The operator norm of random rectangular Toeplitz matrices

Radosław Adamczak*

Abstract

We consider rectangular $N \times n$ Toeplitz matrices generated by sequences of centered independent random variables and provide bounds on their operator norm under the assumption of finiteness of p-th moments ($p > 2$). We also show that if $N \gg n \log n$ then with high probability such matrices preserve the Euclidean norm up to an arbitrarily small error.

1 Introduction

Generalities on random Toeplitz matrices In recent years, following a question raised in [3] certain amount of work has been devoted to the study of random Toeplitz matrices, i.e. Toeplitz matrices determined by sequences of independent random variables. In particular in [7, 9] the convergence of the spectral measure for random symmetric Toeplitz matrices has been established while [5] provides a corresponding result for the spectral measure of XX^T, where X is a nonsymmetric random Toeplitz matrix. In both cases the limiting spectral distribution has unbounded support, which raises the question of the behaviour of the spectral norm of the matrix. In [14] it has been shown that if the underlying random variables are subgaussian and of mean zero, then the operator norm of an $n \times n$ matrix is of the order $\sqrt{n \log n}$. This result has been extended to matrices with bounded variance coefficients in [1], where also a strong law of large numbers with the normalization by expectation has been established. Although both papers consider symmetric matrices, their methods easily generalize to the non-symmetric square ones. Recently in [16], precise asymptotics of the operator norm have been found in the symmetric case. It turns out that if T_n is an $n \times n$ random symmetric Toeplitz matrix with mean zero, variance one coefficients with bounded p-th moments ($p > 2$), then

$$\frac{\|T_n\|_{\ell_2^n \to \ell_2^n}}{\sqrt{2n \log n}} \xrightarrow{L_p} \|S(x,y)\|_{2\to 4}^2,$$

where $S(x,y) = \frac{\sin(\pi(x-y))}{\pi(x-y)}$ is the sine kernel and $\|S(x,y)\|_{2\to 4}$ denotes the norm of the integral operator associated with it, acting from $L_2(\mathbb{R})$ to $L_4(\mathbb{R})$.

In this article we present results on the behaviour of the operator norm of a rectangular $N \times n$ random Toeplitz matrix with independent coefficients in terms of the matrix size. When n and N are of the same order of magnitude, this question can be easily reduced to the square

*Institute of Mathematics, University of Warsaw, ul. Banacha 2, 02-097 Warszawa, Poland. E-mail: radamcz@mimuw.edu.pl. Research partially supported by MNiSW Grant no. N N201 397437.

case, however for general matrices there seems to be no corresponding estimates in the literature. We remark that some results can be obtained from Theorem III.4 in [15], where a more general problem of estimating singular values of submatrices of a square random Toeplitz matrix is considered. This estimate however, when specialized to our problem is not optimal in the whole range of parameters (we discuss it briefly in the sequel).

Our main result gives estimates on the operator norm with optimal dependence on n and N. Additionally, in the case of tall matrices we provide conditions under which a properly scaled Toeplitz matrix preserves the Euclidean norm up to a small error.

Notation and the main result Throughout the article we will consider a random Toeplitz $N \times n$ matrix

$$T = [T_{ij}]_{1 \le i \le N, 1 \le j \le n} = [X_{i-j}]_{1 \le i \le N, 1 \le j \le n},$$

where $X_{1-n}, X_{2-n}, \ldots, X_{N-1}$ is a sequence of independent random variables.

We will denote absolute constants by C, and constants depending on some parameters (say a) by C_a. In both cases the value of a constant may differ between distinct occurrences.

We write ℓ_2^k for $\mathbb{R}^k$ equipped with the standard Euclidean structure (the corresponding inner product will be denoted by $\langle \cdot, \cdot \rangle$). For an $N \times n$ matrix A, by $\|A\|_{\ell_2^n \to \ell_2^N}$ we denote the operator norm of A acting between the spaces ℓ_2^n and ℓ_2^N, i.e. $\|A\|_{\ell_2^n \to \ell_2^N} = \sup_{x \in S^{n-1}} \sup_{y \in S^{N-1}} \langle Ax, y \rangle$.

Having desribed the notation, we are now ready to state our main result which is

Theorem 1.1. *Let $(X_i)_{1-n \le i \le N-1}$ be independent random variables with $\mathbb{E}X_i = 0$, $\mathbb{E}X_i^2 = 1$ and $\|X_i\|_p \le L$ for some $p > 2$. Then*

$$\mathbb{E}\|T\|_{\ell_2^n \to \ell_2^N} \le C_p L(\sqrt{N \vee n} + \sqrt{(N \wedge n) \log(N \wedge n)}). \tag{1}$$

Moreover for any $\delta, \varepsilon \in (0,1)$ if $N > C_{L,p,\delta,\varepsilon} n \log n$, then with probability at least $1 - \delta$, for all $x \in \mathbb{R}^n$,

$$(1 - \varepsilon)|x| \le |\frac{1}{\sqrt{N}} T_n x| \le (1 + \varepsilon)|x|. \tag{2}$$

We postpone the proof of the above theorem to Section 2.

