A PROBABILITY ARGUMENT IN FAVOR OF IGNORING SMALL SINGULAR VALUES

ALBRECHT BÖTTCHER*, DANIEL POTTS†, AND DAVID WENZEL‡

Abstract. If the matrix of a square linear system is nonsingular but has very small singular values, then tiny perturbations of the right-hand side may cause drastic changes in the solution. We show that the probability for this to happen is very close to zero if sufficiently many singular values of the matrix are bounded away from zero.

Key words. condition number, probability argument, linear system, singular value

AMS subject classifications. 65F35, 15A12, 60H25, 65F22

1. Introduction. Let A be a nonzero real or complex $n \times n$ matrix, $A \in M_n(\mathbb{K})$ where $\mathbb{K} = \mathbb{R}$ of $\mathbb{K} = \mathbb{C}$. Pick $p \in \mathbb{K}^n$ and put $Ap = y$. Suppose $\widetilde{p} \in \mathbb{K}^n$ is a perturbation to p and $A\widetilde{p} = \widetilde{y}$. We denote by $\|\cdot\|$ the ℓ^2 norm on $\mathbb{K}^n$. A basic question is whether $\|\delta p\| := \|\widetilde{p}\,\|/\|p\|$ may be large if $\|\delta y\| := \|\widetilde{y}\|/\|y\|$ is small. To tackle this question, let $A = USV$ be the singular value decomposition of A. We assume that $S = \operatorname{diag}(s_1, \ldots, s_n)$ with $0 \leq s_1 \leq \ldots \leq s_n$. The number $s_n/s_1 \in (0, \infty]$ is called the (spectral) condition number of A and it is well known that

$$\|\delta p\| \leq \frac{s_n}{s_1}\|\delta y\|. \tag{1.1}$$

There exist p and $\widetilde{p}$ such that in (1.1) equality holds. Thus, if s_1 is very small, then the system $Ap = y$ is ill-conditioned in the sense that $\|\delta p\|/\|\delta y\|$ may become very large. It is also well known that in practice theoretically ill-conditioned systems often behave better than one would expect. The purpose of this paper is to provide a probabilistic argument that reveals that equality in (1.1) is a rare event if the matrix dimension n is at least moderately large.

To be more precise, fix $p \in \mathbb{K}^n$ and take $\widetilde{p}$ randomly from the ball

$$\widetilde{\varrho}\,\mathbb{B}^n_{\mathbb{K}} := \{z \in \mathbb{K}^n : |z_1|^2 + \ldots + |z_n|^2 \leq \widetilde{\varrho}^{\,2}\}$$

with the uniform distribution. We show that if $\mathbb{K} = \mathbb{C}$, then

$$\mathbb{P}\left(\frac{\|\delta p\|}{\|\delta y\|} \leq 4\,\frac{s_n}{s_{[n/2]+1}}\right) \geq 1 - \frac{1}{\sqrt{n}}\,\left(\frac{1}{2}\right)^n \tag{1.2}$$

for all $n \geq 2$, where $\mathbb{P}(E)$ denotes the probability of the event E. Thus, if half of the singular values of A are separated away from zero, then $\|\delta p\|/\|\delta y\|$ does not exceed a reasonable bound with high probability.

Here are two examples. First, let $A \in M_{60}(\mathbb{C})$ be the Toeplitz matrix generated by the function χ which equals 1 on $[0, 2/3)$ and 0 on $[2/3, 1)$ (see the left of Figure 1.1). The singular values (in decreasing order) are shown in the right of Figure 1.1. The

*Fakultät für Mathematik, Technische Universität Chemnitz, D - 09107 Chemnitz, Germany (aboettch@mathematik.tu-chemnitz.de)

†Fakultät für Mathematik, Technische Universität Chemnitz, D - 09107 Chemnitz, Germany (potts@mathematik.tu-chemnitz.de)

‡Fakultät für Mathematik, Technische Universität Chemnitz, D - 09107 Chemnitz, Germany (david.wenzel@s2000.tu-chemnitz.de)

smallest singular value s_1 is strictly positive, but according to Matlab the quotient s_n/s_1 equals $3.79 \cdot 10^{16}$. On the other hand, the right picture of Figure 1.1 tells us that $s_{[n/2]+1}$ is approximately equal to s_n and hence (1.2) implies that

$$\mathbb{P}\left(\frac{\|\delta p\|}{\|\delta y\|} \le 4\right) \ge 1 - \frac{1}{\sqrt{60}}\left(\frac{1}{2}\right)^{60}.$$

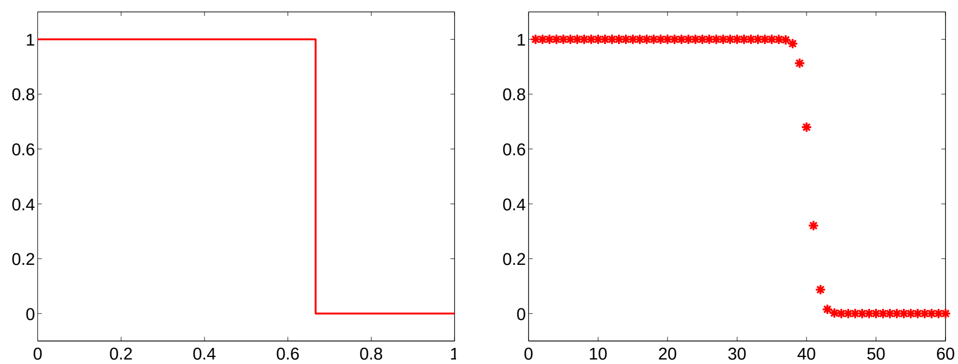

FIG. 1.1. *The graph of a function and the singular values of the* 60×60 *Toeplitz matrix generated by this function.*

