Discrete orthogonal polynomials and superlinear convergence of Krylov subspace methods in numerical linear algebra

Bernhard Beckermann
Laboratoire Paul Painlevé UMR 8524 (ANO-EDP)
UFR Mathématiques – M3, UST Lille,
F-59655 Villeneuve d'Ascq CEDEX, France
e-mail: bbecker@math.univ-lille1.fr

Summerschool OPSF, July 2004, Madrid
Revised version, Mai 10, 2005

Abstract

We give a theoretical explanation for superlinear convergence behavior observed while solving large symmetric systems of equations using the Conjugate Gradient method, or other Krylov subspace methods. We present a new bound on the relative error after n iterations. This bound is valid in an asymptotic sense, when the size N of the system grows together with the number of iterations. The bound depends on the asymptotic eigenvalue distribution and on the ratio n/N. Similar bounds are given for the task of approaching eigenvalues of large symmetric matrices via Ritz values.

Our findings are related to some recent results concerning asymptotics of discrete orthogonal polynomials due to Rakhmanov and Dragnev & Saff, followed by many other authors. An important tool in these investigations is a constrained energy problem in logarithmic potential theory.

The present notes are intended to be selfcontained (even if sometimes the proofs are incomplete and we refer to the original literature for details): the first part about Krylov subspace methods should be accessible for people from the orthogonal polynomial community, also for those who do not know much about numerical linear algebra. In the second part we gather the necessary tools from logarithmic potential theory, and recall the basic results on the nth root asymptotics of discrete orthogonal polynomials. Finally, in the third part we discuss the fruitful relationship between these two fields and give several illustrating examples.

Contents

Chapter 1

Background in Numerical Linear Algebra

1.1 Introduction

The Conjugate Gradient (CG) method is widely used for solving systems of linear equations $Ax = b$ with a positive definite symmetric matrix A. The CG method is popular as an iterative method for large systems, stemming e.g. from the discretisation of boundary value problems for elliptic PDEs. The rate of convergence of CG depends on the distribution of the eigenvalues of A. A well-known upper bound for the error e_n in the A-norm after n steps is

$$\frac{\|e_n\|_A}{\|e_0\|_A} \leq 2\left(\frac{\sqrt{\kappa}-1}{\sqrt{\kappa}+1}\right)^n \tag{1.1.1}$$

where e_0 is the initial error and the condition number $\kappa = \lambda_{\max}/\lambda_{\min}$ is the ratio of the two extreme eigenvalues of A. In practical situation, this bound is too pessimistic, and one observes an increase in the convergence rate as n increases. This phenomenon is known as superlinear convergence of the CG method. It is the purpose of this work to give an explanation for this behavior in an asymptotic sense, following [Kui00a, BeKu99, BeKu00, BeKu02].

As we will see in Section 1.4 below, the CG convergence behavior is determined by asymptotics of discrete orthogonal polynomials, and can be bounded above in terms of asymptotics of discrete L_∞ extremal polynomials. More generally, consider the extremal polynomials $T_{n,p}(z) = z^n + \textit{lower powers}$ with regard to some discrete L_p–norm

$$||w_n \cdot T_{n,p}||_{L_p(E_n)} = \min\{||w_n \cdot P||_{L_p(E_n)} : P(z) = z^n + \textit{lower powers}\}, \tag{1.1.2}$$

where

$$||f||_{L_\infty(E_n)} := \sup_{z\in E_n} |f(z)|, \quad ||f||_{L_p(E_n)} := \left[\sum_{z\in E_n} |f(z)|^p\right]^{1/p}, \tag{1.1.3}$$

$0 < p < \infty$, with E_n being suitable finite or countable subsets of the complex plane, $\#E_n \geq n+1$, and $w_n(z)$, $z \in E_n$, being (sufficiently fast decreasing) positive numbers.

For the case $p = 2$ of monic discrete orthogonal polynomials, examples include the discrete Chebyshev polynomials [Rak96] (choose $w_n = 1$, $E_n = \{0, 1, ..., n\}$) or other classical

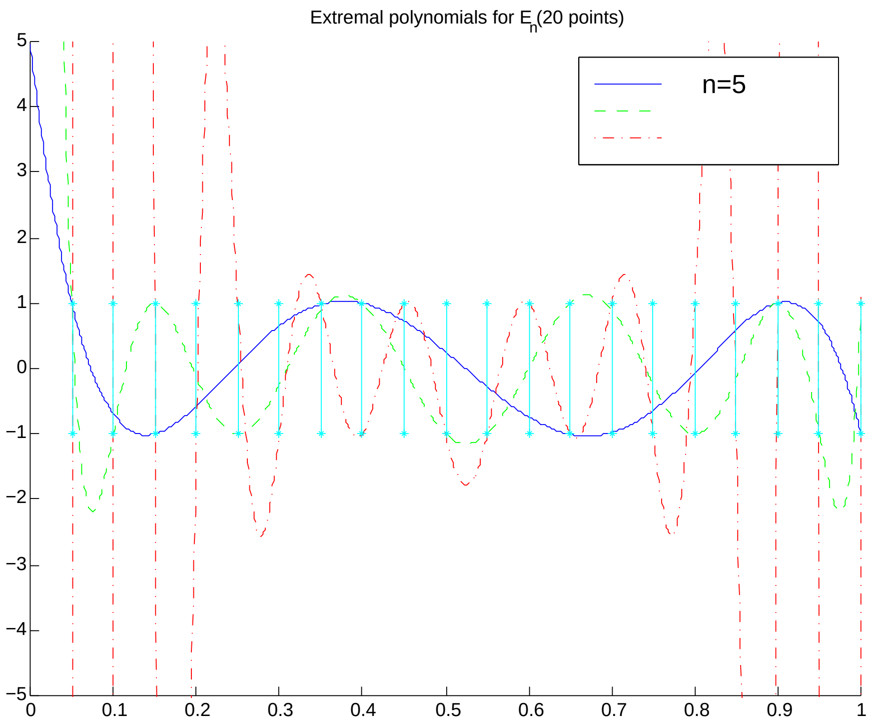

Figure 1.1: *The polynomials* $T_{n,\infty}$ *(after normalization) for* $n = 5, 10, 18$ *for* E *consisting of* 20 *equidistant points and trivial weight* $w = 1$.

families like Krawtchouk or Meixner polynomials [DaSa98, DrSa97, KuVA99], see for instance the review in [KuRa98]. A study of asymptotics of such polynomials has some important applications, e.g., in coding theory, in random matrix theory [Joh00], or in the study of the continuum limit of the Toda lattice [DeMc98].

It was Rakhmanov [Rak96] who first observed that a particular constrained (weighted) energy problem in complex potential theory (see Section 2.2) may furnish a method for calculating the nth root asymptotics of extremal polynomials with respect to so–called *ray sequences* obtained by a suitable renormalization of the sets E_n. Further progress has been made by Dragnev and Saff for real sets E_n being uniformly bounded [DrSa97]; they also obtained asymptotics for discrete L_p–norms with $0 < p \leq \infty$. Generalizations for unbounded real sets E_n and exponentially decreasing weights have been discussed by Kuijlaars and Van Assche [KuVA99] ($0 < p \leq \infty$) and Kuijlaars and Rakhmanov [KuRa98] ($p = 2$). Damelin and Saff [DaSa98] studied the case $p = \infty$ for more general classes of weights. Complex possibly unbounded sets E_n and even more general weights have been discussed in [Be00a]. Here it is also shown that two conjectures of Rakhmanov [KuRa98] are true concerning some separation assumption for the sets E_n.

We will explain in Section 2.3 below how some energy problem with constraint and external field will enable us to describe the nth root asymptotics of the polynomials $T_{n,p}$ and the norms $||w_N \cdot T_{n,p}||_{L_p(E_N)}$. What makes the asymptotic analysis difficult is the fact that a polynomial can be small on a discrete set without being uniformly small in the convex hull of this discrete set. To illustrate this observation, we have chosen $E = \{j/20 : j = 1, ..., 20\}$ and the trivial weight $w = 1$, and have drawn the normalized extremal polynomials $T_{n,\infty}/||T_{n,\infty}||_{L_\infty(E)}$ for $n = 5, 10, 18$ in Figure 1.1. We see that, for

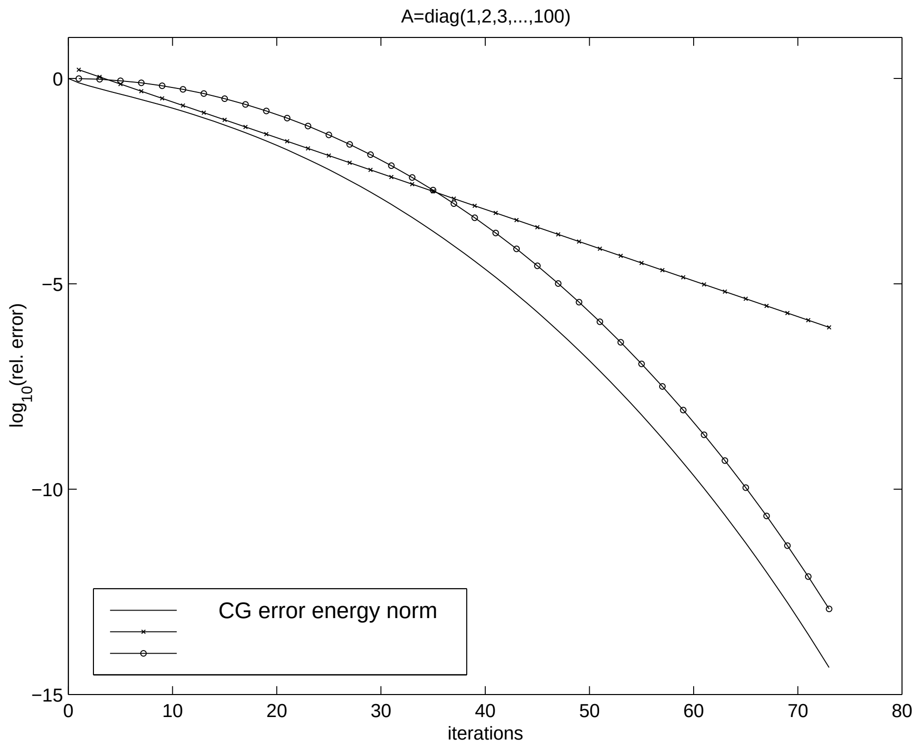

Figure 1.2: *The CG error curve versus the two upper bounds for the system* $Ax = b$ *with* $A = \text{diag}\,(1, 2, \ldots, 100)$*, random solution* x*, and initial residual* $r_0 = (1, ..., 1)^T$*.*

$n = 4$, the polynomial is uniformly small on $[1/20, 20/20]$, but this is no longer true for $n = 10$ or $n = 18$.

For the same reason, the classical CG error bound (1.1.1) gives satisfactory results for small iterations, but can be a crude overestimation in a later stage. Indeed, for small n, a polynomial $p \in P_n$ with $p(0) = 1$ that is small on the spectrum of A has to be uniformly small on the full interval $[\lambda_{\min}, \lambda_{\max}]$ as well. When n gets larger, however, a better strategy for p is to have some of its zeros very close to some of the eigenvalues of A, thereby annihilating the value of p at those eigenvalues, while being uniformly small on a subcontinuum of $[\lambda_{\min}, \lambda_{\max}]$ only.

As an illustration we look at the case of a matrix A with 100 equally spaced eigenvalues $1, 2, \ldots, 100$. The error curve computed for this example is the solid line in Figure 1.2. See also [DTT98, page 560]. The classical error bound given by (1.1.1) with $\kappa = 100$ is the straight line in Figure 1.1. For smaller values of n, the classical error bound gives an excellent approximation to the actual error. The other curve (the one with the dots) is the asymptotic bound for the error proposed in [BeKu99, Corollary 3.2]. This curve follows the actual error especially well for $n \geq 40$, the region of superlinear convergence.

The observations made above have been well known in the numerical linear algebra community, see for instance the monographs [Fi96, Gr97, Nev93, Saa96, TrBa97] or the original articles [AxLi86a, AxLi86b, Gre79, SlvS96, vSvV86]. Eigenvalues far away from the rest of the spectrum (so-called *outliers*) have been treated an a separate manner improving (1.1.1) [AxLi86a, AxLi86b]. The strategy described above to get a polynomial being small on the (discrete) spectrum was known as convergence of some Ritz values

[Gre79, SlvS96, vSvV86]. In addition, the researchers have been aware of the fact that logarithmic potential theory helps in describing or bounding the rate of convergence [DTT98]. There was also a vague idea about what is a "favorable eigenvalue distribution" in order to get a pronounced superlinear convergence [DTT98, TrBa97]. However, precise analytic formulas seemed to occur for the first time only in [Kui00a, BeKu99, BeKu00, BeKu02].

Properly speaking, the concept of superlinear convergence for the CG method applied to a single linear system does not make sense. Indeed, in the absence of roundoff errors, the iteration will terminate latest after N steps if N is the size of the system. Also the notion that the eigenvalues are distributed according to some continuous distribution is problematic when considering a single matrix.

Therefore we are not going to consider a single matrix A, but instead a sequence $(A_N)_N$ of positive definite symmetric matrices. The matrix A_N has size $N \times N$, and we are interested in asymptotics for large N. These matrices need to have an *asymptotic eigenvalue distribution*.

The rest of this manuscript is organized as follows: In §1.2 we present several Krylov subspace methods and fix notations. Subsequently, we introduce polynomial language for explaining the link between convergence theory for Krylov subspace methods, and classical extremal problems in the theory of orthogonal polynomials. We shortly describe the general case in §1.3, and then analyze in more detail the case of hermitian matrices in §1.4.

Following [Kui00a, BeKu99, BeKu00, BeKu02, Be00b], we describe in §2 and §3 how logarithmic potential theory may help to analyze the convergence of Krylov subspace methods. Some facts about the weighted energy problem are recalled in §2.1, but here some additional reading would be helpful, see for instance [MaFi04]. The constrained weighted energy problem is discussed in some more details in §2.2, and used in §2.3 in order to describe nth root asymptotics of discrete L_p extremal polynomials.

The link to the convergence of Krylov subspace methods for hermitian matrices is presented in §3.1 and §3.2, where also several illustrating numerical examples are given. The aim of §3.3 and §3.4 is to show that many classes of structured matrices have an asymptotic eigenvalue distribution. We will consider in particular matrices coming from the discretization of (elliptic) partial differential equations in $\mathbb{R}^2$. A generalization to higher dimension is possible, but for the sake of simplicity we omit details.

1.2 Conjugate gradients, Lanczos, and Ritz values

For solving $Ax = b$ with A being a sparse large matrix of size $N \times N$, one often makes use of Krylov subspace methods which only require to compute matrix vector products with A, the latter can be often implemented in a very efficient manner. In this section we do not attempt to give a complete account of Krylov subspace methods, the interested reader should consult the monographs [Fi96, Gr97, Nev93, Saa96, TrBa97]. We just recall the basic definitions and some elementary properties on the rate of convergence. Here, very much in the spirit of the Lille reseach group and in particular of Claude Brezinski (see also [Fi96]), we will use polynomial language, which should make the theory also more accessible for people coming from orthogonal polynomials.

In what follows we will always suppose exact arithmetic and ignore errors due to floating point operations. In particular, we will find that several Krylov subspace methods are mathematically equivalent for symmetric A. However, their implementation differs quite a lot, and thus the results may change in a floating point environment. The link between convergence of Krylov space methods and loss of precision is subjet of actual research, see for instance the recent work of Strakos or Meurant, e.g., [Meu92, StTi02, Str01].

A Krylov subspace method consists of computing a sequence $x_0, x_1, x_2, ...$ of approximate solutions of $Ax = b$, with residual

$$r_n = r(x_n) = b - Ax_n.$$

The philosophy behind these methods is that (x_n) "converges quickly" to the solution $A^{-1}b$, i.e., x_n is a "good" approximation already for $n \ll N$. The iterates satisfy

$$x_n \in x_0 + \mathcal{K}_n(A, r_0)$$

with the Krylov space

$$\mathcal{K}_n(A, c) = \text{span}\{c, Ac, A^2c, ..., A^{n-1}c\} = \{p(A)c : p \text{ a polynomial of degree} \leq n-1\}.$$

Notice that $r_n \in r_0 + A\mathcal{K}_n(A, r_0)$, and thus

$$r_n = \frac{q_n(A)r_0}{q_n(0)} \tag{1.2.1}$$

for a certain polynomial q_n of degree n. The Krylov subspace method in question is now defined by imposing on the residual either some minimization property (MinRES, GMRES, CG) or some orthogonality property (projection methods, FOM, Lanczos, CR).

Definition 1.2.1 The nth iterate x_n^{GMRES} of GMRES is the unique argument realizing

$$\min\{||r(x)|| : x \in x_0 + \mathcal{K}_n(A, r_0)\}.$$

The nth iterate x_n^{FOM} of FOM is defined by

$$r_n^{FOM} \perp \mathcal{K}_n(A, r_0).$$

For some vector y, the nth iterate x_n^L of the Lanczos method is defined by

$$r_n^L \perp \mathcal{K}_n(A^*, y)$$

(in case of $y = r_0$ we speak of the symmetric Lanczos method).

For real symmetric positive definite A, the nth iterate x_n^{CG} of the method of conjugate gradients (CG) is the unique argument realizing

$$\min\{||r(x)||_{A^{-1}} : x \in x_0 + \mathcal{K}_n(A, r_0)\}, \qquad ||c||_{A^{-1}} = \sqrt{c^* A^{-1} c}.$$

□

If A is invertible, and s.p.d., respectively, the functions $x \mapsto ||r(x)||^2$, and $x \mapsto ||r(x)||^2_{A^{-1}}$, respectively, are strictly convex, and thus the iterates of CG and GMRES exist and are unique. In contrast, it may happen that the nth iterate of FOM does not exist, see Corollary 1.3.4.

Exercise 1.2.2 *There exists $N' = N'(B, r_0)$ such that, for all $n \geq 0$,*

$$dim\ \mathcal{K}_n(A, r_0) = \min\{n, N'\}.$$

Hint: Try first diagonal A, and use the fact the matrix $(r_0, Ar_0, A^2r_0, ..., A^{n-1}r_0)$ is some diagonal matrix times some Vandermonde matrix. Try then diagonalisable A. □

Exercise 1.2.3 *If $x_0 = 0$, and A is invertible, show that $A^{-1}b \in \mathcal{K}_{N'}(A, r_0)$.* □

It follows from the preceding two exercises that, for $x_0 = 0$ and $n = N'$, the iterates x_n^{FOM}, x_n^{GMRES} and x_n^{CG} give the exact solution of $Ax = b$, but of course we hope that we have a good approximation already much earlier.

There is a link between the size of the residuals of FOM and GMRES given by the following result. A proof is immediate once we have the representation (1.2.1) in terms of orthogonal polynomials, see Section 1.3.

Exercise 1.2.4 *Show that*

$$\frac{1}{||r_n^{GMRES}||^2} = \sum_{j=0}^{n} \frac{1}{||r_j^{FOM}||^2}.$$

□

Definition 1.2.5 The Arnoldi basis $v_1, v_2, ..., v_{N'}$ is such that, for all $n = 1, ..., N'$, the vectors $v_1, ..., v_n$ form an orthonormal basis of $\mathcal{K}_n(A, r_0)$ (obtained by the Arnoldi method: orthogonalize Av_n against $v_1, ..., v_n$, and divide the resulting vector by its norm). We also define the matrices

$$\begin{aligned} &V_n := (v_1, v_2, ..., v_n) \in \mathbb{C}^{N \times n}, \\ &J_n := V_n^* A V_n \in \mathbb{C}^{n \times n}, \qquad \widehat{J}_n = V_{n+1}^* A V_n \in \mathbb{C}^{(n+1) \times n}. \end{aligned}$$

Finally, the eigenvalues of the projected matrix J_n are called nth Ritz values of A. □

Exercise 1.2.6 *Show that J_n and $\hat{J}_n$ are upper Hessenberg (all elements at position (j, k) with $k < j - 1$ are equal to zero). Furthermore, show that*

$$n < N' : \qquad V_{n+1}\widehat{J}_n = AV_n. \tag{1.2.2}$$

Finally, in the case of hermitian A, show that J_n is symmetric (and tridiagonal). □

Remark 1.2.7 By construction we have for $n = N'$ that $V_{N'}J_{N'} = AV_{N'}$. As a consequence, denoting by $\Lambda(B)$ the spectrum of some matrix B, we have that the columns of $V_{N'}$ span an A-invariant subspace, and $\Lambda(J_{N'}) \subset \Lambda(A)$. □

Remark 1.2.8 In case of real symmetric A, it follows from Definition 1.2.1 that the symmetric Lanczos method and FOM are mathematically equivalent, i.e., $x_n^{FOM} = x_n^L = x_n^{CR}$, the last denoting the iterates of the conjugate residual method. Also, in this case the GMRES method reduces to the so-called method MinRES (minimal residuals). Finally, we will show in Corollary 1.4.2 below that, in case of symmetric positive definite A, the CG iterates coincide with the symmetric Lanczos iterates. □

1.3 Krylov subspace methods and discrete orthogonal polynomials: non symmetric data

One may show that the nth residual polynomial q_n of (1.2.1) of the Lanczos method is given by the denominator of the nth Padé approximant at infinity of the rational function

$$\pi_N(z) = y^*(zI - A)^{-1} r_0 = \sum_{j=0}^{\infty} z^{-j-1} y^* A^j r_0, \tag{1.3.1}$$

see for instance [Bre72, §3.6] or for the symmetric case [GoSt94]. Hence there is a link between Lanczos method and formal orthogonal polynomials (polynomials being orthogonal with respect to some linear form).

In this Section we will concentrate on FOM/GMRES for non symmetric A. Denote by $\mathcal{P}$ the set of polynomials with complex coefficients, and by $\mathcal{P}_n$ the set of polynomials of degree at most n with complex coefficients. For two polynomials P, Q, we consider the sesquilinear form

$$\prec P, Q \succ = (P(A) r_0)^* Q(A) r_0.$$

The following exercise shows that we have a scalar product

Exercise 1.3.1 *Let $N' = N'(A, r_0)$ as in Exercise 1.2.2. Show that for all $P \in \mathcal{P}_{N'-1} \setminus \{0\}$ we have $\prec P, P \succ > 0$, and that there exists a unique monic polynomial Q of degree N' with $\prec Q, Q \succ = 0$.* □

As a consequence, we can define uniquely orthonormal polynomials p_n, $n = 0, 1, ..., N'$, verifying

$$j = 0, ..., N'-1: \quad p_j(z) = k_j z^j + \text{ lower powers}, \qquad k_j > 0, \tag{1.3.2}$$

$$j, k = 0, ..., N'-1: \quad \prec p_j, p_k \succ = \delta_{j,k}, \tag{1.3.3}$$

$$\text{for all } P \in \mathcal{P}: \quad \prec P, p_{N'} \succ = 0, \qquad p_{N'}(z) = z^{N'} + \text{ lower powers} \tag{1.3.4}$$

(we put $k_{N'} = 1$). These orthonormal polynomials are known to satisfy a (full) recurrence: there exists an upper Hessenberg matrix $J_{N'}$ such that

$$z(p_0, p_1, ..., p_{N'-1})(z) = (p_0, p_1, ..., p_{N'-1})(z) J_{N'} + \frac{k_{N'-1}}{k_{N'}} p_{N'}(z)(0, ..., 0, 1). \tag{1.3.5}$$

The Hessenberg matrix J_n occurred already earlier in Definition 1.2.5. Indeed, this is not an inconsistency in notation, as it becomes clear from the following

Exercise 1.3.2 *The Arnoldi basis is given by the vectors*

$$v_j := p_{j-1}(A)r_0, \quad j = 1, ..., N',$$

and $p_{N'}(A)r_0 = 0$. *In particular, the matrices* $J_{N'}$ *in Definition 1.2.5 and in (1.3.5) coincident, and the matrix* J_n *in Definition 1.2.5 is just the nth principal submatrix of* $J_{N'}$. *Finally, for* $1 \leq n \leq N'$ *we have*

$$z(p_0, p_1, ..., p_{n-1})(z) = (p_0, p_1, ..., p_{n-1})(z)J_n + \frac{k_{n-1}}{k_n} p_n(z)(0, ..., 0, 1). \qquad (1.3.6)$$

□

As a (more or less immediate) consequence of Exercise 1.3.2, we have the following two interpretations in terms of orthogonal polynomials. A proof is left to the reader.

Corollary 1.3.3 *The nth Ritz values of A are given by the zeros of the orthonormal polynomial* p_n. □

As we will see below, $\prec \cdot, \cdot \succ$ can be a discrete Sobolev inner product, its support being a subset of the spectrum of A. Approaching the spectrum of A by Ritz values means that we approach the support of some scalar product by the zeros of the underlying orthogonal polynomials, something familiar for people from the OP community (at least if the support is real, see Section 1.4).

Corollary 1.3.4 *The nth iterate of FOM exists if and only if* $p_n(0) \neq 0$. *In this case*

$$r_n^{FOM} = \frac{p_n(A)r_0}{p_n(0)}, \qquad \frac{1}{||r_n^{FOM}||} = |p_n(0)|.$$

□

We may also describe the residuals of GMRES in terms of orthonormal polynomials. For this we need some preliminary remarks.

Definition 1.3.5 The nth Szegő kernel of the scalar product $\prec \cdot, \cdot \succ$ is defined by

$$K_{n,2}(x, y) = \sum_{j=0}^{n} \overline{p_j(x)} p_j(y).$$

□

It is a well-known fact (see for instance the "bible" of Szegő) that

$$\min_{P \in \mathcal{P}_n} \frac{\prec P, P \succ}{|P(0)|^2} = \frac{1}{K_{n,2}(0,0)}, \quad \text{attained for the polynomial } P(z) = K_{n,2}(0, z). \qquad (1.3.7)$$

By construction of the scalar product $\prec \cdot, \cdot \succ$ and by (1.2.1) we have for $n < N'$

$$\min_{P \in \mathcal{P}_n} \frac{\prec P, P \succ}{|P(0)|^2} = \min\{||r(x)||^2 : x \in x_0 + \mathcal{K}_n(A, r_0)\},$$

leading to the following characterization

Corollary 1.3.6 *For the nth iterate of GMRES, $n = 0, 1, ..., N' - 1$ we have*

$$r_n^{GMRES} = \frac{K_{n,2}(0, A)r_0}{K_{n,2}(0,0)}, \qquad \frac{1}{||r_n^{GMRES}||^2} = K_{n,2}(0,0).$$

□

Exercise 1.3.7 *Let $\mu_{j,k} := (A^j r_0)^* A^k r_0$. Show that*

$$\frac{||r_1^{GMRES}||^2}{||r_0^{GMRES}||^2} = \frac{\mu_{0,0}\mu_{1,1} - |\mu_{0,1}|^2}{\mu_{0,0}\mu_{1,1}}.$$

Conclude that the following algorithm (called GMRES(1))

> *choose any y_0*
> *for $k = 0, 1, ...$ until "convergence" do*
> > *compute y_{k+1} by one iteration of GMRES with starting vector $x_0 = y_k$*

converges ($y_k \to A^{-1}b$ for $k \to \infty$) for all b and y_0 if and only if for all $y \neq 0$ we have $y^* A y \neq 0$.
Hint: if you do not find a direct proof, look at [Gr97]. □

We terminate this section by showing that the scalar product can be (but does not need to be) a Sobolev inner product with finite support, possibly in the complex plane. Finally, we show that in case of a normal matrix A we have an important simplification, leading to discrete orthogonal polynomials with possibly complex support.

