The eigenvalue distribution of products of Toeplitz matrices - clustering and attraction

Stefano Serra-Capizzano, Debora Sesana, Elizabeth Strouse

The eigenvalue distribution of products of Toeplitz matrices - clustering and attraction

Stefano Serra-Capizzano, Debora Sesana, and Elizabeth Strouse [1]

Abstract

We use a recent result concerning the eigenvalues of a generic (non Hermitian) complex perturbation of a bounded Hermitian sequence of matrices to prove that the asymptotic spectrum of the product of Toeplitz sequences, whose symbols have a real-valued essentially bounded product h, is described by the function h in the "Szegö way". Then, using Mergelyan's theorem, we extend the result to the more general case where h belongs to the Tilli class. The same technique gives us the analogous result for sequences belonging to the algebra generated by Toeplitz sequences, if the symbols associated with the sequences are bounded and the global symbol h belongs to the Tilli class. A generalization to the case of multilevel matrix-valued symbols and a study of the case of Laurent polynomials not necessarily belonging to the Tilli class are also given.

Key words: matrix sequence, joint eigenvalue distribution, Toeplitz matrix
AMS Classification (2000): 15A18, 15A12, 47B36, 47B65

1 Introduction and basic notations

Let $L^2(\mathbb{T})$ be the usual Hilbert space of square-integrable functions on the circle $\mathbb{T} = \{z \in \mathbb{C} : |z| = 1\}$ and let H^2 be the Hardy space composed of those functions in $L^2(\mathbb{T})$ whose negative Fourier coefficients are equal to zero. The Toeplitz operator with "symbol" the function f is the operator $T_f : H^2 \longrightarrow H^2$ defined by $T_f(g) = P(fg)$ where P is orthogonal projection from L^2 to H^2. Such an operator is bounded if and only if $f \in L^\infty(\mathbb{T})$ = the space of (essentially) bounded functions on the circle, and its infinite matrix $T(f)$ in the canonical orthonormal basis $\mathcal{B} = \{1, z, z^2, \ldots\}$ is constant along the diagonals, that is, it is of the form $T(f) = [\hat{f}_{j-r}]_{r,j=1}^{\infty}$, $\hat{f}_k$ being the Fourier coefficients of f defined by equation (8) in Section 1.2; see [9].

Now, let f be any integrable function on $\mathbb{T}$ and, for each $n \in \mathbb{N}$, let $T_n(f)$ be the $n \times n$ matrix $[\hat{f}_{j-r}]_{r,j=1}^{n}$. The sequence of operators on H^2 associated with the sequence $\{T_n(f)\}_{n=1}^{\infty}$ is an obvious approximating sequence for the Toeplitz operator T_f when $f \in L^\infty(\mathbb{T})$ and so we call $\{T_n(f)\}_{n=1}^{\infty}$ a *Toeplitz sequence*. It is natural to ask how the spectrum or set of eigenvalues $\Lambda_n = \{\lambda_1, \lambda_2, \ldots, \lambda_n\}$ of $T_n(f)$ is related to the spectrum of $T(f)$ if $f \in L^\infty(\mathbb{T})$ or, even if $f \in L^1(\mathbb{T})$, to study the "convergence" of the sequence of sets $\{\Lambda_n\}_{n=1}^{\infty}$ (or that of the sequence $\{\Gamma_n\}_{n=1}^{\infty}$ where Γ_n is the set of singular values of the matrix $T_n(f)$). An essential result concerning the sequence of sets of eigenvalues is the famous Szegö theorem which says that, if f is real-valued and essentially bounded then

$$\lim_{n\to\infty} \frac{1}{n} \sum_{\lambda \in \Lambda_n} F(\lambda) = \frac{1}{2\pi} \int_{[-\pi,\pi]} F(f(\exp(it)))dt, \qquad i^2 = -1, \tag{1}$$

for every continuous function F with compact support (see, for example, [14]).

In the 1990's, independently, Tilli and Tyrtyshnikov/Zamarashkin showed [35, 38] that equation (1) holds for any integrable real-valued function f. The corresponding result for a

[1] The work of the first and of the second author is supported in part by MIUR grant no. 2006017542

complex-valued function f and the sequence of sets of its singular values (with $|f|$ in the place of f) was first obtained by Parter (continuous times uni-modular symbols [20]), Avram (essentially bounded symbols [1]), and by Tilli and Tyrtyshnikov/Zamarashkin [35, 38], independently, when the symbol f is just integrable. (The book [8] gives a synopsis of all these results in Chapters 5 and 6 and other interesting facts in Chapter 3 concerning the relation between the *pseudospectrum* of $\{T_n(f)\}_{n=1}^{\infty}$ and that of $T(f)$).

The relation (1) was established for a more general class of test functions F in [35, 26, 6] and the case of functions f of several variables (the *multilevel* case) and matrix-valued functions was studied in [35] and in [23] in the context of preconditioning (other related results were established by Linnik, Widom, Doktorski, see Section 6.9 in [8]).

Further extensions of the Szegö result can be considered. An important direction of research is represented by the case of variable Toeplitz sequences or generalized locally Toeplitz sequences (see the pioneering work by Kac, Murdoch and Szegö [17] and by Parter [21], and, more recently, papers [34, 27, 28, 32, 4]). Another important direction is represented by the algebra generated by Toeplitz sequences and this is the main subject of this note, with special attention to the case of eigenvalues, the case of singular values being already known (see [8, 28] and references therein).

However we have to be careful: a simple yet striking example where the eigenvalue result does *not* hold is given by the Toeplitz sequence related to the function $\exp(-it)$, $i^2 = -1$, which has only zero eigenvalues so that the condition (1) means that $F(0) = \frac{1}{2\pi}\int_{[-\pi,\pi]} F(\exp(it))dt$ which is far from being satisfied for all continuous functions with compact support, even though it *is* satisfied for harmonic functions (in cases like this one it is better to consider the pseudospectrum, see [8]). In fact, Tilli was able to show that, if f is any complex-valued integrable function, then the condition (1) holds for all harmonic test functions F [33] and that it is even satisfied by all continuous functions with compact support as long as the symbol f satisfies a certain geometric condition. More specifically, the function f must be essentially bounded and such that its (essential) range does not disconnect the complex plane and has empty interior, see [36]. We call this set of functions the *Tilli class*. In other contexts such a property is informally called "thin spectrum". It is clear that the set of all real-valued L^∞ functions is properly included in the Tilli class. In the final part of this paper, we will discuss some interesting relationships between the Tilli class and the restrictions to $\mathbb{T}^d$, $d \geq 1$, of the Hardy space H^∞.

However we should emphasize that the importance of thin spectrum was already known in the operator theory community before the work by Tilli. This is clear from example 5.39 in [8] and from paper [39], see the top of page 390. In fact, many of the results above have appeared in different forms in different articles using different proofs. One approach uses operator theory (a summary and guide to which can be found in [8] and [5]) and another a lot of basic linear algebra techniques. In particular the proofs and the approach introduced by Tilli are extremely clean as observed in the MathSciNet review of [36] by Estelle Basor. Indeed an interesting aspect of papers [37, 38, 36, 27, 28] relies on the fact that the tools are essentially all from finite dimensional linear algebra and therefore are perhaps more accessible to engineers who are very interested in such questions (see [13]).

In reality, as a matter of fact, the different communities often find it difficult to communicate with each other. Hence one of our objectives in this paper is to provide an approach which is accessible to both groups - and also (to some extent) to parts of the nonmathematical community.

In this article we study the asymptotic spectral behavior of a product of Toeplitz sequences (in the usual, matrix valued, and multilevel cases), by using and extending tools from matrix theory and finite dimensional linear algebra.

It is well known that the product of Toeplitz operators is rarely equal to a Toeplitz operator (see [9] and work by the third author in [19]), but, it turns out that the sequence of eigenvalues or singular values of the product of two Toeplitz sequences is often related to the product of the two symbols in a Szegö-type way. For the singular values the result is known as long as all the involved symbols are essentially bounded and, in fact, for any linear combination of products of Toeplitz operators, the distribution function is exactly the linear combination of the products of the symbols of the sequences: the latter goes back to the work of Roch and Silbermann (see Sections 4.6 and 5.7 in [8]). The previous results have been extended by considering integrable symbols, not necessarily bounded [27, 28], and by considering (pseudo) inversion and the related algebra of sequences (see [28, 30]). Of course for the eigenvalues much less is known, and one simple reason is that much less is true, as another basic example discussed at the beginning of Section 2 in [29] shows. However, quite recently, using the Ky Fan-Mirski theorem which says that the real (or imaginary) parts of the eigenvalues are majorized by the eigenvalues of the real (or imaginary) part of the matrix (see [2]), the first author obtained a method for deducing the eigenvalue distribution of sequences obtained as generic perturbations of Hermitian sequences, when the trace norm of the perturbation is asymptotically negligible with respect to size of the involved matrices (see Theorem 2.3, Theorem 2.4 and [12]). We recall that a real vector v of size n is said to be majorized by a real vector w of the same size if, for each k, the sum of the largest k entries of v is bounded by the sum of the k largest entries of w and equality holds for $k = n$.

By using Lemma 3.2 in [12], Golinskii and the first author proved that the eigenvalues of a non Hermitian complex perturbation of a Jacobi matrix sequence, which are not necessarily real, are still distributed as the real-valued function $2\cos t$ on $[0, \pi]$, which characterizes the non-perturbed case where the Jacobi sequence is of course real and symmetric: see [12], and [16, 28] for a further application of Lemma 3.2 in [12] to a (pseudo) differential setting. In this paper, we apply these results to certain products of Toeplitz sequences, then discuss, apply and extend more general tools introduced by Tilli [36] and based on the Mergelyan theorem, see [22]. Furthermore, the case of Laurent polynomials not necessarily in the Tilli class is sketched and a generalization to the case of multilevel Toeplitz sequences and sequences $T_n(f)$ where f is a matrix-valued function is also given: we have to emphasize that these multilevel and matrix-valued extensions are of interest in the Engineering context where the number of levels refers to multiple inputs (Multi-Input systems) and size of the basic blocks, i.e., the size of the matrix-valued symbol refers to multiple outputs (Multi-Output systems). Following the Engineering terminology, we are talking of SIMO and MIMO systems, see e.g. [11, 15] for details and references therein.

We proceed as follows. In Section 2 we discuss some relationships between the notions of distribution in the sense of eigenvalues and clustering/attracting properties of matrix sequences. In particular, Theorems 2.3 and 2.4 give tools for working with non-Hermitian perturbations of Hermitian matrix sequences. In Section 3, as a straightforward consequence of these results, we obtain the distribution of the eigenvalues of Toeplitz sequences products when the linear combination of the products of the symbols is a real-valued (Hermitian-valued) essentially bounded function. Finally, in Section 4 we introduce some tools based on the Mergelyan theorem and use them in Section 5 to deal with more complicated cases, that of the Tilli class and of sequences belonging to the algebra generated by Toeplitz sequences, when the global symbol lies in the Tilli class. A generalization to the case of matrix-valued symbols is also given together with a more specific study in the case of Laurent polynomials. The conclusion in Section 6 ends the paper.

1.1 Basic notations

We begin with some formal definitions. A square complex matrix A always can be uniquely written as a Hermitian matrix plus a skew-Hermitian matrix (in analogy to case of scalar complex numbers). More precisely, by defining A^* the complex conjugate and transpose of the matrix A, we have

$$\begin{aligned} A &= \mathrm{Re}(A) + i\ \mathrm{Im}(A), \qquad i^2 = -1, \\ \mathrm{Re}(A) &= (A + A^*)/2, \\ \mathrm{Im}(A) &= (A - A^*)/(2i), \end{aligned}$$

where $\mathrm{Re}(A)$ and $\mathrm{Im}(A)$ are Hermitian matrices so that $i\ \mathrm{Im}(A)$ is skew-Hermitian. Moreover, given any Hermitian matrix B, by Schur (see [2]), we find $B = U\mathrm{diag}(d_1, \ldots, d_n)U^*$, where every d_j, $j = 1, \ldots, n$, is real and U is unitary. Then the Hermitian positive semi-definite matrices B^+ and B^- are defined as

$$B^+ = U\mathrm{diag}(d_1^+, \ldots, d_n^+)U^*, \qquad B^- = U\mathrm{diag}(d_1^-, \ldots, d_n^-)U^*,$$

with $d^+ = \max\{0, d\}$, $d^- = \max\{0, -d\}$, and

$$B = B^+ - B^-.$$

For A an $n \times n$ matrix over $\mathbb{C}$ with singular values $\sigma_1(A), \ldots, \sigma_n(A)$, and $p \in [1, \infty]$ we define $\|A\|_p$, the Schatten p-norm of A to be the ℓ^p norm of the vector of the singular values

$$\|A\|_p = \left[\sum_{k=1}^{n} (\sigma_k(A))^p\right]^{\frac{1}{p}}.$$

We will be especially interested in the norm $\|\cdot\|_1$ which is known as the trace norm, and the norm $\|\cdot\|_\infty$ which is equal to the usual operator norm $\|\cdot\|$

$$\|A\| = \sup_{\|x\|=1} \|Ax\|.$$

For any $n \times n$ matrix A with eigenvalues $\lambda_j(A)$, $j = 1, \ldots, n$, we set

$$\Lambda_n = \{\lambda_1(A), \lambda_2(A), \ldots, \lambda_n(A)\}.$$

Then, for any function F defined on $\mathbb{C}$, the symbol $\Sigma_\lambda(F, A)$ stands for the mean

$$\Sigma_\lambda(F, A) := \frac{1}{n} \sum_{j=1}^{n} F\left(\lambda_j(A)\right) = \frac{1}{n} \sum_{\lambda \in \Lambda_n} F(\lambda),$$

and the symbol $\Sigma_\sigma(F, A)$ denotes the corresponding expression with the singular values replacing the eigenvalues. Throughout this paper we speak of *matrix sequences* as sequences $\{A_n\}$ where A_n is an $n \times n$ matrix and *Toeplitz sequences* as matrix sequences of the form $\{A_n\}$ with $A_n = T_n(f)$ and

$$T_n(f) = [\hat{f}_{j-r}]_{r,j=1}^{n},$$

where f is an integrable function and $\hat{f}_k$ are the Fourier coefficients of f defined by equation (8).

