The Asymptotics of Optimal (Equiripple) Filters

Jianhong Shen and Gilbert Strang
Department of Mathematics
Massachusetts Institute of Technology
Cambridge, MA 02139-4307

Dedicated to the memory of Wolfgang Fuchs

Abstract— **For equiripple filters, the relation among the filter length $N+1$, the transition bandwidth $\Delta\omega$, and the optimal passband and stopband errors δ_p and δ_s has been a secret for more than twenty years. This paper is aimed to solve this mystery. We derive the exact asymptotic results in the weight-free case $\delta_p = \delta_s = \delta$, which enables us to interpret and improve the existing empirical formulas. Our main results are finally combined into formula (17). In the transition band, the filter response is discovered to be asymptotically close to a scaled error function. The main tools are potential theory in the complex plane and asymptotic analysis.**

Keywords— **Asymptotics, equiripple filters, Kaiser's formula, Green's function, optimality.**

I. Introduction

It is a familiar (and happy) fact that the equiripple property of an optimal lowpass filter suggests a good algorithm for designing that filter. This is the Remez-Parks-McClellan algorithm (see Cheney [2] and Parks and McClellan [11]), which iteratively pushes down the error at its maximum point. Eventually the error has equal magnitudes and alternating signs at $N+2$ points. Since no polynomial of degree N can have $N+1$ sign changes, this equiripple filter cannot be improved at all $N+2$ points. It is optimal (in the minimax sense). The algorithm is directly available in MATLAB as `remez.m` and is very widely used.

The designer begins with a passband (ending at frequency ω_p) and a stopband (starting at ω_s) and an acceptable error. This paper considers first the weight-free case with equal errors in the passband and stopband: $\delta_p = \delta_s = \delta$. The transition bandwidth $\Delta\omega = \omega_s - \omega_p$ is critical to the relation of the filter length $N+1 = 2n+1$ to the distance δ from an ideal one-zero response. A useful formula derived experimentally by Kaiser [7] suggests an appropriate filter length. There are similar formulas in Rabiner and Gold [12] and Vaidyanathan [13]. Kaiser's is the simplest and most characteristic:

$$N \simeq \frac{20\log_{10}\delta^{-1} - 13}{2.324\Delta\omega} \tag{1}$$

For this value of N, the Remez algorithm yields the frequency response $H(\omega)$ closest to the ideal "one-zero function" $F(\omega)$ on the union of passband $|\omega| \le \omega_p$ and stopband $|\pi - \omega| \le \pi - \omega_s$. The code outputs the coefficients $h[0], \cdots, h[N]$ of this optimal lowpass filter, for which the error is approximately δ (See Fig. 1).

Our paper analyzes this relation of δ to N (or n). The error decays exponentially, $\delta \approx e^{-n\beta}/\sqrt{n}$, and

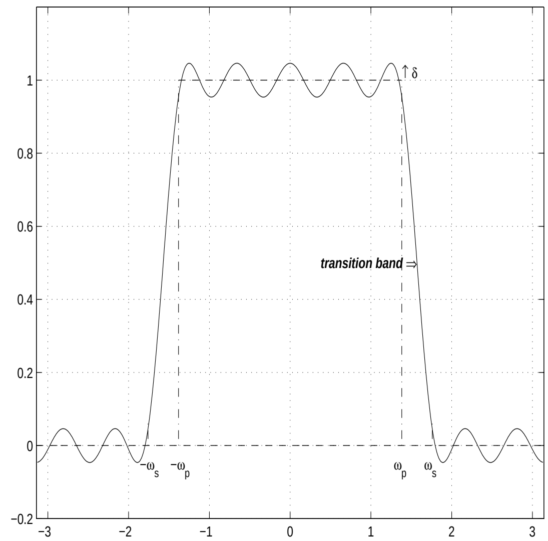

Fig. 1. The frequency response of the optimal FIR lowpass filter with $N+1=21$ coefficients and $\omega_p = 0.44\pi$, $\omega_s = 0.56\pi$.

the key problem is to compute the exponent $\beta = \beta(\omega_p, \omega_s)$. The leading term of β is controlled by $\Delta\omega = \omega_s - \omega_p$ and our asymptotic result is close to Kaiser's experiments for small δ, see Eq.(15):

$$N \simeq \frac{20\log_{10}\delta^{-1} - 10\log_{10}\log_{10}\delta^{-1}}{2.171\Delta\omega}$$

This *asymptotic* result is later modified to the *semi-empirical* formula (17), which applies to a wide range of practical parameters and is hence recommended to replace Kaiser's empirical formula.

Kaiser also discovered a nearly optimal family of filters based on the I_0–sinh function. An empirical formula similar to (1) was also established by Kaiser [7] for this family. The constant in the denominator becomes slightly smaller, which increases N. This family was analyzed theoretically by Wolfgang Fuchs in [5] and we return to it in Section 5.

The fundamental tool in the analysis is the *Green's function* $g(z)$, which solves Laplace's equation on the complement of two intervals (with a pole at infinity). This function has a unique critical point σ, and β is actually $g(\sigma)$. Since our intervals are real, the critical point is also real (and it lies in the transition band). But our problem is emphatically one of *complex* and not *real* analysis. The oscillations of a real polynomial prove that an equiripple filter is optimal, but for more information we must go deeper into the complex plane!

We give references to fundamental work of Walsh [14] and Widom [15] and Fuchs [4]. The virtue of complex analysis (Appendix D) is to permit contours of integration to be deformed. Then the leading term in an integral with a large parameter can be computed by the method of steepest descent.

The Green's function has an elementary form only in the symmetric case, when $\omega_s + \omega_p = \pi$. Then the critical frequency is $\omega_c = \pi/2$, at the center of the transition band. Our analysis is most complete in this symmetric case. Our task in all other cases (when $g(z)$ becomes an elliptic function) is to recapture the same form, in which $N\Delta\omega$ plays such a key role.

We also present early results on *nearly optimal filters*, for which δ_n is of the same order as the optimal error sequence. Unlike equiripple filters, nearly optimal filters may have closed forms and allow fast algorithms. For the symmetric and weight-free case, we propose an explicit set of interpolation points. This leads to the discovery of the asymptotic behavior of optimal filters *in the transition band.* The frequency response is close to an error function. The limit as $n \to \infty$ is the ideal brick wall filter with cutoff at the critical frequency ω_c.

Our paper has been organized as follows. Section 2 introduces an important result due to Fuchs in approximation theory. The complete asymptotic relation among design parameters in the symmetric case $\omega_p + \omega_s = \pi$ is derived in section 3. For the non-symmetric case, theoretical results as well as a MATLAB algorithm for the crucial geometric constant β are described in section 4. Asymptotic analysis is also carried out for the case of narrow transition band. In section 5, our results are compared with that of Fuchs on Kaiser's window family of filters. In section 6, we show the numerical comparison between our asymptotic formula and Kaiser's empirical one. Section 7 describes the asymptotic behavior of optimal filters in the transition band. Some proofs are included in the appendix.

