2012

TR-2012013: Condition Numbers of Random Toeplitz and Circulant Matrices

Victor Y. Pan

Guoliang Qian

Condition Numbers of Random Toeplitz and Circulant Matrices *

Victor Y. Pan[1,2],[a] and Guoliang Qian[2],[b]

[1] Department of Mathematics and Computer Science
Lehman College of the City University of New York
Bronx, NY 10468 USA
[2] Ph.D. Programs in Mathematics and Computer Science
The Graduate Center of the City University of New York
New York, NY 10036 USA
[a] victor.pan@lehman.cuny.edu
http://comet.lehman.cuny.edu/vpan/
[b] gqian@gc.cuny.edu

Abstract

Estimating the condition numbers of random structured matrices is a well known challenge (cf. [SST06]), linked to the design of efficient randomized matrix algorithms in [PGMQ], [PIMR10], [PQ10], [PQ12], [PQZa], [PQa], [PQZb], [PQZC], [PY09]. We deduce such estimates for Gaussian random Toeplitz and circulant matrices. The former estimates can be surprising because the condition numbers grow exponentially in n as $n \to \infty$ for some large and important classes of $n \times n$ Toeplitz matrices [BG05], whereas we prove the opposit for Gaussian random Toeplitz matrices. Our formal estimates are in good accordance with our numerical tests, except that circulant matrices tend to be even better conditioned according to the tests than according to our formal study.

2000 Math. Subject Classification: 15A52, 15A12, 65F22, 65F35

Key Words: Condition numbers, Random matrices, Toeplitz matrices, Circulant matrices

1 Introduction

It is well known that random matrices tend to be well conditioned [D88], [E88], [ES05], [CD05], [SST06], [B11], and this property can be exploited for advancing matrix computations (see e.g., [PGMQ], [PIMR10], [PQ10], [PQ12], [PQa], [PQZa], [PQZb], [PQZC] [PY09]). Exploiting matrix structure in these applications was supported empirically in the latter papers and formally in [T11]. An important step in this direction is the estimation of the condition numbers of structured matrices stated as a challenge in [SST06]. We reply to this challenge by estimating the condition numbers of Gaussian random Toeplitz and circulant matrices both formally (see Sections 6 and 7) and experimentally (see Tables 8.1–8.4). Our study shows that Gaussian random Toeplitz circulant matrices

*The results of this paper have been presented at the ACM-SIGSAM International Symposium on Symbolic and Algebraic Computation (ISSAC '2011), San Jose, CA, 2011, the 3nd International Conference on Matrix Methods in Mathematics and Applications (MMMA 2011) in Moscow, Russia, June 22-25, 2011, the 7th International Congress on Industrial and Applied Mathematics (ICIAM 2011), in Vancouver, British Columbia, Canada, July 18-22, 2011, the SIAM International Conference on Linear Algebra, in Valencia, Spain, June 18-22, 2012, and the Conference on Structured Linear and Multilinear Algebra Problems (SLA2012), in Leuven, Belgium, September 10-14, 2012 do not tend to be ill conditioned and the condition numbers of Gaussian random circulant $n \times n$ matrices tend to grow extremely slow as n grows large. Our numerical tests (the contribution of the second author) are in good accordance with our formal estimates, except that circulant matrices tended to be even better conditioned in the tests than according to our formal study. Our results on Toeplitz matrices are quite surprising because the condition numbers grow exponentially in n as $n \to \infty$ for some large and important classes of $n \times n$ Toeplitz matrices [BG05], which is opposit to the behavior of Gaussian random Toeplitz $n \times n$ matrices as we proved and consistently observed in our tests. Clearly, our study of Toeplitz matrices can be equally applied to Hankel matrices.

We organize our paper as follows. We recall some definitions and basic results on general matrix computations in the next section and on Toeplitz, Hankel and circulant matrices in Section 3. We define Gaussian random matrices and study their ranks and extremal singular values in Section 4. In Sections 6 and 7 we extend this study to Gaussian random Toeplitz and circulant matrices, respectively. In Section 8 we cover numerical tests, which constitute the contribution of the second author. In Section 9 we recall some applications of random circulant and Toeplitz matrices, which provide some implicit empirical support for our estimates for their condition numbers. We end with conclusions in Section 10.

2 Some definitions and basic results

Except for Theorem 3.2 and its application in the proof of Theorem 7.1 we work in the field $\mathbb{R}$ of real numbers. Next we recall some customary definitions of matrix computations [GL96], [S98]. A^T is the transpose of a matrix A. $||A||_h$ is its h-norm for $h = 1, 2, \infty$. We write $||A||$ to denote the 2-norm $||A||_2$. We have

$$\frac{1}{\sqrt{m}}||A||_1 \le ||A|| \le \sqrt{n}||A||_1, \quad ||A||_1 = ||A^T||_\infty, \quad ||A||^2 \le ||A||_1||A||_\infty, \tag{2.1}$$

for an $m \times n$ matrix A,

$$||AB||_h \le ||A||_h||B||_h \text{ for } h = 1, 2, \infty \text{ and any matrix product } AB. \tag{2.2}$$

Define an *SVD* or *full SVD* of an $m \times n$ matrix A of a rank ρ as follows,

$$A = S_A \Sigma_A T_A^T. \tag{2.3}$$

Here $S_A S_A^T = S_A^T S_A = I_m$, $T_A T_A^T = T_A^T T_A = I_n$, $\Sigma_A = \operatorname{diag}(\widehat{\Sigma}_A, O_{m-\rho,n-\rho})$, $\widehat{\Sigma}_A = \operatorname{diag}(\sigma_j(A))_{j=1}^{\rho}$, $\sigma_j = \sigma_j(A) = \sigma_j(A^T)$ is the jth largest singular value of a matrix A for $j = 1, \dots, \rho$, and we write $\sigma_j = 0$ for $j > \rho$. These values have the minimax property

$$\sigma_j = \max_{\dim(\mathbb{S})=j} \min_{\mathbf{x}\in\mathbb{S},\ ||\mathbf{x}||=1} ||A\mathbf{x}||,\ j = 1, \dots, \rho, \tag{2.4}$$

where $\mathbb{S}$ denotes linear spaces [GL96, Theorem 8.6.1].

