Electronic Journal of Probability

Vol. 8 (2003) Paper no. 7, pages 1–29.

Journal URL
http://www.math.washington.edu/~ejpecp/
Paper URL
http://www.math.washington.edu/~ejpecp/EjpVol8/paper7.abs.html

THE NORM OF THE PRODUCT OF A LARGE MATRIX AND A RANDOM VECTOR

Albrecht Böttcher
Fakultät für Mathematik, TU Chemnitz, 09107 Chemnitz, Germany
aboettch@mathematik.tu-chemnitz.de

Sergei Grudsky [1]
Departamento de Matemáticas, CINVESTAV del I.P.N., Apartado Postal 14-740, 07000 México, D.F., México
grudsky@math.cinvestav.mx, grudsky@aaanet.ru

Abstract Given a real or complex $n \times n$ matrix A_n, we compute the expected value and the variance of the random variable $\|A_n x\|^2/\|A_n\|^2$, where x is uniformly distributed on the unit sphere of $\mathbf{R}^n$ or $\mathbf{C}^n$. The result is applied to several classes of structured matrices. It is in particular shown that if A_n is a Toeplitz matrix $T_n(b)$, then for large n the values of $\|A_n x\|/\|A_n\|$ cluster fairly sharply around $\|b\|_2/\|b\|_\infty$ if b is bounded and around zero in case b is unbounded.

Keywords condition number, matrix norm, random vector, Toeplitz matrix

AMS subject classification (2000) Primary 60H30, Secondary 15A60, 47B35, 65F35

Submitted to EJP on September 9, 2002. Final version accepted on May 14, 2003.

[1]This author was partially supported by CONACYT grant, Cátedra Patrimonial, Nivel II, No. 010286 and by a grant of the Saxon Ministry for Science and Art.

1 Introduction

Let $\|\cdot\|$ be the Euclidean norm in $\mathbf{R}^n$. For a real $n \times n$ matrix A_n, the spectral norm $\|A_n\|$ is defined by

$$\|A_n\| = \max_{\|x\|=1} \|A_n x\| = \max_{0<\|x\|\le 1} \frac{\|A_n x\|}{\|x\|}.$$

Let $s_1 \le s_2 \le \ldots \le s_n$ be the singular values of A_n, that is, the eigenvalues of $(A^\top A)^{1/2}$. The set $\{\|A_n x\|/\|A_n\| : \|x\| = 1\}$ coincides with the segment $[s_1/s_n, 1]$. We show that for a randomly chosen unit vector x the value of $\|A_n x\|^2/\|A_n\|^2$ typically lies near

$$\frac{1}{s_n^2}\,\frac{s_1^2 + \ldots + s_n^2}{n}. \tag{1}$$

Notice that $s_n = \|A_n\|$ and that $s_1^2 + \ldots + s_n^2 = \|A_n\|_{\mathrm{F}}^2$, where $\|A_n\|_{\mathrm{F}}$ is the Frobenius (or Hilbert-Schmidt) norm. Thus, if $\|A_n\| = 1$, then for a typical unit vector x the value of $\|A_n x\|^2$ is close to $\|A_n\|_{\mathrm{F}}^2/n$. The purpose of this paper is to use this observation in order to examine the most probable values of $\|A_n x\|/(\|A_n\|\,\|x\|)$ for several classes of large structured matrices A_n.

Our interest in the problem considered here arose from a talk by Siegfried Rump at a conference in Marrakesh in 2001. Let $M_n(\mathbf{R})$ denote the real $n \times n$ matrices and let $\mathrm{Circ}_n(\mathbf{R})$ stand for the circulant matrices in $M_n(\mathbf{R})$. For an invertible matrix $A_n \in \mathrm{Circ}_n(\mathbf{R})$, define the unstructured condition number $\kappa(A_n, x)$ of A_n at a vector $x \in \mathbf{R}^n$ as $\lim_{\varepsilon\to 0} \sup \|\delta x\|/(\varepsilon\|x\|)$, the supremum over all δx such that $(A_n + \delta A_n)(x + \delta x) = A_n x$ for some $\delta A_n \in M_n(\mathbf{R})$ with $\|\delta A_n\| \le \varepsilon\|A_n\|$, and define the structured condition number $\kappa^{\mathrm{circ}}(A_n, x)$ as $\lim_{\varepsilon\to 0} \sup \|\delta x\|/(\varepsilon\|x\|)$, this time the supremum over all δx such that $(A_n + \delta A_n)(x + \delta x) = A_n x$ for some $\delta A_n \in \mathrm{Circ}_n(\mathbf{R})$ with $\|\delta A_n\| \le \varepsilon\|A_n\|$. A well known result by Skeel says that $\kappa(A_n, x) = \|A_n\|\,\|A_n^{-1}\|$ (for every $A_n \in M_n(\mathbf{R})$), and in his talk Rump proved that

$$\kappa^{\mathrm{circ}}(A_n, x) = \frac{\|A_n\|\,\|A_n^{-1}x\|}{\|x\|}$$

(see also [9], [14]). Thus,

$$\frac{\kappa^{\mathrm{circ}}(A_n, x)}{\kappa(A_n, x)} = \frac{\|A_n^{-1}x\|}{\|A_n^{-1}\|\,\|x\|}, \tag{2}$$

which naturally leads to the question on the value taken by (2) at a typical x.

2 General Matrices

Let $B_n = \{x \in \mathbf{R}^n : \|x\| \le 1\}$ and $S_{n-1} = \{x \in \mathbf{R}^n : \|x\| = 1\}$. For a given matrix $A_n \in M_n(\mathbf{R})$, we consider the random variable

$$X_n(x) = \frac{\|A_n x\|}{\|A_n\|},$$

where x is uniformly distributed on S_{n-1}.

For $k \in \mathbf{N}$, the expected value of X_n^k is

$$EX_n^k = \frac{1}{|S_{n-1}|} \int_{S_{n-1}} \frac{\|A_n x\|^k}{\|A_n\|^k} \, d\sigma(x),$$

where $d\sigma$ is the surface measure on S_{n-1}. The variance of X_n^k is

$$\sigma^2 X_n^k = E\left(X_n^k - EX_n^k\right)^2 = EX_n^{2k} - \left(EX_n^k\right)^2.$$

As the following lemma shows, there is no difference between taking x uniformly on a sphere or in a ball.

Lemma 2.1 *For every natural number k,*

$$\frac{1}{|S_{n-1}|} \int_{S_{n-1}} \frac{\|A_n x\|^k}{\|A_n\|^k} \, d\sigma(x) = \frac{1}{|B_n|} \int_{B_n} \frac{\|A_n x\|^k}{\|A_n\|^k \|x\|^k} \, dx.$$

Proof. Using spherical coordinates, $x = rx'$ with $x' \in S_{n-1}$, we get

$$\int_{B_n} \frac{\|A_n x\|^k}{\|x\|^k} dx = \int_0^1 \int_{S_{n-1}} \frac{r^k \|A_n x'\|^k}{r^k} r^{n-1} \, d\sigma(x') dr = \frac{1}{n} \int_{S_{n-1}} \|A_n x'\|^k \, d\sigma(x'),$$

and since

$$|S_{n-1}| = \frac{2\pi^{n/2}}{\Gamma(n/2)} \quad \text{and} \quad |B_n| = \frac{\pi^{n/2}}{\Gamma(n/2+1)}$$

and thus $|S_{n-1}|/n = |B_n|$, the assertion follows. ■

The following result is undoubtedly known. As we have not found an explicit reference, we cite it with a full proof.

Theorem 2.2 *If $A_n \neq 0$, then*

$$EX_n^2 = \frac{1}{s_n^2} \frac{s_1^2 + \ldots + s_n^2}{n}, \tag{3}$$

$$\sigma^2 X_n^2 = \frac{2}{n+2} \frac{1}{s_n^4} \left(\frac{s_1^4 + \ldots + s_n^4}{n} - \left(\frac{s_1^2 + \ldots + s_n^2}{n} \right)^2 \right). \tag{4}$$

Proof. Let $A_n = U_n D_n V_n$ be the singular value decomposition. Thus, U_n and V_n are orthogonal matrices and $D_n = \operatorname{diag}(s_1, \ldots, s_n)$. By Lemma 2.1,

$$EX_n^2 = \frac{1}{|B_n|} \int_{B_n} \frac{\|U_n D_n V_n x\|^2}{\|U_n D_n V_n\|^2 \|x\|^2} \, dx$$

$$= \frac{1}{|B_n|} \int_{B_n} \frac{\|D_n V_n x\|^2}{\|D_n\|^2 \|V_n x\|^2} \, dx = \frac{1}{|B_n|} \int_{B_n} \frac{\|D_n x\|^2}{\|D_n\|^2 \|x\|^2} \, dx \tag{5}$$

$$= \frac{1}{|B_n|} \int_{B_n} \frac{s_1^2 x_1^2 + \ldots + s_n^2 x_n^2}{s_n^2 (x_1^2 + \ldots + x_n^2)} \, dx_1 \ldots dx_n; \tag{6}$$

notice that in (5) we first made the substitution $V_n x = y$ and then changed the notation y back to x. By symmetry, the integrals

$$\frac{1}{|B_n|}\int_{B_n}\frac{x_j^2}{x_1^2+\ldots+x_n^2}\,dx$$

are independent of j and hence they are all equal to $1/n$. This proves (3). In analogy to (6),

$$EX_n^4 = \frac{1}{|B_n|}\int_{B_n}\frac{(s_1^2x_1^2+\ldots+s_n^2x_n^2)^2}{s_n^4(x_1^2+\ldots+x_n^2)^2}\,dx_1\ldots dx_n. \tag{7}$$

A formula by Liouville (see, e.g., [7, No. 676]) states that if $\lambda < (p_1+\ldots+p_n)/2$, then

$$\begin{aligned}&\underset{\substack{x_1,\ldots,x_n\geq 0\\ x_1^2+\ldots+x_n^2\leq 1}}{\int\cdots\int}\frac{x_1^{p_1-1}\ldots x_n^{p_n-1}}{(x_1^2+\ldots+x_n^2)^\lambda}\,dx_1\ldots dx_n\\ &\qquad = \frac{1}{2^n\left(\frac{p_1+\ldots+p_n}{2}-\lambda\right)}\,\frac{\Gamma\left(\frac{p_1}{2}\right)\ldots\Gamma\left(\frac{p_n}{2}\right)}{\Gamma\left(\frac{p_1+\ldots+p_n}{2}\right)}.\end{aligned} \tag{8}$$

From (8) we obtain

$$\begin{aligned}&\frac{1}{|B_n|}\int_{B_n}\frac{x_j^4}{(x_1^2+\ldots+x_n^2)^2}\,dx\\ &= \frac{\Gamma\left(\frac{n}{2}+1\right)}{\pi^{n/2}}\,\frac{2^n}{2^n\left(\frac{n-1}{2}+\frac{5}{2}-2\right)}\,\frac{\Gamma\left(\frac{1}{2}\right)^{n-1}\Gamma\left(\frac{5}{2}\right)}{\Gamma\left(\frac{n-1}{2}+\frac{5}{2}\right)} = \frac{3}{n(n+2)},\end{aligned}$$

$$\begin{aligned}&\frac{1}{|B_n|}\int_{B_n}\frac{x_j^2x_k^2}{(x_1^2+\ldots+x_n^2)^2}\,dx\\ &= \frac{\Gamma\left(\frac{n}{2}+1\right)}{\pi^{n/2}}\,\frac{2^n}{2^n\left(\frac{n-2}{2}+\frac{3}{2}+\frac{3}{2}-2\right)}\,\frac{\Gamma\left(\frac{1}{2}\right)^{n-2}\Gamma\left(\frac{3}{2}\right)^2}{\Gamma\left(\frac{n-2}{2}+\frac{3}{2}+\frac{3}{2}\right)} = \frac{1}{n(n+2)},\end{aligned}$$

whence, by (7),

$$\begin{aligned}EX_n^4 &= \sum_{j=1}^n\frac{s_j^4}{s_n^4}\,\frac{3}{n(n+2)}+2\sum_{j<k}\frac{s_j^2s_k^2}{s_n^4}\,\frac{1}{n(n+2)}\\ &= \frac{1}{n(n+2)}\,\frac{1}{s_n^4}\,\left(2(s_1^4+\ldots+s_n^4)+(s_1^2+\ldots+s_n^2)^2\right).\end{aligned} \tag{9}$$

Since $\sigma^2 X_n^2 = EX_n^4 - (EX_n^2)^2$, formula (4) follows from (3) and (9). $\blacksquare$

From (4) we see that always $\sigma^2 X_n^2 \le \frac{2}{n+2}$. Thus, by Chebyshev's inequality,

$$P\left(\left|X_n^2 - \frac{1}{s_n^2}\frac{s_1^2 + \ldots + s_n^2}{n}\right| \ge \varepsilon\right) \le \frac{2}{(n+2)\varepsilon^2}$$

for each $\varepsilon > 0$ and

$$P\left(\left|X_n^2 - \frac{1}{s_n^2}\frac{s_1^2 + \ldots + s_n^2}{n}\right| \ge \frac{1}{n^{1/2-\delta}}\right) < \frac{2}{n^{2\delta}}$$

for each $\delta > 0$. This reveals that for large n the values of $\|A_n x\|^2/(\|A_n\|^2\|x\|^2)$ cluster around (1).

