Operators
and
Matrices
Volume 9, Number 3 (2015), 549–561
doi:10.7153/oam-09-33

A GENERAL TOOL FOR DETERMINING THE ASYMPTOTIC SPECTRAL DISTRIBUTION OF HERMITIAN MATRIX–SEQUENCES

CARLO GARONI, STEFANO SERRA-CAPIZZANO AND PARIS VASSALOS

(*Communicated by A. Böttcher*)

Abstract. We consider sequences of Hermitian matrices with increasing dimension, and we provide a general tool for determining the asymptotic spectral distribution of a 'difficult' sequence $\{A_n\}_n$ from the one of 'simpler' sequences $\{B_{n,m}\}_n$ that approximate $\{A_n\}_n$ when $m \to \infty$. The tool is based on the notion of an approximating class of sequences (a.c.s.), which was inspired by the work of Paolo Tilli and the second author, and it is applied here in a more general setting. An a.c.s.-based proof of the famous Szegö theorem on the spectral distribution of Toeplitz matrices is finally presented.

1. Introduction

In the last decades, an approximation theory for sequences of matrices with increasing dimension has been developed, having in mind the spectral approximation. Such a topic has both theoretical and practical interest. For instance, in a numerical analysis context, when a large linear system with coefficient matrix A_n is given, we would like to have a procedure for constructing a matrix P_n such that: (a) the cost for solving a linear system with matrix P_n is of the same order as the matrix-vector product $A_n\mathbf{x}$; (b) $A_n - P_n$ or $P_n^{-1}A_n - I$ is spectrally close to the null matrix, with I being the identity matrix. Of course, the notion of spectral closeness has to be made precise. Looking at the convergence properties of important and popular solvers, preconditioned Krylov and multigrid, just to mention the most successful, it is evident that such a notion should take into consideration two aspects:

1. two matrices are spectrally close if their difference is small in some norm (preferably an induced or operator norm);

2. two matrices are spectrally close if the rank of the difference is small with respect to the matrix size.

Mathematics subject classification (2010): 47B06, 15A60, 15B05.

Keywords and phrases: Approximating class of sequences, spectral distribution, Szegö theorem, Toeplitz matrices.

The work of the first and second author has been partly financed by the national group GNCS-INDAM. The work of the second author has been also partly supported by a grant 'Becoming the Number One – 2014' of the Knut and Alice Wallenberg Foundation. The work of the third author has been co-financed by the European Union (European Social Fund – ESF) and Greek national funds through the Operational Program 'Education and Lifelong Learning' of the National Strategic Reference Framework (NSRF) – Research Funding Program: Thales. Investing in knowledge society through the European Social Fund.

Paper OaM-09-33

The second aspect may seem less natural, but, in fact, it is largely used in numerical analysis. As an example, the Sherman–Morrison–Woodbury formula [7, p. 50] is a fast way for obtaining the inverse of a low-rank correction of a given invertible matrix; the convergence speed of the preconditioned conjugate gradient method is not affected, or only mildly affected, by a small number of spectral outliers whose presence is due to a low-rank correction matrix [1].

The concept of spectral closeness is also important for determining the asymptotic spectral distribution of sequences formed by Hermitian matrices with increasing dimension. Roughly speaking, if $\{A_n\}_n$ is a 'difficult' sequence and if $\{B_{n,m}\}_n$ is a 'simple' sequence[1] that approximates $\{A_n\}_n$ (in the sense of items 1–2 above) when $m \to \infty$, then the asymptotic spectral distribution of $\{A_n\}_n$ can be computed as the limit of the asymptotic spectral distribution of $\{B_{n,m}\}_n$ for $m \to \infty$.

This note is devoted to the mathematical foundation of this idea, which, in fact, is presented here on a completely abstract level. In this respect, our main result (Theorem 4) reflects our effort to provide a very general tool for computing asymptotic spectral distributions. It is important to stress that this tool is based on the notion of an approximating class of sequences (a.c.s.), which is due to the second author [10], but was originally inspired by the work of Paolo Tilli on Locally Toeplitz sequences [13]. We also illustrate how the tool can be used in order to derive the famous Szegö theorem on the spectral distribution of Toeplitz matrices.

The note is organized in two sections: the first contains the definition of a.c.s. and the main results, while the second provides an a.c.s.-based proof of the Szegö theorem.

2. Definitions and results

We begin by introducing the notion of an approximating class of sequences [10, Definition 2.1]. Throughout this note, the word 'matrix-sequence' is used as a synonym of 'sequence of matrices'. Moreover, $\|X\|$ denotes the spectral (Euclidean) norm of the matrix X, i.e., the maximum singular value of X.

DEFINITION 1. (Approximating class of sequences) Let $\{A_n\}_n$ be a matrix-sequence, with A_n of size d_n tending to infinity. An approximating class of sequences (a.c.s.) for $\{A_n\}_n$ is a sequence of matrix-sequences $\{\{B_{n,m}\}_n : m\}$ such that, for each m,

$$A_n = B_{n,m} + R_{n,m} + N_{n,m} \qquad \forall n \geqslant n_m \tag{1}$$

where $\operatorname{rank}(R_{n,m}) \leqslant \rho(m)d_n$, $\|N_{n,m}\| \leqslant \nu(m)$, the quantities $n_m, \rho(m), \nu(m)$ depend only on m and $\lim\limits_{m\to\infty} \rho(m) = \lim\limits_{m\to\infty} \nu(m) = 0$.