Fact 2.1. *If A_0 is a submatrix of a matrix A, then $\sigma_j(A) \ge \sigma_j(A_0)$ for all j.*

Proof. [GL96, Corollary 8.6.3] implies the claimed bound where A_0 is any block of columns of the matrix A. Transposition of a matrix and permutations of its rows and columns do not change singular values, and thus we can extend the bounds to all submatrices A_0. □

$A^+ = T_A \operatorname{diag}(\widehat{\Sigma}_A^{-1}, O_{n-\rho,m-\rho})S_A^T$ is the Moore–Penrose pseudo-inverse of the matrix A of (2.3), and

$$||A^+|| = 1/\sigma_\rho(A) \tag{2.5}$$

for a matrix A of a rank ρ.

$\kappa(A) = \frac{\sigma_1(A)}{\sigma_\rho(A)} = ||A||\ ||A^+||$ is the condition number of an $m \times n$ matrix A of a rank ρ. Such matrix is *ill conditioned* if $\sigma_1(A) \gg \sigma_\rho(A)$ and is *well conditioned* otherwise. See [D83], [GL96, Sections 2.3.2, 2.3.3, 3.5.4, 12.5], [H02, Chapter 15], [KL94], and [S98, Section 5.3] on the estimation of the norms and condition numbers of nonsingular matrices.

3 Toeplitz, Hankel and f-circulant matrices

A *Toeplitz* $m \times n$ matrix $T_{m,n} = (t_{i-j})_{i,j=1}^{m,n}$ (resp. Hankel matrices $H = (h_{i+j})_{i,j=1}^{m,n}$) is defined by its first row and first (resp. last) column, that is by the vector $(t_h)_{h=1-n}^{m-1}$ (resp. $(h_g)_{g=2}^{m+n}$) of dimension $m+n-1$. We write $T_n = T_{n,n} = (t_{i-j})_{i,j=1}^{n,n}$ (see equation (3.1) below).

$\mathbf{e}_i$ is the ith coordinate vector of dimension n for $i = 1, \dots, n$. The reflection matrix $J = J_n(\mathbf{e}_n \mid \dots \mid \mathbf{e}_1)$ is the Hankel $n \times n$ matrix defined by its first column $\mathbf{e}_n$ and its last column $\mathbf{e}_1$. We have $J = J^T = J^{-1}$.

A lower *triangular Toeplitz* $n \times n$ matrix $Z(\mathbf{t}) = (t_{i-j})_{i,j=1}^{n}$ (where $t_k = 0$ for $k < 0$) is defined by its first column $\mathbf{t} = (t_h)_{h=0}^{n-1}$. We write $Z(\mathbf{t})^T = (Z(\mathbf{t}))^T$. $Z = Z_0 = Z(\mathbf{e}_2)$ is the downshift $n \times n$ matrix (see (3.1)). We have $Z\mathbf{v} = (v_i)_{i=0}^{n-1}$ and $Z(\mathbf{v}) = Z_0(\mathbf{v}) = \sum_{i=1}^{n} v_i Z^{i-1}$ for $\mathbf{v} = (v_i)_{i=1}^{n}$ and $v_0 = 0$,

$$T_n = \begin{pmatrix} t_0 & t_{-1} & \cdots & t_{1-n} \\ t_1 & t_0 & \ddots & \vdots \\ \vdots & \ddots & \ddots & t_{-1} \\ t_{n-1} & \cdots & t_1 & t_0 \end{pmatrix}, \quad Z = \begin{pmatrix} 0 & & \dots & & 0 \\ 1 & \ddots & & & \\ \vdots & \ddots & \ddots & & \vdots \\ & & \ddots & 0 & \\ 0 & & \dots & 1 & 0 \end{pmatrix}, \quad Z_f = \begin{pmatrix} 0 & & \dots & & f \\ 1 & \ddots & & & \\ \vdots & \ddots & \ddots & & \vdots \\ & & \ddots & 0 & \\ 0 & & \dots & 1 & 0 \end{pmatrix}. \quad (3.1)$$

Combine the equations $||Z(\mathbf{v})||_1 = ||Z(\mathbf{v})||_\infty = ||\mathbf{v}||_1$ with (2.1) to obtain

$$||Z(\mathbf{v})|| \le ||\mathbf{v}||_1. \quad (3.2)$$

Theorem 3.1. *Write $T_k = (t_{i-j})_{i,j=0}^{k-1}$ for $k = n, n+1$.*

(a) Let the matrix T_n be nonsingular and write $\mathbf{p} = T_n^{-1}\mathbf{e}_1$ and $\mathbf{q} = T_n^{-1}\mathbf{e}_n$. If $p_1 = \mathbf{e}_1^T\mathbf{p} \ne 0$, then $p_1 T_n^{-1} = Z(\mathbf{p})Z(J\mathbf{q})^T - Z(Z\mathbf{q})Z(ZJ\mathbf{p})^T$.

In parts (b) and (c) below let the matrix T_{n+1} be nonsingular and write $\widehat{\mathbf{v}} = (v_i)_{i=0}^{n} = T_{n+1}^{-1}\mathbf{e}_1$, $\mathbf{v} = (v_i)_{i=0}^{n-1}$, $\mathbf{v}' = (v_i)_{i=1}^{n}$, $\widehat{\mathbf{w}} = (w_i)_{i=0}^{n} = T_{n+1}^{-1}\mathbf{e}_{n+1}$, $\mathbf{w} = (w_i)_{i=0}^{n-1}$, and $\mathbf{w}' = (w_i)_{i=1}^{n}$.

(b) If $v_0 \ne 0$, then the matrix T_n is nonsingular and $v_0 T_n^{-1} = Z(\mathbf{v})Z(J\mathbf{w}')^T - Z(\mathbf{w})Z(J\mathbf{v}')^T$.

(c) If $v_n \ne 0$, then the matrix $T_{1,0} = (t_{i-j})_{i=1,j=0}^{n,n-1}$ is nonsingular and $v_n T_{1,0}^{-1} = Z(\mathbf{w})Z(J\mathbf{v}')^T - Z(\mathbf{v})Z(J\mathbf{w}')^T$.