A brief discussion of optimality One can easily see that the estimates of Theorem 1.1 are of the right order. Indeed, if X_i are independent Rademacher variables then the Euclidean length of the first column of the matrix is $\sqrt{N}$, while the Euclidean length of the first row is $\sqrt{n}$, which gives $\|T\|_{\ell_2^n \to \ell_2^N} \ge \sqrt{N \vee n}$. Moreover the matrix T contains a square Toeplitz submatrix with independent coefficients of size $(N \wedge n) \times (N \wedge n)$. By a straightforward modification of the argument presented in [14] for the symmetric case (see Theorem 3 therein), one can see that the operator norm of this submatrix is bounded from below by $c\sqrt{(N \wedge n) \log(N \wedge n)}$ for some absolute constant c (in fact instead of mimicking the proof one can also easily reduce the problem to the symmetric case). Standard symmetrization arguments allow to extend such estimates to other sequences of independent random variables satisfying a uniform lower bound on the absolute first moment (cf. the proof of Theorem 6 in [1]).

Further remarks The constants C_p in Theorem 1.1 explode when $p \to 2$, contrary to known inequalities on the operator norm of symmetric Toeplitz matrices. We present here a simple proposition, whose proof is based on general methods of probability in Banach spaces, which gives an estimate weaker than that of Theorem 1.1, but under the assumption of finiteness of the second moment of X_i's only. It's proof is deferred to Section 3.

Proposition 1.2. *Let $(X_i)_{1-n\le i\le N-1}$ be independent random variables with $\mathbb{E}X_i = 0$ and $\mathbb{E}X_i^2 = 1$. Then*

$$\mathbb{E}\|T\|_{\ell_2^n\to\ell_2^N} \le C(\sqrt{N\vee n} + \sqrt[4]{(N\wedge n)(N\vee n)}\sqrt{\log(N\wedge n)}).$$

Restricting our attention to the case $N \ge n$, we see that the above proposition gives an estimate of the same order as Theorem 1.1 (up to constants independent of n and N) if $N \le Cn$ or $N \ge cn\log^2 n$. In the former case the operator norm behaves like in the square case i.e. is of the order $\sqrt{n\log n}$, whereas in the latter one it is of the order $\sqrt{N}$, the same as the Euclidean length of a single column of the matrix. In the intermediate regime, i.e when $n \ll N \ll n\log^2 n$ one looses a logarithmic factor. It is natural to conjecture that the operator norm is of the order $\sqrt{N} + \sqrt{n\log n}$ for all $N \ge n$, however we do not know how to prove it without additional assumptions on higher moments of X_i's. As for the property (2), clearly it cannot hold just under the assumptions of the above proposition without some stronger integrability assumptions, since assuming just $\mathbb{E}X_i = 0, \mathbb{E}X_i^2 = 1$ still does not exclude the possibility that with probability close to one $X_i = 0$ for all i. Let us also remark that an estimate of the same order as in Proposition 1.2 can be obtained for matrices generated by Rademacher or Gaussian sequences using inequalities presented in [15] (as already mentioned in the introduction). In fact it can be also obtained by a modification of the proof of Theorem 1.1, however the argument presented in Section 3 is more concise.

2 Proof of Theorem 1.1

Without loss of generality we may assume that $N \ge n \ge 2$. In the main part of the proof we will not work with the original Toeplitz matrix, but with its modification, which will be more convenient for the calculations. Consider thus the matrix

$$\Gamma = [\Gamma_{ij}]_{1\le i\le N, 1\le j\le n}, \tag{3}$$

where $\Gamma_{ij} = T_{ij} = X_{i-j}$ if $j \le i \le N-n+j$ and $\Gamma_{ij} = 0$ otherwise. Let us note that T and Γ differ just by two "corners" of Toeplitz type and thus $\mathbb{E}\|T-\Gamma\|_{\ell_2^n\to\ell_2^N}$ can be estimated by means of results for square Toeplitz matrices. More precisely, by using Proposition 4.1 from the Appendix, we obtain

$$\mathbb{E}\|T-\Gamma\|_{\ell_2^n\to\ell_2^N} \le C\Big(\sum_{i\le -1 \text{ or } i\ge N-n+2} \mathbb{E}X_i^2\Big)^{1/2}\sqrt{\log n} \le C\sqrt{n\log n}.$$

Therefore for both assertions made in Theorem 1.1, the contribution from the corners is negligible (for the first part it is a direct consequence of the above inequality, whereas for the second part it follows easily by the above estimate and Chebyshev's inequality).

Denote the standard basis of ℓ_2^n and ℓ_2^N by $(e_j)_{j=1}^n$ and $(E_j)_{j=1}^N$ respectively and let $A_i\colon \ell_2^n \to \ell_2^N$, $i = 0, \ldots, N-n$ be the linear operator such that for all $1 \le j \le n$, $A_i e_j = E_{i+j}$ (in the sequel we will identify operators with their matrices in standard basis). Then $\Gamma = \sum_{i=0}^{N-n} X_i A_i$ and so

$$\Gamma^T \Gamma = \sum_{0 \le i,j \le N-n} X_i X_j A_i^T A_j. \tag{4}$$

Note that $A_i^T E_k = 0$ if $k < i+1$ or $k > i+n$ and $A_i^T E_k = e_{k-i}$ if $i+1 \le k \le n+i$. Therefore

$$\langle \Gamma^T \Gamma e_l, e_k \rangle = \sum_{i=0 \vee (l-k)}^{(N-n-(k-l)) \wedge (N-n)} X_i X_{k-l+i}. \tag{5}$$

In particular $\Gamma^T\Gamma$ is a symmetric Toeplitz matrix.

