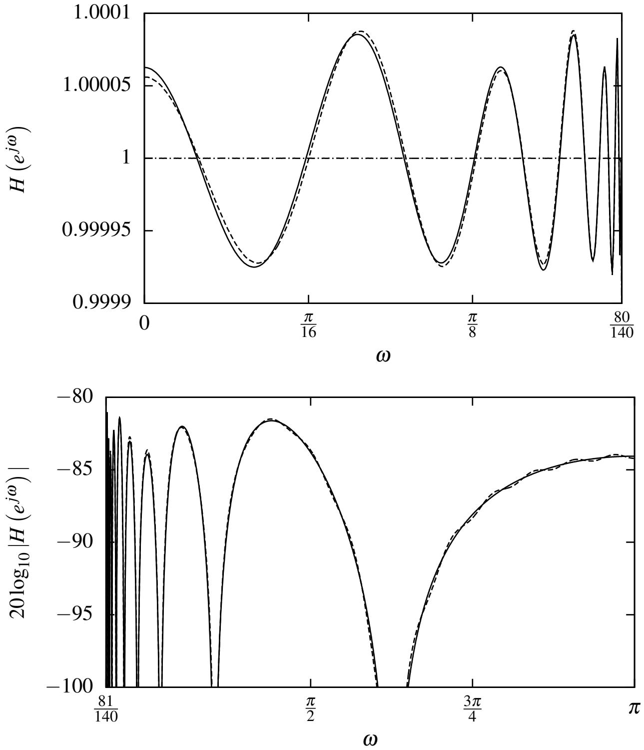


Figure III.1. Frequency response of the lowpass filters H_d (dash-dot line), H (solid line), and $\widetilde{H}$ (dashed line) in the passband (top) and the stopband (bottom). $\widetilde{H}$ is an excellent approximation of H, making their graphs almost indistinguishable.

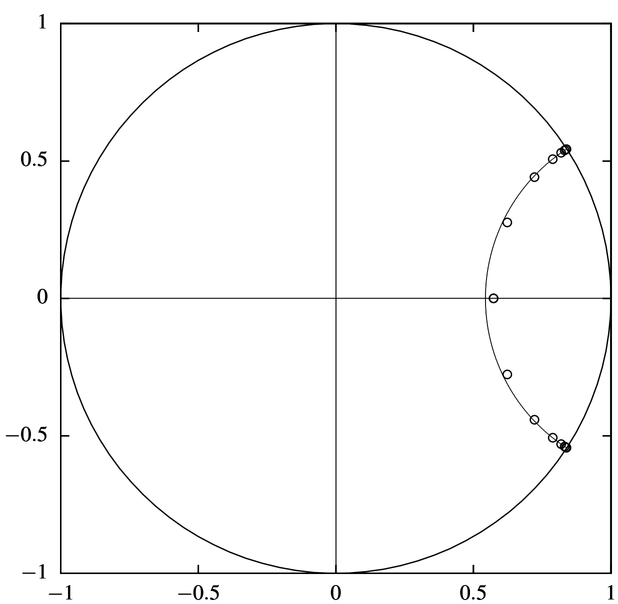


Figure III.2. Poles of the sub-optimal lowpass filter H_d (small dots) and the equivalent near-optimal filter H (open circles). The poles of H_d are so closely spaced that they appear to form a solid arc.

Table I
POLES AND WEIGHTS OF THE LOWPASS FILTER H IN SECTION III, AND THE NUMBER OF FACTORS REQUIRED FOR EACH POLE IN $\widetilde{H}$. THE CONSTANT TERM IS $w_0 = -0.18305$.

Pole z_m	Weight w_m	Factors
$0.83828+0.54323j$	$6.1855\text{e-}7-3.5136\text{e-}4j$	14
$0.83610+0.54180j$	$3.6497\text{e-}6-5.5533\text{e-}4j$	12
$0.83131+0.53862j$	$2.5683\text{e-}5-1.4427\text{e-}3j$	11
$0.81892+0.53003j$	$1.7411\text{e-}4-3.8056\text{e-}3j$	9
$0.78836+0.50650j$	$1.2072\text{e-}3-9.9390\text{e-}3j$	8
$0.72188+0.44108j$	$7.9194\text{e-}3-2.4407\text{e-}2j$	7
$0.62320+0.27635j$	$4.0707\text{e-}2-4.1560\text{e-}2j$	5
0.57310	8.2954e-2	5

FIR filters in (II.13) also has a subband interpretation. Each pole γ_m within the unit disk represents a subband: the subband is centered at the pole's argument $\arg\gamma_m$, and its bandwidth depends on the pole's proximity to the unit circle. The formula (II.9) yields an efficient FIR filter for each subband. Thus, one may view our near-optimal IIR filters as near-optimal subband decompositions of desired frequency responses.