The following definition is motivated by the Szegö and Tilli theorems characterizing the spectral approximation of a Toeplitz operator (in certain cases) by the spectra of the elements of the natural approximating matrix sequence A_n, where A_n is formed by the first n rows and columns of the matrix representation of the operator.

Definition 1.1. Let $\mathcal{C}_0(\mathbb{C})$ be the set of continuous functions with bounded support defined over the complex field, d a positive integer, and θ a complex-valued measurable function defined on a set $G \subset \mathbb{C}^d$ of finite and positive Lebesgue measure $m(G)$. Here G will be equal to $\mathbb{T}^d$. A matrix sequence $\{A_n\}$ is said to be *distributed* (*in the sense of the eigenvalues*) *as the pair* (θ, G), or to *have the distribution function* θ, if, $\forall F \in \mathcal{C}_0(\mathbb{C})$, the following limit relation holds

$$\lim_{n\to\infty} \Sigma_\lambda(F, A_n) = \frac{1}{m(G)} \int_G F(\theta(t))\, dt. \tag{2}$$

Whenever (2) holds $\forall F \in \mathcal{C}_0(\mathbb{C})$ we say that $\{A_n\} \sim_\lambda (\theta, G)$.

If (2) holds for every $F \in \mathcal{C}_0(\mathbb{R}_0^+)$ in place of $F \in \mathcal{C}_0(\mathbb{C})$, with the singular values $\sigma_j(A_n)$, $j = 1, \ldots, n$, in place of the eigenvalues, and with $|\theta(t)|$ in place of $\theta(t)$, we say that $\{A_n\} \sim_\sigma (\theta, G)$ or that the matrix sequence $\{A_n\}$ is *distributed* (*in the sense of the singular values*) *as the pair* (θ, G): more specifically for every $F \in \mathcal{C}_0(\mathbb{R}_0^+)$ we have

$$\lim_{n\to\infty} \Sigma_\sigma(F, A_n) = \frac{1}{m(G)} \int_G F(|\theta(t)|)\, dt, \tag{3}$$

with

$$\Sigma_\sigma(F, A_n) := \frac{1}{n} \sum_{j=1}^{n} F\,(\sigma_j(A_n)).$$

Furthermore, in order to treat block Toeplitz matrices, we consider measurable functions $\theta : G \to \mathcal{M}_N \equiv \mathcal{M}_{NN}$, where $\mathcal{M}_{NM}$ is the space of $N \times M$ matrices with complex entries and a function is considered to be measurable if and only if the component functions are. In this case $\{A_n\} \sim_\lambda (\theta, G)$ means that $M = N$ and

$$\lim_{n\to\infty} \Sigma_\lambda(F, A_n) = \frac{1}{m(G)} \int_G \frac{\sum_{j=1}^{N} F\,(\lambda_j(\theta(t)))}{N}\, dt, \tag{4}$$

$\forall F \in \mathcal{C}_0(\mathbb{C})$, where the $\lambda_i(\theta(t))$ in equation (4) are the eigenvalues of the matrix $\theta(t)$. When considering θ taking values in $\mathcal{M}_{NM}$, we say that $\{A_n\} \sim_\sigma (\theta, G)$ when for every $F \in \mathcal{C}_0(\mathbb{R}_0^+)$ we have

$$\lim_{n\to\infty} \Sigma_\sigma(F, A_n) = \frac{1}{m(G)} \int_G \frac{\sum_{j=1}^{\min\{N,M\}} \left(F(\lambda_j(\sqrt{\theta(t)\theta^*(t)}))\right)}{\min\{N, M\}}\, dt.$$

Finally we say that two sequences $\{A_n\}$ and $\{B_n\}$ are *equally distributed* in the sense of eigenvalues (λ) or in the sense of singular values (σ) if, $\forall F \in \mathcal{C}_0(\mathbb{C})$, we have

$$\lim_{n\to\infty} [\Sigma_\nu(F, B_n) - \Sigma_\nu(F, A_n)] = 0, \qquad \text{with} \quad \nu = \lambda \text{ or } \nu = \sigma.$$

Notice that two sequences having the same distribution function are equally distributed. On the other hand, two equally distributed sequences may not be associated with a distribution function at all: consider any diagonal matrix sequence $\{A_n\}$ and let $\{B_n\}$ be a sequence of the form $B_n = A_n + \epsilon_n I_n$ with $\epsilon_n \to 0$ when $n \to \infty$. Then, if the original $\{A_n\}$ does not have an eigenvalue distribution function (e.g. $A_n = (-1)^n I_n$), we will have $\{A_n\}$ and $\{B_n\}$ equally distributed, even though it is impossible to associate a distribution function with either of them. On the other hand, if one of them has a distribution function, then the other necessarily has the same one. This is easy to show using the definitions (or see [25, Remark 6.1]).

Now, notice that a matrix sequence $\{A_n\}$ is distributed as the pair (θ, G) if and only if the sequence of linear functionals $\{\phi_n\}$ defined by $\phi_n(F) = \sum_\lambda(F, A_n)$ converges weak-* to the functional $\phi(F) = \frac{1}{m(G)} \int_G F(\theta(t))\, dt$ as in (2). In order to describe what this really means about the asymptotic qualities of the spectrum, we will derive more concrete characterizations of $\{\Lambda_n\}$ such as "clustering" and "attraction", where, as above, Λ_n is the set of eigenvalues of A_n.

Definition 1.2. A matrix sequence $\{A_n\}$ is *strongly clustered at* $s \in \mathbb{C}$ (in the eigenvalue sense), if for any $\varepsilon > 0$ the number of the eigenvalues of A_n off the disc

$$D(s, \varepsilon) := \{z : |z - s| < \varepsilon\}, \tag{5}$$

can be bounded by a pure constant q_ε possibly depending on ε, but not on n. In other words

$$q_\varepsilon(n, s) := \#\{j : \lambda_j(A_n) \notin D(s, \varepsilon)\} = O(1), \quad n \to \infty.$$

If every A_n has only real eigenvalues (at least for large n) then we may assume that s is real and that the disc $D(s, \varepsilon)$ is the interval $(s - \varepsilon, s + \varepsilon)$. A matrix sequence $\{A_n\}$ is said to be *strongly clustered at a nonempty closed set* $S \subset \mathbb{C}$ (in the eigenvalue sense) if for any $\varepsilon > 0$

$$q_\varepsilon(n, S) := \#\{j : \lambda_j(A_n) \notin D(S, \varepsilon)\} = O(1), \quad n \to \infty, \tag{6}$$

where $D(S, \varepsilon) := \cup_{s \in S} D(s, \varepsilon)$ is the ε-neighborhood of S. If every A_n has only real eigenvalues, then S is a nonempty closed subset of $\mathbb{R}$. We replace the term "strongly" by "weakly", if

$$q_\varepsilon(n, s) = o(n), \qquad \big(q_\varepsilon(n, S) = o(n)\big), \quad n \to \infty,$$

in the case of a point s or a closed set S. Finally, if we replace eigenvalues with singular values, we obtain all the corresponding definitions for singular values.

It is clear that $\{A_n\} \sim_\lambda (\theta, G)$, with $\theta \equiv s$ equal to a constant function if and only if $\{A_n\}$ is weakly clustered at $s \in \mathbb{C}$ (for more results and relations between the notions of equal distribution, equal localization, spectral distribution, spectral clustering etc., see [25, Section 4]). We introduce one more notion concerning the eigenvalues of a matrix sequence.

Definition 1.3. Let $\{A_n\}$ be a matrix sequence and let Λ_n be the set of eigenvalues of the matrix A_n. We say that $\{A_n\}$ is *strongly attracted by* $s \in \mathbb{C}$ if

$$\lim_{n\to\infty} \operatorname{dist}(s, \Lambda_n) = 0, \tag{7}$$

where $\operatorname{dist}(X, Y)$ is the usual Euclidean distance between two subsets X and Y of the complex plane. Furthermore, if we order the eigenvalues according to their distance from s, i.e.,

$$|\lambda_1(A_n) - s| \le |\lambda_2(A_n) - s| \le \cdots \le |\lambda_n(A_n) - s|,$$

then we say that the attraction to s is of order $r(s) \in \mathbb{N}$, $r(s) \ge 1$ is a fixed number, if

$$\lim_{n\to\infty} |\lambda_{r(s)}(A_n) - s| = 0, \qquad \liminf_{n\to\infty} |\lambda_{r(s)+1}(A_n) - s| > 0,$$

and that the attraction is of order $r(s) = \infty$ if

$$\lim_{n\to\infty} |\lambda_j(A_n) - s| = 0,$$

for every fixed j. Finally, one defines weak attraction by replacing lim with lim inf in (7).

It is not hard to see that, if $\{A_n\}$ is at least weakly clustered at a point s, then s strongly attracts $\{A_n\}$ with infinite order. Indeed, if there is an attraction of finite order to s then

$$\lim_{n\to\infty} \frac{\#\{\lambda \in \Lambda_n : \ \lambda \notin D(s,\delta)\}}{n} = 1,$$

for some $\delta > 0$ and this is impossible if $\{A_n\}$ is weakly clustered at s. On the other hand, there are sequences which are strongly attracted by s with infinite order, but not even weakly clustered at s. Indeed, the notion of weak clustering does not tell anything concerning weak attraction or attraction of finite order.

Remark 1.4. It is easy to see that any of the notions introduced in this section for eigenvalues has a natural analogue for singular values, as explicitly described for the concept of distribution in (2) and (3).

1.2 Toeplitz sequences: definition and previous distribution results

Let f be an integrable function on $\mathbb{T}^d$ the d-fold Cartesian product of the unit circle in the complex plane. The Fourier coefficients of f are given by:

$$\hat{f}_j = \hat{f}_{(j_1,\dots,j_d)}(f) = \frac{1}{(2\pi)^d} \int_{[-\pi,\pi]^d} f(e^{it_1},\dots,e^{it_d}) \exp(-i(j_1t_1 + \cdots + j_dt_d))\, dt_1 \cdots dt_d, \qquad (8)$$

for integers j_ℓ such that $-\infty < j_\ell < \infty$ for $1 \le \ell \le d$. If f is a matrix-valued function of d variables whose component functions are all integrable, then the $(j_1,\dots,j_d)$-th Fourier coefficient is considered to be the matrix whose (r,s)-th entry is the $(j_1,\dots,j_d)$-th Fourier coefficient of the function $[f(e^{it_1},\dots,e^{it_d})]_{r,s}$.

In the following, for the sake of readability, we shall often write n for the d-tuple $(n_1,\dots,n_d)$, $\hat{n} = n_1 \cdots n_d$, z_r for the function $\exp(it_r)$, and z^j, $j = (j_1,\dots,j_d)$, for the monomial $z_1^{j_1} \cdots z_d^{j_d}$. We write $n \to \infty$ to indicate that $\min_{1\le r\le d} n_r \to \infty$.

Now, for $f : \mathbb{T}^d \to \mathcal{M}_{MN}$, we define the $M\hat{n} \times N\hat{n}$ *multilevel Toeplitz matrix* as in [35] by

$$T_n(f) = \sum_{j_1=-n_1+1}^{n_1-1} \cdots \sum_{j_d=-n_d+1}^{n_d-1} J_{n_1}^{(j_1)} \otimes \cdots \otimes J_{n_d}^{(j_d)} \otimes \hat{f}_{(j_1,\dots,j_d)}(f), \qquad (9)$$

where $\otimes$ denotes the tensor or Kronecker product of matrices and $J_m^{(\ell)}$, $(-m+1 \le \ell \le m-1$, is the $m \times m$ matrix whose (i,j)th entry is 1 if $i - j = \ell$ and 0 otherwise; thus $\{J_{-m+1},\dots,J_{m-1}\}$ is the natural basis for the space of $m \times m$ Toeplitz matrices. In the usual multilevel indexing language, we say that $[T_n(f)]_{r,j} = \hat{f}_{j-r}$ where $(1,\dots,1) \le j, r \le n = (n_1,\dots,n_d)$, i.e., $1 \le j_\ell \le n_\ell$ for $1 \le \ell \le d$. To translate the multilevel notation into the usual notation we convert the pair (j,r) of $\hat{n}$-tuples into the pair (t,s) of positive integers using the formulas below:

$$s = (r_1 - 1)\frac{\hat{n}}{n_1} + (r_2 - 1)\frac{\hat{n}}{n_1n_2} + \cdots + (r_{d-1} - 1)\frac{\hat{n}}{n_1 \cdots n_{d-1}} + r_d,$$

$$t = (j_1 - 1)\frac{\hat{n}}{n_1} + (j_2 - 1)\frac{\hat{n}}{n_1 n_2} + \cdots + (j_{d-1} - 1)\frac{\hat{n}}{n_1 \cdots n_{d-1}} + j_d.$$

Operator theorists probably prefer to interpret this matrix in terms of the usual Toeplitz operator on H^2 of the polydisc. To do this, we let E_n be the subspace of H^2 spanned by the set of monomials of degree "less than" z^n, that is:

$$E_n = \mathrm{span}\{z^j\}_{0 \le j_1 \le n_1 - 1, \ldots, 0 \le j_d \le n_d - 1}.$$

Then, if P_n is orthogonal projection from H^2 onto E_n and we define $T_n^f(g) = P_n(fg)$ from E_n to E_n it is not hard to see that $T_n^f = P_n T_f P_n$ where T_f is the usual Toeplitz operator, and that $T_n(f)$ is the matrix of T_n^f in the basis

$$\{z^j : \ j = (j_1, \ldots, j_d), j_1 = 0, \ldots, n_1 - 1; \ldots ; j_d = 0, \ldots, n_d - 1\}.$$

This is the perspective that we shall use in proving Lemma 2.5. See [35] for a detailed matrix definition and [8] for an explanation with examples in the case $d = 2$.