Example 1.3.8 For some parameter $\rho > 0$, consider the matrix

$$A = XBX^{-1}, \qquad B = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 1 & 1 & 0 & 0 \\ 0 & 0 & -1 & 0 \\ 0 & 0 & 1 & -1 \end{bmatrix}, \qquad X = \begin{bmatrix} 1 & 0 & 0 & 0 \\ \rho & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ \rho & 0 & 0 & 1 \end{bmatrix},$$

and $r_0 = Xc$, $c = (1, 0, 1, 0)^t$. Notice that B is in Jordan form, and hence $\Lambda(A) = \{-1, 1\}$. It is not difficult to show that, for any polynomial Q,

$$Q(A)r_0 = XQ(B)c = X \begin{bmatrix} Q(1) \\ Q'(1) \\ Q(-1) \\ Q'(-1) \end{bmatrix} = \begin{bmatrix} Q(1) \\ \rho Q(1) + Q'(1) \\ Q(-1) \\ \rho Q(1) + Q'(-1) \end{bmatrix}.$$

Thus we obtain the following simplification for the scalar product

$$\begin{aligned} \prec P, Q \succ \;&=\; \overline{P(1)}Q(1) + \overline{\rho P(1) + P'(1)}(\rho Q(1) + Q'(1)) \\ &\quad + \overline{P(-1)}Q(-1) + \overline{\rho P(1) + P'(-1)}(\rho Q(1) + Q'(-1)), \end{aligned}$$

which reduces to a (discrete) Sobolev inner product in the case $\rho = 0$. □

One may indeed show that the scalar product always can be represented in terms of a linear combination of the values and the derivatives of P, Q at the points of the spectrum of $J_{N'}$. Let us have a closer look at a case where the derivatives do not occur.

Theorem 1.3.9 *If A is normal (i.e., $A^*A = AA^*$), with matrix of right eigenvectors given by X_N and its spectrum by $\Lambda(A) = \{\lambda_1, ...\lambda_N\}$ then with $X_N^{-1} r_0 = (\beta_j)_j$,*

$$P, Q \in \mathcal{P}: \qquad \prec P, Q \succ = \sum_{j=1}^{N} |\beta_j|^2 \overline{P(\lambda_j)} Q(\lambda_j). \tag{1.3.8}$$

Conversely, if the scalar product has a such representation then at least the projected matrix $J_{N'}$ is normal, and $\Lambda(J_{N'}) = \{\lambda_j : \beta_j \neq 0\} \subset \Lambda(A)$.

Proof. If A is normal then its matrix of eigenvectors X_N is unitary, that is,

$$X_N^* X_N = I, \qquad X_N^* A X_N = \Lambda = \text{diag } (\lambda_1, ..., \lambda_N).$$

Observing that for some polynomial P we have

$$P(A) = X_N P(\Lambda) X_n^*, \qquad \text{with} \;\; P(\Lambda) = \text{diag } (P(\lambda_1), ..., P(\lambda_N)),$$

we obtain by writing $\beta = X_n^* r_0 = X_n^{-1} r_0$

$$P, Q \in \mathcal{P}_n: \qquad \prec P, Q \succ = \beta^* P(\Lambda)^* Q(\Lambda) \beta = \sum_{j=1}^{N} |\beta_j|^2 \overline{P(\lambda_j)} Q(\lambda_j).$$

In order to show the converse result, suppose that there are N^* distinct λ_j with $\beta_j \neq 0$, say, the terms corresponding to $\lambda_1, ..., \lambda_{N^*}$. Then we may rewrite the sum in (1.3.8) as

$$P, Q \in \mathcal{P}: \qquad \prec P, Q \succ = \sum_{j=1}^{N^*} |\beta_j^*|^2 \overline{P(\lambda_j)} Q(\lambda_j)$$

for suitable coefficients β_j^* and distinct $\lambda_1, ..., \lambda_{N^*}$. From Exercise 1.3.1 together with (1.3.4) we learn that $N^* = N'$, and $p_{N'}(z) = (z - \lambda_1)...(z - \lambda_{N'})$. It follows from Corollary 1.3.3 that $J_{N'}$ has the distinct eigenvalues $\lambda_1, ..., \lambda_{N'}$, and the relation $\Lambda(J_{N'}) \subset \Lambda(A)$ was established in Remark 1.2.7. Consider the matrix

$$Y := (p_k(\lambda_j))_{j=1,...,N',k=0,...,N'-1},$$

then from (1.3.5) we know that $Y J_{N'} = DY$, $D = \text{diag } (\lambda_1, ..., \lambda_{N'})$, and from the representation of the scalar product together with (1.3.3) we learn that

$$Y^* \,\text{diag }(|\beta_1^*|^2, ..., |\beta_{N'}^*|^2)\, Y = I.$$

Hence, up to a certain normalization of the rows, the left eigenvector matrix of $J_{N'}$ is unitary, implying that $J_{N'}$ is normal. $\square$

Analyzing in more detail the preceding proof we obtain the following

Corollary 1.3.10 *If A is normal then also $J_{N'}$ is normal, with $\Lambda(J_{N'}) \subset \Lambda(A)$. Furthermore, $J_{N'}$ has N' distinct eigenvalues, and*

$$P, Q \in \mathcal{P}: \qquad \prec P, Q \succ = \sum_{\lambda \in \Lambda(J_{N'})} w(\lambda)^2 \overline{P(\lambda)} Q(\lambda), \tag{1.3.9}$$

$$w(\lambda)^2 = \frac{1}{K_{N',2}(\lambda, \lambda)} > 0, \qquad \sum_{\lambda \in \Lambda(J_{N'})} w(\lambda)^2 = ||r_0||^2.$$

1.4 Krylov subspace methods and discrete orthogonal polynomials: symmetric data

From now on we will always suppose (up to Remark 1.4.9) that our matrix of coefficients A is hermitian, i.e., $A = A^*$, with spectrum $\lambda_1 \leq \lambda_2 \leq ...\lambda_N$. From Exercise 1.2.6 it follows that $J_{N'}$ with $N' = N'(A, r_0)$ as in Exercise 1.2.2 is an hermitian upper Hessenberg matrix with positive entries $k_n/k_{n+1} > 0$ on the first subdiagonal. However, such a matrix is necessarily tridiagonal, of the form

$$J_n = \begin{bmatrix} b_0 & a_0 & 0 & \cdots & \cdots & 0 \\ a_0 & b_1 & a_1 & 0 & \cdots & 0 \\ 0 & a_1 & b_2 & a_2 & \ddots & \vdots \\ \vdots & \ddots & \ddots & \ddots & \ddots & 0 \\ 0 & \cdots & 0 & a_{n-3} & b_{n-2} & a_{n-2} \\ 0 & \cdots & \cdots & 0 & a_{n-2} & b_{n-1} \end{bmatrix}, \qquad b_n \in \mathbb{R},\ a_n = \frac{k_n}{k_{n+1}} > 0. \tag{1.4.1}$$

Thus (1.3.5) becomes a three term recurrence: for $n = 0, ..., N' - 1$

$$zp_n = a_n p_{n+1} + b_n p_n + a_{n-1} p_{n-1}, \qquad p_0(z) = \frac{1}{||r_0||}, \qquad p_{-1}(z) = 0. \tag{1.4.2}$$

In particular it follows that the orthonormal polynomials p_n have real coefficients, and are orthonormal with respect to the linear functional c acting on the space of polynomials via

$$c(P) = \prec 1, P \succ = r_0^* P(A) r_0, \tag{1.4.3}$$

i.e., $c(p_j p_k) = \delta_{j,k}$. Also, since hermitian matrices are in particular normal, we have the representation (1.3.9) for the scalar product (and thus for the linear functional), showing that there is classical orthogonality on the real line (Theorem of Favard) with respect to a positive measure with finite support $\Lambda(J_{N'})$ (or a positive linear functional).

In the following Theorem we put together some elementary properties of such orthogonal polynomials. If you do not remember them, please try to prove them (not necessarily in the indicated order) or look them up in any standard book about OP.

Theorem 1.4.1 **(a)** *For $1 \leq n \leq N'$, the zeros of p_n are simple and real, say,*

$$x_{1,n} < x_{2,n} < ... < x_{n,n}.$$

(b) Interlacing property: *The zeros of p_n and p_{n+1} interlace*

$$1 \leq j \leq n < N' : \qquad x_{j,n+1} < x_{j,n} < x_{j,n+1}.$$

(c) Separation property:

$$1 \leq j < n < n' \leq N' : \qquad \text{there exists } j' \text{ such that} \quad x_{j,n} \leq x_{j',n'} < x_{j+1,n}.$$

(d) Christoffel-Darboux: *For $n < N'$ and $x, y \in \mathbb{C}$*

$$K_{n,2}(\overline{x}, y) = \sum_{j=0}^{n} p_j(x)p_j(y) = a_n \frac{p_{n+1}(x)p_n(y) - p_{n+1}(y)p_n(x)}{x-y}$$

and for $x \in \mathbb{R}$

$$K_{n,2}(x, x) = \sum_{j=0}^{n} p_j(x)p_j(x) = a_n(p_{n+1}'(x)p_n(x) - p_{n+1}(x)p_n'(x)) > 0.$$

(e) Associated linear functionals: *For $\gamma \in \mathbb{R}\setminus\Lambda(J_{N'})$, consider the linear functional $\widetilde{c}(P) = c(\widetilde{P})$, $\widetilde{P}(z) = (z-\gamma)P(z)$. Then $K_{n,2}(\gamma, \cdot)$ is an nth orthogonal polynomial with respect to $\widetilde{c}$.*

(f) Gaussian Quadrature: *For any $0 < n < N'$ and any polynomial of degree not exceeding $2n-1$*

$$c(P) = \sum_{j=1}^{n} \frac{P(x_{j,n})}{K_{n-1,2}(x_{j,n}, x_{j,n})}.$$

(g) Stieltjes functions and Padé approximation: *The rational function*

$$\pi_n(z) = ||r_0||^2 \sum_{j=1}^{n} \frac{1}{K_{n-1,2}(x_{j,n}, x_{j,n})} \frac{1}{z - x_{j,n}} = ||r_0||^2 (e_0, (zI_n - J_n)^{-1} e_0)$$

with e_0 being the first canonical vector of suitable size is of denominator degree n, of numerator degree $n-1$, has real simple poles and positive residuals, and

$$\pi_{n+1}(z) - \pi_n(z) = \frac{1}{a_n p_n(z) p_{n+1}(z)} = \mathcal{O}(z^{-2n-1})_{z\to\infty}.$$

In particular, π_n is the nth Padé approximant at infinity of the Stieltjes function $z \mapsto r_0^*(zI - A)^{-1} r_0$, and $\pi_{N'}$ coincides with this function. Finally, p_n is an nth Padé denominator. □

Theorem 1.4.1(g) gives a link between Padé and Krylov subspace methods, compare with the remarks around (1.3.1). More precisely, we have the following link between CG, FOM (symmetric Lanczos) and Padé.

Corollary 1.4.2 *For symmetric positive definite A, the methods CG, FOM and the symmetric Lanczos method are mathematically equivalent, more precisely,*

$$||r_n^{CG}|| = \frac{1}{|p_n(0)|}, \qquad ||r_n^{CG}||_{A^{-1}}^2 = \pi_n(0) - \pi_{N'}(0).$$

Proof. Define the scalar product $\prec \cdot, \cdot \succ^*$ by replacing in (1.3.9) the term $w(\lambda)^2$ by $w^*(\lambda)^2 = w(\lambda)^2/\lambda$. According to Definition 1.2.1 the residual polynomial q_n of CG in (1.2.1) is the (up to scaling) unique solution of the extremal problem

$$\min_{P \in \mathcal{P}_n} \frac{\prec P, P \succ^*}{|P(0)|^2} = \frac{1}{K_{n,2}^*(0,0)} \tag{1.4.4}$$

with the Szegő function $K_{n,2}^*$ of the new scalar product. By (1.3.7) the minimum is attained by $q_n(z) = K_{n,2}^*(0,z)$, the latter being proportional to p_n by Theorem 1.4.1(e). Hence, from (1.2.1) and Corollary 1.3.4 we may conclude that $x_n^{CG} = x_n^{FOM}$, implying in particular our claim for $||r_n^{CG}||$.

In order to show the link between CG and Padé, recall that $(zI - A)V_{N'} = V_{N'}(zI - J_{N'})$, and hence

$$(r_j^{FOM})^*(zI-A)^{-1}r_k^{FOM} = \frac{v_{j+1}^*(zI-A)^{-1}v_{k+1}}{p_j(0)p_k(0)} = \frac{[(zI-J_{N'})^{-1}]_{j+1,k+1}}{p_j(0)p_k(0)}.$$

It is a well-known fact that the inverse of a Jacobi matrix can be expressed in terms of Padé approximants

$$[(zI-J_{N'})^{-1}]_{j+1,k+1} = p_j(z)p_k(z)[\pi_{N'}(z) - \pi_{\max\{j,k\}}(z)],$$

see for instance [Wal73, §60] or the survey paper [Meu92]. Combining these too findings for $z = 0$ leads to our claim. □

The representation of the CG error as Padé error has been applied successfully by Golub, Meurant, Strakos and others [GoMe97, GoSt94, Meu98] to estimate/bound from below the nth CG error after having computed the $(n+p)$th iterate for $p > 1$

$$||r_n^{CG}||_{A^{-1}}^2 \geq \pi_n(0) - \pi_{n+p}(0) = \sum_{j=n}^{n+p-1} |\frac{1}{a_j p_j(0) p_{j+1}(0)}|.$$

In particular, the authors show that these a posteriori bounds are shown to be reliable even in finite precision arithmetic. A similar results has been already mentioned implicitly in the original paper of Hestenes and Stiefel [HeSt94].

We should mention that Theorem 1.4.1(a),(b),(c) provides already a quite precise idea about how Ritz values do approach the real spectrum $\Lambda(J_{N'}) \subset \Lambda(A)$, or, equivalently, how poles of Padé approximants approach the poles of a rational function with positive residuals. However, up to now there is no information about the rate of convergence. For this rate of convergence of Ritz values we have the following result which roughly says that, provided that a certain polynomial extremal problem depending on λ_k gives a "small" value, there is at least one Ritz value "close" to λ_k. We do not claim that there is an eigenvalue "close" to each Ritz value. Indeed there exist examples with $\Lambda(A) = \Lambda(-A)$ where 0 is a Ritz value for each odd n which might be far from $\Lambda(A)$.

Theorem 1.4.3 *[Be00b] If $\lambda_k \leq x_{1,n}$ then*

$$x_{1,n} - \lambda_k = \min\Big\{\frac{\sum_{\lambda\in\Lambda(J_{N'})\setminus\{\lambda_k\}} w(\lambda)^2(\lambda - x_{1,n})|q(\lambda)|^2}{w(\lambda_k)^2|q(\lambda_k)|^2} : \deg q < n, q(\lambda_k) \neq 0\Big\}.$$

Here the minimum is attained for the polynomial $q(x) = p_n(x)/(x - x_{1,n})$.

If $\lambda_k \in [x_{1,n}, x_{n,n}]$, say, $x_{\kappa-1,n} \leq \lambda_k \leq x_{\kappa,n}$, then

$$(\lambda_k - x_{\kappa-1,n})(x_{\kappa,n} - \lambda_k) = \min\Big\{\frac{\sum_{\lambda\in\Lambda(J_{N'})\setminus\{\lambda_k\}} w(\lambda)^2(\lambda - x_{\kappa-1,n})(\lambda - x_{\kappa,n})|q(\lambda)|^2}{w(\lambda_k)^2|q(\lambda_k)|^2} : \deg q < n-1, |q(\lambda_k)| \neq 0\Big\}.$$

Here the minimum is attained for the polynomial $q(x) = p_n(x)/[(x - x_{\kappa-1,n})(x - x_{\kappa,n})]$.

Proof. We will show here the first part of the assertion; similar arguments may be applied to establish the second part. If q is a polynomial of degree less than n with $q(\lambda_{k,N}) \neq 0$ and $p(x) = (x - x_{1,n}) \cdot q(x) \cdot \overline{q(\overline{x})}$, then $c(p) \geq 0$ by the Gaussian quadrature formula of Theorem 1.4.1(f). Hence the right hand side of (1.3.9) is ≥ 0, and thus

$$x_{1,n} - \lambda_k \leq \frac{\sum_{\lambda \in \Lambda(J_{N'}) \setminus \{\lambda_k\}} w(\lambda)^2 (\lambda - x_{1,n}) |q(\lambda)|^2}{w(\lambda_k)^2 |q(\lambda_k)|^2}.$$

Finally, notice that for the choice $q(x) = p_{n,N}(x)/(x - x_{1,n,N})$ we have $c(p) = 0$ again by Theorem 1.4.1(f), and thus there is equality in the above estimate. $\square$

There are several possibilities to relate the result of Theorem 1.4.3 to more classical extremal problems. The term $|\lambda - x_{j,n}|$ could be bounded by $2\,||A||$, and taking into account (1.3.9) and (1.3.7), we obtain for instance

$$(\lambda_k - x_{\kappa-1,n})(x_{\kappa,n} - \lambda_k) \leq 4\,||A||^2 \Big(\frac{K_{N',2}(\lambda_k, \lambda_k)}{K_{n-2,2}(\lambda_k, \lambda_k)} - 1 \Big), \quad \text{if } x_{\kappa-1,n} \leq \lambda_k \leq x_{\kappa,n}.$$

It follows that the distance to at least one of the Ritz values become small if all $|p_j(\lambda_k)|^2$ for $j \geq n-1$ are small compared to $1/||r_0||^2 = K_{0,2}(\lambda_k, \lambda_k)$.

Another possibility could be to relate (1.3.7) to some extremal problem with respect to the maximum norm: for some integer $n \geq 0$, some $z \in \mathbb{C}$ and some compact set $S \subset \mathbb{C}$, consider the quantity

$$E_n(z, S) = \min_{p \in P_n} \max_{\lambda \in S} \frac{|p(\lambda)|}{|p(z)|}. \tag{1.4.5}$$

Clearly, $E_n(z, S)$ is decreasing in n and increasing in S, and $E_0(z, S) = 1$. Also, for any $a, b \neq 0$ there holds $E_n(a, a + bS) = E_n(0, S)$. The motivation for studying this extremal problem comes from the following observation

Exercise 1.4.4 *For the inner product (1.3.9), show the following link between E_n and the Szegő function*

$$z \in \mathbb{C}: \qquad \frac{1}{K_{n,2}(z, z)} \leq ||r_0||^2\, E_n(z, \Lambda(J_{N'}))^2.$$

If in addition $\Lambda(A) \subset (0, +\infty)$, show that

$$\frac{1}{|p_n(0)|} \leq ||r_0||_{A^{-1}}\, E_n(0, \Lambda(J_{N'})).$$

$\square$

Also, by dropping the negative terms in the sums occurring in Theorem 1.4.3 we obtain the following upper bound for the rate of convergence of Ritz values.

Corollary 1.4.5 *We have the following upper bounds: if $\lambda_k \leq x_{1,n}$ then*

$$x_{1,n} - \lambda_k \leq 2\,||A||\,\frac{||r_0||^2}{w(\lambda_k)^2} E_{n-1}(\lambda_k, \Lambda(J_{N'}) \setminus (-\infty, x_{1,n}])^2,$$

If $x_{\kappa-1,n} \leq \lambda_k \leq x_{\kappa,n}$ then

$$(\lambda_k - x_{\kappa-1,n})(x_{\kappa,n} - \lambda_k) \leq 4\,||A||^2\,\frac{||r_0||^2}{w(\lambda_k)^2} E_{n-2}(\lambda_k, \Lambda(J_{N'}) \setminus [x_{\kappa-1,n}, x_{\kappa,n}])^2$$

□

The preceding Theorem gives the idea that eigenvalues λ_k sufficiently away from the rest of the spectrum of $\Lambda(A)$ (so-called outliers) and having a sufficiently large eigencomponent $w(\lambda_k)$ should be well approximated by Ritz values. However, as we will see later, for convergence it will be sufficient that there are not "too many" eigenvalues close to λ_k.

Let us further discuss the extremal problem (1.4.5). In the case $S \subset \mathbb{R}$ and real z not lying in the convex hull of S, it is known that the polynomial $T_{n,\infty}$ of (1.1.2) (put $w_n = 1$) is extremal for (1.4.5), see for instance [Fi96]. The latter is uniquely characterized by a so-called alternant, that is, the extremal polynomial attains its maximum on S at least $n + 1$ times, with alternating sign. For the sake of completeness, let us discuss in more detail the case of an interval.

Lemma 1.4.6 *If $0 < a < b$, the value $E_n(0, [a, b])$ is attained (up to a linear transformation) for the Chebyshev polynomial of the first kind $T_n(\cos(\phi)) = \cos(n\phi)$, and*

$$E_n(0, [a, b]) = \frac{2}{y^n + y^{-n}} \leq 2y^n, \qquad y = \frac{\sqrt{b/a} - 1}{\sqrt{b/a} + 1} < 1.$$

Proof. We first notice that $E_n(0, [a, b]) = E_n(z, [-1, 1])$ with $z = (b + a)/(b - a) > 1$. If T_n would not be extremal, then there is a polynomial P of degree n with

$$P(z) = T_n(z), \qquad \text{and} \;\; M := ||P||_{L_\infty([-1,1]} < ||T_n||_{L_\infty([-1,1]} = 1.$$

If follows that the polynomial $R := T_n - P$ satisfies $R(z) = 0$, and for $j = 0, 1, ..., n$

$$\begin{aligned}
&(-1)^j\, R(\cos(\frac{\pi j}{n})) = (-1)^j\, T_n(\frac{\pi j}{n})) - (-1)^j\, P(\cos(\frac{\pi j}{n})) \\
&= 1 - (-1)^j\, P(\cos(\frac{\pi j}{n})) \geq 1 - M > 0.
\end{aligned}$$

Hence R must have $n + 1$ roots, but is a polynomial of degree n, a contradiction. Thus T_n is indeed extremal. It remains to compute its value at z, where we use the recurrence

$$T_{n+1}(z) = (y + y^{-1})T_n(z) - T_{n-1}(z), \quad T_0 = 1, \quad T_1(z) = z = \frac{y + 1/y}{2}.$$

□

In [Fi96] one also finds a discussion of the case of S being a union of two intervals. Here the solution may be estimated in terms of Weierstass elliptic functions, see also [Akh90]. A simple upper bound is discussed in the following exercise.

Exercise 1.4.7 *Using the preceding result, show that, for $0 < a < b$,*

$$E_{2n}(0, [-b, -a] \cup [a, b]) \leq 2\Big(\frac{b/a - 1}{b/a + 1}\Big)^n.$$

Derive from this relation an explicit bound for $(\lambda_k - x_{\kappa-1,n})(x_{\kappa,n} - \lambda_k)$ in terms of the distance of λ_k to the rest of the spectrum of A. □

A combination of Exercise 1.4.4 with Remark 1.2.8, Corollary 1.3.6 and Corollary 1.4.2 (and its proof) leads to the following estimates.

Corollary 1.4.8 *For hermitian A,*

$$\frac{||r_n^{GMRES}||}{||r_0^{GMRES}||} = \frac{||r_n^{MinRES}||}{||r_0^{MinRES}||} \leq E_n(0, \Lambda(J_{N'})) \leq E_n(0, \Lambda(A)).$$

Moreover, for symmetric positive definite A

$$\frac{||r_n^{CG}||_{A^{-1}}}{||r_0^{CG}||_{A^{-1}}} = \frac{||r_n^{FOM}||_{A^{-1}}}{||r_0^{FOM}||_{A^{-1}}} \leq E_n(0, \Lambda(J_{N'})) \leq E_n(0, \Lambda(A)).$$

The bounds of Corollary 1.4.8 should be considered as worst case bounds since they do not take into account the particular choice of the starting residual. However, it is known [Gre79] that they cannot be sharpened in the following sense: one may give $\tilde{A}, \tilde{r}_0$ with $E_n(0, \Lambda(A)) = E_n(0, \Lambda(\tilde{A}))$ such that there is equality in Corollary 1.4.8 for these new data.

Notice also that we obtain a proof of (1.1.1) by replacing $\Lambda(A)$ by its convex hull in Corollary 1.4.8, and by applying Lemma 1.4.6. However, following this approach we forget completely about the fine structure of the spectrum. It is the aim of the following sections to analyze more pecisely in terms of logarithmic potential theory how the actual distribution of eigenvalues helps us to improve (1.1.1).

Remark 1.4.9 Let us terminate this section by commenting briefly on different approaches of bounding the residual of Krylov subspace methods in case of not necessarily normal A. We start from the observation that

$$\frac{||r_n^{GMRES}||}{||r_0^{GMRES}||} \leq \min\{\frac{||P(A)||}{|P(0)|} : \ P \text{ is a polynomial of degree} \leq n\}. \tag{1.4.6}$$

For diagonalizable A, the right hand side may be bounded above by $E_n(0, \Lambda(A))$ times the condition number of the matrix of eigenvectors of A, see for instance [Saa96]. However, for matrices far from being normal, this condition number might be quite large. There are mainly two attempts to overcome this difficulties (see for instance [Gr97] and the references therein), the first being based on the so-called ϵ-pseudo-spectrum

$$\Lambda_\epsilon(A) := \{z \in \mathbb{C} : ||(zI - A)^{-1}|| \geq \frac{1}{\epsilon}\}$$

in terms of $E_n(0, \Lambda_\epsilon(A))$, the second one on the field of values

$$W(A) = \{\frac{y^* A y}{y^* y} : y \in \mathbb{C}^N, y \neq 0\},$$

which by the Theorem of Haussdorf is a compact and convex set. For a convex set S it can be shown that

$$\exp(-ng_S(0)) \leq E_n(0,S) \leq 3\,\exp(-ng_S(0)) \tag{1.4.7}$$

g_S denoting the Green function of some compact subset of the complex plane, see Section 2.1. Here the left hand inequality is just the Bernstein-Walsh inequality (see for instance [MaFi04]), and the right-hand bound can be shown with help of the corresponding Faber polynomials. A recent and quite deep result of M. Crouzeix [Cro95] (found after the publication of [Gr97]) says that there is a universal constant $C < 34$ such that, for any square matrix A and any polynomial P,

$$||P(A)|| \leq C \max_{z\in W(A)} |P(z)|. \tag{1.4.8}$$

Hence combining (1.4.6), (1.4.7), and (1.4.8), we obtain

$$\frac{||r_n^{GMRES}||}{||r_0^{GMRES}||} \leq 3\,C\,\exp(-ng_{W(A)}(0)),$$

compare with [BGT04]. The constant C can be made smaller by replacing $W(A)$ in (1.4.8) by some larger convex set, see [BGT04, BeCr05]. A different approach via estimating directly $||P(A)||$ for a suitable Faber polynomial P allows even to establish the sharper bound [Be05]

$$\frac{||r_n^{GMRES}||}{||r_0^{GMRES}||} \leq 3\,\exp(-ng_{W(A)}(0)).$$

Notice however that such bounds are typically interesting for small n since they do not allow to describe a superlinear rate of convergence.