II. Leading Order For δ_n

A. Leading Order for General Problem

We now present Fuchs' result on polynomial approximation on several domains in the complex plane. Let K be a compact domain with disjoint simply connected components $K_1, \cdots, K_m$. Our problem is to approximate by polynomials the function $f(z)$ that equals $h_i(z)$ on the component K_i. (The $h_i(z)$ are entire functions and not all identical.) The minimum error in the maximum norm is δ_n when the polynomials have degree at most n:

$$\delta_n = \min_{p \in P_n} \max_{z \in K} |f(z) - p(z)|.$$

Here P_n denotes the space of all polynomials of degree not greater than n.

Theorem 1 (Fuchs) *There exist a non-negative integer q, a positive number β, and two positive constants A_- and A_+, such that*

$$A_- n^{q-\frac{1}{2}} \exp(-n\beta) \leq \delta_n \leq A_+ n^{q-\frac{1}{2}} \exp(-n\beta). \quad (2)$$

Remark 1 The nonnegative integer q is determined by the objective function $f(z)$ and domain K together. It is the multiplicity of a particular critical point as a zero of a difference $h_i(z) - h_j(z)$ (Fuchs [4] gives details). Our case will automatically have $q = 0$, since $h_0(z) = 1$ and $h_1(z) = 0$.

The exponent β is a geometric constant, entirely determined by K. For $m = 2$, β is *Green's logarithmic radius of the unique critical point* of K^c. Its meaning will be explained immediately.

B. Potential Theory in the Complex Plane

Let $G(z, s)$ be the Green's function for the Laplacian on the complement K^c, which is completely characterized by the following properties:

- $G(z, s)$ is harmonic over K^c except at $z = s$, where $G(\cdot, s)$ behaves like $-\ln|z-s|$ (or $\ln|z|$ when $s = \infty$).
- For any fixed s, $G(z, s)$ goes to zero as z approaches ∂K^c, the boundary of K^c.

We are particularly interested in $g(z) = G(z, \infty)$. For any $z \in K^c$, $g(z)$ is called its *Green's logarithmic radius*, and is denoted by $|z|_K$. The function g has exactly $m-1$ *critical points* ordered by $|\sigma_1|_K \leq |\sigma_2|_K \leq \cdots \leq |\sigma_{m-1}|_K$ inside the domain K^c (Nevanlinna [10]). A critical point of g (or of K^c) means that the gradient at σ is zero. Geometrically, the level line of g through σ is self-intersected at σ (see Fig. 2). Then β in Fuchs' theorem is given by

$$\beta = |\sigma_I|_K : \ f \text{ can be continued analytically on } |z|_K < |\sigma_I|_K \text{ but not on } |z|_K < |\sigma_{I+1}|_K. \quad (3)$$

Since the h_i are not identical, σ_I does exist. When $m = 2$, β must be $|\sigma|_K = g(\sigma)$ at the unique critical point σ.

Remark 2 For optimal polynomial approximation, real analysis yields the famous "Alternation Theorem" (the equiripple property and exchange algorithm, see Cheney [2] and Rabiner and Gold [12]). The deeper asymptotic problems require complex analysis and potential theory. We recommend the

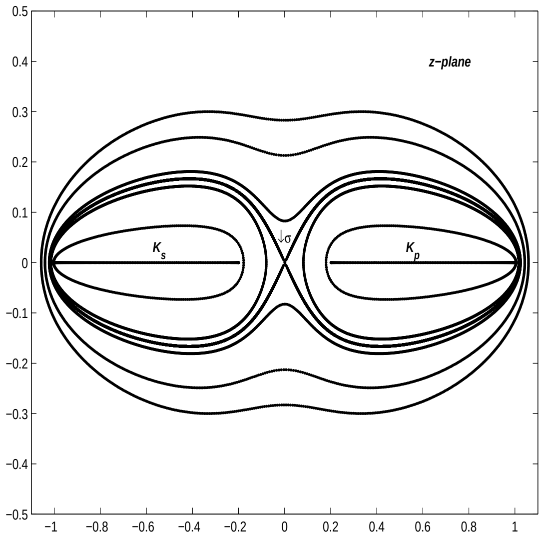

Fig. 2. The level lines and critical point σ of the Green's function associated to a domain $K = [-1, -b] \cup [a,\ 1]$. Here we show the case $a = b = 0.2$. By symmetry $\sigma = 0$.

classical monographs by Walsh [14] and Henrici [6]. Appendix D contains a short survey on the connection between optimal polynomial approximation and potential theory.

C. Leading Order for δ_n

It is natural to work in the $x = \cos\omega$ domain. Let $x_p = \cos\omega_p$ and $x_s = \cos\omega_s$. The passband and stopband become $K_p = [x_p,\ 1]$, $K_s = [-1,\ x_s]$. Then $K = K_p \cup K_s$ and σ is the unique critical point of K^c.

Lemma 1 *There exist two positive constants A_- and A_+ such that for all n*

$$A_- n^{-\frac{1}{2}} \exp(-n|\sigma|_K) \leq \delta_n \leq A_+ n^{-\frac{1}{2}} \exp(-n|\sigma|_K) \tag{4}$$

Proof: Use Theorem 1 for this special case of $m = 2$. By equation (3), $\beta = |\sigma|_K$. On the other hand, $h_0 \equiv 1$ and $h_1 \equiv 0$. Therefore $z = \sigma$ is a zero of order $q = 0$ of $h_0(z) - h_1(z)$. ■

Remark 3 Our K has only two free parameters x_p and x_s (or ω_p and ω_s). So we will also use the function symbol $\beta(x_p, x_s)$ or $\beta(\omega_p, \omega_s)$. To determine the leading order of δ_n, we have to compute β explicitly, which is the task of the next two sections. Lemma 1 leads to the following theorem in terms of logarithms. Its proof has been placed in Appendix A.

Theorem 2 *For long equiripple filters ($n \gg 1$), the asymptotic error satisfies*

$$n \simeq \frac{\ln \delta_n^{-1} - \frac{1}{2} \ln \ln \delta_n^{-1}}{\beta(x_p, x_s)} \tag{5}$$

Remark 4 The empirical formulas (Kaiser [7], Rabiner and Gold [12], and Vaidyanathan [13]) only catch the leading term $\ln \delta_n^{-1}$. They do not capture the correct β or the double logarithm term due to the factor $n^{-1/2}$ (which is overshadowed by the exponential term in all experiments).