For the sake of clarity, whenever the extension from scalar to matrix valued functions is simple enough, we shall prove our theorems (especially the ones concerning multilevel Toeplitz) only in the case $M = N = 1$.

The asymptotic distribution of eigenvalues and singular values of a sequence of Toeplitz matrices has been deeply studied in the last century, and strictly depends on the generating function f (see, for example, [8, 38, 35] and references therein). Now, let $\{f_{\alpha,\beta}\}$ be a finite set of $L^1(\mathbb{T}^d)$ functions and define the measurable function h by:

$$h = \textstyle\sum_{\alpha=1}^{\rho} \prod_{\beta=1}^{q_\alpha} f_{\alpha,\beta}^{s(\alpha,\beta)}, \qquad s(\alpha,\beta) \in \{\pm 1\}, \tag{10}$$

where $f_{\alpha,\beta}$ is sparsely vanishing (i.e., the Lebesgue measure of the set of its zeros is zero) when $s(\alpha,\beta) = -1$. The function h may not belong to L^1 in which case $\{T_n(h)\}$ is not defined according to the rule in (9) simply because the formula (8) is not well-defined. However we can still consider the sequence of matrices $\{\sum_{\alpha=1}^{\rho} \prod_{\beta=1}^{q_\alpha} T_n^{s(\alpha,\beta)}(f_{\alpha,\beta})\}$. In [30] it has been proved that

$$\{\textstyle\sum_{\alpha=1}^{\rho} \prod_{\beta=1}^{q_\alpha} T_n^{s(\alpha,\beta)}(f_{\alpha,\beta})\} \sim_\sigma (h, \mathbb{T}^d),$$

and

$$\{\textstyle\sum_{\alpha=1}^{\rho} \prod_{\beta=1}^{q_\alpha} T_n^{s(\alpha,\beta)}(f_{\alpha,\beta})\} \sim_\lambda (h, \mathbb{T}^d),$$

if the matrices $\sum_{\alpha=1}^{\rho} \prod_{\beta=1}^{q_\alpha} T_n^{s(\alpha,\beta)}(f_{\alpha,\beta})$ are Hermitian, at least for n large enough (which implies necessarily that $N = M$). In this context, the symbol $T_n^{s(\alpha,\beta)}(f_{\alpha,\beta})$ with $s(\alpha,\beta) = -1$ and $f_{\alpha,\beta}$ sparsely vanishing means that we are (pseudo) inverting the matrix in the sense of Moore-Penrose (see [2]), since $T_n(f_{\alpha,\beta})$ is not necessarily invertible, but the number of zero singular values is at most $o(\hat{n})$, for $n \to \infty$.

Notice that in defining the symbol h when matrix-valued symbols are involved, it is necessary to consider compatible dimensions and also one has to be careful in respecting the correct ordering in the products, owing to the lack of commutativity in the matrix context.

When $\rho = 1$, $N = M = 1$, and $q_1 = 1$ this result concerns standard Toeplitz sequences and is attributed to Tyrtyshnikov, Zamarashkin, and Tilli [37, 38, 35]; see also [24] and references therein for the evolution of the subject. The case where $s(\alpha,\beta) = 1$ for every α and β is considered and solved in [24] by using matrix theory techniques. We stress that the Hermitian case where h is defined as in (10) has been treated in two different ways in [30] and in [28], for both singular values and eigenvalues. In this paper we are interested in the more difficult eigenvalue setting, when Hermitianity is lost.

Remark 1.5. It should be noted that, according to [28], the distribution result for singular values holds for any sequence belonging to the algebra generated by Toeplitz sequences with $L^1(\mathbb{T}^d)$ symbols, where the allowed algebraic operations are linear combination, product, and (pseudo) inversion. In order to formally define this algebra $\mathcal{A}_T$ we say that $\mathcal{A}_T = \bigcup_{j=0}^{\infty} \mathcal{A}_T^{(j)}$ where Toeplitz sequences with $L^1(\mathbb{T}^d)$ symbols form the set $\mathcal{A}_T^{(0)}$ and $\{A_n\} \in \mathcal{A}_T^{(j)}$, $j \geq 1$, if there exists a finite set of sequences $\{A_n^{(\alpha,\beta)}\}$ with measurable symbols $f_{\alpha,\beta}$ belonging to $\mathcal{A}_T^{(k)}$, $0 \leq k < j$, such that

$$A_n = \sum_{\alpha=1}^{\rho} \prod_{\beta=1}^{q_\alpha} \left(A_n^{(\alpha,\beta)}\right)^{s(\alpha,\beta)}, \qquad s(\alpha,\beta) \in \{\pm 1\},$$

where every sequence which is (pseudo) inverted ($s(\alpha,\beta) = -1$) should have sparsely vanishing symbol; the new symbol of $\{A_n\}$ is recursively defined as

$$h = \sum_{\alpha=1}^{\rho} \prod_{\beta=1}^{q_\alpha} f_{\alpha,\beta}^{s(\alpha,\beta)}, \qquad s(\alpha,\beta) \in \{\pm 1\}.$$

The general result in [28] is that $\{A_n\} \sim_\sigma (h, \mathbb{T}^d)$ and $\{A_n\} \sim_\lambda (h, \mathbb{T}^d)$, if all the matrices A_n are Hermitian, at least for n large enough.

Finally it is worth mentioning that the above results also hold when starting from the set of block multilevel sequences generated by matrix-valued $N \times M$ symbols; see [28] for general integrable symbols (i.e. all the singular values of the symbol are integrable on $\mathbb{T}^d$) and [7] for the case of bounded symbols with $M = N$ and without pseudo inversion, but where the distribution result for eigenvalues is extended to the case in which the involved sequences are normal (the Hermitian case for general integrable symbols and with pseudo inversion can be found in [28]).

2 Eigenvalue distribution and clustering

Let us recall the notion of the essential range which plays an important role in the study of the asymptotic properties of the spectrum.

Definition 2.1. Given a measurable complex-valued function θ defined on a Lebesgue measurable set G, the *essential range of* θ is the set $\mathcal{S}(\theta)$ of points $s \in \mathbb{C}$ such that, for every $\varepsilon > 0$, the Lebesgue measure of the set $\theta^{(-1)}(D(s,\varepsilon)) := \{t \in G : \ \theta(t) \in D(s,\varepsilon)\}$ is positive, with $D(s,\varepsilon)$ as in (5). The function θ is *essentially bounded* if its essential range is bounded. Furthermore, if θ is real-valued, then the essential supremum (infimum) is defined as the supremum (infimum) of its essential range. Finally if the function θ is $N \times N$ matrix-valued and measurable, then the essential range of θ is the union of the essential ranges of the complex-valued eigenvalues $\lambda_j(\theta)$, $j = 1, \ldots, N$.

We note that $\mathcal{S}(\theta)$ is clearly a closed set - it's easy to see that its complement is open. Next we discuss the relationship between the notions introduced in the last section and the essential range.

Theorem 2.2. *Let θ be a measurable function defined on G with finite and positive Lebesgue measure, and $\mathcal{S}(\theta)$ be the essential range of θ. Let $\{A_n\}$ be a matrix sequence distributed as θ in the sense of eigenvalues; in that case, defining Λ_n to be the set of eigenvalues of A_n, the following facts are true:*

a) *$\mathcal{S}(\theta)$ is a weak cluster for $\{A_n\}$;*

b) *each point $s \in \mathcal{S}(\theta)$ strongly attracts Λ_n with infinite order $r(s) = \infty$;*

c) *there exists a sequence $\{\lambda^{(n)}\}$, where $\lambda^{(n)}$ is an eigenvalue of A_n, such that* $\liminf_{n\to\infty} |\lambda^{(n)}| \geq \|\theta\|_\infty$.

The same statements holds in the case of a $N \times N$ matrix-valued function θ.

Proof. For items **a)** and **b)** see [12], Theorem 2.4, for a proof. Then notice that, by **b)**, each point $s \in \mathcal{S}(\theta)$ is a limit of a sequence $\{\lambda^{(n)}\}$ where $\lambda^{(n)}$ is an eigenvalue of A_n. Hence item **c)** follows from the definition of $\mathcal{S}(\theta)$. The extension to the matrix-valued case is trivial. □

The following result, based on a Mirski theorem (see Proposition III, Section 5.3 of [2]), establishes a link between distributions of non-Hermitian perturbations of Hermitian matrix sequences and the distribution of the original sequence.

Theorem 2.3. [12][Theorem 3.4] *Let $\{B_n\}$ and $\{C_n\}$ be two matrix sequences, where B_n is Hermitian and $A_n = B_n + C_n$. Assume further that $\{B_n\}$ is distributed as (θ, G) in the sense of the eigenvalues, where G is of finite and positive Lebesgue measure, both $\|B_n\|$ and $\|C_n\|$ are uniformly bounded by a positive constant $\widehat{C}$ independent of n, and $\|C_n\|_1 = o(n)$, $n \to \infty$. Then θ is real-valued and $\{A_n\}$ is distributed as (θ, G) in the sense of the eigenvalues. In particular, if $\mathcal{S}(\theta)$ is the essential range of θ, then $\{A_n\}$ is weakly clustered at $\mathcal{S}(\theta)$, and $\mathcal{S}(\theta)$ strongly attracts the spectra of $\{A_n\}$ with an infinite order of attraction for any of its points.*

The next theorem is a slight extension of a theorem from [12] concerning strong clustering.

Theorem 2.4. *Let $\{B_n\}$ and $\{C_n\}$ be two matrix sequences, where B_n is Hermitian and $A_n = B_n + C_n$. Let E be a compact subset of the real line. Assume that $\{B_n\}$ is strongly clustered at E, $\|C_n\|_1 = O(1)$, $n \to \infty$ and $\|A_n\|$ is uniformly bounded by a positive constant $\widehat{C}$ independent of n. Then $\{A_n\}$ is strongly clustered at E.*

Proof. The case where the compact set E is a union of m disjoint closed intervals (possibly, degenerate) has been treated in [12] (Theorem 3.6). The general case follows since, for the notion of strong clustering we have to consider the ϵ fattening of E, or $D(E, \epsilon)$ defined as in relation (6). It is clear that for every compact set E, the closure of $D(E, \epsilon)$ is a finite union of closed intervals and so the general case is reduced to that handled in [12]. □

Now we give a simple technical result which is useful in our subsequent study and which is due to SeLegue: it can explicitly be found in Lemma 5.16 in [8]. We present an elementary matrix proof as an alternative to the (elementary) operator theory proof given in [8] . This proof seems to be the most natural one to extend to the multi-level case, as explained in the proof of Lemma 2.6.

Lemma 2.5. *Let $f, g \in L^\infty(\mathbb{T})$, $A_n = T_n(f)T_n(g)$, and let $h = fg$. Then $\|A_n - T_n(h)\|_1 = o(n)$.*

Proof. In order to estimate $\|A_n - T_n(h)\|_1$, i.e., the Schatten 1 norm of $A_n - T_n(h)$, we will use some classical results from approximation theory.

For a given $\theta \in L^1(\mathbb{T})$, let $p_{k,\theta}$ be its Cesaro sum of degree k, i.e., the arithmetic average of Fourier sums of order q with $q \leq k$ (see [41, 3]). From standard trigonometric series theory we know that $p_{k,\theta}$ converges in L^1 norm to θ as k tends to infinity and also that $\|p_{k,\theta}\|_{L^\infty} \leq \|\theta\|_{L^\infty}$, whenever $\theta \in L^\infty(\mathbb{T})$ with $L^\infty(\mathbb{T}) \subset L^1(\mathbb{T})$. Furthermore, the norm inequality $\|T_n(\theta)\|_p \leq ((2\pi)^{-1}n)^{1/p}\|\theta\|_{L^p}$ holds for every $\theta \in L^p(\mathbb{T})$ if $1 \leq p \leq \infty$ (see [1, 31], Corollary 4.2). Now, by adding and subtracting and by using the triangle inequality several times we get:

$$\begin{aligned} \|A_n - T_n(h)\|_1 \quad \leq \quad & \|A_n - T_n(p_{k,f})T_n(g)\|_1 + \|T_n(p_{k,f})T_n(g) - T_n(p_{k,f})T_n(p_{k,g})\|_1 + \\ & + \|T_n(p_{k,f})T_n(p_{k,g}) - T_n(p_{k,f}p_{k,g})\|_1 + \|T_n(p_{k,f}p_{k,g}) - T_n(h)\|_1, \end{aligned} \tag{11}$$

and, by using Hölder inequalities for the Schatten p norms $\|XY\|_1 \leq \|X\|_1\|Y\|$ and the previously mentioned norm inequality from [31], we infer that

$$\begin{aligned} \|A_n - T_n(p_{k,f})T_n(g)\|_1 &= \|(T_n(f) - T_n(p_{k,f}))T_n(g)\|_1 \\ &\leq \|T_n(f) - T_n(p_{k,f})\|_1\|T_n(g)\| \\ &\leq n(2\pi)^{-1}\|f - p_{k,f}\|_{L^1}\|g\|_{L^\infty}; \\ \|T_n(p_{k,f})T_n(g) - T_n(p_{k,f})T_n(p_{k,g})\|_1 &= \|T_n(p_{k,f})(T_n(g) - T_n(p_{k,g}))\|_1 \\ &\leq \|T_n(g) - T_n(p_{k,g})\|_1\|T_n(p_{k,f})\| \\ &\leq \|T_n(g - p_{k,g})\|_1\|p_{k,f}\|_{L^\infty} \\ &\leq n(2\pi)^{-1}\|g - p_{k,g}\|_{L^1}\|f\|_{L^\infty}; \\ \|T_n(p_{k,f}p_{k,g}) - T_n(h)\|_1 &= \|T_n(h - p_{k,f}p_{k,g})\|_1 \\ &\leq n(2\pi)^{-1}\|h - p_{k,f}p_{k,g}\|_{L^1}. \end{aligned}$$

