On the sharpness of an asymptotic error estimate for Conjugate Gradients

B. Beckermann and A.B.J. Kuijlaars[1]

Laboratoire d'Analyse Numérique et d'Optimisation, UFR IEEA - M3, UST Lille, F-59655 Villeneuve d'Ascq CEDEX, France, email: bbecker@ano.univ-lille1.fr

Department of Mathematics, Katholieke Universiteit Leuven, Celestijnenlaan 200B, 3001 Leuven, email: arno@wis.kuleuven.ac.be

Abstract

Recently, the authors obtained an upper bound on the error for the conjugate gradient method, which is valid in an asymptotic setting as the size of the linear systems tends to infinity. The estimate depends on the asymptotic distribution of eigenvalues, and the ratio between the size and the number of iterations. Such error bounds are related to the existence of polynomials with value 1 at 0 whose supnorm on the spectrum is as small as possible. A possible strategy for constructing such a polynomial p is to select a set S, to specify that every eigenvalue outside S is a zero of p, and then to minimize the supnorm of p on S. Here we show that this strategy can never give a better asymptotic upper bound than the one we obtained before. We also discuss the case where equality is met.

AMS subject classification: 65F10, 65E05, 31A99, 41A10.
Key words: Conjugate gradient method, logarithmic potentials, constrained equilibrium problem.

1 Introduction

The Conjugate Gradient (CG) method is widely used for solving large symmetric positive definite linear systems $Ax = b$. Error bounds for CG depend on the spectrum $\Lambda(A)$ of A. The error e_n after n steps of CG iteration satisfies (see e.g. [11, 13, 23])

$$\frac{\|e_n\|_A}{\|e_0\|_A} \leq E_n(\Lambda(A)), \tag{1.1}$$

where, for any closed set $S \subset \mathbb{R}$, the quantity $E_n(S)$ is defined by

$$E_n(S) = \min_{p \in \mathcal{P}_n} \max_{\lambda \in S} |p(\lambda)|, \tag{1.2}$$

and $\|x\|_A = \sqrt{x^T A x}$ is the A-norm. In (1.2) the minimum is taken over the class $\mathcal{P}_n$ of polynomials p of degree at most n with $p(0) = 1$. The bound (1.1) is sharp in the sense that for given A and n, there exists a starting vector such that equality holds, [12].

If the spectrum $\Lambda(A)$ is known then the right-hand side of (1.1) can be computed (at least numerically). Usually, however, one does not have that precise information. Depending on what one knows about $\Lambda(A)$ one may derive from (1.1) more or less accurate information about the error.

[1]This research is supported in part by FWO research project G.0278.97 and by a research grant of the Fund for Scientific Research – Flanders.

For example, if one only knows that all the eigenvalues belong to some interval $[a, b]$, then the best one can say is that

$$E_n(\Lambda(A)) \subset E_n([a, b]). \tag{1.3}$$

Indeed, it is known that there is a set of eigenvalues $\Lambda(A) \subset [a, b]$ such that equality holds in (1.3). The quantity $E_n([a, b])$ can be evaluated exactly with the use of Chebyshev polynomials of the first kind, see for example [13]. It leads to the well-known estimate

$$E_n([a, b]) \leq 2 \left(\frac{\sqrt{b} - \sqrt{a}}{\sqrt{b} - \sqrt{a}} \right)^n . \tag{1.4}$$

More generally, if one knows that all eigenvalues are contained in a set S, then

$$E_n(\Lambda(A)) \leq E_n(S). \tag{1.5}$$

Since the limit

$$\lim_{n \to \infty} \frac{1}{n} \log E_n(S) = -g_S(0) \tag{1.6}$$

exists, (the limit is equal to the negative of the Green function $g_S(z)$ associated with S, evaluated at $z = 0$), the inequality (1.5) leads to the following approximate inequality valid for large n,

$$\frac{1}{n} \log E_n(\Lambda(A)) \lessapprox - g_S(0), \tag{1.7}$$

provided $\Lambda(A) \subset S$. See [10, 11, 17] for more information on the relation between Green functions from potential theory and error estimates in matrix iteration.