We will now state the main technical proposition of the paper, which will allow us to use standard symmetrization techniques in the proof of Theorem 1.1.

Proposition 2.1. *Let $N \ge n$ be two positive integers, $a_0, \ldots, a_{N-n}$ be real numbers and $g_0, \ldots, g_{N-n}$ be independent standard Gaussian variables. Define a symmetric $n \times n$ Toeplitz matrix $M = [M_{kl}]_{k,l \le n}$, where $M_{kk} = 0$ and for $k \ne l$,*

$$M_{kl} = Y_{|k-l|} := \sum_{i=0 \vee (l-k)}^{(N-n-(k-l)) \wedge (N-n)} a_i a_{k-l+i} g_i g_{k-l+i}.$$

Then

$$\begin{aligned}\mathbb{E}\|M\|_{\ell_2^n \to \ell_2^n} \le & C\Big(\sum_{0 \le i,j \le N-n} a_i^2 a_j^2 \mathbf{1}_{\{1 \le |i-j| \le n-1\}}\Big)^{1/2} \sqrt{\log n} \\ & + C \max_{0 \le k \le \lceil (N-n+1)/n \rceil - 1} \Big(\sum_{\substack{i \ne j \\ kn \le i,j \le ((k+2)n-1) \wedge (N-n)}} a_i^2 a_j^2 \Big)^{1/2} \log n.\end{aligned}$$

Proof of Proposition 2.1. Note first that without loss of generality we can assume that $N-n+1 \ge 2n$ and $N-n+1$ is divisible by n (we may simply enlarge N and put zeros as the new a_i's).

Since M is a symmetric Toeplitz matrix, to estimate the operator norm we may use the same strategy as in [14], i.e. relate the operator norm of M to the supremum of a random trigonometric polynomial for which we will use the entropy method. The main difference between our case and [14] is the fact that the coefficients of the polynomial will not be independent and the related supremum will be a chaos of degree 2, which will result in additional terms appearing in the entropy integral. Similarly as in [14] by extending M to an infinite Laurent matrix $[Y_{|k-l|}\mathbf{1}_{\{1 \le |k-l| \le n-1\}}]_{k,l \in \mathbb{Z}}$ and then noting that it corresponds to a multiplier on the circle we obtain that

$$
\begin{aligned}
\|M\|_{\ell_2^n\to\ell_2^n} &\le 2\sup_{0\le x\le 1}\Big|\sum_{j=1}^{n-1}Y_j\cos(2\pi jx)\Big| = 2\sup_{0\le x\le 1}\Big|\sum_{i=0}^{N-n-1}\sum_{j=1}^{(N-n-i)\wedge(n-1)} a_ia_{i+j}g_ig_{i+j}\cos(2\pi jx)\Big| \\
&= \sup_{0\le x\le 1}\Big|\sum_{0\le i,j\le N-n} B^x_{ij}g_ig_j\Big| =: \sup_{0\le x\le 1}|S_x|,
\end{aligned}
$$

where for $x\in[0,1]$, the matrix $B^x = [B^x_{ij}]_{i,j=0}^{N-n}$ is defined by

$$
B^x_{ij} = a_ia_j\cos(2\pi|i-j|x)\mathbf{1}_{\{1\le|i-j|\le n-1\}}.
$$

By Proposition 4.2 in the Appendix we obtain that

$$
\mathbb{P}(|S_x-S_y|\ge t)\le 2\exp\Big(-\frac{1}{C}\min\Big(\frac{t^2}{\|B^x-B^y\|_{HS}^2},\frac{t}{\|B^x-B^y\|_{\ell_2^{N-n+1}\to\ell_2^{N-n+1}}}\Big)\Big)
$$

and so, by Proposition 4.3 we get

$$
\|M\|_{\ell_2^n\to\ell_2^n}\le C\Big(\mathbb{E}|S_0| + \int_0^\infty\sqrt{\log([0,1],d_1,\varepsilon)}d\varepsilon + \int_0^\infty \log N([0,1],d_2,\varepsilon)d\varepsilon\Big), \tag{6}
$$

where $d_1(x,y) = \|B^x-B^y\|_{HS}$, $d_2(x,y) = \|B^x-B^y\|_{\ell_2^{N-n+1}\to\ell_2^{N-n+1}}$ and for a metric space $(\mathcal{X},d)$, $N(\mathcal{X},d,\varepsilon)$ denotes the minimum number of closed balls with radius ε covering $\mathcal{X}$.