A few remarks are in order.

Remark 7. Many existing IIR design algorithms (see [17] for an early example or [18] for a more recent one) contain a computationally expensive step to ensure that the IIR filter is causal (i.e., all poles lie within the unit disk), which obviously precludes a filter with linear phase. Our FIR approximation algorithm shows that such restrictions are not necessary, since a non-causal IIR filter may be efficiently approximated by a FIR filter to any desired accuracy. The emphasis, then, should be on minimizing the number of poles rather than ensuring that all poles lie within the unit disk.

Remark 8. Approximation by splines is another excellent method of producing sub-optimal IIR filters. They are especially useful for producing more complicated frequency responses. Splines have accurate and efficient rational approximations, so it is easy to obtain a sub-optimal IIR filter $H_d(z)$ given a sequence of spline coefficients. We ensure our reduced filters $H(z)$ are efficient by choosing splines of sufficiently high degree, so they have many continuous derivatives. Finally, we note that the spline expansion coefficients may be obtained rapidly using the algorithm in [19] and [20, Appx.], which makes use of the Fast Fourier Transform.

Remark 9. By combining the "building block" function $F(w;\delta,N)$ with the standard frequency transformations used to construct IIR filters from the classical analog filters (see, e.g., [14]), our method generates the common frequency selective filters (viz., lowpass, bandpass, highpass, and bandstop). We demonstrate further uses of $F(w;\delta,N)$ in Section V.

Remark 10. Many desired impulse responses, such as those requiring linear phase, are two-sided; they decay to the left and right of a central maximum. Conceptually (and sometimes numerically) a causal IIR approximation is obtained by windowing, truncating, and shifting the desired two-sided impulse response to the right, which (in most cases) puts the maximum amplitude significantly to the right of the origin. Directly applying this approach, as in [11], [12], produces serious numerical difficulties effectively precluding the optimality implied by the underlying AAK theory. For example, to obtain the same quality approximation as in our approach, a causal IIR filter must have a similar number of accurate impulse response coefficients as in the causal FIR filter $z^{-2^{N_{max}+1}+1}\widetilde{H}(z)$ (see Remark 6). Thus, for high accuracy, the central peak of the impulse response would be located far away from the origin. This would cause serious numerical difficulties in approximating such sequence by an efficient causal IIR filter. This may explain why examples in the literature for such approximations only deal with low accuracy filters.

IV. Filter Implementations

There are several possible implementations of our FIR approximations of IIR filters. The choice depends on the implementation medium (hardware vs. software), on the purpose of the filter, and on the filter itself. For example, if we have an IIR filter of the form

$$H(z) = \frac{P(z)}{Q(z)}, \tag{IV.1}$$

then [5] shows how to replace $1/Q(z)$ by a cascade of FIR factors, where the application of each factor requires only a single multiplication and addition. Alternatively, we may begin with an IIR filter expressed in partial fractions,

$$H(z) = w_0 + \sum_{m=1}^{M^{\text{in}}} \frac{w_m^{\text{in}}}{1-\gamma_m^{\text{in}}/z} + \sum_{m=1}^{M^{\text{out}}} \frac{w_m^{\text{out}}}{1-z/\gamma_m^{\text{out}}},$$

which is the form produced by the algorithms in Sections II-A and II-B. One implementation path is to rewrite this filter in the form (IV.1), then realize the FIR approximation as a single cascade. An alternative is to approximate each term in the partial fraction expansion separately, obtaining a FIR approximation of the form (II.13). In this way, each pole may be applied in parallel. Such an implementation is especially advantageous for software-based realizations given the current prevalence of multiprocessors. We will discuss this type of parallel realization in some detail, then conclude with several remarks about other implementation considerations.

