# THE RANDOM PAVING PROPERTY FOR
# UNIFORMLY BOUNDED MATRICES

JOEL A. TROPP

ABSTRACT. This note presents a new proof of an important result due to Bourgain and Tzafriri that provides a partial solution to the Kadison–Singer problem. The result shows that every unit-norm matrix whose entries are relatively small in comparison with its dimension can be paved by a partition of constant size. That is, the coordinates can be partitioned into a constant number of blocks so that the restriction of the matrix to each block of coordinates has norm less than one half. The original proof of Bourgain and Tzafriri involves a long, delicate calculation. The new proof relies on the systematic use of symmetrization and (noncommutative) Khintchine inequalities to estimate the norms of some random matrices.

## 1. INTRODUCTION

This note presents a new proof of a result about the paving problem for matrices. Suppose that $\boldsymbol{A}$ is an $n \times n$ matrix. We say that $\boldsymbol{A}$ has an $(m, \varepsilon)$-*paving* if there exists a partition of the set $\{1, 2, \ldots, n\}$ into $m$ blocks $\{\sigma_1, \sigma_2, \ldots, \sigma_m\}$ so that

$$\left\| \sum\nolimits_{j=1}^{m} \boldsymbol{P}_{\sigma_j} \boldsymbol{A} \boldsymbol{P}_{\sigma_j} \right\| \leq \varepsilon \left\| \boldsymbol{A} \right\|$$

where $\boldsymbol{P}_{\sigma_j}$ denotes the diagonal projector onto the coordinates listed in $\sigma_j$. Since every projector in this note is diagonal, we omit the qualification from here onward. As usual, $\|\cdot\|$ is the norm on linear operators mapping $\ell_2^n$ to itself.

The fundamental question concerns the paving of matrices with a zero diagonal (i.e., *hollow matrices*).

**Problem 1** (Paving Problem)**.** *Fix $\varepsilon \in (0, 1)$. Is there a constant $m = m(\varepsilon)$ so that, for sufficiently large $n$, every hollow $n \times n$ matrix has an $(m, \varepsilon)$-paving?*

Anderson [And79] has shown that the Paving Problem is equivalent with the Kadison–Singer problem, a major open question in operator theory. It is closely related to significant problems in harmonic analysis and other areas of mathematics and engineering. See [CT06] for an intriguing discussion.

At present, the strongest results on the paving problem are due to Bourgain and Tzafiri [BT91]. For a fixed $\varepsilon$, they established that

(1) every hollow matrix of size $n \times n$ can be paved with at most $m = \mathrm{O}(\log n)$ blocks and

(2) every square matrix whose entries are relatively small compared with its dimension can be paved with a constant number of blocks.

Let us present a precise statement of their second result. We use the notation $[\![n]\!] \overset{\text{def}}{=} \{1, 2, \ldots, n\}$.

**Theorem 2** (Bourgain–Tzafriri). *Fix $\gamma > 0$ and $\varepsilon \in (0, 1)$. There exists a positive integer $m = m(\gamma, \varepsilon)$ so that, for all $n \geq N(\gamma, \varepsilon)$, the following statement holds. Suppose that $\boldsymbol{A}$ is an $n \times n$ unit-norm matrix with uniformly bounded entries:*

$$|a_{jk}| \leq \frac{1}{(\log n)^{1+\gamma}} \qquad \text{for } j, k = 1, 2, \ldots, n.$$

*Then there is a partition of the set $[\![n]\!]$ into $m$ blocks $\{\sigma_1, \sigma_2, \ldots, \sigma_m\}$ such that*

$$\left\| \sum\nolimits_{j=1}^{m} \boldsymbol{P}_{\sigma_j} \boldsymbol{A} \boldsymbol{P}_{\sigma_j} \right\| \leq \varepsilon$$

*where $\boldsymbol{P}_{\sigma_j}$ is the projector onto the coordinates listed in $\sigma_j$. The number $m$ satisfies the bound*

$$m \leq \mathrm{C}\varepsilon^{-\mathrm{C}/\min\{1,\gamma\}}$$

*where $\mathrm{C}$ is a positive universal constant.*

The proof of this result published in [BT91] hinges on a long and delicate calculation of the supremum of a random process. This computation involves a difficult metric entropy bound based on some subtle iteration arguments.

This note shows that the central step in the known proof can be replaced by another approach based on symmetrization and noncommutative Khintchine inequalities. This method for studying random matrices is adapted from Rudelson's article [Rud99]. Even though it is simple and elegant, it leads to sharp bounds in many cases. By itself, Rudelson's technique is not strong enough, so we must also also invoke a method from Bourgain and Tzafrari's proof to complete the argument. As we go along, we indicate the provenance of various parts of the argument.

## 2. Problem Simplifications

Let us begin with some problem simplifications. The reductions in this section were all proposed by Bourgain and Tzafriri; we provide proofs for completeness.

The overall strategy is to construct the paving with probabilistic tools. The first proposition shows that we can leverage a moment estimate for the norm of a random submatrix to build a paving. The idea is to permute the coordinates randomly and divide them into blocks. The moment bound shows that, if we restrict the matrix to the coordinates in a random block, then it has small spectral norm.