III. Symmetric Case

In the next section, we shall see that $\beta(x_p, x_s)$ generally has no description by elementary functions. In the symmetric case $x_p + x_s = 0$, or $\omega_p + \omega_s = \pi$, the Green's function simplifies and β can be computed explicitly. Several elementary properties will be useful (referred to as Property 1,2,3 later):

1. (*Unit disk*) The Green's function for the domain $|w| \leq 1$ with source $s = 0$ is $-\ln|w|$.
2. (*Conformal equivalence*) Suppose $w = f(z)$ is a conformal mapping from a domain K_z onto a domain K_w. Assume that f is continuous up to the boundary and $f(\partial K_z) \subseteq \partial K_w$. Let z_0 be an interior point of K_z and $w_0 = f(z_0)$. Suppose $g_0(w)$ is the Green's function for K_w corresponding to source w_0. Then $g_0(f(z))$ is the Green's function of K_z corresponding to source z_0.
3. (*Pullback by covering mapping*) In Property 2, suppose that f is an analytic mapping, but z_0 is the only preimage of w_0 and all the other condi-

tions still hold. Then $g_0(f(z))/d$ is still the corresponding Green's function provided that z_0 is the $(d-1)$-multiple zero or pole of $f'(z)$.

Lemma 2 (Green's function: symmetric case) *Suppose $\omega_p + \omega_s = \pi$. Then $x_p = -x_s = a > 0$. The Green's function $g(z)$ for K^c corresponding to source $s = \infty$ is*

$$-\frac{1}{2}\ln\left|\frac{2}{1-a^2}[\, z^2 - a^2 - \sqrt{(z^2-a^2)\,(z^2-1)}\,\,] - 1\right|$$

Here the square root has K as its branch line and takes a positive value at $z = 2$.

Proof: Define

$$\phi(Z) = \frac{2}{1-a^2}[Z - a^2 - \sqrt{(Z-a^2)\,(Z-1)}\,] - 1$$

Here $Z = z^2$ folds K into a single interval $I = [a^2,\ 1]$ in the Z–plane. The inverse Joukowski transform $w = \phi(Z)$ maps the complement of I onto the unit disk D_w in the w–plane, and maps $Z = \infty$ to $w = 0$. Let $f(z) = \phi(z^2)$. Then this lemma is a direct conclusion from Properties 1 and 3 with $d = 2$. ■

Lemma 3 *Suppose $x_p = -x_s = a$. Then the exponent in the error formula is*

$$\begin{aligned}\beta &= \frac{1}{2}\ln\frac{1+x_p}{1-x_p} \\ &= \frac{1}{2}\ln\frac{1+\cos\omega_p}{1-\cos\omega_p} = \ln\cot\frac{\omega_p}{2}\end{aligned} \tag{6}$$

Proof: By symmetry, the unique critical point for K^c must be $\sigma = 0$. Therefore

$$\beta = g(0) = \frac{1}{2}\ln\frac{1+a}{1-a}.$$

■

The combination of (5) and (6) can be used for design problems when $\omega_p + \omega_s = \pi$. Notice that even in the symmetric case, β is not strictly linear in $\Delta\omega$. However, Kaiser's idea of linear approximation to β as shown in the denominator of his formula (1) is good for most applications. The estimated coefficient 2.324 can be improved by our asymptotic analysis. We now look for a theoretical formula in the symmetric case that is similar to Kaiser's.

Suppose $\Delta\omega \ll 1$. Noticing $\omega_p = \pi/2 - \Delta\omega/2$, we have by (6)

$$\beta \simeq x_p = \cos(\pi/2 - \Delta\omega/2) \simeq \Delta\omega/2.$$

In (5), we replace β by $\Delta\omega/2$ and rewrite it in terms of decibels by changing the logarithm to base 10. By ignoring the $O(1)$ term (compared with logarithms of δ_n^{-1}), we obtain the following.

Theorem 3 (Asymptotic Relation of N to δ_n) *Assume that $\omega_p + \omega_s = \pi$ and ω_p is close to $\pi/2$. Then the order is related to the ripple height δ_n by*

$$N = 2n \simeq \frac{20\log_{10}\delta_n^{-1} - 10\log_{10}\log_{10}\delta_n^{-1}}{(5\log_{10}e)\ \Delta\omega}. \tag{7}$$

Remark 5

(a) Numerically $\ln\cot(\omega_p/2)$ is close to $\Delta\omega/2$ except when $\Delta\omega$ is close to π (see Fig. 3). For most applications, $\Delta\omega$ is small. Hence $\Delta\omega/2$ is a satisfactory approximation to β. In fact, when $\Delta\omega = \pi/4$, the relative error is only $(\beta - \Delta\omega/2)/\beta \simeq 2.6\%$.

(b) Kaiser's linear coefficient 2.324 is larger than our corrected value $5\log_{10}e \simeq 2.171$. The relative error is $(2.324 - 2.171)/2.171 \simeq 7\%$. This slope deviation can be detected in Figures 5 and 6.

(c) Since the second leading term for n is a double logarithm, the number "13" in Kaiser's formula is not correct theoretically. However, it does reveal the fact that the second leading term changes very slowly. Practically we only deal with δ_n ranging from 10^{-1} to 10^{-16}. Then the double logarithm in (7) goes from 0 to 16 (and 13 is inside this range).

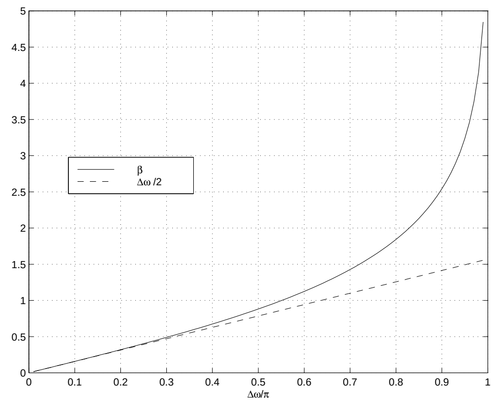

Fig. 3. $\beta = \ln\cot\frac{\omega_p}{2} \simeq \frac{\Delta\omega}{2}$. The dashed line corresponds to $\beta \simeq \frac{\Delta\omega}{2}$ and the solid line is the true $\beta(\omega_p)$ with $\omega_p = \frac{\pi}{2} - \frac{\Delta\omega}{2}$. The horizontal axis shows $\Delta\omega/\pi$.

IV. General Case

In the non-symmetric case, $\beta(x_p, x_s)$ is no longer an elementary function. In this section, we first describe the theoretical approach to determine β, and then create a numerical algorithm using MATLAB to compute it.

A. Conformal Equivalence to Annulus

Lack of symmetry ($\omega_p + \omega_s \neq \pi$) makes K^c a non-trivial doubly connected domain (DCD). Hence one has to turn to the general theory. A famous theorem says that any DCD is conformally equivalent to an annulus $A_r : r < |w| < 1$ (see Nehari [9]). The "modulus" r is uniquely determined by the DCD. In our case, this conformal mapping can be obtained in closed form using elliptic functions. The inverse mapping $z = f(w)$ is (Kober [8])

$$\frac{1+x_p}{2} - \frac{1-x_p}{2}\,\frac{\mathrm{sn}^2(\frac{K'}{\pi}\ln w;k) + \mathrm{sn}^2(\frac{K'}{\pi}\ln s;k)}{\mathrm{sn}^2(\frac{K'}{\pi}\ln w;k) - \mathrm{sn}^2(\frac{K'}{\pi}\ln s;k)}.$$

Here $0 < s < 1$, and $f(s) = \infty$.