As mentioned in Section II-C and demonstrated by the lowpass filter $H(z)$ constructed in Section III, the poles of IIR filters with real valued, even- or odd-symmetric impulse responses either have non-zero imaginary part and appear with 4-fold symmetry (conjugate-reciprocal pairs inside and outside the unit disk) or are purely real and have 2-fold reciprocal symmetry. A similar symmetry exists for the weights. With this in mind, we pick representative poles inside the upper half of the unit disk or on the real axis, and write our IIR filter as

$$\begin{aligned} H(z) = w_0 &+ \sum_{m=1}^{M^{\text{real}}} \frac{a_m}{1-p_m/z} + \frac{a_m}{1-p_m z} \\ &+ \sum_{m=1}^{M^{\text{cpx}}} \frac{\alpha_m}{1-\rho_m/z} + \frac{\overline{\alpha}_m}{1-\overline{\rho}_m/z} + \frac{\alpha_m}{1-\rho_m z} + \frac{\overline{\alpha}_m}{1-\overline{\rho}_m z}, \end{aligned} \tag{IV.2}$$

where w_0, a_m, p_m are real and $|p_m| < 1$; α_m, ρ_m are complex, $|\rho_m| < 1$ and $\operatorname{Im}\rho_m > 0$. For the real terms in the first sum, we use (II.9) to write

$$\begin{aligned} &\frac{a}{1-p/z} + \frac{a}{1-pz} = \\ &\qquad \left[2a - ap(z+z^{-1})\right] \prod_{n=0}^{\infty} \left[1 + p^{2^{n+1}} + p^{2^n}\left(z^{2^n} + z^{-2^n}\right)\right]. \end{aligned}$$

The infinite product may be truncated with bounded error using Proposition 5. For the complex terms in the second sum, we write

$$\begin{aligned} &\frac{\alpha}{1-\rho/z} + \frac{\overline{\alpha}}{1-\overline{\rho}/z} + \frac{\alpha}{1-\rho z} + \frac{\overline{\alpha}}{1-\overline{\rho} z} = \\ &\qquad \left[b_0 + b_1(z+z^{-1}) + b_2(z^2+z^{-2})\right] \\ &\qquad \times \prod_{n=0}^{\infty} \left[c_{0,n} + c_{1,n}\left(z^{2^n} + z^{-2^n}\right) + c_{2,n}\left(z^{2^{n+1}} + z^{-2^{n+1}}\right)\right], \end{aligned}$$

where the real coefficients are given by

$$\begin{aligned} b_0 &= 4\left(\operatorname{Re}\alpha + |\rho|^2 \operatorname{Re}\alpha + \operatorname{Re}\left(\rho^2\overline{\alpha}\right)\right) \\ b_1 &= -2\left(\operatorname{Re}(\rho\alpha) + 2\operatorname{Re}(\rho\overline{\alpha}) + |\rho|^2 \operatorname{Re}(\rho\overline{\alpha})\right) \\ b_2 &= 2|\rho|^2 \operatorname{Re}\alpha \\ c_{0,n} &= 1 + \left(2\operatorname{Re}\left(\rho^{2^n}\right)\right)^2 + |\rho|^{2^{n+2}} \\ c_{1,n} &= 2\operatorname{Re}\left(\rho^{2^n}\right)\left(1 + |\rho|^{2^{n+1}}\right) \\ c_{2,n} &= |\rho|^{2^{n+1}}. \end{aligned}$$

Just as for the real poles, Proposition 5 may be used to truncate the infinite product.

In this way, each real pole is approximated as a cascade where each factor requires 2 additions and 1 multiplication (we factor out the terms $1 + p^{2^{n+1}}$), followed by a factor requiring 2 additions and 2 multiplications. Each complex pole is approximated as a cascade with factors requiring 4 real additions and 2 real multiplications (where we factor out the terms $c_{0,n}$) followed by a factor requiring 4 real additions and 3 real multiplications. Let N_m^{real} denote the number of factors needed to approximate the real pole p_m, and N_m^{cpx} denote the number of factors needed for the complex pole ρ_m (see (II.13)). Then the total computational cost of a parallel implementation is

$$\begin{aligned} \#\text{Adds} &= 4N^{\text{cpx}} + 5M^{\text{cpx}} + 2N^{\text{real}} + 3M^{\text{real}} \\ \#\text{Mults.} &= 2N^{\text{cpx}} + 3M^{\text{cpx}} + N^{\text{real}} + 2M^{\text{real}} + 1, \end{aligned}$$

where

$$N^{\text{cpx}} = \sum_{m=1}^{M^{\text{cpx}}} N_m^{\text{cpx}} \quad \text{and} \quad N^{\text{real}} = \sum_{m=1}^{M^{\text{real}}} N_m^{\text{real}},$$

which includes the costs of the constant term w_0 and combining the output of each parallel component. We emphasize that these are real additions and multiplications, even though the poles and weights are generally complex.

We conclude this section with a series of remarks.

Remark 11. For software-based realizations, the parallel structure of our FIR approximations is simple to implement and yields fast codes. For hardware-based realizations, the serial cascade structure in [5] may also be considered.