**Proposition 3** (Random Paving Principle). *Fix an integer $m$, and let $n = km$ for an integer $k$. Let $\boldsymbol{A}$ be an $n \times n$ unit-norm matrix, and suppose that $\boldsymbol{P}$ is a projector onto exactly $k$ coordinates, chosen uniformly at random from the set $[\![n]\!]$. If, for $p \geq \log n$, we have the estimate*

$$(\mathbb{E}\|\boldsymbol{P}\boldsymbol{A}\boldsymbol{P}\|^p)^{1/p} \leq \varepsilon,$$

*then there exists a partition of the set $[\![n]\!]$ into $m$ blocks $\{\sigma_1, \sigma_2, \ldots, \sigma_m\}$, each of size $k$, such that*

$$\left\| \sum\nolimits_{j=1}^{m} \boldsymbol{P}_{\sigma_j} \boldsymbol{A} \boldsymbol{P}_{\sigma_j} \right\| \leq 3\varepsilon$$

*where $\boldsymbol{P}_{\sigma_j}$ is the projector onto the coordinates listed in $\sigma_j$.*

*Proof.* Consider a random permutation $\pi$ of the set $[\![n]\!]$. For $j = 1, 2, \ldots, m$, define

$$\sigma_j(\pi) = \{\pi(jk - k + 1), \pi(jk - k + 2), \ldots, \pi(jk)\}.$$

For each $j$, the projector $\boldsymbol{P}_{\sigma_j(\pi)}$ onto the coordinates in $\sigma_j(\pi)$ is a restriction to $k$ coordinates, chosen uniformly at random. The hypothesis implies that

$$\mathbb{E}\max_{j=1,2,\ldots,m} \left\| \boldsymbol{P}_{\sigma_j(\pi)} \boldsymbol{A} \boldsymbol{P}_{\sigma_j(\pi)} \right\|^p \leq m\varepsilon^p.$$

There must exist a permutation $\pi_0$ for which the left-hand side is smaller than its expectation. For the partition with blocks $\sigma_j = \sigma_j(\pi_0)$, we have

$$\left\| \sum\nolimits_{j=1}^{m} \boldsymbol{P}_{\sigma_j} \boldsymbol{A} \boldsymbol{P}_{\sigma_j} \right\| = \max_j \left\| \boldsymbol{P}_{\sigma_j} \boldsymbol{A} \boldsymbol{P}_{\sigma_j} \right\| \le m^{1/p} \varepsilon.$$

The equality holds because the coordinate blocks are disjoint. Finally, we have $m^{1/p} \le \mathrm{e}$ because $m \le n$ and $p \ge \log n$. □

This proposition shows that it is sufficient to study the restriction to a random set of coordinates of size $k$. Although this dependent coordinate model is conceptually simple, it would lead to severe inconveniences later in the proof. We prefer instead to study an independent coordinate model for the projector where the *expected number* of coordinates equals $k$. Fortunately, the two models are equivalent for our purposes.

**Proposition 4** (Random Coordinate Models). *Fix an integer $m$, and let $n = km$ for an integer $k$. Let $\boldsymbol{A}$ be an $n \times n$ matrix. Suppose that $\boldsymbol{P}$ is a projector onto $k$ coordinates, chosen uniformly at random from $[\![n]\!]$, and suppose that $\boldsymbol{R}$ is a projector onto a random set of coordinates from $[\![n]\!]$, where each coordinate appears independently with probability $k/n$. For $p > 0$, it holds that*

$$\left( \mathbb{E} \, \| \boldsymbol{P} \boldsymbol{A} \boldsymbol{P} \|^p \right)^{1/p} \le \left( 2 \, \mathbb{E} \, \| \boldsymbol{R} \boldsymbol{A} \boldsymbol{R} \|^p \right)^{1/p}.$$

The reduction to the independent coordinate model also appears in Bourgain and Tzafriri's paper with a different proof. The following attractive argument is drawn from [CR06, Sec. 3].

*Proof.* For a coordinate projector $\boldsymbol{R}$, denote by $\sigma(\boldsymbol{R})$ the set of coordinates onto which it projects. We can make the following computation:

$$\mathbb{P} \left\{ \| \boldsymbol{R} \boldsymbol{A} \boldsymbol{R} \|^p > t \right\} \ge \sum\nolimits_{j=k}^{n} \mathbb{P} \left\{ \| \boldsymbol{R} \boldsymbol{A} \boldsymbol{R} \|^p > t \mid \#\sigma(\boldsymbol{R}) = j \right\} \cdot \mathbb{P} \left\{ \#\sigma(\boldsymbol{R}) = j \right\}$$

$$\ge \mathbb{P} \left\{ \| \boldsymbol{R} \boldsymbol{A} \boldsymbol{R} \|^p > t \mid \#\sigma(\boldsymbol{R}) = k \right\} \cdot \sum\nolimits_{j=k}^{n} \mathbb{P} \left\{ \#\sigma(\boldsymbol{R}) = j \right\}$$

$$\ge \frac{1}{2} \mathbb{P} \left\{ \| \boldsymbol{P} \boldsymbol{A} \boldsymbol{P} \|^p > t \right\}.$$

The second inequality holds because the spectral norm of a submatrix is smaller than the spectral norm of the matrix. The third inequality relies on the fact [JS68, Thm. 3.2] that the medians of the binomial distribution BINOMIAL$(k/n, n)$ lie between $k - 1$ and $k$. Integrate with respect to $t$ to complete the argument. □

## 3. The Main Result

On account of these simplifications, it suffices to prove the following theorem. In the sequel, $\boldsymbol{R}_\delta$ denotes a square, diagonal matrix whose diagonal entries are independent and identically distributed 0–1 random variables with common mean $\delta$. The dimensions of $\boldsymbol{R}_\delta$ conform to its context.

**Theorem 5.** *Fix $\gamma > 0$ and $\varepsilon \in (0, 1)$. There exists a positive integer $m = m(\gamma, \varepsilon)$ so that, for all $n \ge N(\gamma, \varepsilon)$, the following statement holds. Suppose that $\boldsymbol{A}$ is an $n \times n$ unit-norm matrix with uniformly bounded entries:*

$$|a_{jk}| \le \frac{1}{(\log n)^{1+\gamma}} \qquad \text{for } j, k = 1, 2, \ldots, n.$$

*Let $\delta = 1/m$. For $p = 2 \cdot \lceil \log n \rceil$, we have*

$$\left( \mathbb{E} \, \| \boldsymbol{R}_\delta \boldsymbol{A} \boldsymbol{R}_\delta \|^p \right)^{1/p} \le \varepsilon. \tag{3.1}$$

*The number $m$ satisfies the bound*

$$m \le (0.01\varepsilon)^{-2(1+\gamma)/\gamma}.$$

An example of Bourgain and Tzafriri shows that the number $\gamma$ cannot be removed from the bound $(\log n)^{-(1+\gamma)}$ on the matrix entries [BT91, Ex. 2.2]. Fix $\delta \in (0,1)$. For each $n \geq N(\delta)$, they exhibit an $n \times n$ matrix $\boldsymbol{A}$ with unit norm and bounded entries:

$$|a_{jk}| \leq \frac{2\log(1/\delta)}{\log n}.$$

For this matrix, $\mathbb{E}\,\|\boldsymbol{R}_\delta \boldsymbol{A} \boldsymbol{R}_\delta\| \geq 1/2$. In particular, it has no constant-size random paving when $\varepsilon$ is small.