The three parameters r, s, k are given by Freund [3] as functions of x_p and x_s.

$$k = \sqrt{\frac{2(x_p - x_s)}{(1+x_p)(1-x_s)}} \tag{8}$$

$$r = \exp\left(-\frac{\pi K_c(k)}{K'(k)}\right) \tag{9}$$

$$s = \exp\left(-\frac{\pi K_i(k)}{K'(k)}\right). \tag{10}$$

The elliptic functions $\mathrm{sn}(u;k)$, $K_c(k)$, $K_i(k)$, and $K'(k)$ are defined in Appendix B.

B. Green's Function and β

Let $g_A(w)$ denote the Green's function for the annulus A_r corresponding to the source s. Then by Property 2, $g(z) = g_A(f^{-1}(z))$ is the Green's function for K^c corresponding to $s = \infty$. Let σ and σ_A denote the unique critical points of $g(z)$ and $g_A(w)$. Then $\sigma = f(\sigma_A)$ since f preserves level lines. Hence $\beta(x_p,\ x_s) = g(\sigma) = g_A(\sigma_A)$.

Define $\lambda = \ln s/\ln r \in (0,\ 1)$. For any c inside the unit circle, the symbol $[c] = [c](w)$ denotes the Möbius transform of the unit disk associated with c:

$$[c](w) = \frac{w - c}{c^* w - 1}.$$

Then $g_A(w)$ is given by Akhiezer [1] as

$$\lambda \ln|w| - \ln|[s]| - \sum_{j=1}^{\infty}\left(\ln|[r^{2j}s]| + \ln|[\frac{r^{2j}}{s}]|\right).$$

The partial sum from 1 to J of this infinite series converges on A_r with rate $O(r^{2J})$. For small r this is quite satisfactory. However when the transition band is narrow, r defined by (8) is close to 1. So the following form of $g_A(w)$ is much better numerically:

$$\lambda \ln|w| - \ln s - \ln|S^+(w)| + \ln|S^-(w)| \tag{11}$$

Now the partial sums of S^+ and S^- from $-J$ to J give greater accuracy $O(r^{J^2})$:

$$S^+(w) = \sum_{j=-\infty}^{\infty} r^{j^2}\left(\frac{-w}{rs}\right)^j \tag{12}$$

$$S^-(w) = \sum_{j=-\infty}^{\infty} r^{j^2}\left(\frac{-ws}{r}\right)^j. \tag{13}$$

Our MATLAB code uses this form for $g_A(w)$.

Theoretically, the unique critical point σ_A can be located as the zero of the gradient vector ∇g_A. This generally requires substantial computation. The following theorem changes it to a one-dimensional optimization problem.

Theorem 4 (β by Optimization) *Consider*

$$g_A(x) = \lambda \ln(-x) - \ln s - \ln S^{+}(x) + \ln S^{-}(x) \quad (14)$$

for $-1 \le x \le -r$. *Then* $\beta(x_p, x_s) = \max g_A(x)$.

Proof: By definition, $g_A(x) \ge 0$ and $g_A(-1) = g_A(-r) = 0$. Hence $g_A(x)$ reaches its maximum value inside $(-1,\ -r)$. On the other hand, since $g_A(w)$ is symmetric with respect to y $(w = x + iy)$, $\partial g_A/\partial y$ must be zero along $(-1,\ -r)$. Therefore $\partial g_A(w)/\partial x = 0$ immediately implies a critical point of g_A. Since there is only one critical point σ_A, it must yield the maximum of $g_A(x)$. Hence

$$\beta(x_p, x_s) = g_A(\sigma_A) = \max g_A(x).$$

■

C. Algorithm and MATLAB Code

The complete elliptic function called `ellipk` in MATLAB can be used to compute K_c and K'. For the incomplete elliptic function K_i, we apply MATLAB integration `quad8` to the function `for_call`, which is simply $1/\sqrt{1 - m\sin^2 x}$ with $m = k^2$. Set $\theta = \sin^{-1}\alpha$ (α is defined in Appendix B (iii).) Then

"$K_i =$ quad8('for_call', $0,\ \theta,\ 1e-14,\ [\,],\ m$);"

computes K_i to the precision 10^{-14}. This yields r and s from (9) and (10) (by `RS.m`). Then `Green.m` uses (11)–(13) to compute the Green's function g_A on the annulus A_r. Our last program `betak.m` applies the minimization `fmin` to $-g_A(x)$ defined in (14) and finally finds β. We distinguish `betak` from MATLAB's `beta`. These MATLAB functions are available upon request.

D. Asymptotics for narrow transition

When we compute β numerically, we don't know its exact behavior as a function of ω_p and ω_s. To compare with earlier empirical formulas, we apply asymptotic analysis to β when the transition bandwidth is narrow ($\Delta\omega \ll 1$) and fixed. In practice, this narrow transition is preferred. We measure ω_p and ω_s from the mid-frequency $\omega_m = \frac{1}{2}(\omega_p + \omega_s)$:

$$\omega_p = \omega_m - \frac{\Delta\omega}{2} \qquad \text{and} \qquad \omega_s = \omega_m + \frac{\Delta\omega}{2}.$$

Since $\Delta\omega$ is fixed, $\beta(\omega_p, \omega_s)$ becomes a function only of ω_m and is denoted by $\beta(\omega_m)$.

Theorem 5 *The leading term of* $\beta(\omega_m)$ *is* $\beta(\pi/2)$ *in the range* $\Delta\omega \ll \min(\omega_m,\ \pi - \omega_m)$. *Practically, the range can be taken as (see Fig. 4):*

$$\Delta\omega \ < \ \omega_m \ < \ \pi \ - \ \Delta\omega.$$

The proof is in Appendix C.

It is Kaiser's empirical formula (1) that led to our discovery of Theorem 5. In turn, our asymptotic result provides a theoretical support to the *form* of his empirical formula. The transition bandwidth $\Delta\omega$ is crucial and the position ω_m of the transition band has small effect. Our analysis gives the correct constant in the leading term, and also the next term. With the help of Theorem 5, Theorem 4 generalizes to the non-symmetric case.

