Circulant and Toeplitz Matrices in Compressed Sensing

Holger Rauhut
Hausdorff Center for Mathematics and Institute for Numerical Simulation
University of Bonn, Endenicher Allee 60, D-53115 Bonn
rauhut@hcm.uni-bonn.de

***Abstract*—Compressed sensing seeks to recover a sparse vector from a small number of linear and non-adaptive measurements. While most work so far focuses on Gaussian or Bernoulli random measurements we investigate the use of partial random circulant and Toeplitz matrices in connection with recovery by ℓ_1-minization. In contrast to recent work in this direction we allow the use of an arbitrary subset of rows of a circulant and Toeplitz matrix. Our recovery result predicts that the necessary number of measurements to ensure sparse reconstruction by ℓ_1-minimization with random partial circulant or Toeplitz matrices scales linearly in the sparsity up to a $\log$-factor in the ambient dimension. This represents a significant improvement over previous recovery results for such matrices. As a main tool for the proofs we use a new version of the non-commutative Khintchine inequality.**

I. Introduction

Compressed sensing is a recent concept in signal processing where one seeks to reconstruct efficiently a sparse signal from a minimal number of linear and non-adaptive measurements [1]. So far various measurement matrices have been investigated, most of them random matrices. Among these are Bernoulli and Gaussian matrices [2] (with independent ± 1 or standard normal entries) as well as partial Fourier matrices [3], [4], [5]. Recently, Bajwa et al. [6] (see also [7]) studied Toeplitz type and circulant matrices in the context of compressed sensing where the entries of the vector generating the Toeplitz or circulant matrix are chosen at random according to a suitable probability distribution. Compared to Bernoulli or Gaussian matrices random Toepliz and circulant matrices have the advantage that they require a reduced number of random numbers to be generated. More importantly, there are fast matrix-vector multiplication routines which can be exploited in recovery algorithms. Furthermore, they arise naturally in certain applications such as identifying a linear time-invariant system [8].

Basis Pursuit (ℓ_1-minimization) is one of the major approaches to efficiently recover a sparse vector. This technique is quite well understood by now. Modern optimization algorithms [9] such as LARS [10] (sometimes called homotopy method) are reasonably fast.

Bajwa et al. [6], [8] estimated the so-called restricted isometry constants of a random Toeplitz type or circulant matrix which then allows to provide recovery guarantees for ℓ_1-minimization. However, their bound is very pessimistic compared to related estimates for Bernoulli / Gaussian or partial Fourier matrices. More precisely, the estimated number of measurements grows with the sparsity squared, while one would rather expect a linear scaling. Indeed, this is also suggested by numerical experiments. We close the theoretical gap by providing recovery guarantees for ℓ_1-minimization in connection with circulant and Toeplitz type matrices where the necessary number of measurements scales linearly with the sparsity. However, we do not make use of the restricted isometry constants and a good estimate of the latter is therefore still open.

II. Sparse recovery with circulant and Toeplitz matrices

For a vector $x \in \mathbb{R}^N$ we let $\operatorname{supp} x = \{j, x_j \neq 0\}$ denote its support and $\|x\|_0 = |\operatorname{supp} x|$ the number of non-zero entries. It is called s-sparse if $\|x\|_0 \leq s$. We aim at recovering x from $y = Ax \in \mathbb{R}^n$ where A is a suitable $n \times N$ measurement matrix and $n < N$. A natural strategy is to consider ℓ_0-minimization,

$$\min_x \|x\|_0 \quad \text{subject to } Ax = y. \tag{1}$$

Unfortunately this combinatorial optimization problem is NP hard in general [11]. Therefore, we solve instead the convex problem

$$\min \|x\|_1 \quad \text{subject to } Ax = y, \tag{2}$$

where the ℓ_p-norm is defined as usual, $\|x\|_p = (\sum_{j=1}^N |x_j|^p)^{1/p}$. It is by now well understood that the solutions of both minimization problems often coincide and are equal to the original vector x, see e.g. [12], [13], [1], [14], [15]. A by now popular result [12], [16], [17] states that indeed (2) (stably) recovers all s-sparse x from $y = Ax$ provided the restricted isometry constant $\delta_{2s} \leq \delta < \sqrt{2} - 1$. The latter means that

$$(1-\delta)\|x\|_2^2 \leq \|Ax\|_2^2 \leq (1+\delta)\|x\|_2^2$$

for all $2s$-sparse vectors x. It is known [2] that random Gaussian or Bernoulli matrices, i.e. $n \times N$ matrices with independent and normal distributed or Bernoulli distributed entries, satisfy this condition with probability at least $1-\epsilon$ provided $s \leq C_1 n \log(N/s) + C_2 \log(\epsilon^{-1})$.

We consider the following types of measurement matrices. For $b = (b_0, b_1, \ldots, b_{N-1}) \in \mathbb{R}^N$ we let its associated circulant matrix $S = S^b \in \mathbb{R}^{N \times N}$ with entries $S_{i,j} = b_{j-i \mod N}$, where $i, j = 1, \ldots, N$. Similarly, for a vector $c = (c_{-N+1}, c_{-N+2}, \ldots, c_{N-1})$ its associated Toeplitz matrix $T = T^c \in \mathbb{R}^{N \times N}$ has entries $T_{i,j} = c_{j-i}$, where $i, j = 1, \ldots, N$. Now we choose an arbitrary subset $\Omega \subset \{1, \ldots, N\}$ of cardinality $n < N$ and let the partial circulant matrix $S_\Omega = S_\Omega^b \in \mathbb{R}^{n \times N}$ be the submatrix of S consisting of the rows indexed by Ω. The partial Toeplitz matrix $T_\Omega = T_\Omega^c \in \mathbb{R}^{n \times N}$ is defined similarly. In this paper the vectors b and c will always be random vectors with independent Bernoulli ± 1 entries.

Of particular interest is the case $N = nK$ for some $K \in \mathbb{N}$ and $\Omega = \{K, 2K, \ldots, nK\}$. Then the application of S_Ω^b and T_Ω^c corresponds to (periodic or non-periodic) convolution with the sequence b (or c, respectively) followed by a downsampling by a factor of K. This setting was studied numerically in [18] by Tropp et al. (using orthogonal matching pursuit instead of ℓ_1-minimization). Also of interest is the case $\Omega = \{1, 2, \ldots, n\}$ which was investigated in [6], [8] by Bajwa et al., who showed that the restricted isometry constant of T_Ω^c satisfies $\delta_s \leq \delta$ with high probability (w.h.p.) provided $n \geq C_\delta s^2 \log(N/s)$. As a byproduct of the proof of our main result we give an alternative proof that $\delta_s \leq \delta$ holds w.h.p. under the condition $n \geq C\delta^{-2} s^2 \log^2(N)$. However, we strongly believe that this bound is not optimal due to the quite pessimistic quadratic scaling in s. Our main result shows that one can achieve recovery w.h.p. by ℓ_1-minimization, if $n \geq Cs \log^2(N)$.