Proof. See [GS72] on parts (a) and (b); see [GK72] on part (c). □

$Z_f = Z + f\mathbf{e}_1^T\mathbf{e}_n$ for a scalar $f \ne 0$ denotes the $n \times n$ matrix of *f-circular shift* (see (3.1)). An *f-circulant matrix* $Z_f(\mathbf{v}) = \sum_{i=1}^{n} v_i Z_f^{i-1}$ is a special Toeplitz $n \times n$ matrix defined by its first column vector $\mathbf{v} = (v_i)_{i=1}^{n}$ and a scalar f. f-circulant matrix is called *circulant* if $f = 1$ and *skew circulant* if $f = -1$. By replacing f with 0 we arrive at a lower triangular Toeplitz matrix $Z(\mathbf{v})$. The following theorem implies that the inverses (wherever they are defined) and pairwise products of f-circulant $n \times n$ matrices are f-circulant and can be computed in $O(n \log n)$ flops.

Theorem 3.2. *(See [CPW74].) We have $Z_1(\mathbf{v}) = \Omega^{-1}D(\Omega\mathbf{v})\Omega$. More generally, for any $f \ne 0$, we have $Z_{f^n}(\mathbf{v}) = U_f^{-1}D(U_f\mathbf{v})U_f$ where $U_f = \Omega D(\mathbf{f})$, $\mathbf{f} = (f^i)_{i=0}^{n-1}$, $D(\mathbf{u}) = \mathrm{diag}(u_i)_{i=0}^{n-1}$ for a vector $\mathbf{u} = (u_i)_{i=0}^{n-1}$, $\Omega = (\omega_n^{ij})_{i,j=0}^{n-1}$ is the $n \times n$ matrix of the discrete Fourier transform at n points, $\omega_n = \exp(\frac{2\pi}{n}\sqrt{-1})$ being a primitive n-th root of 1, and $\Omega^{-1} = \frac{1}{n}(\omega_n^{-ij})_{i,j=0}^{n-1} = \frac{1}{n}\Omega^H$.*

Hankel $m \times n$ matrices $H = (h_{i+j})_{i,j=1}^{m,n}$ can be defined equivalently as the products $H = TJ_n$ or $H = J_m T$ of $m \times n$ Toeplitz matrices T and the Hankel reflection matrices $J = J_m$ or J_n. Note that $J = J^{-1} = J^T$ and obtain the following simple fact.

Fact 3.1. *For $m = n$ we have $T = HJ$, $H^{-1} = JT^{-1}$ and $T^{-1} = JH^{-1}$ if $H = TJ$, whereas $T = JH$, $H^{-1} = JT^{-1}$ and $T^{-1} = H^{-1}J$ if $H = JT$. Furthermore in both cases $\kappa(H) = \kappa(T)$.*

By using the equations above we can readily extend any Toeplitz matrix inversion algorithm to Hankel matrix inversion and vice versa, preserving the flop count and condition numbers. E.g. $(JT)^{-1} = T^{-1}J$, $(TJ)^{-1} = JT^{-1}$, $(JH)^{-1} = H^{-1}J$ and $(HJ)^{-1} = JH^{-1}$.

4 Gaussian random matrices and their ranks

Definition 4.1. $F_\gamma(y) = Probability\{\gamma \le y\}$ *(for a real random variable γ) is the* cumulative distribution function (cdf) *of γ evaluated at y. $F_{g(\mu,\sigma)}(y) = \frac{1}{\sigma\sqrt{2\pi}}\int_{-\infty}^{y}\exp(-\frac{(x-\mu)^2}{2\sigma^2})dx$ for a Gaussian random variable $g(\mu,\sigma)$ with a mean μ and a positive variance σ^2, and so*

$$\mu - 4\sigma \le y \le \mu + 4\sigma \text{ with a probability near 1.} \tag{4.1}$$

Definition 4.2. *A matrix (or a vector) is a* Gaussian random matrix (or vector) *with a mean μ and a positive variance σ^2 if it is filled with independent identically distributed Gaussian random variables, all having the mean μ and variance σ^2. $\mathcal{G}_{\mu,\sigma}^{m\times n}$ is the set of such Gaussian random $m \times n$ matrices (which are* standard *for $\mu = 0$ and $\sigma^2 = 1$). By restricting this set to Toeplitz or f-circulant matrices we obtain the sets $\mathcal{T}_{\mu,\sigma}^{m\times n}$ and $\mathcal{Z}_{f,\mu,\sigma}^{n\times n}$ of* Gaussian random Toeplitz *and* Gaussian random f-circulant matrices*, respectively.*

Definition 4.3. *$\chi_{\mu,\sigma,n}(y)$ is the cdf of the norm $||\mathbf{v}|| = (\sum_{i=1}^n v_i^2)^{1/2}$ of a Gaussian random vector $\mathbf{v} = (v_i)_{i=1}^n \in \mathcal{G}_{\mu,\sigma}^{n\times 1}$. For $y \ge 0$ we have $\chi_{0,1,n}(y) = \frac{2}{2^{n/2}\Gamma(n/2)}\int_0^y x^{n-1}\exp(-x^2/2)dx$ where $\Gamma(h) = \int_0^\infty x^{h-1}\exp(-x)dx$, $\Gamma(n+1) = n!$ for nonnegative integers n.*

The total degree of a multivariate monomial is the sum of its degrees in all its variables. The total degree of a polynomial is the maximal total degree of its monomials.

Lemma 4.1. *[DL78], [S80], [Z79]. For a set Δ of a cardinality $|\Delta|$ in any fixed ring let a polynomial in m variables have a total degree d and let it not vanish identically on this set. Then the polynomial vanishes in at most $d|\Delta|^{m-1}$ points.*

We assume that Gaussian random variables range over infinite sets Δ, usually over the real line or its interval. Then the lemma implies that a nonzero polynomial vanishes with probability 0. Consequently a square Gaussian random general, Toeplitz or circulant matrix is nonsingular with probability 1 because its determinant is a polynomials in the entries. Likewise rectangular Gaussian random general, Toeplitz and circulant matrices have full rank with probability 1. Hereafter, wherever this causes no confusion, we assume by default that *Gaussian random general, Toeplitz and circulant matrices have full rank*.

5 Extremal singular values of Gaussian random matrices

Besides having full rank with probability 1, Gaussian random matrices in Definition 4.2 are likely to be well conditioned [D88], [E88], [ES05], [CD05], [B11], and even the sum $M + A$ for $M \in \mathbb{R}^{m\times n}$ and $A \in \mathcal{G}_{\mu,\sigma}^{m\times n}$ is likely to be well conditioned unless the ratio $\sigma/||M||$ is small or large [SST06].