Theorem 6 *If* $\Delta\omega \ll \min(\omega_m,\ \pi - \omega_m)$, *then*

$$N = 2n \simeq \frac{20\log_{10}\delta_n^{-1} - 10\log_{10}\log_{10}\delta_n^{-1}}{(5\log_{10} e)\ \Delta\omega}. \quad (15)$$

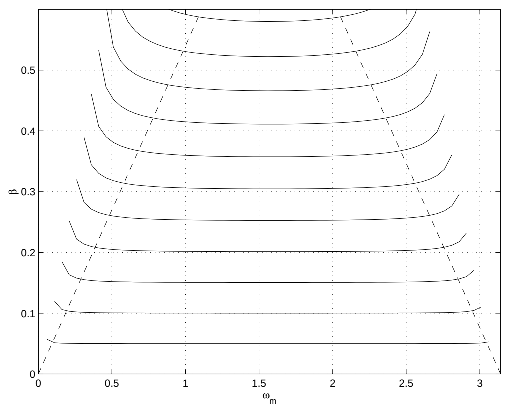

Fig. 4. Theorem 5: $\beta(\omega_m) \simeq \beta(\pi/2)$. Each solid horizontal line represents $\beta(\omega_m)$ when $\Delta\omega$ is fixed. From the bottom to the top, $\Delta\omega = 0.1 : 0.1 : 1.1$. The segments bounded by the diagonal dashed lines show the practical range inside which $\beta(\omega_m) \simeq \beta(\pi/2)$.

In practice, this yields good results for $\Delta\omega \leq \pi/4$ *and* $\omega_m \in [\Delta\omega,\ \pi - \Delta\omega]$ *by Remark 5 and Theorem 5.*

V. Kaiser's filters are near optimal

Besides the equiripple filters, another popular way of designing FIR filters is the *window method.* The ideal one-zero lowpass filter is IIR. In the frequency domain, we convolve this ideal response with the window response. The frequency response of a window is often a damped wave. The narrowness of the main lobe and side lobes determines the quality of the resulting FIR filter.

Kaiser used the non-linearly scaled zeroth–order modified Bessel function to create a family of windows with good properties. They are of limited duration in the time domain and have most of their energy concentrated at low frequency. (This is the core idea of modern wavelet analysis.) Most important, these filters are *nearly optimal*: the error sequence has the same order as that of the optimal approximation.

First, Kaiser [7] established an empirical formula for his windows when $\delta < 0.1$:

$$N \simeq \frac{20 \log_{10} \delta^{-1} - 8}{2.285\ \Delta\omega}.$$

(We have converted from Δf to $\Delta\omega = 2\pi\Delta f$.) Fuchs, Kaiser, and Landau [5] proved that for large window parameter α,

$$\delta \simeq \left[\frac{8}{\pi\Delta\omega}\right]^{\frac{1}{2}} N^{-\frac{1}{2}} \exp\left(-\frac{\Delta\omega}{4}N\right).$$

Comparing with our Eq.(2) Kaiser's windows are indeed nearly optimal (but not exactly, since $\beta \simeq \frac{\Delta\omega}{2}$ is only an approximation). Similar to the way we have proved Theorem 2, Fuchs showed that

$$N \simeq \frac{20 \log_{10} \delta^{-1} - 10 \log_{10} \log_{10} \delta^{-1}}{(5 \log_{10} e)\Delta\omega}.$$

This is exactly (15).

The Chebyshev optimal filter is completely characterized by the equiripple property. The underlying mechanism of Fuchs' result is that for large α, the side-lobes have approximately the same L_1 norms (same areas). This makes Kaiser's filters near equiripple and hence near optimal.

VI. Numerical Experiments

We use the MATLAB function `remez.m` to compute the minimal error δ_N corresponding to each N. The result is then used to test Kaiser's empirical formula and our asymptotic formula.

A. Narrow Transition

For narrow transition (this practically extends to $\Delta\omega \leq \pi/4$), Theorem 6 gives the first two leading terms of N. However, to make Eq. (15) accurate even for small N, we have to know the constant A_n appearing in the proof of Theorem 2. This means that we have to add a constant term (independent of δ_n) in the numerator of Eq. (15). Finding A_n is a mathematically open problem, but our numerical experiments indicate that we can take this constant term as $20\log_{10}\pi$. Then the following formula applies to all N:

$$N = 2n \simeq \frac{20\log_{10}(\pi\delta_n)^{-1} \; - 10\log_{10}\log_{10}\delta_n^{-1}}{(5\log_{10}e)\;\Delta\omega}. \tag{16}$$

This is very accurate for small $\Delta\omega$. Our experiments have $\Delta\omega = 0.02\pi, 0.04\pi, \cdots, 0.10\pi$ and $\omega_m = \pi/2$. For each $\Delta\omega$, first we use `remez.m` to compute the N–δ relation exactly. With this result we test the predictions by Kaiser's empirical formula (1) and our asymptotic formula (16). The test results are plotted in Fig. 5. It shows that Eq. (16) is more accurate.

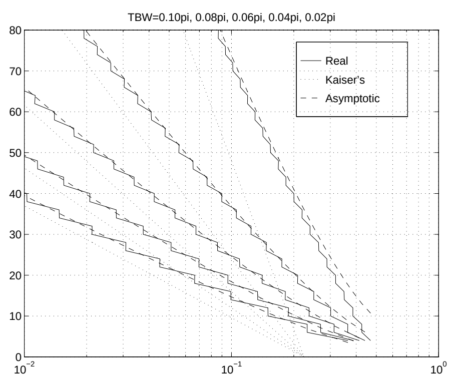

Fig. 5. Comparison of Kaiser's formula and formula (16). There are five sets of curves in the plot, one for each $\Delta\omega$. From right to left, $\Delta\omega = (0.02 : 0.02 : 0.10)\pi$. Each set contains three lines— solid, dotted, and dashed, corresponding to the real N–δ relation, Kaiser's empirical prediction, and the asymptotic prediction by Eq.(16).

B. Modified to Include Wide Transition

For wide transition, say $\Delta\omega \simeq 0.5\pi$, both Kaiser's formula and formula (16) assume that β is a linear function of $\Delta\omega$. Generally we need the original $\beta(\omega_m) \simeq \beta(\pi/2) = \ln\cot(\pi - \Delta\omega)/4$ in the denominator. Then the following formula is very accurate even for wide transition:

$$N = 2n \simeq \frac{20\log_{10}(\pi\delta_n)^{-1} \; - 10\log_{10}\log_{10}\delta_n^{-1}}{(10\log_{10}e)\;\ln\cot\frac{\pi-\Delta\omega}{4}}. \tag{17}$$

The experiments for wide transition are plotted in Fig. 6, with δ on the horizontal axis and N on the vertical.

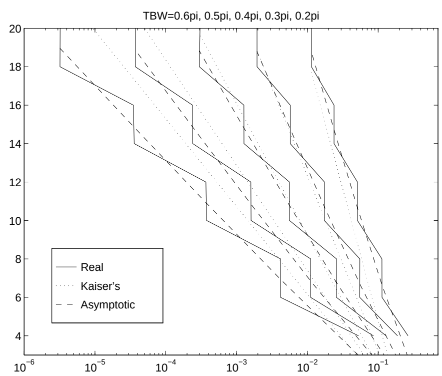

Fig. 6. Comparison of Kaiser's formula and formula (17). The five sets of curves now correspond to wider transitions $\Delta\omega = (0.2 : 0.1 : 0.6)\pi$.