In the following recovery theorem we use a random partial circulant or Toeplitz matrix A_Ω^b or T_Ω^c in the sense that the entries of the vector b or c are independent Bernoulli ± 1 random variables. Furthermore, the signs of the non-zero entries of the s-sparse vector

x are chosen at random according to a Bernoulli distribution as well. In contrast to previous work [6], [18] Ω is allowed to be an arbitrary subset of $\{1, \ldots, N\}$ of cardinality n.

Theorem II.1. *Let $\Omega \subset \{1, 2, \ldots, N\}$ be an arbitrary (deterministic) set of cardinality n. Let $x \in \mathbb{R}^N$ be s-sparse such that the signs of its non-zero entries are Bernoulli ± 1 random variables. Choose $b \in \mathbb{R}^N$ to be a random vector whose entries are ± 1 Bernoulli variables. Let $y = S_\Omega^b x \in \mathbb{R}^n$. There exists a constant $C > 0$ such that*

$$n \geq C s \log^3(N/\epsilon)$$

implies that with probability at least $1 - \epsilon$ the solution of the ℓ_1-minimization problem (2) coincides with x.

The same statement holds with T_Ω^c in place of S_Ω^b where $c \in \mathbb{R}^{2N-1}$ is a random vector with Bernoulli ± 1 entries.

Ignoring the log-factor the necessary number of samples ensuring recovery by ℓ_1-minimization scales linearly with the sparsity s. The power 3 at the log-term can very likely be improved to 1, and moreover, it seems also possible to remove the randomness assumption on the non-zero coefficients of x. We postpone such improvements as well as an investigation of the restricted isometry constants to possible future contributions. The remainder of the paper is concerned with the proof of Theorem II.1.

III. Proof of Theorem II.1

An essential ingredient of the proof is the following recovery theorem for ℓ_1-minimization due to Fuchs [19] and Tropp [20]. For a matrix A we denote by a_ρ its columns and by A_Λ the submatrix consisting only of the columns index by Λ.

Theorem III.1. *Suppose that $y = Ax$ for some x with $\operatorname{supp} x = \Lambda$. If*

$$|\langle A_\Lambda^\dagger a_\rho, \operatorname{sgn}(x_\Lambda)\rangle| < 1 \quad \textit{for all } \rho \notin \Lambda\,, \tag{3}$$

then x is the unique solution of the Basis Pursuit problem (2). Here, $A_\Lambda^\dagger$ denotes the Moore-Penrose pseudo-inverse of A_Λ.

A crucial step in applying this theorem is to show that the ℓ_2-norm of $A_\Lambda^\dagger a_\rho$ in (3) is small. To this end one expands

$$\|A_\Lambda^\dagger a_\rho\|_2 = \|(A_\Lambda^* A_\Lambda)^{-1} A_\Lambda^* a_\rho\|_2 = \|(A_\Lambda^* A_\Lambda)^{-1}\|_{2\to 2} \|A_\Lambda^* a_\rho\|_2, \tag{4}$$

where $\|\cdot\|_{2\to2}$ denotes the operator norm on ℓ_2. The second term can be estimated in terms of the coherence of A, which is defined to be the largest absolute inner product of different columns of A, $\mu = \max_{\rho\neq\lambda} |\langle a_\rho, a_\lambda\rangle|$. Indeed,

$$\|A_\Lambda^* a_\lambda\|_2 = \left(\sum_{\lambda\in\Lambda} |\langle a_\lambda, a_\rho\rangle|^2\right)^{1/2} \leq \sqrt{|\Lambda|}\mu.$$

The coherence of a random Toeplitz or circulant matrix can be bounded as follows.

Proposition III.2. *Let μ be the coherence of the random partial circulant matrix $\frac{1}{\sqrt{n}} S_\Omega^b \in \mathbb{R}^{n\times N}$ or Toeplitz matrix $\frac{1}{\sqrt{n}} T_\Omega^c \in \mathbb{R}^{n\times N}$ where b and c are Rademacher series and Ω has cardinality n. Then with probability at least $1-\epsilon$ the coherence satisfies*

$$\mu \leq 4\frac{\log(2N^2/\epsilon)}{\sqrt{n}}.$$

The proof is contained in Section V. This proposition easily implies the following (probably non-optimal) estimate of the restricted isometry constants of S_Ω^b or T_Ω^c contained also in [8] with a different proof.

Corollary III.3. *Let $\frac{1}{\sqrt{n}} S_\Omega^b, \frac{1}{\sqrt{n}} T_\Omega^c \in \mathbb{R}^{n\times N}$ be the randomly generated normalized partial circulant and Toeplitz matrix generated from Rademacher series and δ_s be their restricted isometry constant. Assume that*

$$n \geq 16\delta^{-2} s^2 \log^2(2N^2/\epsilon).$$

Then with probability at least $1-\epsilon$ it holds $\delta_s \leq \delta$.

<u>**Proof:**</u> Combine the bound $\delta_s \leq (s-1)\mu$ (which easily follows from Gershgorin's disk theorem) with the estimate above on the coherence of $A = \frac{1}{\sqrt{n}} S_\Omega^b$ or $A = \frac{1}{\sqrt{n}} T_\Omega^c$. ■

As suggested by (4) we also need an estimate of the operator norm of the inverse of $A_\Lambda^* A_\Lambda$. To this end we bound the smallest and largest eigenvalue of this matrix.

Theorem III.4. *Let $\Omega, \Lambda \subset \{1, \ldots, N\}$ with $|\Omega| = n$ and $|\Lambda| = s$. Let $b \in \mathbb{R}^N$ and $c \in \mathbb{R}^{2N-1}$ be Rademacher series. Denote either $A = \frac{1}{\sqrt{n}} S_\Omega^b$ or $A = \frac{1}{\sqrt{n}} T_\Omega^c$. Assume*

$$n \geq \tilde{C}\delta^{-2} s \log^2(4s/\epsilon), \tag{5}$$

where $\tilde{C} = 4\pi^2 \approx 39.48$. Then with probability at least $1-\epsilon$ the minimal and maximal eigenvalues $\lambda_{\min}$ and $\lambda_{\max}$ of $A_\Lambda^* A_\Lambda$ satisfy

$$1 - \delta \leq \lambda_{\min} \leq \lambda_{\max} \leq 1 + \delta.$$

Note that the above theorem holds for a fixed subset Λ and random coefficients b or c. It does not imply that for given b or c the estimate holds uniformly for all subsets Λ, which would be equivalent to having an estimate for the restricted isometry constants of $\frac{1}{\sqrt{n}} S_\Omega^b$ or $\frac{1}{\sqrt{n}} T_\Omega^c$. (Note that taking a union bound over all subsets Λ would yield an estimate essentially worse than Corollary III.3.)