The following theorem states an upper bound proportional to y on the cdf $F_{1/||A^+||}(y)$, that is on the probability that the smallest positive singular value $1/||A^+|| = \sigma_l(A)$ of a Gaussian random matrix A is less than a nonnegative scalar y (cf. (2.5)) and consequently on the probability that the norm $||A^+||$ exceeds a positive scalar x. The stated bound still holds if we replace the matrix A by $A - B$ for any fixed matrix B, and for $B = O_{m,n}$ the bounds can be strengthened by a factor $y^{|m-n|}$ [ES05], [CD05].

Theorem 5.1. *Suppose $A \in \mathcal{G}_{\mu,\sigma}^{m\times n}$, $B \in \mathbb{R}^{m\times n}$, $l = \min\{m, n\}$, $x > 0$, and $y \ge 0$. Then $F_{\sigma_l(A-B)}(y) \le 2.35\sqrt{l}y/\sigma$, that is $Probability\{||(A - B)^+|| \ge 2.35x\sqrt{l}/\sigma\} \le 1/x$.*

Proof. For $m = n$ this is [SST06, Theorem 3.3]. Apply Fact 2.1 to extend it to any pair $\{m, n\}$. □

The following two theorems supply lower bounds $F_{||A||}(z)$ and $F_{\kappa(A)}(y)$ on the probabilities that $||A|| \le z$ and $\kappa(A) \le y$ for two scalars y and z, respectively, and a Gaussian random matrix A. We do not use the second theorem, but state it for the sake of completeness and only for square $n \times n$ matrices A. The theorems imply that the functions $1 - F_{||A||}(z)$ and $1 - F_{\kappa(A)}(y)$ decay as $z \to \infty$ and $y \to \infty$, respectively, and that the two decays are exponential in $-z^2$ and proportional to $\sqrt{\log y}/y$, respectively. For small values $y\sigma$ and a fixed n the lower bound of Theorem 5.3 becomes negative, in which case the theorem becomes trivial. Unlike Theorem 5.1, in both theorems we assume that $\mu = 0$.

Theorem 5.2. *[DS01, Theorem II.7]. Suppose $A \in \mathcal{G}_{0,\sigma}^{m\times n}$, $h = \max\{m, n\}$ and $z \geq 2\sigma\sqrt{h}$. Then $F_{||A||}(z) \geq 1 - \exp(-(z - 2\sigma\sqrt{h})^2/(2\sigma^2))$, and so the norm $||A||$ is likely to have order $\sigma\sqrt{h}$.*

Theorem 5.3. *[SST06, Theorem 3.1]. Suppose $0 < \sigma \leq 1$, $y \geq 1$, $A \in \mathcal{G}_{0,\sigma}^{n\times n}$. Then the matrix A has full rank with probability 1 and $F_{\kappa(A)}(y) \geq 1 - (14.1 + 4.7\sqrt{(2\ln y)/n})n/(y\sigma)$.*

Proof. See [SST06, the proof of Lemma 3.2]. □

6 Extremal singular values of Gaussian random Toeplitz matrices

A matrix $T_n = (t_{i-j})_{i,j=1}^{n}$ is the sum of two triangular Toeplitz matrices

$$T_n = Z(\mathbf{t}) + Z(\mathbf{t}_-)^T,\ \mathbf{t} = (t_i)_{i=0}^{n-1},\ \mathbf{t}_- = (t'_{-i})_{i=0}^{n-1},\ t'_0 = 0. \tag{6.1}$$

If $T_n \in \mathcal{T}_{\mu,\sigma}^{n\times n}$, then T_n has $2n-1$ pairwise independent entries in $\mathcal{G}_{\mu,\sigma}$. Thus (3.2) implies that

$$||T_n|| \leq ||Z(\mathbf{t})|| + ||Z(\mathbf{t}_-)^T|| \leq ||\mathbf{t}||_1 + ||\mathbf{t}_-||_1 = ||(t_i)_{i=1-n}^{n-1}||_1 \leq \sqrt{2n-1}\ ||(t_i)_{i=1-n}^{n-1}||.$$

Recall Definition 4.2 and obtain

$$F_{||T_n||}(y) \geq \chi_{\mu,\sigma,2n-1}(y/\sqrt{2n-1}). \tag{6.2}$$

Next we estimate the norm $||T_n^{-1}||$ for $T_n \in \mathcal{T}_{\mu,\sigma}^{n\times n}$.

Lemma 6.1. *[SST06, Lemma A.2]. For a nonnegative scalar y, a unit vector $\mathbf{t} \in \mathbb{R}^{n\times 1}$, and a vector $\mathbf{b} \in \mathcal{G}_{\mu,\sigma}^{n\times 1}$, we have $F_{|\mathbf{t}^T\mathbf{b}|}(y) \leq \sqrt{\frac{2}{\pi}}\frac{y}{\sigma}$.*

Remark 6.1. *The latter bound is independent of μ and n; it holds for any μ even if all coordinates of the vector $\mathbf{b}$ are fixed except for a single coordinate in $\mathcal{G}_{\mu,\sigma}$.*

Theorem 6.1. *Given a matrix $T_n = (t_{i-j})_{i,j=1}^{n} \in \mathcal{T}_{\mu,\sigma}^{n\times n}$, assumed to be nonsingular (cf. Section 4), write $p_1 = \mathbf{e}_1^T T_n^{-1}\mathbf{e}_1$. Then $F_{1/||p_1 T_n^{-1}||}(y) \leq 2n\alpha\beta$ for two random variables α and β such that*

$$F_\alpha(y) \leq \sqrt{\frac{2n}{\pi}}\frac{y}{\sigma} \text{ and } F_\beta(y) \leq \sqrt{\frac{2n}{\pi}}\frac{y}{\sigma} \text{ for } y \geq 0. \tag{6.3}$$