So finally, we would recommend Eq. (17) for all design problems with either wide or narrow transition $\Delta\omega$, and symmetric or non-symmetric bands.

C. One Example

We compare the accuracy of the formulas through a real design problem. Suppose that $\omega_p = .5\pi$, and $\omega_s = .54\pi$. We want an equiripple filter whose passband and stopband errors are $\delta_p = \delta_s = \delta = 0.02$.

By Kaiser's formula (1), the filter length should be $N_K = 72$. The exchange algorithm

$$H_K = \mathrm{remez}(\mathrm{N_K}, [0\ .5\ .54\ 1],\ [1\ 1\ 0\ 0])$$

gives the impulse response of the equiripple filter H_K. The actual ripple height is $\delta_K = 0.0255$. Hence the relative design error is

$$r_K = \frac{|\delta - \delta_K|}{\delta} = 27.5\%.$$

The corresponding data using our asymptotic formulas (17) or (16) are:

$$N_A = 80, \qquad \delta_A = 0.0192, \qquad r_A = 4\%.$$

VII. Response in the transition band

This section describes the asymptotic behavior of the equiripple filter response $H_N^{opt}(\omega)$ inside the transition band $\omega_p \le |\omega| \le \omega_s$ as the filter length $N+1 = 2n+1$ increases. This behavior reveals the convergence of impulse responses to the ideal 0-1 filter with passband $|\omega| \le \omega_c$. Alan Oppenheim brought this problem to our attention. The results can be stated simply, but the mathematics behind them is more complicated. For proofs we refer to our forthcoming paper "Nearly Optimal Approximation on Two Intervals."

A. Nearly Optimal Filters

A family of FIR filters $H_N(\omega)$ of length $N+1$, is said to be *nearly optimal* if its error sequence

$$e_N = \|H_N(\omega) - I(\omega)\|$$

is of the same order as the optimal error sequence. This means that $e_N \le C\delta_N$ for a fixed C.

Nearly optimal filters serve two purposes. Unlike equiripple filters, they may have closed forms and allow direct mathematical analysis. Their properties should give an approximation to their counterparts (the optimal equiripple filters). Second, by relaxing the optimality, we may have a better design algorithm, such as direct interpolation. For the symmetric case $\omega_p + \omega_s = \pi$, we do find such an interpolation scheme.

Theorem 7 *Suppose $\omega_p + \omega_s = \pi$ and $x_p = \cos\omega_p = a > 0$. Define $2k$ points $x_j^{\pm}, j = 1, 2, \cdots, k$ by*

$$x_j^{\pm} = \pm\left[\frac{1+a^2}{2} + \frac{1-a^2}{2}\cos\frac{j-\frac{1}{2}}{k}\pi\right]^{\frac{1}{2}}.$$

Let $p_n(x)$ denote the unique polynomial of degree $n = 2k-1$ interpolating 1 at each x_j^+ and 0 at each x_j^-. Let $N = 2n$ and define

$$H_N(\omega) = p_n(\cos\omega).$$

Then $H_N(\omega)$ is a sequence of nearly optimal filters.

B. Asymptotics in the Transition Band

With the help of $H_N(\omega)$ just constructed, we find the following asymptotic form of the equiripple filter $H_N^{opt}(\omega)$ in the transition band.

Theorem 8 *Let $\omega_m = \dfrac{\omega_p + \omega_s}{2}$ be the mid frequency in the transition band. For $\Delta\omega = \omega_s - \omega_p \ll 1$, the leading term of $H_N^{opt}(\omega)$ on $\omega_p \le \omega \le \omega_s$ is given by*

$$H_N^{opt}(\omega) \approx \operatorname{erf}\left(\sqrt{\frac{N\beta}{4}}\,\frac{\omega_m - \omega}{\omega_s - \omega_m}\right). \qquad (18)$$

Here $\beta \approx \Delta\omega/2$ is the geometric constant appearing in previous sections and the error function $\operatorname{erf}(x)$ *is defined by*

$$\operatorname{erf}(x) = \frac{1}{\pi}\int_{-\infty}^{x} e^{-t^2}\,dt.$$

Practically, this approximation is very satisfactory for a wide range of transition bandwidths. Fig. 7 shows the case of $\Delta\omega = .1\pi$, for both symmetric and non-symmetric bands.

Computational experiment guided by our error function formula leads to the following *semi-empirical* formula for the weighted case. In minimizing the maximum deviation from the ideal filter, the stopband error is weighted by W. In practice, W can be 100. The optimal filter with heights $\delta_p = W\delta_s$ is still denoted by $H_N^{opt}(\omega)$. Then the leading term approximation in the transition band is:

$$H_N^{opt}(\omega) \simeq \operatorname{erf}\left(\sqrt{\frac{N\beta}{4}}\,\frac{\omega_m - \omega - S_N(W)}{\omega_s - \omega_m}\right),$$

with

$$S_N(W) = \frac{\ln W + \frac{1}{2}\ln\ln W}{2N}.$$

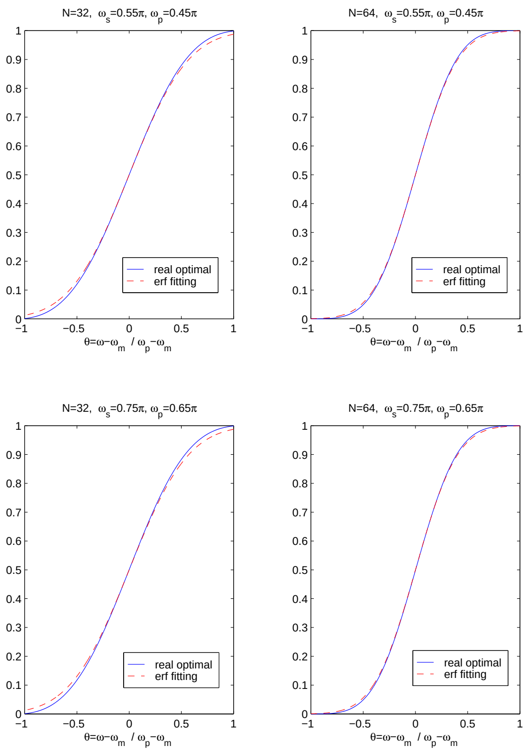

Fig. 7. Closeness to the error function. The four windows show the scaled transition band: $\theta = \dfrac{\omega-\omega_m}{\omega_p-\omega_m}$. For example, ω_p now corresponds to $\theta = 1$. The solid lines represent the optimal equiripple filters, and the dashed lines show the error function (18). For the top two, $\omega_s = .55\pi$ and $\omega_p = .45\pi$ with $N = 32, 64$. For the bottom two, $\omega_s = .75\pi$ and $\omega_p = .65\pi$ with $N = 32, 64$. The fitting improves as the filter length N increases.

This experimental expression for the shift $S_N(W)$ has successfully predicted the impulse response of optimal filters in the transition band. They still converge to the ideal filter with cutoff frequency $\omega \approx \omega_m - S_N(W)$ for narrow transition band.