Roughly speaking, saying that $\{\{B_{n,m}\}_n : m\}$ is an a.c.s. for $\{A_n\}_n$ means that A_n is equal to $B_{n,m}$ plus a small-rank matrix (with respect to the size d_n) plus a small-norm matrix. Lemma 1 shows that, if A_n and $B_{n,m}$ are Hermitian, then the small-rank

[1]Here, 'simple' has to be intended in the sense of the related computational complexity and, for instance, in the sense of sparse vs. dense, shift-invariant vs. smoothly shift-variant, Toeplitz vs. quasi-Toeplitz, circulant vs. Toeplitz, etc.

matrix $R_{n,m}$ and the small-norm matrix $N_{n,m}$ in the splitting (1) may be supposed to be Hermitian.

LEMMA 1. *Let $\{A_n\}_n$ be a sequence of Hermitian matrices, with A_n of size $d_n \to \infty$, and let $\{\{B_{n,m}\}_n : m\}$ be an a.c.s. for $\{A_n\}_n$ formed by Hermitian matrices (i.e., every $B_{n,m}$ is Hermitian). Then, for every m, we have*

$$A_n = B_{n,m} + R_{n,m} + N_{n,m} \qquad \forall n \geqslant n_m$$

where $R_{n,m}, N_{n,m}$ are Hermitian, $\mathrm{rank}(R_{n,m}) \leqslant \rho(m) d_n$, $\|N_{n,m}\| \leqslant \nu(m)$, *the quantities n_m, $\rho(m)$, $\nu(m)$ depend only on m and $\lim\limits_{m\to\infty} \rho(m) = \lim\limits_{m\to\infty} \nu(m) = 0$.*

Proof. Take the real part in (1) and use the inequalities $\mathrm{rank}(\Re(X)) \leqslant 2\,\mathrm{rank}(X)$ and $\|\Re(X)\| \leqslant \|X\|$ to conclude that, by replacing $R_{n,m}, N_{n,m}$ with $\Re(R_{n,m}), \Re(N_{n,m})$ (if necessary), we can assume $R_{n,m}, N_{n,m}$ to be Hermitian. $\square$

Now we turn to the main theorems (Theorems 2 and 4), which provide a general tool for determining the asymptotic spectral distribution of a 'difficult' matrix-sequence $\{A_n\}_n$ formed by Hermitian matrices, starting from the knowledge of the asymptotic spectral distribution of simpler matrix-sequences $\{B_{n,m}\}_n$, $m = 1,2,3,\ldots$, again formed by Hermitian matrices. For any Hermitian matrix $X \in \mathbb{C}^{m\times m}$, the eigenvalues of X are arranged in non-increasing order: $\lambda_1(X) \geqslant \ldots \geqslant \lambda_m(X)$; moreover, we set (by convention) $\lambda_j(X) := +\infty$ if $j \leqslant 0$ and $\lambda_j(X) := -\infty$ if $j \geqslant m+1$. The following interlacing theorem [2, p. 63] is needed in the proof of Theorem 2.

THEOREM 1. *Let $Y = X + E$, where $X, E \in \mathbb{C}^{m\times m}$ are Hermitian. Let $k^+, k^- \geqslant 0$ be respectively the number of positive and the number of negative eigenvalues of E, i.e.*

$$k^+ := \#\{j \in \{1,\ldots,m\} : \lambda_j(E) > 0\}, \qquad k^- := \#\{j \in \{1,\ldots,m\} : \lambda_j(E) < 0\}.$$

Then

$$\lambda_{j-k^+}(X) \geqslant \lambda_j(Y) \geqslant \lambda_{j+k^-}(X), \qquad \forall j = 1,\ldots,m.$$

In particular, if $\mathrm{rank}(E) \leqslant k$, *then*

$$\lambda_{j-k}(X) \geqslant \lambda_j(Y) \geqslant \lambda_{j+k}(X), \qquad \forall j = 1,\ldots,m.$$

If $H : \mathbb{R} \to \mathbb{R}$, we define $H(\infty) := \lim\limits_{x\to\infty} H(x)$ (whenever the limit exists). Similarly, $H(-\infty) := \lim\limits_{x\to-\infty} H(x)$. Moreover, we denote by $C_c(\mathbb{R})$ (resp. $C_c(\mathbb{C})$) the space of complex-valued continuous functions defined over $\mathbb{R}$ (resp. $\mathbb{C}$) and with bounded support. Finally, we set $C_c^1(\mathbb{R}) := C_c(\mathbb{R}) \cap C^1(\mathbb{R})$, where $C^1(\mathbb{R})$ is the space of complex-valued functions F defined on $\mathbb{R}$ and such that the real and imaginary parts, $\Re(F)$ and $\Im(F)$, are of class $C^1(\mathbb{R})$ in the classical sense.

THEOREM 2. *Let $\{A_n\}_n$ be a sequence of Hermitian matrices, with A_n of size $d_n \to \infty$. Assume that*

1. *$\{\{B_{n,m}\}_n : m\}$ is an a.c.s. for $\{A_n\}_n$ formed by Hermitian matrices;*

2. *for every m and every $F \in C_c^1(\mathbb{R})$, there exists $\lim_{n\to\infty} \frac{1}{d_n}\sum_{j=1}^{d_n} F(\lambda_j(B_{n,m})) =: \phi_m(F) \in \mathbb{C}$;*

3. *for every $F \in C_c^1(\mathbb{R})$, there exists $\lim_{m\to\infty} \phi_m(F) =: \phi(F) \in \mathbb{C}$.*

Then, for all $F \in C_c^1(\mathbb{R})$,

$$\exists \lim_{n\to\infty} \frac{1}{d_n}\sum_{j=1}^{d_n} F(\lambda_j(A_n)) = \phi(F). \tag{2}$$

