The Potential Theory of Several Intervals and Its Applications

J. Shen,[1] G. Strang,[2] and A. J. Wathen[3]

[1] Department of Mathematics, UCLA,
Los Angeles, CA 90095

[2] Department of Mathematics, MIT,
Cambridge, MA 02139

[3] Oxford University Computing Laboratory,
Parks Road, Oxford OX1 3QD, UK

Abstract. Motivated both by digital filter design and polynomial-based matrix iteration methods in numerical mathematics, we study the Green's function for a domain consisting of disjoint closed intervals. The key tool is the Schwarz-Christoffel map. Asymptotic analysis produces simple and useful leading terms for the Green's function and the equilibrium distribution. Our results are applied to optimal lowpass filters and matrix iterations.

Key Words. Optimal polynomials, digital filters, matrix iterations, Green's function, equilibrium distribution, Schwarz-Christoffel mapping.

AMS Classification. Primary: 41A10, 65E05; Secondary: 65D99, 65F10.

Running Head. The complete title.

1 Introduction

Domains with disjoint closed intervals appear naturally in these applications: optimal design of digital filters (Fuchs, Kaiser and Landau [10], and Shen and Strang [14]), and polynomial based matrix iterations (Eiermann, Niethammer and Varga [4], Eiermann, Li and Varga [3], Freund [7], and Wathen, Fischer and Silvester [15, 16]). Both applications involve optimal polynomial approximations, with or without constraints. Their analysis requires knowledge of the Green's function and equilibrium distribution of the underlying (real) domain in the complex plane.

1.1 Design of digital bandpass filters

A *digital filter* is specified by a trigonometric polynomial

$$H(e^{i\omega}) = \sum h[k]e^{-ik\omega}.$$

In signal processing and especially in image processing, symmetric filters are preferred. For simplicity, we will center these "linear phase" filters around the index $k = 0$, so that $h[-k] = h[k]$ and the frequency response $H(e^{i\omega})$ is real.

An *ideal bandpass filter* D is defined for several distinct bands of interest. These are disjoint closed intervals $J_1, J_2, \cdots, J_n$ of $[0, \pi]$. D is an even and 2π–periodic zero-one function:

$$D(\omega) = \begin{cases} c_k & \text{on the interval } J_k \\ 0 & \text{elsewhere on } [0,\pi] \end{cases}$$

The value is $c_k = 1$ in the *passbands* and $c_k = 0$ in the *stopbands*. D is "ideal" because it cannot be realized by digital filters of finite length.

A natural question is: given a fixed length $N = 2m+1$, which filter gives the best approximation to D? The error is usually evaluated in the maximum norm

$$\|D - H\|_J = \max_{\omega \in J} |D(\omega) - H(e^{i\omega})|,$$

where $J = \cup_{k=1}^{n} J_k$. The symmetry assumption eventually leads to the following standard polynomial approximation problem (with $x = \cos\theta$):

$$\text{Minimize } \|D^*(x) - p(x)\|_K \text{ over all polynomials } p(x) \text{ of degree } \leq m.$$

Here $K = \cos(J)$ and $D^*(x) = D(\cos^{-1} x)$. Notice that K is a subset of $[-1, 1]$, consisting of disjoint closed intervals.

Numerically, the optimal polynomial is computed by the *Remez exchange algorithm*, developed for filter design as the *Parks-McClellan* algorithm. The MATLAB code realizing this algorithm is `remez.m` . Probably it was Fuchs who first studied this multi-interval approximation problem systematically. He obtained the leading order of the optimal error $\|D - H\|$ in [8], and in [9], he studied the distribution pattern of the extremal points (where the error is achieved). Shen and Strang [14] extended these results to improve an empirical formula of Kaiser widely used in signal processing. We will return to this in Section 4.

1.2 Polynomial based matrix iteration

In this short section, we present two classical matrix iteration methods, closely connected to optimal polynomial approximations on a multi-interval domain.

1.2.1 Semi-iterative method (SIM)

The primitive iteration for solving $A\mathbf{x}^* = \mathbf{b}$ is based on

$$\mathbf{x}_{m+1} = T\mathbf{x}_m + \mathbf{b}, \qquad \text{with } A = I - T.$$

Convergence requires a spectral radius $\rho(T) < 1$. The error vector $\mathbf{e}_m = \mathbf{x}_m - \mathbf{x}^*$ satisfies $\mathbf{e}_m = T^m\mathbf{e}_0$.

If the spectrum $\Lambda(T)$ is clustered in a set K, the *semi-iterative method* can improve the acceleration of convergence very efficiently (Eiermann, Niethammer, and Varga [4]). In the example of Davis and Hageman [1], K is cross-shaped. From the signal processing point of view, the method finds a filter $p(z) = c_0z^m + \cdots + c_m$ with $p(1) = 1$ (the "lowpass" condition) for each integer m, by which a new vector $\mathbf{y}_m$ is generated from filtering the $\mathbf{x}_i$:

$$\mathbf{y}_m = c_0\mathbf{x}_m + c_1\mathbf{x}_{m-1} + \cdots + c_m\mathbf{x}_0.$$

Denote the new error $\mathbf{y}_m - \mathbf{x}^*$ by $\mathbf{f}_m$. Then $\mathbf{f}_m = p(T)\mathbf{e}_0$. This key equation implies that the best $p(z)$ minimizes $\|p(T)\|$. (See Driscoll, Toh and Trefethen [2] for the most detailed discussion.) In practice, when the only information available is that the eigenvalues of T are contained in a known domain K, the best possible choice for $p(z)$ comes from the following minimax optimization (over-constrained):

$$\min_{p\, of degree\, m,\, p(1)=1} \max_{z\in K} |p(z)|.$$

The convergence analysis involves the Green's function of K. Domains formed from several intervals are of particular interest in applications (Eiermann, Li and Varga [3]).