Chapter 2

Extremal problems in complex potential theory and nth root asymptotics of OP

2.1 Energy problems with external field

The energy problem with external field has been successfully applied in order to describe asymptotics of orthogonal polynomials on unbounded sets such as Hermite or Freud polynomials. Since this subject has already discussed in detail in [MaFi04], we recall here without proof the basic concepts. Also, for the sake of a simplified presentation, we will restrict ourselves to compact regular real sets and continuous external fields.

Given some compact $\Sigma \subset \mathbb{R}$, we denote by $\mathcal{M}_t(\Sigma)$ the set of Borel measures μ with support $\operatorname{supp}(\mu) \subset \Sigma$ and mass $||\mu|| := \mu(\Sigma) = t$. The logarithmic potential and the energy of a measure $\mu \in \mathcal{M}_t(\Sigma)$ are given by

$$U^\mu(z) = \int \log(\frac{1}{|x-z|})\, d\mu(x), \quad I(\mu) = \iint \log(\frac{1}{|x-y|})\, d\mu(x)\, d\mu(y).$$

Notice that, for a monic polynomial P of degree n, the expression $-\log(|P|^{1/n})$ coincides with the logarithmic potential of some discrete probability measure, which in case of distinct roots hass mass $1/n$ at each root of P. On the other hand, such discrete measures are dense in the set of Borel measures, explaining why the tool of logarithmic potential theory is suitable for studying nth root asymptotics.

We define more generally for $\mu, \nu \in \mathcal{M}_x(\Sigma)$ the mutual energy by the expression

$$I(\mu, \nu) = \iint \log(\frac{1}{|x-y|})\, d\mu(x)\, d\nu(y) \in (0, +\infty].$$

The mutual energy is lower semi-continuous, that is, given two sequences $(\mu_n), (\nu_n) \subset \mathcal{M}_x(\Sigma)$, which converge in weak star topology to μ, and ν, respectively (written by $\mu_n \xrightarrow{*} \mu$), we have

$$\liminf_{n\to\infty} I(\mu_n, \nu_n) \geq I(\mu, \nu). \tag{2.1.1}$$

From this one can deduce as an exercise the *Principle of descent*

$$\liminf_{n\to\infty} U^{\mu_n}(z_n) \geq U^{\mu}(z), \quad \text{if } \mu_n \xrightarrow{*} \mu,\ z_n \to z \text{ for } n \to \infty. \tag{2.1.2}$$

As explained for instance in [MaFi04], the capacity of a compact set E is defined by the minimization problem

$$\mathrm{cap}\,(E) := \exp(-\min\{I(\mu) : \mu \in \mathcal{M}_1(E)\}).$$

If $\mathrm{cap}\,(E) > 0$, or, equivalently, if there is an $\mu \in \mathcal{M}_1(E)$ with finite energy, then on may prove using the strict convexity of $\mu \mapsto I(\mu)$ and Helly's Theorem (weak compactness of $\mathcal{M}_x(E)$) that there is a unique measure ω_E called Robin measure realizing the minimum in the definition of the capacity. The Green function of a compact set E is defined by

$$g_E(z) = \log(\frac{1}{\mathrm{cap}\,(E)}) - U^{\omega_E}(z),$$

behaving at infinity like $\log(|z|) - \log(\mathrm{cap}\,(E)) + o(1)_{z\to\infty}$, being harmonic in $\mathbb{C} \setminus E$, subharmonic and ≥ 0 in $\mathbb{C}$, and equal to zero quasi everywhere on E, i.e., in $E \setminus E_0$, with $\mathrm{cap}\,(E_0) = 0$. Conversely, the Green function can be also uniquely characterized by these properties. For instance, for an interval we have (see, e.g., [MaFi04])

$$g_{[a,b]}(z) = \log(|\frac{2z-b-a}{b-a} + \sqrt{(\frac{2z-b-a}{b-a})^2 - 1}|), \qquad \frac{d\omega_{[a,b]}}{dx}(x) = \frac{1}{\pi\sqrt{(x-a)(b-x)}}. \tag{2.1.3}$$

A compact set E is called regular (with respect to the Dirichlet problem) if g_E is identically zero on E (and hence g_E is continuous in $\mathbb{C}$ by the principle of continuity [SaTo97, Theorem II.3.5]). The link with the so-called Wiener condition and the regularity with respect to the Dirichlet problem is a nice piece of harmonic analysis, we refer the interested reader for instance to [SaTo97, Appendix A], where it is also shown that regularity is a local property

$$E \text{ regular},\ x \in E,\ r > 0 \quad \Longrightarrow \quad E \cap \{y \in \mathbb{C} : |y - x| \leq r\} \text{ is regular.} \tag{2.1.4}$$

Here we only mention a sufficient condition in the following (a little bit difficult) exercise, the interested reader should compare with [NiSo88, § 5.4.3].

Exercise 2.1.1 *Let $E \subset \mathbb{C}$ be compact, and suppose that for each $x \in E$ there exists $C(x), K(x) > 0$ such that, for all $r > 0$ sufficiently small,*

$$\mathrm{cap}\,(\{y \in E : |y - x| \leq r\}) \geq C(x)\, r^{K(x)}.$$

Show that, for any $x \in E$, there exists $(\mu_n) \in \mathcal{M}_1(E)$ with $\mu_n \xrightarrow{*} \delta_x$ (the Dirac measure), and logarithmic potential V^{μ_n} being bounded uniformly in n by some integrable function (use the maximum principle for logarithmic potentials). Also, show that $I(\mu_n, \omega_E) \to U^{\omega_E}(x)$, and deduce that E is regular. □

Exercise 2.1.2 *Using the fact that* $\mathrm{cap}\,([a,b]) = (b-a)/4$, *show that a finite union of compact non degenerate intervals is regular. Does it remain regular if one adds an additional point?* □

Logarithmic potential theory has a nice electrostatic interpretation in $\mathbb{R}^2$ (or for cylinder symmetric configurations in $\mathbb{R}^3$, that is, a mass point in $\mathbb{R}^2$ corresponds to an infinite wire in $\mathbb{R}^3$ with uniform charge). Given a positive unit charge at zero, its electric potential is given by U^{δ_0}. Under this point of view, $\mu \in \mathcal{M}_1(E)$ represents a (static) distribution of a positive unit charge on some set E, with electric potential U^μ and electric energy $I(\mu)$. In physics, the equilibrium state is always described as the one having minimal energy. Thus, ω_E may be considered as the equilibrium distribution of a positive unit charge on a conducting material E, which by physical reasons should have a constant electric potential on E. Also, the fact that $\operatorname{supp}(\omega_E)$ is subset of the outer boundary of E (see for instance [MaFi04]) is known in physics as the Faraday principle.

One may wonder about what kind of equilibrium distribution is obtained if there is some additional fixed external field, induced for instance by some negative charge on some isolator outside of E. This problem has been already considered a long time ago by Gauß. Mathematically speaking, we have to solve the following problem:

Definition 2.1.3 Let $\Sigma \subset \mathbb{R}$ be a regular compact set, $t > 0$, and $Q \in \mathcal{C}(\Sigma)$. For $\mu \in \mathcal{M}_t(\Sigma)$, consider the weighted energy $I^Q(\mu) = I(\mu) + 2\int Q\, d\mu$.

We consider the problem of finding

$$W_{t,Q,\Sigma} := \inf\{I^Q(\mu) : \mu \in \mathcal{M}_t(\Sigma)\},$$

and, if possible, an extremal measure $\mu_{t,Q,\Sigma}$ avec $I^Q(\mu_{t,Q,\Sigma}) = W_{t,Q,\Sigma}$. □

If the external field Q is repealing positive charges, it may happen that the support of some extremal measure is a proper subset of Σ. By physical arguments, it should happen that there is a unique equilibrium, and that the corresponding potential is constant on the part of Σ which is charged by the equilibrium measure, and larger than this constant else on Σ. Indeed, one may give a mathematical proof of this statement.

Theorem 2.1.4 *Let $\Sigma \subset \mathbb{R}$ be a regular compact set, $t > 0$, and $Q \in \mathcal{C}(\Sigma)$. The extremal measure $\mu_{t,Q,\Sigma}$ of Definition 2.1.3 exists and is unique.*

Moreover, with $w = w_{t,Q,\Sigma} := W_{t,Q,\Sigma} - \int Q\, d\mu_{t,Q,\Sigma}$ and $\mu = \mu_{t,Q,\Sigma}$ there holds

$$U^\mu(z) + Q(z) \geq w \qquad \text{for } z \in \Sigma, \text{ and} \tag{2.1.5}$$
$$U^\mu(z) + Q(z) \leq w \qquad \text{for } z \in \operatorname{supp}(\mu). \tag{2.1.6}$$

Conversely, if there is a measure $\mu \in \mathcal{M}_t(\Sigma)$ and a constant w such that (2.1.5) and (2.1.6) hold, then $\mu = \mu_{t,Q,\Sigma}$ and $w = w_{t,Q,\Sigma}$.

Remark 2.1.5 It follows from (2.1.5) and (2.1.6) that the potential of $\mu_{t,Q,\Sigma}$ equals $w_{t,W,\Sigma} - Q$ on $\operatorname{supp}(\mu_{t,Q,\Sigma})$, the latter being continuous. Thus, by the principle of continuity [SaTo97, Theorem II.3.5], the potential of $\mu_{t,Q,\Sigma}$ is continuous. □

The main ideas for the proof of Theorem 2.1.4 are discussed in the following exercise.

Exercise 2.1.6 *Let $\mathcal{N}$ be some closed convex subset of $\mathcal{M}_t(\Sigma)$, containing at least one element μ with $I(\mu) < \infty$.*

(a) *Show that I^Q is strictly convex.*

(b) *Using the Theorem of Helly, show that $W_{\mathcal{N}} = \min\{I^Q(\mu) : \mu \in \mathcal{N}\}$ is attained for some $\mu_{\mathcal{N}} \in \mathcal{N}$. Why such a mesure must be unique?*

(c) *By discussing $\nu = s\mu + (1-s)\mu_{\mathcal{N}}$ for $s \to 0+$, show that $\mu_{\mathcal{N}}$ is uniquely characterized by the property*

$$\mu \in \mathcal{N} : \qquad w_{\mathcal{N}} := I(\mu_{\mathcal{N}}) + \int Q\, d\mu_{\mathcal{N}} \leq I(\mu, \mu_{\mathcal{N}}) + \int Q\, d\mu.$$

□

The only property of Theorem 2.1.4 not being an immediate consequence of the preceding exercise is the fact that the equilibrium measure satisfies the equilibrium conditions (2.1.5) and (2.1.6). For showing this property, one applies a principle known as *Principle of domination* for logarithmic potentials.

Theorem 2.1.7 *If $\mu, \nu \in \mathcal{M}(\Sigma)$ with finite energy, if $||\nu|| \leq ||\mu||$, and if, for some constant C, the relation $U^\mu \leq U^\nu + C$ holds μ-almost everywhere, then this relation holds for all $z \in \mathbb{C}$.*

Proof. See [SaTo97, Theorem II.3.2]. □

Theorem 2.1.4 together with Theorem 2.1.7 allow us to derive the following result known as the weighted Bernstein-Walsh inequality.

Corollary 2.1.8 *With the assumptions of Theorem 2.1.4, let $w(x) = \exp(-Q(x))$. Then for any polynomial of degree at most n and for all $z \in \mathbb{C}$*

$$\frac{|P(z)|}{||w^n P||_{L_\infty(E)}} \leq \exp(n w_{1,Q,\Sigma} - n U^{\mu_{1,Q,\Sigma}}(z)).$$

Proof. Exercice. □

Remark 2.1.9 For a trivial weight $w = 1$, Corollary 2.1.8 reduces to the classical Bernstein-Walsh inequality, which in terms of the function E_n of (1.4.5) can be rewritten as $E_n(z, S) \geq \exp(-n g_S(z))$. However, for applications in numerical linear algebra we need not lower but upper bounds for E_n. Depending on the "smoothness" of the set S, the Bernstein-Walsh inequality is more or less sharp, for instance, from Lemma 1.4.6 and the explicit formula of (2.1.3) we learn that $E_n(z, [a,b]) \leq 2\exp(-n g_{[a,b]}(z))$. For general sets S of positive capacity and $z \notin S$ ony may show (see, e.g., [NiSo88, Section V.5.3]) that $E_n(z, S)^{1/n}$ tends to $\exp(-g_S(z))$ for $n \to \infty$. □

We learn from Corollary 2.1.8 that the weighted maximum norm of a polynomial P lives on a possibly proper compact subset of Σ, that is, $||w^n P||_{L_\infty(E)} = ||w^n P||_{L_\infty(S^*(1,Q,\Sigma))}$ with

$$S^*(t, Q, \Sigma) := \{z \in E : U^{\mu_{t,Q,\Sigma}}(z) + Q(z) \leq w_{t,Q,\Sigma}\}. \qquad (2.1.7)$$

The link between $S^*(t, Q, \Sigma)$ and the support of the equilibrium measure is studied in [SaTo97, Theorem IV.4.1], and in more detail by Buyarov and Rakhmanov [BuRa99]. We state parts of their findings without proof.

Theorem 2.1.10 ([BuRa99]) *Let Σ, Q be as in Theorem 2.1.4, and define the sets*

$$S(t) := \mathrm{supp}\,(\mu_{t,Q,\Sigma}), \qquad t \geq 0.$$

The sets $S(t)$ are increasing in t, with $\cap_{\tau>t} S(\tau) = S^*(t, Q, \Sigma)$ for all t, and $S(t) = S^*(t, Q, \Sigma)$ for almost all t. Furthermore, for all $t > 0$ and $z \in \mathbb{C}$,

$$\mu_{t,Q,\Sigma} = \int_0^t \omega_{S(\tau)}\, d\tau, \qquad w_{t,Q,\Sigma} - U^{\mu_{1,Q,\Sigma}}(z) = \int_0^t g_{S(\tau)}(z)\, d\tau.$$

□

Theorem 2.1.10 tells us that all extremal quantities are completely determined once one knows $S(t)$ for all $t > 0$. Notice that $S(t)$ may consist of several intervals, or even have a Cantor-like structure. The determination of $S(t)$ for particular classes of external fields Q is facilitated by tools like the F-functional of Rakhmanov-Maskhar-Saff, see [MaFi04] or [SaTo97, Section IV.1.11], but remains in general a quite difficult task.

A quite interesting example for the preceding findings is the case of an exponential weight $w(x) = \exp(-x^2)$, described below, or more generally Freud weights. This example can be considered as the starting point for the research on weighted energy problems in the last twenty years.

Example 2.1.11 Given $\alpha > 0$, let $w(x) = \exp(-\gamma_\alpha |x|^\alpha)$ (i.e., $Q(x) = \gamma_\alpha |x|^\alpha$), with

$$\gamma(\alpha) = \int_0^1 \frac{u^{\alpha-1}}{\sqrt{1-u^2}}\, du = \frac{\Gamma(\alpha/2)\Gamma(1/2)}{2\Gamma((\alpha+1)/2)}$$

($\gamma_2 = 1$), and define the probability measure μ with $\mathrm{supp}\,(\mu) = [-1, 1]$ by the weight function

$$\frac{d\mu}{d\lambda}(\lambda) = s(\alpha, \lambda) = \frac{\alpha}{\pi} \int_{|\lambda|}^1 \frac{u^{\alpha-1}}{\sqrt{u^2 - \lambda^2}}\, du.$$

On shows [SaTo97, Theorem IV.5.1] that

$$U^\mu(x) + Q(x) \begin{cases} = w := \log(2) + 1/\alpha & \text{for } x \in [-1, 1], \\ > w = \log(2) + 1/\alpha & \text{for } x \in \mathbb{R} \setminus [-1, 1]. \end{cases}$$

Hence, by Theorem 2.1.4, for any compact Σ containing $[-1, 1]$ we have $\mu_{1,Q,\Sigma} = \mu$, and $S^*(1, Q, \Sigma) = [-1, 1]$. In particular, for any polynomial P of degree at most n we get from Corollary 2.1.8 that

$$||w^n P||_{L^\infty(\Sigma)} = ||w^n P||_{L^\infty(\mathbb{R})} = ||w^n P||_{L^\infty([-1,1])}.$$

Using the linear transformation $y = x \cdot (n\gamma_\alpha)^{1/\alpha}$, it follows that

$$||e^{-|y|^\alpha} P(y)||_{L^\infty(\mathbb{R})} = ||e^{-|y|^\alpha} P(y)||_{L^\infty([-(n\gamma_\alpha)^{1/\alpha}, (n\gamma_\alpha)^{1/\alpha}])}.$$

□

Exercise 2.1.12 *Relate the findings of Example 2.1.11 to those of Theorem 2.1.10.* □

Remark 2.1.13 Using the results mentioned in Example 2.1.11 one may show for instance that the zeros of the Hermite orthogonal polynomials, after division by $\sqrt{n}$, have an asymptotic distribution given by the weight function $s(2,\lambda) = (2/\pi)\sqrt{1-\lambda^2}$ on $[-1,1]$. We refer the reader to [SaTo97, Section III.6] for more precise results on nth root asymptotics for L_p-extremal polynomials (including the case $p = 2$ of orthogonal polynomials) with respect to varying weights w^n. □

We terminate this section by discussing two particular examples of external fields given by the negative potential of a measure.

Exercise 2.1.14 *If $Q(z) = -U^\sigma(z)$ for some $\sigma \in \mathcal{M}(\Sigma)$ with continuous potential, show that $\mu_{t,Q,\Sigma} = \sigma + (t - ||\sigma||)\omega_\Sigma$ pour tout $t \geq ||\sigma||$.* □

Things are becoming more exciting if the external field is a negative potential of some measure σ with compact support outside of Σ. Here the extremal measure $\mu_{||\sigma||,-U^\sigma,\Sigma}$ can be considered as a sort of projection of σ onto Σ, more precisely, we obtain the "problem of balayage" [SaTo97, Section II.4] studied already by H. Poincaré and Ch. de la Vallée-Poussin: find a measure of the same mass as σ supported on Σ with potential coinciding (up to some constant) with U^σ on Σ. In terms of electrostatics, we look for a positive unit charge on some conductor being in equilibrium with some fixed negative unit charge (on some isolator).

Exercise 2.1.15 *Let $\Sigma \subset \mathbb{R}$ be compact and regular, $t > 0$, $Q(z) = -U^\sigma(z)$ for some measure σ with compact support* $\mathrm{supp}\,(\sigma) \not\subset \Sigma$*, finite energy, and potential continuous on Σ (the latter being true for instance if* $\mathrm{supp}\,(\sigma) \cap \Sigma$ *is empty), and write shorter* $S(t) = \mathrm{supp}\,(\mu_{t,Q,\Sigma})$.

(a) *Show that* $\mathrm{cap}\,(S(t)) > 0$*. Use the maximum principle for subharmonic functions for showing that $g_{S(t)}(z) > 0$ for all $z \notin S(t)$.*

(b) *Let $\Delta \subset \Sigma$ some Borel set with* $\mathrm{cap}\,(\Delta) = 0$*, and $\nu \in \mathcal{M}(\Sigma)$ with finite energy. Show that $\nu(\Delta) = 0$.*

(c) *By applying Theorem 2.1.7, show that*

$$z \in \mathbb{C}: \qquad U^{\mu_{t,Q,\Sigma}}(z) + Q(z) \leq w_{t,Q,\Sigma} + (||\sigma|| - t)\, g_{S(t)}(z).$$

(d) *Using the maximum principle for subharmonic functions and (c), show that $S(t) = \Sigma$ for all $t \geq ||\sigma||$. Deduce that the balayage problem onto Σ has a unique solution given by $\mu_{||\sigma||,-U^\sigma,\Sigma}$.*

(e) *In the case $t < ||\sigma||$, show that $\mu_{t,Q,\Sigma} + (||\sigma|| - t)\,\omega_{S(t)}$ is the balayage of σ onto $S(t)$.*

□

Exercise 2.1.16 *Let $\Sigma \subsetneqq \Sigma' \subset \mathbb{R}$ be compact and regular. What is the balayage measure of $\omega_{\Sigma'}$ onto Σ?* □

For a regular compact $\Sigma \subset \mathbb{R}$, we may define the Green function with pole at $a \in \mathbb{C} \setminus \Sigma$ by the (balayage) formula of a Dirac measure

$$g_\Sigma(z, a) = w_{1,Q,\Sigma} - U^{\mu_{1,Q,\Sigma}}(z) - Q(z), \qquad Q(z) = -U^{\delta_a}(z),$$

compare with [SaTo97, Section II.4]. One may show [SaTo97, Eqn. (II.4.31)] that it is possible to integrate the preceding formula with respect to a: provided that $Q(z) = -U^\sigma(z)$, we have

$$w_{||\sigma||,Q,\Sigma} - U^{\mu_{||\sigma||,Q,\Sigma}}(z) - Q(z) = \int g_\Sigma(z, a)\, d\sigma(a). \tag{2.1.8}$$

Provided that explicit formulas for $g_\Sigma(\cdot, \cdot)$ are available, this formula can be exploited to derive explicit formulas for the density of the balayage measure, by recovering the measure from its potential [SaTo97, Chapter II.1]. E.g., for $x \in \Sigma = [a, b]$, an interval, and $\sigma(\Sigma) = 0$, corresponding formulas are given in [SaTo97, Corollary IV.4.12]

$$\frac{d\mu_{||\sigma||,-U^\sigma,[a,b]}}{dx}(x) = \frac{1}{\pi} \int \frac{\sqrt{|y-a|\,|y-b|}}{|y-x|\,\sqrt{(x-a)(b-x)}}\, d\sigma(y). \tag{2.1.9}$$

2.2 Energy problems with constraint and external field

Discrete Chebyshev polynomials are orthonormal with respect to the scalar product

$$\prec P, Q \succ = \sum_{z \in E_N} w_N(z)^2\, \overline{P(z)}\, Q(z)$$

with $E_N = \{0, 1, ..., N\}$ and $w_N(z) = 1$. Further systems of "classical" discrete orthogonal polynomials contain

$$\begin{array}{llll} \text{Meixner polynomials} : & E_N = \{0,1,2,...\}, & w_N(k)^2 = \dfrac{c^k (b)_k}{k!} \\ \text{Charlier polynomials} : & E_N = \{0,1,2,...\}, & w_N(k)^2 = \dfrac{c^k e^{-c}}{k!} \end{array}$$

Krawtchouk polynomials, discrete Freud polynomials, discrete Hahn polynomials, see for instance [Chi78, DrSa97, DaSa98, KuVA99, KuRa98] and the references therein.

It was Rakhmanov [Rak96] who first observed that the nth root of the nth discrete Chebyshev polynomial (and other discrete orthogonal polynomials) for so-called ray sequences, that is, $n, N \to \infty$ in such a manner that $n/N \to t \in (0, 1)$, can be described in terms of a constrained weighted equilibrium problem in logarithmic potential theory. Recall from Chapter 1.3 that asymptotics of discrete Chebyshev polynomials are closely related to the convergence behavior of Krylov subspace methods applied to a matrix with equally spaced eigenvalues, and to the convergence of its Ritz values. Other domains of applications for asymptotics of discrete orthogonal polynomials include coding theory and discrete dynamical systems.

Similar to the approach for Hermite polynomials (c.f. Remark 2.1.13), for obtaining nth root asymptotics it is first required to scale the set E_N by some appropriate power of N.

The resulting supports will then have an asymptotic distribution for $N \to \infty$ which can be described by some Borel measure σ. Furthermore, after scaling, the weights w_N will behave like $w(z)^N$ for some appropriate weight which can be written as $w = \exp(-Q)$. The constrained energy problem considered by Rakhmanov [Rak96] consists in minimizing the logarithmic energy $I(\mu)$, where μ is some probability measure satisfying in addition the constraint that $\sigma - \mu$ is some nonnegative measure. The set of such measures will be denoted by

$$\mathcal{M}_t^\sigma := \{\mu \geq 0 : ||\mu|| = t, \sigma - \mu \geq 0\}$$

where $0 < t \leq ||\sigma||$. In our context it will be useful to introduce a weighted analogue of this problem. Its unique solution has been characterized by Dragnev and Saff [DrSa97, Theorem 2.1 and Remark 2.3], and further investigated by several other authors. We summarize some of their findings in Theorem 2.2.1 below, here additional regularity assumptions on σ and Q enable us to obtain a simplified statement.