Notice also that $\sigma^2 X_n^2$ can be written in the symmetric forms

$$\sigma^2 X_n^2 = \frac{2}{n+2}\frac{1}{s_n^4}\sum_{i<j}\left(\frac{s_j^2 - s_i^2}{n}\right)^2 = \frac{1}{n+2}\frac{1}{s_n^4}\sum_{i,j=1}^{n}\left(\frac{s_i^2 - s_j^2}{n}\right)^2.$$

Obvious modifications of the proof of Theorem 2.2 show that Theorem 2.2 remains true for complex matrices on $\mathbf{C}^n$ with the ℓ^2 norm.

Example 2.3 Let

$$A_n = \begin{pmatrix} 1 & 1 & \ldots & 1 \\ 1 & 1 & \ldots & 1 \\ \ldots & \ldots & \ldots & \ldots \\ 1 & 1 & \ldots & 1 \end{pmatrix}. \tag{10}$$

The singular values of A_n are $0, \ldots, 0, n$ ($n-1$ zeros). Hence $\|A_n\| = n$, and the inequality $\|A_n x\|^2 \le \|A_n\|^2\|x\|^2$ is the well-known inequality

$$(x_1 + \ldots + x_n)^2 \le n(x_1^2 + \ldots + x_n^2),$$

which is valid for arbitrary real numbers $x_1, \ldots, x_n$. From Theorem 2.2 we deduce that

$$EX_n^2 = \frac{1}{n}, \quad \sigma^2 X_n^2 = \frac{2}{n+2}\frac{1}{n}\left(1 - \frac{1}{n}\right) \le \frac{2}{n^2}. \tag{11}$$

For $EX_n^2 = 1/n \le \varepsilon/2$ we therefore obtain from Chebyshev's inequality that

$$P\left(\frac{(x_1 + \ldots + x_n)^2}{n(x_1^2 + \ldots + x_n^2)} \ge \varepsilon\right) = P(X_n^2 \ge \varepsilon) \le P\left(|X_n^2 - EX_n^2| \ge \frac{\varepsilon}{2}\right) \le \frac{8}{n^2\varepsilon^2}.$$

Thus, the inequality

$$(x_1 + \ldots + x_n)^2 \le \varepsilon n(x_1^2 + \ldots + x_n^2),$$

is true with probability at least $1 - 8/(n^2\varepsilon^2)$. For instance, we have

$$(x_1 + \ldots + x_n)^2 \le \frac{n}{2}(x_1^2 + \ldots + x_n^2),$$

with probability at least 90% for $n \geq 18$ and with probability at least 99% for $n \geq 57$, and the inequality

$$(x_1 + \ldots + x_n)^2 \leq \frac{n}{100}\,(x_1^2 + \ldots + x_n^2),$$

is true with probability at least 90% whenever $n \geq 895$ and with probability at least 99% provided $n \geq 2829$. We will return to the present example in Example 7.5. ■

The following lemma will prove useful when studying concrete classes of matrices. We denote by $\|\cdot\|_{\mathrm{tr}}$ the trace norm, that is, the sum of the singular values.

Lemma 2.4 *Let $\{A_n\}_{n=1}^{\infty}$ be a sequence of matrices $A_n \in M_n(\mathbf{K})$, where $\mathbf{K} = \mathbf{R}$ or $\mathbf{K} = \mathbf{C}$. If*

$$\frac{\|A_n\|_{\mathrm{tr}}}{n} = O(1) \quad \textit{and} \quad \|A_n\| \to \infty,$$

then $EX_n^2 \to 0$ as $n \to \infty$.

Proof. Let $s_1(A_n) \leq s_2(A_n) \leq \ldots \leq s_n(A_n)$ be the singular values of A_n and note that $s_n(A_n) = \|A_n\|$. By assumption, there is a a finite constant M such that

$$\frac{1}{n}\sum_{j=1}^{n} s_j(A_n) \leq M$$

for all n. Fix $\varepsilon \in (0,1)$, for instance, $\varepsilon = 1/2$. Let N_n denote the number of all j for which $s_j(A_n) \geq M\|A_n\|^{1-\varepsilon}$. Then

$$M \geq \frac{1}{n}\sum_{j=1}^{n} s_j(A_n) \geq \frac{1}{n}\,N_n M \|A_n\|^{1-\varepsilon},$$

whence $N_n \leq n/\|A_n\|^{1-\varepsilon}$ and thus, by Theorem 2.2,

$$\begin{aligned} EX_n^2 &= \frac{1}{n s_n^2(A_n)}\sum_{j=1}^{n} s_j^2(A_n) \leq \frac{(n-N_n)M^2\|A_n\|^{2-2\varepsilon}}{n\|A_n\|^2} + \frac{N_n\|A_n\|^2}{n\|A_n\|^2} \\ &\leq \frac{M^2}{\|A_n\|^{2\varepsilon}} + \frac{1}{\|A_n\|^{1-\varepsilon}} = o(1) \end{aligned}$$

because $\|A_n\| \to \infty$. ■

We remark that if $EX_n^2 \to 0$, then $P(X_n \geq \varepsilon) = O(1/n)$ for each $\varepsilon > 0$: we have $EX_n^2 \leq \varepsilon^2/2$ for all $n \geq n_0$ and hence

$$\begin{aligned} P(X_n \geq \varepsilon) &= P(X_n^2 \geq \varepsilon^2) \leq P\left(X_n^2 - EX_n^2 \geq \frac{\varepsilon^2}{2}\right) \\ &= P\left(|X_n^2 - EX_n^2| \geq \frac{\varepsilon^2}{2}\right) \leq \frac{4\sigma^2 X_n^2}{\varepsilon^4} \leq \frac{8}{(n+2)\varepsilon^4}. \end{aligned}$$

3 Toeplitz Matrices with Bounded Symbols

We need one more simple auxiliary result.

Lemma 3.1 *Let $EX_n^2 = \mu_n^2$ and suppose $\mu_n \to \mu$ as $n \to \infty$. If $\varepsilon > 0$ and $|\mu_n - \mu| < \varepsilon$, then*

$$P(|X_n - \mu| \geq \varepsilon) \leq \frac{\sigma^2 X_n^2}{\mu_n^2(\varepsilon - |\mu_n - \mu|)^2}.$$

Proof. We have

$$\begin{aligned} P(|X_n - \mu| \geq \varepsilon) &\leq P\Big(|X_n - \mu_n| \geq \varepsilon - |\mu_n - \mu|\Big) \\ &\leq P\Big(|X_n - \mu_n|(X_n + \mu_n) \geq \mu_n(\varepsilon - |\mu_n - \mu|)\Big) \\ &= P\Big(|X_n^2 - \mu_n^2| \geq \mu_n(\varepsilon - |\mu_n - \mu|)\Big), \end{aligned}$$

and the assertion is now immediate from Chebyshev's inequality. ■

Now let A_n be a Toeplitz matrix, that is, $A_n = T_n(b) := (b_{j-k})_{j,k=1}^n$, where

$$b_k = \int_0^{2\pi} b(e^{i\theta}) e^{-ij\theta} \frac{d\theta}{2\pi} \quad (j \in \mathbf{Z}). \tag{12}$$

Clearly, (12) makes sense for every $b \in L^1$ on the complex unit circle $\mathbf{T}$. Throughout this section we assume that b is a function in L^∞. The Avram-Parter theorem says that in this case

$$\lim_{n\to\infty} \frac{s_1^k + \ldots + s_n^k}{n} = \|b\|_k^k := \int_0^{2\pi} |b(e^{i\theta})|^k \frac{d\theta}{2\pi} \tag{13}$$

for every natural number k (see [1], [2], [4], [11]). It is also well known that $s_n = \|T_n(b)\| \to \|b\|_\infty$ as $n \to \infty$ (see [2] or [4], for example). In what follows we always assume that b does not vanish identically. In Theorems 3.2 to 3.5, the constants hidden in the O's depend of course on ε and δ, respectively.

Theorem 3.2 *Let $b \in L^\infty$ and suppose $|b|$ is not constant almost everywhere. Then for each $\varepsilon > 0$, there is an $n_0 = n_0(\varepsilon)$ such that*

$$P\left(\left|\frac{\|T_n(b)x\|}{\|T_n(b)\|\,\|x\|} - \frac{\|b\|_2}{\|b\|_\infty}\right| \geq \varepsilon\right) \leq \frac{3}{n+2}\,\frac{1}{\varepsilon^2}\,\frac{\|b\|_4^4 - \|b\|_2^2}{\|b\|_2^2\,\|b\|_\infty^2} \tag{14}$$

for all $n \geq n_0$. If, in addition, b is a rational function, then for each $\delta > 0$,

$$P\left(\left|\frac{\|T_n(b)x\|}{\|T_n(b)\|\,\|x\|} - \frac{\|b\|_2}{\|b\|_\infty}\right| \geq \frac{1}{n^{1/2-\delta}}\right) = O\left(\frac{1}{n^{2\delta}}\right). \tag{15}$$

Proof. Since $|b|$ is not constant, it follows that $\|b\|_4 > \|b\|_2$. Put

$$\mu_n = \frac{1}{s_n}\sqrt{\frac{s_1^2 + \ldots + s_n^2}{n}}, \quad \mu = \frac{\|b\|_2}{\|b\|_\infty}.$$

From (13) we know that $\mu_n \to \mu$. Moreover, (13) and Theorem 2.2 imply that

$$\frac{n+2}{2}\,\sigma^2 X_n^2 \to \frac{1}{\|b\|_\infty^4}\left(\|b\|_4^4 - \|b\|_2^4\right).$$

Thus, Lemma 3.1 shows that

$$P(|X_n - \mu| \geq \varepsilon) \leq \frac{3}{n+2}\,\frac{1}{\|b\|_\infty^4}\left(\|b\|_4^4 - \|b\|_2^4\right)\frac{1}{\mu^2\varepsilon^2}$$

for all sufficiently large n, which is (14). If b is a rational function, we even have

$$\frac{s_1^k + \ldots + s_n^k}{n} = \|b\|_k^k + O\left(\frac{1}{n}\right) \tag{16}$$

for every natural number k and

$$s_n = \|b\|_\infty + O\left(\frac{1}{n^2}\right) \tag{17}$$

(see, e.g., [2]). It follows that $\mu_n = \mu + O(1/n)$, and hence Lemma 3.1 gives

$$P\left(\left|\frac{\|T_n(b)x\|}{\|T_n(b)\|\,\|x\|} - \frac{\|b\|_2}{\|b\|_\infty}\right| \geq \frac{1}{n^{1/2-\delta}}\right) \leq \frac{3}{n+2}\,\frac{\|b\|_4^4 - \|b\|_2^4}{\|b\|_\infty^4}\,\frac{1}{\mu^2}\,n^{1-2\delta},$$

which yields (15). ■

Theorem 3.3 *Let $b \in L^\infty$ and suppose $|b|$ is constant almost everywhere. Then*

$$P\left(\frac{\|T_n(b)x\|}{\|T_n(b)\|\,\|x\|} \leq 1 - \varepsilon\right) = o\left(\frac{1}{n}\right)$$

for each $\varepsilon > 0$. If, in addition, b is a rational function, then for each $\delta > 0$,

$$P\left(\frac{\|T_n(b)x\|}{\|T_n(b)\|\,\|x\|} \leq 1 - \frac{1}{n^{1-\delta}}\right) = O\left(\frac{1}{n^{2\delta}}\right)$$

Proof. In the case at hand, $\mu = 1$ and $\|b\|_4 = \|b\|_2$. From (13) and Theorem 2.2 we infer that

$$\mu_n \to 1 \quad \text{and} \quad \frac{n+2}{2}\,\sigma^2 X_n^2 = o(1).$$

Lemma 3.1 therefore gives

$$P(X_n \leq 1 - \varepsilon) \leq \frac{3}{n+2}\,o(1)\,\frac{1}{\varepsilon^2} = o\left(\frac{1}{n}\right).$$

If b is rational, we have (16) and (17). Thus,

$$\mu_n = 1 + O\left(\frac{1}{n}\right) \quad \text{and} \quad \frac{n+2}{2}\,\sigma^2 X_n^2 = O\left(\frac{1}{n}\right).$$

Consequently, by Lemma 3.1,

$$P\left(X_n \le 1 - \frac{1}{n^{1-\delta}}\right) \le \frac{3}{n+2}\, O\left(\frac{1}{n}\right) n^{2-2\delta} = O\left(\frac{1}{n^{2\delta}}\right). \quad \blacksquare$$

We now consider the case where A_n is the inverse of a Toeplitz matrix. Suppose b is a continuous function on $\mathbf{T}$ and b has no zeros on $\mathbf{T}$. Let $\operatorname{wind} b$ denote the winding number of b about the origin.