Thus, we see that the sum of the first, second and fourth terms of (11) equals $\epsilon(k)n$ where, since the Cesaro operator converges to the identity in the L^1 topology, we have

$$\lim_{k\to\infty} \epsilon(k) = 0.$$

We treat the third term of (11) in a different way. Let E_n and P_n be defined as in Section 1.2; recall that $p_{k,f}$ and $p_{k,g}$ are Laurent (or trigonometric) polynomials of degree at most k, that is they are of the form:

$$p_{k,f}(e^{it}) = \sum_{j=-k}^{k} a_j \exp(ijt), \qquad p_{k,g}(e^{it}) = \sum_{j=-k}^{k} b_j \exp(ijt).$$

Now, we recall that $T_n(p_{k,g})$ is the matrix of the Toeplitz operator $T_n^{p_{k,g}}$ where $T_n^{p_{k,g}} = P_n T_{p_{k,g}} P_n$; with $T_{p_{k,g}}$ the classical Toeplitz operator on H^2 and P_n orthogonal projection on the space E_n of analytic polynomials of degree less than n. So, since for $k \leq \ell \leq n-1-k$, the function $\phi(e^{it}) = p_{k,g}(e^{it})\exp(i\ell t)$ is in E_n we see that:

$$T_n^{p_{k,g}}(\exp(i\ell t)) = P_n(p_{k,g}(e^{it})\exp(i\ell t)) = p_{k,g}(e^{it})\exp(i\ell t), \qquad (k \leq \ell \leq n-k-1),$$

and so

$$T_n^{p_{k,f}} T_n^{p_{k,g}}(\exp(i\ell t)) = P_n(p_{k,f}(e^{it})p_{k,g}(e^{it})\exp(i\ell t)) = T_n^{p_{k,f}p_{k,g}}(\exp(i\ell t)), \qquad (k \leq \ell \leq n-1-k).$$

This means that the image of the operator $T_n^{p_{k,f}p_{k,g}} - T_n^{p_{k,f}}T_n^{p_{k,g}}$ is generated by the image of the set $\{\exp(i\ell t)\}_{0\leq\ell\leq k-1 \text{ or } n-k\leq\ell\leq n-1}$ and so its dimension is less than or equal to $2k$. Thus, since $T_n(p_{k,f}p_{k,g}) - T_n(p_{k,f})T_n(p_{k,g})$ is just the matrix of the operator $T_n^{p_{k,f}p_{k,g}} - T_n^{p_{k,f}}T_n^{p_{k,g}}$ in the basis $\{\exp(i\ell t)\}$, we see that the rank of the matrix $T_n(p_{k,f}p_{k,g}) - T_n(p_{k,f})T_n(p_{k,g})$ is at most $2k$.

Now, since the trace norm is bounded by the the rank times the spectral or operator norm, we see that:

$$\begin{aligned}
\|T_n(p_{k,f})T_n(p_{k,g}) - T_n(p_{k,f}p_{k,g})\|_1 &\leq 2k\|T_n(p_{k,f})T_n(p_{k,g}) - T_n(p_{k,f}p_{k,g})\| \\
&\leq 2k(\|T_n(p_{k,f})\|\|T_n(p_{k,g})\| + \|T_n(p_{k,f}p_{k,g})\|) \\
&\leq 2k(\|p_{k,f}\|_{L^\infty}\|p_{k,g}\|_{L^\infty} + \|p_{k,f}p_{k,g}\|_{L^\infty}) \\
&\leq 2k(\|p_{k,f}\|_{L^\infty}\|p_{k,g}\|_{L^\infty} + \|p_{k,f}\|_{L^\infty}\|p_{k,g}\|_{L^\infty}) \\
&= 4k\|p_{k,f}\|_{L^\infty}\|p_{k,g}\|_{L^\infty} \\
&\leq 4k\|f\|_{L^\infty}\|g\|_{L^\infty},
\end{aligned}$$

for each $k \in \mathbb{N}$. Thus, if $M = 4\|g\|_{L^\infty}\|f\|_{L^\infty}$ and $\epsilon(k)$ is defined above, then

$$\|A_n - T_n(h)\|_1 \leq \epsilon(k)n + kM, \tag{12}$$

for each $k \in \mathbb{N}$. Now, for each $\epsilon > 0$, by first choosing k_0 so that $\epsilon(k_0) < \frac{\epsilon}{2}$ then choosing $\tilde{N} > \frac{2Mk_0}{\epsilon}$, we see that $n \geq \tilde{N}$ gives $\frac{\|A_n - T_n(h)\|_1}{n} \leq \epsilon$ which finishes the proof. □

Next, we notice that the reasoning above applies to multilevel Toeplitz matrices. Let T_n represent the multilevel $\hat{n}$ by $\hat{n}$ Toeplitz matrix with symbol f (as in Section 1.2).

Lemma 2.6. *Let $f, g \in L^\infty(\mathbb{T}^d)$, $n = (n_1, \ldots, n_d) \in \mathbb{N}^d$ and $\hat{n} = n_1 n_2 \cdots n_d$. Then for $A_n = T_n(f)T_n(g)$ and $h = fg$ we have:*

$$\|A_n - T_n(h)\|_1 = o(\hat{n}).$$

The only part of the proof which is slightly different from that of Lemma 2.5 is the treatment of the third term of (11). To get the analogous inequality, we consider the multi-variable equivalents P_n and E_n, $n = (n_1, \ldots, n_d)$, and see that if $k = (k_1, \ldots, k_d)$, $p_{k,f}$ and $p_{k,g}$ are the multivariate Laurent polynomials approximating f and g, then for $k_1 \leq \ell_1 \leq n_1 - 1 - k_1, \ldots, k_d \leq \ell_d \leq n_d - 1 - k_d$ we have (writing $z_r = \exp(it_r)$, and $z^\ell = z_1^{\ell_1} \cdots z_d^{\ell_d}$)

$$z^\ell p_{k,g} \in E_n,$$

and so, for $k_1 \leq \ell_1 \leq n_1 - 1 - k_1, \ldots, k_d \leq \ell_d \leq n_d - 1 - k_d$, we find

$$T_n(p_{k,f})T_n(p_{k,g})(z^\ell) = T_n(p_{k,f}p_{k,g})(z^\ell).$$

Thus, by the same logic as in the proof of the one variable case, the rank of the matrix

$$T_n(p_{k,f})T_n(p_{k,g}) - T_n(p_{k,f}p_{k,g}),$$

is less than $\hat{n} - (n_1 - 2k_1) \cdots (n_d - 2k_d)$. So, setting

$$\gamma(k) = \hat{n} - (n_1 - 2k_1) \cdots (n_d - 2k_d),$$

we can replace equation (12) with the equation:

$$\|A_n - T_n(h)\|_1 \leq \epsilon(k)\hat{n} + \gamma(k)M,$$

for each $k \in \mathbb{N}^d$, and choose, for $\epsilon > 0$, a d-tuple k such that $\epsilon(k) < \frac{\epsilon}{2}$ and an $\tilde{N}$ such that $\tilde{N} > \frac{2M\gamma(k)}{\epsilon}$. Then, if $\hat{n} > \tilde{N}$ we will have

$$\frac{\|A_n - T_n(h)\|_1}{\hat{n}} < \epsilon,$$

which shows that $\|A_n - T_n(h)\|_1 = o(\hat{n})$ and finishes the proof.

3 Preliminary results

We start with the case of a sequence $\{A_n\}$ where $A_n = T_n(f)T_n(g)$; $f, g \in L^\infty(\mathbb{T})$ such that fg is real-valued (even though f and g are not necessarily real-valued); for the simpler, all real-valued case, see [30]. The idea is to look at A_n as the Hermitian matrix $T_n(h)$, $h = fg$, plus a correction term C_n such that $\|C_n\|_1 = o(n)$ as $n \to \infty$, where each of the matrix sequences is uniformly bounded in operator norm (see Lemma 2.5). This will permit us to use the powerful Theorem 2.3.

Theorem 3.1. *Let $f, g \in L^\infty(\mathbb{T})$ be such that $h = fg$ is real-valued. Then $\{A_n\} \sim_\lambda (h, \mathbb{T})$ with $A_n = T_n(f)T_n(g)$, $\mathcal{S}(h)$ is a weak cluster for $\{A_n\}$, and any $s \in \mathcal{S}(h)$ strongly attracts the spectra of $\{A_n\}$ with infinite order.*

Proof. It is well known (see [14]) that $\{T_n(h)\} \sim_\lambda (h, \mathbb{T})$ and $\|T_n(\theta)\| \le \|\theta\|_{L^\infty}$ for every $\theta \in L^\infty(\mathbb{T})$. Thus $\|T_n(h)\| \le \|h\|_{L^\infty}$ and $\|A_n\| \le \|T_n(f)\|\|T_n(g)\| \le \|f\|_{L^\infty}\|g\|_{L^\infty}$. As a consequence, since $\|A_n - T_n(h)\|_1 = o(n)$ by Lemma 2.5, the desired results follow by applying Theorem 2.3 with $B_n = T_n(h)$, $C_n = A_n - T_n(h)$, and invoking Theorem 2.2. □

Now we once again notice that the same theorem holds for multilevel Toeplitz matrices.

Theorem 3.2. *Let $d \in \mathbb{N}^+$ and let $f, g \in L^\infty(\mathbb{T}^d)$ be such that $h = fg$ is real-valued. Then, if*

$$A_n = T_n(f)T_n(g),$$

we have that $\{A_n\} \sim_\lambda (h, \mathbb{T}^d)$, $\mathcal{S}(h)$ is a weak cluster for $\{A_n\}$, and any $s \in \mathcal{S}(h)$ strongly attracts the spectra of $\{A_n\}$ with infinite order.

Proof. In 1993, Tyrtyshnikov showed that the relation (1) holds for multilevel Toeplitz sequences (see [[8], Theorem 6.41]) so that we once again have $\{T_n(h)\} \sim_\lambda (h, \mathbb{T}^d)$. Also, by the definition of the Toeplitz operators it is again true that

$$\|T_n(h)\| \le \|h\|_{L^\infty},$$

and $\|A_n\| \le \|T_n(f)\|\|T_n(g)\| \le \|f\|_{L^\infty}\|g\|_{L^\infty}$. As a consequence, since $\|A_n - T_n(h)\|_1 = o(\hat{n})$ by Lemma 2.6, the desired results follow by applying Theorem 2.3 with $B_n = T_n(h)$ and $C_n = A_n - T_n(h)$, and invoking Theorem 2.2. □

Remark 3.3. Let f, g, h and A_n be defined as in Theorem 3.1 and suppose that either f or g is a Laurent polynomial of degree q. Then, if $h = fg$, by the same type of reasoning as above, $A_n - T_n(h)$ has rank less than or equal to q. Therefore, again using the fact that the sequences $\{\|A_n\|\}$ and $\{\|T_n(h)\|\}$ are both bounded by $\|f\|_{L^\infty}\|g\|_{L^\infty}$ and the Schur decomposition, it follows that $\|A_n - T_n(h)\|_1 \le 4q\|f\|_{L^\infty}\|g\|_{L^\infty}$. As a consequence, since $\mathcal{S}(h)$ is a compact real set, Theorem 2.4 implies that $\mathcal{S}(h)$ is a strong cluster for the spectra of $\{A_n\}$.

Remark 3.4. Lemma 2.5 and Theorem 3.1 remain valid in a block multidimensional setting, i.e., when considering symbols belonging to $L^\infty_N(\mathbb{T}^d)$ with $d \ge 2$, $N \ge 2$. In fact, we can follow verbatim the same proof as in Lemma 2.5 (see also [7]) and in Theorem 3.1 since all the involved tools concerning the Cesaro operator and the trace norm estimates have a natural counterpart in several dimensions and in the matrix-valued setting (see [41, 31]). The only change is of notational type: in fact all the terms $o(n)$ will become $o(\hat{n})$, since the involved dimensions in the multidimensional Toeplitz setting are defined as $N\hat{n}$, with $\hat{n} = n_1 \cdots n_d$ and with $n = (n_1, \ldots, n_d)$ being a multi-index, see Section 1.2.

In light of the previous remark, it is natural to state the following generalizations without proof.

Theorem 3.5. *Let $f, g \in L_N^\infty(\mathbb{T}^d)$ such that $h = fg$ is Hermitian-valued (real-valued for $N = 1$). Then $\{A_n\} \sim_\lambda (h, \mathbb{T}^d)$ with $A_n = T_n(f)T_n(g)$, $\mathcal{S}(h)$ is a weak cluster for $\{A_n\}$, and any $s \in \mathcal{S}(h)$ strongly attracts the spectra of $\{A_n\}$ with infinite order.*

Theorem 3.5 is the basis for the subsequent general result on the algebra generated by Toeplitz sequences with $L_N^\infty(\mathbb{T}^d)$ symbols. Its proof works by induction on the structure of h and of A_n and, more specifically, Theorem 3.5 is used for the basis of induction and for the inductive step. We do not furnish further details since, under mild additional assumptions, the same statement is proved carefully in Section 5 in the more general case where h belongs to the Tilli class. We recall that Hermitian-valued (real-valued if $N = 1$) $L_N^\infty(\mathbb{T}^d)$ functions form a proper subset of the Tilli class.