*Proof of Theorem 2 from Theorem 5.* Fix $\gamma$ and $\varepsilon$. Let $m$ be the integer guaranteed by Theorem 5, and assume that $n$ is sufficiently large. Suppose we are given an $n \times n$ matrix with unit norm and uniformly bounded entries. If necessary, augment the matrix with zero rows and columns so that its dimension is a multiple of $m$.

Apply Proposition 4 to transfer the estimate (3.1) to the dependent coordinate model. The Random Paving Principle shows that the augmented matrix has an $(m, 6\varepsilon)$-paving. Discard the zero rows and columns to complete the proof of Theorem 2. $\qquad\square$

## 4. Proof of Theorem 5

In this section, we establish Theorem 5. The proofs of the supporting results are postponed to the subsequent sections.

Fix $\gamma > 0$ and $\varepsilon \in (0,1)$. We assume for convenience that $n \geq 8$, and we suppose that $\boldsymbol{A}$ is an $n \times n$ matrix with unit norm and uniformly bounded entries:

$$|a_{jk}| \leq \frac{1}{(\log n)^{1+\gamma}} \stackrel{\text{def}}{=} \mu.$$

In the sequel, the symbol $\mu$ always abbreviates the uniform bound. Finally, set $p = 2 \cdot \lceil \log n \rceil$.

The major task in the proof is to obtain an estimate for the quantity

$$E(\varrho) \stackrel{\text{def}}{=} \left(\mathbb{E}\,\|\boldsymbol{R}_\varrho \boldsymbol{A} \boldsymbol{R}_\varrho\|^p\right)^{1/p},$$

where $\varrho$ is not too small. This estimate is accomplished with decoupling, symmetrization, and noncommutative Khintchine inequalities. This approach is adapted from work of Rudelson [Rud99] and Rudelson–Vershynin [RV07]. Given this estimate for $E(\varrho)$, we extrapolate the value of $E(m^{-1})$ for a large constant $m = m(\gamma, \varepsilon)$. This step relies on an elegant method due to Bourgain and Tzafriri.

Before continuing, we instate a few more pieces of notation. The symbol $\|\cdot\|_{1,2}$ denotes the norm of an operator mapping $\ell_1^n$ to $\ell_2^n$. For a matrix $\boldsymbol{X}$ expressed in the standard basis, $\|\boldsymbol{X}\|_{1,2}$ is the maximum $\ell_2^n$ norm achieved by a column of $\boldsymbol{X}$. The norm $\|\boldsymbol{X}\|_{\max}$ calculates the maximum absolute value of an entry of $\boldsymbol{X}$.

4.1. **Step 1: Decoupling.** As in Bourgain and Tzafriri's work, the first step involves a classical decoupling argument. First, we must remove the diagonal of the matrix. Since the entries of $\boldsymbol{A}$ do not exceed $\mu$, it follows that $\|\operatorname{diag} \boldsymbol{A}\| \leq \mu$. Define

$$\boldsymbol{B} = \frac{1}{1+\mu}(\boldsymbol{A} - \operatorname{diag} \boldsymbol{A}).$$

Note that $\boldsymbol{B}$ has a zero diagonal and that $\|\boldsymbol{B}\| \leq 1$. Furthermore,

$$|b_{jk}| < \mu \qquad \text{for } j, k = 1, 2, \ldots, n.$$

With this definition,

$$E(\varrho) \leq \|\operatorname{diag} \boldsymbol{A}\| + (1 + \mu)\left(\mathbb{E}\,\|\boldsymbol{R}_\varrho \boldsymbol{B} \boldsymbol{R}_\varrho\|^p\right)^{1/p}.$$

The expectation on the right-hand side cannot exceed one, so we have

$$E(\varrho) \leq 2\mu + \left(\mathbb{E}\,\|\boldsymbol{R}_\varrho \boldsymbol{B} \boldsymbol{R}_\varrho\|^p\right)^{1/p}.$$

Now, we may replace the projector $\boldsymbol{R}_\varrho$ by a pair of independent projectors by invoking the following result.

**Proposition 6.** *Let $\boldsymbol{B}$ be a square matrix with a zero diagonal, and let $p \geq 1$. Then*

$$(\mathbb{E} \|\boldsymbol{R}_\varrho \boldsymbol{B} \boldsymbol{R}_\varrho\|^p)^{1/p} \leq 20 \left( \mathbb{E} \|\boldsymbol{R}_\varrho \boldsymbol{B} \boldsymbol{R}'_\varrho\|^p \right)^{1/p}$$

*where the two random projectors on the right-hand side are independent.*

See [BT87, Prop. 1.1] or [LT91, Sec. 4.4] for the simple proof.

We apply Proposition 6 to reach

$$E(\varrho) \leq 2\mu + 20 \left( \mathbb{E} \|\boldsymbol{R}_\varrho \boldsymbol{B} \boldsymbol{R}'_\varrho\|^p \right)^{1/p}. \tag{4.1}$$

4.2. **Step 2: Norm of a Random Restriction.** The next step of the proof is to develop a bound on the spectral norm of a matrix that has been restricted to a random subset of its columns. The following result is due to Rudelson and Vershynin [RV07], with some inessential modifications by the author.