Note that $\operatorname{diam}([0,1],d_1)\le 2\sqrt{\sum_{0\le i,j\le N-n}a_i^2a_j^2\mathbf{1}_{\{1\le|i-j|\le n-1\}}} =: D_1$. Also, using the Lipschitz property of the cosine function, we get that

$$
d_1(x,y)^2\le 4\pi^2\sum_{0\le i,j\le N-n}a_i^2a_j^2(i-j)^2\mathbf{1}_{\{1\le|i-j|\le n-1\}}|x-y|^2,
$$

which gives $N([0,1],d_1,\varepsilon)\le C\Delta_1/\varepsilon$ for $\varepsilon\le D_1$, where

$$
\Delta_1^2 = \sum_{0\le i,j\le N-n}a_i^2a_j^2(i-j)^2\mathbf{1}_{\{1\le|i-j|\le n-1\}}.
$$

We thus obtain

$$
\begin{aligned}
\int_0^\infty\sqrt{\log([0,1],d_1,\varepsilon)}d\varepsilon &\le \int_0^{D_1}\sqrt{\log\Big(\frac{C\Delta_1}{\varepsilon}\Big)}d\varepsilon \\
&= \frac{C\Delta_1}{\sqrt{2}}\int_{\sqrt{2\log(C\Delta_1/D_1)}}^\infty t^2e^{-t^2/2}dt \\
&\le D_1\sqrt{\log(C\Delta_1/D_1)} + \sqrt{\pi}D_1 \\
&\le CD_1\sqrt{\log n},
\end{aligned} \tag{7}
$$

where in the last inequality we used the estimate $\Delta_1\le nD_1$.

Let us now estimate the other integral on the right-hand side of (6). Note that B^x's are band matrices and they may be decomposed as $B^x = B_1^x + B_2^x + B_3^x$, where B_1^x is the block diagonal part of B^x with blocks of size $n \times n$, whereas B_2^x and B_3^x correspond respectively to the part of B^x below and above the block diagonal. More formally,

$$B_1^x = [B_{ij}^x \mathbf{1}_{\{\lfloor i/n\rfloor = \lfloor j/n\rfloor\}}]_{i,j=0}^{N-n}, B_2^x = [B_{ij}^x \mathbf{1}_{\{\lfloor i/n\rfloor = \lfloor j/n\rfloor+1\}}]_{i,j=0}^{N-n}, B_3^x = [B_{ij}^x \mathbf{1}_{\{\lfloor i/n\rfloor+1 = \lfloor j/n\rfloor\}}]_{i,j=0}^{N-n}.$$

The matrix $B_1^x - B_1^y$ consists of $(N-n+1)/n$ blocks and the Hilbert-Schmidt norm of the k-th block ($k = 1, \ldots, (N-n+1)/n$) is bounded by

$$\begin{aligned}
&\Big(\sum_{i,j=(k-1)n}^{kn-1} a_i^2 a_j^2 (\cos(2\pi|i-j|x) - \cos(2\pi|i-j|y))^2 \mathbf{1}_{\{1\le|i-j|\le n-1\}}\Big)^{1/2} \\
&\le 2\pi \Big(\sum_{\substack{i\neq j \\ (k-1)n\le i,j\le kn-1}} a_i^2 a_j^2 (i-j)^2\Big)^{1/2} |x-y|.
\end{aligned}$$

Thus for $x, y \in [0,1]$,

$$\|B_1^x - B_1^y\|_{\ell_2^{N-n+1}\to\ell_2^{N-n+1}} \le 2\pi|x-y| \max_{1\le k\le (N-n+1)/n} \Big(\sum_{\substack{i\neq j \\ (k-1)n\le i,j\le kn-1}} a_i^2 a_j^2 (i-j)^2\Big)^{1/2}.$$

By a similar estimate for all $x \in [0,1]$,

$$\|B_1^x - B_1^y\|_{\ell_2^{N-n+1}\to\ell_2^{N-n+1}} \le 2 \max_{1\le k\le (N-n+1)/n} \Big(\sum_{\substack{i\neq j \\ (k-1)n\le i,j\le kn-1}} a_i^2 a_j^2\Big)^{1/2}.$$

Bounds on B_2^x and B_3^x can be obtained in an analogous way, by exploring their block-diagonal structure (the blocks are not on the main diagonal, but still the operator norm of the whole matrix is the maximum of operator norms of individual blocks). Therefore we obtain

$$\begin{aligned}
\mathrm{diam}([0,1], d_2) \le & \max_{1\le k\le (N-n+1)/n} \Big(\sum_{\substack{i\neq j \\ (k-1)n\le i,j\le kn-1}} a_i^2 a_j^2\Big)^{1/2} \\
& + \max_{1\le k\le (N-n+1)/n-1} \Big(\sum_{\substack{(k-1)n\le j\le kn-1 \\ kn\le i\le (k+1)n-1}} a_i^2 a_j^2\Big)^{1/2} \\
& + \max_{1\le k\le (N-n+1)/n-1} \Big(\sum_{\substack{kn\le i\le (k+1)n-1 \\ (k-1)n\le j\le kn-1}} a_i^2 a_j^2\Big)^{1/2} \\
& \le 3 \max_{0\le k\le ((N-n+1)/n)-2} \Big(\sum_{\substack{i\neq j \\ kn\le i,j\le (k+2)n-1}} a_i^2 a_j^2\Big)^{1/2} =: D_2
\end{aligned}$$

and

$$d_2(x,y) \le C \max_{0\le k\le ((N-n+1)/n)-2} \Big(\sum_{\substack{i\neq j \\ kn\le i,j\le (k+2)n-1}} a_i^2 a_j^2 (i-j)^2 \Big)^{1/2} |x-y| =: \Delta_2 |x-y|,$$

which allows us to write

$$N([0,1], d_2, \varepsilon) \le \frac{\Delta_2}{\varepsilon}$$

for $\varepsilon \le D_2$. Thus

$$\int_0^\infty \log N([0,1], d_2, \varepsilon) d\varepsilon \le \int_0^{D_2} \log(\Delta_2 \varepsilon^{-1}) d\varepsilon = D_2 \log(\Delta_2) - D_2 \log D_2 + D_2 \le C D_2 \log n, \tag{8}$$

where in the last inequality we used the estimate $\Delta_2 \le CnD_2$.