1.2.2 Minimal residual method (MR)

To solve $A\mathbf{x} = \mathbf{b}$ (already preconditioned or not), one chooses $\mathbf{x}_m$ in the m-th Krylov space generated from $\mathbf{b}$:

$$\mathcal{K}_m = \text{span}\{\mathbf{b}, A\mathbf{b}, \cdots, A^{m-1}\mathbf{b}\}.$$

To minimize the norm of $\mathbf{r}_m = \mathbf{b} - A\mathbf{x}_m$ is to solve the following optimization problem:

$$\min_{p\ of\ degree\ m,\, p(0)=1} \|p(A)\mathbf{b}\|.$$

The corresponding residual vector $\mathbf{r}_m$ and iterate $\mathbf{x}_m$ are what is computed by the MR method. When it is only known that the spectrum $\Lambda(A)$ lies in some domain K, the best choice of $p(z)$ solves the minimax problem:

$$\min_{p\ of\ degree\ m,\, p(0)=1} \max_{x\in K} |p(x)|,$$

the solution of which gives a bound on the convergence of MR. Notice that this time, the constraint becomes $p(0) = 1$, instead of $p(1) = 1$.

If A is discretized from a self-adjoint differential operator, its eigenvalues are normally real. Further information can restrict $\Lambda(A)$ to specific intervals (as in the case of Wathen, Fischer, and Silvester [15]) – in particular, there must be at least two intervals if A is indefinite. The Green's function of a several-interval domain K provides useful information for the convergence analysis.

1.3 The Green's function

To begin with, let us formulate the problem abstractly. Given $2n$ points between -1 and 1:

$$-1 < a_1 < b_1 < a_2 < \cdots < a_n < b_n < 1,$$

we can define $n+1$ disjoint intervals:

$$K_1 = [-1, a_1],\ K_2 = [b_1, a_2],\ \cdots,\ K_{n+1} = [b_n, 1];\ \text{with } K = \cup_{j=1}^{n+1} K_j.$$

Between the K_j are n "gaps" (or "transition bands" in signal processing)

$$I_1 = (a_1, b_1), I_2 = (a_2, b_2), \cdots, I_n = (a_n, b_n);\ \text{with } I = \cup_{k=1}^{n} I_k.$$

Thus $I \cup K = [-1, 1]$. The complement of K in the complex plane is $K^c = \mathbb{C} \backslash K$. The Green's function $g(z)$ is the unique function with these properties: (1) $g(z)$ is harmonic on $K^c \backslash \infty$; (2) near $z = \infty$, $g(z) - \ln\ |z|$ is finite; (3) $g(z)$ is continuous up to the boundary of K^c (which is K in this case), and (4) $g(z) = 0$ on K. The main purpose of this paper is to study this function, its related functions, and their applications in the two classes of target problems. For convenience, $g(z)$ is simply called the Green's function of K.

Our main tool throughout is an integral formula based on a particular Schwarz-Christoffel map (SCM), which maps the upper half plane onto a special polygon domain. This integral formula first appeared in Widom [17] and the beautiful connections to SCM are developed by Embree and Trefethen [5]. The SCM is a suitable geometric tool for studying the Green's function and equilibrium distribution of a multi-interval domain. Each analytical property has a simple geometric picture. To the best of our knowledge, it was Trefethen who first introduced these connections.

The advantage of the SCM idea can be seen more clearly by reviewing related work. Earlier approaches of Eiermann, Li and Varga [3] and Shen and Strang [14] involve elementary polynomial transforms: for the special case of 2 intervals their approaches required a symmetry property not needed in the SCM method. Freund [7] and Fischer [6] turned to a conformal map involving elliptic functions and converted the domain to an annulus. Wathen, Fischer and Silvester [15, 16] avoided Green's functions, but employed a technical perturbation in the two-interval case for which a differential equation can be established. The SCM method is more elementary, universal, and better suited for asymptotic analysis.

In the presentation, special attention has been paid to two aspects. First, the two-interval case has been singled out because of its importance in applications and its simplicity in analysis. Second, domains with narrow gap intervals are given special care since asymptotic analysis can lead to simple and useful leading terms.

We also borrow some ideas from the Hilbert space theory and probability theory, to inspire deeper insights into this analytical problem.

1.4 Organization

Section 2 studies the Green's function for the two-interval case. The main result is the "square root law", which characterizes the special asymptotic behavior of the Green's function on a narrow gap. Section 3 discusses the Green's function and critical polynomial for a general multi-interval domain. In Section 4, we demonstrate two applications in digital filter design and numerical analysis for the Stokes equation. Section 5 studies the equilibrium distribution and its applications in filter design.

2 The Square Root Law for Two Intervals

In this section, we study the Green's function for the two-interval case:

$$K = [-1, a] \cup [b, 1], \qquad -1 < a < b < 1.$$

This case is particularly important in practice, and simpler to analyze than the general multi-interval case. For nice computational results, we refer to Embree and Trefethen [5].

2.1 The SCM and Green's function

Trefethen's central idea is to profit from the symmetry of K with respect to the real axis, by working with its (simply connected) complement K^c in the upper half plane.

In that upper half plane, define a one-parameter SCM family (ϕ_s) by

$$w = \phi_s(z) = \int_a^z \frac{(s-u)du}{\sqrt{(1-u^2)(b-u)(u-a)}}, \qquad s \in I = (a, b).$$

We take the $\sqrt{}$ branch that is positive for all $u \in I$. Under this choice, the image $\phi_s(I)$ must be a subset of the real line in the w-plane. By the general theory of SCM, ϕ_s maps the upper half plane onto a polygon in the w-plane. Its vertices are ∞ and the images of -1, a, s, b, and 1, denoted by C, A, S, B, D. The interior angles of the polygon at C, A, S, B, D are $\pi/2, \pi/2, 2\pi, \pi/2$, and $\pi/2$. With the knowledge that $\phi_s(a) = A = 0$ and $\phi_s(a, s)$ is a subset of the positive half axis, we conclude that the polygon must have the shape and orientation plotted in Figure 1.