In the second example we consider a matrix $A \in M_{121}(\mathbb{C})$ that arises in a sampling problem on a linogram grid; see, e.g., [6], [9]. The linogram grid and the singular values are plotted in Figure 1.2. Matlab gives that s_n/s_1 is about $1.97 \cdot 10^{17}$, but Figure 1.2 reveals that $s_n/s_{[n/2]+1}$ is approximately 5.2, so that (1.2) gives

$$\mathbb{P}\left(\frac{\|\delta p\|}{\|\delta y\|} \le 20.8\right) \ge 1 - \frac{1}{11}\left(\frac{1}{2}\right)^{121}.$$

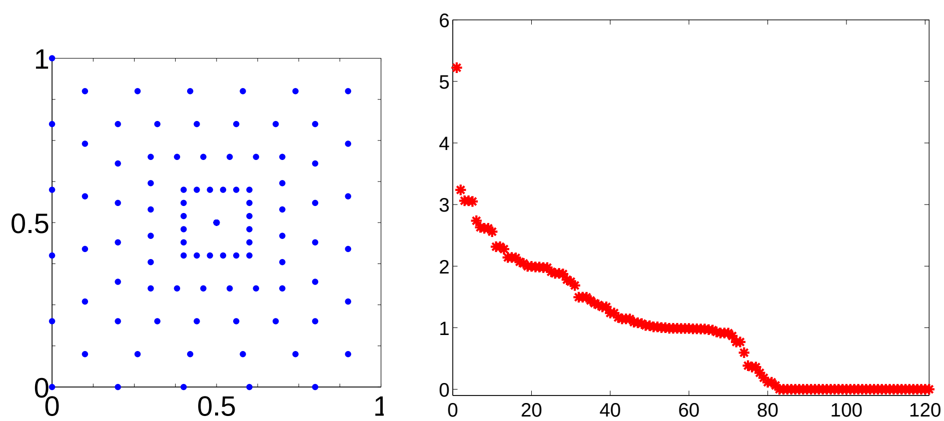

FIG. 1.2. *The linogram grid generated by the parameters* $N = 5$, $T = 10$, $R = 10$ *and the singular values of the* 121×121 *matrix that arises in the sampling of a bivariate trigonometric polynomial of degree* 5 *on the linogram grid.*

To avoid misunderstandings, we emphasize that we prove that $\|\delta p\|/\|\delta y\|$ is with high probability not a large number if $s_n/s_{[n/2]+1}$ is, say, not larger than 10 or 50 and the perturbation $\widetilde{p}$ is drawn from a ball centered at the origin with the uniform distribution. This does not exclude that in practical computations a small value of s_1 may nevertheless cause the known problems.

2. The general result. We begin with a simple estimate. Recall that $A = USV$ is the singular value decomposition.

PROPOSITION 2.1. *If* $s_{k+1} > 0$, *then*

$$\frac{\|\delta p\|^2}{\|\delta y\|^2} \le \frac{s_n^2}{s_{k+1}^2}\,\frac{1}{1-(|(V\widetilde{p})_1|^2+\ldots+|(V\widetilde{p})_k|^2)/\|V\widetilde{p}\|^2}.$$

Proof. We have

$$\begin{aligned}\|\delta y\|^2 &= \frac{\|\widetilde{y}\|^2}{\|y\|^2} = \frac{\|U^*\widetilde{y}\|^2}{\|U^*y\|^2} = \frac{\|SV\widetilde{p}\|^2}{\|SVp\|^2}\\ &= \frac{s_1^2|(V\widetilde{p})_1|^2+\ldots+s_n^2|(V\widetilde{p})_n|^2}{s_1^2|(Vp)_1|^2+\ldots+s_n^2|(Vp)_n|^2} \ge \frac{s_{k+1}^2(|(V\widetilde{p})_{k+1}|^2+\ldots+|(V\widetilde{p})_n|^2)}{s_n^2(|(Vp)_1|^2+\ldots+|(Vp)_n|^2)}\\ &= \frac{s_{k+1}^2}{s_n^2}\,\frac{\|V\widetilde{p}\|^2-|(V\widetilde{p})_1|^2-\ldots-|(V\widetilde{p})_k|^2}{\|Vp\|^2}\end{aligned}$$

and hence

$$\frac{\|\delta p\|^2}{\|\delta y\|^2} = \frac{\|\widetilde{p}\|^2}{\|p\|^2}\,\frac{\|y\|^2}{\|\widetilde{y}\|^2} \le \frac{s_n^2}{s_{k+1}^2}\,\frac{\|Vp\|^2}{\|p\|^2}\,\frac{\|\widetilde{p}\|^2}{\|V\widetilde{p}\|^2-|(V\widetilde{p})_1|^2-\ldots-|(V\widetilde{p})_k|^2},$$

which gives the assertion because $\|Vp\| = \|p\|$ and $\|\widetilde{p}\| = \|V\widetilde{p}\|$. $\square$

Now suppose the perturbation $\widetilde{p}$ is randomly taken from the ball $\widetilde{\varrho}\mathbb{B}_{\mathbb{K}}^n$ with the uniform distribution. Then $V\widetilde{p}$ is uniformly distributed on $\widetilde{\varrho}\mathbb{B}_{\mathbb{K}}^n$ and the random vector $q := V\widetilde{p}/\|V\widetilde{p}\|$ is uniformly distributed on

$$\mathbb{S}_{\mathbb{K}}^{n-1} := \{z \in \mathbb{K}^n : |z_1|^2+\ldots+|z_n|^2 = 1\}.$$

In terms of q, the estimate of Proposition 2.1 reads

$$\frac{\|\delta p\|^2}{\|\delta y\|^2} \le \frac{s_n^2}{s_{k+1}^2}\,\frac{1}{1-(|q_1|^2+\ldots+|q_k|^2)}.$$

It results that, for every $\varepsilon \in (0,1)$,

$$\mathbb{P}\left(\frac{\|\delta p\|^2}{\|\delta y\|^2} \le \frac{1}{\varepsilon^2}\,\frac{s_n^2}{s_{k+1}^2}\right) \ge \mathbb{P}\Big(|q_1|^2+\ldots+|q_k|^2 \le 1-\varepsilon^2\Big). \tag{2.1}$$

We subsequently make use of the formula $|\mathbb{S}_{\mathbb{R}}^{m-1}| = 2\pi^{m/2}/\Gamma(m/2)$.