Now we are ready to complete the proof of Theorem II.1 on the basis of Proposition III.2 and Theorem III.4. We proceed similarly as in [21, Theorem 14]. Hoeffding's inequality states that

$$\mathbb{P}\big(|\sum_j \epsilon_j a_j| \geq u\|a\|_2\big) \leq 2e^{-u^2/2}. \tag{6}$$

By our assumption on the random phases $\epsilon_\lambda = \operatorname{sgn}(x_\lambda)$, the scalar product on the left hand side of (3) is precisely of the above form with $a = A_\Lambda^\dagger a_\rho = (A_\Lambda^* A_\Lambda)^{-1} A_\Lambda^* a_\rho$. Theorem III.4 implies that the smallest eigenvalue of $A_\Lambda^* A_\Lambda$ is bounded from below by $1-\delta$ with probability at least $1-\epsilon$ provided condition (5) holds; hence, $\|(A_\Lambda^* A_\Lambda)^{-1}\|_{2\to2} \leq \frac{1}{1-\delta}$. Plugging this into (4) yields

$$\|A_\Lambda^\dagger a_\rho\|_2 \leq \frac{1}{1-\delta}\sqrt{s}\mu. \tag{7}$$

Following Theorem III.1 the probability that recovery fails can be estimated by

$$\begin{aligned}
&\mathbb{P}\big(|\langle A_\Lambda^\dagger a_\rho, R_\Lambda \operatorname{sgn}(x)\rangle| \geq 1 \text{ for some } \rho \notin \Lambda\big)\\
&\leq \mathbb{P}\big(\,|\langle A_\Lambda^\dagger a_\rho, R_\Lambda \operatorname{sgn}(x)\rangle| \geq 1 \text{ for some } \rho \notin \Lambda \;\big|\mu \leq \frac{\alpha}{\sqrt{n}}\\
&\qquad \& \,\lambda_{\min} \geq 1-\delta\,\big) + \mathbb{P}\big(\mu > \frac{\alpha}{\sqrt{n}}\big) + \mathbb{P}\big(\lambda_{\min} < 1-\delta\big)\\
&\leq \sum_{\rho\notin\Lambda} \mathbb{P}\big(\,|\langle A_\Lambda^\dagger a_\rho, R_\Lambda \operatorname{sgn}(x)\rangle| \geq 1 \;\big|\; \mu \leq \frac{\alpha}{\sqrt{n}} \,\&\, \lambda_{\min} \geq 1-\delta\,\big)\\
&\qquad + \mathbb{P}\big(\mu > \frac{\alpha}{\sqrt{n}}\big) + \mathbb{P}\big(\lambda_{\min} < 1-\delta\big).
\end{aligned}$$

Under the assumption $\mu \leq \frac{\alpha}{\sqrt{n}}$ equation (7) implies that for $u =$

$\frac{(1-\delta)\sqrt{n}}{\alpha\sqrt{s}}$ we have $u\|A_\Lambda^\dagger a_\rho\|_2 \le 1$, so (6) gives

$$\mathbb{P}\big(|\langle A_\Lambda^\dagger a_\rho, R_\Lambda \operatorname{sgn}(x)\rangle| \ge 1 \;\big|\; \mu \le \frac{\alpha}{\sqrt{n}} \,\&\, \lambda_{\min} \ge 1-\delta \big) \le 2\exp\left(-\frac{(1-\delta)^2}{2\alpha^2}\frac{n}{s}\right). \quad (8)$$

Setting $\alpha = 4\log(2N^2/\epsilon)$ Theorem III.2 yields

$$\mathbb{P}(\mu \ge \alpha/\sqrt{n}) \le \epsilon.$$

Now we choose $\delta = 1/2$. Under condition (5), which reads

$$n \ge 4\tilde{C} s \log^2(s/\epsilon), \quad (9)$$

we have $\mathbb{P}(\lambda_{\min} \ge 1-\delta) \le \epsilon$. Hence, under the above conditions we obtain

$$\mathbb{P}\big(|\langle A_\Lambda^\dagger a_\rho, R_\Lambda \operatorname{sgn}(x)\rangle| \ge 1 \text{ for some } \rho \notin \Lambda\big) \le 2N\exp\left(-\frac{1}{8\log^2(2N^2/\epsilon)}\frac{n}{s}\right) + 2\epsilon. \quad (10)$$

The first term is less than ϵ provided $n \ge 8s\log^2(2N^2/\epsilon)\log(2N/\epsilon)$, or

$$n \ge C_1 s \log^3(N/\epsilon) \quad (11)$$

for a suitable constant C_1. Conditions (9) and (11) are both satisfied if

$$n \ge C s \log^3(N/\epsilon)$$

for a suitable constant C, in which case the probability that recovery by ℓ_1-minimization is less than 3ϵ. This completes the proof.

IV. Non-commutative Khintchine inequalities

Both the proof of Proposition III.2 as well as the proof of Theorem III.4 are based on versions of the Khintchine inequality. Let us first state the non-commutative Khintchine inequality due to Lust-Piquard [22] and Buchholz [23], see also [21]. To this end we introduce Schatten class norms on matrices. Denoting by $\sigma(A)$ the vector of singular values of a matrix A, the S_p-norm is defined as

$$\|A\|_{S_p} := \|\sigma(A)\|_p,$$

where $\|\cdot\|_p$ is the usual ℓ_p-norm, $1 \le p \le \infty$.

Theorem IV.1. *Let (A_k) be a finite sequence of matrices of the same dimension and let (g_k) be a sequence of independent standard Gaussian random variables. Then for $m \in \mathbb{N}$,*

$$\left[\mathbb{E}\left\|\sum_k g_k A_k\right\|_{S_{2m}}^{2m}\right]^{1/2m} \le B_m \max\left\{\left\|\left(\sum_k A_k A_k^*\right)^{1/2}\right\|_{S_{2m}}, \left\|\left(\sum_k A_k^* A_k\right)^{1/2}\right\|_{S_{2m}}\right\},$$

with optimal constant

$$B_m = \left(\frac{(2m)!}{2^m m!}\right)^{\frac{1}{2m}}$$

Using the contraction principle for Bernoulli random variables, see [24, eq. (4.8)], we obtain the non-commutative Khintchine inequality for Bernoulli random variables [22].