Let us now note that $S_0 = \sum_{0\le i,j\le N-n} a_i a_j g_i g_j \mathbf{1}_{\{1\le |i-j|\le n-1\}}$ and so by independence of g_i's,

$$\mathbb{E}|S_0| \le \sqrt{\mathbb{E}|S_0|^2} = \sqrt{2 \sum_{0\le i,j\le N-n} a_i^2 a_j^2 \mathbf{1}_{\{1\le |i-j|\le n-1\}}},$$

which together with (6), (7) and (8) ends the proof of the proposition. □

Conclusion of the proof of Theorem 1.1. As explained at the beginning of this section, it suffices to prove the corresponding statements for $N \ge n$ and the matrix Γ defined by (3) instead of T.

We have

$$\Gamma^T \Gamma = \Big(\Gamma^T\Gamma - \Big(\sum_{i=0}^{N-n} X_i^2\Big)\mathrm{Id}_n\Big) + \Big(\sum_{i=0}^{N-n} X_i^2\Big)\mathrm{Id}_n. \tag{9}$$

Denote the first term on the right hand side above by $\tilde{M} = [\tilde{M}_{kl}]_{l,l\le n}$. From (5) it follows that $\tilde{M}_{kk} = 0$ and for $k \neq l$,

$$\tilde{M}_{kl} = \sum_{i=0\vee(l-k)}^{(N-n-(k-l))\wedge(N-n)} X_i X_{k-l+i},$$

thus $\tilde{M}$ is a tetrahedral chaos of order two (with matrix coefficients). Let $g_0, \ldots, g_{N-n}$ be i.i.d. standard Gaussian variables independent of the sequence (X_i). By Proposition 4.4 from the Appendix we obtain

$$\mathbb{E}\|\tilde{M}\|_{\ell_2^n\to\ell_2^n} \le \pi^2 \mathbb{E}\|M\|_{\ell_2^n\to\ell_2^n},$$

where the matrix M is defined as in Proposition 2.1 with $a_i = X_i$. Therefore, applying this proposition conditionally on (X_i) we get

$$\begin{aligned} \mathbb{E}\|\tilde{M}\|_{\ell_2^n\to\ell_2^n} \le & C\mathbb{E}\Big(\sum_{0\le i,j\le N-n} X_i^2 X_j^2 \mathbf{1}_{\{1\le|i-j|\le n-1\}}\Big)^{1/2}\sqrt{\log n} \\ & + C\mathbb{E}\max_{0\le k\le \lceil (N-n+1)/n\rceil -1} \Big(\sum_{kn\le i\le ((k+2)n-1)\wedge(N-n)} X_i^2\Big)\log n \end{aligned} \tag{10}$$

(note that we have enlarged the second summand of the estimate given in Proposition 2.1 by adding the diagonal terms).

Bounding the first summand on the right hand side of the above inequality is easy. By Jensen's inequality, independence and the assumption $\mathbb{E}X_i^2 = 1$ we get

$$\mathbb{E}\Big(\sum_{0\le i,j\le N-n} X_i^2X_j^2\mathbf{1}_{\{1\le|i-j|\le n-1\}}\Big)^{1/2}\sqrt{\log n} \le \sqrt{2Nn\log n}. \tag{11}$$

Let us now take care of the second term. Denote

$$Z_k = \sum_{kn\le i\le((k+2)n-1)\wedge(N-n)} X_i^2,\ \bar{Z}_k = Z_k - \mathbb{E}Z_k,$$

$k = 0,1,\ldots,\lceil(N-n+1)/n\rceil - 1$.

Set $q = (p/2)\wedge 2 \le 2$. Let also $\varepsilon_0,\ldots,\varepsilon_{N-n}$ be independent Rademacher variables, independent of the sequence (X_i). By standard symmetrization techniques and the Khintchine inequality we have

$$\begin{aligned}\mathbb{E}|\bar{Z}_k|^q &\le 2^q\mathbb{E}|\sum_{kn\le i\le((k+2)n-1)\wedge(N-n)} \varepsilon_iX_i^2|^q \le C\mathbb{E}(\sum_{kn\le i\le((k+2)n-1)\wedge(N-n)} X_i^4)^{q/2}\\ &\le C\mathbb{E}(\sum_{kn\le i\le((k+2)n-1)\wedge(N-n)} X_i^{2q}) \le C(\sum_{kn\le i\le((k+2)n-1)\wedge(N-n)} L^{2q}) \le CnL^{2q},\end{aligned}$$

where in the 3-rd inequality we used the fact $q \le 2$ and in the fourth one, $2q \le p$ and the definition of L. We thus get