Having other additional information about the eigenvalues one may improve on the often crude upper bounds (1.3)–(1.4) and (1.7). Indeed, there is an important number of publications where the authors discuss and estimate the quantity $E_n(\Lambda(A))$, see for instance [10, 12] and the references therein. A well-known example is the case of a few outliers in the spectrum. To obtain an upper bound in the case of l outliers, $E_n(\Lambda(A))$ is estimated by constructing a polynomial $p \in \mathcal{P}_n$ that is the product of a polynomial of degree l having a zero at each of the outliers, and a suitably scaled and shifted Chebyshev polynomial of degree $n - l$, see e.g. [13, p.53].

What other knowledge about the spectrum of A could we have? In a recent paper [3] the authors have taken the point of view that one might know the asymptotic distribution of eigenvalues. By this we mean that the eigenvalues are distributed according to some known measure σ. For example, the eigenvalues could be (more or less) equally spaced on some interval $[a, b]$, in which case σ would be the uniform Lebesgue measure on $[a, b]$ properly scaled to have total mass one.

To formulate this notion precisely, it is necessary to consider a sequence of symmetric positive definite systems $A_N x = b_N$ with increasing dimensions N, instead of a single linear system. Such a sequence of systems might arise from discretizations of elliptic PDEs, with N related to the mesh size of the discretization. In such cases one might not know the exact location of the individual eigenvalues, but the asymptotic eigenvalue distribution, as defined below, might be known, see e.g. the recent works of Serra and Tilly [20, 21, 22]. Also in the context of symmetric Toeplitz systems the asymptotic distribution of eigenvalues is known, see e.g. [5, 24].

Definition 1.1 The Borel measure σ on $[0, \infty)$ is called the *asymptotic eigenvalue distribution* of the sequence of symmetric positive definite matrices (A_N) if for all $b > a > 0$ with $\sigma(\{a\}) = \sigma(\{b\}) = 0$, we have

$$\lim_{N\to\infty} \frac{\#(\Lambda(A_N) \cap [a,b])}{N} = \sigma([a,b]), \tag{1.8}$$

where in the left-hand side of (1.8) each eigenvalue of A_N is counted once, irrespective of its multiplicity.

The measure σ has total mass $\|\sigma\| \leq 1$. Since we do not count the eigenvalues according to their multiplicities, the measure σ need not be a probability measure and the total mass could be strictly less than one.

An equivalent way to express that σ is the asymptotic eigenvalue distribution of the matrices A_N is that

$$\lim_{N\to\infty} \frac{1}{N} \sum_{\lambda\in\Lambda(A_N)} f(\lambda) = \int f\, d\sigma \tag{1.9}$$

holds, for every bounded continuous function f on $[0, \infty)$. A third equivalence is to say that the measures σ_N defined by

$$\sigma_N = \frac{1}{N} \sum_{\lambda\in\Lambda(A_N)} \delta_\lambda \tag{1.10}$$

with δ_λ the Dirac point mass at λ, converge in weak sense to σ.

The problem we considered in [3] was the following.

Problem 1.2 Let a measure σ on $[0, \infty)$ be given and a number $t \in (0, \|\sigma\|)$. Let (A_N) be a sequence of symmetric positive definite matrices with σ as asymptotic eigenvalue distribution. The problem is to give a sharp upper bound for $E_n(\Lambda(A_N))$ for n and N large, where $n = n_N$ depends on N in such a way that

$$\lim_{N\to\infty} \frac{n_N}{N} = t. \tag{1.11}$$

Under additional assumptions on σ and the sequence (A_N), see Section 2 below, the asymptotic upper bound for $E_n(\Lambda(A_N))$ found in [3] was

$$\limsup_{\substack{n,N\to\infty \\ n/N\to t}} \frac{1}{n} \log E_n(\Lambda(A_N)) \leq -\frac{1}{t} \int_0^t g_{S(\tau)}(0)\, d\tau, \tag{1.12}$$

for some decreasing family of sets $S(t)$, $0 < t < \|\sigma\|$. The estimate (1.12) is optimal if the eigenvalues of the matrices A_N are well-separated in a certain precise sense, see [3, Theorem 2.2]. Like the basic estimate (1.7), the estimate (1.12) uses Green functions evaluated at 0. But now the right-hand side of (1.12) is an average of such Green functions evaluated at 0. The sets $S(t)$ arise from a constrained energy problem in logarithmic potential theory studied first by Rakhmanov [18] in the context of zero distributions of discrete orthogonal polynomials. See Section 2 for a more precise description.