Proof. Recall from part (a) of Theorem 3.1 that $p_1 T_n^{-1} = Z(\mathbf{p})Z(J\mathbf{q})^T - Z(Z\mathbf{q})Z(ZJ\mathbf{p})^T$. Therefore $||p_1 T_n^{-1}|| \leq ||Z(\mathbf{p})||\ ||Z(J\mathbf{q})^T|| + ||Z(Z\mathbf{q})||\ ||Z(ZJ\mathbf{p})^T||$ for $\mathbf{p} = T_n^{-1}\mathbf{e}_1$, $\mathbf{q} = T_n^{-1}\mathbf{e}_n$, and $p_1 = \mathbf{p}^T\mathbf{e}_1$. It follows that $||p_1 T_n^{-1}|| \leq ||Z(\mathbf{p})||\ ||Z(J\mathbf{q})|| + ||Z(Z\mathbf{q})||\ ||Z(ZJ\mathbf{p})||$ since $||A|| = ||A^T||$ for all matrices A. Furthermore $||p_1 T_n^{-1}|| \leq ||\mathbf{p}||_1\ ||J\mathbf{q}||_1 + ||Z\mathbf{q}||_1\ ||ZJ\mathbf{p}||_1$ due to (3.2). Clearly $||J\mathbf{v}||_1 = ||\mathbf{v}||_1$ and $||Z\mathbf{v}||_1 \leq ||\mathbf{v}||_1$ for every vector $\mathbf{v}$, and so (cf. (2.1))

$$||p_1 T_n^{-1}|| \leq 2||\mathbf{p}||_1\ ||\mathbf{q}||_1 \leq 2n||\mathbf{p}||\ ||\mathbf{q}||. \tag{6.4}$$

By definition the vector $\mathbf{p}$ is orthogonal to the vectors $T_n\mathbf{e}_2, \ldots, T_n\mathbf{e}_n$, whereas $\mathbf{p}^T T_n\mathbf{e}_1 = 1$ (cf. [SST06]). Consequenty the vectors $T_n\mathbf{e}_2, \ldots, T_n\mathbf{e}_n$ uniquely define the vector $\mathbf{u} = \mathbf{p}/||\mathbf{p}||$, whereas $|\mathbf{u}^T T_n\mathbf{e}_1| = 1/||\mathbf{p}||$. The last coordinate t_{n-1} of the vector $T_n\mathbf{e}_1$ is independent of the vectors $T_n\mathbf{e}_2, \ldots, T_n\mathbf{e}_n$ and consequently of the vector $\mathbf{u}$. Apply Remark 6.1 to estimate the cdf of the random variable $\alpha = 1/||\mathbf{p}|| = |\mathbf{u}^T T_n\mathbf{e}_1|$ and obtain that $F_\alpha(y) \leq \sqrt{\frac{2n}{\pi}}\frac{y}{\sigma}$ for $y \geq 0$.

Likewise the $n-1$ column vectors $T\mathbf{e}_1, \dots, T_{n-1}$ define the vector $\mathbf{v} = \beta\mathbf{q}$ for $\beta = 1/||\mathbf{q}|| = |\mathbf{v}^T T_n \mathbf{e}_n|$. The first coordinate t_{1-n} of the vector $T_n\mathbf{e}_n$ is independent of the vectors $T\mathbf{e}_1, \dots, T_{n-1}$ and consequently of the vector $\mathbf{v}$. Apply Remark 6.1 to estimate the cdf of the random variable β and obtain that $F_\beta(y) \le \sqrt{\frac{2n}{\pi}}\frac{y}{\sigma}$ for $y \ge 0$. Finally combine these bounds on the cdfs $F_\alpha(y)$ and $F_\beta(y)$ with (6.4). □

By applying parts (b) and (c) of Theorem 3.1 instead of its part (a), we similarly deduce the bounds $||v_0 T_{n+1}^{-1}|| \le 2\alpha\beta$ and $||v_n T_{n+1}^{-1}|| \le 2\alpha\beta$ for two pairs of random variables α and β that satisfy (6.3) for $n+1$ replacing n. We have $p_1 = \frac{\det T_{n-1}}{\det T_n}$, $v_0 = \frac{\det T_n}{\det T_{n+1}}$, and $v_n = \frac{\det T_{0,1}}{\det T_{n+1}}$ for $T_{0,1} = (t_{i-j})_{i=0,j=1}^{n-1,n}$. Next we bound the geometric means of the ratios $|\frac{\det T_{h+1}}{\det T_h}|$ for $h = 1, \dots, k-1$. $1/|p_1|$ and $1/|v_0|$ are such ratios for $k = n-1$ and $k = n$, respectively, whereas the ratio $1/|v_n|$ is similar to $1/|v_0|$, under slightly distinct notation.

Theorem 6.2. *Let $T_h \ne O$ denote $h \times h$ matrices for $h = 1, \dots, k$ whose entries have absolute values at most t for a fixed scalar or random variable t, e.g. for $t = ||T||$. Furthermore let $T_1 = (t)$. Then the geometric mean $(\prod_{h=1}^{k-1} |\frac{\det T_{h+1}}{\det T_h}|)^{1/(k-1)} = \frac{1}{t}|\det T_k|^{1/(k-1)}$ is at most $k^{\frac{1}{2}(1+\frac{1}{k-1})}t$.*

Proof. The theorem follows from Hadamard's upper bound $|\det M| \le k^{k/2}t^k$, which holds for any $k \times k$ matrix $M = (m_{i,j})_{i,j=1}^k$ with $\max_{i,j=1}^k |m_{i,j}| \le t$. □

The theorem says that the geometric mean of the ratios $|\det T_{h+1}/\det T_h|$ for $h = 1, \dots, k-1$ is not greater than $k^{0.5+\epsilon(k)}t$ where $\epsilon(k) \to 0$ as $k \to \infty$. Furthermore if $T_n \in \mathcal{T}_{\mu,\sigma}^{n\times n}$ we can write $t = ||T||$ and apply (6.2) to bound the cdf of t.

7 Extremal singular values of Gaussian random circulant matrices

Next we estimate the norms of a random Gaussian f-circulant matrix and its inverse.

Theorem 7.1. *Assume $y \ge 0$ and a circulant $n \times n$ matrix $T = Z_1(\mathbf{v})$ for $\mathbf{v} \in \mathcal{G}_{\mu,\sigma}^{n\times 1}$. Then*
(a) $F_{||T||}(y) \ge \chi_{\mu,\sigma,n}(\sqrt{\frac{2}{n}}y)$ for $\chi_{\mu,\sigma,n}(y)$ in Definition 4.3 and (b) $F_{1/||T^{-1}||}(y) \le \sqrt{\frac{2}{\pi}}\frac{ny}{\sigma}$.