The following result is undoubtedly known to probabilists (see [4] for $K = 1$ and see also [5, pp. 300] for a related discussion). As we have not been able to find the result as it is stated in the literature, we cite it with a full proof.

THEOREM 2.2. *If* $(q_1,\ldots,q_N)$ *is taken from* $\mathbb{S}_{\mathbb{R}}^{N-1}$ *with the uniform distribution, then for* $1 \le K \le N-1$ *the density function of the random vector* $(q_1,\ldots,q_K)$ *is*

$$\frac{|\mathbb{S}_{\mathbb{R}}^{N-K-1}|}{|\mathbb{S}_{\mathbb{R}}^{N-1}|}\,(1-x_1^2-\ldots-x_K^2)^{(N-K-2)/2}.$$

Proof. For a measurable set $\Omega \subset \mathbb{R}^K$, we define

$$\Omega^* = \{(x_1, \dots, x_N) \in \mathbb{S}_{\mathbb{R}}^{N-1} : (x_1, \dots, x_K) \in \Omega\}.$$

We then have

$$\mathbb{P}\Big((q_1, \dots, q_K) \in \Omega\Big) = \mathbb{P}\Big((q_1, \dots, q_N) \in \Omega^*\Big) = \frac{1}{|\mathbb{S}_{\mathbb{R}}^{N-1}|} \int_{\Omega^*} d\sigma,$$

where $d\sigma$ is the surface measure on $\mathbb{S}_{\mathbb{R}}^{N-1}$. Put

$$\Omega' = \Big\{(x_1, \dots, x_{N-1}) \in \mathbb{B}_{\mathbb{R}}^{N-1} : (x_1, \dots, x_{N-1}, \sqrt{1 - x_1^2 - \dots - x_{N-1}^2}) \in \Omega^*\Big\}.$$

The set Ω^* is composed of two congruent pieces given by

$$x_N = \pm\sqrt{1 - x_1^2 - \dots - x_{N-1}^2}, \quad (x_1, \dots, x_{N-1}) \in \Omega'.$$

Consequently,

$$\frac{1}{|\mathbb{S}_{\mathbb{R}}^{N-1}|} \int_{\Omega^*} d\sigma = \frac{2}{|\mathbb{S}_{\mathbb{R}}^{N-1}|} \int_{\Omega'} \sqrt{1 + \left(\frac{\partial x_N}{\partial x_1}\right)^2 + \dots + \left(\frac{\partial x_N}{\partial x_{N-1}}\right)^2}\, dx_1 \dots dx_{N-1}$$
$$= \frac{2}{|\mathbb{S}_{\mathbb{R}}^{N-1}|} \int_{\Omega'} \frac{dx_1 \dots dx_{N-1}}{\sqrt{1 - x_1^2 - \dots - x_{N-1}^2}}$$
$$= \frac{2}{|\mathbb{S}_{\mathbb{R}}^{N-1}|} \int_{\Omega} \left(\int_{\varrho \mathbb{B}_{\mathbb{R}}^{N-1-K}} \frac{dx_{K+1} \dots dx_{N-1}}{\sqrt{1 - x_1^2 - \dots - x_{N-1}^2}}\right) dx_1 \dots dx_K$$

where $\varrho := \sqrt{1 - x_1^2 - \dots - x_K^2}$. It follows that the density function is

$$\frac{2}{|\mathbb{S}_{\mathbb{R}}^{N-1}|} \int_{\varrho \mathbb{B}_{\mathbb{R}}^{N-1-K}} \frac{dx_{K+1} \dots dx_{N-1}}{\sqrt{1 - x_1^2 - \dots - x_{N-1}^2}},$$

which can also be written as

$$\frac{2}{|\mathbb{S}_{\mathbb{R}}^{N-1}|} \int_{\varrho \mathbb{B}_{\mathbb{R}}^{N-1-K}} \frac{dx_{K+1} \dots dx_{N-1}}{\sqrt{\varrho^2 - x_{K+1}^2 - \dots - x_{N-1}^2}}. \tag{2.2}$$

To compute (2.2), we use the well known formula

$$\int_{R\mathbb{B}_{\mathbb{R}}^m} f\Big(\sqrt{x_1^2 + \dots + x_m^2}\Big) dx_1 \dots dx_m = \frac{2\,\pi^{m/2}}{\Gamma\left(\frac{m}{2}\right)} \int_0^R r^{m-1} f(r)\, dr, \tag{2.3}$$

for which see, e.g., [8, pp. 396–397]. This formula shows that (2.2) equals

$$
\begin{aligned}
&\frac{2}{|\mathbb{S}_{\mathbb{R}}^{N-1}|}\,\frac{2\,\pi^{(N-1-K)/2}}{\Gamma\left(\frac{N-1-K}{2}\right)}\int_0^{\varrho} r^{N-K-2}\,\frac{dr}{\sqrt{\varrho^2-r^2}}\\
&=\frac{2}{|\mathbb{S}_{\mathbb{R}}^{N-1}|}\,\frac{2\,\pi^{(N-1-K)/2}}{\Gamma\left(\frac{N-1-K}{2}\right)}\,\frac{\Gamma\left(\frac{N-1-K}{2}\right)\Gamma\left(\frac{1}{2}\right)}{2\,\Gamma\left(\frac{N-K}{2}\right)}\,\varrho^{N-K-2}\\
&=\frac{1}{|\mathbb{S}_{\mathbb{R}}^{N-1}|}\,\frac{2\,\pi^{(N-K)/2}}{\Gamma\left(\frac{N-K}{2}\right)}\,\varrho^{N-K-2}\\
&=\frac{1}{|\mathbb{S}_{\mathbb{R}}^{N-1}|}\,|\mathbb{S}_{\mathbb{R}}^{N-K-1}|\,(1-x_1^2-\ldots-x_K^2)^{(N-K-2)/2}. \qquad \square
\end{aligned}
$$