Theorem 3.6. *Let $f_{\alpha,\beta} \in L_N^\infty(\mathbb{T}^d)$ with $\alpha = 1, \dots, \rho$, $\beta = 1, \dots, q_\alpha$, $\rho, q_\alpha < \infty$. Assume that the function*

$$h = \sum_{\alpha=1}^{\rho} \prod_{\beta=1}^{q_\alpha} f_{\alpha,\beta},$$

is Hermitian-valued (real-valued for $N = 1$) and consider the sequence $\{A_n\}$ with $A_n = \sum_{\alpha=1}^{\rho} \prod_{\beta=1}^{q_\alpha} T_n(f_{\alpha,\beta})$. Then $\{A_n\} \sim_\lambda (h, \mathbb{T}^d)$, $\mathcal{S}(h)$ is a weak cluster for $\{A_n\}$, and any $s \in \mathcal{S}(h)$ strongly attracts the spectra of $\{A_n\}$ with infinite order.

Remark 3.7. Theorem 3.6 nicely complements the analysis by Böttcher and coauthors in [7]. In fact in [7] the authors require that the given sequence $\{A_n\}$ is normal, i.e., every A_n satisfies $A_n^* A_n = A_n A_n^*$. This technical assumption may be difficult to verify except in the Hermitian case. For the Hermitian setting see also [28] and Remark 1.5.

4 Further tools for general matrix sequences

Before generalizing the results of the previous section to the case where the product symbol h belongs to the Tilli class, we establish a series of general results for matrix sequences. In particular, we give some generalizations of Theorem 2.2 from [12]. We begin by stating this theorem in a slightly different, but equivalent way. The basic ideas used here come from the paper [36], where the same questions were considered in a Toeplitz context. First we give the results which come directly from Theorem 2.2 of [12] (in fact the ideas are taken from [36], but there were also known in a certain form to the operator theory community (see [39], top of page 390) and were extensively developed by Böttcher, Roch, SeLegue, Silbermann etc, see [8]).

Theorem 4.1. [12] *Let $\{A_n\}$ be a matrix sequence and S a subset of $\mathbb{C}$. If:*

(a1) *S is a compact set and $\mathbb{C}\backslash S$ is connected;*
(a2) *the matrix sequence $\{A_n\}$ is weakly clustered at S;*
(a3) *the spectra Λ_n of A_n are uniformly bounded, i.e., $\exists C \in \mathbb{R}^+$ such that $|\lambda| < C$, $\lambda \in \Lambda_n$, for all n;*
(a4) *there exists a function θ measurable, bounded, and defined on a set G of positive and finite Lebesgue measure, such that, for every positive integer L, we have $\lim_{n\to\infty} \frac{\mathrm{tr}(A_n^L)}{n} = \frac{1}{m(G)} \int_G \theta^L(t)\, dt$, i.e., relation (2) holds with F being any polynomial of an arbitrary fixed degree;*
(a5) *the essential range of θ is contained in S;*

then relation (2) is true for every continuous function F with bounded support which is holomorphic in the interior of S. If it is also true that the interior of S is empty then the sequence $\{A_n\}$ is distributed as θ on its domain G, in the sense of the eigenvalues.

Next, we show that the hypotheses **(a3)** and (a slightly stronger form of) **(a4)** imply **(a1)**, **(a2)**, and **(a3)** for the set S defined by "filling in" the essential range of the function θ from **(a4)** (or its strengthened version). This will show that, when our set $\mathcal{S}(\theta)$ has empty interior our matrix sequence has the desired distribution. When we say "filling in" we mean taking the "Area" in the following sense:

Definition 4.2. Let K be a compact subset of $\mathbb{C}$. We define

$$Area(K) = \mathbb{C}\backslash U,$$

where U is the (unique) unbounded connected component of $\mathbb{C}\backslash K$.

Theorem 4.3. *Let $\{A_n\}$ be a matrix sequence. If*

(b1) *the spectra Λ_n of A_n are uniformly bounded, i.e., $\exists C \in \mathbb{R}^+$ such that $|\lambda| < C$, $\lambda \in \Lambda_n$, for all n;*
(b2) *there exists a function θ measurable, bounded, and defined on a set G of positive and finite Lebesgue measure, such that, for all positive integers L and l, we have $\lim_{n\to\infty} \frac{\mathrm{tr}((A_n^*)^l A_n^L)}{n} = \frac{1}{m(G)} \int_G \overline{\theta^l(t)}\theta^L(t)\, dt$;*

then $\mathcal{S}(\theta)$ is compact, the matrix sequence $\{A_n\}$ is weakly clustered at $Area(\mathcal{S}(\theta))$, and relation (2) is true for every continuous function F with bounded support which is holomorphic in the interior of $S = Area(\mathcal{S}(\theta))$.

If it is also true that $\mathbb{C}\backslash\mathcal{S}(\theta)$ is connected and the interior of $\mathcal{S}(\theta)$ is empty then the sequence $\{A_n\}$ is distributed as θ on its domain G, in the sense of the eigenvalues.

Proof. Since θ is bounded, $\mathcal{S}(\theta)$ is bounded, and so, since the essential range is always closed, the set $\mathcal{S}(\theta)$ is compact. Hence we can define $S = Area(\mathcal{S}(\theta))$.

We prove that S is a weak cluster for the spectra of $\{A_n\}$. First, we notice that the compact set $S_C = \{z \in \mathbb{C} : |z| \leq C\}$ is a strong cluster for the spectra of $\{A_n\}$ since by **(b1)** it contains all the eigenvalues. Moreover C can be chosen such that S_C contains S. Therefore, we will have proven that S is a weak cluster for $\{A_n\}$ if we prove that, for every $\varepsilon > 0$, the compact set $S_C\backslash D(S,\varepsilon)$ contains at most only $o(n)$ eigenvalues, with $D(S,\varepsilon)$ as in Definition 1.2. By compactness, for any $\delta > 0$, there exists a finite covering of $S_C\backslash D(S,\varepsilon)$ made of balls $D(z,\delta)$, $z \in S_C\backslash S$ with $D(z,\delta) \cap S = \emptyset$, and so, it suffices to show that, for a particular δ, at most $o(n)$ eigenvalues lie in $D(z,\delta)$. Let $F(t)$ be the characteristic function of the compact set $\overline{D(z,\delta)}$. Then restricting our attention to the compact set $\overline{D(z,\delta)} \bigcup S$, Mergelyan's theorem implies that for each $\epsilon > 0$ there exists a polynomial P_ϵ such that $|F(t) - P_\epsilon(t)|$ is bounded by ϵ on $\overline{D(z,\delta)} \bigcup S$. Therefore, setting $\gamma_n(z,\delta)$ equal to the number of eigenvalues of A_n belonging to $\overline{D(z,\delta)}$, we find

$$
\begin{aligned}
(1-\epsilon)\gamma_n(z,\delta) &\le \sum_{i=1}^{n} F(\lambda_i)|P_\epsilon(\lambda_i)| && (13)\\
&\le \left(\sum_{i=1}^{n} F^2(\lambda_i)\right)^{1/2} \left(\sum_{i=1}^{n} |P_\epsilon(\lambda_i)|^2\right)^{1/2} && (14)\\
&= \left(\sum_{i=1}^{n} F(\lambda_i)\right)^{1/2} \left(\sum_{i=1}^{n} |P_\epsilon(\lambda_i)|^2\right)^{1/2} && (15)\\
&= (\gamma_n(z,\delta))^{1/2} \left(\sum_{i=1}^{n} |P_\epsilon(\lambda_i)|^2\right)^{1/2} && (16)\\
&\le (\gamma_n(z,\delta))^{1/2} \, \|P_\epsilon(A_n)\|_2 && (17)\\
&= (\gamma_n(z,\delta))^{1/2} \, (\mathrm{tr}(P_\epsilon^*(A_n)P_\epsilon(A_n)))^{1/2} && (18)\\
&= (\gamma_n(z,\delta))^{1/2} \left(\mathrm{tr}\left(\sum_{l,L=0}^{M} \overline{c_l} c_L (A_n^*)^l A_n^L\right)\right)^{1/2} && (19)\\
&= (\gamma_n(z,\delta))^{1/2} \left(\sum_{l,L=0}^{M} \overline{c_l} c_L \mathrm{tr}((A_n^*)^l A_n^L)\right)^{1/2}, && (20)
\end{aligned}
$$

where inequality (13) follows from the definition of F and from the approximation properties of P_ϵ, inequality (14) is Cauchy-Schwartz, relations (15)–(16) come from the definitions of F and $\gamma_n(z,\delta)$, (17) is a consequence of the Schur decomposition and of the unitary invariance of the Schatten norms, identities (18)–(20) follow from the entry-wise definition of the Schatten 2 norm (the Frobenius norm), from the monomial expansion of the polynomial P_ϵ, and from the linearity of the trace.

Given $\epsilon_2 > 0$, we choose $\epsilon_1 > 0$ so that equation

$$
\epsilon_1 \sum_{l,L=0}^{M} |c_l||c_L| \le \epsilon_2,
$$

is true and then we choose N so that for $n > N$, equation

$$
\left| \frac{\mathrm{tr}((A_n^*)^l A_n^L)}{n} - \frac{1}{m(G)} \int_G \overline{\theta^l(t)} \theta^L(t)\, dt \right| < \epsilon_1,
$$

is true. Then, picking up from equation (20), we have

$$
\begin{aligned}
(1-\epsilon)\gamma_n(z,\delta) &\le (\gamma_n(z,\delta))^{1/2} \left(n \left(\epsilon_2 + \frac{1}{m(G)} \int_G \sum_{l,L=0}^{M} \overline{c_l} c_L (\overline{\theta^l(t)} \theta^L(t))\, dt \right)\right)^{1/2} && (21)\\
&= (\gamma_n(z,\delta))^{1/2} \left(n \left(\epsilon_2 + \frac{1}{m(G)} \int_G |P_\epsilon(\theta(t))|^2\, dt \right)\right)^{1/2} && (22)\\
&\le (\gamma_n(z,\delta))^{1/2}\, n^{1/2} (\epsilon^2 + \epsilon_2)^{1/2}, && (23)
\end{aligned}
$$

where inequality (21) is assumption **(b2)**, the latter two inequalities are again consequences of the monomial expansion of P_ϵ and of the approximation properties of P_ϵ over the area delimited by the range of θ, and ϵ_2 is arbitrarily small. So, choosing $\epsilon_2 = \epsilon^2$, we see that (13)–(23) imply that, for n sufficiently large,

$$\gamma_n(z,\delta) \le 2n\epsilon^2(1-\epsilon)^{-2},$$

which means that: $\gamma_n(z,\delta) = o(n)$.

Thus, hypotheses **(a1)**–**(a5)** of Theorem 4.1 hold with $S = Area(\mathcal{S}(\theta))$, which is necessarily compact and with connected complement, and consequently the first conclusion of Theorem 4.1 holds. Finally if $\mathbb{C}\backslash\mathcal{S}(\theta)$ is connected and the interior of $\mathcal{S}(\theta)$ is empty then $Area(\mathcal{S}(\theta)) = \mathcal{S}(\theta)$ and so all the hypotheses of Theorem 4.1 are satisfied, and so we conclude that the sequence $\{A_n\}$ is distributed in the sense of the eigenvalues as θ on its domain G. □

Now, we give a second version, replacing hypotheses **(a1)**–**(a5)** with only **(a3)**, **(a4)**, and a condition on the Schatten p norm for a certain p.

Theorem 4.4. *Let $\{A_n\}$ be a matrix sequence and S a subset of $\mathbb{C}$. If*

(c1) *the spectra Λ_n of A_n are uniformly bounded, i.e., $|\lambda| < C$, $\lambda \in \Lambda_n$, for all n;*
(c2) *there exists a function θ measurable, bounded, and defined over G having positive and finite Lebesgue measure, such that, for every positive integer L, we have $\lim_{n\to\infty} \frac{\mathrm{tr}(A_n^L)}{n} = \frac{1}{m(G)} \int_G \theta^L(t)\,dt$;*
(c3) *there exist a constant $\widehat{C}$ and a positive real number $p \in [1,\infty)$, independent of n, such that $\|P(A_n)\|_p^p \le \widehat{C} n \frac{1}{m(G)} \int_G |P(\theta(t))|^p\,dt$ for every fixed polynomial P independent of n and for every n large enough;*

then the matrix sequence $\{A_n\}$ is weakly clustered at $Area(\mathcal{S}(\theta)) := \mathbb{C}\backslash U$ (see Definition 4.2) and relation (2) is true for every continuous function F with bounded support which is holomorphic in the interior of $S = Area(\mathcal{S}(\theta))$. If, moreover

(c4) $\mathbb{C}\backslash\mathcal{S}(\theta)$ *is connected and the interior of $\mathcal{S}(\theta)$ is empty;*

then the sequence $\{A_n\}$ is distributed as θ on its domain G, in the sense of the eigenvalues.