It is the aim of this paper to approach Problem 1.2 from a slightly different, perhaps more natural perspective. The main result is that the asymptotic bound (1.12) is best possible in a different sense as well.

Suppose σ, (A_N), and t are as in Problem 1.2. In order to estimate $E_n(\Lambda(A_N))$, we are guided by the well-known fact that the optimal CG polynomial has some of its zeros very close to some of the eigenvalues. This is certainly true for isolated eigenvalues, i.e., the outliers mentioned above. However, also eigenvalues in dense parts of the spectrum might be well approximated. Indeed, both numerical experiments and theoretical results [2, 14] show that this happens in a certain fixed part of the spectrum depending on σ and t if $n, N \to \infty$ in such a manner that n/N has a limit t.

We do not know in advance which eigenvalues are well approximated. However, for a given closed set $S \subset [0, \infty)$, we could start from the assumption that the eigenvalues outside S are well approximated, while the ones in S are not. Then a suitable strategy to estimate $E_n(\Lambda(A_N))$ is to construct a polynomial $p_n \in \mathcal{P}_n$ in the following way:

(1) 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.

(2) The other zeros of p_n are free and they are chosen with the aim to minimize $\|p_n\|_S$.

(3) We estimate $E_n(\Lambda(A_N)) \leq \|p_n\|_S$.

Here and in the following we use $\|p_n\|_S = \max_{\lambda \in S} |p_n(\lambda)|$ to denote the supnorm on 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 [12] and [10, 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, $\|p_n\|_S$ may be much bigger than $\|p_n\|_{\Lambda(A_N)}$. On the other hand, choosing a small set S means that according to (1) we fix a lot of zeros of p_n and so we loose a lot of freedom in our choice for minimizing $\|p_n\|_S$.

The main result of this paper is that the above strategy (1)–(3) cannot produce a better asymptotic bound on $E_n(\Lambda(A_N))$ than (1.12). Under three not very restrictive conditions, stated in Section 2, we show that for any closed set S and any choice of $p_n \in \mathcal{P}_n$ satisfying (1) above, the estimate

$$\liminf_{\substack{n,N\to\infty \\ n/N\to t}} \frac{1}{n} \log \|p_n\|_S \geq -\frac{1}{t} \int_0^t g_{S(\tau)}(0)\, d\tau \tag{1.13}$$

holds. The precise statement is in Theorem 3.1 below.

Remark 1.3 In the proof of Theorem 2.1 in [3] we constructed for a given $\epsilon > 0$, a set S and polynomials p_n according to the strategy (1)–(3), such that

$$\limsup_{\substack{n,N\to\infty \\ n/N\to t}} \frac{1}{n} \log \|p_n\|_S \leq -\frac{1}{t} \int_0^t g_{S(\tau)}(0)\, d\tau + \epsilon. \tag{1.14}$$

It also transpires from that proof that the choice $S = S(t)$ leads to equality in (1.14) provided that $\sigma(\partial S(t)) = 0$, where $\partial S(t)$ denotes the boundary of $S(t)$ in $\mathbb{R}$. Since σ does not have point masses (cf. Condition 2 below), this last condition is certainly satisfied if $S(t)$ is a finite union of disjoint intervals, which was the case in all explicit examples considered in [2, 3, 14].

We believe that in pathological cases where the set $S(t)$ has a Cantor like structure, it might happen that equality cannot be achieved in (1.13) for any choice of the set S and polynomials p_n. However, as a guiding principle, it is safe to say that $S(t)$ is the optimal choice for the strategy (1)–(3).

2 The constrained energy problem

We suppose that we are in the situation described by Problem 1.2. So we have a sequence (A_N) of symmetric positive definite matrices with σ as asymptotic eigenvalue distribution. In addition we assume that the Conditions 1, 2, and 3, stated below, are satisfied.

Condition 1 The eigenvalues of the matrices A_N are uniformly bounded, so that they are all in a fixed interval $[0, R]$ for some $R > 0$.

It follows from Condition 1 that σ has a compact support in $[0, R]$.

The second condition is stated in terms of the logarithmic potential U^σ of σ, which is the function

$$U^\sigma : \mathbb{C} \to (-\infty, +\infty] : \lambda \mapsto \int \log \frac{1}{|\lambda - \lambda'|} d\sigma(\lambda'). \tag{2.1}$$

For any finite Borel measure μ with compact support, the logarithmic potential U^μ is superharmonic on $\mathbb{C}$ (in particular lower semi-continuous), and harmonic on $\mathbb{C} \setminus \operatorname{supp}(\sigma)$, see [19].