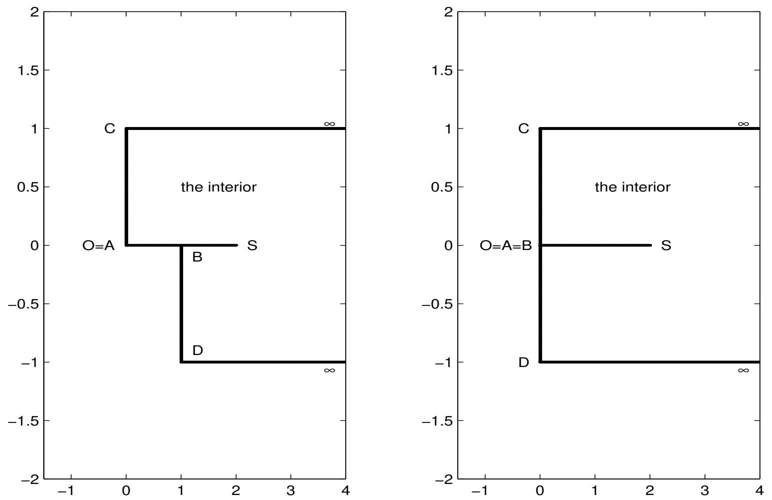

Figure 1: The image of the upper half plane and the interval $-1 < a < s < b < 1$ under a general Schwarz-Christoffel map ϕ_s (left) and with the critical parameter $s = \sigma$ (right).

A critical parameter $s = \sigma$ is needed to achieve $B = A = 0$. This amounts to requiring

$$0 = \phi_\sigma(b) = \int_a^b \frac{(\sigma - x)dx}{\sqrt{(1-x^2)(b-x)(x-a)}}.$$

The integral form leads to a probability interpretation of the critical parameter σ.

Proposition 1 (The critical parameter) *Define*

$$\gamma = \int_a^b \frac{dx}{\sqrt{(1-x^2)(b-x)(x-a)}}.$$

Let X be a random variable supported in (a, b) and with probability density function $\rho(x) = [(1-x^2)(b-x)(x-a)]^{-1/2}/\gamma$. Then $\sigma = \mathbf{E}\{X\}$, the mean value (or expectation) of X.

For the critical parameter, denote ϕ_σ simply by ϕ. Define

$$g(z) = \begin{cases} \mathrm{Re}\phi(z) & \mathrm{Im}z \geq 0 \\ g(\overline{z}) & \mathrm{Im}z < 0 \end{cases}.$$

Proposition 2 *$g(z)$ is the Green's function of K.*

Proof. From the definition, the image of the two separated intervals $\phi(K)$ is a subset of the imaginary axis (see Figure 1). Hence $g(K) = \{0\}$. Since $\phi(\mathbb{R}\backslash K)$ consists of three horizontal open lines, $\phi(z)$ can be analytically continued locally near any $x \in \mathbb{R}\backslash K$ (by Schwarz's Reflection Principle). This shows that $g(z)$ is harmonic on K^c. Finally, from our choice of the $\sqrt{}$ branch, $\phi(z) = \ln z + c_0 + c_1/z + \cdots$, near $z = \infty$. Hence $g(z) - \ln|z|$ is finite near $z = \infty$. The proof is complete. □

Recall that $z = \sigma$ is a *critical point* of $g(z)$ if the level line through σ is self-intersected. Equivalently, the gradient of $g(z)$ vanishes at σ.

Corollary 1 (The Green's function on the gap) *For all $x \in I = (a, b)$,*

$$g(x) = \int_a^x \frac{(\sigma - t)dt}{\sqrt{(1-t^2)(b-t)(t-a)}}.$$

Especially, $z = \sigma$ is the unique critical point *of $g(z)$.*

Proof. For all $x \in I$, $\phi(x)$ is real. Hence $g(x) = \phi(x)$ is given by the above integral. The same integral shows $\partial g/\partial x(\sigma) = 0$. Since $\partial g/\partial y(\sigma) = 0$ follows automatically from the vertical symmetry of $g(z)$, σ is a critical point of $g(z)$. Uniqueness follows from the general principle that the Green's function for an $n+1$-multiply connected domain has exactly n critical points (Nevanlinna [12]). □

2.2 Asymptotics for a small gap

Domains with a small gap interval (a, b) arise from both digital filter design and matrix iterations. Define the midpoint c and the half-width δ by:

$$c = \frac{a+b}{2} \qquad \text{and} \qquad \delta = \frac{b-a}{2}.$$

From now on, we assume that c belongs to a **fixed** compact set of $(-1,\ 1)$, and $\delta \to 0$.

The change of variables $\theta = (x-c)/\delta$ is useful. Then for any $f(t)$,

$$\int_a^x \frac{f(t)dt}{\sqrt{(1-t^2)(b-t)(t-a)}} = \int_{-1}^{\theta} \frac{f(c+\delta s)ds}{\sqrt{(1-s^2)(1-(c+\delta s)^2)}}. \tag{1}$$

Lemma 1 *The location of the critical point is*

$$\sigma = c + O(\delta^2).$$

A similar result was also proved in Wathen, Fischer and Silvester [15], but in a quite different context.

Proof of Lemma 1. Following the notation of Proposition 1, and using Eq. (1), we have

$$\sigma - c = \mathbf{E}\{X - c\} = \frac{1}{\gamma}\int_{-1}^{1} \frac{ds}{\sqrt{1-s^2}} \frac{\delta s}{\sqrt{1-(c+\delta s)^2}}.$$

Since c is assumed to be in a compact set of $(-1, 1)$, the following infinite series converges uniformly for small t:

$$\frac{t}{\sqrt{1-(c+t)^2}} = c_1 t + c_2 t^2 + \cdots .$$

Therefore

$$\sigma - c = \delta^2 \frac{c_2}{\gamma} \int_{-1}^{1} \frac{s^2 ds}{\sqrt{1-s^2}} + \cdots .$$

The proof is complete by noticing that γ and c_k are both of order $O(1)$. $\square$

Proposition 3 (The square root law) *Suppose the midpoint c lies in a fixed compact set of $(-1, 1)$. Then uniformly for all $x \in (a, b)$,*

$$\begin{aligned} g(x) &= \frac{1}{\sqrt{1-c^2}}\sqrt{(b-x)(x-a)} + O(\delta^2) \\ &= \sqrt{(\omega-\omega_b)(\omega_a-\omega)} + O(\Delta\omega^2). \end{aligned}$$

Here $\omega = \cos^{-1} x$, $\omega_a = \cos^{-1} a$, $\omega_b = \cos^{-1} b$, and $\Delta\omega = \omega_a - \omega_b$ (the "transition bandwidth" in signal processing).