**Proposition 7** (Rudelson–Vershynin). *Let $\boldsymbol{X}$ be an $n \times n$ matrix, and suppose that $p \geq 2 \log n \geq 2$. Then*

$$(\mathbb{E} \|\boldsymbol{X} \boldsymbol{R}_\varrho\|^p)^{1/p} \leq 3\sqrt{p} \left( \mathbb{E} \|\boldsymbol{X} \boldsymbol{R}_\varrho\|_{1,2}^p \right)^{1/p} + \sqrt{\varrho} \|\boldsymbol{X}\|.$$

The proof of Proposition 7 depends on a lemma of Rudelson that bounds the norm of a Rademacher sum of rank-one, self-adjoint matrices [Rud99]. This lemma, in turn, hinges on the noncommutative Khintchine inequality [LP86, Buc01]. See Section 5 for the details.

To account for the influence of $\boldsymbol{R}'_\varrho$, we apply Proposition 7 with $\boldsymbol{X} = \boldsymbol{R}_\varrho \boldsymbol{B}$. Inequality (4.1) becomes

$$E(\varrho) \leq 2\mu + 60\sqrt{p} \left( \mathbb{E} \|\boldsymbol{R}_\varrho \boldsymbol{B} \boldsymbol{R}'_\varrho\|_{1,2}^p \right)^{1/p} + 20\sqrt{\varrho} \left( \mathbb{E} \|\boldsymbol{R}_\varrho \boldsymbol{B}\|^p \right)^{1/p}.$$

We invoke Proposition 7 again with $\boldsymbol{X} = \boldsymbol{B}^*$ to reach

$$E(\varrho) \leq 2\mu + 60\sqrt{p} \left( \mathbb{E} \|\boldsymbol{R}_\varrho \boldsymbol{B} \boldsymbol{R}'_\varrho\|_{1,2}^p \right)^{1/p} + 60\sqrt{\varrho p} \left( \mathbb{E} \|\boldsymbol{B}^* \boldsymbol{R}_\varrho\|_{1,2}^p \right)^{1/p} + 20\varrho \|\boldsymbol{B}^*\|.$$

Discard the projector $\boldsymbol{R}'_\varrho$ from the first expectation by means of the observation

$$\|\boldsymbol{R}_\varrho \boldsymbol{B} \boldsymbol{R}'_\varrho\|_{1,2} \leq \|\boldsymbol{R}_\varrho \boldsymbol{B}\|_{1,2}.$$

In words, the maximum column norm of a matrix exceeds the maximum column norm of any submatrix. We also have the bound

$$\|\boldsymbol{B}^* \boldsymbol{R}_\varrho\|_{1,2} \leq \|\boldsymbol{B}^*\|_{1,2} \leq \|\boldsymbol{B}^*\| \leq 1$$

because the spectral norm dominates the maximum $\ell_2^n$ norm of a column. The inequality $\varrho \leq \sqrt{\varrho}$ yields

$$E(\varrho) \leq 2\mu + 60\sqrt{p} \left( \mathbb{E} \|\boldsymbol{R}_\varrho \boldsymbol{B}\|_{1,2}^p \right)^{1/p} + 80\sqrt{\varrho p}. \tag{4.2}$$

4.3. **Step 3: Estimate of Maximum Column Norm.** To complete our estimate of $E(\varrho)$, we must bound the remaining expectation. The following result does the job.

**Proposition 8.** *Let $\boldsymbol{X}$ be an $n \times n$ matrix, and suppose that $p \geq 2 \log n \geq 4$. Then*

$$\left( \mathbb{E} \|\boldsymbol{R}_\varrho \boldsymbol{X}\|_{1,2}^p \right)^{1/p} \leq 3\sqrt{p} \|\boldsymbol{X}\|_{\max} + \sqrt{\varrho} \|\boldsymbol{X}\|_{1,2}.$$

The proof of Proposition 8 uses only classical methods, including symmetrization and scalar Khintchine inequalities. A related bound appears inside the proof of Proposition 2.5 in [BT91]. Turn to Section 6 for the argument.

Apply Proposition 8 to the remaining expectation in (4.2) to find that

$$E(\varrho) \leq 2\mu + 180p \, \|\boldsymbol{B}\|_{\max} + 60\sqrt{\varrho p} \, \|\boldsymbol{B}\|_{1,2} + 80\sqrt{\varrho p}.$$

As above, the maximum column norm $\|\boldsymbol{B}\|_{1,2} \leq 1$. The entries of $\boldsymbol{B}$ are uniformly bounded by $\mu$. Recall $p = 2 \cdot \lceil \log n \rceil$ to conclude that

$$E(\varrho) \leq 550\mu \log n + 250\sqrt{\varrho \log n}, \tag{4.3}$$

taking into account $\lceil \log n \rceil \leq 1.5 \log n$ whenever $n \geq 8$.

The result in (4.3) is not quite strong enough to establish Theorem 5. In the theorem, the relation between the size $m$ of the paving and the proportion $\delta$ of columns is $\delta = 1/m$. The parameter $\varrho$ also represents the proportion of columns selected. Unfortunately, when we set $\varrho = 1/m$, we find that the bound in (4.3) is trivial unless $\varrho$ is smaller than $c/\log n$, which suggests that $m$ grows logarithmically with $n$. To prove the result, however, we must obtain a bound for $m$ that is independent of dimension.

4.4. **Step 4: Extrapolation.** To finish the argument, we require a remarkable fact uncovered by Bourgain and Tzafriri in their work. Roughly speaking, the value of $E(\varrho)^p$ is comparable with a polynomial of low degree. It is possible to use the inequality (4.3) to estimate the coefficients of this polynomial. We can then extrapolate to obtain a nontrivial estimate of $E(\delta)^p$, where $\delta$ is a small constant.