If $\operatorname{wind} b = 0$, then $T_n(b)$ is invertible for all sufficiently large n and

$$\|T_n^{-1}(b)\| \to \|T^{-1}(b)\| \text{ as } n \to \infty \tag{18}$$

(see, e.g., [2] or [4]). We remark that $T^{-1}(b) - T(b^{-1})$ is compact, so that

$$\|T^{-1}(b)\| \ge \|T(b^{-1})\| = \|b^{-1}\|_\infty \ge \|b^{-1}\|_2.$$

Theorem 3.4 *Suppose* $\operatorname{wind} b = 0$. *Then*

$$P\left(\left|\frac{\|T_n^{-1}(b)x\|}{\|T_n^{-1}(b)\|\,\|x\|} - \frac{\|b^{-1}\|_2}{\|T^{-1}(b)\|}\right| \ge \varepsilon\right) = O\left(\frac{1}{n}\right)$$

for each $\varepsilon > 0$. *If, in addition,* b *is rational, then*

$$P\left(\left|\frac{\|T_n^{-1}(b)x\|}{\|T_n^{-1}(b)\|\,\|x\|} - \frac{\|b^{-1}\|_2}{\|T^{-1}(b)\|}\right| \ge \frac{1}{n^{1/2-\delta}}\right) = O\left(\frac{1}{n^{2\delta}}\right)$$

for each $\delta \in (0, 1/2)$.

Proof. The singular values of $T_n^{-1}(b)$ are $1/s_j$ $(j = 1, \dots, n)$. Thus, by Theorem 2.2,

$$\begin{aligned} EX_n^2 &= \frac{1}{n}\, s_1^2 \left(\frac{1}{s_1^2} + \ldots + \frac{1}{s_n^2}\right), \\ \sigma^2 X_n^2 &= \frac{2}{n(n+2)}\, s_1^4 \left(\frac{1}{s_1^4} + \ldots + \frac{1}{s_n^4} - \frac{1}{n}\left(\frac{1}{s_1^2} + \ldots + \frac{1}{s_n^2}\right)^2\right). \end{aligned}$$

Since b has no zeros on $\mathbf{T}$, the Avram-Parter formula (13) also holds for negative integers k. This formula for $k = -2$ and (18) imply that

$$\mu_n^2 := EX_n^2 \to \|T^{-1}(b)\|^{-2} \|b^{-1}\|_2^2 =: \mu^2.$$

As always $\sigma^2 X_n^2 \le 2/(n+2)$, we obtain from Lemma 3.1 that

$$P(|X_n - \mu| \ge \varepsilon) \le \frac{3}{n+2}\,\frac{1}{\mu^2}\,\frac{1}{\varepsilon^2} = O\left(\frac{1}{n}\right).$$

In the case where b is rational, one can sharpen (18) and (13) to

$$\|T_n^{-1}(b)\| = \|T^{-1}(b)\| + O\left(\frac{\log n}{n}\right),$$
$$\frac{1}{n}\left(\frac{1}{s_1^k} + \ldots + \frac{1}{s_n^k}\right) = \|b^{-1}\|_k^k + O\left(\frac{1}{n}\right)$$

(see [2] and [4, Theorem 5.18]). Hence $\mu_n = \mu + O(\log n/n)$, and Lemma 3.1 shows that

$$P\left(|X_n - \mu| \geq \frac{1}{n^{2\delta}}\right) \leq \frac{3}{n+2}\,\frac{1}{\mu^2}\,n^{1-2\delta} = O\left(\frac{1}{n^{2\delta}}\right). \quad \blacksquare$$

If $|\operatorname{wind} b| = k \geq 1$, then $T_n(b)$ need not be invertible for all sufficiently large n. We therefore consider the Moore-Penrose inverse $T_n^+(b)$, which coincides with $T_n^{-1}(b)$ in the case of invertibility.

Theorem 3.5 *Suppose b is rational and $|\operatorname{wind} b| \geq 1$. Then*

$$P\left(\frac{\|T_n^+(b)x\|}{\|T_n^+(b)\|\,\|x\|} \geq \varepsilon\right) = O\left(\frac{1}{n^2}\right)$$

for each $\varepsilon > 0$ and

$$P\left(\frac{\|T_n^+(b)x\|}{\|T_n^+(b)\|\,\|x\|} \geq \frac{1}{n^{1/2-\delta}}\right) = O\left(\frac{1}{n^{4\delta}}\right)$$

for each $\delta > 0$.

Proof. The so-called splitting phenomenon, discovered by Roch and Silbermann [13] (see also [2]), tells us that if $|\operatorname{wind} b| = k \geq 1$, then k singular values of $T_n(b)$ converge to zero with exponential speed,

$$s_\ell \leq Ce^{-\gamma n} \quad (\gamma > 0) \quad \text{for } \ell \leq k,$$

while the remaining singular values stay away from zero,

$$s_\ell \geq \lambda > 0 \text{ for } \ell \geq k+1.$$

Thus, the singular values of $T_n^+(b)$ are

$$\underbrace{0, \ldots, 0}_{j-1}, \frac{1}{s_n}, \ldots, \frac{1}{s_{j+1}}, \frac{1}{s_j}$$

with $0 < s_j \leq s_{j+1} \leq \ldots \leq s_n$ and $j \leq k$, and from Theorem 2.2 we infer that

$$\begin{aligned}
EX_n^2 &= \frac{1}{n}\,s_j^2\left(\frac{1}{s_j^2} + \ldots + \frac{1}{s_n^2}\right) \\
&= \frac{1}{n}\left(\frac{s_j^2}{s_j^2} + \ldots + \frac{s_j^2}{s_k^2}\right) + \frac{s_j^2}{n}\left(\frac{1}{s_{k+1}^2} + \ldots + \frac{1}{s_n^2}\right) \\
&\leq \frac{1}{n}\,(k-j+1) + \frac{C^2e^{-2\gamma n}}{n}\,\frac{n-k}{\lambda^2} \\
&\leq \frac{k}{n} + \frac{C^2e^{-2\gamma n}}{\lambda^2} \leq \frac{k+1}{n}
\end{aligned}$$

for all sufficiently large n. Also by Theorem 2.2,

$$\sigma^2 X_n^2 \le \frac{2}{n(n+2)}\, s_j^4 \left(\frac{1}{s_j^4} + \ldots + \frac{1}{s_n^4}\right),$$

and since, analogously,

$$s_j^4 \left(\frac{1}{s_j^4} + \ldots + \frac{1}{s_n^4}\right) \le k + \frac{C^4 e^{-4\gamma n}(n-k)}{\lambda^4} \le k+1,$$

we get $\sigma^2 X_n^2 = O(1/n^2)$. If $\varepsilon > 0$, then

$$P(X_n \ge \varepsilon) = P(X_n^2 \ge \varepsilon^2) \le P\left(X_n^2 \ge \frac{k+1}{n} + \frac{\varepsilon^2}{2}\right)$$

for all sufficiently large n, and thus,

$$\begin{aligned} P(X_n \ge \varepsilon) &\le P\left(X_n^2 \ge EX_n^2 + \frac{\varepsilon^2}{2}\right) \\ &\le P\left(|X_n^2 - EX_n^2| \ge \frac{\varepsilon^2}{2}\right) \le \frac{4}{\varepsilon^4}\,\sigma^2 X_n^2 = O\left(\frac{1}{n^2}\right). \end{aligned}$$

Similarly, for large n,

$$\begin{aligned} P\left(X_n \ge \frac{1}{n^{1/2-\delta}}\right) &= P\left(X_n^2 \ge \frac{1}{n^{1-2\delta}}\right) \\ &\le P\left(X_n^2 \ge \frac{k+1}{n} + \frac{1}{2n^{1-2\delta}}\right) \le P\left(X_n^2 \ge EX_n^2 + \frac{1}{2n^{1-2\delta}}\right) \\ &\le P\left(|X_n^2 - EX_n^2| \ge \frac{1}{2n^{1-2\delta}}\right) \le 4n^{2-4\delta}\sigma^2 X_n^2 = O\left(\frac{1}{n^{4\delta}}\right). \quad \blacksquare \end{aligned}$$

4 Circulant Matrices

Now suppose b is a trigonometric polynomial. Then $T_n(b)$ is a banded matrix for all sufficiently large n. For these n, we change $T_n(b)$ to a circulant matrix by adding appropriate entries in the upper-right and lower-left corner blocks. For example, if

$$T_6(b) = \begin{pmatrix} b_0 & b_{-1} & 0 & 0 & 0 & 0 \\ b_1 & b_0 & b_{-1} & 0 & 0 & 0 \\ b_2 & b_1 & b_0 & b_{-1} & 0 & 0 \\ 0 & b_2 & b_1 & b_0 & b_{-1} & 0 \\ 0 & 0 & b_2 & b_1 & b_0 & b_{-1} \\ 0 & 0 & 0 & b_2 & b_1 & b_0 \end{pmatrix},$$

then

$$C_6(b) = \begin{pmatrix} b_0 & b_{-1} & 0 & 0 & b_2 & b_1 \\ b_1 & b_0 & b_{-1} & 0 & 0 & b_2 \\ b_2 & b_1 & b_0 & b_{-1} & 0 & 0 \\ 0 & b_2 & b_1 & b_0 & b_{-1} & 0 \\ 0 & 0 & b_2 & b_1 & b_0 & b_{-1} \\ b_{-1} & 0 & 0 & b_2 & b_1 & b_0 \end{pmatrix}.$$

We have

$$C_n(b) = U_n^* \,\text{diag}\,(b(1), b(\omega_n), \dots, b(\omega_n^{n-1}))\, U_n \tag{19}$$

where U_n is a unitary matrix and $\omega_n = e^{2\pi i/n}$. Thus, the singular values of $C_n(b)$ are $|b(\omega_n^j)|$ $(j = 0, \dots, n-1)$. The only trigonometric polynomials b of constant modulus are $b(t) = \alpha t^k$ $(t \in \mathbf{T})$ with $\alpha \in \mathbf{C}$, and in this case $\|C_n(b)x\| = |\alpha|\, \|x\|$ for all x.

Theorem 4.1 *Assume that $|b|$ is not constant. Then for each $\varepsilon > 0$ there exists an $n_0 = n_0(\varepsilon)$ such that*

$$P\left(\left| \frac{\|C_n(b)x\|}{\|C_n(b)\|\, \|x\|} - \frac{\|b\|_2}{\|b\|_\infty} \right| \geq \varepsilon \right) \leq \frac{3}{n+2} \frac{1}{\varepsilon^2} \frac{\|b\|_4^4 - \|b\|_2^2}{\|b\|_2^2\, \|b\|_\infty^2}$$

for all $n \geq n_0$.