Proof. The second line follows from the first by a change of variables to $x = \cos\omega$. To prove the first line, Lemma 1 implies

$$\begin{aligned} g(x) &= \int_{-1}^{\theta} \frac{(\sigma-c)-\delta s}{\sqrt{(1-s^2)(1-(c+\delta s)^2)}}\, ds. \\ &= -\delta \int_{-1}^{\theta} \frac{sds}{\sqrt{(1-s^2)(1-(c+\delta s)^2)}} + O(\delta^2). \end{aligned}$$

Suppose for small t,

$$\frac{1}{\sqrt{1-(c+t)^2}} = \frac{1}{\sqrt{1-c^2}} + c_1 t + \cdots .$$

Then

$$g(x) = -\frac{\delta}{\sqrt{1-c^2}} \int_{-1}^{\theta} \frac{sds}{\sqrt{1-s^2}} + O(\delta^2) = \frac{\delta}{\sqrt{1-c^2}}\sqrt{1-\theta^2} + O(\delta^2).$$

This completes the proof since $\delta^2(1-\theta^2) = (b-x)(x-a)$. $\square$

In certain cases, we have to allow $\epsilon = 1 - c$ or $\epsilon = 1 + c$ to be small too. Set $r = \delta/\epsilon$. A modification of the above proof leads to the stronger version.

Theorem 1 *Suppose $r \ll 1$. Then uniformly for all $x \in (a, b)$,*

$$g(x) = \frac{1}{\sqrt{1-c^2}}\sqrt{(b-x)(x-a)} + O(r^2\sqrt{\epsilon}).$$

Here c denotes the midpoint $(a + b)/2$.

3 The Critical Polynomial for a Multi-interval Domain

In this section, we study the Green's function for a domain with multiple interval components. K still denotes the union of $n+1$ disjoint intervals $K_1, K_2, \cdots, K_{n+1}$, and I is the union of all gap intervals $I_1, I_2, \cdots, I_n$. Thus $K \cup I = [-1, 1]$. Since K^c is $n+1$-multiply connected, the Green's function $g(z)$ must have n critical points $\sigma_1 < \sigma_2 < \cdots < \sigma_n$. The symmetry of the domain demands that all critical points sit along the real axis. It is clear that there is one critical point on each gap I_k.

3.1 The configuration polynomial and critical polynomial

We define the *configuration polynomial* $Q(z)$ by

$$Q(z) = (z^2 - 1) \prod_{k=1}^{n} (z - a_k)(z - b_k).$$

This is a monic polynomial of degree $2n+2$ and contains all the information of K. One useful property is that Q is positive on all n gaps, and in fact

$$Q(x) > 0 \qquad \text{for all } x \in \mathbb{R} \backslash K.$$

The *critical polynomial* $P(z)$ is

$$P(z) = (z - \sigma_1)(z - \sigma_2) \cdots (z - \sigma_n).$$

It is also monic and of degree n. We can suppose

$$P(z) = z^n - c_1 z^{n-1} - c_2 z^{n-2} - \cdots - c_n$$

for some set of coefficients $c_1, \cdots, c_n$.

Theorem 2 (The Green's function) *For any x in the gap I_k,*

$$g(x) = (-1)^{n+1-k} \int_{a_k}^{x} \frac{P(t)}{\sqrt{Q(t)}} dt.$$

Proof. The proof is exactly the same as the two-interval case. □

This integral form first appeared in Widom [17]. The major advantage of the SCM approach adopted here is its clear geometric meaning. Many quantities related to the Green's function have very simple correspondences in the $w = \phi(z)$ plane.

3.2 Critical polynomial: a linear algebra approach

In this subsection, we illustrate one way to compute the critical polynomial $P(x)$ for a given configuration polynomial $Q(x)$. By finding the roots of $P(x)$, we can then find all the critical points of the Green's function, which are the necessary information in most applications. In the next subsection, we provide another geometric way to compute $P(x)$.

Since the Green's function $g(x)$ vanishes on K, we have

$$\int_{a_k}^{b_k} \frac{P(t)}{\sqrt{Q(t)}} dt = 0, \tag{2}$$

for all $k = 1, \cdots, n$. Assume

$$P(x) = x^n - c_1 x^{n-1} - c_2 x^{n-2} - \cdots - c_n.$$

Then the next theorem asserts that $P(x)$ can be obtained by solving a linear system of equations.

Theorem 3 $\mathbf{c} = (c_1, c_2, \cdots, c_n)'$ *is the unique solution to the* n *by* n *linear system* $M\mathbf{c} = \mathbf{b}$*. Here the configuration matrix* $M = (M_{jk})$ *and vector* $\mathbf{b} = (\mathbf{b}_j)$ *are defined by*

$$M_{jk} = \int_{I_j} \frac{t^{n-k} dt}{\sqrt{Q(t)}}, \qquad \mathbf{b}_j = \int_{I_j} \frac{t^n dt}{\sqrt{Q(t)}}.$$

Proof. By applying Eq.(2) for $k = 1, 2, \cdots, n$, the coefficient vector $\mathbf{c}$ is easily seen to solve $M\mathbf{c} = \mathbf{b}$. Uniqueness follows from the following lemma. □

Lemma 2 *The configuration matrix is non-singular.*

Proof. Otherwise, we can find a non-zero polynomial $q(t)$ of degree no more than $n - 1$ such that

$$\int_{I_k} \frac{q(t) dt}{\sqrt{Q(t)}} = 0$$

for $k = 1, 2, \cdots, n$. Therefore $q(t)$ must change its sign on each gap I_k, implying that $q(t)$ has at least one zero on each I_k. It is impossible since $q(t)$ can have at most $n-1$ zeros. □

3.3 Critical polynomial: a geometric approach

This subsection computes the critical polynomial $P(x)$ in a geometric way based on the orthogonal projection in a Hilbert space.