Remark 12. A non-causal filter lacking a symmetric impulse response does not possess symmetry of poles inside and outside the unit disk. In this situation, the poles inside the unit disk may be applied using the standard recursive equations and the poles outside the unit disk using an appropriate FIR approximation.

Remark 13. Lowpass filters used in downsampling applications, such as digital tuning or sigma/delta A/D conversion, are an important special case. These filters are characterized by a narrow passband, narrow transition band, and tight error tolerances. Lowpass FIR filters designed by our method are especially convenient in these situations. Since factors in the cascade have terms z^{2^n}, we can apply a factor then decimate by two prior to applying the next factor. This approach dramatically reduces the memory and number of arithmetic operations required to implement the filter. We note that strategically interlacing decimation and filtering stages has been used with great success in the field of multirate signal processing (see [21] and references therein).

V. Design Examples

A. Frequency Selective Filters

We now turn to a more complicated sub-optimal filter, and thereby obtain a near-optimal filter that could not easily be obtained by other means. Let us consider a "staircase" filter $H_d(z)$ constructed by using the Möbius transform (III.3) together with the function

$$\frac{1}{2}F\left(w;\frac{3}{8},25\right)+\frac{1}{2}F\left(w;\frac{5}{8},25\right),$$

where $F(w;\delta,N)$ is defined in (III.2) (see Fig. V.1). The sub-optimal filter $H_d(z)$ is a real-valued IIR filter with 200 poles. We reduce their number by choosing a con-eigenvalue of $\sigma \approx 4.6\times 10^{-4}$ in Algorithm 3 to obtain a new filter $H(z)$ with 26 poles, of which 13 are inside the unit disk and 13 are outside. The approximation error $\left|H_d(e^{j\omega})-H(e^{j\omega})\right|$ is shown in Fig. V.1. The error displays almost exact equioscillation, consistent with our claim that IIR filters produced by our method are near-optimal. The pre- and post-reduction pole locations are shown in Fig. V.2, where only poles inside the unit disk are displayed. The pole pattern is complicated enough that it is not clear how one would produce these poles by other means. The poles, weights, and number of factors required in an FIR approximation with error less than 10^{-3} are shown in Table II (only poles inside the upper part of the unit disk are listed).

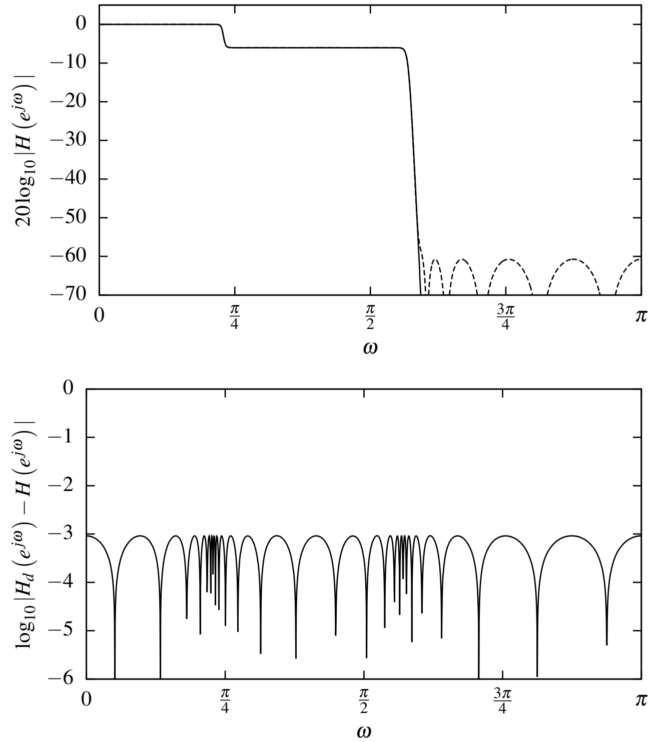


Figure V.1. The "staircase" filter H_d (solid line) and the approximation H (dashed line) (top). The equioscillation approximation error (bottom) shows that H is near-optimal.