Theorem 2.2.1 (see [DrSa97]) *Let Q be a continuous real–valued function on some closed set $\Sigma \subset \mathbb{C}$, $w := \exp(-Q)$, and, if Σ is unbounded, suppose that $Q(z) - \log|z| \to +\infty$ for $|z| \to \infty$. Furthermore, let σ be a positive measure with $\mathrm{supp}\,(\sigma) \subset \Sigma$, such that, for any compact $K \subset \mathrm{supp}\,(\sigma)$, the restriction $\sigma|_K$ of σ to K has a continuous potential. Finally, let $0 < t < ||\sigma||$.*

Then for the extremal problem

$$W_{t,Q,\sigma} := \inf\{I^Q(\mu) : \mu \in \mathcal{M}_t^\sigma\}$$

there exists a unique measure $\mu_{t,Q,\sigma} \in \mathcal{M}_t^\sigma$ with $W_{t,Q,\sigma} = I^Q(\mu_{t,Q,\sigma})$, and this extremal measure has compact support. Furthermore, there exists a constant $w = w_{t,Q,\sigma}$ such that for $\mu = \mu_{t,Q,\sigma}$ we have the equilibrium conditions

$$U^\mu(z) + Q(z) \geq w \qquad \text{for } z \in \mathrm{supp}\,(\sigma - \mu), \text{ and} \tag{2.2.1}$$

$$U^\mu(z) + Q(z) \leq w \qquad \text{for } z \in \mathrm{supp}\,(\mu). \tag{2.2.2}$$

Conversely, if $\mu \in \mathcal{M}_t^\sigma$ has compact support and satisfies the equlibrium conditions (2.2.1), (2.2.2), for some constant w then $\mu = \mu_{t,Q,\sigma}$. □

In terms of electrostatics, we may consider $\mu_{t,Q,\sigma}$ as the equilibrium distribution on $\mathrm{supp}\,(\sigma)$ of a positive charge of mass t in the presence of an external field Q, but here $\mathrm{supp}\,(\sigma)$ is no longer conducting: indeed $\mu \leq \sigma$ imposes a constraint on the maximum charge per unit. As a consequence, the corresponding weighted potential is no longer constant on the whole part of $\mathrm{supp}\,(\sigma)$ charged by our extremal measure: we may have a strictly smaller weighted potential at the part $\mathrm{supp}\,(\sigma) \setminus \mathrm{supp}\,(\sigma - \mu_{t,Q,\sigma})$ where the constraint is active. However, in the free part $\mathrm{supp}\,(\sigma - \mu_{t,Q,\sigma}) \cap \mathrm{supp}\,(\mu_{t,Q,\sigma})$ the weighted potential is still constant.

We should comment on the proof of Theorem 2.2.1. For showing existence and uniqueness of the extremal measure, we can follow the reasoning of Exercice 2.1.6, at least for compact Σ (the growth condition on Q can be shown to imply that it is sufficient to consider compact Σ). A proof of the equivalent characterization by the equilibrium conditions (2.2.1) and (2.2.2) uses Exercice 2.1.6(c) and Theorem 2.1.7, as well as the following observation.

Exercise 2.2.2 *Let ν be a measure with compact support and continuous potential. Use twice the principle of descent for showing that any measure $\mu \geq 0$ with $\mu \leq \nu$ also has a continuous potential (c.f. [Rak96]). Show also that μ has no mass points.* $\square$

Remark 2.2.3 The extremal constant $w_{t,Q,\sigma}$ is not necessarily unique [DrSa97, Example 2.4], but will be unique if $\mathrm{supp}\,(\sigma)$ is connected, or, more generally, if $\mathrm{supp}\,(\mu_{t,Q,\sigma})$ and $\mathrm{supp}\,(\sigma - \mu_{t,Q,\sigma})$ have a non-empty intersection. $\square$

Example 2.2.4 For the constrain $\frac{d\sigma}{dx}(x) = \alpha x^{\alpha-1}$ on $\mathrm{supp}\,(\sigma) = [0, +\infty)$, $\alpha > 1/2$, and the external field $Q(x) = \gamma \cdot x^\alpha$, it is shown in [KuVA99, Theorem 2.1] that

$$\mu_{t,Q,\sigma} = \int_0^t \omega_{[a(\tau),b(\tau)]}\, d\tau, \qquad w_{t,Q,\sigma} - U^{\mu_{t,Q,\sigma}}(z) = \int_0^t g_{[a(\tau),b(\tau)]}(z)\, d\tau, \tag{2.2.3}$$

$$\mathrm{supp}\,(\mu_{t,Q,\sigma}) = [0, b(t)], \qquad \mathrm{supp}\,(\mu_{t,Q,\sigma}) \cap \mathrm{supp}\,(\sigma - \mu_{t,Q,\sigma}) = [a(t), b(t)], \tag{2.2.4}$$

where $0 \leq a(t) = t^\alpha a_0 < b(t) = t^\alpha b_0$ are solutions of the system

$$0 = \frac{1}{\pi} \int_{a(t)}^{b(t)} \frac{Q'(x)\, dx}{\sqrt{(b(t) - x)(x - a(t))}} - \int_{x \leq a(t)} \frac{d\sigma(x)}{\sqrt{(\beta(t) - x)(\alpha(t) - x)}}, \tag{2.2.5}$$

$$t = \frac{1}{\pi} \int_{a(t)}^{b(t)} \frac{xQ'(x)\, dx}{\sqrt{(b(t) - x)(x - b(t))}} - \int_{x \leq a(t)} \frac{x\, d\sigma(x)}{\sqrt{(b(t) - x)(a(t) - x)}}. \tag{2.2.6}$$

$\square$

Exercise 2.2.5 *In case of compact* $\mathrm{supp}\,(\sigma)$*, show the following property of duality*

$$\mu_{t,Q,\sigma} + \mu_{||\sigma||-t,\tilde{Q},\sigma} = \sigma, \qquad \textit{where } \tilde{Q} := -Q - U^\sigma.$$

$\square$

Let us compare our extremal problem to the unconstrained one of Definition 2.1.3. In case of a "sufficiently large" constraint we clearly see by comparing Theorem 2.2.1 with Theorem 2.1.4 that the following implication holds

$$\sigma \geq \mu_{t,Q,\Sigma} \qquad \Longrightarrow \qquad \mu_{t,Q,\sigma} = \mu_{t,Q,\Sigma}. \tag{2.2.7}$$

Of course, the same conclusion is true if $\mathrm{supp}\,(\sigma - \mu_{t,Q,\sigma}) = \Sigma$. In what follows we will consider sometimes the case of a trivial external field $Q = 0$ (and hence compact $\mathrm{supp}\,(\sigma)$). Here the following result will be helpful relating the constrained energy problem with trivial weight to an unconstrained weighted extremal problem and more precisely to a balayage problem.

Lemma 2.2.6 *Under the assumptions of Theorem 2.2.1, if $Q = 0$ then*

$$\sigma - \mu_{t,0,\sigma} = \mu_{||\sigma||-t,\tilde{Q},\tilde{\Sigma}}, \qquad \textit{where } \tilde{Q} := -U^\sigma \textit{ and } \tilde{\Sigma} := \mathrm{supp}\,(\sigma),$$

and moreover

$$\mathrm{supp}\,(\mu_{t,0,\sigma}) = \mathrm{supp}\,(\sigma).$$

Proof. Write shorter $S := \mathrm{supp}\,(\mu_{t,0,\sigma})$. Recall that g_S equals zero on S up to some set of capacity zero. From (2.2.2) with $Q = 0$ and Exercise 2.1.15(b) we conclude that

$$U^{\mu_{t,0,\sigma}}(z) + g_S(z) \leq w := w_{t,0,\sigma}$$

holds $\mu_{t,0,\sigma}$–everywhere on S, and the principle of domination of Theorem 2.1.7 tells us that the above inequality is true for all $z \in \mathbb{C}$. In particular, for $z \notin S$ we have $U^{\mu_{t,0,\sigma}}(z) < w$ by Exercise 2.1.15(a). Comparing with (2.2.1) shows that $z \notin \mathrm{supp}\,(\sigma - \mu_{t,0,\sigma}) \supset \mathrm{supp}\,(\sigma) \setminus S$. Hence $S = \mathrm{supp}\,(\sigma)$, as claimed in the assertion. Moreover, from (2.2.1) and (2.2.2) we see that $\sigma - \mu_{t,0,\sigma}$ satisfies the equilibrium conditions (2.1.5), (2.1.6) corresponding to the external field $\tilde{Q}$ on $\tilde{\Sigma}$ with the normalization $||\sigma|| - t$, and hence is equal to $\mu_{||\sigma||-t,\tilde{Q},\tilde{\Sigma}}$ by Theorem 2.1.4. □

Exercise 2.2.7 *Suppose that $(0, T) \ni t \mapsto S(t) \subset \mathbb{C}$ with $S(t)$ compact and regular decreasing sets, i.e., $S(t') \subset S(t)$ for $t' > t$, and consider the constraint*

$$\sigma(x) = \int_0^T \omega_{S(t)}(x)\, dt, \qquad T = ||\sigma||. \tag{2.2.8}$$

Show that

$$w_{t,0,\sigma} - U^{\mu_{t,0,\sigma}}(z) = \int_0^t g_{S(\tau)}(z)\, d\tau. \tag{2.2.9}$$

Hint: verify equilibrium conditions. □

Remark 2.2.8 As shown in [BeKu99, Theorem 2.1], it follows from Theorem 2.1.10 and Lemma 2.2.6 that the following more general statement is valid:

In case of a trivial external field $Q = 0$, the compact sets $S(t) := \mathrm{supp}\,(\sigma - \mu_{t,0,\sigma})$ are decreasing in t, any constraint σ has the integral representation (2.2.8), and formula (2.2.9) is true. □

Remark 2.2.9 It is an open problem of establishing integral formulas of Buyarov-Rakhmanov type in the case of the constrained weighted extremal problem for general external fields Q. However, beside the preceding remark, there is another case where such formulas may be established (compare with [KuMc00, Lemma 3.1, Theorem 3.3, Proof of Lemma 6.2], [DeMc98, Chapter 4], [Kui00b, Proposition 4.1] and [BeKu02]):

Let $\mathrm{supp}\,(\sigma) = [A, B]$, $Q(A) = 0$, and suppose that the functions Q and $\widetilde{Q}$ defined by

$$\widetilde{Q}(x) = -Q(x) - U^{\sigma}(x)$$

are continuous in $[A, B]$ and have a continuous derivative in (A, B). Suppose in addition that the functions $x \mapsto (x - A)Q'(x)$ and $x \mapsto (B - x)\widetilde{Q}'(x)$ are increasing functions on $[A, B]$. Then (2.2.3) and (2.2.4) hold, with $A \leq a(t) < b(t) \leq B$ defined by (2.2.5), (2.2.6). □

Some other results on the interval case for the constrained unweighted energy problem are given in [KuDr99, Theorem 2], [Kui00a, Theorem 5.1], [BeKu99, Lemma 3.1], in particular one may find (systems of) integral equations for determining the endpoints of $S(t)$, and links with the technique of balayage since, by Exercise 2.1.15 and Lemma 2.2.6, the measure $\sigma - \mu_{t,0,\sigma} + t\omega_{S(t)}$ coincides with the balayage of σ onto $S(t)$.

2.3 Asymptotics for discrete orthogonal polynomials

Let $T_{n,p}$ be the extremal polynomials of (1.1.2). In this section we describe how the nth root asymptotic of the extremal constants $||w_N \cdot T_{n,p}||_{L_p(E_N)}$ for ray sequences $n, N \to \infty$, $n/N \to t$, as well as the asymptotic distribution of zeros of $T_{n,p}$ may be expressed in terms of the solution of the constrained weighted energy problem of Section 2.2.

For some discrete set E_N we define the corresponding counting measure

$$\nu_N(E_N) = \frac{1}{N} \sum_{z \in E_N} \delta_z,$$

a discrete measure where each element of E_N is charged by the mass $1/N$. Similarly, given a polynomial P with set of zeros Z, we write $\nu_N(P) := \nu_N(Z)$ for the corresponding normalized zero counting measure (where we count zeros according to their multiplicities). As usual, for a sequence of discrete sets $(E_N)_N$ we write $\nu_N(E_N) \xrightarrow{*} \sigma$ if for any continuous function f with compact support there holds

$$\lim_{N\to\infty} \int f(z) d\nu_N(z) = \int f(z)\, d\sigma(z), \quad \text{where} \quad \int f(z) d\nu_N(z) = \frac{1}{N} \sum_{z \in E_N} f(z).$$

Finally, for discrete sets E_N, F_N we define the discrete mutual energy

$$I_N(E_N, F_N) = \frac{1}{N^2} \sum_{x \in E_N} \sum_{y \in F_N, y \neq x} \log(\frac{1}{|x-y|}),$$

the mutual energy between two systems E_N and F_N of positive masspoints.

Exercise 2.3.1 *Suppose that $\nu_N(E_N) \xrightarrow{*} \mu$, $\nu_N(F_N) \xrightarrow{*} \nu$. Show the semi-continuity*

$$\liminf_{N\to\infty} I_N(E_N, F_N) \geq I(\mu, \nu).$$

Hint: consider the regularized kernel $(x, y) \mapsto \max\{\eta, \log(\frac{1}{|x-y|})\}$ for $\mathbb{R} \ni \eta \to +\infty$. □

Weak asymptotics of discrete L_p-extremal polynomials have been a subject of a number of publications, see [DrSa97, Theorem 3.3] (for real compact Σ), [DaSa98, Theorem 2.5] (for $p = \infty$ and real Σ), [KuVA99, Theorem 7.4 and Lemma 8.3] (for $0 < p \leq \infty$ and real Σ), [KuRa98, Theorem 7.1] (for $p = 2$ and real Σ, see also [KuRa99]), and finally [Be00a, Theorem 1.3] (for $0 < p \leq \infty$ and complex supports). We summarize these findings in the following (a bit technical) Theorem.

Theorem 2.3.2 ([Be00a]) *Let $0 < p \leq \infty$. Furthermore, let Σ, σ, Q, $w = \exp(-Q)$ be as in Theorem 2.2.1, and $t \in (0, ||\sigma||)$ with* $\operatorname{supp}(\mu_{t,Q,\sigma}) \cap \operatorname{supp}(\sigma - \mu_{t,Q,\sigma}) \neq \emptyset$.

Suppose that the sets $E_N \subset \Sigma$ and the weights $w_N(z) \geq 0$, $z \in E_N$, $N \geq 0$, satisfy the conditions

$$\nu_N(E_N) \xrightarrow{*} \sigma, \tag{2.3.1}$$

$$\limsup_{N\to\infty} \sup_{z \in E_N} \frac{w_N(z)^{1/N}}{\tilde{w}(z)} \leq 1, \tag{2.3.2}$$

for some $\tilde{w} \in \mathcal{C}(\Sigma)$ *with* $|z|\tilde{w}(z) \to 0$ *for* $|z| \to \infty$ *and if for any compact* K *there holds*

$$\limsup_{N\to\infty} \sup_{z\in E_N\cap K} \frac{w_N(z)^{1/N}}{w(z)} \leq 1, \tag{2.3.3}$$

then

$$\limsup_{n,N\to\infty, n/N\to t} ||w_N \cdot T_{n,p}||_{L_p(E_N)}^{1/N} \leq \exp(-w_{t,Q,\sigma}), \tag{2.3.4}$$

and for any $z \in \mathbb{C}$ *satisfying*

$$\lim_{N\to\infty} U^{\nu_N(V(z)\cap E_N)}(z) = U^{\sigma|_{V(z)}}(z) \tag{2.3.5}$$

for some open neighborhood $V(z)$ *of* z *with* $\sigma(\partial V(z)) = 0$ *there holds*

$$\limsup_{n,N\to\infty, n/N\to t} \Big[\min_{\deg P\leq n} \frac{||w_N \cdot P||_{L_p(E_N)}}{|P(z)|}\Big]^{1/N} \leq \exp(U^{\mu_{t,Q,\sigma}}(z) - w_{t,Q,\sigma}), \tag{2.3.6}$$

If moreover there exists some bounded open neigborhood V *of* $\mathrm{supp}\,(\mu_{t,Q,\sigma})$ *with*

$$\lim_{N\to\infty} I_N(V \cap E_N, V \cap E_N) = I(\sigma|_V), \tag{2.3.7}$$

if for any compact K *there holds*

$$\lim_{N\to\infty} \sup_{z\in E_N\cap K} |w_N(z)^{1/N} - w(z)| = 0, \tag{2.3.8}$$

and if in the case $p < \infty$ *there exists a* $p' \in (0, p)$ *with*

$$\limsup_{N\to\infty} \big[||[z\, w]^N||_{L_{p'}(E_N)}\big]^{1/N} < \infty, \tag{2.3.9}$$

then (2.3.4) and (2.3.6) are sharp, more precisely, we have

$$\lim_{n,N\to\infty, n/N\to t} ||w_N \cdot T_{n,p}||_{L_p(E_N)}^{1/N} = \exp(-w_{t,Q,\sigma}), \tag{2.3.10}$$

$$\lim_{n,N\to\infty, n/N\to t} \Big[\min_{\deg P\leq n} \frac{||w_N \cdot P||_{L_p(E_N)}}{|P(z)|}\Big]^{1/N} = \exp(U^{\mu_{t,Q,\sigma}}(z) - w_{t,Q,\sigma}), \tag{2.3.11}$$

with z *as in (2.3.5), and finally, if the two-dimensional Lebesgue measure of* Σ *is zero, there holds*

$$\nu_N(T_{n,p}) \xrightarrow{*} \mu_{t,Q,\sigma} \quad \textit{for } n, N \to \infty, n/N \to t. \tag{2.3.12}$$

The statement of this Theorem simplifies considerably for compact Σ (why?). Since then E_N has $\mathcal{O}(N)$ elements, we can use classical inequalities between Hölder norms showing that $L_p(E_N)$–norms for two different p are equivalent up to a factor being some power of N (which of course will vanish once we take Nth roots). Therefore it is evident that the right hand side of (2.3.4), (2.3.6), (2.3.10), (2.3.11), and (2.3.12) do not depend on p.

We will comment in Remark 2.3.5 below on the different assumptions and variations proposed by different authors and subsequently give the main ideas of the proof of Theorem 2.3.2 for compact Σ (for a proof for general Σ we refer the reader to [Be00a, Theorem 1.3 and Theorem 1.4(c)]). Let us first have a look at the examples mentioned in the introduction of Section 2.2.

Example 2.3.3 [Discrete Chebyshev polynomials] After scaling (dividing the support by N) we obtain assumption (2.3.1) with σ being the Lebesgue measure on $[0,1]$, having a continuous potential. Here conditions (2.3.2), (2.3.3), (2.3.8) and (2.3.9) are trivially true with $Q=0$. The interested reader may check that also condition (2.3.7) holds, and that, by (2.1.3),

$$\sigma(x) = \int_0^1 \omega_{S(t)}(x)\,dt, \qquad S(t) = [\frac{1}{2} - \frac{\sqrt{1-t^2}}{2}, \frac{1}{2} + \frac{\sqrt{1-t^2}}{2}].$$

Thus explicit formulas for $\mu_{t,0,\sigma}$ are given in Exercice 2.2.7, in particular one obtains from (2.3.11) for $p=+\infty$ (compare with [BeKu99, Corollary 3.2])

$$\lim_{n,N\to\infty, n/N\to t} \log(E_n(0,\{1/N,2/N,...,N/N\})^{1/N}) \tag{2.3.13}$$
$$= w_{t,0,\sigma} - U^{\mu_{t,0,\sigma}}(0) = -\frac{(1+t)\log(1+t)+(1-t)\log(1-t)}{2}.$$

□

Example 2.3.4 [Meixner polynomials] Here, again after division of the support by N, we obtain for σ the Lebesgue measure on $[0,+\infty)$. By Stirling's formula, we have for $x=k/N$ and $w(x)=\exp(x\log(c)/2)$

$$\frac{w_N(x)^{1/N}}{w(x)} = c^{-x/2}[\frac{c^{Nx}\Gamma(Nx+b)}{\Gamma(Nx+1)}]^{1/(2N)} = 1+o(1)_{N\to\infty}$$

uniformly for x in some compact. Hence (2.3.2), (2.3.3), (2.3.8) and (2.3.9) are true with $Q(x)=x\log(1/c)/2$. The corresponding equilibrium measure $\mu_{t,Q,\sigma}$ has been given in Example 2.2.4. Here we may explicitly solve the system (2.2.5), (2.2.6), of integral equations, and $a_0=(1-\sqrt{c})/(1+\sqrt{c})=1/b_0$. □

Remark 2.3.5 Let us shortly comment on the different assumptions of Theorem 2.3.2 and relate them with related conditions proposed by other authors. Conditions (2.3.1), (2.3.3) and (2.3.8) allow to relate our discrete L_p norm to the extremal problem with data σ and Q.

Conditions (2.3.2) and (2.3.9) insure the finiteness of $||w_n P||_{L_p(E_n)}$ for a polynomial of degree at most n, at least for sufficiently large n. Such an additional condition is required for $p<\infty$ for controlling the contribution to the L_p norm of in modulus large elements of E_n. Stronger sufficient conditions for (2.3.9) in case of unbounded Σ have been discussed in [KuVA99] and [Be00a, Lemma 2.7].

Finally, by considering the example $E_N = F_N \cup (e^{-N}+F_N)$, $F_N=\{0,1/N,2/N,...,N/N\}$ it becomes clear that our asymptotic bounds (2.3.4), (2.3.6) cannot be sharp since they do not take into account the clustering of points of the support, compare also the discussion in [KuRa98, Section 8]. Rakhmanov [Rak96] considered the additional separation condition

$$\liminf_{N\to\infty} \inf_{x,y\in E_N\cap K, x\neq y} N\cdot|x-y| > 0$$

for all compact sets K. The weaker condition

$$\lim_{N\to\infty} \max_{y\in K\cap E_N} \Big| \prod_{x\in E_N\cap K, x\neq y} |y-x|^{1/N} - \exp(-U^{\sigma|_K}(y)) \Big| = 0$$

for any compact K was proposed in [DrSa97] (see also [DaSa98, KuVA99, KuRa98] for some generalizations). It may be shown [DrSa97, Lemma 3.2] that, e.g., sets of zeros of suitable orthogonal polynomials satisfy this condition. One may show that any of these two conditions imply (2.3.7). This latter separation condition (2.3.7) was conjectured to be sufficient by Rakhmanov at the Sevilla OPSF conference [KuRa98, Conjectures 2 and 3], and proved to be sufficient later in [Be00a]. $\square$

We terminate this section by giving the main ideas of the proof of Theorem 2.3.2 for compact Σ, compare also with [BeKu99, Theorem 2.1 and Theorem 2.2] for $Q = 0$ and [BeKu02, Theorem 2.2] for general Q. For general Σ the reader may consult the statements [Be00a, Theorem 1.3 and Theorem 1.4(c)] and their proofs.

Proof of (2.3.4), (2.3.6). Given $\epsilon > 0$, it is sufficient to construct a sequence of monic polynomials p_N of degree $n = n(N)$ with $n(N)/N \to t$ for $N \to \infty$, such that

$$\limsup_{N\to\infty} ||w_N p_N||_{L_\infty(E_N)}^{1/N} \leq e^{\epsilon - w_{t,Q,\sigma}}, \quad \lim_{N\to\infty} |p_N(0)|^{1/N} = e^{-U^{\mu_{t,Q,\sigma}}(0)}, \tag{2.3.14}$$

where we suppose that (2.3.5) holds for $z = 0$ and some $V(0)$ (notice that (2.3.5) is true for any $z \notin \Sigma$ by assumption (2.3.1) since then $x \mapsto \log(|x - z|)$ is continuous in Σ). We will choose the zeros of p_N in E_N.

First notice that, by assumption (2.3.3), it is sufficient to show (2.3.14) for $w_N = w^N$. The main idea of the proof is that one is able to discretize $\mu_{t,Q,\sigma}$ with help of points in E_N: there exist sets E_N^* with

$$\mathrm{card}(E_N^*) = n(N), \qquad E_N^* \subset E_N, \qquad \nu_N(E_N^*) \xrightarrow{*} \mu_{t,Q,\sigma} \quad \text{for } N \to \infty, \tag{2.3.15}$$

see [BeKu99, Lemma A.1] for real Σ and [Be00a, Lemma 2.1(d)]. Consider the polynomial p_N with simple zeros given by the elements of E_N^*, and the compact set

$$K_\epsilon = \{\lambda \in \Sigma : U^{\mu_{t,Q,\sigma}}(\lambda) + Q(z) \leq w_{t,Q,\sigma} - \epsilon\}, \tag{2.3.16}$$

then from Theorem 2.2.1 and from the uniqueness of the extremal constant $w_{,Q,\sigma}$ we know that $\mu_{t,Q,\sigma}(K_\epsilon) = \sigma(K_\epsilon) < t$. Hence only $o(N)$ elements of $E_N \cap K_\epsilon$ are not in $E_N^* \cap K_\epsilon$, but more than $o(N)$ elements lie in $E_N^* \setminus K_\epsilon$. Hence, by possibly exchanging $o(N)$ elements we may add to (2.3.15) the additional requirement that

$$E_N \cap K_\epsilon = E_N^* \cap K_\epsilon,$$

implying that

$$||w^N p_N||_{L_\infty(E_N)}^{1/N} = ||w^N p_N||_{L_\infty(E_N \setminus K_\epsilon)}^{1/N} =: \exp(-Q(\zeta_N) - U^{\nu_N(E_N^*)}(\zeta_N))$$

for some $\zeta_N \in E_N \setminus K_\epsilon$. By going to subsequences if necessary, we may suppose that $\zeta_N \to \zeta \in \Sigma$, and hence by the principle of descent (2.1.2) and by continuity of Q

$$\limsup_{N\to\infty} ||w^N p_N||_{L_\infty(E_N)}^{1/N} \leq \sup_{\zeta \in \Sigma \setminus K_\epsilon} \exp(-Q(\zeta) - U^{\mu_{t,Q,\sigma}}(\zeta)) \leq e^{\epsilon - w_{t,Q,\sigma}}$$

the last inequality following from the definition of K_ϵ. We also have from the principle of descent and from (2.3.15) that

$$\limsup_{N\to\infty} |p_N(0)|^{1/N} = \limsup_{N\to\infty} \exp(-U^{\nu_N(E_N^*)}(0)) \leq e^{-U^{\mu_{t,Q,\sigma}}(0)}.$$

The assumption $\sigma(\partial V(0)) = 0$ and thus $\mu_{t,Q,\sigma}(\partial V(0)) = 0$ allows us to conclude that

$$\nu_N(E_N^* \setminus V(0)) \xrightarrow{*} \mu_{t,Q,\sigma} - \mu_{t,Q,\sigma}|_{V(0)}, \qquad \nu_N(V(0) \cap (E_N \setminus E_N^*)) \xrightarrow{*} \sigma|_{V(0)} - \mu_{t,Q,\sigma}|_{V(0)},$$

and hence again by the principle of descent and by (2.3.5)

$$\begin{aligned}
&\limsup_{N\to\infty} U^{\nu_N(E_N^*)}(0) \\
&= \lim_{N\to\infty} U^{\nu_N(E_N^*\setminus V(0))}(0) + \lim_{N\to\infty} U^{\nu_N(E_N\cap V(0))}(0) - \liminf_{N\to\infty} U^{\nu_N(V(0)\cap(E_N\setminus E_N^*))}(0) \\
&\geq U^{\mu_{t,Q,\sigma}-\mu_{t,Q,\sigma}|_{V(0)}}(0) + U^{\sigma|_{V(0)}}(0) - U^{\sigma|_{V(0)}-\mu_{t,Q,\sigma}|_{V(0)}}(0) = U^{\mu_{t,Q,\sigma}}(0),
\end{aligned}$$

showing (2.3.14). □

Proof of (2.3.10), (2.3.11). It is shown implicitly in [Be00a, Lemma 2.1(c) and Lemma 2.2] that we may suppose that the set V in (2.3.7) satisfies $\sigma(\partial V) = 0$. Let $F_N \subset E_N$ with $\nu_N(F_N) \xrightarrow{*} \mu$. We claim that

$$F_N \subset E_N \cap V, \quad \nu_N(F_N) \xrightarrow{*} \mu \quad \Longrightarrow \quad \lim_{N\to\infty} I_N(F_N, F_N) = I(\mu, \mu). \tag{2.3.17}$$

Indeed, since $\sigma(\partial V) = 0$, we have that $\nu_N(E_N \cap V) \xrightarrow{*} \sigma|_V$, and hence with $F_N' := (E_N \cap V) \setminus F_N$, $\nu_N(F_N') \xrightarrow{*} \sigma|_V - \mu$ we have

$$\begin{aligned}
&\limsup_{N\to\infty} I_N(F_N, F_N) = \limsup_{N\to\infty} \Big(I_N(E_N \cap V, E_N \cap V) - I_N(F_N', F_N') - 2I_N(F_N, F_N') \Big) \\
&\leq \limsup_{N\to\infty} (E_N \cap V, E_N \cap V) - \liminf_{N\to\infty} I_N(F_N', F_N') - 2 \liminf_{N\to\infty} I_N(F_N, F_N') \\
&\leq I(\sigma|_V) - I(\sigma|_V - \mu) - 2I(\mu, \sigma|_V - \mu) = I(\mu),
\end{aligned}$$

where in the last inequality we have applied Exercise 2.3.1 and (2.3.7). From Exercise 2.3.1 it also follows that $\liminf_N I_N(F_N, F_N) \geq I(\mu)$, showing that (2.3.17) holds.