Condition 2 The logarithmic potential U^σ is a continuous, real-valued function on $\mathbb{C}$.

Condition 2 is a regularity condition on σ. For example, it does not allow σ to have point masses. In applications, σ will typically have a density with respect to Lebesgue measure. Condition 2 is satisfied for example if the density is continuous.

Condition 3 The limit (1.9) also holds for the function $f(\lambda) = \log \lambda$. That is,

$$\lim_{N\to\infty} \frac{1}{N} \sum_{\lambda \in \Lambda(A_N)} \log \lambda = \int \log \lambda \, d\sigma(\lambda).$$

This condition prevents eigenvalues from approaching 0 too fast as $N \to \infty$. It is equivalent to (recall the definition of σ_N in (1.10))

$$\lim_{N\to\infty} U^{\sigma_N}(0) = U^\sigma(0). \tag{2.2}$$

If the eigenvalues of the matrices A_N are all simple (or, more generally, if σ is a probability measure), then there is also an equivalent expression in terms of the determinants of A_N, namely

$$\lim_{N\to\infty} [\det(A_N)]^{1/N} = \exp\left(-U^\sigma(0)\right).$$

The sets $S(t)$ appearing in (1.12) are characterized using a *constrained energy problem* where σ acts as the constraint. The constrained energy problem is to minimize

$$\iint \log \frac{1}{|\lambda - \lambda'|} d\mu(\lambda) d\mu(\lambda') \tag{2.3}$$

among all Borel measures on $[0, R]$ that satisfy

$$0 \leq t\mu \leq \sigma, \qquad \int d\mu = 1. \tag{2.4}$$

The condition $t\mu \leq \sigma$ means that $t\mu(B) \leq \sigma(B)$ for every Borel set B.

Because of Condition 2 there is a unique minimizer for (2.3), (2.4), see [15, 18], denoted here by μ_t. It satisfies the variational conditions

$$U^{\mu_t}(\lambda) \begin{cases} = F_t & \text{for } \lambda \in \operatorname{supp}(\sigma - t\mu_t), \\ \leq F_t & \text{for } \lambda \in \mathbb{C} \setminus \operatorname{supp}(\sigma - t\mu_t), \end{cases} \tag{2.5}$$

for some constant F_t, (a Lagrange multiplier arising from the condition $\int d\mu = 1$). The functional (2.3) on Borel measures μ is strictly convex, so that (2.5) actually characterizes the minimizer μ_t. Now the sets $S(t)$ are defined as

$$S(t) = \operatorname{supp}(\sigma - t\mu_t). \tag{2.6}$$

Note that the minimizer μ_t and the set $S(t)$ also depend on σ. However, we do not indicate this in the notation.

The following theorem is the main result of [3].

Theorem 2.1 [3, Theorem 2.1] *Let (A_N) be a sequence of symmetric positive definite matrices with asymptotic eigenvalue distribution σ, such that Conditions 1, 2, and 3 are satisfied. Then, for every $t \in (0, \|\sigma\|)$,*

$$\limsup_{\substack{n,N\to\infty \\ n/N\to t}} E_n(\Lambda(A_N)) \leq -\frac{1}{t}\int_0^t g_{S(\tau)}(0)\, d\tau, \tag{2.7}$$

where the sets $S(t)$ are defined by (2.6).

Rakhmanov [18] proved that the Ritz values (zeros of the CG polynomial p_n) are asymptotically distributed according to the measure μ_t as $n, N \to \infty$ with $n/N \to t$. From this it follows that there are roughly speaking as many eigenvalues of A_N outside $S(t)$ as there are Ritz values. See [2, 14] for estimates on the distance between Ritz values and eigenvalues in this setting. These results were our main guides in the proof of Theorem 2.1, together with the following lemma.

Lemma 2.2 [3, Theorem 2.1] *We have*

$$\frac{1}{t}\int_0^t g_{S(\tau)}(0)\, d\tau = F_t - U^{\mu_t}(0) \tag{2.8}$$

where F_t is the constant from (2.5).

The proof of the lemma is based on fundamental ideas due to Buyarov and Rakhmanov [6].