VIII. Conclusions

We have proved the *asymptotic* formula (15), which relates all the key parameters in the design of equiripple filters. To be applicable to all cases of transition bandwidth, formula (15) is modified to formula (17). The numerical experiments confirm its accuracy. Filter designers can use this formula!

Acknowledgments

The authors would like to thank Jim Kaiser and Alan Oppenheim for their interest in this problem.

APPENDIX.

A. Proof of Theorem 2 (Section II)

Suppose $\delta_n = A_n n^{-\frac{1}{2}} \exp(-n\beta)$, with $A_- \le A_n \le A_+$ by Lemma 1. Taking the natural logarithm yields

$$\ln \delta_n^{-1} = -\ln A_n + \frac{1}{2}\ln n + n\beta. \tag{19}$$

The dominant term on the right is $n\beta$, which must equal the dominant term on the left. Hence $\ln \delta_n^{-1} \simeq n\beta$. This determines the leading term. To find the next term, assume $n = \dfrac{\ln \delta_n^{-1}}{\beta} + \Delta n$. By (19),

$$0 = -\ln A_n + \frac{1}{2}\ln n + \beta \Delta n.$$

As $n \gg 1$, the dominant term $\frac{1}{2}\ln n$ can only be balanced by $\beta\Delta n$, since $\ln A_n$ is bounded. Hence

$$\Delta n \simeq -\frac{\frac{1}{2}\ln n}{\beta} \simeq -\frac{\frac{1}{2}\ln\ln \delta_n^{-1}}{\beta}.$$

B. Definitions of elliptic functions (Section IV, A)

(i) $v = \operatorname{sn}(u;\ k)$ is the Jacobian elliptic function with modulus $0 < k < 1$, defined by the incomplete elliptic integral:

$$u = \int_0^v \frac{dx}{\sqrt{(1-x^2)(1-k^2x^2)}}.$$

(ii) $K_c(k) = \operatorname{sn}^{-1}(1;\ k)$ is a complete elliptic integral:

$$K_c(k) = \int_0^1 \frac{dx}{\sqrt{(1-x^2)(1-k^2x^2)}}.$$

$K'(k)$ in the expression of f is defined by $K'(k) = K_c(1-k^2)$, also a complete integral.

(iii) $K_i(k) = \operatorname{sn}^{-1}(\alpha;\ k)$ and $\alpha = \sqrt{\frac{1+x_p}{2}}$.

C. Proof of Theorem 5 (Section IV, C)

(i) The unique critical point σ of $g(z) = G(z, \infty)$ must lie inside (x_s, x_p). If the mid-frequency

$\omega_m = \frac{1}{2}(\omega_p + \omega_s)$ is below $\pi/2$, the stopband K_s is longer than the passband K_p. Hence the Green's function $g(z)$ grows more slowly near K_s. The maximum of g on $[x_s,\ x_p]$ occurs closer to x_p than to x_s, so that $\sigma > x_m$.

(ii) Define

$$d = \frac{1 + x_p x_s}{x_p + x_s} \qquad \text{and} \qquad c = d - \sqrt{d^2 - 1}.$$

Then

$$c \in [x_s,\ x_p] \qquad \text{and} \qquad c = x_m + O(\Delta x^2). \tag{20}$$

The linear fractional transform

$$z' = F(z) = \frac{z - c}{1 - cz}$$

maps $K = K_s \cup K_p$ onto a symmetric domain K' in the z'–plane:

$$K' = K'_s \cup K'_p = [-1,\ x'_s] \cup [x'_p,\ 1].$$

Here $x'_p = F(x_p) = -F(x_s) = -x'_s$.

(iii) Set $s' = F(\infty) = -1/c$ and $\sigma' = F(\sigma)$. Then σ' is the unique critical point of $G'(z',\ s')$. Since the new source s' lies inside $(-\infty, -1)$ and the new domain is symmetric, its Green's function $G'(z', s')$ grows more rapidly near K_s than K_p. Hence the maximum of $G'(\cdot, s')\big|_{(x'_s, x'_p)}$ must occur closer to x'_s, implying that $\sigma' < 0$. Now denote $F(x_m)$ by x'_m. Since F preserves the critical point as well as the order on $[x_s,\ x_p]$, we have by (i) $x'_m < \sigma' < 0$.

But $\Delta\omega \ll \min(\omega_m, \pi - \omega_m)$ implies that

$$x'_m = \frac{x_m - c}{1 - c x_m} \simeq O(x_m - c) = O(\Delta x^2).$$

So finally we have

$$\sigma' = O(\Delta x^2). \tag{21}$$

(iv) For the symmetric case in Lemma 3, when $x'_p \ll 1$,

$$G'(x', \infty) \simeq \sqrt{(x'_p)^2 - (x')^2} + O((x'_p)^2) \tag{22}$$

for all x' in $[-x'_p,\ x'_p]$. And since $z' \to x'_p/z'$ maps K' onto itself, Property 2 yields

$$G'(z',\ s') = G'(x'_p/z',\ x'_p/s'). \tag{23}$$

Therefore finally,

$$\begin{aligned}
\beta(\omega_m) &= G(\sigma,\ \infty) = G'(\sigma',\ s') \\
&= G'(0,\ s') + O(\Delta x^2) && [(21)] \\
&= G'(\infty,\ -x'_p c) + O(\Delta x^2) && [(23)] \\
&= G'(-x'_p c,\ \infty) + O(\Delta x^2) \\
&= \sqrt{(x'_p)^2 - (x'_p c)^2} + O(\Delta x^2) && [(22)] \\
&= x'_p \sqrt{1 - c^2} + O(\Delta x^2) \\
&= \frac{x_p - c}{1 - c x_p} \sqrt{1 - c^2} + O(\Delta x^2) \\
&= \frac{x_p - x_m}{1 - x_m^2} \sqrt{1 - x_m^2} + O(\Delta x^2) && [(20)] \\
&= \frac{\Delta x}{2\sqrt{1 - x_m^2}} + O(\Delta x^2) \\
&= \frac{\Delta\omega}{2} + O(\Delta\omega^2) \\
&= \beta(\frac{\pi}{2}) + O(\Delta\omega^2).
\end{aligned}$$

(v) The numerical results displayed in Fig. 4 show that practically, in the whole range of $[\Delta\omega,\ \pi - \Delta\omega]$, $\beta(\omega_m) \simeq \beta(\pi/2)$ is a satisfactory approximation.

D. The connection between polynomial approximation and complex potential theory (Section II, B)

We explain how polynomial approximation on a "reasonable" complex domain K is related to the complex potential on the complement K^c.