Proof. The technique of this proof is taken from [10, Proposition 2.3], where an analogous result was proved for the singular values instead of the eigenvalues. We first observe that it suffices to prove (2) for those test functions $F \in C_c^1(\mathbb{R})$ that are real-valued. Indeed, any (complex-valued) $F \in C_c^1(\mathbb{R})$ can be decomposed as $F = \Re(F) + \mathrm{i}\Im(F)$, where $\Re(F), \Im(F) \in C_c^1(\mathbb{R})$. Thus, once we have proved (2) for all real-valued functions in $C_c^1(\mathbb{R})$, we have

$$\begin{aligned}\lim_{n\to\infty} \frac{1}{d_n}\sum_{j=1}^{d_n} F(\lambda_j(A_n)) &= \lim_{n\to\infty} \frac{1}{d_n}\sum_{j=1}^{d_n} [\Re(F(\lambda_j(A_n))) + \mathrm{i}\Im(F(\lambda_j(A_n)))] \\ &= \phi(\Re(F)) + \mathrm{i}\phi(\Im(F)) = \phi(F),\end{aligned}$$

where the last equality holds by the linearity of the functional ϕ, which follows from its definition.

Now, let $F \in C_c^1(\mathbb{R})$ be real-valued. For all n, m we have

$$\begin{aligned}\left|\frac{1}{d_n}\sum_{j=1}^{d_n} F(\lambda_j(A_n)) - \phi(F)\right| &\leqslant \left|\frac{1}{d_n}\sum_{j=1}^{d_n} F(\lambda_j(A_n)) - \frac{1}{d_n}\sum_{j=1}^{d_n} F(\lambda_j(B_{n,m}))\right| \\ &+ \left|\frac{1}{d_n}\sum_{j=1}^{d_n} F(\lambda_j(B_{n,m})) - \phi_m(F)\right| + |\phi_m(F) - \phi(F)|. \end{aligned}\tag{3}$$

By hypothesis, the second term in the right-hand side tends to 0 for $n \to \infty$, while the third one tends to 0 for $m \to \infty$. Therefore, if we prove that

$$\lim_{m\to\infty} \limsup_{n\to\infty} \left|\frac{1}{d_n}\sum_{j=1}^{d_n} F(\lambda_j(A_n)) - \frac{1}{d_n}\sum_{j=1}^{d_n} F(\lambda_j(B_{n,m}))\right| = 0, \tag{4}$$

then, passing first to the $\limsup_{n\to\infty}$ and then to the $\lim_{m\to\infty}$ in (3), we get the thesis.

In conclusion, we only have to prove (4). To this end, we recall that $\{\{B_{n,m}\}_n : m\}$ is an a.c.s. for $\{A_n\}_n$ and that $A_n, B_{n,m}$ are Hermitian as in Lemma 1. Hence, for every m,

$$A_n = B_{n,m} + R_{n,m} + N_{n,m} \qquad \forall n \geqslant n_m$$

where $R_{n,m}, N_{n,m}$ are Hermitian, $\mathrm{rank}(R_{n,m}) \leqslant \rho(m)d_n$, $\|N_{n,m}\| \leqslant \nu(m)$, the quantities n_m, $\rho(m)$, $\nu(m)$ depend only on m and $\lim_{m\to\infty} \rho(m) = \lim_{m\to\infty} \nu(m) = 0$. We can then write, for every m and every $n \geqslant n_m$,

$$\begin{aligned}
&\left| \frac{1}{d_n} \sum_{j=1}^{d_n} F(\lambda_j(A_n)) - \frac{1}{d_n} \sum_{j=1}^{d_n} F(\lambda_j(B_{n,m})) \right| \\
&\leqslant \left| \frac{1}{d_n} \sum_{j=1}^{d_n} F(\lambda_j(A_n)) - \frac{1}{d_n} \sum_{j=1}^{d_n} F(\lambda_j(B_{n,m}+R_{n,m})) \right| \\
&\quad + \left| \frac{1}{d_n} \sum_{j=1}^{d_n} F(\lambda_j(B_{n,m}+R_{n,m})) - \frac{1}{d_n} \sum_{j=1}^{d_n} F(\lambda_j(B_{n,m})) \right|. \qquad (5)
\end{aligned}$$

We will consider separately the two terms in the right-hand side of (5), and we will show that each of them is bounded from above by a quantity depending only on m and tending to 0 as $m \to \infty$. After this, (4) is proved and the thesis follows.

In order to estimate the first term in the right-hand side of (5), we use the Weyl's perturbation theorem; see [2, p. 63]. We have

$$\begin{aligned}
&\left| \frac{1}{d_n} \sum_{j=1}^{d_n} F(\lambda_j(A_n)) - \frac{1}{d_n} \sum_{j=1}^{d_n} F(\lambda_j(B_{n,m}+R_{n,m})) \right| \\
&\leqslant \frac{1}{d_n} \sum_{j=1}^{d_n} \left| F(\lambda_j(A_n)) - F(\lambda_j(B_{n,m}+R_{n,m})) \right| \\
&\leqslant \frac{1}{d_n} \sum_{j=1}^{d_n} \|F'\|_\infty \left| \lambda_j(A_n) - \lambda_j(B_{n,m}+R_{n,m}) \right| \\
&\leqslant \|F'\|_\infty \|A_n - B_{n,m} - R_{n,m}\| = \|F'\|_\infty \|N_{n,m}\| \leqslant \|F'\|_\infty \nu(m),
\end{aligned}$$

which tends to 0 as $m \to \infty$.