**Proposition 9** (Bourgain–Tzafriri). *Let $\boldsymbol{X}$ be an $n \times n$ matrix with $\|\boldsymbol{X}\| \leq 1$. Suppose that $p$ is an even integer with $p \geq 2\log n$. Choose parameters $\delta \in (0,1)$ and $\varrho \in (0, 0.5)$. For each $\lambda \in (0,1)$, it holds that*

$$(\mathbb{E} \, \|\boldsymbol{R}_\delta \boldsymbol{X} \boldsymbol{R}_\delta\|^p)^{1/p} \leq 60 \left[ \delta^\lambda + \varrho^{-\lambda} (\mathbb{E} \, \|\boldsymbol{R}_\varrho \boldsymbol{X} \boldsymbol{R}_\varrho\|^p)^{1/p} \right].$$

The proof depends essentially on a result of V. A. Markov that bounds the coefficients of a polynomial in terms of its maximum value. See Section 7 for the details.

Recall now that

$$\mu \leq \frac{1}{(\log n)^{1+\gamma}}.$$

Set the proportion $\varrho = (\log n)^{-1-2\gamma}$, and introduce these quantities into (4.3) to obtain

$$(\mathbb{E} \, \|\boldsymbol{R}_\varrho \boldsymbol{A} \boldsymbol{R}_\varrho\|^p)^{1/p} \leq 800(\log n)^{-\gamma}.$$

Proposition 9 shows that

$$(\mathbb{E} \, \|\boldsymbol{R}_\delta \boldsymbol{A} \boldsymbol{R}_\delta\|^p)^{1/p} \leq 60\delta^\lambda + 48000(\log n)^{\lambda(1+2\gamma)-\gamma}$$

for every value of $\lambda$ in $(0,1)$. Make the selection $\lambda = \gamma/(2+2\gamma)$. Since the exponent on the logarithm is strictly negative, it follows for sufficiently large $n$ that

$$(\mathbb{E} \, \|\boldsymbol{R}_\delta \boldsymbol{A} \boldsymbol{R}_\delta\|^p)^{1/p} \leq 100 \, \delta^{\gamma/(2+2\gamma)}.$$

To make the right-hand side less than a parameter $\varepsilon$, it suffices that $\delta \leq (0.01\varepsilon)^{2(1+\gamma)/\gamma}$. Therefore, any value

$$m \geq (0.01\varepsilon)^{-2(1+\gamma)/\gamma}$$

is enough to establish Theorem 5.

## 5. Proof of Random Restriction Estimate

In this section, we establish Proposition 7. The difficult part of the estimation is performed with the noncommutative Khintchine inequality. This result was originally discovered by Lust-Picquard [LP86]. We require a sharp version due to Buchholz [Buc01] that provides the optimal order of growth in the constant.

Before continuing, we state a few definitions. Given a matrix $\boldsymbol{X}$, let $\sigma(\boldsymbol{X})$ denote the vector of its singular values, listed in weakly decreasing order. The *Schatten p-norm* $\|\cdot\|_{S_p}$ is calculated as

$$\|\boldsymbol{X}\|_{S_p} = \|\boldsymbol{\sigma}\|_p$$

where $\|\cdot\|_p$ denotes the $\ell_p$ vector norm.

A *Rademacher random variable* takes the two values $\pm 1$ with equal probability. A *Rademacher sequence* is a sequence of independent Rademacher variables.

**Proposition 10** (Noncommutative Khintchine Inequality). *Let $\{\boldsymbol{X}_j\}$ be a finite sequence of matrices of the same dimension, and let $\{\varepsilon_j\}$ be a Rademacher sequence. For each $p \geq 2$,*

$$\left[\mathbb{E}\left\|\sum_j \varepsilon_j \boldsymbol{X}_j\right\|_{S_p}^p\right]^{1/p} \leq \mathrm{C}_p \max\left\{\left\|\left(\sum_j \boldsymbol{X}_j \boldsymbol{X}_j^*\right)^{1/2}\right\|_{S_p}, \left\|\left(\sum_j \boldsymbol{X}_j^* \boldsymbol{X}_j\right)^{1/2}\right\|_{S_p}\right\}, \qquad (5.1)$$

*where $\mathrm{C}_p \leq 2^{-0.25}\sqrt{\pi/\mathrm{e}}\,\sqrt{p}$.*

This proposition is a corollary of Theorem 5 of [Buc01]. In this work, Buchholz shows that the noncommutative Khintchine inequality holds with a Gaussian sequence in place of the Rademacher sequence. He computes the optimal constant when $p$ is an even integer:

$$\mathrm{C}_{2n} = \left(\frac{(2n)!}{2^n n!}\right)^{1/2n}.$$

One extends this result to other values of $p$ using Stirling's approximation and an interpolation argument. The inequality for Rademacher variables follows from the contraction principle.

In an important paper [Rud99], Rudelson showed how to use the noncommutative Khintchine inequality to study the moments of a Rademacher sum of rank-one matrices.

**Lemma 11** (Rudelson). *Suppose that $\boldsymbol{x}_1, \boldsymbol{x}_2, \ldots, \boldsymbol{x}_n$ are the columns of a matrix $\boldsymbol{X}$. For any $p \geq 2 \log n$, it holds that*

$$\left(\mathbb{E}\left\|\sum_{j=1}^n \varepsilon_j \boldsymbol{x}_j \boldsymbol{x}_j^*\right\|^p\right)^{1/p} \leq 1.5\sqrt{p}\,\|\boldsymbol{X}\|_{1,2}\,\|\boldsymbol{X}\|,$$

*where $\{\varepsilon_j\}$ is a Rademacher sequence.*

*Proof.* First, bound the spectral norm by the Schatten $p$-norm.

$$E \stackrel{\text{def}}{=} \left(\mathbb{E}\left\|\sum_{j=1}^n \varepsilon_j \boldsymbol{x}_j \boldsymbol{x}_j^*\right\|^p\right)^{1/p} \leq \left(\mathbb{E}\left\|\sum_{j=1}^n \varepsilon_j \boldsymbol{x}_j \boldsymbol{x}_j^*\right\|_{S_p}^p\right)^{1/p}.$$

Apply the noncommutative Khintchine inequality to obtain

$$E \leq \mathrm{C}_p \left\|\left(\sum_{j=1}^n \|\boldsymbol{x}_j\|_2^2\, \boldsymbol{x}_j \boldsymbol{x}_j^*\right)^{1/2}\right\|_{S_p}.$$