$$\begin{aligned}&\mathbb{E}\max_{0\le k\le\lceil(N-n+1)/n\rceil-1}\Big(\sum_{kn\le i\le((k+2)n-1)\wedge(N-n)} X_i^2\Big)\\ &\le \max_{0\le k\le\lceil(N-n+1)/n\rceil-1}\Big(\sum_{kn\le i\le((k+2)n-1)\wedge(N-n)} \mathbb{E}X_i^2\Big) + \mathbb{E}\max_{0\le k\le\lceil(N-n+1)/n\rceil-1}|\bar{Z}_k|\\ &\le 2n + (\sum_{0\le k\le\lceil(N-n+1)/n\rceil-1} \mathbb{E}|\bar{Z}_k|^q)^{1/q} \le 2n + (C\frac{N}{n}nL^{2q})^{1/q} \le 2n + CL^2N^{1/q},\end{aligned}$$

which together with (10) and (11) gives

$$\mathbb{E}\|\tilde{M}\|_{\ell_2^n\to\ell_2^n} \le C(\sqrt{Nn\log n} + n\log n + L^2N^{1/q}\log n) \le C_pL^2(N + n\log n), \tag{12}$$

where we used that $q > 1$. Going back to (9), we see that it remains to estimate the second term on the right hand side. Clearly $\mathbb{E}\sum_{i=0}^{N-n} X_i^2 = N-n$, which together with (12) gives $\mathbb{E}\|\Gamma\|^2_{\ell_2^n\to\ell_2^N} = \mathbb{E}\|\Gamma^T\Gamma\|_{\ell_2^n\to\ell_2^n} \le C_pL^2(N + n\log n)$ (recall that $L \ge 1$) and proves the first assertion of the theorem.

To prove the second part, note that in the same way as for $\bar{Z}_k$ above, we get

$$\mathbb{E}|\sum_{i=0}^{N-n}(X_i^2 - 1)| \le CN^{1/q}L^2.$$

Thus by the first inequality of (12), one obtains

$$\mathbb{E}\|\Gamma^T\Gamma - (N-n+1)\mathrm{Id}_n\|_{\ell_2^n\to\ell_2^n} \le C(\sqrt{N\log n} + n\log n + L^2N^{1/q}\log n),$$

which gives

$$\mathbb{E}\|\frac{1}{N}\Gamma^T\Gamma - \mathrm{Id}_n\|_{\ell_2^n\to\ell_2^n} \le \varepsilon\delta$$

for $N \ge C_{L,p,\delta,\varepsilon} n\log n$. By Markov's inequality this yields

$$\|\frac{1}{N}\Gamma^T\Gamma - \mathrm{Id}_n\|_{\ell_2^n\to\ell_2^n} \le \varepsilon,$$

with probability at least $1-\delta$, which (after a suitable change of ε) easily implies the second part of the theorem. □

Remark In the proof above we did not try to obtain explicit dependence of the constant $C_{L,p,\delta,\varepsilon}$ on the parameters. Certain suboptimal estimates can be clearly read from the proof. Moreover, once the expectations of the variables involved are estimated, one can use general concentration inequalities for sums of independent random variables and suprema of polynomial chaoses to get a better estimate on the constants (depending on integrability properties of the underlying sequence of random variables). We do not pursue this direction here.

3 Proof of Proposition 1.2

We can again assume that $N \ge n \ge 2$ and prove the corresponding estimate for the matrix Γ. Going back to the equality (4) we obtain

$$\mathbb{E}\|\Gamma^T\Gamma\|_{\ell_2^n\to\ell_2^n} \le \mathbb{E}\|\sum_{0\le i\le N-n} X_i^2 A_i^T A_i\|_{\ell_2^n\to\ell_2^n} + \mathbb{E}\|\sum_{0\le i\neq j\le N-n} X_iX_jA_i^TA_j\|_{\ell_2^n\to\ell_2^n}.$$

Since $A_i^TA_i = \mathrm{Id}_n$ and $\mathbb{E}X_i^2 = 1$, the first term on the right hand side equals $N-n+1 \le N$. To bound the second term we use the fact that the space of $n\times n$ matrices equipped with the operator norm has type 2 constant bounded by $C\sqrt{\log n}$ (see Proposition 4.6 in the Appendix). Thus we can use Proposition 4.5 from the Appendix and get

$$(\mathbb{E}\|\sum_{0\le i\neq j\le N-n} X_iX_jA_i^TA_j\|_{\ell_2^n\to\ell_2^n})^2 \le C\Big(\sum_{0\le i\neq j\le N-n}\|A_i^TA_j\|_{\ell_2^n\to\ell_2^n}^2\Big)\log^2 n.$$

Note that if $|i-j|\ge n$ then $A_iA_j^T = 0$, moreover for all i,j, $\|A_iA_j^T\|_{\ell_2^n\to\ell_2^n} \le 1$, which together with the above inequality gives

$$\mathbb{E}\|\sum_{0\le i\neq j\le N-n} X_iX_jA_i^TA_j\|_{\ell_2^n\to\ell_2^n} \le C\sqrt{Nn}\log n$$

Combining this with the previous estimates we get $\mathbb{E}\|\Gamma\|^2_{\ell_2^n\to\ell_2^N} = \mathbb{E}\|\Gamma^T\Gamma\|_{\ell_2^n\to\ell_2^n} \le C(N + \sqrt{nN}\log n$, which ends the proof.