In order to estimate the second term in the right-hand side of (5), we will use the interlacing Theorem 1. We first observe that F can be expressed as the difference between two nonnegative, non-decreasing, bounded functions:

$$F = H - K, \qquad H(x) := \int_{-\infty}^{x} (F')_+(t)dt, \qquad K(x) := \int_{-\infty}^{x} (F')_-(t)dt,$$

where $(F')_+ := \max(F',0)$ and $(F')_- := \max(-F',0)$. Hence, for the second term in the right-hand side of (5) we have

$$\begin{aligned}
&\left| \frac{1}{d_n} \sum_{j=1}^{d_n} F(\lambda_j(B_{n,m}+R_{n,m})) - \frac{1}{d_n} \sum_{j=1}^{d_n} F(\lambda_j(B_{n,m})) \right| \\
&\leqslant \left| \frac{1}{d_n} \sum_{j=1}^{d_n} H(\lambda_j(B_{n,m}+R_{n,m})) - \frac{1}{d_n} \sum_{j=1}^{d_n} H(\lambda_j(B_{n,m})) \right| \\
&\quad + \left| \frac{1}{d_n} \sum_{j=1}^{d_n} K(\lambda_j(B_{n,m}+R_{n,m})) - \frac{1}{d_n} \sum_{j=1}^{d_n} K(\lambda_j(B_{n,m})) \right|. \qquad (6)
\end{aligned}$$

Defining $r_{n,m} := \operatorname{rank}(R_{n,m}) \leqslant \rho(m)d_n$, Theorem 1 gives

$$\lambda_{j-r_{n,m}}(B_{n,m}) \geqslant \lambda_j(B_{n,m}+R_{n,m}) \geqslant \lambda_{j+r_{n,m}}(B_{n,m}), \qquad \forall j=1,\ldots,d_n,$$

and, moreover, it is clear from our notation that

$$\lambda_{j-r_{n,m}}(B_{n,m}) \geqslant \lambda_j(B_{n,m}) \geqslant \lambda_{j+r_{n,m}}(B_{n,m}), \qquad \forall j=1,\ldots,d_n.$$

Therefore, recalling the monotonicity and nonnegativity of H,

$$\begin{aligned}
&\left|\frac{1}{d_n}\sum_{j=1}^{d_n} H(\lambda_j(B_{n,m}+R_{n,m})) - \frac{1}{d_n}\sum_{j=1}^{d_n} H(\lambda_j(B_{n,m}))\right| \\
&\leqslant \frac{1}{d_n}\sum_{j=1}^{d_n} \left|H(\lambda_j(B_{n,m}+R_{n,m})) - H(\lambda_j(B_{n,m}))\right| \\
&\leqslant \frac{1}{d_n}\sum_{j=1}^{d_n} \left|H(\lambda_{j-r_{n,m}}(B_{n,m})) - H(\lambda_{j+r_{n,m}}(B_{n,m}))\right| \\
&= \frac{1}{d_n}\sum_{j=1}^{d_n} H(\lambda_{j-r_{n,m}}(B_{n,m})) - \frac{1}{d_n}\sum_{j=1}^{d_n} H(\lambda_{j+r_{n,m}}(B_{n,m})) \\
&= \frac{1}{d_n}\sum_{j=1-r_{n,m}}^{d_n-r_{n,m}} H(\lambda_j(B_{n,m})) - \frac{1}{d_n}\sum_{j=1+r_{n,m}}^{d_n+r_{n,m}} H(\lambda_j(B_{n,m})) \\
&= \frac{1}{d_n}\sum_{j=1-r_{n,m}}^{r_{n,m}} H(\lambda_j(B_{n,m})) - \frac{1}{d_n}\sum_{j=d_n-r_{n,m}+1}^{d_n+r_{n,m}} H(\lambda_j(B_{n,m})) \\
&\leqslant \frac{1}{d_n}\sum_{j=1-r_{n,m}}^{r_{n,m}} H(\lambda_j(B_{n,m})) \leqslant \frac{2r_{n,m}H(\infty)}{d_n} \leqslant 2\rho(m)\|H\|_\infty.
\end{aligned}$$

Similarly, one can show that the second term in the right-hand side of (6) is bounded from above by $2\rho(m)\|K\|_\infty$, implying that the quantity in (6), namely the second term in the right-hand side of (5), is less than or equal to $2(\|H\|_\infty + \|K\|_\infty)\rho(m)$. Since the latter tends to 0 as $m \to \infty$, the thesis is proved. □

The only unpleasant point about Theorem 2 is that, in traditional formulations of asymptotic spectral distribution results, the usual set of test functions F is $C_c(\mathbb{C})$ or $C_c(\mathbb{R})$, but not $C_c^1(\mathbb{R})$; see, e.g., Definition 2 below or [6, Definition 3.1]. However, this point is readily settled in Theorem 4, where we prove that, under the same hypotheses of Theorem 2, if the second and third assumptions are met for every $F \in C_c(\mathbb{R})$, then (2) holds for every $F \in C_c(\mathbb{R})$. For the proof of Theorem 4, we shall use the following corollary of the Banach-Steinhaus theorem [9].

THEOREM 3. *Let $\mathscr{E},\mathscr{F}$ be normed vector spaces, with $\mathscr{E}$ a Banach space, and let $T_n : \mathscr{E} \to \mathscr{F}$ be a sequence of continuous linear operators. Assume that, for all $x \in \mathscr{E}$, there exists $\lim_{n\to\infty} T_n x =: Tx \in \mathscr{F}$. Then,*

- $\sup \|T_n\| < \infty$;
- $T : \mathscr{E} \to \mathscr{F}$ *is a continuous linear operator with* $\|T\| \leqslant \liminf_{n\to\infty} \|T_n\|$.