Proof. For the matrix $T = Z_1(\mathbf{v})$ we have both equation (6.1) and the bound $||\mathbf{t}_-||_1 \le ||\mathbf{t}||_1$, and so $||T||_1 \le 2||\mathbf{t}||_1$. Now part (a) of the theorem follows similarly to (6.2). To prove part (b) recall Theorem 3.2 and write $B = \Omega T \Omega^{-1} = D(\mathbf{u})$, $\mathbf{u} = (u_i)_{i=0}^{n-1} = \Omega\mathbf{v}$. We have $\sigma_j(T) = \sigma_j(B)$ for all j because $\frac{1}{\sqrt{n}}\Omega$ and $\sqrt{n}\Omega^{-1}$ are unitary matrices. By combining the equations $u_i = \mathbf{e}_i^T\Omega\mathbf{v}$, the bounds $||\Re(\mathbf{e}_i^T\Omega)|| \ge 1$ for all i, and Lemma 6.1, deduce that $F_{|\Re(u_i)|}(y) \le \sqrt{\frac{2}{\pi}}\frac{y}{\sigma}$ for $i = 1, \dots, n$. We have $F_{\sigma_n(B)}(y) = F_{\min_i |u_i|}(y)$ because $B = \operatorname{diag}(u_i)_{i=0}^{n-1}$, and clearly $|u_i| \ge |\Re(u_i)|$. □

Remark 7.1. *Our extensive experiments suggest that the estimates of Theorem 7.1 are overly pessimistic (cf. Table 8.4).*

Combining Theorem 3.2 with minimax property (2.4) implies that

$$\frac{1}{g(f)}\sigma_j(Z_1(\mathbf{v})) \le \sigma_j(Z_f(\mathbf{v})) \le g(f)\sigma_j(Z_1(\mathbf{v}))$$

for all vectors $\mathbf{v}$, scalars $f \ne 0$, $g(f) = \max\{|f|^2, 1/|f|^2\}$, and $j = 1, \dots, n$. Thus we can readily extend the estimates of Theorem 7.1 to f-circulant matrices for $f \ne 0$. In particular Gaussian random f-circulant matrices tend to be well conditioned unless $f \approx 0$ or $1/f \approx 0$.

8 Numerical Experiments

Our numerical experiments with random general, Hankel, Toeplitz and circulant matrices have been performed in the Graduate Center of the City University of New York on a Dell server with a dual core 1.86 GHz Xeon processor and 2G memory running Windows Server 2003 R2. The test Fortran code was compiled with the GNU gfortran compiler within the Cygwin environment. Random numbers were generated with the random_number intrinsic Fortran function, assuming the uniform probability distribution over the range $\{x : -1 \leq x < 1\}$. The tests have been designed by the first author and performed by his coauthors.

We have computed the condition numbers of random general $n \times n$ matrices for $n = 2^k$, $k = 5, 6, \ldots$, with entries sampled in the range $[-1, 1)$ as well as complex general, Toeplitz, and circulant matrices whose entries had real and imaginary parts sampled at random in the same range $[-1, 1)$. We performed 100 tests for each class of inputs, each dimension n, and each nullity r. Tables 8.2–8.4 display the test results. The last four columns of each table display the average (mean), minimum, maximum, and standard deviation of the computed condition numbers of the input matrices, respectively. Namely we computed the values $\kappa(A) = ||A||\ ||A^{-1}||$ for general, Toeplitz, and circulant matrices A and the values $\kappa_1(A) = ||A||_1\ ||A^{-1}||_1$ for Toeplitz matrices A. We computed and displayed in Table 8.3 the 1-norms of Toeplitz matrices and their inverses rather than their 2-norms to facilitate the computations in the case of inputs of large sizes. Relationships (2.1) link the 1-norms and 2-norms to one another, but the empirical data in Table 8.1 consistently show even closer links, in all cases of general, Toeplitz, and circulant $n \times n$ matrices A where $n = 32, 64, \ldots, 1024$.

Table 8.1: The norms of random general, Toeplitz and circulant $n \times n$ matrices and of their inverses

matrix A	n	$\|\|A\|\|_1$	$\|\|A\|\|_2$	$\frac{\|\|A\|\|_1}{\|\|A\|\|_2}$	$\|\|A^{-1}\|\|_1$	$\|\|A^{-1}\|\|_2$	$\frac{\|\|A^{-1}\|\|_1}{\|\|A^{-1}\|\|_2}$
General	32	1.9×10^1	1.8×10^1	1.0×10^0	4.0×10^2	2.1×10^2	1.9×10^0
General	64	3.7×10^1	3.7×10^1	1.0×10^0	1.2×10^2	6.2×10^1	2.0×10^0
General	128	7.2×10^1	7.4×10^1	9.8×10^{-1}	3.7×10^2	1.8×10^2	2.1×10^0
General	256	1.4×10^2	1.5×10^2	9.5×10^{-1}	5.4×10^2	2.5×10^2	2.2×10^0
General	512	2.8×10^2	3.0×10^2	9.3×10^{-1}	1.0×10^3	4.1×10^2	2.5×10^0
General	1024	5.4×10^2	5.9×10^2	9.2×10^{-1}	1.1×10^3	4.0×10^2	2.7×10^0
Toeplitz	32	1.8×10^1	1.9×10^1	9.5×10^{-1}	2.2×10^1	1.3×10^1	1.7×10^0
Toeplitz	64	3.4×10^1	3.7×10^1	9.3×10^{-1}	4.6×10^1	2.4×10^1	2.0×10^0
Toeplitz	128	6.8×10^1	7.4×10^1	9.1×10^{-1}	1.0×10^2	4.6×10^1	2.2×10^0
Toeplitz	256	1.3×10^2	1.5×10^2	9.0×10^{-1}	5.7×10^2	2.5×10^2	2.3×10^0
Toeplitz	512	2.6×10^2	3.0×10^2	8.9×10^{-1}	6.9×10^2	2.6×10^2	2.6×10^0
Toeplitz	1024	5.2×10^2	5.9×10^2	8.8×10^{-1}	3.4×10^2	1.4×10^2	2.4×10^0
Circulant	32	1.6×10^1	1.8×10^1	8.7×10^{-1}	9.3×10^0	1.0×10^1	9.2×10^{-1}
Circulant	64	3.2×10^1	3.7×10^1	8.7×10^{-1}	5.8×10^0	6.8×10^0	8.6×10^{-1}
Circulant	128	6.4×10^1	7.4×10^1	8.6×10^{-1}	4.9×10^0	5.7×10^0	8.5×10^{-1}
Circulant	256	1.3×10^2	1.5×10^2	8.7×10^{-1}	4.7×10^0	5.6×10^0	8.4×10^{-1}
Circulant	512	2.6×10^2	3.0×10^2	8.7×10^{-1}	4.5×10^0	5.4×10^0	8.3×10^{-1}
Circulant	1024	5.1×10^2	5.9×10^2	8.7×10^{-1}	5.5×10^0	6.6×10^0	8.3×10^{-1}