Proof. The proof goes as in Theorem 4.3 until relation (13). Then with q the conjugate of p (i.e., $1/q + 1/p = 1$) we have

$$\begin{aligned}
(1-\epsilon)\gamma_n(z,\delta) &\le \left(\sum_{i=1}^n F^q(\lambda_i)\right)^{1/q} \left(\sum_{i=1}^n |P_\epsilon(\lambda_i)|^p\right)^{1/p} && (24)\\
&= \left(\sum_{i=1}^n F(\lambda_i)\right)^{1/q} \left(\sum_{i=1}^n |P_\epsilon(\lambda_i)|^p\right)^{1/p} && (25)\\
&= (\gamma_n(z,\delta))^{1/q} \left(\sum_{i=1}^n |P_\epsilon(\lambda_i)|^p\right)^{1/p} && (26)\\
&\le (\gamma_n(z,\delta))^{1/q}\, \|P_\epsilon(A_n)\|_p && (27)\\
&\le (\gamma_n(z,\delta))^{1/q} \left(\frac{\widehat{C}n}{m(G)} \int_G |P_\epsilon(\theta(t))|^p\,dt\right)^{1/p} && (28)\\
&\le (\gamma_n(z,\delta))^{1/q}\, (\widehat{C}n)^{1/p}\epsilon, && (29)
\end{aligned}$$

where relation (24) is the Hölder inequality, relations (25)–(26) come from the definitions of F and $\gamma_n(z,\delta)$, (27) comes from the fact that, for any square matrix, the vector with the moduli of the eigenvalues is weakly-majorized by the vector of the singular values (see [2] for the precise definition and for the result), inequality (28) is assumption **(c3)** (which holds for any polynomial of fixed degree), and finally inequality (29) follows from the approximation properties of P_ϵ over the area delimited by the range of θ. Therefore

$$\gamma_n(z,\delta) \leq \widehat{C} n \epsilon^p (1-\epsilon)^{-p},$$

and since ϵ is arbitrary we have the desired result, i.e., $\gamma_n(z,\delta) = o(n)$.

The rest of the proof is the same as in Theorem 4.3. □

The next result tells us that the key assumption **(c3)** follows from the distribution in the singular value sense of $\{P(A_n)\}$ and that the latter is equivalent to the very same limit relation with only polynomial test functions. We should mention here that the distribution results in the singular value sense are much easier to obtain and to prove [35, 38, 34, 27, 28], thanks to the higher stability of singular values under perturbations [40].

Theorem 4.5. *Using the notation of Section 2, if the sequence $\{A_n\}$ is uniformly bounded in spectral norm then $\{A_n\} \sim_\sigma (\theta, G)$ is true whenever condition (3) holds for all polynomial test functions. Moreover, if $\{P(A_n)\} \sim_\sigma (P(\theta), G)$ for every polynomial P then claim **(c3)** is true for every value $p \in [1, \infty)$, for every $\epsilon > 0$ where $\widehat{C} = 1 + \epsilon$ and for n larger than a fixed value $\bar{n}_\epsilon$.*

Proof. The first claim is proved by using the fact that one can approximate any continuous function defined on a compact set contained in the (positive) real line by polynomials. The second claim follows from taking as test function the function z^p, with positive p, and exploiting the limit relation from the assumption $\{P(A_n)\} \sim_\sigma (P(\theta), G)$. Indeed, the sequence $\{P(A_n)\}$ is uniformly bounded since $\{A_n\}$ is, so we are allowed to use as test functions continuous functions with no restriction on the support. Therefore, by definition (see (3)), $\{P(A_n)\} \sim_\sigma (P(\theta), G)$ implies that

$$\lim_{n\to\infty} \frac{1}{n} \sum_{j=1}^{n} \sigma_j^p(P(A_n)) = \frac{1}{m(G)} \int_G |P(\theta(t))|^p dt.$$

Hence, by observing that $\sum_{j=1}^{n} \sigma_j^p(P(A_n))$ is by definition $\|P(A_n)\|_p^p$ and using the definition of limit, we see that, for every $\epsilon > 0$, there exists an integer $\bar{n}_\epsilon$ such that

$$\|P(A_n)\|_p^p \leq n \frac{1+\epsilon}{m(G)} \int_G |P(\theta(t))|^p dt, \quad \forall n \geq \bar{n}_\epsilon.$$

The latter inequality coincides with **(c3)** with $\widehat{C} = 1 + \epsilon$ and every $p \in [1, \infty)$. □

5 The Tilli class and the algebra generated by Toeplitz sequences

As discussed in Section 1.1, we can write any matrix A in the form

$$\mathrm{Re}(A)^+ - \mathrm{Re}(A)^- + i\ \mathrm{Im}(A)^+ - i\ \mathrm{Im}(A)^-,$$

where the four matrices $\mathrm{Re}(A)^+$, $\mathrm{Re}(A)^-$, $\mathrm{Im}(A)^+$, $\mathrm{Im}(A)^-$ are positive semi-definite so that their trace coincides with the trace norm. As a consequence it is not difficult to see that

$$|\mathrm{tr}(A)| \leq 2\|A\|_1. \tag{30}$$

Now we are ready to state and prove two important lemmas. An alternative proof using operator theory methods can be found in [7].

Lemma 5.1. *Let $f_\alpha \in L^\infty(\mathbb{T}^d)$, $\alpha = 1,\dots,\rho$, $\rho < \infty$, $d \geq 1$, let $A_n = \prod_{\alpha=1}^{\rho} T_n(f_\alpha) := T_n(f_1)T_n(f_2)\cdots T_n(f_\rho)$, $n = (n_1,\dots,n_d)$, and let $h = \prod_{\alpha=1}^{\rho} f_\alpha$. Then*

$$\|A_n - T_n(h)\|_1 = o(\hat{n}), \qquad \hat{n} = n_1\cdots n_d, \tag{31}$$

$$\lim_{n\to\infty} \frac{\mathrm{tr}(A_n)}{\hat{n}} = \frac{1}{(2\pi)^d}\int_{[-\pi,\pi]^d} h(e^{it_1},\dots,e^{it_d})\,dt_1\cdots dt_d. \tag{32}$$

Proof. For proving (31) we proceed by induction on the positive integer ρ. If $\rho = 1$ then there is nothing to prove since $A_n - T_n(h)$ is the null matrix. For $\rho > 1$, we write $A_n = \left(\prod_{\alpha=1}^{\rho-1} T_n(f_\alpha)\right) T_n(f_\rho)$, where, by the inductive step, we have $\prod_{\alpha=1}^{\rho-1} T_n(f_\alpha) = T_n(h_{\rho-1}) + E_{n,\rho-1}$ with $h_{\rho-1} = \prod_{\alpha=1}^{\rho-1} f_\alpha$ and $\|E_{n,\rho-1}\|_1 = o(\hat{n})$. As a consequence

$$A_n = T_n(h_{\rho-1})T_n(f_\rho) + E_{n,\rho-1}T_n(f_\rho),$$

where

$$\|E_{n,\rho-1}T_n(f_\rho)\|_1 \leq \|E_{n,\rho-1}\|_1\|T_n(f_\rho)\| \leq \|E_{n,\rho-1}\|_1\|f_\rho\|_{L^\infty},$$

by the Hölder inequality $\|XY\|_1 \leq \|X\|_1\|Y\|$ and by the inequality $\|T_n(g)\| \leq \|g\|_{L^\infty}$, see e.g. [8]. Furthermore, thanks to Lemma 2.6 , we have

$$\|T_n(h_{\rho-1})T_n(f_\rho) - T_n(h)\|_1 = o(\hat{n}),$$

since $h = h_{\rho-1}f_\rho$. In conclusion $A_n = T_n(h) + E_{n,\rho}$ where $E_{n,\rho} = E_{n,\rho-1}T_n(f_\rho) + T_n(h_{\rho-1})T_n(f_\rho) - T_n(h)$ so that by the triangle inequality $\|E_{n,\rho}\|_1 = o(\hat{n})$, and therefore the proof of the first part is concluded.

The proof of the second part, i.e, relation (32) is plain since the statement is a straightforward consequence of the first part. In fact

$$\mathrm{tr}(T_n(h)) = \hat{n}\hat{h}_0 = \frac{\hat{n}}{(2\pi)^d}\int_{[-\pi,\pi]^d} h(e^{it_1},\dots,e^{it_d})\,dt_1\cdots dt_d,$$

where $\hat{h}_0$ is the Fourier coefficient defined in (8), and, by (30) and (31),

$$\mathrm{tr}(A_n) = \mathrm{tr}(T_n(h)) + o(\hat{n}) = \frac{\hat{n}}{(2\pi)^d}\int_{[-\pi,\pi]^d} h(t)\,dt + o(\hat{n}),$$

which implies (32). □

Lemma 5.2. *Let $f_{\alpha,\beta} \in L^\infty(\mathbb{T}^d)$ with $\alpha = 1,\dots,\rho$, $\beta = 1,\dots,q_\alpha$, $\rho, q_\alpha < \infty$, $d \geq 1$, and let $n = (n_1,\dots,n_d)$ and $\hat{n} = n_1\cdots n_d$. Set*

$$A_n = \sum_{\alpha=1}^{\rho}\prod_{\beta=1}^{q_\alpha} T_n(f_{\alpha,\beta}),$$

and $h = \sum_{\alpha=1}^{\rho}\prod_{\beta=1}^{q_\alpha} f_{\alpha,\beta}$. Then $\|A_n - T_n(h)\|_1 = o(\hat{n})$ and

$$\lim_{n\to\infty} \frac{\mathrm{tr}\,(A_n)}{\hat{n}} = \frac{1}{(2\pi)^d}\int_{[-\pi,\pi]^d} h(e^{it_1},\dots,e^{it_d})\,dt_1\cdots dt_d. \tag{33}$$

Proof. The first claim is a trivial consequence of Lemma 5.1. For the second claim, just observe that the linearity of the trace operator and of the limit operation implies that (33) is equivalent to the statement that

$$\sum_{\alpha=1}^{\rho} \lim_{n\to\infty} \frac{1}{\hat{n}} \mathrm{tr} \left(\prod_{\beta=1}^{q_\alpha} T_n(f_{\alpha,\beta}) \right) = \sum_{\alpha=1}^{\rho} \frac{1}{(2\pi)^d} \int_{[-\pi,\pi]^d} \prod_{\beta=1}^{q_\alpha} f_{\alpha,\beta}(e^{it_1}, \ldots, e^{it_d})\, dt_1 \cdots dt_d.$$

Hence, setting $g_\alpha = \prod_{\beta=1}^{q_\alpha} f_{\alpha,\beta}$, $\alpha = 1, \ldots, \rho$, the desired result follows from

$$\lim_{n\to\infty} \frac{1}{\hat{n}} \mathrm{tr} \left(\prod_{\beta=1}^{q_\alpha} T_n(f_{\alpha,\beta}) \right) = \frac{1}{(2\pi)^d} \int_{[-\pi,\pi]^d} g_\alpha(e^{it_1}, \ldots, e^{it_d})\, dt_1 \cdots dt_d, \tag{34}$$

which is a consequence of Lemma 5.1. □

Theorem 5.3. *Let $f_{\alpha,\beta} \in L^\infty(\mathbb{T}^d)$ with $\alpha = 1, \ldots, \rho$, $\beta = 1, \ldots, q_\alpha$, $\rho, q_\alpha < \infty$, $d \geq 1$. Assume that the function*

$$h = \sum_{\alpha=1}^{\rho} \prod_{\beta=1}^{q_\alpha} f_{\alpha,\beta},$$

belongs to the Tilli class and consider the sequence $\{A_n\}$ with $A_n = \sum_{\alpha=1}^{\rho} \prod_{\beta=1}^{q_\alpha} T_n(f_{\alpha,\beta})$. Then $\{A_n\} \sim_\lambda (h, \mathbb{T}^d)$, $\mathcal{S}(h)$ is a weak cluster for $\{A_n\}$, and any $s \in \mathcal{S}(h)$ strongly attracts the spectra of $\{A_n\}$ with infinite order.

Proof. We choose to apply Theorem 4.4. Assumption **(c1)** is easily obtained by repeated applications of the triangle inequality to the infinity norm of the function h since the module of the eigenvalues is dominated by the infinity norm of the symbol. Statement **(c3)** is true for every p by Theorem 4.5, since $\{P(A_n)\} \sim_\sigma (P(h), \mathbb{T}^d)$ for every fixed polynomial P (see Remark 1.5); assumption **(c4)** is verified with $\theta = h$ since h belongs to the Tilli class. The only thing left is statement **(c2)** which is a consequence of Lemma 5.2, since any positive power of linear combinations of products is still a linear combination of products. Therefore $\{A_n\} \sim_\lambda (h, \mathbb{T}^d)$ by Theorem 4.4 and the proof is completed by invoking **a)** and **b)** from Theorem 2.2. □

5.1 The Tilli class in the case of matrix-valued symbols

With the same tools we can easily give the generalization of Theorem 5.3 to the case of $N \times N$ matrix valued symbols. Lemmas 5.1 and 5.2 are easy to extend and indeed this extension can be found in [7]. The only key point is to define the Tilli class in this context. We say that f belongs to the $N \times N$ matrix-valued Tilli class if f is essentially bounded (i.e. this is true for any entry of f) and if the union of the ranges of the eigenvalues of f has empty interior and does not disconnect the complex plane. We have to observe that the case where $f(t)$ is diagonalizable, by a constant transformation independent of t, is special in the sense that the Szegö-type distribution result holds under the milder assumption the every eigenvalue of f (now a scalar complex-valued function) belongs to the standard Tilli class. This leaves open the question whether this weaker requirement is sufficient in general.

Finally we remark that such results can be seen as a generalization of the analysis by Böttcher and coauthors in [7], with the advantage that the technical and difficult assumption of normality is dropped.

5.2 The role of thin spectrum in the case of Laurent polynomials

In this section we treat the case where the symbol f of our Toeplitz operator is a Laurent polynomial, i.e.,

$$f(z) = \sum_{j=-r}^{s} \hat{f}_j z^j, \qquad z \in \mathbb{T}.$$

Given a Laurent polynomial f and given a value $\rho > 0$, we denote by $f^{[\rho]}$ the function

$$f^{[\rho]}(z) = \sum_{j=-r}^{s} \hat{f}_j \rho^j z^j. \tag{35}$$

Clearly $f^{[\rho]}$ is still a Laurent polynomial and, if we define the $n \times n$ matrix D_ρ by:

$$D_\rho = \begin{bmatrix} 1 & & & & \\ & \rho & & & \\ & & \rho^2 & & \\ & & & \ddots & \\ & & & & \rho^{n-1} \end{bmatrix}, \qquad \rho > 0, \tag{36}$$

then a straightforward computation shows that

$$D_\rho T_n(f) D_\rho^{-1} = T_n(f^{[\rho]}). \tag{37}$$

Now, if f is any Laurent polynomial, then, as shown in the book [5] the eigenvalues of the sequence $\{T_n(f)\}$ cluster along a certain set called the Schmidt-Spitzer set, and denoted by $\Lambda(f)$. It was shown by Hirschmann (Theorems 11.16 and 11.17 of the book [5]), that, under certain hypotheses,

$$\{T_n(f)\} \sim_\lambda (\theta_f, G_f), \tag{38}$$

where θ_f is a suitable function supported on $G_f = \bigcap_{\rho>0} Area(\mathcal{S}(f^{[\rho]}))$, and where $f^{[\rho]}$ is defined as in (35).