The constrained energy problem (2.3)–(2.4) was first studied by Rakhmanov [18] to describe the asymptotic zero distribution of polynomials that are extremal with respect to a discrete measure. Extensions and refinements are due to Dragnev and Saff [8, 9], Kuijlaars and Van Assche [16], and Beckermann [1]. See [15] for a survey. From a different point of view, the constrained energy problem was also studied by Deift and McLaughlin [7].

Remark 2.3 Given the asymptotic distribution σ it is in general a nontrivial task to find the sets $S(t)$ and to calculate the asymptotic bound of (2.7). Complicated examples can generate very complicated behavior. Only for some simple examples we know $S(t)$ explicitly. For instance, for equidistant eigenvalues on $[0,1]$, and thus $d\sigma = dx$, Rakhmanov [18] showed that

$$S(t) = \left[1/2 - \sqrt{1-t^2}/2, 1/2 + \sqrt{1-t^2}/2\right], \qquad t \in (0,1),$$

which via (2.7) leads to the estimate

$$\limsup_{\substack{n,N\to\infty\\ n/N\to t}} E_n(\Lambda(A_N)) \leq -\frac{(1+t)\log(1+t) + (1-t)\log(1-t)}{2t}$$

see [3, Section 3].

Finite difference discretization of the two-dimensional Poisson equation on the unit square leads to matrices with an asymptotic eigenvalue distribution on the interval $[0,8]$ for which the sets $S(t)$ were computed explicitly as

$$S(t) = [4 - 4\cos(\pi t), 4 + 4\cos(\pi t)], \qquad t \in (0, 1/2),$$

and this leads to the asymptotic error estimate

$$\limsup_{\substack{n,N\to\infty\\ n/N\to t}} E_n(\Lambda(A_N)) \leq -\frac{1}{t}\int_0^t \log\left(\tan\left(\frac{\pi}{4}(1+2\tau)\right)\right) d\tau,$$

see [3, Section 5]. Other explicit examples include Toeplitz matrices [3, Section 4] and ultraspherical distributions of eigenvalues [14].

3 The main theorem

In this section we state and prove the main result of this paper.

Theorem 3.1 *Let (A_N) be a sequence of symmetric positive definite matrices with σ as asymptotic eigenvalue distribution. Assume that the Conditions* 1, 2, *and* 3 *are satisfied. Let $t \in (0, \|\sigma\|)$, and let for each N, a natural number $n = n_N$ be given such that $n/N \to t$ as $N \to \infty$. Further, let $S \subset [0,R]$ be a closed set such that*

$$\#(\Lambda(A_N) \setminus S) \leq n \tag{3.1}$$

for every N. Assume that for each N, a polynomial $p_n \in \mathcal{P}_n$ is given such that every eigenvalue of A_N outside S is a zero of p_n. Then

$$\liminf_{N\to\infty} \frac{1}{n}\log\|p_n\|_S \geq -\frac{1}{t}\int_0^t g_{S(\tau)}(0)\, d\tau. \tag{3.2}$$

Remark 3.2 Observe that because of (3.1) polynomials p_n as described in the theorem exist.

Proof. We note first that we may assume that $0 \notin S$. Indeed, if $0 \in S$, then $\|p_n\|_S \geq |p_n(0)| = 1$, and (3.2) holds. So we assume in the rest of the proof that $0 \notin S$.

Define for each N,

$$m = m_N = \#(\Lambda(A_N) \setminus S)$$

and

$$r_m(\lambda) = \prod_{\lambda' \in \Lambda(A_N) \setminus S} \left(1 - \frac{\lambda}{\lambda'}\right). \tag{3.3}$$

Since p_n has a zero at each point of $\Lambda(A_N) \setminus S$ we may then write

$$p_n = q_{n-m} r_m \tag{3.4}$$

with $q_{n-m} \in \mathcal{P}_{n-m}$. Without loss of generality we may assume that q_{n-m} has $n-m$ zeros all contained in the convex hull of S. Indeed, if this were not the case, then we could modify q_{n-m} by moving a zero closer to the convex hull of S, thereby reducing the uniform norm $\|p_n\|_S$ on S. Note that this implies in particular that p_n has exact degree n.

Assume that the theorem is false so that (3.2) does not hold. Then by (2.8) there exists an $\epsilon > 0$ such that

$$\frac{1}{n} \log \|p_n\|_S \leq U^{\mu_t}(0) - F_t - \epsilon \tag{3.5}$$

for infinitely many N, say for $N \in \mathcal{N}_0 \subset \mathbb{N}$.