Proof. The proof is analogous to the proof of (14). Note that now (13) amounts to the fact that the integral sum

$$\frac{s_1^k + \ldots + s_n^k}{n} = \sum_{j=0}^{n-1} |b(e^{2\pi i j/n})|^k \frac{1}{n}$$

converges to the Riemann integral

$$\int_0^1 |b(e^{2\pi i\theta})|^k d\theta = \int_0^{2\pi} |b(e^{i\theta})|^k \frac{d\theta}{2\pi} =: \|b\|_k^k.$$

Furthermore, it is obvious that $s_n = \max |b(\omega_n^j)| \to \|b\|_\infty$. ■

If b has no zeros on $\mathbf{T}$, then (19) shows that

$$C_n^{-1}(b) = U_n^* \,\text{diag}\,(b^{-1}(1), b^{-1}(\omega_n), \dots, b^{-1}(\omega_n^{n-1}))\, U_n$$

and hence the argument of the proof of Theorem 4.1 delivers

$$P\left(\left| \frac{\|C_n^{-1}(b)x\|}{\|C_n^{-1}(b)\|\, \|x\|} - \frac{\|b^{-1}\|_2}{\|b^{-1}\|_\infty} \right| \geq \varepsilon \right) \leq \frac{3}{n+2} \frac{1}{\varepsilon^2} \frac{\|b^{-1}\|_4^4 - \|b^{-1}\|_2^2}{\|b^{-1}\|_2^2\, \|b^{-1}\|_\infty^2} \tag{20}$$

for all sufficiently large n.

Example 4.2 Put $b(t) = 2 + \alpha + t + t^{-1}$, where $\alpha > 0$ is small. Thus,

$$C_5(b) = \begin{pmatrix} 2+\alpha & 1 & 0 & 0 & 1 \\ 1 & 2+\alpha & 1 & 0 & 0 \\ 0 & 1 & 2+\alpha & 1 & 0 \\ 0 & 0 & 1 & 2+\alpha & 1 \\ 1 & 0 & 0 & 1 & 2+\alpha \end{pmatrix}.$$

If n is large, then $\|C_n(b)\| \approx 4$ and $\|C_n^{-1}(b)\| \approx 1/\alpha$. Consequently, for the condition numbers defined in the introduction we have

$$\kappa(C_n(b), x) = \|C_n(b)\|\, \|C_n^{-1}(b)\| \approx \frac{4}{\alpha}$$

and

$$\kappa^{\rm circ}(C_n(b),x) = \frac{\|C_n^{-1}(b)x\|}{\|C_n^{-1}(b)\|\,\|x\|}\,\kappa(C_n(b),x) \approx \frac{4}{\alpha}\,\frac{\|C_n^{-1}(b)x\|}{\|C_n^{-1}(b)\|\,\|x\|}.$$

From (20) we therefore obtain that if n is sufficiently large, then with probability near 1,

$$\kappa^{\rm circ}(C_n(b),x) \approx \frac{4}{\alpha}\,\frac{\|b^{-1}\|_2}{\|b^{-1}\|_\infty} = \frac{4}{\alpha}\,\frac{\alpha^{1/4}(2+\alpha)^{1/2}}{(4+\alpha)^{3/4}} \approx \frac{\alpha^{1/4}}{2}\,\frac{4}{\alpha}.$$

For $\alpha = 0.01$ this gives

$$\kappa(C_n(b),x) \approx 400, \quad \kappa^{\rm circ}(C_n(b),x) \approx 63$$

with probability near 1, and for $\alpha = 0.0001$ we get

$$\kappa(C_n(b),x) \approx 40000, \quad \kappa^{\rm circ}(C_n(b),x) \approx 2000$$

with probability near 1. ■

5 Hankel Matrices

We begin with a general result.

Theorem 5.1 *Let $A = (a_{jk})_{j,k=1}^\infty$ be an infinite matrix and put $A_n = (a_{jk})_{j,k=1}^n$. If A induces a compact operator on ℓ^2, then $EX_n^2 \to 0$ as $n \to \infty$.*

Proof. We denote by $\mathcal{F}_j^\infty$ and $\mathcal{F}_j^n$ the operators of rank at most j on ℓ^2 and $\mathbf{C}^n$, respectively. The approximation numbers σ_j of A and A_n are defined by

$$\begin{aligned} \sigma_j(A) &= {\rm dist}\,(A,\mathcal{F}_j^\infty) = \inf\{\|A - F_j\| : F_j \in \mathcal{F}_j^\infty\}, \quad (j = 0,1,2,\dots),\\ \sigma_j(A_n) &= {\rm dist}\,(A_n,\mathcal{F}_j^n) = \inf\{\|A - F_j\| : F_j \in \mathcal{F}_j^n\}, \quad (j = 0,1,\dots,n-1). \end{aligned}$$

Note that the approximation numbers of A_n are just the singular values in reverse order, $s_{n-j}(A_n) = \sigma_j(A_n)$ for $j = 0,\dots,n-1$. Let P_n stand for the orthogonal projection onto the first n coordinates. If $F_j \in \mathcal{F}_j^\infty$, then $P_nF_jP_n$ may be identified with a matrix in $\mathcal{F}_j^n$. Furthermore, a matrix $F_j \in \mathcal{F}_j^n$ may be thought of as a matrix of the form $P_nF_jP_n$ with $F_j \in \mathcal{F}_j^\infty$. Thus, $\mathcal{F}_j^n = P_n\mathcal{F}_j^\infty P_n$ and it follows that

$$\begin{aligned} s_{n-j}(A_n) = \sigma_j(A_n) &= \inf\{\|P_nAP_n - F_j\| : F_j \in \mathcal{F}_j^n\}\\ &= \inf\{\|P_nAP_n - P_nF_jP_n\| : F_j \in \mathcal{F}_j^\infty\}\\ &\le \inf\{\|A - F_j\| : F_j \in \mathcal{F}_j^\infty\} = \sigma_j(A). \end{aligned}$$

From Theorem 2.2 we therefore obtain

$$EX_n^2 = \frac{1}{n}\sum_{k=1}^n s_k^2(A_n) = \frac{1}{n}\sum_{j=0}^{n-1} s_{n-j}^2(A_n) \le \frac{1}{n}\sum_{j=0}^{n-1}\sigma_j^2(A). \tag{21}$$

But if A is compact, then $\sigma_j(A) \to 0$ as $j \to \infty$. This implies that the right-hand side of (21) goes to zero as $n \to \infty$. ■

An interesting concrete situation is the case where $A = H(b)$ is the Hankel matrix $(b_{j+k-1})_{j,k=1}^{\infty}$ generated by the Fourier coefficients of a function $b \in L^1$. If b is continuous, then the Hankel matrix $H(b)$ induces a compact operator and hence, by Theorem 5.1, $EX_n^2 \to 0$. The following result shows that, surprisingly, $EX_n^2 \to 0$ for all Hankel matrices with L^1 symbols (notice that such matrices need not even generate bounded operators).

Theorem 5.2 *Let $b \in L^1$ and let A_n be the $n \times n$ principal section of $H(b)$. Then $EX_n^2 \to 0$.*

Proof. As shown by Fasino and Tilli [6], [15],

$$\lim_{n\to\infty} \frac{F(s_1) + \ldots + F(s_n)}{n} = F(0)$$

for every uniformly continuous and bounded function F on $\mathbf{R}$. Suppose first that $H(b)$ induces a bounded operator. Then $\|A_n\| \le \|H(b)\| =: d < \infty$, which implies that all singular values of A_n lie in the segment $[0, d]$. Thus, letting F be a smooth and bounded function such that $F(x) = x^2$ on $[0, d]$, we deduce that $\sum s_j^2/n \to 0$. Since b is not identically zero, there is an N such that $\|A_N\| > 0$. It follows that $0 < \|A_N\| \le \|A_n\| = s_n$ for all $n \ge n$. Thus, $\sum s_j^2/(ns_n^2) \to 0$, and Theorem 2.2 gives the assertion.

Now suppose that $H(b)$ is not bounded. We claim that then $\|A_n\| \to \infty$. Indeed, the sequence $\{\|A_n\|\}$ is monotonically increasing: $\|A_n\| \le \|A_{n+1}\|$ for all n. If there exists a finite constant M such that $\|A_n\| \le M$ for all n, then $\{A_n x\}$ is a convergent sequence for each $x \in \ell^2$. The Banach-Steinhaus theorem (= uniform boundedness principle) therefore implies that the operator A defined by $Ax := \lim A_n x$ is bounded on ℓ^2. But A is clearly given by the matrix $H(b)$. This contradiction proves that $\|A_n\| \to \infty$. Finally, Fasino and Tilli [6], [15] proved that always

$$\frac{1}{n}\|A_n\|_{\mathrm{tr}} \le 2\|b\|_1.$$

Lemma 2.4 now shows that $EX_n^2 \to 0$. ■

6 Toeplitz Matrices with Unbounded Symbols

Following Tyrtyshnikov and Zamarashkin [18], we consider Toeplitz matrices generated by so-called Radon measures. Thus, given a function $\beta : [-\pi, \pi] \to \mathbf{C}$ of bounded variation, we define

$$(d\beta)_k = \frac{1}{2\pi}\int_{-\pi}^{\pi} e^{-ik\theta} d\beta(\theta), \tag{22}$$

the integral understood in the Riemann-Stieltjes sense, and we put

$$T_n(d\beta) = ((d\beta)_{j-k})_{j,k=1}^n.$$

If β is absolutely continuous, then $\beta' \in L^1[-\pi,\pi]$ and $(d\beta)_k$ is nothing but the kth Fourier coefficient of β', defined in accordance with (12). Consequently, in this case $T_n(d\beta)$ is just what we denoted by $T_n(\beta')$ in Section 3.

For general β we have $\beta = \beta_{\rm a} + \beta_{\rm j} + \beta_{\rm s}$ where $\beta_{\rm a}$ is absolutely continuous with $\beta_{\rm a}' \in L^1[-\pi,\pi]$, $\beta_{\rm j}$ is the "jump part", that is, a function of the form

$$\beta_{\rm j}(\theta) = \sum_{\theta_\ell < \theta} h_\ell, \quad \sum_\ell |h_\ell| < \infty,$$

with an at most countable set $\{\theta_1, \theta_2, \dots\} \subset [-\pi,\pi)$, and $\beta_{\rm s}$ is the "singular part", that is, a continuous function of bounded variation whose derivative vanishes almost everywhere. This decomposition is unique up to constant additive terms. After partial integration (see, e.g., [7, No. 577]), formula (22) can be written more explicitly as

$$\begin{aligned}(d\beta)_k &= \frac{1}{2\pi}\int_{-\pi}^{\pi} e^{-ik\theta}\beta_{\rm a}'(\theta)d\theta + \frac{1}{2\pi}\sum_\ell h_\ell e^{-i\theta_\ell k} \\ &\quad + \frac{(-1)^k}{2\pi}\Big(\beta_{\rm s}(\pi) - \beta_{\rm s}(-\pi)\Big) + \frac{ik}{2\pi}\int_{-\pi}^{\pi} e^{-ik\theta}\beta_{\rm s}(\theta)d\theta.\end{aligned}$$

In particular, if $\beta(\theta) = 0$ for $\theta \in [-\pi, 0]$ and $\beta(\theta) = 2\pi$ for $\theta \in (0,\pi]$, then $(d\beta)_k = 1$ for all k, that is, $T_n(d\beta)$ is the matrix (10).

Theorem 6.1 *Let $\beta = \beta_{\rm a} + \beta_{\rm j} + \beta_{\rm s}$ be a nonconstant function of bounded variation and put $A_n = T_n(d\beta)$. Then EX_n^2 converges to a limit as $n \to \infty$. This limit is positive if and only if $\beta_{\rm a}' \in L^\infty[-\pi,\pi]$ and $\beta_{\rm j} = \beta_{\rm s} = 0$.*

Proof. If $\beta = \beta_{\rm a}$ with $\beta_{\rm a}' \in L^\infty[-\pi,\pi]$, then EX_n^2 converges to $\|\beta_{\rm a}'\|_2/\|\beta_{\rm a}'\|_\infty \neq 0$ due to Theorems 3.2 and 3.3.