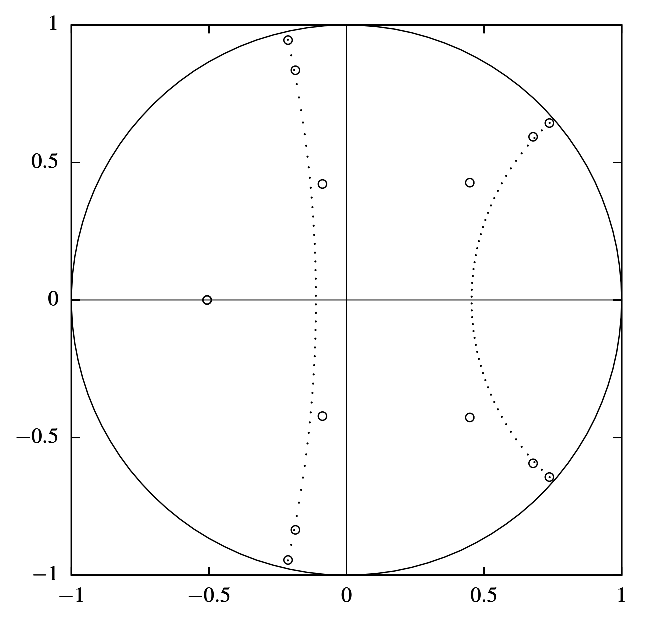


Figure V.2. Poles of the sub-optimal "staircase" filter H_d (small dots) and equivalent near-optimal filter H (open circles).

B. Quadrature Mirror Filters

Our approach allows us to widen the range of useful properties in the design of FIR Quadrature Mirror Filters (QMFs). The perfect reconstruction condition requires the low-pass filter of the QMF pair to satisfy

$$H(z)H(z^{-1})+H(-z)H(-z^{-1})=1. \qquad \text{(V.1)}$$

Table II
POLES AND WEIGHTS OF THE "STAIRCASE" FILTER H IN SECTION V-A, AND THE NUMBER OF FACTORS REQUIRED FOR EACH POLE IN $\widetilde{H}$. THE CONSTANT TERM IS $w_0 = 0.60057$.

Pole z_m	Weight w_m	Factors
$0.73729+0.64330j$	$2.6451\text{e-}5-3.7663\text{e-}3j$	9
$-0.21254+0.94461j$	$7.2305\text{e-}6-5.5944\text{e-}3j$	8
$0.67809+0.59327j$	$-4.3178\text{e-}4-1.1810\text{e-}2j$	7
$-0.18571+0.83540j$	$6.8318\text{e-}4-1.7861\text{e-}2j$	6
$0.44780+0.42670j$	$-1.3985\text{e-}2-7.0914\text{e-}2j$	4
$-8.7442\text{e-}2+0.42165j$	$-2.9205\text{e-}2-0.16703j$	4
-0.50671	$-1.4759\text{e-}2$	3

Such filters give rise to filter banks, and, with simple additional constraints, to orthonormal wavelet bases. Filter banks provide methods for efficiently applying operators to signals, in particular, operators that in the standard representation result in very long filters, such as fractional derivatives or the Hilbert transform (see, e.g., [22]). Filter banks have proven useful for applications in signal processing, numerical analysis, and data compression (see, e.g., [23]).

Depending on the application, we may request different properties of the filter (V.1). Algebraically, many of these properties are interrelated and several are mutually exclusive. For example, no FIR QMF can be symmetric but nothing prevents the design of symmetric IIR QMFs. We note that many such restrictions on properties of QMFs are fragile; i.e., for any finite accuracy these restrictions disappear, and we use this fact as a tool for the design of approximate QMFs with the desired properties. Some examples may be found in [5] and here we construct approximate IIR and FIR QMFs that are symmetric (i.e., have linear phase), efficient, and have attractive flatness and subband isolation properties.

In [24] a particularly interesting family of symmetric IIR QMFs is introduced,

$$E_{4N}(z) = \frac{(1+z)^{2N}\left((1+z)^{2N}+(-1)^N\sqrt{2}(1-z)^{2N}\right)}{(1+z)^{4N}+(1-z)^{4N}+(-1)^N\sqrt{2}(1-z^2)^{2N}}, \quad \text{(V.2)}$$

where the positive integer parameter N simultaneously controls the flatness of the passband and stopband and the width of the transition region. It may be that the value N required to achieve a sufficiently narrow transition band results in a filter that is excessively flat. We show how to use our method to obtain an efficient FIR approximation of the original QMF that retains the desired sharpness but gains efficiency by reducing the excessive flatness. An example of such a QMF frequency response is illustrated in Fig. V.3.