Corollary IV.2. *Let (A_k) be a finite sequence of matrices of the same dimension and let (ϵ_k) be a sequence of independent Bernoulli ± 1 random variables. Then for $m \in \mathbb{N}$,*

$$\left[\mathbb{E}\left\|\sum_k \epsilon_k A_k\right\|_{S_{2m}}^{2m}\right]^{1/2m} \le C_m \max\left\{\left\|\left(\sum_k A_k A_k^*\right)^{1/2}\right\|_{S_{2m}}, \left\|\left(\sum_k A_k^* A_k\right)^{1/2}\right\|_{S_{2m}}\right\}, \quad (12)$$

with constant

$$C_m = \sqrt{\frac{\pi}{2}}\left(\frac{(2m)!}{2^m m!}\right)^{\frac{1}{2m}}.$$

In the scalar case the factor $\sqrt{\pi/2}$ can be removed. However, it is not clear yet whether this is true also in the non-commutative situation.

The following theorem extends the non-commutative Khintchine inequality to a second order chaos variable. Its proof uses decoupling and Corollary IV.2.

Theorem IV.3. *Let $A_{j,k} \in \mathbb{C}^{r\times t}$, $j,k = 1,\ldots,N$, be matrices with $A_{j,j} = 0$, $j = 1,\ldots,N$. Let ϵ_k, $k = 1,\ldots,N$ be independent Bernoulli random variables. Then for $m \in \mathbb{N}$ it holds*

$$\left[\mathbb{E}\left\|\sum_{j,k=1}^N \epsilon_j\epsilon_k A_{j,k}\right\|_{S_{2m}}^{2m}\right]^{1/2m} \le D_m \max\left\{\left\|\left(\sum_{j,k=1}^N A_{j,k}A_{j,k}^*\right)^{1/2}\right\|_{S_{2m}}, \left\|\left(\sum_{j,k=1}^N A_{j,k}^* A_{j,k}\right)^{1/2}\right\|_{S_{2m}}, \|F\|_{S_{2m}}\right\},$$

where F is the block matrix $F = (A_{j,k})_{j,k=1}^N$ and the constant

$$D_m = 2^{1/2m} 2\pi C_m^2 = 2^{1/2m} 2\pi \left(\frac{(2m)!}{2^m m!}\right)^{1/m}.$$

At present it is not clear whether the term $\|F\|_{S_{2m}}$ can be omitted above. At least, there is no a priori inequality between any of the terms in the maximum. The proof of the theorem is based on the following decoupling lemma, see [25, Proposition 1.9] or [26, Theorem 3.1.1].

Lemma IV.4. *Let ξ_j, $j = 1,\ldots,N$, be a sequence of independent random variables with $\mathbb{E}\xi_j = 0$ for all $j = 1,\ldots,N$. Let $A_{j,k}$, $j,k = 1,\ldots,N$, be a double sequence of elements in a Banach space with norm $\|\cdot\|$, where $A_{j,j} = 0$ for all $j = 1,\ldots,N$. Then for $1 \le p < \infty$*

$$\mathbb{E}\left\|\sum_{j,k=1}^N \xi_j\xi_k A_{j,k}\right\|^p \le 4^p \mathbb{E}\left\|\sum_{j,k=1}^N \xi_j\xi_k' A_{j,k}\right\|^p,$$

where ξ' denotes an independent copy of the sequence $\xi = (\xi_j)$.

Proof of Theorem IV.3. We apply Lemma IV.4 followed by the non-commutative Khintchine inequality (12),

$$
\begin{aligned}
E &:= \mathbb{E}\left\|\sum_{j,k=1}^{N} \epsilon_j \epsilon_k A_{j,k}\right\|_{S_{2m}}^{2m} \\
&\leq 4^{2m} \mathbb{E}_\epsilon \mathbb{E}_{\epsilon'} \left\|\sum_{j,k=1}^{N} \epsilon_j \epsilon_k' A_{j,k}\right\|_{S_{2m}}^{2m} \\
&\leq 4^{2m} C_m^{2m} \mathbb{E}_\epsilon \max\left\{\left\|\left(\sum_{k=1}^{N} B_k(\epsilon)^* B_k(\epsilon)\right)^{1/2}\right\|_{S_{2m}}^{2m},\right. \\
&\qquad \left.\left\|\left(\sum_{k=1}^{N} B_k(\epsilon) B_k(\epsilon)^*\right)^{1/2}\right\|_{S_{2m}}^{2m}\right\}, \qquad (13)
\end{aligned}
$$

where $B_k(\epsilon) := \sum_{j=1}^{N} \epsilon_j A_{j,k}$. We define

$$\widehat{A}_{j,k} = (0|\ldots|0|A_{j,k}|0|\ldots|0) \in \mathbb{C}^{r \times tN}$$

where the non-zero block $A_{j,k}$ is the k-th one, and similarly

$$\widetilde{A}_{j,k} = \big(0|\ldots|0|A_{j,k}^*|0|\ldots|0\big)^* \in \mathbb{C}^{rN \times t}.$$

Then clearly

$$
\begin{aligned}
\widehat{A}_{j,k}\widehat{A}_{j',k'}^* &= \begin{cases} 0 & \text{if } k \neq k', \\ A_{j,k}A_{j',k}^* & \text{if } k = k'. \end{cases}, \qquad (14)\\
\widetilde{A}_{j,k}^*\widetilde{A}_{j',k'} &= \begin{cases} 0 & \text{if } k \neq k', \\ A_{j,k}^*A_{j',k} & \text{if } k = k'. \end{cases}
\end{aligned}
$$

The Schatten class norm satisfies $\|A\|_{S_{2m}} = \|(AA^*)^{1/2}\|_{S_{2m}}$. This allows us to verify that

$$
\begin{aligned}
\left\|\sum_{j=1}^{N} \epsilon_j \sum_{k=1}^{N} \widehat{A}_{j,k}\right\|_{S_{2m}} &= \left\|\left(\sum_{j,j'} \epsilon_j \epsilon_{j'} \sum_{k,k'} \widehat{A}_{j,k}\widehat{A}_{j',k'}^*\right)^{1/2}\right\|_{S_{2m}} \\
&= \left\|\left(\sum_{j,j'} \epsilon_j \epsilon_{j'} \sum_{k} A_{j,k}A_{j',k}^*\right)^{1/2}\right\|_{S_{2m}} \\
&= \left\|\left(\sum_{k} B_k(\epsilon) B_k(\epsilon)^*\right)^{1/2}\right\|_{S_{2m}}.
\end{aligned}
$$