So assume $\beta = \beta_{\rm a} + \beta_{\rm j} + \beta_{\rm s}$. Write $\beta = \beta_1 - \beta_2 + i(\beta_3 - \beta_4)$ with nonnegative functions β_k of bounded variation. We have

$$\frac{1}{n}\|T_n(d\beta)\|_{\rm tr} \le \frac{1}{n}\sum_{k=1}^{4}\|T_n(d\beta_k)\|_{\rm tr},$$

The singular values of the positively semi-definite matrices $T_n(d\beta_k)$ coincide with the eigenvalues. Hence

$$\frac{1}{n}\|\, T_n(\beta_k)\|_{\rm tr} = \frac{1}{n}\operatorname{tr} T_n(\beta_k) = (d\beta_k)_0 = \frac{1}{2\pi}\int_{-\pi}^{\pi} d\beta_k \le \frac{1}{2\pi}\operatorname{Var}\beta_k < \infty$$

(this argument is standard; see, e.g., [18]). Consequently, there is a finite constant M such that

$$\frac{1}{n}\|T_n(d\beta)\|_{\rm tr} \le M$$

for all n. The sequence $\{\|T_n(d\beta)\|\}$ is monotonically increasing, that is, $\|T_n(d\beta)\| \le \|T_{n+1}(d\beta)\|$ for all n. We show that if this sequence is bounded, then necessarily $\beta = \beta_{\rm a}$

with $\beta_{\rm a}' \in L^\infty[-\pi,\pi]$. By virtue of Lemma 2.4, this implies that $EX_n^2 \to 0$ whenever $\beta_{\rm a}' \notin L^\infty[-\pi,\pi]$ or $\beta_{\rm j} \neq 0$ or $\beta_{\rm s} \neq 0$.

Thus, suppose there is a finite constant C such that $\|T_n(d\beta)\| \le C$ for all n. Let $e_j \in \mathbf{C}^n$ be the jth vector of the standard basis. Then

$$\|T_n(d\beta)e_1\|_2^2 = |(d\beta)_0|^2 + \ldots + |(d\beta)_{n-1}|^2 \le C^2,$$
$$\|T_n(d\beta)e_n\|_2^2 = |(d\beta)_0|^2 + \ldots + |(d\beta)_{-(n-1)}|^2 \le C^2$$

for all n, which tells us that there is a function $b \in L^2[-\pi,\pi]$ such that $(d\beta)_k = b_k$ for all k. Since the decomposition of β into the absolutely continuous part, the jump part, and the singular part is unique (up to additive constants), it follows that $\beta = \beta_{\rm a} + \beta_{\rm j} + \beta_{\rm s}$ with $\beta_{\rm a}' = b$ and $\beta_{\rm j} = \beta_{\rm s} = 0$. We are left to show that b is in $L^\infty[-\pi,\pi]$. Using the Banach-Steinhaus theorem as in the proof of Theorem 5.2 we arrive at the conclusion that the Toeplitz matrix $T(b) := (b_{j-k})_{j,k=1}^\infty$ induces a bounded operator on ℓ^2. By a classical theorem of Toeplitz [16] (full proofs are also in [3] and [8]), this happens if and only if b is in $L^\infty[-\pi,\pi]$. ■

Theorem 6.1 reveals in particular that $EX_n^2 \to 0$ if $A_n = T_n(b)$ with $b \in L^1 \setminus L^\infty$. The following theorem concerns a class of Toeplitz matrices with increasing entries. The notation $c_j \simeq d_j$ means that c_j/d_j remains bounded and bounded away from zero.

Theorem 6.2 *Let $A_n = (b_{j-k})_{j,k=1}^n$ where $|b_j| \simeq e^{\gamma j}$ as $j \to +\infty$ and $|b_{-j}| \simeq e^{\delta j}$ as $j \to +\infty$. If one of the numbers γ and δ is positive, then $EX_n^2 = O(1/n)$ as $n \to \infty$.*

Proof. For the sake of definiteness, suppose $\gamma > 0$. We have

$$\begin{aligned} \|A_n\|_{\rm F}^2 &= \sum_{j=0}^{n-1}(n-j)|b_j|^2 + \sum_{j=1}^{n-1}(n-j)|b_{-j}|^2 \\ &\le C_1\sum_{j=0}^{n-1}(n-j)e^{2\gamma j} + C_1\sum_{j=1}^{n-1}(n-j)e^{2\delta j} \end{aligned}$$

with some finite constant C_1. In the cases $\delta > 0$ and $\delta \le 0$, this gives

$$\|A_n\|_{\rm F}^2 \le C_2\left(e^{2\gamma n} + e^{2\delta n}\right) \quad \text{and} \quad \|A_n\|_{\rm F}^2 \le C_2 e^{2\gamma n}$$

with some finite constant C_2, respectively; note that, for example,

$$\sum_{j=0}^{n-1}(n-j)e^{2\gamma j} = \frac{e^{2\gamma n}}{(e^{2\gamma}-1)^2} + O(n).$$

On the other hand, considering $\|A_n e_1\|_2^2$ and $\|A_n e_n\|_2^2$, we see that

$$\|A_n\|^2 \ge \frac{1}{2}\sum_{j=0}^{n-1}|b_j|^2 + \frac{1}{2}\sum_{j=1}^{n-1}|b_{-j}|^2, \tag{23}$$

which shows that there is a finite constant $C_3 > 0$ such that

$$\|A_n\|^2 \geq C_3 \left(e^{2\gamma n} + e^{2\delta n} \right) \quad \text{and} \quad \|A_n\|^2 \geq C_3 e^{2\gamma n}$$

for $\delta > 0$ and $\delta \leq 0$, respectively. Thus, in either case,

$$EX_n^2 = \frac{\|A_n\|_{\mathrm{F}}^2}{n\|A_n\|^2} = O\left(\frac{1}{n}\right). \quad \blacksquare$$

Example 6.3 Let

$$A_n = \begin{pmatrix} a & b\varrho & b\varrho^2 & \dots & b\varrho^{n-1} \\ c\sigma & a & b\varrho & \dots & b\varrho^{n-2} \\ c\sigma^2 & c\sigma & a & \dots & b\varrho^{n-3} \\ \dots & \dots & \dots & \dots & \dots \\ c\sigma^{n-1} & c\sigma^{n-2} & c\sigma^{n-3} & \dots & a \end{pmatrix}.$$

Similar matrices are studied in [17]. In the case $a = b = c$ and $\sigma = \varrho$, such matrices are called Kac-Murdock-Szegö matrices [10]. Suppose that a, b, c are nonzero. If $|\sigma| < 1$ and $|\varrho| < 1$, then EX_n^2 converges to a nonzero limit by Theorem 2.2 and (13). If $|\sigma| > 1$ or $|\varrho| > 1$, we can invoke Theorem 6.2 to deduce that $EX_n^2 \to 0$. Finally, in the two cases where $|\sigma| \leq 1 = |\varrho|$ or $|\varrho| \leq 1 = |\sigma|$, Theorem 6.1 implies that $EX_n^2 \to 0$. $\blacksquare$

7 Appendix: Distribution Functions

The referee suggested that it would be interesting to compute the distribution of X_n^2 in some cases and noted that this can probably be done easily for small n and for the matrix of Example 2.3. The purpose of this section is to address this problem. It will turn out that the referee is right in all respects.

Let $A_n \in M_n(\mathbf{R})$ and let $0 \leq s_1 \leq \ldots \leq s_n$ be the singular values of A_n. Suppose $s_n > 0$. The random variable $X_n^2 = \|A_n x\|^2/\|A_n\|^2$ assumes its values in $[0, 1]$. With notation as in the proof of Theorem 2.2,

$$E_\xi := \left\{ x \in S_{n-1} : \frac{\|A_n x\|^2}{\|A_n\|^2} < \xi \right\} = \left\{ x \in S_{n-1} : \frac{\|D_n V_n x\|^2}{s_n^2} < \xi \right\}.$$

Put $G_\xi = \{x \in S_{n-1} : \|D_n x\|^2/s_n^2 < \xi\}$. Clearly, $G_\xi = V_n(E_\xi)$. Since V_n is an orthogonal matrix, it leaves the surface measure on S_{n-1} invariant. It follows that $|G_\xi| = |V_n(E_\xi)|$ and hence

$$F_n(\xi) := P(X_n^2 < \xi) = P\left(\frac{\|D_n x\|^2}{s_n^2} < \xi \right) = P\left(\frac{s_1^2 x_1^2 + \ldots + s_n^2 x_n^2}{s_n^2} < \xi \right). \qquad (24)$$

This reveals first of all that the distribution function $F_n(\xi)$ depends only on the singular values of A_n.

A real-valued random variable X is said to be $B(\alpha, \beta)$ distributed on (a, b) if

$$P(c \leq X < d) = \int_c^d f(\xi)\, d\xi$$

where the density function $f(\xi)$ is zero on $(-\infty, a]$ and $[b, \infty)$ and equals

$$\frac{(b-a)^{1-\alpha-\beta}}{B(\alpha,\beta)}\,(\xi-a)^{\alpha-1}(b-\xi)^{\beta-1}$$

on (a,b). Here $B(\alpha,\beta)=\Gamma(\alpha)\Gamma(\beta)/\Gamma(\alpha+\beta)$ is the common beta function and it is assumed that $\alpha>0$ and $\beta>0$. The emergence of the beta distribution, of the χ^2 distribution, of elliptic integrals and Bessel functions in connection with uniform distribution on the unit sphere is no surprise and can be found throughout the literature. Thus, the following results are not at all new. However, they tell a nice story and uncover the astonishing simplicity of Theorem 2.2.

We first consider 2×2 matrices, that is, we let $n=2$. From (24) we infer that

$$F_2(\xi)=P\left(\frac{s_1^2}{s_2^2}x_1^2+x_2^2<\xi\right). \tag{25}$$

The constellation $s_1=s_2$ is uninteresting, because $F_2(\xi)=0$ for $\xi<1$ and $F_2(\xi)=1$ for $\xi\ge 1$ in this case.

Theorem 7.1 *If $s_1<s_2$, then the random variable X_2^2 is subject to the $B\left(\frac{1}{2},\frac{1}{2}\right)$ distribution on $(s_1^2/s_2^2,1)$.*

Proof. Put $\tau=s_1/s_2$. By (25), $F_2(\xi)$ is $\frac{1}{2\pi}$ times the length of the piece of the unit circle $x_1^2+x_2^2=1$ that is contained in the interior of the ellipse $\tau^2x_1^2+x_2^2=\xi$. This gives $F_2(\xi)=0$ for $\xi\le\tau^2$ and $F_2(\xi)=1$ for $\xi\ge 1$. Thus, let $\xi\in(\tau^2,1)$. Then the circle and the ellipse intersect at the four points

$$\left(\pm\sqrt{\frac{1-\xi}{1-\tau^2}},\ \pm\sqrt{\frac{\xi-\tau^2}{1-\tau^2}}\right),$$

and consequently,

$$F_2(\xi)=\frac{2}{\pi}\arctan\sqrt{\frac{\xi-\tau^2}{1-\xi}},$$

which implies that $F_2'(\xi)$ equals

$$\frac{1}{\pi}\,(\xi-\tau^2)^{-1/2}(1-\xi)^{-1/2}=\frac{1}{B(1/2,1/2)}\,(\xi-\tau^2)^{-1/2}(1-\xi)^{-1/2}$$

and proves that X_2^2 has the $B\left(\frac{1}{2},\frac{1}{2}\right)$ distribution on $(\tau^2,1)$. ■

In the general case, things are more involved. An idea of the variety of possible distribution functions is provided by the class of matrices whose singular values satisfy

$$0=s_1=\ldots=s_{n-m}<s_{n-m+1}\le\ldots\le s_n \tag{26}$$

with small m. Notice that m is just the rank of the matrix. We put

$$\mu_{n-m+1}=\frac{s_n}{s_{n-m+1}},\quad \mu_{n-m+2}=\frac{s_n}{s_{n-m+2}},\quad \ldots\quad,\quad \mu_n=\frac{s_n}{s_n}\ (=1).$$

Our problem is to find

$$F_n(\xi) = P\left(\frac{x_{n-m+1}^2}{\mu_{n-m+1}^2} + \ldots + \frac{x_n^2}{\mu_n^2} < \xi\right); \tag{27}$$

the dependence of F_n on m and $\mu_{n-m+1}, \ldots, \mu_n$ will be suppressed.