THEOREM 2.3. *Let $(q_1,\ldots,q_K)$ be as in the previous theorem. If $\varepsilon\in(0,1)$, then*

$$
\mathbb{P}\Big(q_1^2+\ldots+q_K^2\le 1-\varepsilon^2\Big)\ge 1-\gamma\,\frac{\Gamma\left(\frac{N}{2}\right)}{\Gamma\left(\frac{N-K}{2}\right)\Gamma\left(\frac{K}{2}\right)}\,\frac{2}{N-K}\,\varepsilon^{N-K},
$$

where $\gamma=1/\sqrt{1-\varepsilon^2}$ for $K=1$ and $\gamma=1$ for $2\le K\le N-1$.

Proof. By Theorem 2.2, the probability in question is

$$
\begin{aligned}
&\frac{|\mathbb{S}_{\mathbb{R}}^{N-K-1}|}{|\mathbb{S}_{\mathbb{R}}^{N-1}|}\int_{x_1^2+\ldots+x_K^2\le 1-\varepsilon^2}(1-x_1^2-\ldots-x_K^2)^{(N-K-2)/2}\,dx_1\ldots dx_K,\\
&=\frac{\pi^{-K/2}\Gamma\left(\frac{N}{2}\right)}{\Gamma\left(\frac{N-K}{2}\right)}\int_{x_1^2+\ldots+x_K^2\le 1-\varepsilon^2}(1-x_1^2-\ldots-x_K^2)^{(N-K-2)/2}\,dx_1\ldots dx_K. \qquad (2.4)
\end{aligned}
$$

Formula (2.3) implies that (2.4) is equal to

$$
\frac{2\,\Gamma\left(\frac{N}{2}\right)}{\Gamma\left(\frac{N-K}{2}\right)\Gamma\left(\frac{K}{2}\right)}\int_0^{1-\eta} r^{K-1}(1-r^2)^{(N-K-2)/2}dr,
$$

where $\eta\in(0,1)$ is defined by $(1-\eta)^2=1-\varepsilon^2$. Taking into account that

$$
\frac{2\,\Gamma\left(\frac{N}{2}\right)}{\Gamma\left(\frac{N-K}{2}\right)\Gamma\left(\frac{K}{2}\right)}\int_0^{1} r^{K-1}(1-r^2)^{(N-K-2)/2}dr=1,
$$

we obtain that

$$
\begin{aligned}
&\mathbb{P}\Big(q_1^2+\ldots+q_K^2>1-\varepsilon^2\Big)=1-\mathbb{P}\Big(q_1^2+\ldots+q_K^2\le 1-\varepsilon^2\Big)\\
&=2\,\frac{\Gamma\left(\frac{N}{2}\right)}{\Gamma\left(\frac{N-K}{2}\right)\Gamma\left(\frac{K}{2}\right)}\int_{1-\eta}^{1} r^{K-1}(1-r^2)^{(N-K-2)/2}dr. \qquad (2.5)
\end{aligned}
$$

The change of variables $r^2=x$ yields that the integral in (2.5) is

$$
\frac{1}{2}\int_{(1-\eta)^2}^{1} x^{(K-2)/2}(1-x)^{(N-K-2)/2}dx,
$$

which does not exceed

$$
\frac{1}{1-\eta}\int_{(1-\eta)^2}^{1}(1-x)^{(N-K-2)/2}dx
$$

for $K = 1$ and

$$\int_{(1-\eta)^2}^{1} (1-x)^{(N-K-2)/2} dx$$

for $K \geq 2$. Since

$$\int_{(1-\eta)^2}^{1} (1-x)^{(N-K-2)/2} dx = \frac{(1-(1-\eta)^2)^{(N-K)/2}}{(N-K)/2} = \frac{2}{N-K}\, \varepsilon^{N-K},$$

we arrive at the assertion. □

Corollary 2.4. *Let $\varepsilon \in (0,1)$ and suppose $s_{k+1} > 0$. Take $\widetilde{p}$ randomly from the ball $\widetilde{\varrho}\mathbb{B}^n_{\mathbb{K}}$ with the uniform distribution. If $\mathbb{K} = \mathbb{R}$, then*

$$\mathbb{P}\left(\frac{\|\delta p\|^2}{\|\delta y\|^2} \leq \frac{1}{\varepsilon^2}\,\frac{s_n^2}{s_{k+1}^2}\right) \geq 1 - \gamma\, \frac{\Gamma\left(\frac{n}{2}\right)}{\Gamma\left(\frac{n-k}{2}\right)\Gamma\left(\frac{k}{2}\right)}\, \frac{2}{n-k}\, \varepsilon^{n-k},$$

where $\gamma = 1/\sqrt{1-\varepsilon^2}$ for $k = 1$ and $\gamma = 1$ for $2 \leq k \leq n-1$. If $\mathbb{K} = \mathbb{C}$, then

$$\mathbb{P}\left(\frac{\|\delta p\|^2}{\|\delta y\|^2} \leq \frac{1}{\varepsilon^2}\,\frac{s_n^2}{s_{k+1}^2}\right) \geq 1 - \frac{\Gamma(n)}{\Gamma(n-k)\Gamma(k)}\, \frac{\varepsilon^{2(n-k)}}{n-k}$$

for $1 \leq k \leq n-1$.