For an arbitrary polynomial p we define the zero counting measure

$$\nu(p) = \sum_{p(\lambda)=0} \delta_\lambda$$

where each zero is counted according to its multiplicity. Then $(\frac{1}{N}\nu(r_m))$ and $(\frac{1}{N}\nu(q_{n-m}))$ are two sequences of measures on $[0, R]$ having total mass less than 1. By Helly's selection theorem, see e.g. [4, 19], we can extract from $\mathcal{N}_0$ an infinite subsequence $\mathcal{N}_1$ such that the following limits exist in the sense of weak convergence of measures on $[0, R]$:

$$\lim_{\substack{N\to\infty \\ N\in\mathcal{N}_1}} \frac{1}{N}\nu(q_{n-m}) =: \nu. \tag{3.6}$$

$$\lim_{\substack{N\to\infty \\ N\in\mathcal{N}_1}} \frac{1}{N}\nu(r_m) =: \rho, \tag{3.7}$$

From (3.4) we have that $\nu(p_n) = \nu(q_{n-m}) + \nu(r_m)$, so that by (3.6)–(3.7),

$$\lim_{\substack{N\to\infty \\ N\in\mathcal{N}_1}} \frac{1}{N}\nu(p_n) = \nu + \rho. \tag{3.8}$$

We now write

$$\begin{aligned} \frac{1}{N} \log |p_n(\lambda)| &= \frac{1}{N}\left(U^{\nu(p_n)}(0) - U^{\nu(p_n)}(\lambda)\right) \\ &= \frac{1}{N}U^{\nu(q_{n-m})}(0) + \frac{1}{N}U^{\nu(r_m)}(0) - \frac{1}{N}U^{\nu(p_n)}(\lambda), \end{aligned} \tag{3.9}$$

and we are going to analyze the three terms on the right-hand side of (3.9) separately as $N \to \infty$ with $N \in \mathcal{N}_1$.

For the first term we recall that the zeros of the polynomials q_{n-m} are all contained in the convex hull of S, and that 0 does not belong to this convex hull. Therefore $\log \lambda$ is a continuous function on the convex hull, and (3.6) implies

$$U^{\nu}(0) = \lim_{\substack{N\to\infty \\ N\in\mathcal{N}_1}} \frac{1}{N} U^{\nu(q_{n-m})}(0). \tag{3.10}$$

Also, by construction, $\sigma_N - \frac{1}{N}\nu(r_m)$ is a positive measure and

$$\sigma_N|_{S^c} = \frac{1}{N}\nu(r_m), \tag{3.11}$$

where $\cdot|_{S^c}$ denotes the restriction to the complement of the set S. Thus $\sigma_N - \frac{1}{N}\nu(r_m)$ is supported on S. Since $\sigma_N - \frac{1}{N}\nu(r_m) \to \sigma - \rho$ as $N \to \infty$ and $0 \notin S$, it follows that

$$U^{\sigma-\rho}(0) = \lim_{\substack{N\to\infty \\ N\in\mathcal{N}_1}} \left(U^{\sigma_N}(0) - \frac{1}{N} U^{\nu(r_m)}(0) \right). \tag{3.12}$$

Taking into account Condition 3, we get from (3.12)

$$U^{\rho}(0) = \lim_{\substack{N\to\infty \\ N\in\mathcal{N}_1}} \frac{1}{N} U^{\nu(r_m)}(0). \tag{3.13}$$

Next, from (3.8) and the lower envelope theorem, see [19, p. 73], we get

$$U^{\nu+\rho}(\lambda) = \liminf_{\substack{N\to\infty \\ N\in\mathcal{N}_1}} \frac{1}{N} U^{\nu(p_n)}(\lambda) \tag{3.14}$$

for quasi every $\lambda \in \mathbb{C}$. Here "quasi every" means "with the possible exception of a set of capacity zero."