9 Implicit empirical support of the estimates of Sections 6 and 7

The papers [PQa] and [PQZa] describe successful applications of randomized circulant and Toeplitz multipliers to some fundamental matrix computations. These applications were bound to fail if the

Table 8.2: The condition numbers $\kappa(A)$ of random $n \times n$ matrices A

n	**input**	**min**	**max**	**mean**	**std**
32	real	2.4×10^1	1.8×10^3	2.4×10^2	3.3×10^2
64	real	4.6×10^1	1.1×10^4	5.0×10^2	1.1×10^3
128	real	1.0×10^2	2.7×10^4	1.1×10^3	3.0×10^3
256	real	2.4×10^2	8.4×10^4	3.7×10^3	9.7×10^3
512	real	3.9×10^2	7.4×10^5	1.8×10^4	8.5×10^4
1024	real	8.8×10^2	2.3×10^5	8.8×10^3	2.4×10^4
2048	real	2.1×10^3	2.0×10^5	1.8×10^4	3.2×10^4

Table 8.3: The condition numbers $\kappa_1(A) = \frac{||A||_1}{||A^{-1}||_1}$ of random Toeplitz $n \times n$ matrices A

n	**min**	**mean**	**max**	**std**
256	9.1×10^2	9.2×10^3	1.3×10^5	1.8×10^4
512	2.3×10^3	3.0×10^4	2.4×10^5	4.9×10^4
1024	5.6×10^3	7.0×10^4	1.8×10^6	2.0×10^5
2048	1.7×10^4	1.8×10^5	4.2×10^6	5.4×10^5
4096	4.3×10^4	2.7×10^5	1.9×10^6	3.4×10^5
8192	8.8×10^4	1.2×10^6	1.3×10^7	2.2×10^6

multipliers were ill conditioned, and so the success gives some implicit empirical support to our probabilistic estimates of Sections 6 and 7 and motivates the effort for proving these estimates.

Namely it is well known that Gaussian elimination with no pivoting fails numerically where the input matrix has an ill conditioned leading block, even if the matrix itself is nonsingular and well conditioned. In our extensive tests in [PQa] and [PQZa] we consistently fixed this problem by means of multiplication by random circulant matrices. This implies that the random circulant matrices tend to be nonsingular and well conditioned for otherwise the products would be singular or ill conditioned.

Likewise in other tests in [PQZa] the column sets of the products A^TG of an $n \times m$ matrix A^T having a numerical rank ρ by random Toeplitz $m \times \rho$ multipliers consisently approximated some bases for the singular spaces associated with the ρ largest singular vaues of the matrix A, and this was readily extended to computing a rank-ρ approximation of the matrix A, which is a fundamental task of matrix computations [HMT11]. Then again one can immediately observe that these tests would have failed numerically if the multipliers and consequently the products were ill conditioned.

10 Conclusions

Estimating the condition numbers of random structured matrices is a well known challenge (cf. [SST06]). We deduce such estimates for Gaussian random Toeplitz and circulant matrices. The former estimates can be surprising because the condition numbers grow exponentially in n as $n \to \infty$ for some large and important classes of $n \times n$ Toeplitz matrices [BG05], whereas we prove the opposit for Gaussian random Toeplitz matrices. Our formal estimates are in good accordance with our numerical tests, except that circulant matrices tended to be even better conditioned in the tests than according to our formal study. The study of the condition number of Hankel matrices is immediately reduced to the study for Toeplitz matrices and vice versa. Can our progress be extended to other important classes of structured matrices?

Acknowledgements: Our research has been supported by NSF Grant CCF–1116736 and PSC

Table 8.4: The condition numbers $\kappa(A)$ of random circulant $n \times n$ matrices A

n	**min**	**mean**	**max**	**std**
256	9.6×10^{0}	1.1×10^{2}	3.5×10^{3}	4.0×10^{2}
512	1.4×10^{1}	8.5×10^{1}	1.1×10^{3}	1.3×10^{2}
1024	1.9×10^{1}	1.0×10^{2}	5.9×10^{2}	8.6×10^{1}
2048	4.2×10^{1}	1.4×10^{2}	5.7×10^{2}	1.0×10^{2}
4096	6.0×10^{1}	2.6×10^{2}	3.5×10^{3}	4.2×10^{2}
8192	9.5×10^{1}	3.0×10^{2}	1.5×10^{3}	2.5×10^{2}
16384	1.2×10^{2}	4.2×10^{2}	3.6×10^{3}	4.5×10^{2}
32768	2.3×10^{2}	7.5×10^{2}	5.6×10^{3}	7.1×10^{2}
65536	2.4×10^{2}	1.0×10^{3}	1.2×10^{4}	1.3×10^{3}
131072	3.9×10^{2}	1.4×10^{3}	5.5×10^{3}	9.0×10^{2}
262144	6.3×10^{2}	3.7×10^{3}	1.1×10^{5}	1.1×10^{4}
524288	8.0×10^{2}	3.2×10^{3}	3.1×10^{4}	3.7×10^{3}
1048576	1.2×10^{3}	4.8×10^{3}	3.1×10^{4}	5.1×10^{3}

CUNY Awards 64512–0042 and 65792–0043.

References

[B11] C. Beltrán, Estimates on the Condition Number of Random, Rank-deficient Matrices, *IMA Journal of Numerical Analysis*, **31, 1**, 25–39, 2011.