The rank of matrix inside the norm does not exceed $n$, so we can bound the Schatten $p$-norm by the spectral norm if we pay a factor of $n^{1/p}$, which does not exceed $\sqrt{\mathrm{e}}$. Afterward, pull the square root out of the norm to find

$$E \leq \mathrm{C}_p\sqrt{\mathrm{e}}\left\|\sum_{j=1}^n \|\boldsymbol{x}_j\|_2^2\, \boldsymbol{x}_j \boldsymbol{x}_j^*\right\|^{1/2}.$$

The summands are positive semidefinite, so the spectral norm of the sum increases monotonically with each scalar coefficient. Therefore, we may replace each coefficient by $\max_j \|\boldsymbol{x}_j\|_2^2$ and use the homogeneity of the norm to obtain

$$E \le \mathrm{C}_p \sqrt{\mathrm{e}} \max_j \|\boldsymbol{x}_j\|_2 \left\| \sum\nolimits_{j=1}^n \boldsymbol{x}_j \boldsymbol{x}_j^* \right\|^{1/2}.$$

The maximum can be rewritten as $\|\boldsymbol{X}\|_{1,2}$, and the spectral norm can be expressed as

$$\left\| \sum\nolimits_{j=1}^n \boldsymbol{x}_j \boldsymbol{x}_j^* \right\|^{1/2} = \|\boldsymbol{X}\boldsymbol{X}^*\|^{1/2} = \|\boldsymbol{X}\|.$$

Recall that $\mathrm{C}_p \le 2^{-0.25} \sqrt{\pi/\mathrm{e}} \sqrt{p}$ to complete the proof. $\qquad\square$

Recently, Rudelson and Vershynin showed how Lemma 11 implies a bound on the moments of the norm of a matrix that is compressed to a random subset of columns [RV07].

**Proposition 12** (Rudelson–Vershynin)**.** *Let* $\boldsymbol{X}$ *be a matrix with* $n$ *columns, and suppose that* $p \ge 2 \log n \ge 2$*. It holds that*

$$(\mathbb{E} \|\boldsymbol{X}\boldsymbol{R}_\varrho\|^p)^{1/p} \le 3\sqrt{p} \left( \mathbb{E} \|\boldsymbol{X}\boldsymbol{R}_\varrho\|_{1,2}^p \right)^{1/p} + \sqrt{\varrho} \, \|\boldsymbol{X}\|.$$

In words, a random compression of a matrix gets its share of the spectral norm plus another component that depends on the total number of columns and on the $\ell_2^n$ norms of the columns.

*Proof.* Let us begin with an overview of the proof. First, we express the random compression as a random sum. Then we symmetrize the sum and apply Rudelson's lemma to obtain an upper bound involving the value we are trying to estimate. Finally, we solve an algebraic relation to obtain an explicit estimate for the moment.

We seek a bound for

$$E \overset{\mathrm{def}}{=} (\mathbb{E} \|\boldsymbol{X}\boldsymbol{R}_\varrho\|^p)^{1/p}.$$

First, observe that

$$E^2 = \left( \mathbb{E} \|\boldsymbol{X}\boldsymbol{R}_\varrho\boldsymbol{X}^*\|^{p/2} \right)^{2/p} = \left( \mathbb{E} \left\| \sum\nolimits_{j=1}^n \varrho_j \boldsymbol{x}_j \boldsymbol{x}_j^* \right\|^{p/2} \right)^{2/p}$$

where $\{\varrho_j\}$ is a sequence of independent 0–1 random variables with common mean $\varrho$. Subtract the mean, and apply the triangle inequality (once for the spectral norm and once for the $L_{p/2}$ norm):

$$E^2 \le \left( \mathbb{E} \left\| \sum\nolimits_{j=1}^n (\varrho_j - \varrho)\boldsymbol{x}_j \boldsymbol{x}_j^* \right\|^{p/2} \right)^{2/p} + \varrho \left\| \sum\nolimits_{j=1}^n \boldsymbol{x}_j \boldsymbol{x}_j^* \right\|.$$

In the sum, write $\varrho = \mathbb{E}\,\varrho_j'$ where $\{\varrho_j'\}$ is an independent copy of the sequence $\{\varrho_j\}$. Draw the expectation out of the norm with Jensen's inequality:

$$E^2 \le \left( \mathbb{E} \left\| \sum\nolimits_{j=1}^n (\varrho_j - \varrho_j')\boldsymbol{x}_j \boldsymbol{x}_j^* \right\|^{p/2} \right)^{2/p} + \varrho \|\boldsymbol{X}\boldsymbol{X}^*\|.$$

The random variables $(\varrho_j - \varrho_j')$ are symmetric and independent, so we may symmetrize them using the standard method, Lemma 6.1 of [LT91].

$$E^2 \le \left( \mathbb{E} \left\| \sum\nolimits_{j=1}^n \varepsilon_j (\varrho_j - \varrho_j')\boldsymbol{x}_j \boldsymbol{x}_j^* \right\|^{p/2} \right)^{2/p} + \varrho \|\boldsymbol{X}\|^2$$

where $\{\varepsilon_j\}$ is a Rademacher sequence. Apply the triangle inequality again and use the identical distribution of the sequences to obtain

$$E^2 \le 2 \left( \mathbb{E} \left\| \sum\nolimits_{j=1}^n \varepsilon_j \varrho_j \boldsymbol{x}_j \boldsymbol{x}_j^* \right\|^{p/2} \right)^{2/p} + \varrho \|\boldsymbol{X}\|^2$$

Writing $\Omega = \{j : \varrho_j = 1\}$, we see that

$$E^2 \leq 2 \left[ \mathbb{E}_\Omega \left( \mathbb{E}_{\boldsymbol{\varepsilon}} \left\| \sum_\Omega \varepsilon_j \boldsymbol{x}_j \boldsymbol{x}_j^* \right\|^{p/2} \right)^{(2/p)(p/2)} \right]^{2/p} + \varrho \left\| \boldsymbol{X} \right\|^2 .$$

Here, $\mathbb{E}_{\boldsymbol{\varepsilon}}$ is the partial expectation with respect to $\{\varepsilon_j\}$, holding the other random variables fixed.