Suppose now the functions $f_{\alpha,\beta}$, $\alpha = 1, \ldots, \nu$, $\beta = 1, \ldots, q_\alpha$, $\nu, q_\alpha < \infty$, are all Laurent polynomials, then the function h defined by

$$h = \textstyle\sum_{\alpha=1}^{\nu} \prod_{\beta=1}^{q_\alpha} f_{\alpha,\beta},$$

is also a Laurent polynomial. We want to prove that if h satisfies the hypotheses of the Hirschmann theorem so that $\{T_n(h)\} \sim_\lambda (\theta_h, G_h)$, then we can obtain the corresponding result for the sequence $\{A_n\}$, i.e., $\{A_n\} \sim_\lambda (\theta_h, G_h)$.

Theorem 5.4. *Let f, g be two Laurent polynomials, $A_n = T_n(f)T_n(g)$ and let $h = fg$. With D_ρ defined as in (36), for each $\rho > 0$, $\|D_\rho A_n D_\rho^{-1} - D_\rho T_n(h) D_\rho^{-1}\|_1 = o(n)$.*

Proof. This is a direct consequence of Lemma 2.5 applied to the functions $f^{[\rho]}$ and $g^{[\rho]}$ since (using (37)) we have

$$D_\rho A_n D_\rho^{-1} = T_n(f^{[\rho]}) T_n(g^{[\rho]}), \quad \text{and} \quad D_\rho T_n(h) D_\rho^{-1} = T_n(h^{[\rho]}),$$

and $f^{[\rho]} g^{[\rho]} = h^{[\rho]}$. □

Lemma 5.5. *Let $f_\alpha \in L^\infty(\mathbb{T})$ be Laurent polynomials with $\alpha = 1, \ldots, \nu$, $\nu < \infty$. Let*

$$h = \prod_{\alpha=1}^{\nu} f_\alpha,$$

be a new Laurent polynomial and let $\{A_n\}$ be defined as $A_n = \prod_{\alpha=1}^{\nu} T_n(f_\alpha)$. For each $\rho > 0$ we have

$$\|D_\rho A_n D_\rho^{-1} - D_\rho T_n(h) D_\rho^{-1}\|_1 = o(n),$$
$$\lim_{n\to\infty} \frac{\mathrm{tr}(D_\rho A_n D_\rho^{-1})}{n} = \frac{1}{2\pi}\int_{[-\pi,\pi]} h^{[\rho]}(t)dt.$$

Proof. The same reasoning as above shows that

$$D_\rho A_n D_\rho^{-1} = \prod_{\alpha=1}^{\nu} T_n(f_\alpha^{[\rho]}) \quad \text{and that} \quad D_\rho T_n(h) D_\rho^{-1} = T_n(h^{[\rho]}),$$

so that the Lemma is a direct consequence of Lemma 5.1 with $d = 1$. □

Lemma 5.6. *Let $f_{\alpha,\beta} \in L^\infty(\mathbb{T})$ be Laurent polynomials with $\alpha = 1, \ldots, \nu$, $\beta = 1, \ldots, q_\alpha$, $\nu, q_\alpha < \infty$. Let*

$$h = \sum_{\alpha=1}^{\nu}\prod_{\beta=1}^{q_\alpha} f_{\alpha,\beta},$$

be a new Laurent polynomial and let $\{A_n\}$ be defined as $A_n = \sum_{\alpha=1}^{\nu}\prod_{\beta=1}^{q_\alpha} T_n(f_{\alpha,\beta})$. For each $\rho > 0$ we have

$$\|D_\rho A_n D_\rho^{-1} - D_\rho T_n(h) D_\rho^{-1}\|_1 = o(n),$$
$$\lim_{n\to\infty} \frac{\mathrm{tr}(D_\rho A_n D_\rho^{-1})}{n} = \frac{1}{2\pi}\int_{[-\pi,\pi]} h^{[\rho]}(t)dt.$$

Proof. Once again, we apply (37) to see that

$$D_\rho A_n D_\rho^{-1} = \sum_{\alpha=1}^{\nu}\prod_{\beta=1}^{q_\alpha} T_n(f_{\alpha,\beta}^{[\rho]}) \quad \text{and that} \quad D_\rho T_n(h) D_\rho^{-1} = T_n\left(\sum_{\alpha=1}^{\nu}\prod_{\beta=1}^{q_\alpha} f_{\alpha,\beta}^{[\rho]}\right), \tag{39}$$

so that a direct application of Lemma 5.2, with $d = 1$, gives the desired result. □

Theorem 5.7. *Let $f_{\alpha,\beta} \in L^\infty(\mathbb{T})$ be Laurent polynomials with $\alpha = 1, \ldots, \nu$, $\beta = 1, \ldots, q_\alpha$, $\nu, q_\alpha < \infty$. Let*

$$h = \sum_{\alpha=1}^{\nu}\prod_{\beta=1}^{q_\alpha} f_{\alpha,\beta},$$

be a new Laurent polynomial and let $\{A_n\}$ be defined as $A_n = \sum_{\alpha=1}^{\nu}\prod_{\beta=1}^{q_\alpha} T_n(f_{\alpha,\beta})$. Denoting by $\mathcal{S}(h^{[\rho]})$ the essential range of $h^{[\rho]}$, for each $\rho > 0$, the set Area$(\mathcal{S}(h^{[\rho]}))$ is a weak cluster for $\{A_n\}$.

Proof. We apply Theorem 4.4 to the sequence $\{D_\rho A_n D_\rho^{-1}\}$ using the equations (39). Condition **(c1)** is obtained by repeatedly applying the triangle inequality to $\| \sum_{\alpha=1}^{\nu} \prod_{\beta=1}^{q_\alpha} f_{\alpha,\beta}^{[\rho]} \|_{L^\infty}$; **(c2)** is a consequence of Lemma 5.6, since any positive integer power of a linear combination of products is still linear combination of products; **(c3)** is true, in light of Theorem 4.5, since $\{P(D_\rho^{-1} A_n D_\rho)\} \sim_\sigma (P(h^{[\rho]}), \mathbb{T})$ for every polynomial P as a consequence of Lemma 5.6. Therefore Theorem 4.4 implies that the sequence $\{D_\rho A_n D_\rho^{-1}\}$ is weakly clustered at $Area(\mathcal{S}(h^{[\rho]}))$. Since A_n has the same eigenvalues as $D_\rho A_n D_\rho^{-1}$ this means that the sequence $\{A_n\}$ is also weakly clustered at $Area(\mathcal{S}(h^{[\rho]}))$. □

Theorem 5.8. *With the same notation as in Theorem 5.7,* $\bigcap_{\rho>0} Area(\mathcal{S}(h^{[\rho]}))$ *is a weak cluster both for* $\{A_n\}$ *and for* $\{T_n(h)\}$*.*

Proof. This follows from Theorem 5.7. □

Now, the Hirschmann theorem (Theorem 11.16, p. 274 in [5]), shows that for h Laurent polynomial satisfying certain assumptions we have

$$\{T_n(h)\} \sim_\lambda (\theta_h, G_h), \tag{40}$$

where θ_h is a suitable function supported on $G_h = \bigcap_{\rho>0} Area(\mathcal{S}(h^{[\rho]}))$, so that $\mathcal{S}(\theta_h) \subseteq \bigcap_{\rho>0} Area(\mathcal{S}(h^{[\rho]}))$. We use this to prove the following result.

Theorem 5.9. *Let* $f_{\alpha,\beta} \in L^\infty(\mathbb{T})$ *be Laurent polynomials with* $\alpha = 1, \ldots, \nu$, $\beta = 1, \ldots, q_\alpha$, $\nu, q_\alpha < \infty$ *and let*

$$h = \sum_{\alpha=1}^{\nu} \prod_{\beta=1}^{q_\alpha} f_{\alpha,\beta},$$

be a new Laurent polynomial satisfying the hypotheses of the Hirschmann theorem. Let $\{A_n\}$ *be defined as* $A_n = \sum_{\alpha=1}^{\nu} \prod_{\beta=1}^{q_\alpha} T_n(f_{\alpha,\beta})$*, and set* $G_h = \bigcap_{\rho>0} Area(\mathcal{S}(h^{[\rho]}))$*. If* $\mathbb{C} \setminus G_h$ *is connected in the complex field and the interior of* G_h *is empty, then* $\{A_n\} \sim_\lambda (\theta_h, G_h)$ *where* θ_h *is the distribution function of* $\{T_n(h)\}$ *indicated in (40), see* [5] *at page 274.*

Proof. We will use Theorem 4.1. First we see that **(a1)** holds since G_h is compact by construction and $\mathbb{C} \setminus G_h$ is connected by the hypotheses. Condition **(a2)** is a consequence of Theorem 5.8; while **(a3)** follows from a repeated application of the triangle inequality to $\| \sum_{\alpha=1}^{\nu} \prod_{\beta=1}^{q_\alpha} f_{\alpha,\beta} \|_{L^\infty}$.

Condition **(a4)** amounts in proving that

$$\lim_{n\to\infty} \frac{\mathrm{tr}(A_n^L)}{n} = \frac{1}{m(G_h)} \int_{G_h} \theta_h^L(t)dt. \tag{41}$$

In fact, from Lemma 5.2, with $d = 1$, we find $A_n = T_n(h) + R_{n,h}$ where $\|R_{n,h}\|_1 = o(n)$ and, in addition, by assumption $\{T_n(h)\} \sim_\lambda (\theta_h, G_h)$ (this second claim is indeed the Hirschmann result).

With these ingredients, we now prove formula (41). Since

$$\mathrm{tr}(X) = \sum_{\lambda \in \Lambda_n(X)} \lambda = \sum_{k=1}^{n} [X]_{k,k},$$

and since tr$(\cdot)$ is a linear functional, the assumption $A_n = T_n(h) + R_{n,h}$ implies that tr$(A_n)-$tr$(T_n(h))=$tr$(R_{n,h})$. Consequently

$$\begin{aligned}\left|\frac{1}{n}\mathrm{tr}(A_n)-\frac{1}{n}\mathrm{tr}(T_n(h))\right| &= \left|\frac{1}{n}\mathrm{tr}(R_{n,h})\right| \\ &\leq_{(\alpha)} \frac{2}{n}\|R_{n,h}\|_1 \\ &\leq_{(\beta)} \frac{2}{n}o(n)=o(1),\end{aligned}$$

where (α) follows from (30) and (β) follows from Lemma 5.2 (with $d=1$). Since $T_n(h)$ is distributed as θ_h over G_h, we infer

$$\lim_{n\to\infty}\frac{1}{n}\mathrm{tr}(A_n)=\lim_{n\to\infty}\frac{1}{n}\mathrm{tr}(T_n(h))=\frac{1}{m(G_h)}\int_{G_h}\theta_h(t)dt,$$

therefore (41) is satisfied in the special case where $L=1$.

Now we consider all nonnegative integers $L>0$. For $L=0,1$ the result is valid, so that we focus our attention to the case where $L\geq 2$. Relation $A_n=T_n(h)+R_{n,h}$ implies

$$\begin{aligned}A_n^L &= (T_n(h)+R_{n,h})^L \\ &= T_n(h)^L+\widetilde{R}_{n,h},\end{aligned}$$

where $\widetilde{R}_{n,h}$ is a term of the form

$$\widetilde{R}_{n,h}=\sum_{X_i\in\{T_n(h),R_{n,h}\}}(X_1\cdots X_L)-T_n(h)^L. \tag{42}$$

In other words the error matrix $\widetilde{R}_{n,h}$ is the sum of all possible combinations of products of j matrices $T_n(h)$ and k matrices $R_{n,h}$, with $j+k=L$ and the exception of $j=L$ (obviously it is understood that all the addends are pairwise different). By using a simple Hölder inequality involving Schatten p norms: $\|XY\|_1\leq\|X\|\|Y\|_1$, for every summand R in (42), we deduce that there exists $j\geq 1$, $k=L-j$ for which

$$\begin{aligned}\|R\|_1 &\leq \|T_n(h)\|^k\|R_{n,h}\|^{j-1}\|R_{n,h}\|_1 \\ &\leq_{(\alpha)} C^kC^{j-1}o(n),\end{aligned} \tag{43}$$

where (α) follows from the assumption:

$$\begin{aligned}&\|T_n(h)\|\leq\|h\|_{L^\infty}\leq C<\infty, \\ &\|R_{n,h}\|=\|A_n-T_n(h)\|\leq C<\infty.\end{aligned}$$

Therefore by the triangle inequality and by applying inequality (43) to any summand in (42), we find $\|\widetilde{R}_{n,h}\|_1\leq\hat{K}o(n)$, with $\hat{K}=\hat{K}(L)$ constant independent of n. Consequently

$\mathrm{tr}(A_n^L) - \mathrm{tr}(T_n(h)^L) = \mathrm{tr}(\widetilde{R}_{n,h})$, and, since $\lambda(X^L) = \lambda^L(X)$, we have

$$\begin{aligned}\left|\frac{1}{n}\mathrm{tr}(A_n^L) - \frac{1}{n}\mathrm{tr}(T_n(h)^L)\right| &= \left|\frac{1}{n}\sum_{\lambda\in\Lambda_n(A_n)}\lambda^L - \frac{1}{n}\sum_{\lambda\in\Lambda_n(T_n(h))}\lambda^L\right| \\ &= \left|\frac{1}{n}\sum_{\lambda\in\Lambda_n(\widetilde{R}_{n,h})}\lambda\right| \\ &\le \frac{2}{n}\|\widetilde{R}_{n,h}\|_1 \\ &\le \frac{2}{n}\hat{K}o(n) = o(1).\end{aligned}$$

Since $T_n(h)$ is distributed as θ_h over G_h, we infer

$$\lim_{n\to\infty}\frac{1}{n}\mathrm{tr}(A_n^L) = \lim_{n\to\infty}\frac{1}{n}\mathrm{tr}(T_n(h)^L) = \frac{1}{m(G_h)}\int_{G_h}\theta_h(t)^L dt.$$

The latter proves that (41) is satisfied for any nonnegative integer L.