Let $\epsilon > 0$, and K_ϵ as in (2.3.16). According to the equilibrium conditions in Theorem 2.2.1 and thanks to continuity we find some open set K with $\operatorname{supp}(\mu_{t,Q,\sigma}) \subset K \subset K_{-\epsilon}$. By possibly replacing K by some smaller set, we may also suppose that $K \subset V$, and that $\sigma(\partial K) = 0$. Finally, notice that $t' := (\sigma - \mu_{t,Q,\sigma})(K) > 0$ since for any ζ in the by assumption non-empty set $\operatorname{supp}(\mu_{t,Q,\sigma}) \cap \operatorname{supp}(\sigma - \mu_{t,Q,\sigma})$ there exists a small neighborhood U of ζ with $U \subset V$, and thus $t' \geq (\sigma - \mu_{t,Q,\sigma})(U) > 0$ by definition of the support.

We now consider the weighted Fekete points, a set Φ_N of $n(N) + 1$ elements of $E_N \cap K$ which minimize the expression

$$I_N(\Phi_N, \Phi_N) + 2 \int Q \, d\nu_N(\phi_N).$$

By discretizing $\mu_{t,Q,\sigma}$ as in the preceding proof with $n(N) + 1$ elements in $E_N \cap K$, we obtain a candidate Φ_N^* with $I_N(\Phi_N^*, \Phi_N^*) + 2 \int Q \, d\nu_N(\phi_N^*) \to I^Q(\mu_{t,Q,\sigma})$ according to (2.3.17). Hence

$$I^Q(\mu_{t,Q,\sigma}) \geq \limsup_{N\to\infty} I_N(\Phi_N, \Phi_N) + 2 \int Q \, d\nu_N(\phi_N).$$

On the other hand, by Exercice 2.3.1 and Theorem 2.2.1,

$$\liminf_{N\to\infty} I_N(\Phi_N, \Phi_N) + 2 \int Q \, d\nu_N(\phi_N) \geq I^Q(\mu_{t,Q,\sigma}),$$

which by the uniqueness of the extremal measure shows that $\nu_N(\phi_N) \xrightarrow{*} \mu_{t,Q,\sigma}$.

According to (2.3.4), (2.3.6) and (2.3.8), the assertions (2.3.10) and (2.3.11) will follow by showing that

$$\begin{aligned}
&\liminf_{N\to\infty} \Big[\min_{\deg P \le n(N)} \frac{||w^N P||_{L_\infty(\Phi_N)}}{|P(0)|} \Big]^{1/N} \ge \exp(U^{\mu_{t,Q,\sigma}}(0) - w_{t,Q,\sigma}), \\
&\liminf_{N\to\infty} \min_{\deg P \le n(N)} ||w^N P||_{L_\infty(\Phi_N)}^{1/N} \ge \exp(-w_{t,Q,\sigma}),
\end{aligned}$$

However, since Φ_N has $n(N)+1$ elements, both expressions on the left can be written explicitly in terms of Lagrange polynomials, a task which we leave as an exercise. Then it is not difficult to see that the above two formulas follow from the principle of descent, and from a fact which we will show now: for any $z_N \in \Phi_N$ with $E_N^* := \Phi_N \setminus \{z_N\}$ there holds

$$\limsup_{N\to\infty} U^{\nu_N(E_n^*)}(z_N) + Q(z_N) \le w_{t,Q,\sigma} + \epsilon. \tag{2.3.18}$$

Write $F_N := (E_N \cap K) \setminus E_N^*$, with $n'(N)$ elements, and observe that

$$\nu_N(F_N \cup E_N) \to \sigma|_K, \quad \nu_N(F_N) \to \sigma|_K - \mu_{t,Q,\sigma}, \quad \frac{n'(N)}{N} \to t' = (\sigma - \mu_{t,Q,\sigma})(K)$$

where $t' > 0$, and, by (2.3.17),

$$\begin{aligned}
\lim_{N\to\infty} I_N(E_N^*, F_N) &= \lim_{N\to\infty} \frac{I_N(E_N^* \cup F_N, E_N^* \cup F_N) - I_N(E_N^*, E_N^*) - I_N(F_N, F_N)}{2} \\
&= I(\mu_{t,Q,\sigma}, \sigma|_K - \mu_{t,Q,\sigma}).
\end{aligned}$$

It follows from the definition of the Fekete points (replace one element of Φ_N by an element of $(E_N \cap V) \setminus \Phi_N$) that

$$z \in F_N : \qquad U^{\nu_N(E_N^*)}(z_N) + Q(z_N) \le U^{\nu_N(E_N^*)}(z) + Q(z).$$

Therefore, we can bound $U^{\nu_N(E_N^*)}(z_N) + Q(z_N)$ above by

$$\frac{1}{n'(N)} \sum_{z \in F_N} [U^{\nu_N(E_N^*)}(z) + Q(z)] = \frac{N}{n'(N)} [I_N(E_N^*, F_N) + \int Q \, d\nu_N(F_N)],$$

the right-hand term tending for $N \to \infty$ to

$$\frac{1}{t'} \Big(I(\mu_{t,Q,\sigma}, \sigma|_K - \mu_{t,Q,\sigma}) + \int Q \, d(\sigma|_K - \mu_{t,Q,\sigma}) \Big) = \frac{1}{t'} \int \Big(U^{\mu_{t,Q,\sigma}} + Q \Big) \, d(\sigma|_K - \mu_{t,Q,\sigma}) \Big)$$

which according to $\operatorname{supp}(\sigma|_K - \mu_{t,Q,\sigma}) \subset K \subset K_{-\epsilon}$ can be bounded above by $w_{t,Q,\sigma} + \epsilon$, as claimed in (2.3.18). □

Proof of (2.3.12). See [Be00a, Theorem 1.3(b)]. □

Chapter 3

Consequences

3.1 Applications to the rate of convergence of CG

As mentioned already in the introduction, we want to provide a better understanding of the superlinear convergence of CG iteration, and in particular to explain the form of the error curve as seen in Figure 1.2, and in all examples considered below. Recall from Corollary 1.4.8 the link between the CG error of a positive definite matrix A of size $N \times N$ having the spectrum $\Lambda(A)$, and the quantity $E_n(0, \Lambda(A))$. We will argue that for large N, the error $E_n(0, \Lambda(A))$ in the polynomial minimization problem (1.4.5) is approximately

$$\frac{1}{N} \log E_n(\Lambda(A)) \approx -\int_0^t g_{S(\tau)}(0) d\tau \tag{3.1.1}$$

where $t = n/N \in (0, 1)$ and $S(\tau)$, $\tau > 0$, is a decreasing family of sets, depending on the distribution of the eigenvalues of A. The sets $S(\tau)$ have the following interpretation: $S(\tau)$ is the subcontinuum of $[\lambda_{\min}, \lambda_{\max}]$ where the optimal polynomial of degree $[\tau N]$ is uniformly small.

From Corollary 1.4.8 and (3.1.1) we find the improved approximation

$$\frac{\|r_n^{CG}\|_{A^{-1}}}{\|r_0^{CG}\|_{A^{-1}}} \lessapprox \rho_t^n \tag{3.1.2}$$

with

$$\rho_t = \exp\left(-\frac{1}{t}\int_0^t g_{S(\tau)}(0) d\tau\right) \tag{3.1.3}$$

depending on n, since $t = n/N$. As the sets $S(\tau)$ are decreasing as τ increase, their Green functions $g_{S(\tau)}(0)$, evaluated at 0, increase with τ. Hence the numbers ρ_t decrease with increasing n, and this explains the effect of superlinear convergence (notice that $\log(\rho_{n/N})$ equals the slope at n of the bound on a semi-logarithmic plot).

Indeed, we will only show that (3.1.2) only holds in an asymptotic sense after taking nth roots. However, in order to be able to take limits, we need to consider sequences of matrices A_N having a joint asymptotic eigenvalue distribution. Such sequences of matrices occur naturally in the context of the discretization of elliptic PDEs, by varying the stepsize or some other parameter of discretization, see Section 3.4. We have the following result [BeKu99, Theorem 2.1].

Theorem 3.1.1 *Let $(A_N)_N$ be a sequence of symmetric invertible matrices, A_N of size $N \times N$, satisfying the conditions*

(i) *There exists a compact Σ and a positive Borel measure σ such that $\Lambda(A_N) \subset \Sigma$ for all N, and $\nu_N(\Lambda(A_N)) \xrightarrow{*} \sigma$ for $N \to \infty$;*

(ii) *σ has a continuous potential;*

(iii) *$U^{\nu_N(\Lambda(A_N))}(0) \to U^{\sigma}(0)$ for $N \to \infty$.*

Define $S(t) := \operatorname{supp}(\sigma - \mu_{t,0,\sigma})$, with the extremal measure $\mu_{t,0,\sigma}$ as in Theorem 2.2.1. Then for $t \in (0, \|\sigma\|)$, we have

$$\limsup_{\substack{n,N\to\infty \\ n/N\to t}} \frac{1}{n} \log E_n(0, \Lambda(A_N)) \leq -\frac{1}{t} \int_0^t g_{S(\tau)}(0)\, d\tau \tag{3.1.4}$$

Proof. Apply Theorem 2.3.2 for $p = \infty$, $w_N = 1$, and $Q = 0$: conditions (i), and (iii), are corresponding to (2.3.1), and (2.3.5) for $z = 0$, respectively. The conditions (2.3.2) and (2.3.3) are trivially true. Then our claim follows from (2.3.6), where according to Remark 2.2.8 we may replace $w_{t,0,\sigma} - U^{\mu_{t,0,\sigma}}(0)$ by our integral formula (2.2.9). □

Remark 3.1.2 According to Theorem 2.3.2, if we have the additional separation condition (2.3.7) on the spacing of eigenvalues, then there is equality in Theorem 3.1.1. In particular, if follows from the comments after Corollary 1.4.8 that the asymptotic CG bound on the right-hand side of (3.1.4) cannot be improved. □

Remark 3.1.3 For determining σ, each λ in $\Lambda(A_N)$ is taken only once, regardless of its multiplicity. Hence it might happen that $||\sigma|| < 1$. However, Theorem 2.3.2 remains valid even if one counts multiplicities, since $E_n(\cdot,\cdot)$ becomes larger if one adds points close to multiple eigenvalues to $\Lambda(A)$. The new conditions obtained form (i)–(iii) counting multiplicities will be referred to as (i)', (ii)', and (iii)'. □

Remark 3.1.4 The condition (ii) is not very restrictive. For example, if σ is absolutely continuous with respect to Lebesgue measure with a bounded density then (ii) is satisfied. It is also satisfied if the density has only logarithmic-type or power-type singularities at a finite number of points. On the other hand, condition (ii) is not satisfied if σ has point masses.

In case of simple eigenvalues, condition (iii) may be rewritten as

$$\lim_{N\to\infty} |\det(A_N)|^{1/N} = \exp(-U^{\sigma}(0))$$

Comparing (iii) with the Principle of descent (2.1.2), we see that this condition prevents too many eigenvalues close to 0. If (iii) would not hold, then the matrices A_N are ill-conditioned and the estimate (3.1.2) may very well fail. □

Remark 3.1.5 In [BeKu00], the following strategy was considered in order to find a polynomial p_n of degree n being small on $\Lambda(A_N)$ and $p_n(0) = 1$ (and thus to find an upper bound for $E_n(0, \Lambda(A_N))$):

Choose some fixed set S. Each eigenvalue of A_N outside the set S is a zero of p_n. This determines a certain fraction of the zeros of p_n. Clearly the set S has to be sufficiently big so that the number of eigenvalues of A_N outside S is less than n. The other zeros of p_n are free and they are chosen with the aim to minimize $\|p_n\|_{L_\infty(S)}$.

Though this strategy of imitating the CG polynomial seems to be natural, it depends very much on a good choice of the set S, see also the discussion in [Gre79] and [DTT98, Section 6]. Indeed, choosing a large set S means that there only few outliers, and their influence on the supnorm $\|p_n\|_S$ is small. Since we only need a polynomial which is small on the discrete set $\Lambda(A_N)$ but not necessarily in the gaps between the eigenvalues, $\|p_n\|_{L_\infty(S)}$ may be much bigger than $\|p_n\|_{\Lambda(A_N)}$. On the other hand, choosing a small set S means that we fix a lot of zeros of p_n and so we loose a lot of freedom in our choice for minimizing $\|p_n\|_{L_\infty(S)}$.

The main result of [BeKu00] is that the above strategy cannot produce a better asymptotic bound on $E_n(0, \Lambda(A_N))$ than Theorem 3.1.1, and that (under some additional assumptions) $S = S(t)$ leads to the same bound. □

Remark 3.1.6 Provided that the sets $S(t)$ of Theorem 3.1.1 are intervals, say, $S(t) = [a(t), b(t)]$, we may give an interpretation of the bound (3.1.2) in terms of *marginal condition numbers*: Since $g_{S(t)}(0)$ is increasing in t, we get from (3.1.3) that

$$\log(\rho_t^n) = -N \int_0^{n/N} g_{S(\tau)}(0) \leq \sum_{j=0}^{n-1} g_{S(j/N)}(0),$$

and, by (2.1.3), estimate (3.1.2) can be rewritten as

$$\frac{\|e_n^{CG}\|_A}{\|e_0^{CG}\|_A} \lessapprox \prod_{j=0}^{n-1} \frac{\sqrt{b(\frac{j}{N})/a(\frac{j}{N})} - 1}{\sqrt{b(\frac{j}{N})/a(\frac{j}{N})} + 1}.$$

Hence the classical bound (1.1.1) is obtained for constant b/a, and we see that the superlinear convergence behavior is obtained if the marginal condition number $b(\frac{j}{N})/a(\frac{j}{N})$ strictly decreases. Indeed, as we will see in Section 3.2, some extremal eigenvalues will be matched by Ritz values, and can be disregarded for the further convergence behavior. □

Remark 3.1.7 For the moment it is not completely clear how to generalize Theorem 3.1.1 to the case of matrices with unbounded spectra and asymptotic eigenvalue distribution given by σ with unbounded support. In this case, we certainly have to impose some growth condition on σ around infinity such that the constraint is active around infinity. □

Let us give some examples illustrating Theorem 3.1.1.

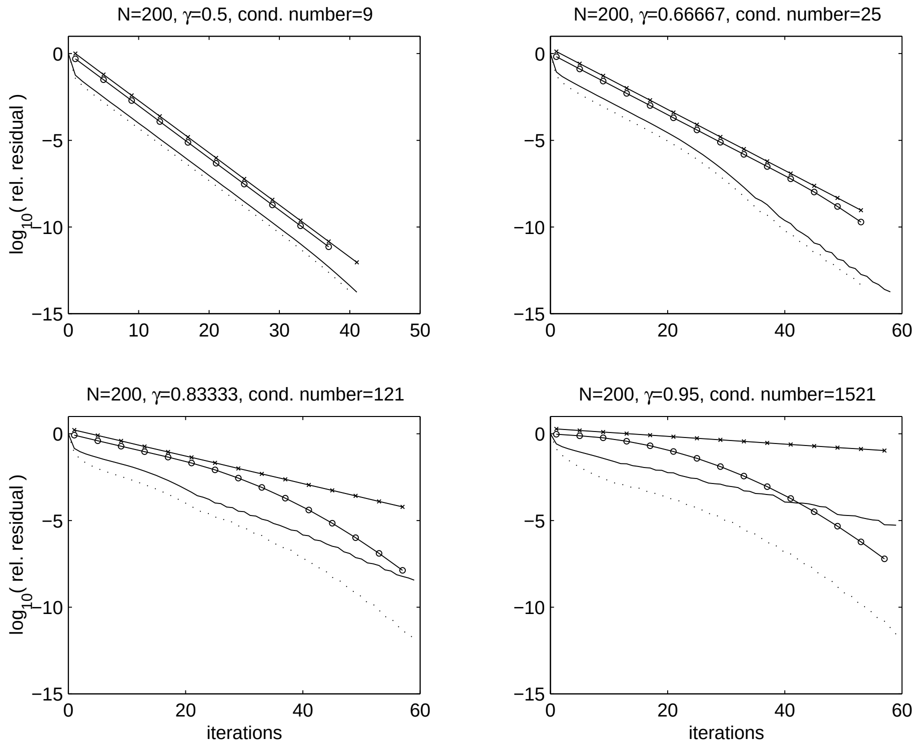

Figure 3.1: *The error curve of CG (solid line) and GMRES (dotted line) versus the classical upper bound (crosses) and our asymptotic upper bound (circles) for the system* $T_{200}x = b$*, with random solution* x*, and initial residual* $r_0 = (1, ..., 1)^T$*. Here* T_N *is the Kac, Murdock and Szegő matrix, with parameter* $\gamma \in \{1/2, 2/3, 5/6, 19/20\}$*.*

Example 3.1.8 The case of equidistant eigenvalues $\Lambda(A_N) = \{1/N, 2/N, ..., N/N\}$ leading to σ being the Lebesgue measure on $[0, 1]$ has already been discussed in Example 2.3.3, see also [BeKu99, Section 3]. For CG we obtain the error curve as well as the bounds (1.1.1), (3.1.3) as displayed in Figure 1.1, see Section 1.1. □

Example 3.1.9 Consider the "worst case" eigenvalues

$$\Lambda(A_N) = \{2 + 2\cos(\pi \frac{j}{N+1}) : j = 1, ..., N\},$$

here conditions (i)–(iii) of Theorem 3.1.1 hold with $\sigma = \omega_{[0,4]}$. Comparing with Exercice 2.2.7 we see that $S(t) = [0, 4]$ for $0 < t < 1$, and thus the bound (3.1.3) is trivial. Indeed, for this example it is known that there are starting residuals such that CG does not lead to a small residual before reaching $n \approx N$. □

Example 3.1.10 For the Toeplitz matrix $A_N := (\gamma^{|j-k|})_{j,k=1,2...,N}$, $0 < \gamma < 1$, of Kac, Murdock and Szegő [KaMuSz53, p. 783] it is shown in [BeKu99, Section 4] (see also Section 3.3 below) that conditions (i)',(ii)',(iii)' hold with

$$\sigma(x) = \frac{1}{x}\,\omega_{[a,1/a]}(x) = \int_0^1 \omega_{[a,b(t)]}(x)\,dt, \quad a = \frac{1-\gamma}{1+\gamma}, \quad b(t) = \begin{cases} 1/a & \text{for } t \le a \\ a/t^2 & \text{for } t \ge a. \end{cases}$$

Numerical experiments for the symmetric positive definite Toeplitz matrix T_{200} of order 200 of Kac, Murdock and Szegő are given in Figure 3.1. The four different plots correspond to the choices $\gamma \in \{1/2, 2/3, 5/6, 19/20\}$ of the parameter. Notice that the CG error curve (solid line) of the last two plots is clearly affected by rounding errors leading to loss of orthogonality, whereas the GMRES relative residual curves (dotted line) behave essentially like predicted by our theory. In particular, the classical bound (1.1.1) (crosses) does no longer describe correctly the size of the relative residual of GMRES for $n \geq 20$ and $\gamma \in \{5/6, 19/20\}$. Experimentally we observe that the range of superlinear convergence starts in the different examples approximately at the iteration indices ≥ 50, 30, 20, and 10, respectively. This has to be compared with the predicted quantity $N \cdot a$ which for the different choices of γ approximately takes the values 66, 40, 29, and 5, respectively. Though theses numbers differ slightly, we observe that the new bound (3.1.3) reflects quite precisely the shape of the relative residual curve, and in particular allows to detect the ranges of linear and of superlinear convergence. $\square$

Example 3.1.11 Consider the two dimensional Poisson equation

$$-\frac{\partial^2 u(x,y)}{\partial x^2} - \frac{\partial^2 u(x,y)}{\partial y^2} = f(x,y)$$

for (x, y) in the unit square $0 < x, y < 1$, with Dirichlet boundary conditions on the boundary of the square. The usual five-point finite difference approximation on the uniform grid

$$(j/(m_x+1), k/(m_y+1)), \qquad j = 0, 1, \ldots, m_x+1, k = 0, 1, \ldots, m_y+1,$$

leads to a linear system of size $N \times N$ where $N = m_x m_y$. After rescaling, the coefficient matrix of the system may be written as a sum of Kronecker products

$$A_N = \frac{(m_x+1)}{(m_y+1)} B_{m_x} \otimes I_{m_y} + \frac{(m_y+1)}{(m_x+1)} I_{m_x} \otimes B_{m_y} \tag{3.1.5}$$

where

$$B_m = \begin{bmatrix} 2 & -1 & 0 & \cdots & 0 \\ -1 & 2 & -1 & \ddots & \vdots \\ 0 & \ddots & \ddots & \ddots & 0 \\ \vdots & \ddots & \ddots & \ddots & -1 \\ 0 & \cdots & 0 & -1 & 2 \end{bmatrix}_{m \times m} \tag{3.1.6}$$

and I_m is the identity matrix of order m. It is well known and easy to verify that the eigenvalues of B_m are

$$\mu_{k,m} = 2 - 2\cos\frac{\pi k}{m+1}, \qquad k = 1, \ldots, m,$$

and that the eigenvalues $\lambda_{j,k}$ of A_N are connected with the eigenvalues of B_m via

$$\lambda_{j,k} = \frac{m_y+1}{m_x+1}\mu_{j,m_x} + \frac{m_x+1}{m_y+1}\mu_{k,m_x}, \qquad j = 1, 2, \ldots, m_x, k = 1, \ldots, m_y. \tag{3.1.7}$$

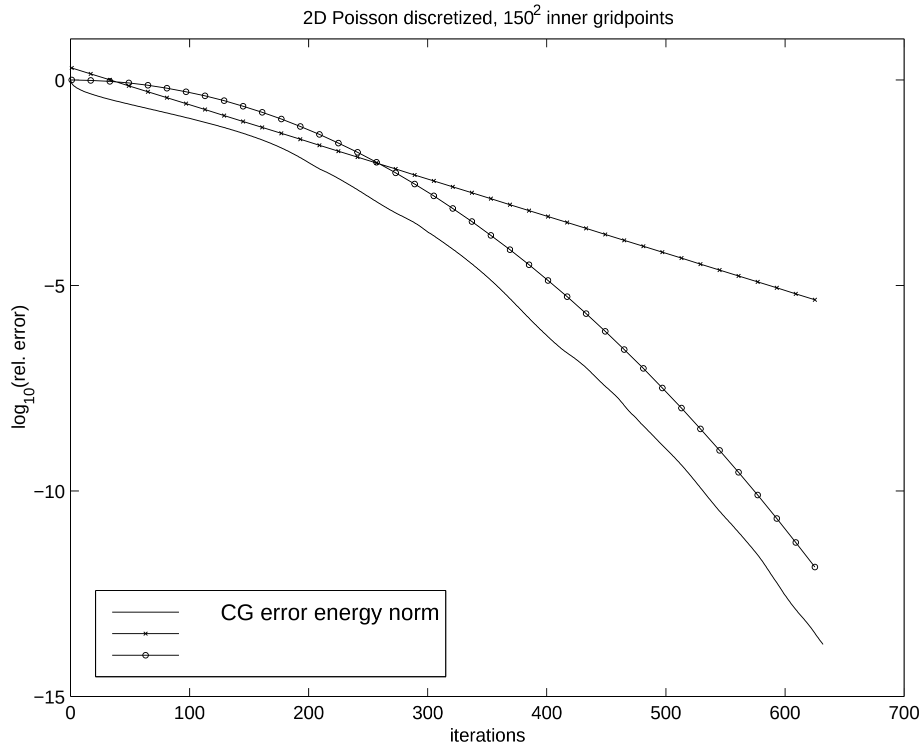

Figure 3.2: *The CG error curve versus the two upper bounds for the system* $A_N x = b$ *resulting from discretizing the 2D Poisson equation on a uniform grid with* $m_x = m_y =$ 150*. We have chosen a random solution* x*, and initial residual* $r_0 = (1, ..., 1)^T$*, and obtain superlinear convergence from the beginning. Notice that the classical upper bound for CG is far too pessimistic for larger iteration indices. For the new bound we have added a factor* $1/2$ *in front of* σ *since* $\lambda_{j,k} = \lambda_{k,j}$*, and we suspect that most of the eigenvalues are of multiplicity* 2.