Similarly, we also verify that

$$\left\|\left(\sum_{k} B_k(\epsilon)^* B_k(\epsilon)\right)^{1/2}\right\|_{S_{2m}} = \left\|\sum_{j=1}^{N} \epsilon_j \sum_{k=1}^{N} \widetilde{A}_{j,k}\right\|_{S_{2m}}.$$

Plugging the above expressions into (13) we can further estimate

$$
\begin{aligned}
E \leq 4^{2m} C_{2m}^{2m} &\left(\mathbb{E}\left\|\sum_{j=1}^{N} \epsilon_j \sum_{k=1}^{N} \widehat{A}_{j,k}\right\|_{S_{2m}}^{2m}\right. \\
&\left. + \mathbb{E}\left\|\sum_{j=1}^{N} \epsilon_j \sum_{k=1}^{N} \widetilde{A}_{j,k}\right\|_{S_{2m}}^{2m}\right).
\end{aligned}
$$

Using Khintchine's inequality (12) once more we obtain

$$
\begin{aligned}
E_1 &:= \mathbb{E}\left\|\sum_{j=1}^{N} \epsilon_j \sum_{k=1}^{N} \widehat{A}_{j,k}\right\|_{S_{2m}}^{2m} \\
&\leq C_m^{2m} \max\left\{\left\|\left(\sum_j \widetilde{B}_j \widetilde{B}_j^*\right)^{1/2}\right\|_{S_{2m}}^{2m},\right. \\
&\qquad \left.\left\|\left(\sum_j \widetilde{B}_j^* \widetilde{B}_j\right)^{1/2}\right\|_{S_{2m}}^{2m}\right\},
\end{aligned}
$$

where $\widetilde{B}_j = \sum_{k=1}^{N} \widehat{A}_{j,k}$. Using (14) we see that

$$\sum_j \widetilde{B}_j \widetilde{B}_j^* = \sum_{k,j} A_{j,k} A_{j,k}^*.$$

Furthermore, with the block matrix

$$F = \begin{pmatrix} \widetilde{B}_1 \\ \widetilde{B}_2 \\ \vdots \\ \widetilde{B}_N \end{pmatrix} = \begin{pmatrix} A_{1,1} & A_{1,2} & \ldots & A_{1,N} \\ A_{2,1} & A_{2,2} & \ldots & A_{2,N} \\ \vdots & \vdots & \vdots & \vdots \\ A_{N,1} & A_{N,2} & \ldots & A_{N,N} \end{pmatrix}$$

we have

$$\|(\sum_k \widetilde{B}_k^* \widetilde{B}_k)^{1/2}\|_{S_{2m}}^{2m} = \|(F^*F)^{1/2}\|_{S_{2m}}^{2m} = \|F\|_{S_{2m}}^{2m}.$$

Hence,

$$E_1 \leq C_m^{2m} \max\left\{\|(\sum_{j,k=1}^{N} A_{j,k}A_{j,k}^*)^{1/2}\|_{S_{2m}}^{2m}, \|F\|_{S_{2m}}^{2m}\right\}.$$

As $\widetilde{A}_{j,k}$ differs from $\widehat{A}_{j,k}$ only by interchanging $A_{j,k}$ with $A_{j,k}^*$ we obtain similarly

$$
\begin{aligned}
E_2 &:= \mathbb{E}\left\|\sum_{j=1}^{N} \epsilon_j \sum_{k=1}^{N} \widetilde{A}_{j,k}\right\|_{S_{2m}}^{2m} \\
&\leq \max\left\{\left\|\left(\sum_{j,k=1}^{N} A_{j,k}^* A_{j,k}\right)^{1/2}\right\|_{S_{2m}}^{2m}, \|F\|_{S_{2m}}^{2m}\right\}.
\end{aligned}
$$

Finally, we obtain

$$
\begin{aligned}
E &\leq 4^{2m} C_m^{2m} (E_1 + E_2) \\
&\leq 2 \cdot 4^{2m} C_m^{4m} \max\left\{\left\|\left(\sum_{j,k=1}^{N} A_{j,k}^* A_{j,k}\right)^{1/2}\right\|_{S_{2m}}^{2m},\right. \\
&\qquad \left.\left\|\left(\sum_{j,k=1}^{N} A_{j,k} A_{j,k}^*\right)^{1/2}\right\|_{S_{2m}}^{2m}, \|F\|_{S_{2m}}^{2m}\right\}.
\end{aligned}
$$

This concludes the proof. ■

Repeating the above proof for the scalar case (which removes the factor $\pi/2$ in the constant) and applying interpolation (see (16) and (17) below) yields the following (compare also [27, Proposition 2.2]).

Corollary IV.5. *Let $a_{j,k} \in \mathbb{C}$, $j,k = 1,\ldots,N$ be numbers with $a_{j,j} = 0$, $j = 1,\ldots,N$. Let ϵ_k, $k = 1,\ldots,N$ be independent Bernoulli ± 1 random variables. Then for $2 \leq p < \infty$ it holds*

$$\left[\mathbb{E}\left|\sum_{j,k=1}^{N} \epsilon_j \epsilon_k a_{j,k}\right|^p\right]^{1/p} \leq d_p \left(\sum_{j,k=1}^{N} |a_{j,k}|^2\right)^{1/2},$$

where the constant

$$d_p = 4^{1/p}(4/e)p.$$

V. Proof of the coherence estimate

Now we are equipped to provide the proof of Proposition III.2. An inner product of two columns s_i, s_ℓ of the normalized matrix $\frac{1}{\sqrt{n}} S_\Omega^b$ has the form

$$\langle s_i, s_\ell \rangle = \frac{1}{n} \sum_{r \in \Omega} b_{i-r \mod N} b_{\ell - r \mod N} = \frac{1}{n} \sum_{j,k=1}^N b_j b_k a_{j,k}^{i,\ell},$$