Proof. This is immediate from (2.1) and Theorem 2.3 with $N = n$ and $K = k$ for $\mathbb{K} = \mathbb{R}$ and with $N = 2n$ and $K = 2k$ for $\mathbb{K} = \mathbb{C}$. □

Example 2.5. Let $\mathbb{K} = \mathbb{C}$ and take $k = [n/2]$, where $[\,\cdot\,]$ stands for the integral part. Stirling's formula tells that

$$\frac{\Gamma(n)}{\Gamma\left(\frac{n}{2}\right)^2} \sim 2^{n-1}\sqrt{\frac{n}{2\pi}},$$

whence

$$\frac{\Gamma(n)}{\Gamma\left(n-\left[\frac{n}{2}\right]\right)\Gamma\left(\left[\frac{n}{2}\right]\right)}\, \frac{\varepsilon^{2(n-[n/2])}}{n-\left[\frac{n}{2}\right]} \leq \frac{2^n}{\sqrt{n}}\, \varepsilon^n \tag{2.6}$$

for all $n \geq n_0$. A careful analysis shows that (2.6) is in fact true for all $n \geq 2$. From Corollary 2.4 we therefore deduce that

$$\mathbb{P}\left(\frac{\|\delta p\|^2}{\|\delta y\|^2} \leq \frac{1}{\varepsilon^2}\,\frac{s_n^2}{s_{[n/2]+1}^2}\right) \geq 1 - \frac{2^n}{\sqrt{n}}\, \varepsilon^n$$

for all $n \geq 2$. Choosing $\varepsilon = 1/2$ and $\varepsilon = 1/4$ we obtain in particular that

$$\begin{aligned} &\mathbb{P}\left(\frac{\|\delta p\|^2}{\|\delta y\|^2} \leq 4\,\frac{s_n^2}{s_{[n/2]+1}^2}\right) \geq 1 - \frac{2}{\sqrt{n}}, \\ &\mathbb{P}\left(\frac{\|\delta p\|^2}{\|\delta y\|^2} \leq 16\,\frac{s_n^2}{s_{[n/2]+1}^2}\right) \geq 1 - \frac{1}{\sqrt{n}}\left(\frac{1}{2}\right)^n. \end{aligned} \tag{2.7}$$

Notice that (2.7) is the same as (1.2). The choice $k = [\alpha n]$ with $\alpha \in (0,1)$ gives

$$\mathbb{P}\left(\frac{\|\delta p\|^2}{\|\delta y\|^2} \le \frac{1}{\varepsilon^2}\,\frac{s_n^2}{s_{[\alpha n]+1}^2}\right) \ge 1 - \sqrt{\frac{\alpha}{1-\alpha}}\,\frac{1}{\sqrt{n}}\,\left(\frac{\varepsilon^{2(1-\alpha)}}{\alpha^\alpha(1-\alpha)^{1-\alpha}}\right)^n,$$

for sufficiently large n, which for $\varepsilon^2 = 2^{-1/(1-\alpha)}(1-\alpha)\alpha^{\alpha/(1-\alpha)}$ becomes

$$\mathbb{P}\left(\frac{\|\delta p\|^2}{\|\delta y\|^2} \le \frac{2^{1/(1-\alpha)}}{(1-\alpha)\alpha^{\alpha/(1-\alpha)}}\,\frac{s_n^2}{s_{[\alpha n]+1}^2}\right) \ge 1 - \sqrt{\frac{\alpha}{1-\alpha}}\,\frac{1}{\sqrt{n}}\,\left(\frac{1}{2}\right)^n. \tag{2.8}$$

3. Toeplitz operators. Fix a function $a \in L^\infty([0,1]^d)$ and let $\{a_k\}_{k\in\mathbb{Z}^d}$ be the sequence of its Fourier coefficients,

$$a_k = \int_{[0,1]^d} a(x)e^{-2\pi i k\cdot x}dx.$$

The Mth Toeplitz operator $T_M(a)$ generated by a is the operator on $\ell^2(\{1,\dots,M\}^d)$ defined by

$$(T_M(a)p)_j = \sum_{k\in\{1,\dots,M\}^d} a_{j-k}p_k \quad (j \in \{1,\dots,M\}^d).$$

We think of $T_M(a)$ as a linear operator on $\mathbb{C}^n$ with $n = M^d$. All properties of Toeplitz operators used in the following can be found in [2] and [3].

As the case where a is identically zero is uninteresting, we assume that a is not the zero function. Then $\|a\|_\infty > 0$. The largest singular value $s_n(T_M(a))$ does not exceed $\|a\|_\infty$ and converges to $\|a\|_\infty$ as $M \to \infty$. The question whether the smallest singular value $s_1(T_M(a))$ stays away from zero as $M \to \infty$ is difficult. An answer is known if $a \in C([0,1]^d)$, but this answer provides an effectively verifiable criterion only for $d = 1$ and some particular cases if $d \ge 2$. We remark that the condition

$$\operatorname{ess\,inf}_{x\in[0,1]^d} |a(x)| > 0$$

is necessary but not sufficient for $s_1(T_M(a))$ to be bounded away from zero as $M \to \infty$. Thus, if

$$\operatorname{ess\,inf}_{x\in[0,1]^d} |a(x)| > 0, \tag{3.1}$$

then certainly $s_1(T_M(a)) \to 0$. The following result shows that independently of whether (3.1) holds or not, $\|\delta p\|/\|\delta y\|$ always remains bounded with a high probability.

Theorem 3.1. *For every $a \in L^\infty([0,1]^d) \setminus \{0\}$ there exists an $\alpha \in (0,1)$ such that $s_n^2/s_{[\alpha n]+1}^2 \le 4$ and hence*

$$\mathbb{P}\left(\frac{\|\delta p\|^2}{\|\delta y\|^2} \le \frac{4\cdot 2^{1/(1-\alpha)}}{(1-\alpha)\alpha^{\alpha/(1-\alpha)}}\right) \ge 1 - \sqrt{\frac{\alpha}{1-\alpha}}\,\frac{1}{\sqrt{n}}\,\left(\frac{1}{2}\right)^n. \tag{3.2}$$

for all sufficiently large n.