We consider the limit relation

$$m_x, m_y \to \infty, \qquad \frac{m_x}{m_y} \to \delta \leq 1,$$

then it is not difficult to see using (3.1.7) that condition (i)' holds with

$$\int f\, d\sigma = \int_0^1 d\phi \int_0^1 d\psi\, f(2\delta(1-\cos(\pi\phi)) + 2\delta^{-1}(1-\cos(\pi\psi)))$$
$$= \frac{1}{\pi^2}\int_0^{4\delta} dx \int_x^{x+4\delta^{-1}} d\lambda \frac{f(\lambda)}{\sqrt{x(4\delta - x)(\lambda - x)(4\delta^{-1} - \lambda + x)}} = \int_0^{4\delta+4\delta^{-1}} \sigma'(\lambda) d\lambda$$

with

$$\sigma'(\lambda) := \frac{1}{\pi^2}\int_{\max\{0,\lambda-4\delta^{-1}\}}^{\min\{4\delta,\lambda\}} \frac{dx}{\sqrt{x(4\delta - x)(\lambda - x)(4\delta^{-1} - \lambda + x)}}.$$

The substitution $x' = 4\delta - x$ shows that $\sigma'(4\delta + 4\delta^{-1} - \lambda) = \sigma'(\lambda)$, and thus we only need to consider the case where $\lambda \geq 2\delta + 2\delta^{-1}$, and hence $\lambda \geq 4\delta$. We now construct a linear fractional transformation $y = T(x)$ with $T(0) = 0$, $T(4\delta) = 1$, $T(\lambda) = \infty$, and hence

$$T(x) = \frac{x}{\lambda - x}\frac{\lambda - 4\delta}{4\delta}, \qquad \gamma := T(\lambda - 4\delta^{-1}) = \frac{1}{16}(\lambda - 4\delta^{-1})(\lambda - 4\delta) \leq 1,$$

and the substitution $y = T(x)$ leads to

$$\sigma'(\lambda) = \frac{1}{4\pi^2} \int_{\max\{\gamma,0\}}^{1} \frac{dy}{\sqrt{y(1-y)(y-\gamma)}} = \frac{1}{\pi} \int_{0,\sin(\pi t/2)\geq\gamma}^{1} \frac{dy}{\sqrt{16\sin^2(\pi t/2) - 16\gamma}}.$$

By substituting γ, we find with $\Delta := 2\delta + 2\delta^{-1}$

$$\sigma = \int_0^1 \omega_{S(t)}\, dt, \qquad S(t) = [\Delta - \sqrt{\Delta^2 - 16\sin^2(\frac{\pi t}{2})}, \Delta + \sqrt{\Delta^2 - 16\sin^2(\frac{\pi t}{2})}],$$

and thus the extremal measures by Exercise 2.2.7. One may compare our findings for $\delta = 25/40$ with numerical experiments presented in Figure 3.11 of Section 3.5 below, where both histograms for $m_x = 25$ and $m_y = 40$ and the density function σ' of the limiting distribution are drawn. Notice that σ has the support $S(0) = [0, 4\delta + 4\delta^{-1}]$, and that σ' has logarithmic singularities at 4δ and at $4\delta^{-1}$.

The sets $S(t)$ have been known before only in the case $\delta = 1$, and thus $\Delta = 4$ [BeKu99, Section 5]. More precisely, we observe that $\lambda_{j,k} = \lambda_{k,j}$, that is, most of the eigenvalues have multiplicity at least 2. Also, $\lambda_{j,m+1-j} = 4$ for all $j = 1, \ldots, m$, and the eigenvalue 4 has multiplicity m. We suspect that $N/2 + o(N)$ eigenvalues have multiplicity 2. In this case, not only condition (i)' but also condition (i) holds, with the new constraint being the half of the old constraint. Since

$$\mu_{t,0,\sigma/2} = \frac{1}{\ }$$

$$\left[\frac{\ }{||r_{0,N}||}\right]^{1/N} \leq 1.$$

Then, for every $t \in (0, ||\sigma||)$,

$$\limsup_{\substack{n,N\to\infty \\ n/N\to t}} \frac{1}{N} \log\left(\frac{\|e_{n,N}^{CG}\|_{A_N}}{\|e_{0,N}^{CG}\|_{A_N}}\right) \leq U^{\mu_{t,Q,\sigma}}(0) - w_{t,Q,\sigma}, \tag{3.1.8}$$

where $\mu_{t,Q,\sigma}$ and $w_{t,Q,\sigma}$ are as in Theorem 2.2.1. A similar bound is valid for the nth relative residual of MINRES.

[1] In case of distinct eigenvalues, the quantity $|(r_{0,N}, v_{j,N})|$ was called before $\beta_{j,N}$.

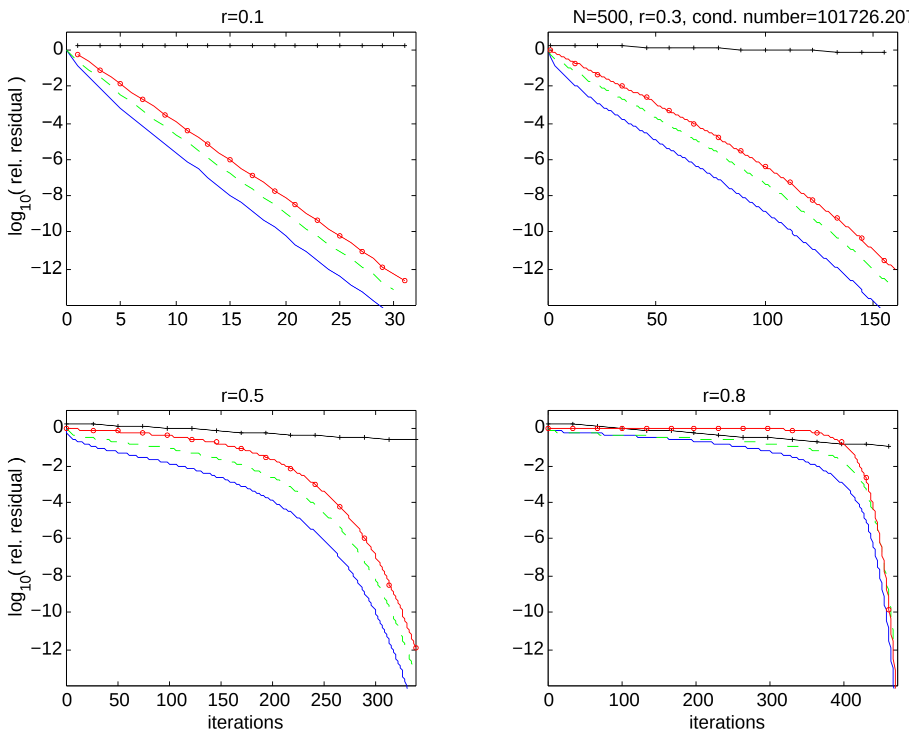

Figure 3.3: *The one dimensional Poisson problem discretized on a uniform grid* ($N = 500$) *for* $f(x) = \sum_{j=1}^{N} r^j \sin(j\pi x)$, $r = 0.1, 0.3, 0.5, 0.8$. *We find the error curve of CG (solid blue line) and the classical bound (1.1) (black line with crosses), and our new asymptotic bound (red line with circles). For comparison we give the MINRES relative residual curve (dashed green line). Notice that, for* $r = 0.8$*, there is hardly any superlinear convergence and one has to reach approximately the dimension of the system in order to achieve full precision.*

Proof. See (1.4.4) and Corollary 1.3.6, and use the fact that from condition (iii) it follows that $[\min_j |\lambda_{j,N}|]^{1/N} \to 1$. □

Example 3.1.13 As a motivating model problem for Theorem 3.1.12, we consider the one dimensional Poisson equation $-u''(x) = f(x)$, $x \in [0, 1]$, with homogeneous Dirichlet boundary conditions $u(0) = u(1) = 0$. The usual central finite difference approximation on the uniform grid $j/(N + 1)$, $j = 0, 1, \ldots, N + 1$, leads to a linear system $A_N x = b_N$ with N equations and unknowns, where $A_N = B_N$ of (3.1.6), and

$$b_N = (N + 1)^2 \cdot \left[f(1/(N + 1)) \quad f(2/(N + 1)) \quad f(3/(N + 1)) \quad \cdots \quad f(N/(N + 1)) \right]^T.$$

Both the one dimensional Poisson problem and the system $A_N x = b_N$ are easy to solve; however, this toy problem can serve to explain convergence behavior observed also in less trivial situations. From Example 3.1.9 we know that conditions (i)–(iii) are satisfied with $\sigma = \omega_{[0,4]}$, and that, for general starting residual, one obtains poor CG convergence, as being confirmed by Figure 3.3.

Here we will be interested in what happens for the CG starting vector 0 (i.e., $r_{0,N} = b_N$)

and particularly smooth functions f, namely

$$f(x) = \sum_{j=1}^{\infty} f_j \sin(\pi j x), \quad x \in [0,1], \quad \text{where} \quad r := \limsup_{j\to\infty} |f_j|^{1/j} \in (0,1).$$

It is shown in [BeKu02, Lemma 3.1] that here condition (iv) holds with

$$Q(x) = \frac{\log(1/r)}{\pi} \arccos(\frac{2-\lambda}{2}).$$

Also, the reader may verify that the assumptions of Remark 2.2.9 hold. As shown in [BeKu02, Section 3], here the integral equations (2.2.5),(2.2.6) can be solved in terms of the complete elliptic integral $K(\cdot)$ and the Jacobi elliptic functions: if $k = k(r)$ is defined by

$$\frac{\log(1/r)}{\pi} K(k) = K(\sqrt{1-k^2}),$$

then

$$a(t) = 4\mathrm{cn}^2((1-t)K(k); k), \qquad b(t) = \alpha(t)/\mathrm{dn}^2((1-t)K(k); k),$$

and we obtain the asymptotic CG error bound of (2.2.3). □

3.2 Applications to the rate of convergence of Ritz values

In order to approximate eigenvalues of large real symmetric matrices A of order N via the Lanczos method with starting vector $r_0 \in \mathbb{R}^N$, one computes the so-called Ritz values, namely, the eigenvalues $x_{1,n} < ... < x_{n,n}$ of the (tridiagonal) matrix Jacobi matrix J_n, see Definition 1.2.5 and Corollary1.3.3. Depending on the eigenvector components $\beta_1, .., \beta_N$ of the starting vector r_0, some of the eigenvalues $\lambda_1 \leq ... \leq \lambda_N$ of A are well approximated by Ritz values even if n is much smaller than the dimension N. Classical results on convergence and on technical details of the Lanczos method may be found in many textbooks. Let us cite here the well-known Kaniel-Page-Saad estimate for extremal eigenvalues [GoVL93, PPV95, Saa96, TrBa97] being a consequence of Corollary 1.4.5 and Lemma 1.4.6

$$|\frac{x_{1,n} - \lambda_1}{\lambda_N - \lambda_1}| \leq \frac{1}{T_{n-1}(1 + 2\frac{\lambda_2 - \lambda_1}{\lambda_N - \lambda_2})^2} \frac{1}{\beta_1^2} \sum_{j=2}^{n} \beta_j^2, \tag{3.2.1}$$

with T_n being the nth Chebyshev polynomial of the first kind. Thus one may expect geometric convergence of the smallest (largest) Ritz value to the smallest (largest) eigenvalue for a fixed matrix A, but the rate of convergence will depend on the size of the eigenvector component β_1, and on the (relative) distance of λ_1 to the other eigenvalues. For an "inner" eigenvalue λ_k lying in the convex hull of the Ritz values, say, $x_{\kappa-1,n} < \lambda_k \leq x_{\kappa,n}$ for some $\kappa = \kappa(k)$, we get by combining Corollary 1.4.5 and Exercice 1.4.7

$$\min_{\ell} |\lambda_k - x_{\ell,n}|^2 \leq |(\lambda_k - x_{\kappa-1,n})(\lambda_k - x_{\kappa,n})| \leq 2\, b^2 \Big[\frac{b/a - 1}{b/a + 1}\Big]^{[n/2]-1} \frac{1}{\beta_k^2} \sum_{j=1, j\neq k}^{n} \beta_j^2, \tag{3.2.2}$$

where $a = \min\{\lambda_{k+1} - \lambda_k, \lambda_k - \lambda_{k-1}\}$, and $b = \max\{\lambda_N - \lambda_k, \lambda_k - \lambda_1\}$. Notice again that we may only expect an interesting rate of convergence if λ_k is well separated from the rest of the spectrum, and if $|\beta_k|/||r_0||$ is sufficiently large.

There exist (worst case) examples A, r_0 with eigenvalue and eigenvector component distribution such that the bounds (3.2.1) or (3.2.2) are (approximately) sharp. However, for matrices occurring in applications one observes quite often that the above bounds greatly overestimate the actual error, even for a judicious choice of the set in Corollary 1.4.5 (for instance a finite union of intervals representing the parts of the real axis where the spectrum of A is relatively dense).

Trefethen and Bau [TrBa97, p. 279] observed a relationship with electric charge distributions, and claimed that the Lanczos iteration tends to converge to eigenvalues in regions of "too little charge" for an equilibrium distribution. This has been made more precise by Kuilaars [Kui00a], who considered as in the preceding section a sequence of symmetric matrices A_N of which are supposed to have an asymptotic eigenvalue distribution

$$\nu_N(\Lambda(A_N)) \xrightarrow{*} \sigma \quad \text{for } N \to \infty. \tag{3.2.3}$$

Then, following Trefethen and Bau, Kuijlaars compared $\omega_{\mathrm{supp}(\sigma)}$ and σ, and considered more precisely the constrained energy problem with external field $Q = 0$ of Section 2.2.

In the remainder of this section we will suppose that A_N has N distinct eigenvalues $\lambda_{1,N} < \ldots < \lambda_{N,N}$ contained all in some compact set Σ. Also, we suppose that the Lanczos method is applied to matrix A_N with starting vector $r_{0,N}$ having eigenvector components $\beta_{1,N}, \ldots, \beta_{N,N}$, and we are interested in measuring the distance of an eigenvalue $\lambda_{j,N}$ to the set of Ritz values $x_{1,n,N} < \ldots < x_{n,n,N}$ obtained in the nth iteration of the Lanczos process. We then have the following result

Theorem 3.2.1 *Suppose that the asymptotic distribution of the spectra of $(A_N)_N$ is given by σ, which has a continuous potential. Let k_N be a sequence of indices such that*

$$\lim_{N\to\infty} \lambda_{k_N,N} = \lambda \tag{3.2.4}$$

and suppose that

$$\lim_{N\to\infty} \frac{1}{N} \sum_{j \neq k_N} \log|\lambda_{k_N,N} - \lambda_{j,N}| = \int \log|\lambda - \lambda'| \, d\sigma(\lambda'), \tag{3.2.5}$$

$$\liminf_{N\to\infty} \Big[\frac{|\beta_{k_N,N}|}{||r_{0,N}||}\Big]^{1/N} =: \rho \in (0,1]. \tag{3.2.6}$$

Then

$$\limsup_{\substack{n,N\to\infty \\ n/N\to t}} \min_j |\lambda_{k_N,N} - x_{j,n,N}|^{1/N} \leq \frac{1}{}$$

$$\overline{\quad ||r_{0,N}|| \quad}\Big]^{1/N} \leq 1,$$

then

$$\limsup_{\substack{n,N\to\infty \\ n/N\to t}} \min_j |\lambda_{k_N,N} - x_{j,n,N}|^{1/N} \leq \frac{1}{}$$

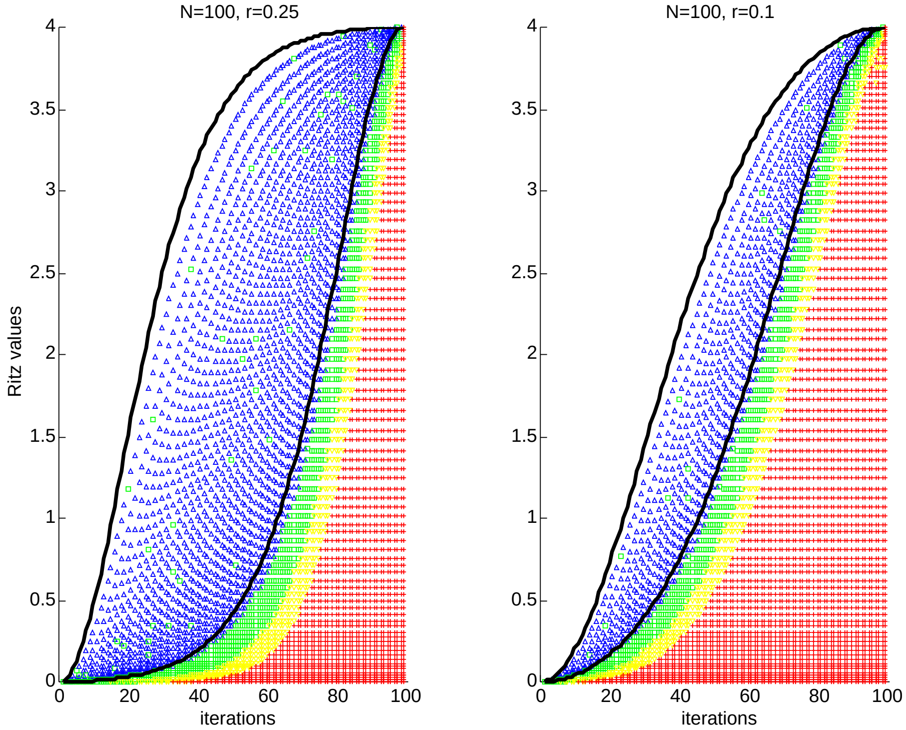

Figure 3.4: *Convergence of Ritz values for our 1D Poisson model problem with particular smooth right hand side. Here* $N = 100$, $x_{0,N} = 0$, *and* $f(x) = \sum_{j=1}^{N} r^j \sin(j\pi x)$, $r \in \{1/4, 1/10\}$. *The two black curves indicate the graphs of* a, b. *We draw in the* n*th column,* $1 \leq n \leq N$, *the position of the* n*th Ritz values within the interval* $[0, 4]$. *Here the color/symbol indicates the distance of the Ritz value to the set of eigenvalues of* A_N, *compare with (*). Notice that nearly all Ritz values are red in the range* $[0, a(t)]$, $t = n/N$, *that there are no Ritz values in* $[b(t), 4]$, *and that in the range* $[a(t), b(t)]$ *hardly any Ritz value converged (up to some exceptions by "accident").*

Proof. We may suppose without loss of generality that $\lambda = 0$. Using the estimate of Corollary 1.4.5, we apply Theorem 2.3.2 for $p = \infty$ to the sets $E_N := \{\lambda_{j,N} - \lambda_{k_N,N} : j \neq k_N\}$. $\square$

Remark 3.2.2 Recall from Remark 2.2.8 that we may replace $w_{t,0,\sigma} - U^{\mu_{t,0,\sigma}}(\lambda)$ in (3.2.7) some mean of the Green functions $g_{S(t)}(\lambda)$, see (2.2.9). Hence, for $\rho = 1$, the right-hand side of (3.2.7) is strictly negative for $\lambda \notin \bigcap_{\tau<t} \operatorname{supp}(\sigma - \mu_{\tau,0,\sigma})$, in correspondence with the heuristic observation of Trefethen and Bau. $\square$

Remark 3.2.3 In case of (3.2.8) one may observe a further phenomenon which for the data of Example 3.1.13 is shown in Figure 3.4 and which is not fully covered by Theorem 3.2.1: all eigenvalues in $\operatorname{supp}(\mu_{t,Q,\sigma}) \setminus \operatorname{supp}(\sigma - \mu_{t,Q,\sigma})$ are fit by Ritz values (the constraint is active), and there is hardly any Ritz value in $\operatorname{supp}(\sigma - \mu_{t,Q,\sigma}) \setminus \operatorname{supp}(\mu_{t,Q,\sigma})$. $\square$

Remark 3.2.4 The first condition (3.2.5) will be true if the eigenvalues $\lambda_{k_N+j,N}$ for $j \neq 0$ do not approach "too fast" $\lambda_{k_N,N}$ when $N \to \infty$. This separation condition has

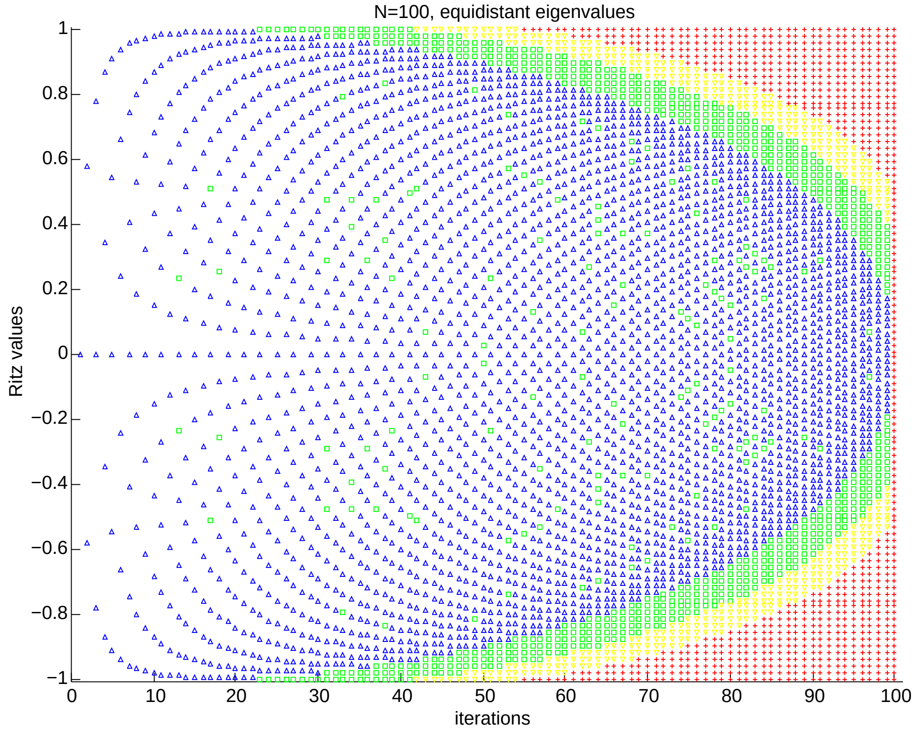

Figure 3.5: *Convergence of Ritz values for* 400 *equidistant eigenvalues in* $[-1, 1]$. *We draw in the nth column,* $1 \leq n \leq N$, *the position of the nth Ritz values within the interval* $[0, 4]$. *Here the color/symbol indicates the distance of the Ritz value to the set of eigenvalues of* A_N, *compare with (*).*

been suggested by Dragnev and Saff [DrSa97, Definition 3.1] as a sufficient condition for insuring nth root asymptotics for discrete orthogonal polynomials. It holds if the distance $|\lambda_{k_N+j,N} - \lambda_{k_N,N}|$ is bounded below by a constant times $|j|/N$ [Rak96] or by some positive power of this quantity; however, it is excluded that two neighboring eigenvalues approach exponentially. Condition (3.2.6) means that the starting vector $r_{0,N}$ has a sufficiently large eigencomponent for the eigenvalue $\lambda_{k_N,N}$. □

Remark 3.2.5 The statement of (3.2.7) can be found in [Be00b, Theorem 2.1(a)]. Before, Kuijlaars [Kui00a] had established a related inequality with $\rho = 1$, and the right-hand side of (3.2.7) being replaced by its square root. As assumption, Kuijlaars imposed that (3.2.5) and (3.2.6) for $\rho = 1$ hold for any set of indices verifying (3.2.4). □

Remark 3.2.6 According to the first part of Corollary 1.4.5, we learn from the proof of Theorem 3.2.1 that the right-hand side of (3.2.7) can be replaced by its square for extremal eigenvalues $k_N = 1$ or $k_N = N$.

Indeed, this is also true for more general situations: suppose that $\Sigma = [A, B]$, and $B' \in \Sigma$ such that $w_{t,0,\sigma} - U^{\mu_{t,0,\sigma}}(\lambda) > 0$ for $\lambda \in [A, B']$. Furthermore suppose that that (3.2.5) and (3.2.6) for $\rho = 1$ hold for any set of indices verifying (3.2.4) and limit $\lambda \in [A, B']$. Then it is not difficult to show that $|\lambda_{k,N} - \lambda_{k+1,N}|^{1/N} \to 1$ for eigenvalues in $[A, B']$. Taking into account (3.2.7) and the separation property of Theorem 1.4.1(c), a little bit

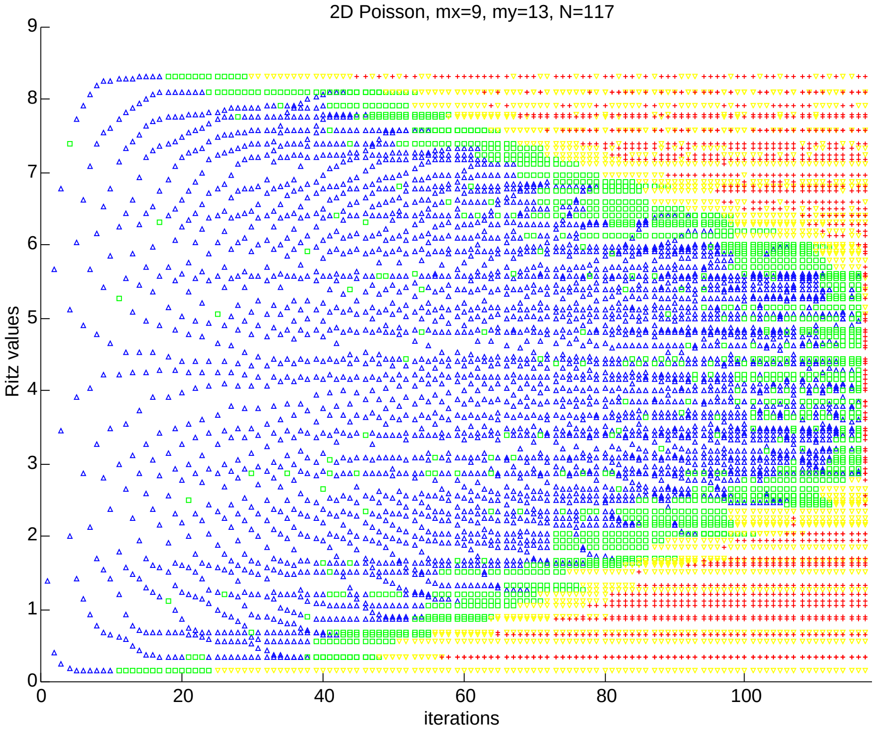

Figure 3.6: *Convergence of Ritz values for the Poisson problem with $m_x = 9$ and $m_y = 13$. Notice that, even for large $n \approx N$, hardly any eigenvalue in $S(1) = [4\delta, 4\delta^{-1}] = [2.77, 5.78]$ is found by Ritz values.*

of combinatorics shows that $x_{j,n,N}$ approaches $\lambda_{j,N}$ exponentially for sufficiently small j. This implies that one of the factors on the left-hand side of the second estimate of Corollary 1.4.5 can be dropped, and we have that

$$\begin{aligned}\limsup_{\substack{n,N\to\infty\\ n/N\to t}} \min_j |\lambda_{k_N,N} - x_{j,n,N}|^{1/N} &= \limsup_{\substack{n,N\to\infty\\ n/N\to t}} |\lambda_{k_N,N} - x_{k_N,n,N}|^{1/N} \\ &\le \exp\Big(2U^{\mu_{t,0,\sigma}}(\lambda) - 2w_{t,0,\sigma}\Big)\end{aligned}$$

provided that $\lambda \in [A, B']$.