where $a_{j,k}^{i,\ell} = 1$ if $(j,k) = (i - r \mod N, \ell - r \mod N)$ for some $r \in \Omega$ and $a_{j,k}^{i,\ell} = 0$ otherwise. Similarly, the inner product of the columns t_i of the normalized matrix $\frac{1}{\sqrt{n}} T_\Omega^c$ can be written as $\langle t_i, t_\ell \rangle = n^{-1} \sum_{j,k=-N+1}^{N-1} c_j c_k \tilde{a}_{j,k}^{i,\ell}$ with $\tilde{a}_{j,k}^{i,\ell} = 1$ if $(j,k) = (i-r, \ell-r) \in \{1,\dots,N\}^2$ for some $r \in \Omega$ and 0 otherwise. Observe that $\sum_{j,k} |a_{j,k}|^2 = \sum_{j,k} |\tilde{a}_{j,k}|^2 = |\Omega| = n$. Now let $b \in \mathbb{R}^N$ and $c \in \mathbb{R}^{2N}$ be Rademacher series. Then Corollary IV.5 yields

$$\begin{aligned} n \left(\mathbb{E}|\langle s_i, s_j\rangle|^p\right)^{1/p} &= \left(\mathbb{E}|\sum_{j,k} b_j b_k a_{j,k}^{i,\ell}|^p\right)^{1/p} \\ &\le 4^{1/p}(4/e)p \left(\sum_{j,k} |a_{j,k}|^2\right)^{1/2} = 4^{1/p}(4/e)p\sqrt{n} \end{aligned}$$

for $p \ge 2$, and the same estimate holds for $\mathbb{E}|\langle t_i, t_j\rangle|^p$. In order to complete the proof we use the following simple and well-known probability estimate, see e.g. [24], [21].

Lemma V.1. *Suppose Z is a positive random variable satisfying $(\mathbb{E}Z^p)^{1/p} \le \alpha \beta^{1/p} p^{1/\gamma}$ for all $p_0 \le p < \infty$ and some $\alpha, \beta, \gamma > 0$. Then for arbitrary $\kappa > 0$,*

$$\mathbb{P}(Z \ge e^\kappa \alpha u) \le \beta e^{-\kappa u^\gamma}$$

for all $u \ge p_0$.

Proof: By Markov's inequality we obtain

$$\mathbb{P}(Z \ge e^\kappa \alpha u) \le \frac{\mathbb{E}Z^p}{(e^\kappa \alpha u)^p} \le \beta \left(\frac{\alpha p^{1/\gamma}}{e^\kappa \alpha u}\right)^p.$$

Choosing $p = u^\gamma$ yields the statement. ∎

Lemma V.1 with the optimal choice $\kappa = 1$ yields

$$\mathbb{P}(n|\langle s_i, s_\ell\rangle| \ge 4\sqrt{n} u) \le 4 e^{-u}$$

for $u \ge 2$. Taking the union bound over all possible pairs of different columns s_i, s_ℓ we obtain

$$\mathbb{P}(\mu \ge 4 n^{-1/2} u) \le 2N^2 e^{-u}.$$

Set the right hand side to ϵ. Then the resulting $u = \log(2N^2/\epsilon) \ge 2$ since we may assume without loss of generality that $N \ge 2$. We obtain

$$\mathbb{P}\left(\mu \ge 4 \frac{\log(2N^2/\epsilon)}{\sqrt{n}}\right) \le \epsilon.$$

The same holds for the coherence of $\frac{1}{\sqrt{n}} T_\Omega^c$.

VI. Proof of Theorem III.4

We introduce the elementary shift operators on $\mathbb{R}^N$, $(S_j x)_\ell = x_{\ell - j \mod N}$, $j = 1, \dots, N$, and

$$(T_j x)_\ell = \begin{cases} x_{\ell-j} & \text{if } 1 \le \ell - j \le N, \\ 0 & \text{otherwise,} \end{cases}$$

for $j = -N+1, \dots, N-1$, $\ell = 1, \dots, N$. Further, denote by $R_\Omega : \mathbb{R}^N \to \mathbb{R}^\Omega$ the operator that restricts a vector to the indices in Ω. Then we can write

$$S_\Omega^b = R_\Omega \sum_{j=1}^N \epsilon_j S_j \quad \text{and} \quad T_\Omega^c = R_\Omega \sum_{j=-N+1}^{N-1} \epsilon_j T_j,$$

where (ϵ_j) is a Rademacher sequence. Denote by A either $\frac{1}{\sqrt{n}} S_\Omega^b$ or $\frac{1}{\sqrt{n}} T_\Omega^c$. We need to prove a bound on the operator norm of $X_\Lambda := A_\Lambda^* A_\Lambda - I_\Lambda$ where I_Λ denotes the identity on $\mathbb{R}^\Lambda$. We introduce $R_\Lambda^* : \mathbb{R}^\Lambda \to \mathbb{R}^N$ to be the extension operator that fills up a vector in $\mathbb{R}^\Lambda$ with zeros outside Λ. Further, we denote by D_j either S_j or T_j. Observe that

$$\begin{aligned} A_\Lambda^* A_\Lambda &= \frac{1}{n} \sum_j \epsilon_j R_\Lambda D_j^* R_\Omega^* \sum_k \epsilon_k R_\Omega D_k R_\Lambda^* \\ &= \frac{1}{n} \sum_{\substack{j,k \\ j \ne k}} \epsilon_j \epsilon_k R_\Lambda D_j^* P_\Omega D_k R_\Lambda^* + \frac{1}{n} R_\Lambda \left(\sum_j D_j^* P_\Omega D_j\right) R_\Lambda^*, \end{aligned}$$

where $P_\Omega = R_\Omega^* R_\Omega$ denotes the projection operator which cancels all components of a vector outside Ω. Here and in the following the sums range either over $\{1,\dots,N\}$ or over $\{-N+1,\dots,N-1\}$ depending on whether we consider circulant or Toeplitz matrices. It is straightforward to check that

$$\sum_j D_j^* P_\Omega D_j = n I_N, \tag{15}$$

where I_N is the identity on $\mathbb{R}^N$. Since $R_\Lambda R_\Lambda^* = I_\Lambda$ we obtain

$$X_\Lambda = \frac{1}{n} \sum_{j \ne k} \epsilon_j \epsilon_i R_\Lambda D_j^* P_\Omega D_k R_\Lambda^* = \frac{1}{n} \sum_{j \ne k} \epsilon_j \epsilon_k A_{j,k}$$

with $A_{j,k} = R_\Lambda D_j^* P_\Omega D_k R_\Lambda^*$. Our goal is to apply Corollary IV.3. To this end we first observe that by (15)

$$\begin{aligned} \sum_j A_{j,k}^* A_{j,\ell} &= R_\Lambda D_k^* P_\Omega \left(\sum_j D_j P_\Lambda D_j^*\right) P_\Omega D_\ell R_\Lambda^* \\ &= s R_\Lambda D_k^* P_\Omega D_\ell R_\Lambda^*. \end{aligned}$$