[BG05] A. Böttcher, S. M. Grudsky, *Spectral Properties of Banded Toeplitz Matrices*, SIAM Publications, Philadelphia, 2005.

[CD05] Z. Chen, J. J. Dongarra, Condition Numbers of Gaussian Random Matrices, *SIAM. J. on Matrix Analysis and Applications*, **27**, 603–620, 2005.

[CPW74] R. E. Cline, R. J. Plemmons, and G. Worm, Generalized Inverses of Certain Toeplitz Matrices, *Linear Algebra and Its Applications,* **8**, 25–33, 1974.

[D83] J. D. Dixon, Estimating Extremal Eigenvalues and Condition Numbers of Matrices, *SIAM J. on Numerical Analysis*, **20, 4**, 812–814, 1983.

[D88] J. Demmel, The Probability That a Numerical Analysis Problem Is Difficult, *Math. of Computation*, **50**, 449–480, 1988.

[DL78] R. A. Demillo, R. J. Lipton, A Probabilistic Remark on Algebraic Program Testing, *Information Processing Letters*, **7**, **4**, 193–195, 1978.

[DS01] K. R. Davidson, S. J. Szarek, Local Operator Theory, Random Matrices, and Banach Spaces, in *Handbook on the Geometry of Banach Spaces* (W. B. Johnson and J. Lindenstrauss editors), pages 317–368, North Holland, Amsterdam, 2001.

[E88] A. Edelman, Eigenvalues and Condition Numbers of Random Matrices, *SIAM J. on Matrix Analysis and Applications*, **9**, **4**, 543–560, 1988.

[ES05] A. Edelman, B. D. Sutton, Tails of Condition Number Distributions, *SIAM J. on Matrix Analysis and Applications*, **27**, **2**, 547–560, 2005.

[GK72] I. Gohberg, N. Y. Krupnick, A Formula for the Inversion of Finite Toeplitz Matrices, *Matematicheskiie Issledovaniia* (in Russian), **7**, **2**, 272–283, 1972.

[GL96] G. H. Golub, C. F. Van Loan, *Matrix Computations*, Johns Hopkins University Press, Baltimore, Maryland, 1996 (third addition).

[GS72] I. Gohberg, A. Sementsul, On the Inversion of Finite Toeplitz Matrices and Their Continuous Analogs, *Matematicheskiie Issledovaniia* (in Russian), **7**, **2**, 187–224, 1972.

[H02] N. J. Higham, *Accuracy and Stability in Numerical Analysis*, SIAM, Philadelphia, 2002 (second edition).

[HMT11] N. Halko, P. G. Martinsson, J. A. Tropp, Finding Structure with Randomness: Probabilistic Algorithms for Constructing Approximate Matrix Decompositions, *SIAM Review*, **53, 2**, 217–288, 2011.

[KL94] C. S. Kenney, A. J. Laub, Small-Sample Statistical Condition Estimates for General Matrix Functions, *SIAM J. on Scientific and Statistical Computing*, **15**, 36–61, 1994.

[PGMQ] V. Y. Pan, D. Grady, B. Murphy, G. Qian, R. E. Rosholt, A. Ruslanov, Schur Aggregation for Linear Systems and Determinants, *Theoretical Computer Science*, *Special Issue on Symbolic–Numerical Algorithms* (D. A. Bini, V. Y. Pan, and J. Verschelde editors), **409**, **2**, 255–268, 2008.

[PIMR10] V. Y. Pan, D. Ivolgin, B. Murphy, R. E. Rosholt, Y. Tang, X. Yan, Additive Preconditioning for Matrix Computations, *Linear Algebra and Its Applications*, **432**, 1070–1089, 2010.

[PQ10] V. Y. Pan, G. Qian, Randomized Preprocessing of Homogeneous Linear Systems of Equations, *Linear Algebra and Its Applications*, **432**, 3272–3318, 2010.

[PQ12] V. Y. Pan, G. Qian, Solving Linear Systems of Equations with Randomization, Augmentation and Aggregation, *Linear Algebra and Its Applications*, **437**, 2851–1876, 2012.

[PQa] V. Y. Pan, G. Qian, A. Zheng, More on the Power of Randomized Matrix Multiplication, Tech. Report TR 2012014, *Ph.D. Program in Computer Science*, *Graduate Center, the City University of New York* (2012), 2012.

[PQZa] V. Y. Pan, G. Qian, A. Zheng, Randomized Preprocessing versus Pivoting, *Linear Algebra and Its Applications*, in print. http://dx.doi.org/10.1016/j.laa.2011.02.052

[PQZb] V. Y. Pan, G. Qian, A. Zheng, Randomized Augmentations and Additive Preprocessing, Tech. Report TR 20120xz, *Ph.D. Program in Computer Science*, *Graduate Center, the City University of New York* (2012), 2012.

[PQZC] V. Y. Pan, G. Qian, A. Zheng, Z. Chen, Matrix Computations and Polynomial Rootfinding with Preprocessing, *Linear Algebra and Its Applications*, **434**, 854–879, 2011.

[PY09] V. Y. Pan, X. Yan, Additive Preconditioning, Eigenspaces, and the Inverse Iteration, *Linear Algebra and Its Applications*, **430**, 186–203, 2009.

[S80] J. T. Schwartz, Fast Probabilistic Algorithms for Verification of Polynomial Identities, *Journal of ACM*, **27**, **4**, 701–717, 1980.

[S98] G. W. Stewart, *Matrix Algorithms, Vol I: Basic Decompositions*, SIAM, 1998.

[SST06] A. Sankar, D. Spielman, S.-H. Teng, Smoothed Analysis of the Condition Numbers and Growth Factors of Matrices, *SIAM J. on Matrix Analysis*, **28**, **2**, 446–476, 2006.

[T11] J. A. Tropp, Improved analysis of the subsampled randomized Hadamard transform, *Adv. Adapt. Data Anal.*, **3, 1–2**, Special Issue, "Sparse Representation of Data and Images," 115-126, 2011.

[Z79] R. E. Zippel, Probabilistic Algorithms for Sparse Polynomials, *Proceedings of EUROSAM'79, Lecture Notes in Computer Science*, **72**, 216–226, Springer, Berlin, 1979.