Example 7.2 In order to illustrate what will follow by a transparent special case, we take $n = 3$ and suppose that the singular values of A_3 satisfy $0 = s_1 < s_2 < s_3$. We put $\mu = s_3/s_2$. Clearly, (27) becomes $F_3(\xi) = P\left(\frac{x_2^2}{\mu^2} + x_3^2 < \xi\right)$. We rename x_1, x_2, x_3 to x, y, z. The preceding equality tells us that $F_3(\xi)$ is $\frac{1}{4\pi}$ times the area of the piece $\widetilde{\Sigma}$ of the sphere $x^2 + y^2 + z^2 = 1$ that is cut out by the elliptic cylinder $y^2/\mu^2 + z^2 < \xi$. Let us assume that $\xi \in (0, 1/\mu^2)$. Then the ellipse $y^2/\mu^2 + z^2 < \xi$ is completely contained in the disk $y^2 + z^2 < 1$. By symmetry, it suffices to consider the part Σ of $\widetilde{\Sigma}$ that lies in the octant $x \geq 0, y \geq 0, z \geq 0$. We have

$$F_3(\xi) = \frac{8}{4\pi} \int_\Sigma d\sigma,$$

and it easily verified (see the proof of Theorem 7.3) that

$$\frac{8}{4\pi} \int_\Sigma d\sigma = \frac{8}{4\pi/3} \int_{\mathrm{co}(0,\Sigma)} dx\, dy\, dz,$$

where $\mathrm{co}(0, \Sigma)$ is the cone $\cup_{s \in \Sigma}[0, s]$ (we prefer volume integrals to surface integrals). A parametrization of Σ is

$$\begin{aligned} y &= \mu r \cos\varphi \\ z &= r \sin\varphi \\ x &= \sqrt{1 - r^2(\mu^2 \cos^2\varphi + \sin^2\varphi)}, \end{aligned}$$

where $r \in [0, \sqrt{\xi}\,)$ and $\varphi \in [0, \pi/2]$. Consequently, a parametrization of $\mathrm{co}(0, \Sigma)$ is given by

$$\begin{aligned} y &= t\mu r \cos\varphi \\ z &= tr \sin\varphi \\ x &= t\sqrt{1 - r^2(\mu^2 \cos^2\varphi + \sin^2\varphi)} \end{aligned}$$

with $t \in [0, 1]$ and r and φ as before. We set $v(\varphi) = \mu^2 \cos^2\varphi + \sin^2\varphi$ and denote the Jacobian $\partial(y, z, x)/\partial(t, r, \varphi)$ by J. By what was said above,

$$F_3(\xi) = \frac{6}{\pi} \int_{\mathrm{co}(0,\Sigma)} dx\, dy\, dz = \frac{6}{\pi} \int_0^{\pi/2} \int_0^{\sqrt{\xi}} \int_0^1 |J|\, dt\, dr\, d\varphi.$$

Writing down J and subtracting r/t times the second column from the first we get

$$\begin{aligned}
J &= \begin{vmatrix} \mu r\cos\varphi & t\mu\cos\varphi & -t\mu r\sin\varphi \\ r\sin\varphi & t\sin\varphi & tr\cos\varphi \\ \sqrt{1-r^2v} & \dfrac{-trv}{\sqrt{1-r^2v}} & t\dfrac{\partial}{\partial\varphi}\sqrt{1-r^2v} \end{vmatrix} \\
&= \begin{vmatrix} 0 & t\mu\cos\varphi & -t\mu r\sin\varphi \\ 0 & t\sin\varphi & tr\cos\varphi \\ \dfrac{1}{\sqrt{1-r^2v}} & \dfrac{-trv}{\sqrt{1-r^2v}} & t\dfrac{\partial}{\partial\varphi}\sqrt{1-r^2v} \end{vmatrix} \\
&= \frac{t^2\mu r}{\sqrt{1-r^2v}}.
\end{aligned}$$

It follows that

$$\begin{aligned}
F_3(\xi) &= \frac{6}{\pi}\int_0^{\pi/2}\int_0^{\sqrt{\xi}}\int_0^1 \frac{t^2\mu r}{\sqrt{1-r^2v(\varphi)}}\,dt\,dr\,d\varphi \\
&= \frac{2\mu}{\pi}\int_0^{\pi/2}\int_0^{\sqrt{\xi}} \frac{r}{\sqrt{1-r^2v(\varphi)}}\,dr\,d\varphi \\
&= \frac{\mu}{\pi}\int_0^{\pi/2}\int_0^{\xi} \frac{1}{\sqrt{1-sv(\varphi)}}\,ds\,d\varphi.
\end{aligned}$$

Thus, the density function $f_3(\xi) = F_3'(\xi)$ is

$$f_3(\xi) = \frac{\mu}{\pi}\int_0^{\pi/2}\frac{d\varphi}{\sqrt{1-\xi v(\varphi)}}.$$

We have

$$\begin{aligned}
1-\xi v(\varphi) &= 1-\xi(\mu^2\cos^2\varphi+\sin^2\varphi) \\
&= 1-\xi\mu^2\cos^2\varphi-\xi+\xi\cos^2\varphi \\
&= (1-\xi)\left(1-\frac{\xi(\mu^2-1)}{1-\xi}\cos^2\varphi\right), \qquad (28)
\end{aligned}$$

whence

$$\begin{aligned}
f_3(\xi) &= \frac{\mu}{\pi\sqrt{1-\xi}}\int_0^{\pi/2}\frac{d\varphi}{\sqrt{1-\dfrac{\xi(\mu^2-1)}{1-\xi}\cos^2\varphi}} \\
&= \frac{\mu}{\pi\sqrt{1-\xi}}\int_0^{\pi/2}\frac{d\varphi}{\sqrt{1-\dfrac{\xi(\mu^2-1)}{1-\xi}\sin^2\varphi}} \\
&= \frac{\mu}{\pi\sqrt{1-\xi}}\,\mathsf{K}\left(\sqrt{\frac{\xi(\mu^2-1)}{1-\xi}}\right)
\end{aligned}$$

with the standard complete elliptic integral K. ■

We now return to the situation given by (26). Let $Q = [0, \pi/2]$. For $\varphi_1, \dots, \varphi_{k-1}$ in Q we introduce the spherical coordinates $\omega_1^{(k)}, \dots, \omega_k^{(k)}$ by

$$\begin{aligned}
\omega_1^{(k)} &= \cos\varphi_1 \\
\omega_2^{(k)} &= \sin\varphi_1 \cos\varphi_2 \\
\omega_3^{(k)} &= \sin\varphi_1 \sin\varphi_2 \cos\varphi_3 \\
&\dots \\
\omega_{k-1}^{(k)} &= \sin\varphi_1 \sin\varphi_2 \dots \sin\varphi_{k-2} \cos\varphi_{k-2} \\
\omega_k^{(k)} &= \sin\varphi_1 \sin\varphi_2 \dots \sin\varphi_{k-2} \sin\varphi_{k-2}.
\end{aligned}$$

Notice that

$$\frac{\partial(r\omega_1^{(k)}, \dots, r\omega_k^{(k)})}{\partial(r, \varphi_1, \dots, \varphi_{k-1})} = r^{k-1} \sin^{k-2}\varphi_1 \sin^{k-3}\varphi_2 \dots \sin\varphi_{k-2}, \tag{29}$$

$$\int_{Q^{k-1}} \sin^{k-2}\varphi_1 \sin^{k-3}\varphi_2 \dots \sin\varphi_{k-2}\, d\varphi_1 \dots d\varphi_{k-1} = \frac{1}{2^{k-1}} \frac{\pi^{k/2}}{\Gamma(k/2)}. \tag{30}$$

We define $v = v(\varphi_1, \dots, \varphi_{m-1})$ by

$$v = \mu_{n-m+1}^2 [\omega_1^{(m)}]^2 + \mu_{n-m+2}^2 [\omega_2^{(m)}]^2 + \dots + \mu_n^2 [\omega_{n-m}^{(m)}]^2.$$

Theorem 7.3 *Let $n \geq 3$ and suppose the singular values of A_n satisfy (26). Then for $\xi \in (0, 1/\mu_{n-m+1}^2)$, the density function of X_n^2 is*

$$f_n(\xi) = c_n\, \xi^{(m-2)/2} \int_{Q^{m-1}} (1 - \xi v(\varphi))^{(n-m-2)/2} \sin^{m-2}\varphi_1 \sin^{m-3}\varphi_2 \dots \sin\varphi_{m-2}\, d\varphi$$

with

$$c_n = \frac{2^{m-1}}{\pi^{m/2}} \frac{\Gamma\left(\dfrac{n}{2}\right)}{\Gamma\left(\dfrac{n-m}{2}\right)} \mu_{n-m+1} \dots \mu_n.$$

Proof. We proceed as in Example 7.2. Let Σ denote the set of all points $(x_1, \dots, x_n)$ for which $x_1^2 + \dots + x_n^2 = 1$, $x_1 \geq 0, \dots, x_n \geq 0$, and

$$\frac{x_{n-m+1}^2}{\mu_{n-m+1}^2} + \dots + \frac{x_n^2}{\mu_n^2} < \xi. \tag{31}$$

We have

$$F_n(\xi) = \frac{2^n}{|S_{n-1}|} \int_\Sigma d\sigma.$$

We prefer to switch from the surface integral to a volume integral. Let $\mathrm{co}(0, \Sigma)$ denote the cone formed by all segments $[0, s]$ with $s \in \Sigma$. Because $|S_{n-1}| = n|B_n|$, we get

$$\frac{1}{|S_{n-1}|} \int_\Sigma d\sigma = \frac{1}{n|B_n|} \int_\Sigma d\sigma = \frac{1}{|B_n|} \int_0^1 \int_\Sigma t^{n-1}\, d\sigma\, dt = \frac{1}{|B_n|} \int_{\mathrm{co}(0,\Sigma)} dx.$$

Since $\xi\mu_{n-m+j}^2 < 1$ for $j = 1, \dots, m$, the ellipsoid (31) is completely contained in the ball

$$x_{n-m+1}^2 + \dots + x_n^2 < 1.$$

We start with the parametrization

$$\begin{aligned}
x_{n-m+1} &= \mu_{n-m+1} r\, \omega_1^{(m)}(\varphi_1, \dots, \varphi_{m-1}) \\
x_{n-m+2} &= \mu_{n-m+2} r\, \omega_2^{(m)}(\varphi_1, \dots, \varphi_{m-1}) \\
&\dots \\
x_n &= \mu_n r\, \omega_m^{(m)}(\varphi_1, \dots, \varphi_{m-1}),
\end{aligned}$$

where $r \in [0, \sqrt{\xi})$ and $(\varphi_1, \dots, \varphi_{m-1}) \in Q^{m-1}$. By the definition of v,

$$x_1^2 + \dots + x_{n-m}^2 = 1 - x_{n-m+1}^2 - \dots - x_n^2 = 1 - r^2 v.$$

Hence, after letting $(\theta_1, \dots, \theta_{n-m-1}) \in Q^{n-m-1}$ and

$$\begin{aligned}
x_1 &= \sqrt{1 - r^2 v(\varphi_1, \dots, \varphi_{m-1})}\, \omega_1^{(n-m)}(\theta_1, \dots, \theta_{n-m-1}) \\
x_2 &= \sqrt{1 - r^2 v(\varphi_1, \dots, \varphi_{m-1})}\, \omega_2^{(n-m)}(\theta_1, \dots, \theta_{n-m-1}) \\
&\dots \\
x_{n-m} &= \sqrt{1 - r^2 v(\varphi_1, \dots, \varphi_{m-1})}\, \omega_{n-m}^{(n-m)}(\theta_1, \dots, \theta_{n-m-1})
\end{aligned}$$

we have accomplished the parametrization of Σ. Finally, on multiplying the right-hand sides of the above expressions for $x_1, \dots, x_n$ by $t \in [0, 1]$, we obtain a parametrization of $\mathrm{co}(0, \Sigma)$. The Jacobian

$$J = \frac{\partial(x_1, \dots, x_n)}{\partial(t, r, \varphi_1, \dots, \varphi_{m-1}, \theta_1, \dots, \theta_{n-m-1})}$$

can be evaluated as in Example 7.2: after subtracting r/t times the second column from the first and taking into account that

$$\sqrt{1 - r^2 v} - r\, \frac{\partial}{\partial r}\, \sqrt{1 - r^2 v} = \frac{1}{\sqrt{1 - r^2 v}}$$

one arrives at a determinant that is the product of an $m \times m$ determinant and an $(n - m) \times (n - m)$ determinant; these two determinants can in turn be computed using (29). What results is

$$\begin{aligned}
|J| &= t^{n-1} r^{m-1} \mu_{n-m+1} \cdots \mu_n\, (1 - r^2 v)^{(n-m-2)/2} \\
&\quad \times\ \sin^{m-2}\varphi_1 \sin^{m-3}\varphi_2 \dots \sin\varphi_{m-2} \\
&\quad \times\ \sin^{n-m-2}\theta_1 \sin^{n-m-3}\theta_2 \dots \sin\theta_{n-m-2}.
\end{aligned}$$