On the gap set $I = I_1 \cup \cdots \cup I_n$, define a measure $d\mu = [Q(t)]^{-1/2} dt$. Then $(I, d\mu)$ is a finite measure space. We work in the Hilbert space $L^2(I, d\mu)$ with the inner product denoted by $\langle \cdot, \cdot \rangle$.

Let $\chi_k(t)$ be the indicator function of I_k. Consider two sets of linearly independent functions in $L^2(I, d\mu)$:

$$\{1, t, \cdots, t^{n-1}\} \qquad \text{and} \qquad \{\chi_1, \chi_2, \cdots, \chi_n\}.$$

The linear space $\mathbb{P}_{n-1}$ spanned by the first set contains all polynomials of degree no more than $n-1$.

The entries of the configuration matrix can be expressed by $M_{jk} = \langle \chi_j,\ t^{n-k} \rangle$. The non-singularity of M implies the existence of a dual basis in $\mathbb{P}_{n-1}$:

Corollary 2 *There exists a unique set of vectors $\{q_1, \cdots, q_n\}$ in $\mathbb{P}_{n-1}$ that is dual to $\{\chi_1, \cdots, \chi_n\}$:*

$$\langle \chi_j,\ q_k \rangle = \delta_{jk}, \qquad 1 \leq j,k \leq n.$$

With this dual basis, the critical polynomial can be computed explicitly.

Proposition 4 *The critical polynomial is given by*

$$P(t) = t^n - \sum_{k=1}^{n} \langle t^n,\ \chi_k \rangle q_n(t).$$

We now illustrate how to obtain the dual basis algorithmically.

Let S denote any non-empty subset of $[n] = \{1, 2, \cdots, n\}$. For each subset S of k elements, we shall define a monic polynomial $P_S(t)$ of degree k, subject to

$$\langle P_S,\ \chi_j \rangle = 0 \qquad \text{for any } \ j \in S. \tag{3}$$

The computation can be realized by the following inductive projection algorithm.

Step 1 For any subset $S = \{j\}$ of one element, define

$$P_S(t) = t - \frac{\langle t,\ \chi_j \rangle}{\mu(I_j)}.$$

Obviously $\langle P_S,\ \chi_j \rangle = 0$.

Step k Suppose at the end of Step $k-1$, we have defined all polynomials $P_S(t)$ subject to condition (3), for all subset S with $k-1$ elements. For any subset S with k elements, define

$$P_S(t) = t^k - \sum_{j \in S} \frac{\langle t^k,\ \chi_j \rangle}{\langle P_{S \setminus j},\ \chi_j \rangle} P_{S \setminus j}(t).$$

This is well-defined since $\langle P_{S\setminus j},\ \chi_j \rangle$ cannot be zero ($P_{S\setminus j}$ has no zero on I_j). Obviously P_S satisfies condition (3).

Proposition 5 *$P_{[n]}(t)$ is the critical polynomial. After normalization by multiplicative constants, the dual basis of $\{\chi_1, \cdots, \chi_n\}$ consists of the polynomials*

$$P_{[n]\setminus 1},\ \cdots,\ P_{[n]\setminus n}.$$

Besides its role in characterizing the critical polynomial and dual basis, this algorithm also works efficiently in practice for $n = 2, 3, 4$. For large n, the algorithm is in no way economic since at least $2^n - 1$ polynomials are to be computed.

3.4 Asymptotics for a small gap I_j

The square root law (Proposition 3) for a small gap still holds for several intervals.

Let us fix an index j. Set $c_j = \frac{1}{2}(a_j + b_j)$ and $\delta_j = \frac{1}{2}(b_j - a_j)$. For simplicity, we assume that all the other gaps $I_k : k \neq j$ are fixed (this restriction can be easily relaxed), and c_j belongs to a compact set of (b_{j-1}, a_{j+1}), and $\delta_j \to 0$. Define

$$Q_j(x) = \frac{Q(x)}{(1-x^2)(b_j - x)(x - a_j)}, \qquad P_j(x) = \prod_{k \neq j} (x - \sigma_k).$$

Then we have the following version of the square root law. Its proof is similar to Proposition 3.

Proposition 6 (The square root law) *Suppose c_j belongs to a fixed compact set of (b_{j-1}, a_{j+1}). Then uniformly for all $x \in I_j = (a_j, b_j)$,*

$$g(x) = \frac{|P_j(c_j)|}{\sqrt{Q_j(c_j)}} \sqrt{(\omega - \omega_j^b)(\omega_j^a - \omega)} + O(\delta_j^2).$$

Here $\omega = \cos^{-1} x$, $\omega_j^a = \cos^{-1} a_j$, and $\omega_j^b = \cos^{-1} b_j$.

Similar results can be established for the delicate case when c_j approaches b_{j-1} or a_{j+1}.

4 Applications of the Square Root Law

In this section, we apply our results to the two target problems mentioned earlier: the design of optimal (equiripple) lowpass filters, and the convergence analysis of the minimum residual (MR) method for solving indefinite linear systems such as arise from the Stokes equation in fluid dynamics. We anticipate more applications in other fields.

4.1 Design of equiripple lowpass filters

Our first application is to give a simple proof of Shen and Strang's result [14].

Recall that the ideal lowpass filter $D(\omega)$ equals 1 on the passband $[0, \omega_p]$ and 0 on the stopband $[\omega_s, \pi]$. With $x = \cos\omega$, the polynomial approximation problem corresponding to each fixed order m is:

$$\text{Minimize} \quad \|D^*(x) - p(x)\|_K \text{ over all polynomials of degree } \leq m.$$

Here $K = [-1, x_s] \cup [x_p, 1]$, with $x_p = \cos(\omega_p)$ and $x_s = \cos(\omega_s)$. In the L^∞ norm, the optimal filter is "equiripple" by Chebyshev's theory. The optimal error is of order $O(m^{-1/2} e^{-m\beta})$, where β is the value of the Green's function $g(x)$ at the unique critical point σ (Fuchs [8]). This exponent β is a function of $\omega_m = \frac{1}{2}(\omega_p + \omega_s)$ and the transition bandwidth $\Delta\omega = \frac{1}{2}(\omega_s - \omega_p)$. Now we are ready to give a simple proof of its leading term.