Using (3.10), (3.13), and (3.14) we get from (3.9)

$$\limsup_{\substack{N\to\infty \\ N\in\mathcal{N}_1}} \frac{1}{N} \log |p_n(\lambda)| = U^{\nu}(0) + U^{\rho}(0) - U^{\nu+\rho}(\lambda), \tag{3.15}$$

for quasi every $\lambda \in \mathbb{C}$. Putting

$$\mu = \frac{1}{t}(\nu + \rho)$$

and recalling that $n/N \to t$, we obtain from (3.15) that

$$\limsup_{\substack{N\to\infty \\ N\in\mathcal{N}_1}} \frac{1}{n} \log |p_n(\lambda)| = U^{\mu}(0) - U^{\mu}(\lambda) \tag{3.16}$$

for quasi every $\lambda \in \mathbb{C}$. From (3.5) and (3.16) it then follows that

$$U^{\mu}(0) - U^{\mu}(\lambda) \le U^{\mu_t}(0) - F_t - \epsilon \tag{3.17}$$

for quasi every $\lambda \in S$. By (2.5) we then have

$$U^{\mu_t}(\lambda) \leq F_t \leq U^{\mu}(\lambda) + U^{\mu_t}(0) - U^{\mu}(0) - \epsilon, \tag{3.18}$$

for quasi every $\lambda \in S$.

By construction and (3.11) we have

$$t\mu = \nu + \rho \geq \rho = \lim_{\substack{N\to\infty \\ N\to\mathcal{N}_1}} \sigma_N|_{S^c} \geq \sigma|_{S^c} \tag{3.19}$$

where for the last inequality we have taken into account that S^c is open, compare with [4, Theorem 2.1]. Also note that $t\mu_t \leq \sigma$, since μ_t is the minimizer for the constrained equilibrium problem (2.3)–(2.4). Thus $\mu_t \leq \mu$ outside the set S.

Subtract from (3.18) the logarithmic potential of $\mu_t|_{S^c}$ (which is finite everywhere, since $t\mu_t \leq \sigma$ and U^σ is continuous, cf. [8, 18]), to obtain

$$U^{\mu_t|_S}(\lambda) \leq U^{\mu-\mu_t|_{S^c}}(\lambda) + U^{\mu_t}(0) - U^{\mu}(0) - \epsilon, \tag{3.20}$$

for quasi every $\lambda \in S$. Since $\mu_t \leq \mu$ outside S, we have that $\mu - \mu_t|_{S^c}$ is a positive measure. We also note that $\mu_t|_S$ has finite energy, and that the total masses of the two measures $\mu_t|_S$ and $\mu - \mu_t|_{S^c}$ are the same. Therefore, we can apply the principle of domination, see [19, p. 104], to (3.20) and it follows that (3.20) holds for all $\lambda \in \mathbb{C}$. Then the inequality (3.18) holds for every $\lambda \in \mathbb{C}$ as well. Taking $\lambda = 0$ in (3.18), however, we arrive at a contradiction.

This contradiction shows that our assumption that (3.2) does not hold must be incorrect. Therefore (3.2) is true, and the theorem is proved. □

Remark 3.3 In Remark 1.3 we already discussed the case of equality in (3.2), and we found that in general a set S with equality need not exist, but that in certain cases equality holds for $S = S(t)$.

Here we complement this remark by analyzing what would happen in case of equality for S. Extending the arguments used in the proof of Theorem 3.1, we show that then necessarily

$$S(t) \subset S \subset \{\lambda \in \mathbb{R} : U^{\mu_t}(\lambda) = F_t\} \cup E \tag{3.21}$$

where E is a set of zero capacity. Note that $S(t) \subset \{\lambda \in \mathbb{R} : U^{\mu_t}(\lambda) = F_t\}$ by (2.5) and in general equality need not hold. However, it is known that

$$\{\lambda \in \mathbb{R} : U^{\mu_t}(\lambda) = F_t\} = \bigcap_{\tau<t} S(\tau),$$

see [6], so that this set is still intimately related to the sets $S(t)$.

In order to prove (3.21) we assume the conditions and notations of Theorem 3.1 and its proof, and we suppose that S is such that there is equality in (3.2). Then we have (instead of (3.5))

$$\frac{1}{n} \log \|p_n\|_S \leq U^{\mu_t}(0) - F_t + \epsilon_N$$

with $\epsilon_N \to 0$ as $N \to \infty$ with $N \in \mathcal{N}_0$. Following the proof of Theorem 3.1, we then find that (3.20) holds with $\epsilon = 0$, and therefore