THEOREM 4. *Let $\{A_n\}_n$ be a sequence of Hermitian matrices, with A_n of size $d_n \to \infty$. Assume that*

1. *$\{\{B_{n,m}\}_n : m\}$ is an a.c.s. for $\{A_n\}_n$ formed by Hermitian matrices;*
2. *for every m and every $F \in C_c(\mathbb{R})$, there exists $\lim_{n\to\infty} \frac{1}{d_n}\sum_{j=1}^{d_n} F(\lambda_j(B_{n,m})) =: \phi_m(F) \in \mathbb{C}$;*
3. *for every $F \in C_c(\mathbb{R})$, there exists $\lim_{m\to\infty} \phi_m(F) =: \phi(F) \in \mathbb{C}$.*

Then $\phi : (C_c(\mathbb{R}), \|\cdot\|_\infty) \to \mathbb{C}$ is a continuous linear functional with $\|\phi\| \leqslant 1$, and, for all $F \in C_c(\mathbb{R})$,

$$\exists \lim_{n\to\infty} \frac{1}{d_n}\sum_{j=1}^{d_n} F(\lambda_j(A_n)) = \phi(F). \tag{7}$$

Proof. For fixed n,m, let

$$\phi_{n,m}(F) := \frac{1}{d_n}\sum_{j=1}^{d_n} F(\lambda_j(B_{n,m})) : (C_c(\mathbb{R}), \|\cdot\|_\infty) \to \mathbb{C}.$$

It is clear that each $\phi_{n,m}$ is a continuous linear functional on $(C_c(\mathbb{R}), \|\cdot\|_\infty)$ with $\|\phi_{n,m}\| \leqslant 1$. Indeed, the linearity of $\phi_{n,m}$ is obvious and the inequality $|\phi_{n,m}(F)| \leqslant \|F\|_\infty$, which is satisfied for all $F \in C_c(\mathbb{R})$, yields the continuity of $\phi_{n,m}$ as well as the bound $\|\phi_{n,m}\| \leqslant 1$. The functional ϕ_m is the pointwise limit of $\phi_{n,m}$ as $n \to \infty$. Hence, by Theorem 3, $\phi_m : (C_c(\mathbb{R}), \|\cdot\|_\infty) \to \mathbb{C}$ is a continuous linear functional on $(C_c(\mathbb{R}), \|\cdot\|_\infty)$ with $\|\phi_m\| \leqslant 1$. The functional ϕ is the pointwise limit of ϕ_m as $m \to \infty$. Hence, again by Theorem 3, ϕ is a continuous linear functional on $(C_c(\mathbb{R}), \|\cdot\|_\infty)$ with $\|\phi\| \leqslant 1$.

Now, fix $F \in C_c(\mathbb{R})$. For all $\varepsilon > 0$ we can find $F_\varepsilon \in C_c^1(\mathbb{R})$ such that $\|F - F_\varepsilon\|_\infty \leqslant \varepsilon$. As a consequence, for all $\varepsilon > 0$ and all n we have

$$\begin{aligned}
&\left|\frac{1}{d_n}\sum_{j=1}^{d_n} F(\lambda_j(A_n)) - \phi(F)\right| \\
\leqslant &\left|\frac{1}{d_n}\sum_{j=1}^{d_n} F(\lambda_j(A_n)) - \frac{1}{d_n}\sum_{j=1}^{d_n} F_\varepsilon(\lambda_j(A_n))\right| \\
&+ \left|\frac{1}{d_n}\sum_{j=1}^{d_n} F_\varepsilon(\lambda_j(A_n)) - \phi(F_\varepsilon)\right| + |\phi(F_\varepsilon) - \phi(F)| \\
\leqslant &\|F - F_\varepsilon\|_\infty + \left|\frac{1}{d_n}\sum_{j=1}^{d_n} F_\varepsilon(\lambda_j(A_n)) - \phi(F_\varepsilon)\right| + |\phi(F_\varepsilon) - \phi(F)|.
\end{aligned}$$

Considering that (7) holds for F_ε by Theorem 2, we have

$$\limsup_{n\to\infty}\left|\frac{1}{d_n}\sum_{j=1}^{d_n}F(\lambda_j(A_n))-\phi(F)\right|\leqslant\varepsilon+|\phi(F_\varepsilon)-\phi(F)|.$$

Passing to the limit as $\varepsilon\to 0$ and taking into account the continuity of ϕ, we obtain

$$\limsup_{n\to\infty}\left|\frac{1}{d_n}\sum_{j=1}^{d_n}F(\lambda_j(A_n))-\phi(F)\right|=0,$$

which means that (7) holds for every $F\in C_c(\mathbb{R})$. □

3. An a.c.s.-based proof of the Szegö theorem on the spectral distribution of Toeplitz matrices

As an application of Theorem 4, we present in this section a new proof of the famous Szegö theorem on the spectral distribution of Toeplitz matrices. This theorem, which originally appeared in [8], has undergone several extensions; see [15, 16, 4, 14]. For the proof of these extensions, other arguments, different from the one used in [8], have been proposed. In particular, Tilli's argument [14] is similar to the one that we are going to present, but it does not make use of the concept of a.c.s., which was introduced later. To our knowledge, an a.c.s.-based proof, like the one that we are going to see in the following, has never appeared in the literature.

In the following, we denote by m the Lebesgue measure on $\mathbb{R}$. Moreover, the word 'measurable' always means 'Lebesgue measurable'.

DEFINITION 2. Let $\{A_n\}_n$ be a sequence of matrices, with A_n of size $d_n\to\infty$, and let $f:D\to\mathbb{C}$ be a measurable function, defined on a measurable set $D\subset\mathbb{R}$ with $0<m(D)<\infty$. We say that $\{A_n\}_n$ has an asymptotic spectral distribution described by f, in symbols $\{A_n\}_n\sim_\lambda f$, if

$$\lim_{n\to\infty}\frac{1}{d_n}\sum_{j=1}^{d_n}F(\lambda_j(A_n))=\phi_f(F),\qquad \forall F\in C_c(\mathbb{C}),\tag{8}$$

where

$$\phi_f(F):=\frac{1}{m(D)}\int_D F(f(x))dx.\tag{9}$$

In the case where $\{A_n\}_n$ is formed by Hermitian matrices and f is real-valued, all the eigenvalues of A_n are real and writing $\{A_n\}_n\sim_\lambda f$ is equivalent to saying that (8) is satisfied for every test function $F\in C_c(\mathbb{R})$, with ϕ_f still defined by (9). Concerning the functional ϕ_f, we record the following property, of interest later on.