Proof. This follows from the (multidimensional version of the) Avram-Parter theorem [10], which says that

$$\lim_{M\to\infty} \frac{1}{n}\sum_{j=1}^n \varphi(s_j(T_M(a))) = \int_{[0,1]^d} \varphi(|a(x)|)\,dx \tag{3.3}$$

for every continuous function $\varphi : \mathbb{R} \to \mathbb{C}$. This theorem tells us that the singular values of $T_M(a)$ are asymptotically distributed as the values of $|a(x)|$ (which is convincingly seen in Figure 1.1), and since $|a(x)| \ge \|a\|_\infty/2$ on a set of positive measure, we conclude that a strictly positive percentage of the singular values of $T_M(a)$ must be greater than $\|a\|_\infty/2$ as $M \to \infty$. Consequently, there is an $\alpha \in (0,1)$ such that $s_{[\alpha n]+1} \ge \|a\|_\infty/2$ for all sufficiently large M. Because $s_n \le \|a\|_\infty$, it results that $s_n^2/s_{[\alpha n]+1}^2 \le 4$ if n is large enough. Given this, (3.2) is a direct consequence of (2.8). □

Remark 3.2. If one does not insist on (3.2) as it is stated, one can proceed more elementary. Namely, (3.3) in the special case where $\varphi(t) = t$ (and with $s_j(T_M(a))$ abbreviated to s_j) implies that

$$\frac{s_1 + \ldots + s_n}{n} > \frac{1}{2} \int_{[0,1]^d} |a(x)|\, dx = \frac{\|a\|_1}{2}$$

for all sufficiently large n. Choose $\alpha \in (0,1)$ so that $(1-\alpha)\|a\|_\infty < \|a\|_1/4$. Then

$$\begin{aligned} \frac{\|a\|_1}{2} < \frac{s_1 + \ldots + s_n}{n} &\le \frac{([\alpha n]+1)s_{[\alpha n]+1} + (n - [\alpha n] - 1)s_n}{n} \\ &\le \frac{([\alpha n]+1)s_{[\alpha n]+1} + (n - [\alpha n] - 1)\|a\|_\infty}{n}, \end{aligned}$$

which shows that

$$\frac{\|a\|_1}{2} \le \alpha \liminf_{n\to\infty} s_{[\alpha n]+1} + (1-\alpha)\|a\|_\infty$$

and thus

$$s_{[\alpha n]+1} \ge \frac{1}{\alpha}\left(\frac{\|a\|_1}{4} - (1-\alpha)\|a\|_\infty\right)$$

for all n large enough. For these n,

$$\frac{s_n^2}{s_{[\alpha n]+1}^2} \le \frac{16\,\alpha^2}{(\|a\|_1/\|a\|_\infty - 4(1-\alpha))^2},$$

which can now be inserted in (2.8).

Example 3.3. One example was already cited in Section 1. To have another example, take $d = 1$ and $a(x) = x^\sigma$ ($\sigma > 0$). Then $\|a\|_\infty = 1$. We have $a(x) \ge 1/2$ if and only if $x \ge \alpha := 1/2^{1/\sigma}$. Thus, we can use (3.2) with this value of α. Notice that in this and any other concrete case, one can appropriately modify the argument of the proof of Theorem 3.1 to get less complicated constants. For instance, we can argue that $a(x) \ge 1/2^\sigma$ for $x \ge 1/2$, which, by the Avram-Parter theorem, implies that

$$s_{[n/2]+1} \ge \sqrt{\frac{16}{17}}\,\frac{1}{2^\sigma}$$

for all sufficiently large n. Estimate (2.7) now yields

$$\mathbb{P}\left(\frac{\|\delta p\|^2}{\|\delta y\|^2} \le 17 \cdot 4^\sigma\right) \ge 1 - \frac{1}{\sqrt{n}}\left(\frac{1}{2}\right)^n.$$

4. Sampling of trigonometric polynomials. Let $\Pi_N^d := \{-N, \dots, N\}^d$ and $n := (2N+1)^d$. We want to find the coefficients p_k ($k \in \Pi_N^d$) of a trigonometric polynomial

$$p(x) = \sum_{k \in \Pi_N^d} p_k e^{2\pi i k \cdot x}$$

from the values $v_j = p(x_j)$ at prescribed points $x_1, \dots, x_r \in [0,1]^d$. Thus, we have to solve the system $Up = v$ with

$$U = \left(e^{2\pi i k \cdot x_j}\right)_{j \in \{1,\dots,r\}, k \in \Pi_N^d}, \quad p = (p_k)_{k \in \Pi_N^d}, \quad v = (v_j)_{j \in \{1,\dots,r\}}.$$

To take into account clusters in the sampling set $\{x_1, \dots, x_r\}$, we introduce the diagonal matrix $W = \mathrm{diag}\,(w_1, \dots, w_r)$ in which w_j are certain weights satisfying $\sum_j w_j = 1$ and consider the $r \times n$ system $\sqrt{W}Up = \sqrt{W}v$. Finally, to get a square system, we pass to the $n \times n$ system $Ap = y$ with $A = U^*WU$ and $y = U^*Wv$.

The matrix A is a Toeplitz matrix. In the notation of Section 3,

$$A = T_{2N+1}(a_N) \quad \text{with} \quad a_N(x) = \sum_{j=1}^{r} w_j e^{2\pi i k \cdot (x - x_j)}.$$

Since both the size and the generating function of A depend on N, the results of Section 3 are not applicable. Fortunately, the argument of Remark 3.2 can be carried over to the situation at hand.