In summary,

$$\begin{aligned}
F_n(\xi) &= \frac{2^n}{|B_n|}\int_{\mathrm{co}(0,\Sigma)} dx \\
&= \frac{2^n}{|B_n|}\int_0^1\int_0^{\sqrt{\xi}}\int_{Q^{m-1}}\int_{Q^{n-m-1}} |J|\, d\theta\, d\varphi\, dr\, dt \\
&= 2c_n\int_0^{\sqrt{\xi}} r^{m-1}\int_{Q^{m-1}}(1-r^2v)^{(n-m-2)/2}\sin^{m-2}\varphi_1\ldots\sin\varphi_{m-2}\, d\varphi\, dr \\
&= c_n\int_0^{\xi} s^{(m-2)/2}\int_{Q^{m-1}}(1-sv)^{(n-m-2)/2}\sin^{m-2}\varphi_1\ldots\sin\varphi_{m-2}\, d\varphi\, ds
\end{aligned}$$

with c_n as in the theorem. Consequently,

$$F_n'(\xi) = c_n\xi^{(m-2)/2}\int_{Q^{m-1}}(1-\xi v)^{(n-m-2)/2}\sin^{m-2}\varphi_1\ldots\sin\varphi_{m-2}\, d\varphi. \quad \blacksquare$$

Corollary 7.4 *Let $n \geq 3$. If $s_1 = \ldots = s_{n-m} = 0$ and $s_{n-m+1} = \ldots = s_n$, then the random variable X_n^2 is $B\left(\frac{m}{2}, \frac{n-m}{2}\right)$ distributed on $(0,1)$.*

Proof. This is the case $\mu_{n-m+1} = \ldots = \mu_n = 1$. The function v is identically 1 and hence, by (30),

$$\begin{aligned}
&\int_{Q^{m-1}}(1-\xi v(\varphi))^{(n-m-2)/2}\sin^{m-2}\varphi_1\ldots\sin\varphi_{m-2}\, d\varphi_1\ldots d\varphi_{m-1} \\
&= (1-\xi)^{(n-m-2)/2}\,\frac{1}{2^{m-1}}\,\frac{\pi^{m/2}}{\Gamma(m/2)}.
\end{aligned}$$

From Theorem 7.3 we therefore deduce that the density function $f_n(\xi)$ is a constant times $\xi^{(m-2)/2}(1-\xi)^{(n-m-2)/2}$. The constant is

$$\frac{2^{m-1}}{\pi^{m/2}}\,\frac{\Gamma\left(\frac{n}{2}\right)}{\Gamma\left(\frac{n-m}{2}\right)}\,\frac{1}{2^{m-1}}\,\frac{\pi^{m/2}}{\Gamma\left(\frac{m}{2}\right)} = \frac{1}{B\left(\frac{m}{2},\frac{n-m}{2}\right)}. \quad \blacksquare$$

Under the hypothesis of Corollary 7.4, the density function of X_n^2 is

$$f_n(\xi) = \frac{1}{B\left(\frac{m}{2},\frac{n-m}{2}\right)}\,\xi^{(m-2)/2}(1-\xi)^{(n-m-2)/2}.$$

It follows that the random variable nX_n^2 has the density

$$\frac{1}{n}f_n\left(\frac{\xi}{n}\right) = \frac{\Gamma\left(\frac{n}{2}\right)}{\Gamma\left(\frac{m}{2}\right)\Gamma\left(\frac{n-m}{2}\right)}\,\frac{1}{n^{m/2}}\,\xi^{(m-2)/2}\left(1-\frac{\xi}{n}\right)^{(n-m-2)/2}$$

on $(0,n)$. If m remains fixed and n goes to infinity, then this has the limit

$$\frac{1}{2^{m/2}\Gamma(m/2)}\,\xi^{(m-2)/2}e^{-\xi/2}, \quad \xi \in (0,\infty),$$

which is the density of the χ_m^2 distribution.

Example 7.5 Let us consider Example 2.3 again. Thus, suppose A_n is the matrix (10). The singular values of A_n are $0,\ldots,0,n$ and hence we can apply Corollary 7.4 with $m=1$ to the situation at hand. It follows that X_n^2 is $B\left(\frac{1}{2},\frac{n-1}{2}\right)$ distributed on the interval $(0,1)$. If X has the $B(\alpha,\beta)$ distribution on $(0,1)$, then

$$EX = \frac{\alpha}{\alpha+\beta}, \quad \sigma^2 X = \frac{\alpha\beta}{(\alpha+\beta)^2(\alpha+\beta+1)}.$$

This yields

$$EX_n^2 = \frac{1}{n}, \quad \sigma^2 X_n^2 = \frac{2(n-1)}{n^2(n+2)},$$

which is in perfect accordance with (11). In Example 2.3 we were able to conclude that $P(X_n^2 \ge \varepsilon) \le 8/(n^2\varepsilon^2)$. Since we know the density, we can now write

$$P(X_n^2 \ge \varepsilon) = \frac{1}{B\left(\dfrac{1}{2},\dfrac{n-1}{2}\right)}\int_\varepsilon^1 \xi^{-1/2}(1-\xi)^{(n-3)/2}\,d\xi.$$

Once partially integrating and using Stirling's formula we obtain

$$P(X_n^2 \ge \varepsilon) = \sqrt{\frac{2}{\pi}}\,\frac{1}{\sqrt{n\varepsilon}}\,(1-\varepsilon)^{(n-1)/2}\left(1+O\left(\frac{1}{n}\right)\right),$$

the O depending on ε. Thus, for each $\varepsilon \in (0,1)$, the probability $P(X_n^2 \ge \varepsilon)$ actually decays exponentially to zero as $n \to \infty$. ■

Example 7.6 Orthogonal projections have just the singular value pattern of Corollary 7.4. This leads to some pretty nice conclusions.

Let E be an N-dimensional Euclidean space and let U be an m-dimensional linear subspace of E. We denote by $\mathcal{P}_U$ the orthogonal projection of E onto U. Then for $y \in E$, the element $\mathcal{P}_U y$ is the best approximation of y in U and we have $\|y\|^2 = \|\mathcal{P}_U y\|^2 + \|y - \mathcal{P}_U y\|^2$. The singular values of $\mathcal{P}_U$ are $N-m$ zeros and m units. Thus, Corollary 7.4 implies that if y is uniformly distributed on the unit sphere of E, then $\|\mathcal{P}_U y\|^2$ has the $B\left(\frac{m}{2},\frac{N-m}{2}\right)$ distribution on $(0,1)$. In particular, if N is large, then $\mathcal{P}_U y$ lies with high probability close to the sphere of radius $\sqrt{\frac{m}{N}}$ and the squared distance $\|y - \mathcal{P}_U y\|^2$ clusters sharply around $1-\frac{m}{N}$.

Now take $E = M_n(\mathbf{R})$. With the Frobenius norm $\|\cdot\|_{\mathrm{F}}$, E is an n^2-dimensional Euclidean space. Let $U = \mathrm{Struct}_n(\mathbf{R})$ denote any class of structured matrices that form an m-dimensional linear subspace of $M_n(\mathbf{R})$. Examples include

$$\begin{aligned}
&\text{the Toeplitz matrices,} \quad \mathrm{Toep}_n(\mathbf{R})\\
&\text{the Hankel matrices,} \quad \mathrm{Hank}_n(\mathbf{R})\\
&\text{the tridiagonal matrices,} \quad \mathrm{Tridiag}_n(\mathbf{R})\\
&\text{the tridiagonal Toeplitz matrices,} \quad \mathrm{TridiagToep}_n(\mathbf{R})\\
&\text{the symmetric matrices,} \quad \mathrm{Symm}_n(\mathbf{R})\\
&\text{the lower-triangular matrices,} \quad \mathrm{Lowtriang}_n(\mathbf{R})\\
&\text{the matrices with zero main diagonal,} \quad \mathrm{Zerodiag}_n(\mathbf{R})\\
&\text{the matrices with zero trace,} \quad \mathrm{Zerotrace}_n(\mathbf{R}).
\end{aligned}$$

The dimensions of these linear spaces are

$$\begin{aligned}
&\dim \mathrm{Toep}_n(\mathbf{R}) = 2n-1, \quad \dim \mathrm{Hank}_n(\mathbf{R}) = 2n-1,\\
&\dim \mathrm{Tridiag}_n(\mathbf{R}) = 3n-2, \quad \dim \mathrm{TridiagToep}_n(\mathbf{R}) = 3,\\
&\dim \mathrm{Symm}_n(\mathbf{R}) = \frac{n^2+n}{2}, \quad \dim \mathrm{Lowtriang}_n(\mathbf{R}) = \frac{n^2+n}{2},\\
&\dim \mathrm{Zerodiag}_n(\mathbf{R}) = n^2-n, \quad \dim \mathrm{Zerotrace}_n(\mathbf{R}) = n^2-1.
\end{aligned}$$

Suppose n is large and $Y_n \in M_n(\mathbf{R})$ is uniformly distributed on the unit sphere on $M_n(\mathbf{R})$, $\|Y_n\|_{\mathrm{F}}^2 = 1$. Let $\mathcal{P}_{\mathrm{Struct}} Y_n$ be the best approximation of Y_n by a matrix in $\mathrm{Struct}_n(\mathbf{R})$. Notice that the determination of $\mathcal{P}_{\mathrm{Struct}} Y_n$ is a least squares problem that can be easily solved. For instance, $\mathcal{P}_{\mathrm{Toep}} Y_n$ is the Toeplitz matrix whose kth diagonal, $k = -(n-1), \ldots, n-1$, is formed by the arithmetic mean of the numbers in the kth diagonal of Y_n. Recall that $\dim \mathrm{Struct}_n(\mathbf{R}) = m$. From what was said in the preceding paragraph, we conclude that $\|\mathcal{P}_{\mathrm{Struct}} Y_n\|_{\mathrm{F}}^2$ is $B\left(\frac{m}{2}, \frac{n^2-m}{2}\right)$ distributed on $(0,1)$. For example, $\|\mathcal{P}_{\mathrm{Toep}} Y_n\|^2$ has the $B\left(\frac{2n-1}{2}, \frac{n^2-2n+1}{2}\right)$ distribution on $(0,1)$. The expected value of the variable $\|Y_n - \mathcal{P}_{\mathrm{Toep}} Y_n\|^2$ is $1-\frac{2}{n}+\frac{1}{n^2}$ and the variance does not exceed $\frac{4}{n^3}$. Hence, Chebyshev's inequality gives

$$P\left(1-\frac{2}{n}+\frac{1}{n^2}-\frac{\varepsilon}{n} < \|Y_n - \mathcal{P}_{\mathrm{Toep}} Y_n\|^2 < 1-\frac{2}{n}+\frac{1}{n^2}+\frac{\varepsilon}{n}\right) \geq 1-\frac{4}{n\varepsilon^2}. \tag{32}$$

Consequently, $\mathcal{P}_{\mathrm{Toep}} Y_n$ is with high probability found near the sphere with the radius $\sqrt{\frac{2}{n}-\frac{1}{n^2}}$ and $\|Y_n - \mathcal{P}_{\mathrm{Toep}} Y_n\|_{\mathrm{F}}^2$ is tightly concentrated around $1-\frac{2}{n}+\frac{1}{n^2}$.