Condition **(a5)** is true since $\mathcal{S}(\theta_h) \subset G_h$; finally G_h has empty interior by hypothesis. Therefore we can apply Theorem 4.1 and we conclude that $\{A_n\} \sim_\lambda (\theta_h, G_h)$. □

5.3 A complex analysis consequence for H^∞ functions

Let us consider the space $\mathcal{H}$ given by L^∞ functions defined on $\mathbb{T}^d$, $d \ge 1$; (where $\mathbb{T}$ is the unit circle in the complex plane) such that the Fourier coefficient $\hat{f}_j$, $j = (j_1, \ldots, j_d) \in \mathbb{Z}^d$, defined as in (8) equals zero if $j_k < 0$ for some k with $1 \le k \le d$.

Theorem 5.10. *If $h \in \mathcal{H}$, $[\mathcal{S}(h)]^C$ is connected, and the interior of $\mathcal{S}(h)$ is empty, then h is necessarily constant almost everywhere.*

Proof. By Theorem 2 of [36] (or equivalently, by Theorem 5.3 with $\rho = 1$ and $q_1 = 1$) we know that $\{T_n(h)\} \sim_\lambda (h, \mathbb{T}^d)$. However $T_n(h)$ is lower triangular with $\hat{h}_0$ on the main diagonal since $\hat{h}_j = 0$ if there exists k, $1 \le k \le d$, with $j_k < 0$. Therefore it is also true that $\{T_n(h)\} \sim_\lambda (\hat{h}_0, \mathbb{T}^d)$, i.e., $h \equiv \hat{h}_0$ and the proof is concluded. □

In other words, if $f \in \mathcal{H}$ and it is not constant almost everywhere, then its essential range necessarily divides the complex field in (at least two) unconnected components or its interior is not empty. Since a function is in $\mathcal{H}$ if and only if it is equal to the boundary values of a function in H^∞ this rigidity is not surprising.

From an operator theory viewpoint the proof is as follows. Since $\mathcal{H}$ is a closed subalgebra of L^∞, the spectrum of h in the subalgebra results from the spectrum of h in L^∞ by filling in holes. Thus, if the first set has no holes, then the two sets coincides and are equal to a set without interior points. As the second set is the closure of h over the polydisc, which contains interior points if h is not constant, it follows that h must be constant.

5.4 Some issues from statistics

This work was begun in search of an analysis of the asymptotic behavior of the function $W_n : C^0(\Pi, \mathbb{R}) \to \mathbb{R}$ where $C^0(\Pi, \mathbb{R})$ is the space of real continuous functions on the circle and W_n is defined by:

$$W_n(f) = \frac{1}{2\pi n}\int_\Pi f(t)\big|\sum_{j=0}^{n} X_j \exp(ijt)\big|^2 dt,$$

where (X_n) is a centered stationary real Gaussian process. If the spectral density of (X_n) is the positive bounded function g then

$$W_n(f) = \frac{1}{n} Y^{(n)} T_n(g)^{\frac{1}{2}} T_n(f) T_n(g)^{\frac{1}{2}} Y^{(n)},$$

where the vector $Y^{(n)}$ has a Gaussian $\mathcal{N}(0, I_n)$ distribution. We hope that our results will help. In fact the matrix $T_n(g)^{\frac{1}{2}} T_n(f) T_n(g)^{\frac{1}{2}}$ is similar to $T_n(g)T_n(f)$ since $T_n(g)$ is Hermitian positive definite. As a consequence in view of item **c)** in Theorem 2.2 and in view of Theorem 3.1, we can claim that the eigenvalue distribution of the the sequence $\{T_n(g)^{\frac{1}{2}} T_n(f) T_n(g)^{\frac{1}{2}}\}$ is $h = fg$ and that its maximal eigenvalue has $\limsup$ bounded from above by $\|f\|_{L^\infty}\|g\|_{L^\infty}$ and $\liminf$ bounded from below by $\|h\|_{L^\infty}$.

6 Concluding remarks and open problems

As a conclusion, we observe that tools from matrix theory (Mirski Theorem, see [2]) and approximation theory in the complex field (Mergelyan Theorem, see [22]), combined with those from asymptotic linear algebra [35, 36, 25] have been crucial in our proof of results concerning the eigenvalue distribution of non Hermitian matrix sequences. In particular, we have employed these tools to deduce general results that we have applied, as a special case, to the algebra generated by Toeplitz sequences. An interesting side effect, already implicitly contained in the Tilli analysis [36], is a characterization of the range of $L^\infty(\mathbb{T}^d)$ functions obtained as restrictions of functions of several complex variables in the Hardy space H^∞.

Some problems remain open. For instance it would be interesting to extend the results of this paper to the case where the involved symbols are not necessarily bounded, but just integrable. As already stressed in [28], in that case, the matrix theoretic approach seems more convenient, since the corresponding Toeplitz operators are not well defined if the symbols are not bounded.

Finally, it should be observed that the conditions described in the Tilli class for the existence of a canonical distribution corresponding to the symbol are sufficient, but not necessary. In fact for $f(t) = \exp(-it)$ the range of f is the complex unit circle, disconnecting the complex plane, while the eigenvalues are all equal to zero. However, if one takes the symbol $f(t)$ in (3.24), p.80 in [8] ($f(t) = \exp(2it)$, $t \in [0, \pi)$, $f(t) = \exp(-2it)$, $t \in [\pi, 2\pi)$), then the range of f is again the complex unit circle, that disconnects the complex plane, but the eigenvalues indeed distribute as the symbol as discussed in Example 5.39, pp. 167-169 in [8]. It would be nice to understand how to discriminate between these two types of generating functions which do not belong to the Tilli class.

References

[1] F. Avram, On bilinear forms on Gaussian random variables and Toeplitz matrices, *Probab. Theory Related Fields*, **79** (1988), pp. 37–45.

[2] R. Bhatia, “Matrix Analysis”, Springer Verlag, New York, 1997.

[3] R. Bhatia, “Fourier Series”, AMS, Providence, 2004.

[4] A. Böttcher, Variable-coefficient Toeplitz matrices with symbols beyond the Wiener algebra, *Heinig's Memorial Volume*, D. Bini, V. Olshevsky, E. Tyrtyshnikov Eds, Birkäuser, in print.

[5] A. Böttcher and S. Grudsky, “Spectral properties of banded Toeplitz matrices”, SIAM, Philadelphia, PA, 2005.

[6] A. Böttcher, S. Grudsky, and E. Maksimenko, The Szegö and Avram-Parter theorems for general test functions, *Comptes Rendus Acad. Sci. Paris - Ser. I*, in print.

[7] A. Böttcher, J. Gutiérrez-Gutiérrez, and P. Crespo, Mass concentration in quasicommutators of Toeplitz matrices, *J. Comput. Appl. Math.*, **205** (2007), pp. 129–148.

[8] A. Böttcher and B. Silbermann, “Introduction to Large Truncated Toeplitz Matrices”, Springer-Verlag, New York, 1999.

[9] L. Brown and P. Halmos, Algebraic properties of Toeplitz operators, *J. Reine Angew. Math.*, **123** (1964), pp. 89–102.

[10] D. Fasino and S. Serra-Capizzano, From Toeplitz matrix sequences to zero distribution of orthogonal polynomials, *Contemp. Math.*, **323** (2003), pp. 329–340.

[11] H. Gazzah, P. Regalia, J.P. Delmas, Asymptotic eigenvalue distribution of block Toeplitz matrices and application to blind SIMO channel identification, *IEEE Trans. Inform. Theory*, **47-3** (2001), pp. 1243–1251.

[12] L. Golinskii and S. Serra-Capizzano, The asymptotic properties of the spectrum of non symmetrically perturbed Jacobi matrix sequences, *J. Approx. Theory*, **144-1** (2007), pp. 84–102.

[13] R.M. Gray, Toeplitz and Circulant Matrices: a Review, *Foundations and Trends in Comm. Inf. Theory*, **2-3** (2006), pp. 1–93.

[14] U. Grenander and G. Szegö, “Toeplitz Forms and Their Applications”, Second Edition, Chelsea, New York, 1984.

[15] J. Gutiérrez-Gutiérrez, P. Crespo, and A. Böttcher, Functions of the banded Hermitian block Toeplitz matrices in signal processing, *Linear Algebra Appl*, **422-2/3** (2007), pp. 788–807.

[16] S. Holmgren, S. Serra-Capizzano, and P. Sundqvist, Can one hear the composition of a drum?, *Mediterr. J. Math.*, **3-2** (2006), pp. 227–249.

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

[18] A.B.J. Kuijlaars and S. Serra-Capizzano, Asymptotic zero distribution of orthogonal polynomials with discontinuously varying recurrence coefficients, *J. Approx. Theory*, **113** (2001), pp. 142–155.

[19] I. Louhichi, E. Strouse, and L. Zakariasy, Products of Toeplitz Operators on the Bergman space, *Integral Equations and Operator Theory*, **54** (2006), pp. 525–539.

[20] S. V. Parter, On the distribution of the singular values of Toeplitz matrices, *Linear Algebra Appl.*, **80** (1986), pp. 115–130.

[21] S. V. Parter, On the eigenvalues of certain generalizations of Toeplitz matrices, *Arch. Rat. Math. Mech.*, **3** (1962), pp. 244–257.

[22] W. Rudin, "Real and Complex Analysis", McGraw-Hill, New York, 1974.

[23] S. Serra-Capizzano, Spectral and computational analysis of block Toeplitz matrices with nonnegative definite generating functions, *BIT*, **39** (1999), pp. 152–175.

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

[25] S. Serra-Capizzano, Spectral behavior of matrix sequences and discretized boundary value problems, *Linear Algebra Appl.*, **337** (2001), pp. 37–78.

[26] S. Serra-Capizzano, Test functions, growth conditions and Toeplitz matrices, *Rendiconti Circolo Mat. Palermo*, **II-68** (2002), pp. 791–795.

[27] S. Serra-Capizzano, Generalized Locally Toeplitz sequences: spectral analysis and applications to discretized Partial Differential Equations, *Linear Algebra Appl.*, **366-1** (2003), pp. 371–402.

[28] S. Serra-Capizzano, The GLT class as a Generalized Fourier Analysis and applications, *Linear Algebra Appl.*, **419-1** (2006), pp. 180–233.

[29] S. Serra-Capizzano, D. Bertaccini and G. H. Golub, How to deduce a proper eigenvalue cluster from a proper singular value cluster in the non normal case, *SIAM J. Matrix Anal. Appl.*, **27-1** (2005), pp. 82–86.

[30] S. Serra-Capizzano and P. Sundqvist, Stability of the notion of approximating class of sequences and applications, *J. Comput. Appl. Math.*, electronic version, DOI 10.1016/j.cam.2007.03.027 (27/3/2007).

[31] S. Serra-Capizzano and P. Tilli, On unitarily invariant norms of matrix valued linear positive operators, *J. Inequalities Appl.*, **7-3** (2002), pp. 309–330.

[32] B. Silbermann and O. Zabroda, Asymptotic behavior of generalized convolutions: an algebraic approach, *J. Integral Equ. Appl.*, **18-2** (2006), pp. 169–196.

[33] P. Tilli, Singular values and eigenvalues of non-Hermitian block Toeplitz matrices, *Linear Algebra Appl.*, **272** (1998), pp. 59–89.

[34] P. Tilli, Locally Toeplitz matrices: spectral theory and applications, *Linear Algebra Appl.*, **278** (1998), pp. 91–120.

[35] P. Tilli, A note on the spectral distribution of Toeplitz matrices, *Linear Multilin. Algebra*, **45** (1998), pp. 147–159.

[36] P. Tilli, Some results on complex Toeplitz eigenvalues, *J. Math. Anal. Appl.*, **239-2** (1999), pp. 390–401.

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

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

[39] H. Widom, Eigenvalue distribution of nonselfadjoint Toeplitz matrices and the asymptotics of Toeplitz determinants in the case of nonvanishing index, *Oper. Theory Adv. Appl.*, **48** (1990), pp. 387–421.

[40] J. Wilkinson, “The Algebraic Eigenvalue Problem”, Claredon Press, Oxford, 1965.

[41] A. Zygmund, “Trigonometric Series”, Cambridge University Press, Cambridge, 1959.

Stefano Serra-Capizzano,
Department of Physics and Mathematics,
University of “Insubria”,
Via Valleggio 11, 22100 Como, Italy;
e-mail: stefano.serrac@uninsubria.it.

Debora Sesana,
Department of Physics and Mathematics,
University of “Insubria”,
Via Valleggio 11, 22100 Como, Italy;
e-mail: debora.sesana@uninsubria.it.

Elizabeth Strouse,
Department of Mathematics,
University of Bordeaux I,
Cours de la Liberation, 33405, Talence-Cedex, France;
e-mail: strouse@math.u-bordeaux1.fr.