$$U^{\mu_t}(\lambda) \leq U^{\mu}(\lambda) + U^{\mu_t}(0) - U^{\mu}(0)$$

for all $\lambda \in \mathbb{C}$. Then the function $h := U^{\mu} - U^{\mu_t}$ is superharmonic on $\mathbb{C} \setminus S$, and bounded from below by $h(0)$ on $\mathbb{C}$. Since 0 belongs to $\mathbb{C} \setminus S$, and $\mathbb{C} \setminus S$ is connected, it follows from the generalized minimum principle for superharmonic functions [19, p. 39] that h is constant on $\mathbb{C} \setminus S$. By the unicity theorem for logarithmic potentials [19, p. 97] we then conclude that $\mu = \mu_t$. Hence $t\mu = t\mu_t \leq \sigma$. Since by (3.19) we have $t\mu \geq \sigma$ outside S, it follows that $t\mu = \sigma$ outside S and therefore

$$S(t) = \mathrm{supp}\,(\sigma - t\mu_t) = \mathrm{supp}\,(\sigma - t\mu) \subset S.$$

This establishes the first inclusion of (3.21)

For the second inclusion, we note that by (3.16) and the fact that $\mu_t = \mu$, we have

$$\limsup_{\substack{N\to\infty \\ N\in\mathcal{N}_1}} \frac{1}{n} \log |p_n(\lambda)| = U^{\mu_t}(0) - U^{\mu_t}(\lambda) \tag{3.22}$$

for quasi every $\lambda \in \mathbb{C}$. Let E denote the exceptional set, so that E has zero capacity. Then if $\lambda \in S \setminus E$, we have both (3.22) and

$$\begin{aligned} \limsup_{\substack{N\to\infty \\ N\in\mathcal{N}_1}} \frac{1}{n} \log |p_n(\lambda)| &\leq \limsup_{\substack{N\to\infty \\ N\in\mathcal{N}_1}} \frac{1}{n} \log \|p_n\|_S \\ &\leq U^{\mu_t}(0) - F_t. \end{aligned}$$

Hence $U^{\mu_t}(\lambda) \geq F_t$ and in view of (2.5) this implies $U^{\mu_t}(\lambda) = F_t$. So we have the second inclusion of (3.21).

References

[1] B. Beckermann, On a conjecture of E.A. Rakhmanov, Constr. Approx. 16 (2000), 427–448.

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

[3] B. Beckermann and A.B.J. Kuijlaars, Superlinear convergence of Conjugate Gradients, submitted.

[4] P. Billingsley, Convergence of probability measures, Second edition, John Wiley & Sons, New York, 1999.

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

[6] V.S. Buyarov and E.A. Rakhmanov, Families of equilibrium measures in an external field on the real axis, Sb. Math. 190 (1999), 791-802.

[7] P. Deift and K.T-R. McLaughlin, A continuum limit of the Toda lattice, Mem. Amer. Math. Soc. 131 (1998), no. 624.

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

[9] P.D. Dragnev and E.B. Saff, A problem in potential theory and zero asymptotics of Krawtchouk polynomials, J. Approx. Theory 102 (2000), 120–140.

[10] T.A. Driscoll, K-C. Toh, and L.N. Trefethen, From potential theory to matrix iterations in six steps, SIAM Review 40 (1998), 547–578.

[11] B. Fischer, Polynomial based iteration methods for symmetric linear systems, Wiley-Teubner, 1996.

[12] A. Greenbaum, Comparison of splittings used with the conjugate gradient algorithm, Numer. Math. 33 (1979), 181–193.

[13] A. Greenbaum, Iterative methods for solving linear systems, SIAM, Philadelphia, PA, 1997.

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

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

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

[17] O. Nevanlinna, Convergence of iterations for linear equations, Birkhäuser Verlag, Basel, 1993.

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

[19] E.B. Saff and V. Totik, Logarithmic potentials with external fields, Springer-Verlag, Berlin, 1997.

[20] S. Serra Capizzano, A note on the asymptotic spectra of finite difference discretizations of second order elliptic partial differential equations, Asian J. Math., to appear.

[21] S. Serra Capizzano and P. Tilli, From partial differential equations to generalised locally Toeplitz sequences, manuscript.

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

[23] L.N. Trefethen and D. Bau III, Numerical linear algebra, SIAM, Philadelphia, PA, 1997.

[24] E.E. Tyrtyshnikov and N.L. Zamarashkin, Spectra of multilevel Toeplitz matrices: advanced theory via simple matrix relationships, Linear Algebra Appl. 270 (1998), 15–27.