Theorem 4 (Shen and Strang [14, Theorem 5]) *In the range of* $\Delta\omega \ll \min(\omega_m, \pi - \omega_m)$, *the leading term of* $\beta(\omega_m, \Delta\omega)$ *is* $\beta(\pi/2, \Delta\omega)$.

Proof. Without loss of generality, assume $\omega_m \in [0, \pi/2]$. From our Theorem 1,

$$g(x) = \sqrt{(\omega - \omega_p)(\omega_s - \omega)} + O(r^2\sqrt{\epsilon}).$$

Since $\Delta\omega = O(\delta/\sqrt{\epsilon})$, and $\omega_m = O(\sqrt{\epsilon})$, we have

$$O(r^2\sqrt{\epsilon}) = O(\frac{\Delta\omega^2}{\omega_m}),$$

which is an order smaller than $O(\Delta\omega)$ since $\Delta\omega \ll \min(\omega_m, \pi - \omega_m)$. Therefore

$$\beta = g(\sigma) = \max_{x_s \le x \le x_p} g(x) = \max_{\omega_p \le \omega \le \omega_s} \sqrt{(\omega - \omega_p)(\omega_s - \omega)} + o(\Delta\omega) = \frac{\Delta\omega}{2} + o(\Delta\omega).$$

Hence the leading term of β is independent of ω_m. Especially one can take ω_m to be $\pi/2$. $\square$

For the symmetric case ($\omega_m = \pi/2$), Shen and Strang [14] showed that

$$\beta(\pi/2, \Delta\omega) = \ln\cot\frac{\pi - \Delta\omega}{4}.$$

For small $\Delta\omega$, this again gives $\Delta\omega/2$ as the leading order. Numerical evidence showed that taking $\ln\cot\frac{\pi-\Delta\omega}{4}$ as an approximation to $\beta(\omega_m, \Delta\omega)$ is better than $\Delta\omega/2$.

The theorem has played a crucial rule in Shen and Strang's interpretation and improvement of the celebrated empirical formula of Kaiser in optimal filter design (see Shen and Strang [14]).

4.2 Estimation of the asymptotic convergence factor

Wathen, Fischer and Silvester [15] studied the numerical solution of the Stokes problem of fluid dynamics:

$$\begin{aligned} -\nabla^2\mathbf{u} + \text{grad } p = \mathbf{f} \qquad &\text{in} \quad \Omega, \\ \text{div } \mathbf{u} = 0 \qquad &\text{in } \Omega. \end{aligned}$$

With suitable boundary conditions, the equation is discretized (by the finite element method, say) to yield a linear system of equations of the form

$$\begin{pmatrix} A & B^T \\ B & -\beta C \end{pmatrix} \cdot \begin{pmatrix} u \\ p \end{pmatrix} = \begin{pmatrix} f \\ 0 \end{pmatrix},$$

or simply $\mathcal{A}\mathbf{x} = \mathbf{b}$. The matrix C is often zero, but in any case the coefficient matrix is symmetric and *indefinite* so that $\mathcal{A}$ has both positive and negative (real) eigenvalues. For such an indefinite system, the Minimum Residual method is more robust than Conjugate Gradients. It involves only one more vector operation per iteration.

After preconditioning, the eigenvalues of the discrete Stokes operator $\mathcal{A}$ lie in two intervals:

$$K_h = [-a, -bh] \cup [ch^2, d], \qquad a, b, c, d, h > 0.$$

Here h is the characteristic mesh size that discretizes the underlying domain Ω. The asymptotic convergence factor ρ is

$$\rho = \exp(-g(0)).$$

Here $g(x)$ is the Green's function of K_h. The main result of Wathen, Fischer and Silvester is the following.

Theorem 5 (Wathen, Fischer and Silvester [15, Theorem 4.1])

$$\rho \le 1 - \sqrt{bc/ad}\ \ h^{3/2} + O(h^{5/2}).$$

Their proof strategy was based on the equiripple property of the optimal polynomial $p_m(x)$, which is as small as possible on K under the constraint $p_m(0) = 1$. By perturbing the interval a little bit, $p_m(x)$ can have $m+2$ extremal points. This makes it possible to establish a differential equation and apply asymptotic analysis successfully. The proof is complicated though it does yield a better exponent for the higher order terms than the result given below. The important leading terms are identical.

Here we use our square root law in Section 2 to give a simple proof of

$$\rho = 1 - \sqrt{bc/ad}\ \ h^{3/2} + O(h^2). \tag{4}$$

Proof. To apply Proposition 3, normalize the set K_h by introducing

$$z_* = \psi(z) = -1 + 2\frac{z+a}{d+a}.$$

$\psi(z)$ maps K_h to

$$K_* = [-1, a_*] \cup [b_*, 1], \qquad a_* = \psi(-bh), b_* = \psi(ch^2).$$

The gap size is

$$2\delta_* = b_* - a_* = \psi'(0)(ch^2 - (-bh)) + O(h^2) = \frac{2bh}{a+d} + O(h^2),$$

or, $O(\delta_*) = O(h)$. The center point is $c_* = \psi(0) + O(h)$.

Let $g_*(z_*)$ denote the Green's function for the normalized domain. Then

$$\begin{aligned} g(0) = g_*(\psi(0)) &= \frac{1}{\sqrt{1-c_*^2}} \sqrt{[\psi(0)-\psi(-bh)]\,[\psi(ch^2)-\psi(0)]} + O(h^2) \\ &= \frac{1+O(h)}{\sqrt{1-\psi^2(0)}} \sqrt{\psi'(0)bh\cdot\psi'(0)ch^2 + O(h^4)} + O(h^2) \\ &= \frac{\psi'(0)\sqrt{bch^3}}{\sqrt{1-\psi^2(0)}}\,(1+O(h)) + O(h^2) \\ &= \sqrt{bc/ad}\; h^{3/2} + O(h^2). \end{aligned}$$

Therefore

$$\rho = \exp(-g(0)) = 1 - \sqrt{bc/ad}\; h^{3/2} + O(h^2).$$

□

Similarly, by normalizing the domain and applying Theorem 1, one can give a short proof to their second result:

Theorem 6 (Wathen, Fischer and Silvester [16, Theorem 5.1]) *If the domain is $K_h = [-ah^L, -bh^l] \cup [ch^r, d]$ for some positive a, b, c, d, with $L < r$ and $L < l$, then the leading term is*

$$\rho \simeq 1 - \sqrt{bc/da}\;\; h^{(r+l-L)/2}.$$

5 The Equilibrium Distribution and Asymptotics of Extremal Points

5.1 The potential and equilibrium distribution

Closely related to the Green's function is the *equilibrium distribution*. In this section, we give an explicit expression for the equilibrium distribution of a multi-interval domain.