To estimate the large parenthesis, invoke Rudelson's Lemma, conditional on $\Omega$. The matrix in the statement of the lemma is $\boldsymbol{X} \boldsymbol{R}_\varrho$, resulting in

$$E^2 \leq 3\sqrt{p} \left[ \mathbb{E} \left( \left\| \boldsymbol{X} \boldsymbol{R}_\varrho \right\|_{1,2} \left\| \boldsymbol{X} \boldsymbol{R}_\varrho \right\| \right)^{p/2} \right]^{2/p} + \varrho \left\| \boldsymbol{X} \right\|^2 .$$

Apply the Cauchy–Schwarz inequality to find that

$$E^2 \leq 3\sqrt{p} \left( \mathbb{E} \left\| \boldsymbol{X} \boldsymbol{R}_\varrho \right\|_{1,2}^p \right)^{1/p} \left( \mathbb{E} \left\| \boldsymbol{X} \boldsymbol{R}_\varrho \right\|^p \right)^{1/p} + \varrho \left\| \boldsymbol{X} \right\|^2 .$$

This inequality takes the form $E^2 \leq bE + c$. Select the larger root of the quadratic and use the subadditivity of the square root:

$$E \leq \frac{b + \sqrt{b^2 + 4c}}{2} \leq b + \sqrt{c}.$$

This yields the conclusion. $\qquad\square$

## 6. Proof of Maximum Column Norm Estimate

This section establishes the moment bound for the maximum column norm of a matrix that has been restricted to a random set of its rows. We use an approach that is analogous with the argument in Proposition 12. In this case, we require only the scalar Khintchine inequality to perform the estimation. Bourgain and Tzafriri's proof of Proposition 2.5 [BT91] contains a similar bound, developed with a similar argument.

**Proposition 13.** *Assume that $\boldsymbol{X}$ has $n$ columns, and suppose $p \geq 2\log n \geq 4$. Then*

$$\left( \mathbb{E} \left\| \boldsymbol{R}_\varrho \boldsymbol{X} \right\|_{1,2}^p \right)^{1/p} \leq 2^{1.5} \sqrt{p} \left\| \boldsymbol{X} \right\|_{\max} + \sqrt{\varrho} \left\| \boldsymbol{X} \right\|_{1,2} .$$

In words, the $B(\ell_1^n, \ell_2^n)$ norm of a matrix that has been compressed to a random set of rows gets its share of the total, plus an additional component that depends on the number of columns and the magnitude of the largest entry in matrix.

*Proof.* Our strategy is the same as in the proof of Proposition 12, so we pass lightly over the details. Let $\{\varrho_j\}$ be a sequence of independent 0–1 random variables with common mean $\varrho$. We seek a bound for

$$E^2 \stackrel{\text{def}}{=} \left( \mathbb{E} \left\| \boldsymbol{R}_\varrho \boldsymbol{X} \right\|_{1,2}^p \right)^{2/p} = \left( \mathbb{E} \max_{k=1,2,\dots,n} \left| \sum_j \varrho_j \left| x_{jk} \right|^2 \right|^{p/2} \right)^{2/p} .$$

In the sequel, we abbreviate $q = p/2$ and also $y_{jk} = \left| x_{jk} \right|^2$.

First, center and symmetrize the selectors. In the following calculation, $\{\varrho'_j\}$ is an independent copy of the sequence $\{\varrho_j\}$, and $\{\varepsilon_j\}$ is a Rademacher sequence, independent of everything else.

$$
\begin{aligned}
E^2 &\leq \left( \mathbb{E}\max_k \left| \sum_j (\varrho_j - \varrho) y_{jk} \right|^q \right)^{1/q} + \varrho \max_k \left| \sum_j y_{jk} \right| \\
&\leq \left( \mathbb{E}\max_k \left| \sum_j (\varrho_j - \varrho'_j) y_{jk} \right|^q \right)^{1/q} + \varrho \left\| \boldsymbol{X} \right\|_{1,2}^2 \\
&= \left( \mathbb{E}\max_k \left| \sum_j \varepsilon_j (\varrho_j - \varrho'_j) y_{jk} \right|^q \right)^{1/q} + \varrho \left\| \boldsymbol{X} \right\|_{1,2}^2 \\
&\leq 2 \left( \mathbb{E}\max_k \left| \sum_j \varepsilon_j \varrho_j y_{jk} \right|^q \right)^{1/q} + \varrho \left\| \boldsymbol{X} \right\|_{1,2}^2 .
\end{aligned}
$$

The first step uses the triangle inequality; the second uses $\varrho = \mathbb{E}\,\varrho'_j$ and Jensen's inequality; the third follows from the standard symmetrization, Lemma 6.1 of [LT91]. In the last step, we invoked the triangle inequality and the identical distribution of the two sequences.

Next, bound the maximum by a sum and introduce conditional expectations:

$$
E^2 \leq 2 \left( \mathbb{E}_{\boldsymbol{\varrho}} \sum_k \mathbb{E}_{\boldsymbol{\varepsilon}} \left| \sum_j \varepsilon_j \varrho_j y_{jk} \right|^q \right)^{1/q} + \varrho \left\| \boldsymbol{X} \right\|_{1,2}^2 .
$$

Here, $\mathbb{E}_{\boldsymbol{\varepsilon}}$ denotes partial expectation with respect to $\{\varepsilon_j\}$, holding the other random variables fixed. Since $q \geq 2$, we may apply the scalar Khintchine inequality to the inner expectation to obtain

$$
E^2 \leq 2\mathrm{C}_q \left( \mathbb{E}_{\boldsymbol{\varrho}} \sum_k \left| \sum_j \varrho_j y_{jk}^2 \right|^{q/2} \right)^{1/q} + \varrho \left\| \boldsymbol{X} \right\|_{1,2}^2 ,
$$

where the constant $\mathrm{C}_q \leq 2^{0.25} \mathrm{e}^{-1/2} \sqrt{q}$. The value of the constant follows from work of Haagerup [Haa82], combined with Stirling's approximation.