The filter flatness is controlled by the root of order $2N$ at $z=-1$ of $E_{4N}(z)$. To obtain a more efficient, but less flat, IIR filter, we factor our a portion of this high-order root and apply the reduction algorithm from Section II-B to the remaining terms. Observing that $E_{4N}(z)$ is real-valued on the unit circle, we select some integer $S<N$ (which controls the flatness of the new filter) and rewrite $E_{4N}(z)$ as

$$E_{4N}(z) = \left(\frac{1+z}{2}\right)^S\left(\frac{1+z^{-1}}{2}\right)^S \times\left[c+\sum_{n=1}^{2N}\frac{s_n}{1-p_n/z}+\frac{\overline{s}_n}{1-\overline{p}_n z}\right]. \quad \text{(V.3)}$$

We may now reduce the expression in brackets and construct a FIR approximation of the result. For example, we choose $N=20$, yielding an IIR filter with 40 poles inside the unit disk and 40 poles outside. This filter has an appealingly narrow transition band, but the passband is flatter than may be required for many filter bank applications. We select $S=3$ in (V.3) and apply the reduction algorithm of Section II-B to obtain a new IIR filter with only 17 poles inside the unit disk and 17 outside. Finally, we use the FIR approximation algorithm of Section II-C to obtain a FIR filter $\widetilde{E}_{80}(z)$ that approximates $E_{80}(z)$ on the unit circle with an error bounded by 10^{-8}. Because $\widetilde{E}_{80}(z)$ approximates $E_{80}(z)$ so closely, it (approximately) inherits the same properties as $E_{80}(z)$. In particular, $\widetilde{E}_{80}(z)$ is symmetric and approximately satisfies the perfect reconstruction condition (V.1) with an error that does not exceed 10^{-8} on the unit circle. It is also very flat because of the root of order 6 at $z=-1$. The approximate QMF $\widetilde{E}_{80}(z)$ and the perfect reconstruction error

$$\widetilde{E}_{80}(z)\widetilde{E}_{80}(z^{-1})+\widetilde{E}_{80}(-z)\widetilde{E}_{80}(-z^{-1})-1$$

are shown in Fig. V.3. The poles, weights, and number of factors in the FIR approximation are shown in Table III (only the poles in the upper half of the unit disk are listed).

Remark 14. We note that the order of the zero, $2S$, yields an approximate interpolating property for the filter bank [25], [26]. Directly constructing FIR QMFs with this property leads to the so-called Coiflets (design of which which is difficult, see, e.g., [25], [26]) and the resulting filters cannot be symmetric. In comparison, our construction is simple and provides additional properties.

VI. CONCLUSION

We have described a new method of designing accurate and efficient IIR and FIR filters. Our method has several advantages. First, the FIR filters it produces are more efficient than FIR filters constructed by other methods, when such constructions are even possible. Second, many properties (such as symmetric filters satisfying the perfect reconstruction condition) can only be obtained by IIR filters. Our method produces FIR filters that, with any finite accuracy, approximately possess these properties. Third, our filters have a straightforward parallel implementation. Finally, by approximating IIR filters with FIR filters, we can consider IIR filters with properties, such as linear phase, not obtainable by causal IIR filters.

REFERENCES

[1] C. Rader, "DSP history—the rise and fall of recursive digital filters," *IEEE Signal Process. Mag.*, vol. 23, pp. 46–49, Nov. 2006.

[2] G. Beylkin and L. Monzón, "On approximation of functions by exponential sums," *Appl. Comput. Harmon. Anal.*, vol. 19, no. 1, pp. 17–48, 2005.

Table III
POLES, WEIGHTS, AND NUMBER OF FACTORS REQUIRED IN THE FIR APPROXIMATION $\widetilde{E}_{80}$ OF THE IIR QMF E_{80} IN SECTION V-B. THE CONSTANT TERM IS $w_0 = -1.24533870508$.

Pole z_m	Weight w_m	Factors
$1.614962028091702\text{e-}8 + 0.9427643190768825j$	$0.1156203491819011 - 8.058245848823418\text{e-}2j$	9
$1.523267607493497\text{e-}8 + 0.9063478708933764j$	$-6.602258562520147\text{e-}2 - 0.1235199582810649j$	8
$9.452755533145547\text{e-}4 + 0.8038059179307812j$	$0.1435760298523706 - 5.972837054722412\text{e-}3j$	7
$2.385100189342754\text{e-}3 + 0.7758181038788966j$	$-8.466685108153214\text{e-}3 - 0.1405348047690257j$	7
$-4.287984710863575\text{e-}3 + 0.6711924419976365j$	$0.1673367117011427 - 4.488479171905594\text{e-}2j$	6
$1.654221356678399\text{e-}2 + 0.5487007195130383j$	$0.1691065050219905 + 4.884707929391815\text{e-}2j$	5
$3.826740048904385\text{e-}2 + 0.4009109376978119j$	$9.267944965286919\text{e-}2 + 0.1068308245943145j$	5
$4.944326086546647\text{e-}2 + 0.2253118851038638j$	$8.840369128305046\text{e-}3 + 6.662132497247104\text{e-}2j$	4
0.5130884721438124	$-1.582530456670174\text{e-}6$	4