4 Appendix

For reader's convenience we gather here several by now standard results which have been used in the proofs above. For most of them we provide detailed references, however in some cases we haven't been able to find the formulation we need in the literature, so we briefly describe how they follow from available references.

The first proposition gives estimates on the operator norm of a square random Toeplitz matrix. It was proved in [1] for symmetric random Toeplitz matrices. A simple modification of the proof gives the result in the non-symmetric case, however it can be also easily obtained by exploring the type 2 property of the space of symmetric matrices (see Proposition 4.6 below) or e.g. by noncommutative Bernstein inequalities, since a random square Toeplitz matrix can be written as a linear combination with random coefficients of norm one matrices.

Proposition 4.1. *If $N = n$ and $(X_i)_{i \le 1-n \le n-1}$ are independent, centered random variables, then*

$$\mathbb{E}\|T\|_{\ell_2^n \to \ell_2^n} \le C\Big(\sum_{i=1}^{n} \mathbb{E}X_i^2\Big)^{1/2} \sqrt{\log n}.$$

We will now state concentration results on Gaussian chaoses of order 2. In a weaker form they can be traced to [10]. The present formulation can be deduced from results on a Banach space valued case in [4, 2] and appears explicitly in [11] and [12] (where a generalization to chaoses of higher degree has been obtained).

Proposition 4.2. *Let $g_1, g_2, \ldots, g_n$ be independent standard Gaussian random variables and let $A = (a_{ij})_{1 \le i,j \le n}$ be an array of real numbers such that for all $1 \le i \le n$, $a_{ii} = 0$. Then for any $t \ge 0$,*

$$\mathbb{P}\Big(\Big|\sum_{1 \le i,j \le n} a_{ij} g_i g_j\Big| \ge t\Big) \le 2\exp\Big(-\frac{1}{C}\min\Big(\frac{t^2}{\|A\|_{HS}^2}, \frac{t}{\|A\|_{\ell_2^n \to \ell_2^n}}\Big)\Big).$$

The next proposition is a consequence of the previous one, Theorem 1.2.7 in [17] and a standard comparison between γ_p functionals and entropy integrals.

Proposition 4.3. *Consider a set T provided with two distances d_1 and d_2 and a stochastic process $(X_t)_{t \in T}$ such that $\mathbb{E}X_t = 0$ for all $t \in T$ and for all $s, t \in T$ and $u > 0$,*

$$\mathbb{P}(|X_s - X_t| \ge u) \le 2\exp\Big(-\min\Big(\frac{u^2}{d_1(s,t)^2}, \frac{u}{d_2(s,t)}\Big)\Big).$$

Then

$$\mathbb{E}\sup_{s,t \in T} |X_s - X_t| \le C\Big(\int_0^\infty \sqrt{\log N(T, d_1, \varepsilon)}d\varepsilon + \int_0^\infty \log N(T, d_2, \varepsilon)d\varepsilon\Big).$$

Let us now state a simple proposition which combines standard symmetrization techniques with comparison between Gaussian and Rademacher averages.

Proposition 4.4. *Let $X_1, \ldots, X_n$ be independent centered random variables and let $(a_{ij})_{1 \le i \ne j \le n}$ be coefficients from a normed space $(F, \|\cdot\|)$. Finally let $g_1, \ldots, g_n$ be standard Gaussian variables independent of the sequence $X_1, \ldots, X_n$. Then*

$$\mathbb{E}\Big\| \sum_{1 \le i \ne j \le n} a_{ij} X_i X_j \Big\| \le \pi^2 \mathbb{E}\Big\| \sum_{1 \le i \ne j \le n} a_{ij} g_i g_j X_i X_j \Big\|$$

Proof. Let $\varepsilon_1, \ldots, \varepsilon_n$ be independent Rademacher variables, independent of the sequences (X_i), (g_i). Using repetitively (and conditionally) the fact that for any convex function $\varphi\colon \mathbb{R} \to \mathbb{R}$ we have $\mathbb{E}\varphi(X_i) \le \mathbb{E}\varphi(2\varepsilon_i X_i)$, we get

$$\mathbb{E}\Big\| \sum_{1 \le i \ne j \le n} a_{ij} X_i X_j \Big\| \le 4\mathbb{E}\Big\| \sum_{1 \le i \ne j \le n} a_{ij} \varepsilon_i \varepsilon_j X_i X_j \Big\|.$$