In the same spirit, one can show that for all but at most one exceptional eigenvalue in any closed sub-interval of Σ with strictly positive $w_{t,0,\sigma} - U^{\mu_{t,0,\sigma}}$ we have this improved rate of convergence. Finally, there are examples where the rate of convergence for the exceptional eigenvalue is given by (3.2.7). A detailed discussion of these exceptional indices is given in [Be00b]. $\square$

Example 3.2.7 If A_N has equidistant eigenvalues in $[-1, 1]$, we found in Example 3.1.8 that $S(t) = [-\sqrt{1-t^2}, \sqrt{1-t^2}]$. Indeed, as shown in Figure 3.5, the eigenvalues outside the disk are found by the Lanczos method.

The numerical results displayed in Figure 3.5 as well as in subsequent experiments have been obtained by the Lanczos method with full reorthogonalization, in order to prevent

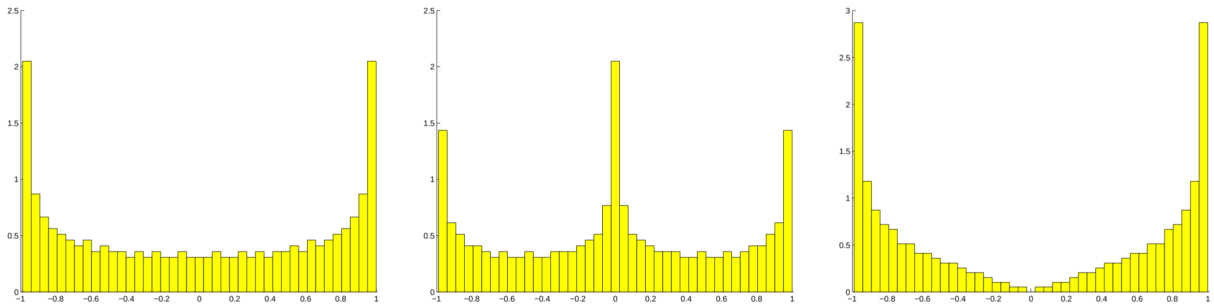

Figure 3.7: *Bar chart for the eigenvalue distribution of* 400 *eigenvalues in the case* $\alpha = 1$ *(equilibrium distribution),* $\alpha = 2$*, and* $\alpha = 1/2$ *(from the left to the right).*

from loss of orthogonality, due to finite precision arithmetic. The following symbols/colors are used to indicate the distance of a given Ritz value to the set of eigenvalues

Color	Symbol	distance between Ritz value and set of eigenvalues
Red	$+$	less than $0.5\,10^{-14}$
Yellow	$\bigtriangledown$	between $0.5\,10^{-14}$ and $0.5\,10^{-8}$
Green	$\square$	between $0.5\,10^{-8}$ and $0.5\,10^{-3}$
Blue	$\bigtriangleup$	larger than $0.5\,10^{-3}$

$(*)$

$\square$

Example 3.2.8 For the Poisson problem of Exercise 3.1.11 and $m_x = 9$, $m_y = 13$, the Ritz values are displayed in Figure 3.6. As in the preceding example, the color/symbol is chosen depending on the distance of the Ritz value to the set of eigenvalues, as a function of the iteration index $n = 1, 2, ..., N = 9 * 13 = 117$. Observe that eigenvalues outside of the set $S(n/N)$ described in Example 3.1.11 are well approximated by Ritz values, but not those in $S(n/N)$. Notice also that, even for large $n \approx N$, hardly any eigenvalue in $S(1) = [4\delta, 4\delta^{-1}] = [2.77, 5.78]$ is well approximated by Ritz values. $\square$

Example 3.2.9 Consider the eigenvalues

$$\lambda_{j,N} = \cos(\pi\frac{2j-1}{2N}) \cdot |\cos(\pi\frac{2j-1}{2N})|^{\alpha-1}, \quad \alpha > 0,$$

and eigencomponents $\beta_{j,N} = 1$, having clearly an asymptotic eigenvalue distribution σ with continuous potential. As mentioned already in Example 3.1.9, no convergence of Ritz values can be expected if $\alpha = 1$. Things become more interesting for $\alpha = 2$ (more eigenvalues close to zero) or for $\alpha = 1/2$ (more eigenvalues close to the endpoints ± 1). This behavior is displayed in Figure 3.7.

In the case $\alpha = 2$ one may show that for $0 < t \leq 1/\sqrt{2}$ we obtain $S(t) = [-1, 1]$, and there is no geometric convergence of Ritz values. For $t \in (1/\sqrt{2}, 1)$, the sets are strictly decreasing and of the form $S(t) = [-b(t), b(t)]$, but the resulting formulas for $b(t)$ are complicated, we omit details. The convergence behavior of the corresponding Ritz values can be found in Figure 3.8, indeed, for $n \leq N/\sqrt{2}$ hardly any eigenvalue is well approximated by a Ritz value.

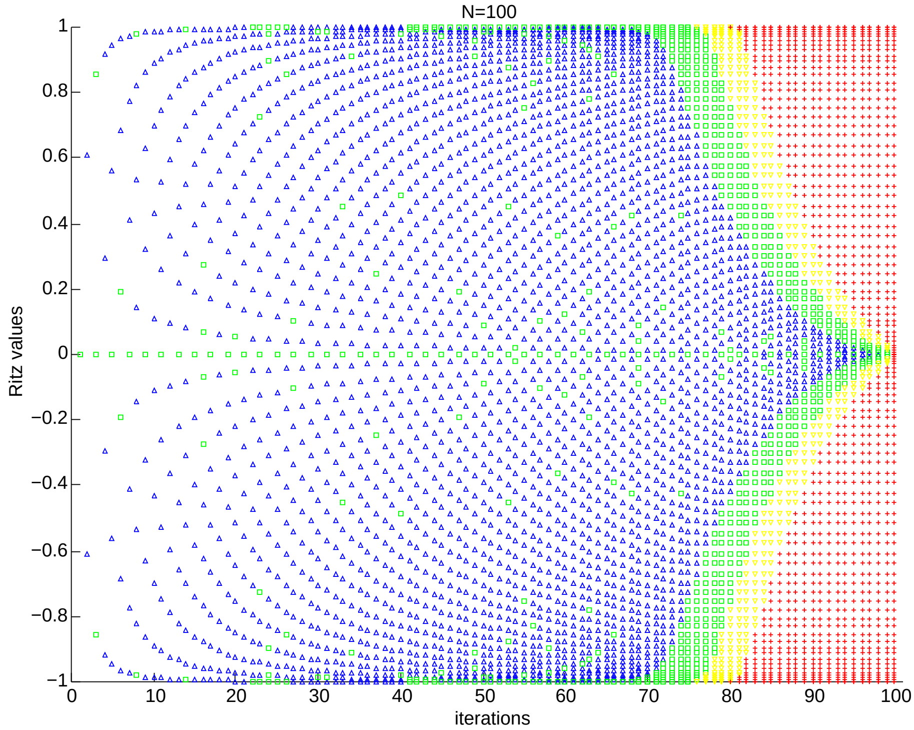

Figure 3.8: *Convergence of Ritz values for "squares" of* 100 *Chebyshev eigenvalues in* $[-1, 1]$ *(*$\alpha = 2$*).*

For the case $\alpha = 1/2$ it is shown in [Be00b] that

$$S(t) = [-1, -r(t)] \cup [r(t), 1], \qquad r(t) = \frac{1 - \cos(\pi t/2)}{1 + \cos(\pi t/2)}.$$

Notice that the eigenvalues (and the eigenvector components) are symmetric with respect to the origin. Thus $p_{2n-1,N}$ is odd, and $p_{2n,N}$ is even. Moreover, $\lambda_{N+1,2N+1} = x_{n+1,2n+1,2N+1} = 0$, and thus here is a perfect rate of convergence. However, the eigenvalue $\lambda_{N+1,2N+1}$ is approached by the Ritz values $x_{n+1,2n,2N+1} = -x_{n,2n,2N+1}$. Comparing with Remark 3.2.6 we get an "exceptional" eigenvalue with a smaller rate of convergence. In contrast, for even N no exceptional eigenvalue occurs, even if the Ritz value $x_{n+1,2n+1,2N} = 0$ is not close to the spectrum. This last example contradicts the widely believed fact that first extremal eigenvalues are found by Ritz values. □

3.3 Circulants, Toeplitz matrices and their cousins

A circulant matrix of order N generated by some exponential polynomial $\phi(\theta) := \phi_0 + \phi_1 e^{i\theta} + ... + \phi_{N-1} e^{(N-1)i\theta}$ are defined by

$$C_N(\phi) = \begin{bmatrix} \phi_0 & \phi_1 & \cdots & \phi_{N-1} \\ \phi_{N-1} & \phi_0 & \cdots & \phi_{N-2} \\ \vdots & \vdots & & \vdots \\ \phi_1 & \phi_2 & \cdots & \phi_0 \end{bmatrix}, \tag{3.3.1}$$

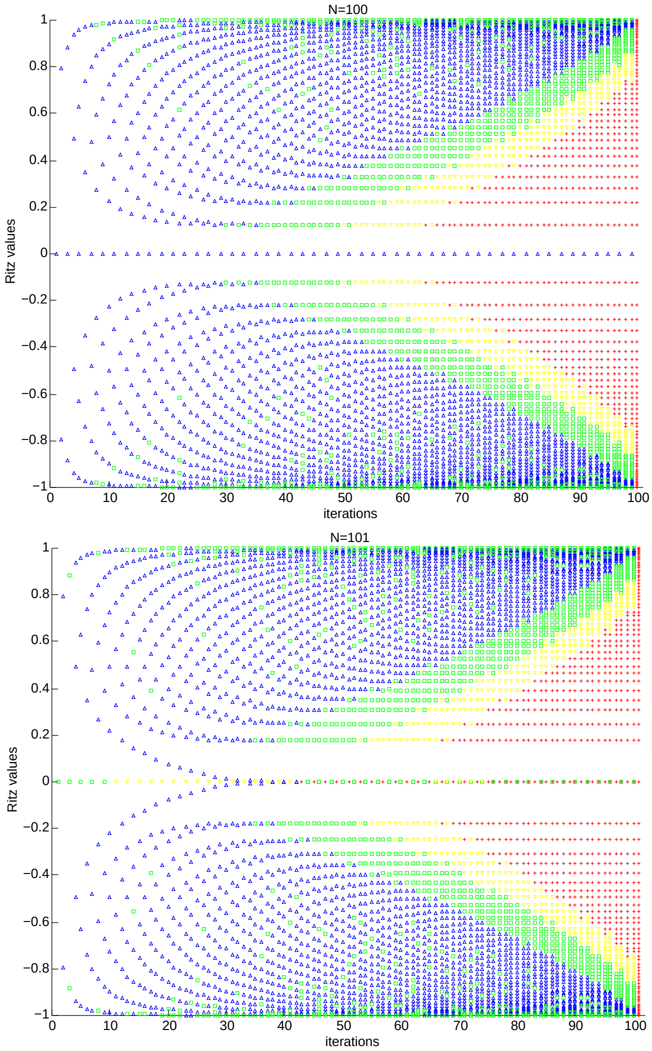

Figure 3.9: *Convergence of Ritz values for "square roots" of $N \in \{100, 101\}$ Chebyshev eigenvalues in $[-1, 1]$ ($\alpha = 1/2$). In the top plot ($N = 100$), the Ritz values $x_{j+1,2j+1,N} = 0$ for odd n are not close to the spectrum. In the bottom plot ($N = 101$) one observes the phenomena of exceptional eigenvalues.*

i.e., $C_N(\phi)$ is constant along diagonals. It is easily seen that $C_N(\phi)$ is diagonalized by the unitary FFT matrix of eigenvectors

$$\Omega_N = \frac{1}{\sqrt{N}} \Big[\exp(\frac{2\pi i j k}{N}) \Big]_{j,k=0,1,...,N-1},$$

with corresponding eigenvalues given by $\phi(\frac{2\pi i(k-1)}{N})$, $k = 1, ..., N$. Notice also that $C_N(\phi)$ is normal, and in addition hermitian if and only if all eigenvalues are real. One easily checks using the explicit knowledge of the eigenvalues that, if $\phi^{(N)}$ is the partial sum of a exponential power series ϕ being absolutely convergent (such symbols ϕ are called of Wiener class), then

$$\nu_N(C_N(\phi^{(N)})) \xrightarrow{*} \sigma_\phi \quad \text{for } N \to \infty, \text{ where} \qquad \int f \, d\sigma_\phi = \frac{1}{2\pi} \int_0^{2\pi} f(\phi(e^{is})) \, ds, \qquad (3.3.2)$$

where here and in what follows we count eigenvalues according to their multiplicities. We speak of a circulant matrix of level two (and by iteration of level t) or a circulant-circulant matrix of order $m_x m_y$ if there is a block structure with m_x^2 blocks as in (3.3.1), with each block being itself a circulant matrix. Thus such matrices are induced by a bi-variate exponential polynomial ϕ with degree in x being equal to $m_x - 1$ and degree in y being equal to $m_y - 1$, and we write $C_{m_x,m_y}(\phi)$. With the Kronecker product

$$A \otimes B = (A_{j,k}B)_{j,k},$$

we see that the matrix of eigenvectors is given by the unitary matrix $\Omega_{m_x} \otimes \Omega_{m_y}$ and the eigenvalues by the expressions $\phi(\exp(\frac{2\pi i(j-1)}{m_x}), \exp(\frac{2\pi i(k-1)}{m_y}))$, $j = 1, ..., m_x$, $k = 1, ..., m_y$. Thus, as in (3.3.2), if $\phi^{(m_x,m_y)}$ is the truncation of a bivariate exponential power series ϕ being absolutely convergent, then for $m_x \to \infty$, $m_y \to \infty$,

$$\nu_{m_x m_y}(C_{m_x,m_y}(\phi^{(m_x,m_y)})) \xrightarrow{*} \sigma_\phi, \qquad \int f \, d\sigma_\phi = \frac{1}{(2\pi)^2} \iint_{[0,2\pi]^2} f(\phi(\theta)) \, d\theta. \tag{3.3.3}$$

Toeplitz matrices are generated by Fourier series

$$T_N(\phi) = \begin{bmatrix} \phi_0 & \phi_1 & \cdots & \phi_{N-1} \\ \phi_{-1} & \phi_0 & \cdots & \phi_{N-2} \\ \vdots & \vdots & & \vdots \\ \phi_{1-N} & \phi_{2_N} & \cdots & \phi_0 \end{bmatrix},, \qquad \phi(\theta) = \sum_{j=-\infty}^{\infty} \phi_j e^{ij\theta} \tag{3.3.4}$$

being again constant along diagonals, and hermitian if ϕ is real-valued. Hence any circulant is Toeplitz, but not reciprocally. We also define Toeplitz-Toeplitz matrices (or level 2 Toeplitz matrices) $T_{m_x,m_y}(\phi)$ induced by some bivariate Fourier series ϕ as a matrix with Toeplitz block structure, each individual block being also of Toeplitz structure. Similarly, we speak of Toeplitz-circulant matrices (Toeplitz block structure with circulant blocks) or Circulant-Toeplitz matrices.

It is well-known that these matrices occur in the discretization by finite differences using the five point stencil of the Poisson PDE on $[0,1]^2$, more precisely we have a (banded) Toeplitz-Toeplitz matrix in case of Dirichlet boundary conditions (compare with Example 3.1.11), and a (banded) circulant-circulant matrix in case of homogeneous Neumann boundary conditions. Toeplitz systems arise also in a variety of other applications, such as signal processing and time series analysis, see [ChNg96] and the references cited therein.

For Toeplitz matrices and their level 2 counterparts, it is in general impossible to give explicit formulas for eigenvalues. However, we may find formulas for the asymptotic eigenvalue distribution, compare for instance with [GrSz84, pp. 63-65], [BoSi99, Theorem 5.10 and Corollary 5.11].

Theorem 3.3.1 *Let ϕ be a univariate absolutely convergent Fourier series (we say that ϕ is of Wiener class) and real-valued. Then $\nu_N(T_N(\phi)) \to \sigma_\phi$ for $N \to \infty$, with σ_ϕ as in (3.3.2).*

If ϕ be a bi-variate absolutely convergent and real-valued Fourier series then for $m_x, m_y \to \infty$ we have that $\nu_{m_x m_y}(T_{m_x,m_y}(\phi)) \to \sigma_\phi$, with σ_ϕ as in (3.3.3).

In the proof of Theorem 3.3.1 we will require the following perturbation result of Tyrtyshnikov [Tyr96] in a form given by Serra Capizzano [Ser02, Proposition 2.3]. See also Tilli [Til98].

Theorem 3.3.2 ([Ser02, Proposition 2.3]) *Let (A_N) be a sequence of Hermitian matrices where A_N has size $N \times N$. Suppose for every $\epsilon > 0$ there exists N_ϵ such that for every $N \geq N_\epsilon$ a splitting*

$$A_N = B_N(\epsilon) + R_N(\epsilon) + \Delta_N(\epsilon)$$

where $B_N(\epsilon)$, $R_N(\epsilon)$ and $\Delta_N(\epsilon)$ are Hermitian matrices so that for $N \geq N_\epsilon$,

$$\text{rank} R_N(\epsilon) \leq C_1(\epsilon) N, \qquad \textit{and} \qquad \|\Delta_N(\epsilon)\| \leq C_2(\epsilon)$$

where $C_1(\epsilon)$ and $C_2(\epsilon)$ are positive constants independent of N with

$$\lim_{\epsilon \to 0} C_1(\epsilon) = \lim_{\epsilon \to 0} C_2(\epsilon) = 0.$$

Suppose that, for every $\epsilon > 0$, the limit $\nu_N(\Lambda(B_N(\epsilon))) \xrightarrow{*} \sigma_\epsilon$ for $N \to \infty$ exists, and that the limit $\sigma_\epsilon \xrightarrow{*} \sigma$ for $\epsilon \to 0$ exists, then $\nu_N(\Lambda(A_N)) \xrightarrow{*} \sigma$ for $N \to \infty$.

Proof. Apply the Courant minimax principle and the Theorem of Bauer and Fike [GoVL93] telling us that if A, B are two hermitian matrices with "small" $||A - B||$, then for each eigenvalue of A there exists an eigenvalue of B which is "close". $\square$

Proof of Theorem 3.3.1. We will give the main idea of proof for the case of a Toeplitz matrix, the arguments for a level 2 Toeplitz matrix are similar. Denote by $\phi^{(N)}$ the Fourier sum obtained from ϕ by taking the $[N^{1/3}]$th partial sum. Since for the p-matrix norm $|| \cdot ||_p$ of a matrix we have

$$(||A||_2)^2 \leq ||A||_\infty \, ||A||_1,$$

and since ϕ is of Wiener class, we find that, for $\epsilon > 0$ and sufficiently large N, we have

$$||\phi - \phi^{(N)}||_{L_\infty([-\pi,\pi])} < \epsilon, \qquad ||T_N(\phi) - T_N(\phi^{(N)})|| < \epsilon.$$

Then $T_N(\phi^{(N)})$ is banded, of bandwidth $\leq N^{1/3}$, and we need to modify at most $N^{2/3}$ entries in order to transform it into a hermitian circular matrix B_N. Since the eigenvalues of this hermitian circular matrix are explicitly known, we obtain the assumptions Theorem 3.3.2 with $\sigma_\epsilon = \sigma_\phi$, and our claim follows from Theorem 3.3.2. $\square$

As seen from the above proof, results similar to Theorem 3.3.1 are true for Toeplitz-circulant or Circulant-Toeplitz matrices.

By Theorem 3.3.1 we see that condition (i)' of Section 3.1 holds for sequences of hermitian (level 2) Toeplitz matrices. Also, condition (ii)' will be true for instance for continuous symbols. Let us shortly comment of condition (iii)' for the case of hermitian positive definite Toeplitz matrices (and hence $\phi \geq 0$). A result of Szegő (see [GrSz84, p. 44 and p. 66]) is that

$$\lim_{N \to \infty} \frac{\det(T_N(\phi))}{\det(T_{N-1}(\phi))} = \exp\left(\frac{1}{2\pi}\int_{-\pi}^{\pi} \log \phi(\theta)\, d\theta\right) \tag{3.3.5}$$

provided that ϕ satisfies the Szegő condition

$$\int_{-\pi}^{\pi} \log \phi(\theta)\, d\theta > -\infty.$$

Notice that this condition can be rewritten as $U^{\sigma_\phi}(0) < +\infty$. Also recall the link to strong asymptotics of orthogonal polynomials on the unit circle (the ratio of determinants in (3.3.5) is linked to the leading coefficient of such orthonormal polynomials).

It follows from (3.3.5) that

$$\lim_{N\to\infty} \log(|\det |T_N(\phi)|^{1/N}) = \frac{1}{2\pi}\int_{-\pi}^{\pi} \log\phi(\theta)\, d\theta = \int \log\lambda\, d\sigma(\lambda) \in \mathbb{R},$$

and the condition (iii') is satisfied.

3.4 Discretization of elliptic PDE's

The asymptotic eigenvalue distribution of matrices obtained by a finite difference discretization of elliptic partial differential equations has been discussed in detail by Serra-Cappizano [Ser03]. Here we will not look for greatest generality, but just have a look on the particular example of a 2D diffusion equation on some polyhedral domain in $\mathbb{R}^2$, discretized by the classical five point stencil.

Let $\Omega \subset [0,1]^2$ be some open polyhedron, and $b : \Omega \to [0, +\infty)$ piecewise continuous. We solve the diffusion problem

$$\operatorname{div}(b\nabla u) = f \quad \text{on } \Omega$$

plus Dirichlet (Neumann) boundary conditions via central finite differences, with stepsizes

$$\Delta x = \frac{1}{m_x + 1}, \qquad \Delta y = \frac{1}{m_y + 1},$$

This leads to a system of linear equations for the unknowns $u_{j,k} \approx u(j\Delta x, k\Delta y)$ with

$$(j,k) \in \{(j,k) \in \mathbb{Z}^2 \mid (j\Delta x, k\Delta y) \in \Omega\},$$

given by

$$\begin{aligned}
&\frac{m_x+1}{m_y+1}\Big[-b_{j-1/2,k}u_{j-1,k} - b_{j+1/2,k}u_{j+1,k} + (b_{j-1/2,k} + b_{j+1/2,k})u_{j,k}\Big] \\
&+\frac{m_x+1}{m_y+1}\Big[-b_{j,k-1/2}u_{j,k-1} - b_{j,k+1/2}u_{j,k+1} + (b_{j,k-1/2} + b_{j,k+1/2})u_{j,k}\Big] = \frac{f_{j,k}}{(m_x+1)(m_y+1)}
\end{aligned}$$

where $b_{j-1/2,k} = c((j-1/2)h, kh)$, etc. Supposing that there are and N gridpoints in Ω, we can write this system as $A_N x = b_N$, where it is known that (at least for b strictly positive) $\operatorname{cond}(A_N)$ grows like $\mathcal{O}(N)$. Notice that for $b = 1$ and $\Omega = (0,1)^2$ and Dirichlet boundary conditions we recover the Toeplitz-Toeplitz matrix of Example 3.1.11. In what follows we will not specify further the (discretisation of the) boundary conditions, since in all cases this only leads to a small rank pertubation of order $\mathcal{O}(\sqrt{N})$, and hence by Theorem 3.3.2 does not affect the asymptotic eigenvalue distribution.

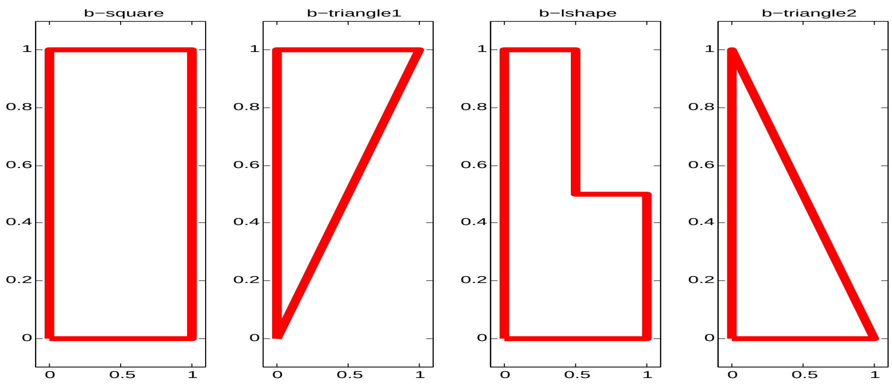

Figure 3.10: *The four different domains, referred to as b-square, b-triangle-1, b-lshape and b-triangle2 (from the left to the right).*

As in Example 3.1.11 we consider the limit relation

$$m_x, m_y \to \infty, \qquad \frac{m_x}{m_y} \to \delta < 1,$$

Again, for $b = 1$ and $\Omega = (0,1)^2$, the asymptotic eigenvalue distribution has been determined in Example 3.1.11, compare also with Theorem 3.3.1 for the symbol

$$\phi(s_1, s_2) = 2\delta(1 - \cos(s_1)) + 2\delta^{-1}(1 - \cos(s_2)). \tag{3.4.1}$$

In the general case we find the following result being a consequence of a more general result of Serra-Cappizano [Ser03].

Theorem 3.4.1 *Under the above assumptions on Ω and b, we have that, for any continuous function f with compact support,*

$$\lim_{N\to\infty} \int f \, d\nu_N(A_N) = \frac{1}{m(\Omega)} \int_\Omega dx \frac{1}{(2\pi)^2} \iint_{[0,2\pi]^2} ds\, f(b(x) \cdot \phi(s))$$

with ϕ as in (3.4.1) and $m(\cdot)$ denoting the two-dimensional Lebesgue measure.