Using (15) once more this yields

$$\sum_{j,k} A_{j,k}^* A_{j,k} = s R_\Lambda \left(\sum_k D_k^* P_\Omega D_k\right) R_\Lambda^* = sn R_\Lambda R_\Lambda^* = sn I_\Lambda.$$

Since the entries of all matrices $A_{j,k}$ are non-negative we get

$$\begin{aligned} \|(\sum_{j \ne k} A_{j,k}^* A_{j,k})^{1/2}\|_{S_{2m}}^{2m} &= \mathrm{Tr}\left(\sum_{j \ne k} A_{j,k}^* A_{j,k}\right)^m \\ &\le \mathrm{Tr}\left(\sum_{j,k} A_{j,k}^* A_{j,k}\right)^m = \mathrm{Tr}\,(sn I_\Lambda)^m = s^{m+1} n^m, \end{aligned}$$

where Tr denotes the trace. Furthermore, since $A_{j,k}^* = A_{k,j}$ we have $\sum_{j \ne k} A_{j,k}^* A_{j,k} = \sum_{j \ne k} A_{j,k} A_{j,k}^*$. Let F denote the block matrix $F = (\tilde{A}_{j,k})_{j,k}$ where $\tilde{A}_{j,k} = A_{j,k}$ if $j \ne k$ and $\tilde{A}_{j,j} = 0$. Using once again that the entries of all matrices are non-negative we obtain

$$\begin{aligned}&\|F\|_{S_{2m}}^{2m} = \mathrm{Tr}\,[(F^*F)^m]\\&= \mathrm{Tr}\left[\sum_{\substack{j_1,j_2,\dots,j_m\\k_1,k_2,\dots,k_m}} \tilde{A}^*_{j_1,k_1}\tilde{A}_{j_1,k_2}\tilde{A}^*_{j_2,k_2}\tilde{A}_{j_2,k_3}\cdots\tilde{A}^*_{j_m,k_m}\tilde{A}_{j_m,k_1}\right]\\&\leq \mathrm{Tr}\sum_{k_1,\dots,k_m}\left[\sum_{j_1}A^*_{j_1,k_1}A_{j_1,k_2}\cdots\sum_{j_m}A^*_{j_m,k_m}A_{j_m,k_1}\right]\\&= s^m\,\mathrm{Tr}\sum_{k_1,\dots,k_m}[R_\Lambda D^*_{k_1}P_\Omega D_{k_2}R^*_\Lambda R_\Lambda D^*_{k_2}P_\Omega D_{k_3}R^*_\Lambda\cdots\\&\qquad\qquad\cdots R_\Lambda D^*_{k_m}P_\Omega D_{k_1}R^*_\Lambda]\,,\end{aligned}$$

where we applied also (15) once more. Using the cyclicity of the trace and applying (15) another time, together with the fact that $T_k = T^*_{-k}$ and $S_k = S^*_{N-k}$, gives

$$\begin{aligned}\|F\|_{S_{2m}}^{2m} \leq s^m\,\mathrm{Tr}\Bigg[\sum_{k_1}D_{k_1}P_\Lambda D^*_{k_1}P_\Omega\sum_{k_2}D_{k_2}P_\Lambda D^*_{k_2}P_\Omega\cdots\\\cdots\sum_{k_n}D_{k_n}P_\Lambda D^*_{k_n}P_\Omega\Bigg] = s^{2m}\,\mathrm{Tr}[P_\Omega] = ns^{2m}.\end{aligned}$$

Since by assumption (5) $s \leq n$ it follows that

$$\begin{aligned}\|F\|_{S_{2m}}^{2m} &\leq \|(\sum_{j\neq k}A^*_{j,k}A_{j,k})^{1/2}\|_{S_{2m}}^{2m}\\&= \|(\sum_{j\neq k}A_{j,k}A^*_{j,k})^{1/2}\|_{S_{2m}}^{2m} \;\leq\; n^m s^{m+1}.\end{aligned}$$

Using $\|X_\Lambda\| = \|X_\Lambda\|_{S_\infty} \leq \|X_\Lambda\|_{S_p}$ and applying the Khintchine inequality in Theorem IV.3 we obtain for an integer m

$$\begin{aligned}&\mathbb{E}\|X_\Lambda\|^{2m} = \mathbb{E}\|A^*_\Lambda A_\Lambda - I_\Lambda\|^{2m} \leq \mathbb{E}\|A^*_\Lambda A_\Lambda - I_\Lambda\|_{S_{2m}}^{2m}\\&= \frac{1}{n^{2m}}\mathbb{E}\|\sum_{j\neq k}\epsilon_j\epsilon_k A_{j,k}\|_{S_{2m}}^{2m} \;\leq\; 2(2\pi)^{2m}\left(\frac{(2m)!}{2^m m!}\right)^2\frac{s^{m+1}}{n^m}.\end{aligned}$$

Stirling's formula gives

$$\frac{(2m)!}{2^m m!} = \frac{\sqrt{2\pi 2m}(2m/e)^{2m}e^{\lambda_{2m}}}{2^m\sqrt{2\pi m}(m/e)^m e^{\lambda_m}} \leq \sqrt{2}\,(2/e)^m m^m, \qquad (16)$$

where $\frac{1}{12m+1} \leq \lambda_m \leq \frac{1}{12m}$. An application of Hölder's inequality yields for $\theta \in [0,1]$ and an arbitrary random variable Z.

$$\begin{aligned}\mathbb{E}|Z|^{2m+2\theta} &= \mathbb{E}[|Z|^{(1-\theta)2m}|Z|^{\theta(2m+2)}]\\&\leq (\mathbb{E}|Z|^{2m})^{1-\theta}(\mathbb{E}|Z|^{2m+2})^{\theta}. \qquad (17)\end{aligned}$$