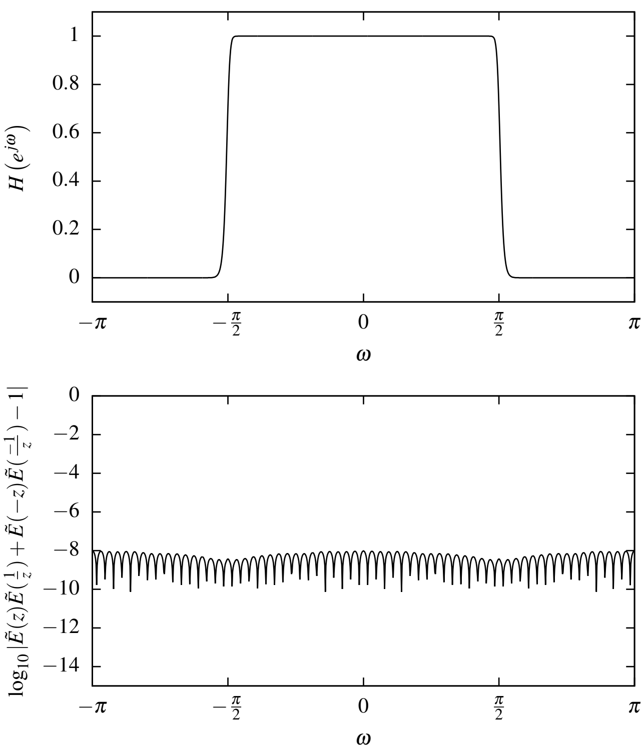


Figure V.3. Approximate QMF $\widetilde{E}_{80}$ (top) and its perfect reconstruction error (bottom).

[3] G. Beylkin and L. Monzón, "Nonlinear inversion of a band-limited Fourier transform," *Appl. Comput. Harmon. Anal.*, vol. 27, no. 3, pp. 351–366, 2009.

[4] T. S. Haut and G. Beylkin, "Fast and accurate con-eigenvalue algorithm for optimal rational approximations," *arXiv:1012.3196v2 [math.NA]*, 2011.

[5] G. Beylkin, "On factored FIR approximation of IIR filters," *Appl. Comput. Harmon. Anal.*, vol. 2, no. 3, pp. 293–298, 1995.

[6] V. M. Adamjan, D. Z. Arov, and M. G. Kreĭn, "Infinite Hankel matrices and generalized Carathéodory-Fejér and I. Schur problems," *Funkcional. Anal. i Priložen.*, vol. 2, no. 4, pp. 1–17, 1968.

[7] V. M. Adamjan, D. Z. Arov, and M. G. Kreĭn, "Infinite Hankel matrices and generalized problems of Carathéodory-Fejér and F. Riesz," *Funkcional. Anal. i Priložen.*, vol. 2, no. 1, pp. 1–19, 1968.

[8] V. M. Adamjan, D. Z. Arov, and M. G. Kreĭn, "Analytic properties of the Schmidt pairs of a Hankel operator and the generalized Schur-Takagi problem," *Mat. Sb. (N.S.)*, vol. 86(128), pp. 34–75, 1971.

[9] G. Beylkin and L. Monzón, "Approximation of functions by exponential sums revisited," *Appl. Comput. Harmon. Anal.*, vol. 28, no. 2, pp. 131–149, 2010.

[10] R. A. Horn and C. R. Johnson, *Matrix analysis*. Cambridge: Cambridge University Press, 1990.

[11] M. Gutknecht, J. Smith, and L. Trefethen, "The Carathéodory-Fejér method for recursive digital filter design," *IEEE Trans. Acoust. Speech Signal Process.*, vol. 31, pp. 1417–1426, Dec. 1983.

[12] B. S. Chen, S. C. Peng, and B. W. Chiou, "IIR filter design via optimal Hankel-norm approximation," *Circuits, Devices and Systems, IEE Proceedings G*, vol. 139, pp. 586–590, Oct. 1992.

[13] P. Kootsookos, R. Bitmead, and M. Green, "The Nehari shuffle: FIR(q) filter design with guaranteed error bounds," *IEEE Trans. Signal Process.*, vol. 40, pp. 1876–1883, Aug. 1992.

[14] T. W. Parks and C. Burrus, *Digital filter design*. Topics in digital signal processing, John Wiley & Sons, Inc., 1987.

[15] Digital Signal Processing Committee. IEEE Acoustics, Speech, and Signal Processing Society, ed., *Programs for digital signal processing*. IEEE Press, 1979.