Proof. By covering Ω by "small" squares $S_{j,N}$ of equal size tending to zero for $N \to \infty$, we may replace A_N by some block diagonal matrix C_N where entries with row/column index corresponding to points in squares $S_{j,N}$ (for the row) and $S_{k,N}$ (for the column) will be replaced by zero if $S_{j,N} \neq S_{k,N}$ or if $S_{j,N} \cup S_{k,N}$ is not a subset of Ω or if b is not continuous on $S_{j,N} \cup S_{k,N}$. By choosing a correct size of the square, we see that the rank of $A_N - C_N$ is $o(N)$. Denote by B_N the matrix obtained from C_N by replacing all b–values of a diagonal block corresponding to the square $S_{j,N}$ by some constant $b(\xi_{j,N})$ with $\xi_{j,N} \in S_{j,N}$. Then $||B_N - C_N||$ is small by continuity. Applying Theorem 3.3.1 for each square and summing up all squares we then find that

$$\lim_{N\to\infty} \int f \, d\nu_N(B_N) = \frac{1}{m(\Omega)} \int_\Omega dx \frac{1}{(2\pi)^2} \iint_{[0,2\pi]^2} ds\, f(b(x) \cdot \phi(s)),$$

and our claim follows from Theorem 3.3.2. $\square$

Remark 3.4.2 It is interesting to observe that, for $b = 1$, we find the same asymptotic eigenvalue distribution as in Example 3.1.11 independently on the domain Ω. $\square$

Remark 3.4.3 Let $M := \sup_\Omega b$, then from Theorem 3.4.1 it becomes clear that the asymptotic eigenvalue distribution is described by some measure σ with $\mathrm{supp}\,(\sigma) = [0, (4\delta + 4\delta^{-1})M] =: \Sigma$. One may also prove that all eigenvalues of A_N lie in Σ. Finally, defining the measure τ by

$$\tau((-\infty, r]) := \frac{m(\{x \in \Omega : b(x) \leq r\})}{m(\Omega)},$$

with support given by the essential range of b, and denoting the extremal measure of Example 3.1.11 by σ_0, we find that σ is obtained by taking the Mellin convolution of τ and σ_0. More precisely, if τ, σ_0 have densities τ', σ_0' then also σ has a density σ', given by

$$\sigma'(y) = \int_{y/M}^{4\delta+4\delta^{-1}} \sigma_0'(x)\tau'(\frac{y}{x})\frac{dx}{x}.$$

If $m := \inf_\Omega b > 0$, we deduce that

$$\sigma'(y) = \frac{1}{4\pi\, m(\Omega)} \int_\Omega \frac{dx}{b(x)} + y\frac{\delta + \delta^{-1}}{32\pi\, m(\Omega)} \int_\Omega \frac{dx}{b(x)^2} + \mathcal{O}(y^2)_{y\to 0}.$$

It is interesting to compare this formula with the Weyl formula for the asymptotic distribution of eigenvalues of the corresponding differential operator. $\square$

Example 3.4.4 We consider the Poisson problem, i.e., $b = 1$, on four different domains $\Omega \subset [0, 1]^2$ displayed in Figure 3.10. On the bottom of Figure 3.11 one may find histograms for the eigenvalue distribution in the case $m_x = 15$, $m_y = 40$, and hence $\delta = 15/40 = 0.375$. In blue we have drawn the density of the asymptotic eigenvalue distribution, which according to Remark 3.4.2 is the same for the four domains. In the upper part of Figure 3.11 we find the convergence history of CG for random starting vector. In all four cases, the actual CG convergence looks quite similar (notice the different scales for the iteration index, since the number of unknowns differ depending on how many grid points are lying in Ω. In all cases we find that the classical and our new asymptotic bound lie above the actual CG error curve, the latter describing quite well the slope of the convergence curve.

Here and also in the next example we proceeded as follows to compute numerically the asymptotic convergence bound (3.1.3): first the quite complicated density of the asymptotic eigenvalue distribution σ of Theorem 3.4.1 was replaced around both endpoints by the first two nontrivial terms in the Taylor expansion around the endpoints (see the red curves). For this new constraint $\widetilde{\sigma}$, we expect that $S(t)$ is an interval. The endpoints of this interval were obtained by solving numerically the corresponding system of integral equations. One finds as yellow curve the density of the extremal measure $\mu_{t,0,\widetilde{\sigma}}$, with t being the ratio of the last iteration index, divided by the number of unknowns. $\square$

Example 3.4.5 As second example we consider the diffusion problem with $b(x, y) = 1 + y$ on the same four different domains $\Omega \subset [0, 1]^2$ displayed in Figure 3.10. On the bottom of Figure 3.12 one may find histograms for the eigenvalue distribution in the case $m_x = m_y = 40$, and hence $\delta = 1$. In blue we have drawn the density of the asymptotic eigenvalue distribution, which has a shape depending on the domain, especially in a neighborhood of the right end point of supp (σ). In the upper part of Figure 3.12 we find the convergence history of CG for random starting vector. Notice that the final iteration index divided by the number of unknowns is 0.138, 0.186, 0.158, 0.198, and thus depends on the domain. Again our new asymptotic bound lie above the actual CG error curve, and describes well its slope. □

3.5 Conclusions

We have seen that there is a fruitful relationship between convergence behavior of Krylov subspace methods in Numerical Linear Algebra and logarithmic potential theory, the link being given by asymptotics of discrete orthogonal polynomials. Thus, in a certain sense, this manuscript contains the next 2-4 steps of the nice introduction paper [DTT98] of Driscoll, Toh and Trefethen entitled *From potential theory to matrix iteration in six steps*.

The linear algebra theory described here can be found in much more details in the textbooks [Fi96, GoVL93, Gr97, Nev93, Saa96, TrBa97], see also the original articles on superlinear convergence [AxLi86a, AxLi86b, Gre79, Mor97, NaEn00, PPV95, SlvS96, vSvV86, Win80]. The potential theoretic tools are from [Rak96, DrSa97, BuRa99, DaSa98, KuDr99, KuVA99, Be00a, Kui00b, KuMc00], see also the textbooks [MaFi04, SaTo97, Ran95, NiSo88, La72]. Finally, the link between these two domains is described in [Kui00a, BeKu99, BeKu00, BeKu02, Be00b].

There are at least two directions of current research: first it would be nice to have a similar theory as in Section 3.4 for finite element discretization of elliptic PDEs, including techniques of grid refinements. For P1 elements, some work of Serra-Capizzano and the author is in progress. What is so attracting about the finite element method is that one proceeds by projection, and therefore there are inequalities between the eigenvalues of A_N and of the continuous differential operator. Thus for instance the Weyl formula should tell us much about superlinear convergence for CG.

In order to make CG perform better, one uses in practice the technique of preconditionning. A quite involved research project is to find asymptotic eigenvalue distributions for such preconditioned matrices. For instance, there is a whole theory about how one should precondition Toeplitz matrices with help of circulant matrices, see for instance [ChNg96]. However, for level 2 Toeplitz matrices the theory is much less developed. What may happen is that there is a clustering of many eigenvalues around some point, and we should zoom this clustering in order to obtain more precise information abound the eigenvalue distribution.

A different popular class of precondition techniques include the incomplete Choleski factorization and its relaxed generalizations, see for instance [Saa96] or the original articles [Ax72, Ch91, ChEl89, El86, MeVdV77, vdV89]. In [Ch91, ChEl89], the (complicated) triangular matrices in the incomplete Choleski factorization were replaced by Circulants, which made it possible to make a more precise analysis for the Poisson problem with periodic boundary conditions. For the model problem of Section 3.4, the asymptotic eigenvalue distribution is determined in some work in progress of Kuijlaars and the author.

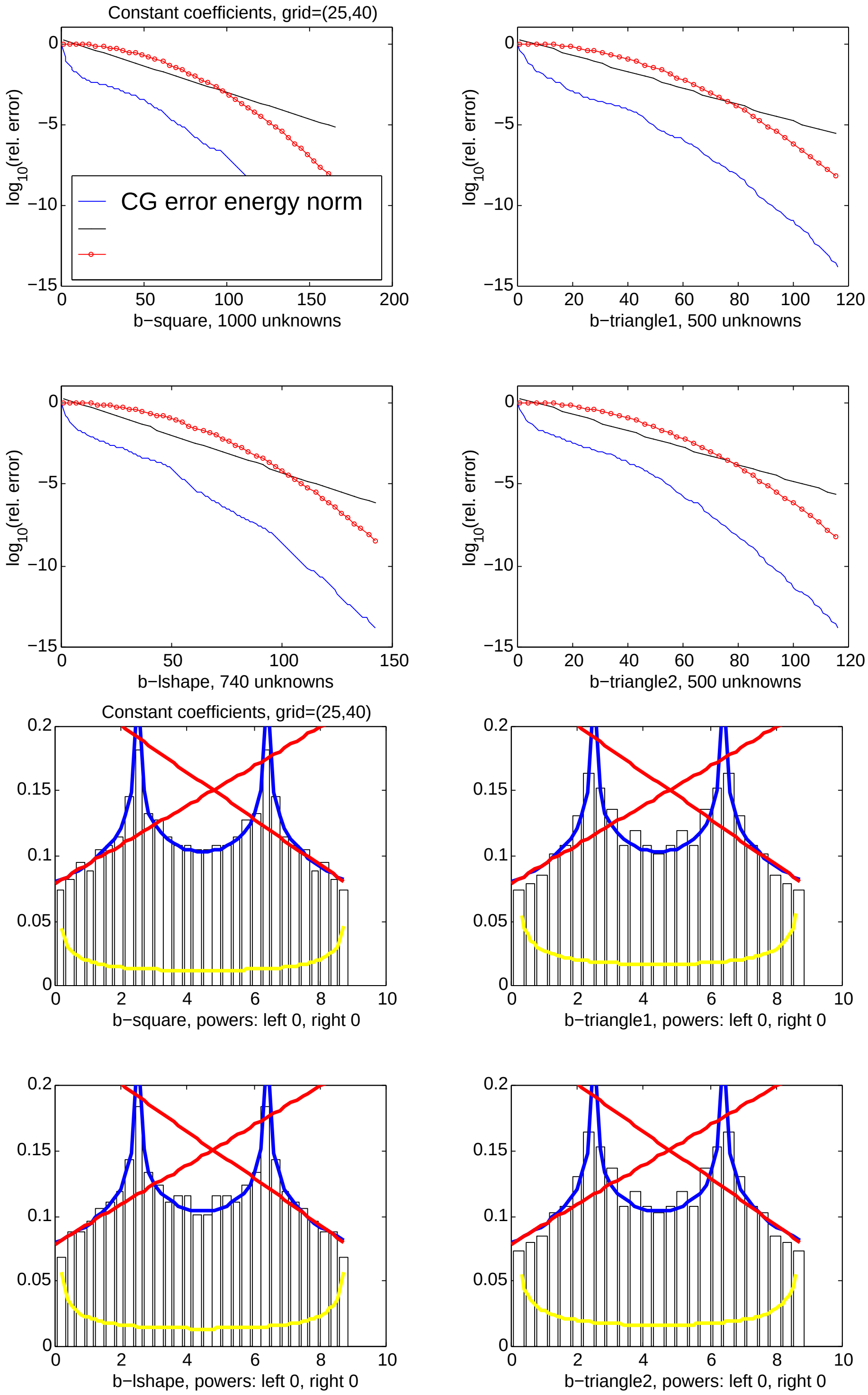

Figure 3.11: *Convergence rate for the diffusion problem with* $b = 1$ *on four different domains, the b-square, b-triangle-1, b-lshape and b-triangle2. We have chosen* $m_x = 25$ *and* $m_y = 40$*. .*

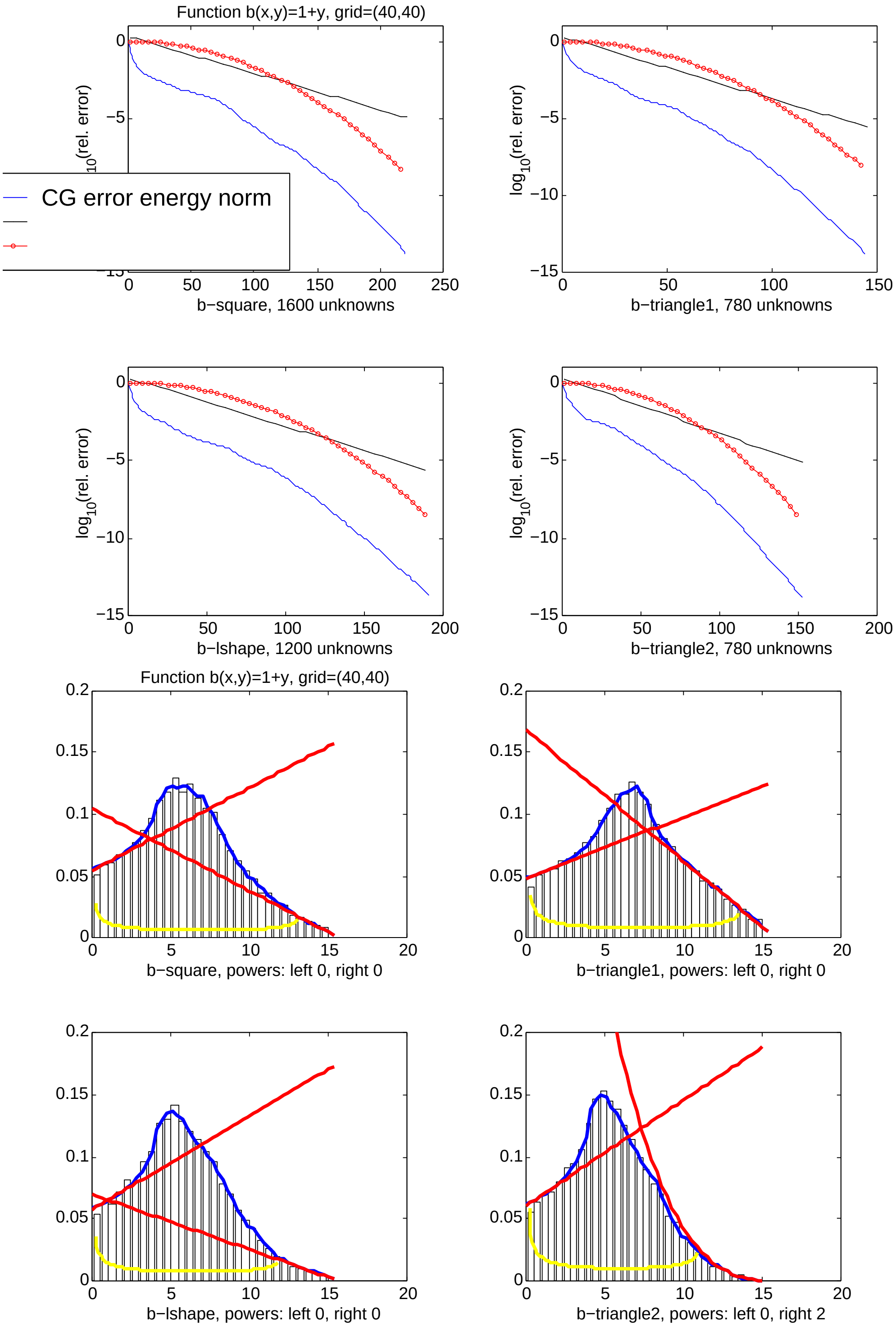

Figure 3.12: *Convergence rate for the diffusion problem with* $b(x, y) = 1 + y$ *on four different domains, the b-square, b-triangle-1, b-lshape and b-triangle2. We have chosen* $m_x = m_y = 40$.

Bibliography

[Akh90] N.I. Akhieser, *Elements of the Theory of Elliptic Functions*, Transl. of Math. Monographs **79**, AMS, Providence RI (1990).

[Ax72] O. Axelsson, A generalized SSOR method, BIT 13 (1972), 443-467.

[AxLi86a] O. Axelsson and G. Lindskog, On the eigenvalue distribution of a class of preconditioning methods, Numer. Math. 48 (1986) 479-498.

[AxLi86b] O. Axelsson and G. Lindskog, On the rate of convergence of the preconditioned conjugate gradient method, Numer. Math. 48 (1986) 499-523.

[Be00a] B. Beckermann, *On a conjecture of E.A. Rakhmanov*, Constr. Approx. 16 (2000), no. 3, 427–448.

[Be00b] B. Beckermann, *A note on the convergence of Ritz values for sequences of matrices*, Publication ANO 408, Université de Lille (2000).

[Be01] B. Beckermann, Complex Jacobi matrices, *J. Comput. Appl. Math.* **127** (2001) 17-65.

[Be05] B. Beckermann, *Image numérique, GMRES et polynômes de Faber.* Manuscript (2005).

[BeCr05] B. Beckermann, M. Crouzeix, *A lenticular version of a von Neumann inequality.* Manuscript (2005).

[BGT04] B. Beckermann, S.A. Goreinov, E.E. Tyrtyshnikov, Some remarks on the Elman estimate for GMRES. Manuscript (2004).

[BeKu99] B. Beckermann and A.B.J. Kuijlaars, *Superlinear Convergence of Conjugate Gradients*, *SIAM J. Num. Anal.* **39** (2001) 300-329.

[BeKu00] B. Beckermann and A.B.J. Kuijlaars, *On the sharpness of an asymptotic error estimate for Conjugate Gradients*, *BIT* **41** (2001), 856–867.

[BeKu02] B. Beckermann and A.B.J. Kuijlaars, *Superlinear CG convergence for special right-hand sides*, *Electr. Trans. Num. Anal.* **14** (2002) 1-19.

[BeSC05] B. Beckermann, S. Serra Capizzano, *On the asymptotic spectrum of Finite Element matrix sequences.* Manuscript (2005).

[BoSi99] A. Böttcher and B. Silbermann, *Introduction to large truncated Toeplitz matrices*, Universitext, Springer Verlag, New York, 1999.

[Bre72] C. Brezinski, Biorthogonality and its Applications to Numerical Analysis, Marcel Dekker, New York (1992).

[BuRa99] V.S. Buyarov and E.A. Rakhmanov, *Families of equilibrium measures with external field on the real axis*, Sb. Math. 190 (1999), pp. 791–802.

[ChEl89] T.F. Chan and H.C. Elman, Fourier analysis of iterative methods for elliptic problems, SIAM Review 31 (1989) 20-49.

[Ch91] T.F. Chan, Fourier analysis of relaxed incomplete factorization preconditioners, SIAM J. Sci. Stat. Comput. 12 (1991) 668-680.

[ChNg96] R.H. Chan and M.K. Ng, *Conjugate Gradient Methods for Toeplitz systems*, SIAM Rev. 38 (1996), pp. 427–482.

[Chi78] T. S. Chihara, *An Introduction to Orthogonal Polynomials*, Gordon and Breach, New York, 1978.

[Cro95] M. Crouzeix, Numerical range, holomorphic calculus and applications. Manuscript (2005).

[DaSa98] S.B. Damelin & E.B. Saff, Asymptotics of Weighted Polynomials on Varying Sets, manuscript (1998).

[DeMc98] P. Deift & K.T-R. McLaughlin, *A continuum limit of the Toda lattice*, Mem. Amer. Math. Soc. 131/624 (1998).

[DrSa97] P.D. Dragnev and E.B. Saff, *Constrained energy problems with applications to orthogonal polynomials of a discrete variable*, J. d'Analyse Math. 72 (1997), pp. 223–259.

[DTT98] T.A. Driscoll, K.-C. Toh and L.N. Trefethen, *From potential theory to matrix iteration in six steps*, SIAM Rev. 40 (1998), pp. 547–578.

[El86] H.C. Elman, A Stability Analysis of Incomplete LU Factorizations, Math. Comp. 47 (1986) 191-217.

[Fi96] B. Fischer, *Polynomial Based Iteration Methods for Symmetric Linear Systems*, Wiley, Teubner, Stuttgart (1996).

[GoMe97] G. Golub, G. Meurant, Matrices, moments and quadratures II; how to compute the norm of the error in iterative methods, *BIT* **37** (1997) 687-705.

[GoSt94] G.H. Golub, Z. Strakos, Estimates in Quadratic Formulas, *Numer. Algo.* **8** (1994) 241-268.

[GoVL93] G.H. Golub, C.F. Van Loan, Matrix Computations, Second Edition, Johns Hopkins University Press, Baltimore, London (1993).

[Gre79] A. Greenbaum, *Comparisions of splittings used with the conjugate gradient algorithm*, Numer. Math. 33 (1979), pp. 181–194.

[Gr97] A. Greenbaum, *Iterative Methods for Solving Linear Systems*, SIAM, Philadelphia PA, 1997.

[GrTr94] A. Greenbaum and L.N. Trefethen, *GMRES/CR and Arnoldi/Lanczos as matrix approximation problems*, SIAM J. Sci. Comp. 15 (1994), 359–368.

[GrSz84] U. Grenander and G. Szegő, *Toeplitz forms and their applications*, 2nd ed., Chelsea, New York, 1984.

[Gus78] I. Gustafsson, A class of first order factorizations, BIT 18 (1978), 142-156.

[HeSt94] M. R. Hestenes and E. L. Stiefel, Methods of conjugate gradients for solving linear systems, *J. Res. Nat. Bur. Standards* **49** (1952) 409-436.

[Joh00] K. Johansson, *Shape fluctuations and random matrices*, Comm. Math. Phys. 209 (2000), 437–476.

[KaMuSz53] M. Kac, W. Murdock, and G. Szegö, *On the eigenvalues of certain Hermitian forms*, J. Ration. Mech. Anal. 2 (1953), pp. 767–800.

[Kui00a] A.B.J. Kuijlaars, *Which eigenvalues are found by the Lanczos method?* SIAM J. Matrix Anal. Appl. 22 (2000), 306–321.

[Kui00b] A.B.J. Kuijlaars, *On the finite gap ansatz in the continuum limit of Toda lattice.* Duke Math. J. 104 (2000), 433–462.

[KuDr99] A.B.J. Kuijlaars and P.D. Dragnev, *Equilibrium problems associated with fast decreasing polynomials*, Proc. Amer. Math. Soc. 127 (1999), pp. 1065–1074.

[KuMc00] A.B.J. Kuijlaars & K.T-R. McLaughlin, *Long time behavior of the continuum limit of the Toda lattice, and the generation of infinitely many gaps from $\mathcal{C}^\infty$ initial data*, Manuscript (2000).

[KuRa98] A.B.J. Kuijlaars and E.A. Rakhmanov, *Zero distributions for discrete orthogonal polynomials*, J. Comp. Appl. Math. 99 (1998), pp. 255–274.

[KuRa99] A.B.J. Kuijlaars and E.A. Rakhmanov, *Zero distributions for discrete orthogonal polynomials (Corrigendum)*, J. Comp. Appl. Math. 104 (1999), p. 213.

[KuVA99] A.B.J. Kuijlaars and W. Van Assche, *Extremal polynomials on discrete sets*, Proc. London Math. Soc. 79 (1999), pp. 191–221.

[La72] N.S. Landkof, *Foundations of Modern Potential Theory.* Grundlehren der Mathematischen Wissenschaften, Vol 190, Springer Verlag, New York, 1972.

[MaFi04] A. Martinez-Finkelstein, Introduction to Potential Theory in the complex plane, Proceedings of this summerschool (2004).

[MeVdV77] J.A. Meijerink and H.A. van der Vorst, An Iterative Solution Method for Linear Systems of Which the Coefficient Matrix is a Symmetric M-Matrix, Math. Comp. 31 (1977) 148-162.

[Meu92] G. Meurant, A review on the inverse of tridiagonal and block tridiagonal symmetric matrices. *SIAM J. of Matrix Anal. and Appl* **13** (1992) 707-728.

[Meu98] G. Meurant, The computation of bounds for the norm of the error in the conjugate gradient algorithm, *Numer. Algo.* **16** (1998) 77-87.

[Meu92] G. Meurant, Computer solution of large linear systems, *Studies in Mathematics and Its Applications, Vol. 28*, North-Holland (1999).

[Mor97] I. Moret, *A note on the superlinear convergence of GMRES*, SIAM J. Numer. Anal. 34 (1997), pp. 513–516.

[NaEn00] A.E. Naiman, S. Engelberg, *A note on conjugate gradient convergence - Part II*, Numer. Math. 85 (2000), pp. 665-683.

[Nev93] O. Nevanlinna, *Convergence of Iterations for Linear Equations*, Birkhäuser, Basel, 1993.

[NiSo88] E.M. Nikishin & V.N. Sorokin, Rational Approximations and Orthogonality, Nauka, Moscow, 1988. Engl. transl. in *Translations of Mathematical Monographs* **92**, Am. Math. Soc., Providence, R.I. (1991).

[PPV95] C. Paige, B. Parlett, and H. van der Vorst, Approximate solutions and eigenvalue bounds from Krylov subspaces, *Num. Lin. Alg. Applics.* **2** (1995) 115-135.

[Ran95] T. Ransford, *Potential theory in the complex plane*, Cambridge University Press, Cambridge, 1995.

[Rak96] E.A. Rakhmanov, *Equilibrium measure and the distribution of zeros of the extremal polynomials of a discrete variable*, Sb. Math. 187 (1996), pp. 1213–1228.

[Saa96] Y. Saad, *Iterative methods for sparse linear systems*, PWS Publishing, Boston, MA, 1996.

[SaTo97] E.B. Saff and V. Totik, *Logarithmic potentials with external fields*, Springer, Berlin, 1997.

[Ser02] S. Serra Capizzano, Distribution results on the algebra generated by Toeplitz sequences: a finite-dimensional approach. Linear Algebra Appl. 328 (2001), 121–130.

[Ser03] S. Serra Capizzano, Generalised locally Toeplitz sequences: spectral analysis and applications to discretized partial differential equations, *Linear Algebra Appl.* **366** (2003) 371-402.

[SlvS96] G.L.G. Sleijpen and A. van der Sluis, *Further results on the convergence behavior of conjugate-gradients and Ritz values*, Linear Algebra Appl. 246 (1996), 233-278.

[vSvV86] A. van der Sluis and H.A. van der Vorst, *The rate of convergence of conjugate gradients*, Numer. Math. 48 (1986), pp. 543–560.

[StTo92] H. Stahl and V. Totik, *General Orthogonal Polynomials*, Cambridge University Press, Cambride, 1992.

[StTi02] Z. Strakos and P. Tichy, On Error Estimation in the Conjugate Gradient Method and Why It Works In Finite Precision Computations, *ETNA* **13** (2002) 56-80.

[Str01] Z. Strakos, Theory of Convergence and Effects of Finite Precision Arithmetic in Krylov Subspace Methods (D. Sc. Thesis), AS CR, 104+156p., Prague (2001).

[Til98] P. Tilli, Locally Toeplitz sequences: spectral properties and applications, Linear Algebra Appl. 278 (1998), 91–120.

[TrBa97] L.N. Trefethen and D. Bau III, *Numerical Linear Algebra*, SIAM, Philadelphia PA, 1997.

[Tyr96] E.E. Tyrtyshnikov, A unifying approach to some old and new theorems on distribution and clustering, Linear Algebra Appl. 232 (1996) 1–43.

[vdV89] H.A. van der Vorst, High performance preconditioning, SIAM J. Sci. Stat. Comput. 10 (1989) 1174-1185.

[Wal73] H. S. Wall, Analytic Theory of Continued Fractions, Chelsea, Bronx NY (1973).

[Win80] R. Winther, *Some superlinear convergence results for the conjugate gradient method*, SIAM J. Numer. Anal. 17 (1980), pp. 14–17.

Index