Combining our estimates above gives

$$\begin{aligned}&\mathbb{E}\|X_\Lambda\|^{2m+2\theta} \leq (\mathbb{E}\|X_\Lambda\|^{2m})^{1-\theta}(\mathbb{E}\|X_\Lambda\|^{2m+2})^{\theta}\\&\leq 4(2\pi)^{2m+2\theta}(2/e)^{2m+2\theta}m^{2m(1-\theta)}(m+1)^{2\theta(m+1)}\frac{s^{m+\theta+1}}{n^{m+\theta}}\\&\leq 4\left(\frac{4\pi}{e}\right)^{2m+2\theta}(m+\theta)^{2m+2\theta}\frac{s^{m+\theta+1}}{n^{m+\theta}},\end{aligned}$$

where we used the inequality between the geometric and arithmetic mean in the third step. In other words, for $p \geq 2$,

$$(\mathbb{E}\|X_\Lambda\|^p)^{1/p} \leq \frac{2\pi}{e}\sqrt{\frac{s}{n}}(4s)^{1/p}p.$$

An application of Lemma V.1 with the optimal value $\kappa = 1$ yields

$$\mathbb{P}\left(\|X_\Lambda\| \geq 2\pi\sqrt{\frac{s}{n}}u\right) \leq 4se^{-u}$$

for all $u \geq 2$. Setting the right hand side equal ϵ shows that $\|X_\Lambda\| \leq \delta$ with probability at least $1-\epsilon$ provided

$$n \geq (2\pi)^2\delta^{-2}s\log^2(4s/\epsilon).$$

This completes the proof of Theorem III.4.

Acknowledgements

This work was supported by the Hausdorff Center for Mathematics, University of Bonn and the WWTF project SPORTS (MA 07-004).

References

[1] D. L. Donoho, "Compressed sensing," *IEEE Trans. Inform. Theory*, vol. 52, no. 4, pp. 1289–1306, 2006.

[2] R. G. Baraniuk, M. Davenport, R. A. DeVore, and M. Wakin, "A simple proof of the restricted isometry property for random matrices," *Constr. Approx.*, vol. 28, no. 3, pp. 253–263, 2008.

[3] E. J. Candès and T. Tao, "Near optimal signal recovery from random projections: universal encoding strategies?" *IEEE Trans. Inform. Theory*, vol. 52, no. 12, pp. 5406–5425, 2006.

[4] H. Rauhut, "Stability results for random sampling of sparse trigonometric polynomials," *IEEE Trans. Inform. Theory*, vol. 54, no. 12, pp. 5661–5670, 2008.

[5] M. Rudelson and R. Vershynin, "Sparse reconstruction by convex relaxation: Fourier and Gaussian measurements," in *Proc. CISS 2006 (40th Annual Conference on Information Sciences and Systems)*, 2006.

[6] W. Bajwa, J. Haupt, G. Raz, S. Wright, and R. Nowak, "Toeplitz-structured compressed sensing matrices." 2007, IEEE Workshop SSP,.

[7] J. Romberg, "Compressive sensing by random convolution," *preprint*, 2008.

[8] W. U. Bajwa, J. Haupt, G. Raz, and R. Nowak, "Compressed channel sensing," in *Proc. CISS'08*, Princeton, 2008.

[9] S. Boyd and L. Vandenberghe, *Convex Optimization.* Cambridge Univ. Press, 2004.

[10] B. Efron, T. Hastie, I. Johnstone, and R. Tibshirani, "Least angle regression," *Ann. Statist.*, vol. 32, no. 2, pp. 407–499, 2004.

[11] G. Davis, S. Mallat, and M. Avellaneda, "Adaptive greedy approximations," *Constr. Approx.*, vol. 13, no. 1, pp. 57–98, 1997.

[12] E. J. Candès, J., T. Tao, and J. Romberg, "Robust uncertainty principles: exact signal reconstruction from highly incomplete frequency information," *IEEE Trans. Inform. Theory*, vol. 52, no. 2, pp. 489–509, 2006.

[13] D. L. Donoho and J. Tanner, "Counting faces of randomly-projected polytopes when the projection radically lowers dimension," *J. Amer. Math. Soc.*, vol. 22, no. 1, pp. 1–53, 2009.

[14] R. Gribonval and M. Nielsen, "Sparse representations in unions of bases," *IEEE Trans. Inform. Theory*, vol. 49, no. 12, pp. 3320–3325, 2003.

[15] H. Rauhut, "Random sampling of sparse trigonometric polynomials," *Appl. Comput. Harmon. Anal.*, vol. 22, no. 1, pp. 16–42, 2007.

[16] E. J. Candès, "The restricted isometry property and its implications for compressed sensing," *Compte Rendus de l'Academie des Sciences, Paris, Serie I*, vol. 346, pp. 589–592, 2008.

[17] S. Foucart and M. Lai, "Sparsest solutions of underdetermined linear systems via ℓ_q-minimization for $0 < q \leq 1$," *Appl. Comput. Harmon. Anal.*, to appear.

[18] J. Tropp, M. Wakin, M. Duarte, D. Baron, and R. Baraniuk, "Random filters for compressive sampling and reconstruction," *Proc. 2006 IEEE Int. Conf. Acoustics, Speech, and Signal Processing*, vol. 3, pp. 872–875, May 2006.

[19] J. J. Fuchs, "On sparse representations in arbitrary redundant bases," *IEEE Trans. Inform. Theory*, vol. 50, no. 6, pp. 1341–1344, 2004.

[20] J. A. Tropp, "Recovery of short, complex linear combinations via l_1 minimization," *IEEE Trans. Inform. Theory*, vol. 51, no. 4, pp. 1568–1570, 2005.

[21] ——, "On the conditioning of random subdictionaries," *Appl. Comput. Harmon. Anal.*, vol. 25, pp. 1–24, 2008.

[22] F. Lust-Piquard, "Inégalites de Khintchine dans $C_p(1 < p < \infty)$," *C. R. Acad. Sci. Paris S'er. I Math.*, vol. 303, pp. 289–292, 1986.

[23] A. Buchholz, "Operator Khintchine inequality in non-commutative probability," *Math. Ann.*, vol. 319, pp. 1–16, 2001.

[24] M. Ledoux and M. Talagrand, *Probability in Banach Spaces. Isoperimetry and Processes.* Berlin, Heidelberg, NewYork: Springer-Verlag, 1991, vol. 23.

[25] J. Bourgain and L. Tzafriri, “Invertibility of 'large' submatrices with applications to the geometry of Banach spaces and harmonic analysis,” *Israel J. Math.*, vol. 57, no. 2, pp. 137–224, 1987.

[26] V. H. de la Peña and E. Giné, *Decoupling, From dependence to independence*, ser. Probability and its Applications (New York). New York: Springer-Verlag, 1999.

[27] T. R. McConnell and M. S. Taqqu, “Decoupling inequalities for multilinear forms in independent symmetric random variables,” *Annals Prob.*, vol. 11, no. 3, pp. 943–951, 1986.