We arrive at the conclusion that nearly all $n \times n$ matrices of Frobenius norm 1 are at nearly the same distance to the set of all $n \times n$ Toeplitz matrices!

This does not imply that the Toeplitz matrices are at the center of the universe. In fact, the conclusion is true for each of the classes $\mathrm{Struc}_n(\mathbf{R})$ listed above. For instance, from Chebyshev's inequality we obtain

$$P\left(\frac{1}{2}-\frac{1}{2n}-\varepsilon < \|Y_n - \mathcal{P}_{\mathrm{Symm}} Y_n\|^2 < \frac{1}{2}-\frac{1}{2n}+\varepsilon\right) \geq 1-\frac{1}{2n^2\varepsilon^2} \tag{33}$$

and

$$P\left(\frac{1}{n^2}-\frac{\varepsilon}{n^2}<\|Y_n-\mathcal{P}_{\mathrm{Zerotrace}}Y_n\|^2<\frac{1}{n^2}+\frac{\varepsilon}{n^2}\right)\geq 1-\frac{2}{n^2\varepsilon^2}.$$

If the expected value of $\|Y_n-\mathcal{P}_{\mathrm{Struct}}Y_n\|^2$ stays away from 0 and 1 as $n\to\infty$, we have much sharper estimates. Namely, Lemma 2.2 of [5] in conjunction with Corollary 7.4 implies that if X_n^2 has the $B\big(\frac{m}{2},\frac{N-m}{2}\big)$ distribution on $(0,1)$, then

$$P\left(X_n^2\leq\sigma\,\frac{m}{N}\right)\leq\left(\sigma e^{1-\sigma}\right)^{m/2},\quad P\left(X_n^2\geq\tau\,\frac{m}{N}\right)\leq\left(\tau e^{1-\tau}\right)^{m/2}\tag{34}$$

for $0<\sigma<1<\tau$. This yields, for example,

$$\begin{aligned}&P\left(\sigma\left(\frac{1}{2}-\frac{1}{2n}\right)<\|Y_n-\mathcal{P}_{\mathrm{Symm}}Y_n\|_{\mathrm{F}}^2<\tau\left(\frac{1}{2}-\frac{1}{2n}\right)\right)\\&\qquad\geq 1-\left(\sigma e^{1-\sigma}\right)^{(n^2+n)/4}-\left(\tau e^{1-\tau}\right)^{(n^2+n)/4}\end{aligned}\tag{35}$$

whenever $0<\sigma<1<\tau$. Clearly, (35) is better than (33). On the other hand, let $\varepsilon>0$ be small and choose τ such that

$$\tau\left(1-\frac{2}{n}+\frac{1}{n^2}\right)=1-\frac{2}{n}+\frac{1}{n^2}+\frac{\varepsilon}{n}.$$

Then

$$\left(\tau e^{1-\tau}\right)^{n-1/2}=1-\frac{\varepsilon^2}{2n}+O\left(\frac{1}{n^2}\right),$$

the O depending on ε, and hence (34) amounts to

$$P\left(\|Y_n-\mathcal{P}_{\mathrm{Toep}}Y_n\|^2\geq 1-\frac{2}{n}+\frac{1}{n^2}+\frac{\varepsilon}{n}\right)\leq 1-\frac{\varepsilon^2}{2n}+O\left(\frac{1}{n^2}\right),$$

which is worse than the Chebyshev estimate (32). ■

Here is another case in which Theorem 7.3 can be made more explicit. The Gaussian hypergeometric function $F(a,b,c;z)$ is defined by

$$F(a,b,c;z)=1+\sum_{k=1}^{\infty}\frac{(a)_k(b)_k}{(c)_k}\,\frac{z^k}{k!},$$

where $(y)_k=y(y+1)\ldots(y+k-1)$.

Corollary 7.7 *Let $n\geq 3$ and $0=s_1=\ldots=s_{n-2}<s_{n-1}<s_n$. Put $\mu=s_n/s_{n-1}$. Then for $\xi\in(0,1/\mu^2)$ the density of the random variable X_n^2 is*

$$\begin{aligned}f_n(\xi)&=\frac{\mu}{\pi}\left(\frac{n}{2}-1\right)(1-\xi)^{(n-4)/2}\int_0^1\left(1-\frac{\xi(\mu^2-1)}{1-\xi}\,x\right)^{(n-4)/2}\frac{dx}{\sqrt{x(1-x)}}&(36)\\&=\mu\left(\frac{n}{2}-1\right)(1-\xi)^{(n-4)/2}F\left(\frac{1}{2},\frac{4-n}{2},1;\frac{\xi(\mu^2-1)}{1-\xi}\right).&(37)\end{aligned}$$

Proof. We have $v(\varphi) = \mu^2 \cos^2 \varphi + \sin^2 \varphi$ and hence (28) yields

$$\begin{aligned}
&\int_0^{\pi/2} (1 - \xi v(\varphi))^{(n-4)/2} d\varphi \\
&= (1-\xi)^{(n-4)/2} \int_0^{\pi/2} \left(1 - \frac{\xi(\mu^2-1)}{1-\xi} \cos^2 \varphi\right)^{(n-4)/2} d\varphi \\
&= \frac{(1-\xi)^{(n-4)/2}}{2} \int_0^1 \left(1 - \frac{\xi(\mu^2-1)}{1-\xi}\, x\right)^{(n-4)/2} \frac{dx}{\sqrt{x(1-x)}}. \qquad (38)
\end{aligned}$$

Combining (38) and Theorem 7.3 we arrive at (36). Formula 2.2.6.1 of [12] gives that (38) equals

$$\frac{(1-\xi)^{(n-4)/2}}{2} B\left(\frac{1}{2}, \frac{1}{2}\right) F\left(\frac{1}{2}, \frac{4-n}{2}, 1; \frac{\xi(\mu^2-1)}{1-\xi}\right).$$

This in conjunction with Theorem 7.3 proves (37). ■

For $y \in (0,1)$, the complete elliptic integrals $\mathsf{K}(y)$ and $\mathsf{E}(y)$ are defined by

$$\mathsf{K}(y) = \int_0^{\pi/2} \frac{d\varphi}{\sqrt{1 - y^2 \sin^2 \varphi}}, \qquad \mathsf{E}(y) = \int_0^{\pi/2} \sqrt{1 - y^2 \sin^2 \varphi}\, d\varphi.$$

For small n's, Corollary 7.7 delivers the following densities on $(0, 1/\mu^2)$:

$$\begin{aligned}
f_3(\xi) &= \frac{\mu}{\pi\sqrt{1-\xi}} \mathsf{K}\left(\sqrt{\frac{\xi(\mu^2-1)}{1-\xi}}\right), \\
f_4(\xi) &= \mu \quad \text{(uniform distribution)}, \\
f_5(\xi) &= \frac{3\mu\sqrt{1-\xi}}{\pi} \mathsf{E}\left(\sqrt{\frac{\xi(\mu^2-1)}{1-\xi}}\right), \\
f_6(\xi) &= 2\mu - \mu(\mu^2+1)\xi,
\end{aligned}$$

and $f_7(\xi)$ equals

$$\frac{5\mu}{3\pi} \sqrt{1-\xi} \left((4 - 2\xi - 2\xi\mu^2)\, \mathsf{E}\left(\sqrt{\frac{\xi(\mu^2-1)}{1-\xi}}\right) - (1 - \xi\mu^2)\, \mathsf{K}\left(\sqrt{\frac{\xi(\mu^2-1)}{1-\xi}}\right)\right).$$

Finally, under the hypothesis of Corollary 7.7 the distribution of nX_n^2 is no longer χ^2 in the limit $n \to \infty$. Indeed, by (36), the density of nX_n^2 is

$$\frac{1}{n} f_n\left(\frac{\xi}{n}\right) = \frac{\mu}{\pi n}\left(\frac{n}{2} - 1\right)\left(1 - \frac{\xi}{n}\right)^{(n-4)/2} \int_0^1 \left(1 - \frac{\xi(\mu^2-1)}{n-\xi}\, x\right)^{(n-4)/2} \frac{dx}{\sqrt{x(1-x)}}$$

and as $n \to \infty$, this converges to

$$\frac{\mu}{2\pi} e^{-\xi/2} \int_0^1 e^{-\xi x(\mu^2-1)/2} \frac{dx}{\sqrt{x(1-x)}}.$$

Formula 2.3.6.2 of [12] tells us that the last expression equals

$$\frac{\mu}{2\pi}\, e^{-\xi/2}\, \pi\, e^{-\xi(\mu^2-1)/4} I_0\left(\frac{\xi(\mu^2-1)}{4}\right) = \frac{\mu}{2}\, e^{-\xi(\mu^2+1)/4} I_0\left(\frac{\xi(\mu^2-1)}{4}\right),$$

where I_0 is the modified Bessel function,

$$I_0(y) := 1 + \sum_{k=1}^{\infty} \left(\frac{1}{k!}\right)^2 \left(\frac{y}{2}\right)^{2k}.$$

Conclusion. It is clear that estimates based on knowledge of the distribution function are in general better than estimates that are obtained from Chebyshev's inequality. Examples 2.3 and 7.5 convincingly demonstrate the superiority of the distribution function over Chebyshev estimates. However, the message of this paper is that, for large n, the ratio $\|A_n x\|^2/(\|A_n\|^2\|x\|^2)$ clusters sharply around a certain number and that this number can be completely identified for important classes of structured matrices. The zoo of distribution functions we encountered above makes us appreciate the beauty of the simple and general Theorem 2.2 and the ease with which we were able to deduce the results of Sections 3 to 6 from this theorem.

Acknowledgment. We thank Florin Avram, Nick Trefethen, and Man-Chung Yeung for encouraging remarks on a baby version of this paper. We are also greatly indebted to David Aldous, Ted Cox, and the referee for giving our attempt at probability a warm reception.

References

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

[2] A. Böttcher and S. Grudsky: *Toeplitz Matrices, Asymptotic Linear Algebra, and Functional Analysis.* Hindustan Boook Agency, New Delhi 2000 and Birkhäuser Verlag, Basel 2000.

[3] A. Böttcher and B. Silbermann: *Analysis of Toeplitz Operators.* Akademie-Verlag, Berlin 1989 and Springer-Verlag, Berlin 1990.

[4] A. Böttcher and B. Silbermann: *Introduction to Large Truncated Toeplitz Matrices.* Springer-Verlag, New York 1999.

[5] S. Dasgupta and A. Gupta: An elementary proof of a theorem of Johnson and Lindenstrauss. *Random Structures & Algorithms* **22** (2003), 60–65.

[6] D. Fasino and P. Tilli: Spectral clustering properties of block multilevel Hankel matrices. *Linear Algebra Appl.* **306** (2000), 155–163.

[7] G. M. Fichtenholz: *Differential- und Integralrechnung, Vol. III.* Deutscher Verlag der Wissenschaften, Berlin 1977.

[8] P. Halmos: *A Hilbert Space Problem Book.* D. van Nostrand, Princeton 1967.

[9] N. J. Higham: *Accuracy and Stability of Numerical Algorithms.* SIAM, Philadelphia, PA 1996.

[10] M. Kac, W. L. Murdock, and G. Szegö: On the eigenvalues of certain Hermitian forms. *J. Rat. Mech. Anal.* **2** (1953), 787–800.

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

[12] A. P. Prudnikov, Yu. A. Brychkov, and O. I. Marichev: *Integrals and Series. Vol. 1. Elementary Functions.* Gordon and Breach, New York 1986.

[13] S. Roch and B. Silbermann: Index calculus for approximation methods and singular value decomposition. *J. Math. Analysis Appl.* **225** (1998), 401–426.

[14] S. M. Rump: Structured perturbations. Part I: Normwise distances. Preprint, April 2002.

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

[16] O. Toeplitz: Zur Theorie der quadratischen und bilinearen Formen von unendlichvielen Veränderlichen. *Math. Ann.* **70** (1911), 351–376.

[17] W. F. Trench: Asymptotic distribution of the spectra of a class of generalized Kac-Murdock-Szegö matrices. *Linear Algebra Appl.* **294** (1999), 181–192.

[18] E. E. Tyrtyshnikov and N. L. Zamarashkin: Toeplitz eigenvalues for Radon measures. *Linear Algebra Appl.* **343/344** (2002), 345–354.