The mesh norm ν of the set $\{x_1, \dots, x_r\}$ is defined as

$$\nu = 2 \max_{x \in [0,1]^d} \min_{j=1,\dots,r} \mathrm{dist}\,(x, x_j)$$

where the distance is taken in the $\|\cdot\|_\infty$ norm and periodicity is factored in, that is,

$$\mathrm{dist}\,(x,y) = \min_{k \in \mathbb{Z}^d} \|x - y + k\|_\infty \quad \text{with} \quad \|z\|_\infty = \max(|z_1|, \dots, |z_d|).$$

In [1], [7] it was shown that the spectral norm $\|A\|$ admits the estimate

$$\|A\| \le (2 + e^{2\pi d N \nu})^2 =: C_n^2. \tag{4.1}$$

THEOREM 4.1. *If $\alpha_n \in (0,1)$ and $1 - \alpha_n < 1/C_n^2$, then*

$$\frac{s_n^2}{s_{[\alpha_n n]+1}^2} \le \left(\frac{2\alpha_n C_n^2}{1 - (1-\alpha_n)C_n^2}\right)^2$$

and hence

$$\mathbb{P}\left(\frac{\|\delta p\|^2}{\|\delta y\|^2} \le \frac{2^{1/(1-\alpha_n)}}{(1-\alpha_n)\alpha_n^{\alpha_n/(1-\alpha_n)}} \left(\frac{2\alpha_n C_n^2}{1-(1-\alpha_n)C_n^2}\right)^2\right)$$
$$\ge 1 - \sqrt{\frac{\alpha_n}{1-\alpha_n}}\,\frac{1}{\sqrt{n}}\left(\frac{1}{2}\right)^n \tag{4.2}$$

for all sufficiently large n

Proof. The matrix A is positive semi-definite and hence its singular values $s_1 \le \ldots \le s_n$ coincide with the eigenvalues. The sum $s_1 + \ldots + s_n$ equals the trace of A, and since A is Toeplitz, the trace of A is $n(a_N)_0 = n \sum_{j=1}^{r} w_j = n$. This together with (4.1) gives $n = s_1 + \ldots + s_n \le ([\alpha_n n] + 1)s_{[\alpha_n n]+1} + (n - [\alpha_n n] - 1)C_n^2$ and thus

$$\liminf_{n\to\infty} s_{[\alpha_n n]+1} \ge \frac{1}{\alpha_n}(1 - (1-\alpha_n)C_n^2) > 0.$$

It follows that $s_{[\alpha_n n]+1} > (1 - (1-\alpha_n)C_n^2)/(2\alpha_n)$ for all n large enough. As $s_n \le C_n^2$, we get the desired estimate for $s_n^2/s_{[\alpha_n n]+1}^2$. Inequality (2.8) now yields (4.2). □

If we choose a sequence of meshes depending on n so that $\nu = O(1/n)$, then the C_n in (4.1) is bounded by a constant independent of n. The α_n in Theorem 4.1 does then also not depend on n, and the theorem's message becomes that the rate $\|\delta p\|/\|\delta y\|$ does not exceed a constant K independent of n with a probability converging exponentially fast to 1. Whether the K delivered by Theorem 4.1 is practically acceptable or astronomically large is another matter. The proof of (4.2) is based on very rough (but simple) arguments that do not aim at a best possible bound K.

EXAMPLE 4.2. Let $d = 2$ and let $\{x_1, \ldots, x_r\}$ be a linogram grid, which is formed by concentric squares centered at $(1/2, 1/2)$. To be more precise, pick natural numbers R and T and put

$$\{x_1, \ldots, x_r\} = \bigcup_{-R/2 \le j \le R/2-1} \; \bigcup_{-T/4 \le t \le T/4-1} \{x_{t,j}^{\mathrm{H}}, x_{t,j}^{\mathrm{V}}\}$$

where

$$x_{t,j}^{\mathrm{H}} = \left(\frac{1}{2} + \frac{j}{R}, \frac{1}{2} + \frac{4t}{T}\,\frac{j}{R}\right), \quad x_{t,j}^{\mathrm{V}} = \left(\frac{1}{2} - \frac{4t}{T}\,\frac{j}{R}, \frac{1}{2} + \frac{j}{R}\right).$$

We take the weights $w_{t,j} = \pi|j|/(TR^2)$. Different choices of the parameters R and T result in matrices A with very different singular value patterns. Figure 1.2 corresponds to the case $N = 5$, $T = 2N$, $R = 2N$. Three more examples are shown in Figures 4.1 to 4.3.

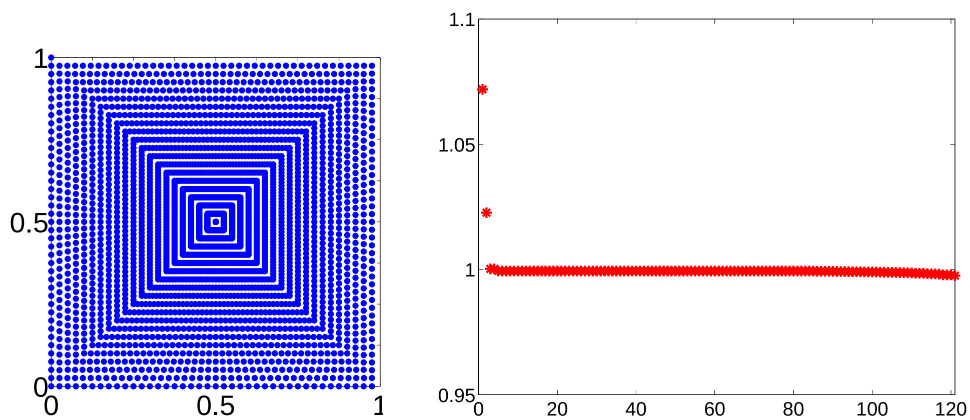

FIG. 4.1. *The linogram grid generated by the parameters $N = 5$, $T = 16N$, $R = 8N$ and the singular values of the 121×121 matrix that arises in the sampling of a bivariate trigonometric polynomial of degree 5 on the linogram grid.*