LEMMA 2. *Let $f_m : D \to \mathbb{C}$ be a sequence of measurable functions, defined on a measurable set $D \subset \mathbb{R}$ with $0 < m(D) < \infty$, and assume that f_m converges in measure to some measurable function $f : D \to \mathbb{C}$. Then,*

$$\phi_{f_m}(F) \to \phi_f(F), \qquad \forall F \in C_c(\mathbb{C}). \tag{10}$$

In particular, if f_m, f are real-valued then

$$\phi_{f_m}(F) \to \phi_f(F), \qquad \forall F \in C_c(\mathbb{R}). \tag{11}$$

Proof. Let $F \in C_c(\mathbb{C})$ and $\varepsilon > 0$. Defining $\{|f_m - f| \geqslant \varepsilon\} := \{x \in D : |f_m(x) - f(x)| \geqslant \varepsilon\}$ and $\{|f_m - f| < \varepsilon\} := \{x \in D : |f_m(x) - f(x)| < \varepsilon\}$, we have

$$\begin{aligned}
|\phi_{f_m}(F) - \phi_f(F)| &\leqslant \frac{1}{m(D)} \int_D |F(f_m(x)) - F(f(x))| dx \\
&= \frac{1}{m(D)} \int_{\{|f_m - f| \geqslant \varepsilon\}} |F(f_m(x)) - F(f(x))| dx \\
&\quad + \frac{1}{m(D)} \int_{\{|f_m - f| < \varepsilon\}} |F(f_m(x)) - F(f(x))| dx \\
&\leqslant \frac{2\|F\|_\infty m(\{|f_m - f| \geqslant \varepsilon\})}{m(D)} + \omega_F(\varepsilon),
\end{aligned} \tag{12}$$

where ω_F is the modulus of continuity of F. Note that $\lim_{m\to\infty} m(\{|f_m - f| \geqslant \varepsilon\}) = 0$ (because $f_m \to f$ in measure) and $\lim_{\varepsilon\to 0} \omega_F(\varepsilon) = 0$ (because F is uniformly continuous by the Heine-Cantor theorem). Hence, passing first to the $\limsup_{m\to\infty}$ and then to the $\lim_{\varepsilon\to 0}$ in (12), we get (10). In the case where f_m, f are real-valued, (10) immediately implies (11), because every $F \in C_c(\mathbb{R})$ is obtained as the restriction to $\mathbb{R}$ of some $\tilde{F} \in C_c(\mathbb{C})$. $\square$

Now, let $f : [-\pi, \pi] \to \mathbb{C}$ be a function in $L^1([-\pi, \pi])$, and denote its Fourier coefficients by

$$f_j := \frac{1}{2\pi} \int_{-\pi}^{\pi} f(x) \mathrm{e}^{-\mathrm{i}jx} dx, \qquad j \in \mathbb{Z}.$$

For every $n \geqslant 1$, define the n-th Toeplitz matrix associated with f as follows:

$$T_n(f) := [f_{i-j}]_{i,j=1}^n.$$

Note that $T_n(\cdot)$ is linear: $T_n(\alpha f + \beta g) = \alpha T_n(f) + \beta T_n(g)$ for $\alpha, \beta \in \mathbb{C}$ and $f, g \in L^1([-\pi, \pi])$. In the case where f is real-valued, all the matrices $T_n(f)$ are Hermitian and the following result holds, which is Szegö's theorem for L^∞ functions and is due to Zamarashkin and Tyrtyshnikov [16] and Tilli [14] in the form cited here.

THEOREM 5. *Let f be a real-valued function in $L^1([-\pi, \pi])$, then $\{T_n(f)\}_n \sim_\lambda f$.*

Our goal is to provide a proof of Theorem 5 based on the notion of a.c.s. and, especially, on Theorem 4. To this end, we need some auxiliary lemmas. For any square matrix $X \in \mathbb{C}^{n\times n}$, we denote by $\|X\|_1$ the Schatten 1-norm (or trace-norm) of X, defined as the sum of the singular values of X; see [2] for the Schatten p-norms of matrices. If $f \in L^1([-\pi,\pi])$, we set

$$\|f\|_{L^1([-\pi,\pi])} := \frac{1}{2\pi}\int_{-\pi}^{\pi} |f(x)|dx.$$

LEMMA 3. *Let $f \in L^1([-\pi,\pi])$ and $n \in \mathbb{N}$, then*

$$\|T_n(f)\|_1 \leqslant 2n\|f\|_{L^1([-\pi,\pi])}. \tag{13}$$

Proof. See [14, Lemma 3.1]. $\square$

The inequality (13) is part of a large family of inequalities involving Toeplitz matrices and Schatten p-norms. In particular, in a finer version of (13), the constant 2 is replaced by 1. We refer the interested reader to [12, Corollary 3.5].

LEMMA 4. *Let $\{\{Z_{n,m}\}_n : m\}$ be a sequence of matrix-sequences, with $Z_{n,m}$ of size n, and assume that, for every m,*

$$\|Z_{n,m}\|_1 \leqslant \alpha(m)n, \qquad \forall n \geqslant n_m,$$

where $\alpha(m)$, n_m depend only on m. Then, for each m,

$$Z_{n,m} = R_{n,m} + N_{n,m}, \qquad \forall n \geqslant n_m,$$

where $\mathrm{rank}(R_{n,m}) \leqslant \sqrt{\alpha(m)}\,n$ and $\|N_{n,m}\| \leqslant \sqrt{\alpha(m)}$.