[16] S. Mitra, A. Mahalonobis, and T. Saramaki, "A generalized structural subband decomposition of FIR filters and its application in efficient FIR filter design and implementation," *IEEE Trans. on Circuits and Systems Part II*, vol. 40, pp. 363–374, June 1993.

[17] A. Chottera and G. Jullien, "A linear programming approach to recursive digital filter design with linear phase," *IEEE Trans. Circuits Syst.*, vol. 29, pp. 139–149, Mar. 1982.

[18] A. Tarczynski, G. Cain, E. Hermanowicz, and M. Rojewski, "A WISE method for designing IIR filters," *IEEE Trans. Signal Process.*, vol. 49, pp. 1421–1432, July 2001.

[19] G. Beylkin and R. Cramer, "Toward multiresolution estimation and efficient representation of gravitational fields," *Celestial Mechanics and Dynamical Astronomy*, vol. 84, no. 1, pp. 87–104, 2002.

[20] B. Jones, G. Born, and G. Beylkin, "Comparisons of the cubed sphere gravity model with the spherical harmonics," *Journal of Guidance, Control, and Dynamics*, vol. 33, no. 2, pp. 415–425, 2010.

[21] R. Crochiere and L. Rabiner, *Multirate digital signal processing*. Prentice-Hall signal processing series, Prentice-Hall, 1983.

[22] G. Beylkin, "On the representation of operators in bases of compactly supported wavelets," *SIAM J. Numer. Anal.*, vol. 29, no. 6, pp. 1716–1740, 1992.

[23] S. Jaffard, Y. Meyer, and R. Ryan, *Wavelets: tools for science & technology*. Philadelphia, PA: Society for Industrial and Applied Mathematics (SIAM), revised ed., 2001.

[24] L. Monzón, "Linear phase perfect reconstruction filters and wavelets with even symmetry," *arXiv:1112.5214 [math.NA]*, 1999.

[25] I. Daubechies, "Orthonormal bases of compactly supported wavelets II. Variations on a theme," *SIAM J. Math. Anal.*, vol. 24, no. 2, pp. 499–519, 1993.

[26] L. Monzón, G. Beylkin, and W. Hereman, "Compactly supported wavelets based on almost interpolating and nearly linear phase filters (coiflets)," *Appl. Comput. Harmon. Anal.*, vol. 7, no. 2, pp. 184–210, 1999.

Gregory Beylkin received the Diploma (equivalent to M.S. degree) in mathematics from Leningrad (Saint Petersburg) University in 1975 and the Ph.D. degree in mathematics from New York University in 1982.

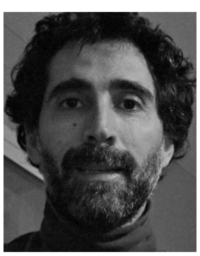
In 1982–1983 he was an Associate Research Scientist at Courant Institute of Mathematical Sciences (NYU). From 1983 to 1991 he was a Member of the Professional Staff at Schlumberger-Doll Research, Ridgefield, CT. Since 1991 he is a Professor with the Department of Applied Mathematics, University of Colorado at Boulder. His research interests include harmonic and numerical analysis, wave propagation, inverse problems, quantum chemistry, gravity modeling, and signal processing.

Dr. Beylkin is a member of the Society for Industrial and Applied Mathematics, the American Mathematical Society, and the Society of Exploration Geophysicists.

Ryan D. Lewis was born in California in 1981. He received the B.S. degree in computer science in 2003 and the M.S. degree in applied mathematics in 2009, both from the University of Colorado at Boulder, and is currently pursuing the Ph.D. degree in applied mathematics at the University of Colorado at Boulder.

He was with Northrop Grumman Corp. from 2003–2010 in Aurora, Colorado and was an intern with WesternGeco (part of Schlumberger Ltd.) in Houston, Texas in 2011. He is currently a consultant for WesternGeco. His interests include approximation algorithms, signal processing, computational optical and seismic imaging, and inverse problems.

Mr. Lewis is a member of the Society for Industrial and Applied Mathematics.



Lucas Monzón received the B.S. in mathematics from the University of Buenos Aires in Argentina in 1988 and the Ph.D. degree in mathematics from Yale University in 1994.

He is a Research Faculty with the Department of Applied Mathematics, University of Colorado at Boulder. His research interests include computational harmonic analysis, approximation theory, signal processing, seismic imaging and inverse problems.

He is currently a consultant for Schlumberger Ltd.

Dr. Monzón is a member of the Society for Industrial and Applied Mathematics.