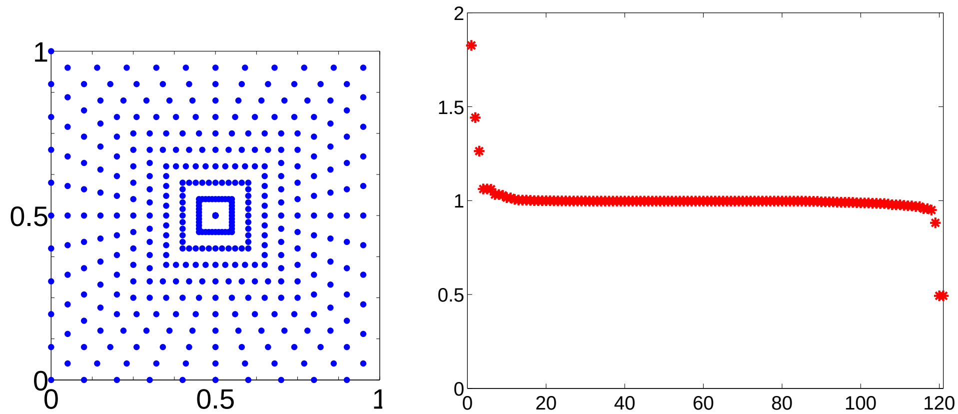

FIG. 4.2. *The linogram grid generated by the parameters $N = 5$, $T = 4N$, $R = 4N$ and the singular values of the 121×121 matrix that arises in the sampling of a bivariate trigonometric polynomial of degree 5 on the linogram grid.*

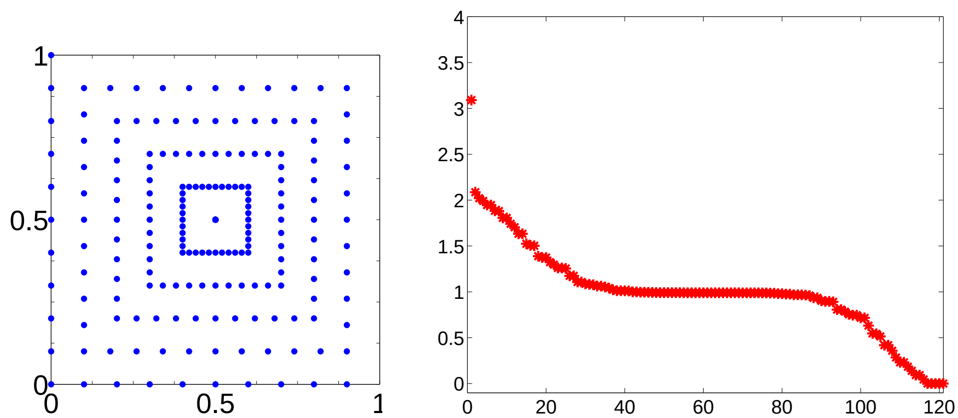

FIG. 4.3. *The linogram grid generated by the parameters $N = 5$, $T = 4N$, $R = 2N$ and the singular values of the 121×121 matrix that arises in the sampling of a bivariate trigonometric polynomial of degree 5 on the linogram grid.*

Let us embark on the case $N = 5$, $T = 2N$, $R = 2N$ (Figure 1.2). The number of points in this grid is $r = TR = 4\,N^2 = 100$. One can easily show that $\nu = 2/N$. Hence (4.1) holds with $C_n = 2 + e^{8\pi} =: C$ and Theorem 4.1 yields that

$$\mathbb{P}\left(\frac{\|\delta p\|}{\|\delta y\|} \le K\right) \ge 1 - L\,\frac{1}{11}\,\left(\frac{1}{2}\right)^{121} \tag{4.3}$$

with constants K and L independent of n. The determination of these constants from Theorem 4.1 runs into a disaster. The inequality $1 - \alpha < 1/C^2$ is satisfied for some α extremely close to 1, and for this α, (4.2) gives (4.3) with K and L about 10^{57} and 10^{11}, respectively. In spite of that the right-hand side of (4.3) is greater than 0.9999. Thus, while theoretically the rate $\|\delta p\|/\|\delta y\|$ turns out to remain bounded with high probability, we here meet a case in which the bound K is unacceptable practically. In Section 1, we used Figure 1.2 to see that actually 20.8 is a bound.

REFERENCES

[1] R. F. Bass and K. Gröchenig, Random sampling of multivariate trigonometric polynomials, *SIAM J. Math. Anal.*, 36:773–795, 2004.

[2] A. Böttcher and B. Silbermann, *Introduction to Large Truncated Toeplitz Matrices*, Springer-Verlag, New York 1998.

[3] A. Böttcher and B. Silbermann, *Analysis of Toeplitz Operators*, 2nd ed., Springer Verlag, Berlin, Heidelberg, New York 2006.

[4] J.D. Dixon, Estimating extremal eigenvalues and condition numbers of matrices, SIAM *J. Numer. Anal.*, 20:812–814, 1983.

[5] N.J.Higham, *Accuracy and Stability of Numerical Algorithms*, SIAM, 1996.

[6] M. Fenn, S. Kunis and D. Potts, A note on the polar FFT, Preprint 2006-5, Technische Universität Chemnitz, 2006.

[7] K. Gröchenig, Reconstruction algorithms in irregular sampling, *Lin. Alg. Appl.*, 199:181–194, 1992.

[8] G. M. Fichtenholz, *Differential- und Integralrechnung III*, fünfte Auflage, Deutscher Verlag d. Wisss., Berlin 1972.

[9] D. Potts and G. Steidl, A new linogram algorithm for computerized tomography, IMA *J. Numer. Anal.*, 21:769–782, 2001.

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