Let $g(z)$ and $\phi(z)$ denote the real and complex potentials on K^c generated by a source at $z = \infty$ and with the boundary ∂K^c grounded. Then $g(z) = \operatorname{Re} \ln \phi(z)$, and $|\phi(z)| = 1$ along ∂K^c. Near $z = \infty$,

$$\phi(z) = cz + a + \frac{b}{z} + \cdots$$

Assume the target function f being approximated is analytic in a neighborhood of K. If K is a disk, the n-th order Taylor expansion of f at the center can be a close approximation to the optimal polynomial. For general K, as we already see in the Parks-McClellan algorithm, the n-th degree optimal polynomial must interpolate f at a set of points $\hat{S}_n = \{\hat{z}_1, \cdots, \hat{z}_{n+1}\} \subset K$ (the crossings with 0 and 1 in the stopband and passband). The difficulty is that there is no simple way to describe $\hat{S}_n$, which apparently depends on both K and f. The key idea of *nearly optimal approximation* is to find a simple set $S_n = \{z_1, \cdots, z_{n+1}\} \subset K$ (for each n), on which interpolation yields a nearly optimal polynomial.

The connection arises from the integral representation of the interpolating polynomial $f_n(z)$:

$$f_n(z) = \int_\Gamma \frac{f(t)}{q_{n+1}(t)} \frac{q_{n+1}(t) - q_{n+1}(z)}{t - z}\, dt. \qquad (24)$$

Here $q_{n+1}(z)$ is the polynomial with roots S_n, and Γ can be any contour (containing K) in the interior of which f is analytic. The approximation error in K is,

$$\delta_n(z) = q_{n+1}(z) \int_\Gamma \frac{f(t)}{q_{n+1}(t)} \frac{dt}{t - z},$$

and

$$|\delta_n(z)| \le C(\Gamma, f)\ \|1/q_{n+1}\|_\Gamma\ |q_{n+1}(z)|.$$

$C(\Gamma, f)$ is a constant independent of n and q_{n+1}, and $\|1/q_{n+1}\|_\Gamma$ is the maximum norm on Γ. We can choose $q_{n+1}(z)$ such that $\|q_{n+1}\|_K = \|q_{n+1}\|_{\partial K} = 1$. Then $\|\delta_n\|_K \le C(\Gamma, f)\|1/q_{n+1}\|_\Gamma$. To minimize δ_n, it is important for $\|1/q_{n+1}(t)\|_\Gamma$ to be as small as possible. This leads to the following *fast growth problem.* Among all polynomials of degree $n+1$ and with $\|\cdot\|_{\partial K} = 1$, find $q_{n+1}(z)$, such that $|q_{n+1}(z)|$ is "as large as possible" for any $z \in K^c$ (thus $\|1/q_{n+1}\|_\Gamma$ can be as small as possible).

There is a natural polynomial of degree $n+1$ with this property. It is the so-called Faber polynomial and is constructed from the complex potential $\phi(z)$.

For any polynomial q_{n+1} of degree n and $\|q_{n+1}\|_{\partial K} = 1$, define $h(z) = q_{n+1}(z)/\phi^{n+1}(z)$. This is analytic on K^c and $h(\infty)$ is finite. Therefore

$$\|h\|_{K^c} = \|h\|_{\partial K} = \|q_{n+1}\|_{\partial K} = 1,$$

since $|\phi(z)| \equiv 1$ on ∂K. This leads to

$$|q_{n+1}(z)| \le |\phi^{n+1}(z)|, \qquad \text{for all } z \in K^c.$$

The growth of $|q_{n+1}(z)|$ on the complement of K^c is dominated by $|\phi^{n+1}(z)|$.

Suppose near $z = \infty$,

$$\phi^{n+1}(z) = az^{n+1} + bz^n + \cdots + c + d/z + \cdots.$$

Then the $n+1$-th Faber polynomial $F_{n+1}(z)$ is defined as the principal part of $\phi^{n+1}(z)$:

$$F_{n+1}(z) = az^{n+1} + bz^n + \cdots + c.$$

It can be shown that

$$\lim_{n\to\infty} \|F_n(z) - \phi^n(z)\|_{K^c} = 0.$$

Therefore, $|F_{n+1}(z)| \simeq |\phi^{n+1}(z)|$ achieves the optimal growth. The roots set S_n of $F_{n+1}(z)$ can be used to interpolate a given function $f(z)$. Such an interpolation is guaranteed to be nearly optimal.

For example, if K is the unit disk centered at the origin, then $F_{n+1}(z) = z^{n+1}$ and the interpolation is simply the n-th order Taylor expansion around 0. Another important example is when K is the interval $[-1,\ 1]$. Then $F_{n+1}(z)$ is the $n+1$-th Chebyshev polynomial $T_{n+1}(z) = \cos((n+1)\cos^{-1} z)$.

When the complement K^c is not simply connected (as in our two-interval case), the analysis is much more complicated but Faber's idea still works.

References

[1] Akhiezer, N.I. *Elements of the theory of elliptic functions*, American Mathematical Society, Providence, RI, 1990

[2] Cheney, E.W. *Introduction to approximation theory*, McGraw-Hill, New York, 1966

[3] Freund, R.W. "On polynomial preconditioning and asymptotic convergence factors for indefinite Hermitian matrices", *Linear Alg. Appl.* **154–156**, 259–288, 1991

[4] Fuchs, W.H.J. "On the degree of Chebyshev approximation on sets with several components", *Izv. Akad. Nauk Armyan. SSR*, **13**, No. 5-6, 396-404, 1978; See also "On Chebyshev approximation on several disjoint intervals", in Bernard Aupetit, ed., *Complex Approximation*, Birkhäuser, pp. 67-74, 1980

[5] Fuchs, W.H.J., Kaiser, J.F., and Landau, H.J. "Asymptotic behavior of a family of window functions used in non–recursive digital filter design", Technical Memorandum, Bell Laboratories, 1980

[6] Henrici, P. *Applied and computational complex analysis*, **3**, Wiley, New York, 1986

[7] Kaiser, J.F. "Nonrecursive digital filter design using the I_0–sinh window function", *Proc. 1974 IEEE Symp. Circuits and Syst.*, 20–23, April, 1974

[8] Kober, H. *Dictionary of conformal representations*, Dover, New York, 1957

[9] Nehari, Z. *Conformal mapping*, McGraw–Hill, New York, 1952

[10] Nevanlinna, R. *Analytic functions*, Springer-Verlag, New York, 1970

[11] Parks, T.W. and McClellan, J.H. "Chebyshev approximation for nonrecursive digital filters with linear phase", *IEEE Trans. on Circuit Theory*, **CT-19**, March 1972

[12] Rabiner, L.R. and Gold, B. *Theory and application of digital signal processing*, Prentice-Hall, Englewood Cliffs, NJ, 1975

[13] Vaidyanathan, P.P. *Multirate systems and filter banks*, Prentice-Hall, Englewood Cliffs, NJ, 1992

[14] Walsh, J.L. *Interpolation and approximation by rational functions in the complex domain* , American Mathematical Society, Providence, RI, 1965

[15] Widom, H. "Extremal polynomials associated with a system of curves in the complex plane", Advances in Mathematics, **3**, 127–232, 1969