Let μ be a probability measure on K. The *potential* generated by μ is

$$V_\mu(z) = \int_K \ln|z-s|\,\mu(ds).$$

The potential must be subharmonic on $\mathbb{C}$ and harmonic away from K.

The *total energy* generated by μ is

$$E_\mu = -\int_K V_\mu(s)\,\mu(ds).$$

The *equilibrium distribution* ν is a special one that minimizes the total energy. The potential associated to the equilibrium distribution is called the *equilibrium potential.* For a "regular" domain, ν exists and is unique. Frostman's Theorem gives a characterization of the equilibrium potential.

Theorem 7 (Frostman's Theorem) *Let K be a "regular" compact set in $\mathbb{C}$ and ν the equilibrium distribution of K. Then*

(i) $V_\nu(z) \geq -E_\nu$ *on* $\mathbb{C}$.

(ii) $V_\nu(z) \equiv -E_\nu$ *on* K.

Conversely, a subharmonic function $V(z)$ with the following two properties must be the equilibrium potential.

(i) $V(z)$ is harmonic on the complement of K and $V(z) - \ln|z| = o(1)$ near $z = \infty$.

(ii) $V(z) = -E$ for all $z \in K$ and a certain constant E.

The inverse problem is solved by the generalized Laplacian: $\nu = \Delta V/2\pi$ (in the sense of generalized functions).

The Green's function $g(z)$ and the equilibrium potential $V_\nu(z)$ are connected by

$$g(z) = V_\nu(z) + E_\nu.$$

Our main result of this section is the expression for the equilibrium measure.

Theorem 8 *Following the notation of section 3, let $Q(z)$ and $P(z)$ denote the configuration polynomial and critical polynomial. Then the equilibrium distribution ν is supported on K and given by*

$$\nu(dx) = C\frac{|P(x)|}{\sqrt{|Q(x)|}}dx,$$

where the positive constant C normalizes ν to be a probability measure. Especially, in terms of the Schwarz-Christoffel map $\phi(z)$, ν is the pullback by $\phi(z)$ of the uniform distribution on $\phi(K)$ (the purely imaginary edge of the polygon domain. See Figure 1).

Proof. The general case is proved in the same fashion as for $K = [-1, a] \cup [b, 1]$. The equilibrium distribution is

$$\nu(dx) = \rho(x)dx.$$

By definition, for all x in the gap $I = (a, b)$,

$$V_\nu(x) = \int_K \ln|x - t|\ \rho(t)dt;$$

and the Green's function is (by Corollary 1)

$$g(x) = \int_a^x \frac{\sigma - t}{\sqrt{Q(t)}}\, dt.$$

Differentiating $g(x) = V_\nu(x) + E_\nu$ yields

$$\frac{\sigma - x}{\sqrt{Q(x)}} = \int_K \frac{\rho(t)}{x - t}\, dt$$

for all $x \in (a, b)$. Define two analytic functions on $\mathbb{C}\backslash K$:

$$\Phi_1(z) = \frac{\sigma - z}{\sqrt{Q(z)}}, \qquad \Phi_2(z) = \int_K \frac{\rho(t)}{z - t}\, dt.$$

Take $\sqrt{Q}$ to be positive on $I = (a, b)$. Since $\Phi_1(z) = \Phi_2(z)$ on I, $\Phi_1(z) \equiv \Phi_2(z)$ for all $z \in \mathbb{C}\backslash K$. Notice that $\Phi_2(z)$ is a Cauchy integral. Therefore, for any $x \in K$ (excluding the end points),

$$\rho(x) = \frac{i}{2\pi}\Big(\Phi_2(x^+) - \Phi_2(x^-)\Big),$$

where $\Phi_2(x^\pm) = \lim_{\delta\to 0^+} \Phi_2(x \pm i\delta)$ (Plemelj's formula). Hence

$$\rho(x) = \frac{i}{2\pi}\Big(\Phi_1(x^+) - \Phi_1(x^-)\Big) = \frac{1}{\pi}\frac{|\sigma - x|}{\sqrt{|Q(x)|}}.$$

This completes the proof of the first part. For the remaining part, notice that

$$d\phi = \frac{\sigma - z}{\sqrt{Q(z)}}, \qquad \text{and so} \qquad \nu(dx) = C|d\phi|.$$

□

Remark. Peherstorfer [13] also obtained the first part of the theorem based on Widom's formula [17] for the complex Green's function. The proof presented here avoids the multivalue problem caused by the multi-connectivity of the domain. The first part can also be obtained from Geronimo and Van Assche's result on polynomial maps [11]. Our second part gives a clear geometric meaning to the equilibrium measure. □

An asymptotic result for the two-interval case can thus be established based on this theorem and Lemma 1 (on the location of σ).

Corollary 3 *Suppose $K = [-1, a] \cup [b, 1]$, and a, b are contained in a fixed compact set of $(-1, 1)$. Let $\delta = (b - a)/2 \ll 1$ and choose $c < d$ in $[-1, a]$ or in $[b, 1]$ with $\epsilon = \min\{|a - d|, |c - b|\} \gg \delta$. Then*

$$\nu[c, d] = (\omega_c - \omega_d)/\pi + O((\delta/\epsilon)^2),$$

where $\omega_c = \cos^{-1} c$ and $\omega_d = \cos^{-1} d$.

5.2 Asymptotics of extremal points and applications

The design of optimal lowpass filters or bandpass filters is realized by the Remez-Parks-McClellan exchange algorithm. The algorithm needs improvement in at least two aspects. First, it has no recursive structure. This forces one to rerun the program to get a filter of length 65 even when the optimal one of length 33 is available. Secondly, the efficiency of the algorithm can be improved if one can initiate it with an approximate set of extremal points that are very close to the real ones.