Proof. The thesis may be somehow derived from the results in [11] (see in particular Theorem 4.4 and Corollaries 4.1–4.2). However, since the derivation is not so plain, we include a short and direct proof for the reader's convenience.

Fix m and $n \geqslant n_m$. Since $\|Z_{n,m}\|_1 \leqslant \alpha(m)n$, the number of singular values of $Z_{n,m}$ that exceed $\sqrt{\alpha(m)}$ cannot be larger than $\sqrt{\alpha(m)}\,n$. Let $Z_{n,m} = U_{n,m}\Sigma_{n,m}V_{n,m}^*$ be a singular value decomposition of $Z_{n,m}$ and write

$$Z_{n,m} = U_{n,m}\Sigma_{n,m}V_{n,m}^* = U_{n,m}\Sigma_{n,m}^{(1)}V_{n,m}^* + U_{n,m}\Sigma_{n,m}^{(2)}V_{n,m}^*,$$

where $\Sigma_{n,m}^{(1)}$ is obtained from $\Sigma_{n,m}$ by setting to 0 all the singular values that are less than or equal to $\sqrt{\alpha(m)}$, while $\Sigma_{n,m}^{(2)} := \Sigma_{n,m} - \Sigma_{n,m}^{(1)}$ is obtained from $\Sigma_{n,m}$ by setting to 0 all the singular values that exceed $\sqrt{\alpha(m)}$. Then

$$Z_{n,m} = R_{n,m} + N_{n,m},$$

where $R_{n,m} := U_{n,m}\Sigma_{n,m}^{(1)}V_{n,m}^*$ and $N_{n,m} := U_{n,m}\Sigma_{n,m}^{(2)}V_{n,m}^*$ satisfy $\mathrm{rank}(R_{n,m}) \leqslant \sqrt{\alpha(m)}\,n$ and $\|N_{n,m}\| \leqslant \sqrt{\alpha(m)}$. $\square$

The next lemma shows that Theorem 5 holds in the case where f is a trigonometric polynomial.

LEMMA 5. *Let p be a real-valued trigonometric polynomial, then $\{T_n(p)\}_n \sim_\lambda p$.*

Proof. Let $p(x) := \sum_{j=-s}^{s} p_j \mathrm{e}^{\mathrm{i}jx}$ be a real-valued trigonometric polynomial. Note that $p_{-j} = \overline{p_j}$ for all $j = -s, \ldots, s$, because p is real. For every $n \geqslant 2s+1$, consider the following decomposition of $T_n(p)$:

$$
\begin{aligned}
T_n(p) &= \begin{bmatrix}
p_0 & \cdots & p_{-s} & & & p_s & \cdots & p_1 \\
\vdots & \ddots & & \ddots & & & \ddots & \vdots \\
p_s & & \ddots & & \ddots & & & p_s \\
& \ddots & & \ddots & & \ddots & & \\
& & \ddots & & \ddots & & \ddots & \\
p_{-s} & & & \ddots & & \ddots & & p_{-s} \\
\vdots & \ddots & & & \ddots & & \ddots & \vdots \\
p_{-1} & \cdots & p_{-s} & & & p_s & \cdots & p_0
\end{bmatrix}
-
\begin{bmatrix}
& & & & p_s & \cdots & p_1 \\
& & & & & \ddots & \vdots \\
& & & & & & p_s \\
& & & & & & \\
p_{-s} & & & & & & \\
\vdots & \ddots & & & & & \\
p_{-1} & \cdots & p_{-s} & & & &
\end{bmatrix} \\
&=: C_n(p) - Z_n(p).
\end{aligned} \tag{14}
$$

$C_n(p)$ is a (Hermitian) circulant matrix and hence its eigenvalues are explicitly known (see [3, p. 33] or [5, Section 3.2]):

$$
\lambda_j(C_n(p)) = p\left(\frac{2\pi(j-1)}{n}\right), \qquad j = 1, \ldots, n.
$$

Therefore, for every test function $F \in C_c(\mathbb{R})$, we have

$$
\begin{aligned}
\lim_{n\to\infty} \frac{1}{n}\sum_{j=1}^{n} F(\lambda_j(C_n(p))) &= \lim_{n\to\infty} \frac{1}{n}\sum_{j=0}^{n-1} F\left(p\left(\frac{2\pi j}{n}\right)\right) = \frac{1}{2\pi}\int_0^{2\pi} F(p(x))dx \\
&= \frac{1}{2\pi}\int_{-\pi}^{\pi} F(p(x))dx,
\end{aligned}
$$

where the last equality holds because p is periodic with period 2π, while the second equality is due to the fact that $\frac{2\pi}{n}\sum_{j=0}^{n-1} F(p(\frac{2\pi j}{n}))$ is a Riemann integral sum for $\int_0^{2\pi} F(p(x))dx$ and converges to this integral as $n \to \infty$, because the function $F(p(x))$ is continuous and hence Riemann integrable on $[0, 2\pi]$. Thus, $\{C_n(p)\}_n \sim_\lambda p$.