Now, by symmetry of g_i and Jensen's inequality,

$$\begin{aligned}
\frac{4}{\pi^2}\mathbb{E}\Big\| \sum_{1 \le i \ne j \le n} a_{ij} \varepsilon_i \varepsilon_j X_i X_j \Big\| &= \mathbb{E}\Big\| \sum_{1 \le i \ne j \le n} a_{ij} \varepsilon_i \varepsilon_j \mathbb{E}_g |g_i g_j| X_i X_j \Big\| \\
&\le \mathbb{E}\Big\| \sum_{1 \le i \ne j \le n} a_{ij} \varepsilon_i g_i \varepsilon_j g_j X_i X_j \Big\| \\
&= \mathbb{E}\Big\| \sum_{1 \le i \ne j \le n} a_{ij} g_i g_j X_i X_j \Big\|,
\end{aligned}$$

which ends the proof. □

Recall that a Banach space $(F, \|\cdot\|)$ is of type 2 if there exists a finite constant T_F, such that for all $a_1, \ldots, a_n \in F$ and independent Rademacher variables $\varepsilon_1, \ldots, \varepsilon_n$, we have

$$\mathbb{E}\|\varepsilon_1 a_1 + \ldots + \varepsilon_n a_n\|^2 \le T_F^2 \sum_{i=1}^n \|a_i\|^2. \tag{13}$$

The next proposition concerns basic properties of polynomial chaoses in spaces of type 2. It is very well known, so we skip the proof and remark only that it consists of the three following steps: 1) decoupling inequalities for chaoses (see e.g. Theorem 3.1.1. in [8]), 2) an iterative application of symmetrization inequalities, 3) iterative application of (13) conditionally on $(X_i)_{i=1}^n$.

Proposition 4.5. *Let $X_1, \ldots, X_n$ be independent centered, variance one random variables and let $(a_{ij})_{1 \le i \ne j \le n}$ be coefficients from a normed space $(F, \|\cdot\|)$ with type 2 constant T_F. Then*

$$\mathbb{E}\| \sum_{1 \le i \ne j \le n} a_{ij} X_i X_j \|^2 \le C T_F^4 \sum_{1 \le i \ne j \le n} \|a_{ij}\|^2.$$

Finally, the last proposition gives the type constant for the space of symmetric $n \times n$ matrices equipped with the operator norm.

Proposition 4.6. *The space $F = S_\infty^n$ of $n \times n$ symmetric matrices equipped with the operator norm has type 2 with constant $T_F \le C\sqrt{\log n}$.*

This proposition follows easily from estimates of type 2 constants for Schatten classes S_p^n given in [18] ($T_2(S_p^n) \le C\sqrt{p}$) and the fact that the Banach-Mazur distance between S_∞^n and S_p^n is equal to $n^{1/p}$ (it is enough to take $p = \log n$). We refer the reader e.g. to [19] for details on the Banach-Mazur distance and geometry of Banach spaces.

References

[1] Adamczak R. *A few remarks on the operator norm of random Toeplitz matrices.* J. Theoret. Probab., 2010, 23, 85-108.

[2] Arcones M. A., Giné, E. *On decoupling, series expansions, and tail behavior of chaos processes.* J. Theoret. Probab., 1993, 6, 101-122.

[3] Bai Z. D. *Methodologies in spectral analysis of large-dimensional random matrices, a review* Statist. Sinica, 1999, 9, 611-677.

[4] Borell C. *On a Taylor series of a Wiener polynomial.* In Seminar Notes on Multiple Stochastic Integration, polynomial chaos and their integration. 1984. Case Western Reserve Univ., Cleveland.

[5] Bose A., Gangopadhyay S., Sen, A. *Limiting spectral distribution of XX' matrices.* Ann. Inst. Henri Poincaré Probab. Stat., 2010, 46, 677-707.

[6] Bose A., Hazra R. S., Saha K. *Spectral norm of circulant-type matrices.* J. Theoret. Probab., 2011, 24, 479-516.

[7] Bryc W., Dembo A., Jiang T. *Spectral measure of large random Hankel, Markov and Toeplitz matrices.* Ann. Probab., 2006, 34, 1-38.

[8] de la Peña, V. H., Giné, E. Decoupling. From Dependence to Independence, Springer-Verlag, 1999.

[9] Hammond C., Miller S. J. *Distribution of eigenvalues for the ensemble of real symmetric Toeplitz matrices.* J. Theoret. Probab., 2005, 18, 537-566.

[10] Hanson D. L., Wright, F. T., *A bound on tail probabilities for quadratic forms in independent random variables.* Ann. Math. Statist., 1971, 42, 1079-1083.

[11] Latała R., *Tail and moment estimates for some types of chaos.* Studia Math., 1999, 135, 39-53.

[12] Latała R., *Estimates of moments and tails of Gaussian chaoses.* Ann. Probab., 2006, 34, 2315-2331.

[13] Ledoux M., Talagrand M. *Probability in Banach spaces* Springer-Verlag, 1991, 23.

[14] Meckes M. W. *On the spectral norm of a random Toeplitz matrix* Electron. Comm. Probab., 2007, 12, 315-325.

[15] Rauhut H. *Circulant and Toeplitz matrices in compressed sensing.* Proceedings SPARS '09. Saint-Malo, France, 2009.

[16] Sen A., Virág B. *The top eigenvalue of the random Toeplitz matrix and the Sine kernel.* Available at http://arxiv.org/abs/1109.5494.

[17] Talagrand M. *The generic chaining.* Springer-Verlag, 2005.

[18] Tomczak-Jaegermann N. *The moduli of smoothness and convexity and the Rademacher averages of trace classes* S_p $(1 < p < \infty)$. Studia Math. 50 (1974), 163–182.

[19] Tomczak-Jaegermann N. *Banach-Mazur distances and finite-dimensional operator ideals.* Longman Scientific & Technical, 1989, 38.