These two problems are closely connected. We are led to investigate the distribution pattern of the extremal points, or equivalently, the zeros of $r_m(x) = D^*(x) - p_m(x)$, where $p_m(x)$ is the optimal polynomial of degree m and $D^*(x)$ is the ideal lowpass filter. With enough information about the zeros, it is possible to design an approximately optimal filter by a once-and-for-all interpolation. It is also possible to provide good initial points for the exchange algorithm.

This has been the major motivation of Fuchs paper [9] and the current section.

Let $F_e(x)$ denote the cumulative distribution function (c.d.f.) of the equilibrium distribution:

$$F_e(x) = \nu(-\infty, x].$$

Let $y_0, y_1, \cdots, y_{m+1}$ denote the $m+2$ extremal alternating points on K of the optimal error $r_m(x)$:

$$r_m(y_i) = \pm\|r_m\|, \qquad r_m(y_i)r_m(y_{i+1}) < 0.$$

By assigning each point y_i a uniform weight $1/(m+2)$, we can define another c.d.f.:

$$F_u(x) = \sum_{i:y_i \le x} \frac{1}{m+2}.$$

The main result of Fuchs is

Theorem 9 (Fuchs [9, Theorem 3]) *Uniformly for all $x \in \mathbb{R}$,*

$$F_u(x) - F_e(x) = O(m^{-1/5}), \qquad \textit{as } m \to \infty.$$

It follows immediately that

Corollary 4 *Let $Z_m(x)$ denote the c.d.f for the zeros of $r_m(x)$ (assigning each zero a weight $1/m$). Then uniformly for all $x \in \mathbb{R}$,*

$$Z_m(x) - F_e(x) = O(m^{-1/5}), \qquad \textit{as } m \to \infty.$$

This is because in between each pair of y_i and y_{i+1} (excluding one i), there exists exactly one zero.

Corollary 5 *Let $[c,d]$ be any interval contained in K. Then the fraction of extremal points (or zeros) on $[c,d]$ is $\nu[c,d]+O(m^{-1/5})$.*

Especially, for the two-interval case, as $m \to \infty$,

$$\#\text{ of zeros on } [-1,a] \Big/ \#\text{ of zeros on } [b,1] \to \int_{-1}^{a} \frac{\sigma - x}{\sqrt{|Q(x)|}}dx \Big/ \int_{b}^{1} \frac{x-\sigma}{\sqrt{|Q(x)|}}dx.$$

Furthermore, if the gap (a,b) is narrow, under the condition of Corollary 3, we have

Proposition 7 *The fraction of extremal points (or zeros) on $[c,d]$ is*

$$(\omega_c - \omega_d)/\pi + O\Big((\delta/\epsilon)^2 + m^{-1/5}\Big).$$

Under the Schwarz-Christoffel map $\phi(z)$, the equilibrium distribution of K becomes the uniform distribution on the imaginary edge $\phi(K)$ in Figure 1. Therefore, the numerical position of the zeros of $r_m(x)$ can be approximated by the preimages of any m points equidistributed along $\phi(K)$. The resulted interpolation leads to *nearly optimal filters* in the sense of Shen and Strang [14]. For numerical examples and a deeper discussion, see Embree and Trefethen [5].

References

[1] J. A. Davis and L. A. Hageman. An iterative method for solving the neutron transport equation in x-y geometry. *SIAM J. Appl. Math.*, 17:149–161, 1969.

[2] T. A. Driscoll, K. C. Toh, and L. N. Trefethen. Matrix iterations: the six gaps beteen potential theory and convergence. *SIAM Review*, to appear.

[3] M. Eiermann, X. Li, and R. S. Varga. On hybrid semi-iterative methods. *SIAM J. Numer. Anal.*, 26:152–168, 1989.

[4] M. Eiermann, W. Niethammer, and R. S. Varga. A study of semi-iterative methods for nonsymmetric systems of linear equations. *Numer. Math.*, 47:505–533, 1985.

[5] M. Embree and L. N. Trefethen. Green's functions for multiply connected domains via Schwarz-Christoffel mapping. Manuscript in preparation, 1998.

[6] B. Fischer. *Polynomial Based Iteration Methods for Symmetric Linear Systems.* Wiley & Sons and Teubner, 1996.

[7] R. W. Freund. On polynomial preconditioning and asymptotic convergence factors for indefinite hermitian matrices. *Linear Alg. Appl.*, 154-156:259–288, 1991.

[8] W. H. J. Fuchs. On the degree of chebyshev approximation on sets with several components. *Izv. Akad. Nauk Armyan. SSR*, 13:396–404, 1978.

[9] W. H. J. Fuchs. On Chebyshev approximation on sets with several components. In D. A. Brannan and J. G. Clunie, editors, *Aspects of Contemporary Complex Analysis*, pages 399–408. Academic Press, 1980.

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

[11] J. S. Geronimo and W. Van Assche. Orthogonal polynomials on several intervals via a polynomial mapping. *Trans. Amer. Math. Soc.*, 308(2):559–581, 1988.

[12] R. Nevanlinna. *Analytic Functions*. Springer-Verlag, New York, 1970.

[13] F. Peherstorfer. Gauss-Tchebycheff quadrature formulas. *Numer. Math.*, 58:273–286, 1990.

[14] J. Shen and G. Strang. The asymptotics of optimal (equiripple) filters. To appear in *IEEE Trans. Sig. Proc.*, 1998.

[15] A. J. Wathen, B. Fischer, and D. J. Silvester. The convergence rate of minimal residual method for the Stokes problem. *Numer. Math.*, 17:121–134, 1995.

[16] A. J. Wathen, B. Fischer, and D. J. Silvester. The convergence of iterative solution methods for symmetric and indefinite linear systems. In D.F. Griffiths, D.J. Higham, and G.A. Watson, editors, *Numerical Analysis 1997*, Pitman Research Notes in Mathematics, pages 230–243. Addison Wesley Longman, 1998.

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