Now, for every n, m, set $A_n := T_n(p)$ and $B_{n,m} := C_n(p)$. We have just proved that $\{B_{n,m}\}_n \sim_\lambda p$ for every m. All the hypotheses of Theorem 4 are then satisfied (with $\phi_m = \phi = \phi_p$, as given by (9) for $f = p$) if $\{\{B_{n,m}\}_n : m\}$ is an a.c.s. for $\{A_n\}_n$. But this is clearly true, because, in view of (14), for every m we have

$$
A_n = B_{n,m} + R_{n,m} + N_{n,m}, \qquad \forall n \geqslant n_m,
$$

where $N_{n,m}$ is the zero matrix and $R_{n,m} := -Z_n(p)$ satisfies $\mathrm{rank}(R_{n,m}) \leqslant 2s \leqslant \rho(m)n$ for all $n \geqslant n_m$, provided that we choose $n_m = m$ and $\rho(m) = 2s/m$. All the hypotheses of Theorem 4 are then satisfied and so $\{T_n(p)\}_n \sim_\lambda p$. $\square$

Proof of Theorem 5. Take a sequence of real trigonometric polynomials p_m such that $p_m \to f$ in $L^1([-\pi,\pi])$. We prove that the assumptions of Theorem 4 are satisfied with

$$A_n = T_n(f), \qquad B_{n,m} = T_n(p_m), \qquad \phi_m = \phi_{p_m}, \qquad \phi = \phi_p.$$

We first note that $T_n(f)$ and $T_n(p_m)$ are Hermitian, because f and p_m are real. By Lemma 5, for every m we have $\{T_n(p_m)\}_n \sim_\lambda p_m$. By Lemma 2, $\phi_{p_m}(F) \to \phi_p(F)$ for all $F \in C_c(\mathbb{R})$, because $p_m \to f$ in $L^1([-\pi,\pi])$ and hence, a fortiori, $p_m \to f$ in measure. It remains to show that $\{\{T_n(p_m)\}_n : m\}$ is an a.c.s. for $\{T_n(f)\}_n$.

By Lemma 3, for every n,m we have

$$\|T_n(f) - T_n(p_m)\|_1 = \|T_n(f-p_m)\|_1 \leqslant 2n\|f-p_m\|_{L^1([-\pi,\pi])} = \alpha(m)n,$$

where $\alpha(m) := 2\|f-p_m\|_{L^1([-\pi,\pi])}$. Thus, by Lemma 4, for every n,m we have

$$T_n(f) - T_n(p_m) = R_{n,m} + N_{n,m},$$

where $\mathrm{rank}(R_{n,m}) \leqslant \sqrt{\alpha(m)}\,n$ and $\|N_{n,m}\| \leqslant \sqrt{\alpha(m)}$. Since $\alpha(m) \to 0$ as $m \to \infty$, $\{\{T_n(p_m)\}_n : m\}$ is an a.c.s. for $\{T_n(f)\}_n$. The thesis now follows from Theorem 4. $\square$

We conclude by saying that a completely analogous proof as the one presented in this section can be given also for the multilevel block version of the Szegö theorem [14, Theorem 2]. Here, we decided to address only the monolevel scalar case in order to avoid technicalities and notational complications, so as to make more clear the ‘a.c.s. idea’ and the way in which Theorem 4 is applied in practice.

REFERENCES

[1] O. AXELSSON, G. LINDSKOG, *On the rate of convergence of the preconditioned conjugate gradient method*, Numer. Math. **48** (1986) 499–523.
[2] R. BHATIA, *Matrix analysis*, Springer-Verlag, New York (1997).
[3] A. BÖTTCHER, S. GRUDSKY, *Spectral properties of banded Toeplitz matrices*, SIAM, Philadelphia (2005).
[4] A. BÖTTCHER, B. SILBERMANN, *Introduction to large truncated Toeplitz matrices*, Springer-Verlag, New York (1999).
[5] P. J. DAVIS, *Circulant matrices*, 2nd Edition, AMS Chelsea Publishing (1994).
[6] C. GARONI, S. SERRA-CAPIZZANO, D. SESANA, *Tools for determining the asymptotic spectral distribution of non-Hermitian perturbations of Hermitian matrix-sequences and applications*, Integr. Equat. Oper. Theory **81** (2015), 213–225.
[7] G. H. GOLUB, C. F. VAN LOAN, *Matrix computations*, 3rd Edition, The Johns Hopkins University Press (1996).
[8] U. GRENANDER, G. SZEGÖ, *Toeplitz forms and their applications*, 2nd Edition, Chelsea, New York (1984).
[9] W. RUDIN, *Real and complex analysis*, 3rd Edition, McGraw-Hill (1987).
[10] S. SERRA-CAPIZZANO, *Distribution results on the algebra generated by Toeplitz sequences: a finite dimensional approach*, Linear Algebra Appl. **328** (2001) 121–130.
[11] S. SERRA-CAPIZZANO, *Spectral behavior of matrix sequences and discretized boundary value problems*, Linear Algebra Appl. **337** (2001) 37–78.
[12] S. SERRA-CAPIZZANO, *More inequalities and asymptotics for matrix valued linear positive operators: the noncommutative case*, Oper. Theory Adv. Appl. **135** (2002) 293–315.

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

[14] P. TILLI, *A note on the spectral distribution of Toeplitz matrices*, Linear and Multilinear Algebra **45** (1998) 147–159.

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

[16] N. L. ZAMARASHKIN, E. E. TYRTYSHNIKOV, *Distribution of eigenvalues and singular values of Toeplitz matrices under weakened conditions on the generating function*, Sb. Math. **188** (1997) 1191–1201.

(Received November 4, 2014)

Carlo Garoni
University of Insubria
Department of Science and High Technology
Via Valleggio 11, 22100 Como, Italy
e-mail: carlo.garoni@uninsubria.it

Stefano Serra-Capizzano
University of Insubria
Department of Science and High Technology
Via Valleggio 11, 22100 Como, Italy
and
Division of Scientific Computing
Department of Information Technology
Uppsala University
Box 337, SE-751 05 Uppsala, Sweden
e-mail: stefano.serrac@uninsubria.it, stefano.serra@it.uu.se

Paris Vassalos
Athens University of Economics and Business
Department of Informatics
GR10434 Greece
e-mail: pvassal@aueb.gr