Bound the outer sum, which ranges over $n$ indices, by a maximum:

$$
E^2 \leq 2^{1.25} \mathrm{e}^{-1/2} n^{1/q} \sqrt{q} \left( \mathbb{E}_{\boldsymbol{\varrho}} \max_k \left| \sum_j \varrho_j y_{jk}^2 \right|^{q/2} \right)^{1/q} + \varrho \left\| \boldsymbol{X} \right\|_{1,2}^2 .
$$

Since $q \geq \log n$, it holds that $n^{1/q} \leq \mathrm{e}$, which implies that the leading constant is less than four. Use Hölder's inequality to bound the sum, and then apply Hölder's inequality again to double the exponent:

$$
\begin{aligned}
E^2 &< 4\sqrt{q} \left( \max_{j,k} y_{jk} \right)^{1/2} \left( \mathbb{E}_{\boldsymbol{\varrho}} \max_k \left| \sum_j \varrho_j y_{jk} \right|^{q/2} \right)^{1/q} + \varrho \left\| \boldsymbol{X} \right\|_{1,2}^2 \\
&\leq 4\sqrt{q} \left( \max_{j,k} y_{jk} \right)^{1/2} \left( \mathbb{E}_{\boldsymbol{\varrho}} \max_k \left| \sum_j \varrho_j y_{jk} \right|^q \right)^{1/2q} + \varrho \left\| \boldsymbol{X} \right\|_{1,2}^2 .
\end{aligned}
$$

Recall that $q = p/2$ and that $y_{jk} = |x_{jk}|^2$. Observe that we have obtained a copy of $E$ on the right-hand side, so

$$
E^2 \leq 2^{1.5} \sqrt{p} \left\| \boldsymbol{X} \right\|_{\max} E + \varrho \left\| \boldsymbol{X} \right\|_{1,2}^2 .
$$

As in the proof of Proposition 12, we take the larger root of the quadratic and invoke the subadditivity of the square root to reach

$$
E \leq 2^{1.5} \sqrt{p} \left\| \boldsymbol{X} \right\|_{\max} + \sqrt{\varrho} \left\| \boldsymbol{X} \right\|_{1,2} .
$$

This is the advertised conclusion.                                                          □

## 7. PROOF OF EXTRAPOLATION BOUND

This section summarizes the argument of Bourgain and Tzafriri that leads to the extrapolation result. The key to the proof is an observation due to V. A. Markov that estimates the coefficients of an arbitrary polynomial in terms of its maximum value [Tim63, Sec. 2.9].

**Proposition 14** (Markov). *Let $r(t) = \sum_{k=0}^{d} c_k t^k$. The coefficients of the polynomial $r$ satisfy the inequality*

$$|c_k| \leq \frac{d^k}{k!} \max_{|t| \leq 1} |r(t)| \leq e^d \max_{|t| \leq 1} |r(t)|.$$

*for each $k = 0, 1, \ldots, d$.*

The proof depends on the minimax property of the Chebyshev polynomial of degree $d$, combined with a careful determination of its coefficients.

**Proposition 15** (Bourgain–Tzafriri). *Let $p$ be an even integer with $p \geq 2 \log n$. Suppose that $\boldsymbol{X}$ is an $n \times n$ matrix with $\|\boldsymbol{X}\| \leq 1$. Choose parameters $\delta \in (0, 1)$ and $\varrho \in (0, 0.5)$. For each $\lambda \in (0, 1)$, it holds that*

$$(\mathbb{E} \|\boldsymbol{R}_\delta \boldsymbol{X} \boldsymbol{R}_\delta\|^p)^{1/p} \leq 60 \left[ \delta^\lambda + \varrho^{-\lambda} (\mathbb{E} \|\boldsymbol{R}_\varrho \boldsymbol{X} \boldsymbol{R}_\varrho\|^p)^{1/p} \right].$$

*For self-adjoint matrices, the constant is halved.*

*Proof.* We assume that $\boldsymbol{X}$ is self-adjoint. For general $\boldsymbol{X}$, apply the final bound to each half of the Cartesian decomposition

$$\boldsymbol{X} = \frac{\boldsymbol{X} + \boldsymbol{X}^*}{2} + \frac{\mathrm{i}(\boldsymbol{X} - \boldsymbol{X}^*)}{2\mathrm{i}}.$$

This yields the same result with constants doubled.

Consider the function

$$F(s) = \mathbb{E} \|\boldsymbol{R}_s \boldsymbol{X} \boldsymbol{R}_s\|^p \qquad \text{with } 0 \leq s \leq 1.$$

Note that $F(s) \leq 1$ because $\|\boldsymbol{R}_s \boldsymbol{X} \boldsymbol{R}_s\| \leq \|\boldsymbol{X}\| \leq 1$. Furthermore, $F$ increases monotonically.

Next, we show that $F$ is comparable with a polynomial. Use the facts that $p$ is even, that $p \geq \log n$, and that $\operatorname{rank} \boldsymbol{X} \leq n$ to check the inequalities

$$F(s) \leq \mathbb{E} \operatorname{trace}(\boldsymbol{R}_s \boldsymbol{X} \boldsymbol{R}_s)^p \leq e^p F(s).$$

It is easy to see that the central member is a polynomial of maximum degree $p$ in the variable $s$. Indeed, one may expand the product and compute the expectation using the fact that the diagonal entries of $\boldsymbol{R}_s$ are independent 0–1 random variables of mean $s$. Therefore,

$$\mathbb{E} \operatorname{trace}(\boldsymbol{R}_s \boldsymbol{X} \boldsymbol{R}_s)^p = \sum_{k=1}^{p} c_k s^k$$

for (unknown) coefficients $c_1, c_2, \ldots, c_p$. The polynomial has no constant term because $\boldsymbol{R}_0 = \boldsymbol{0}$.

We must develop some information about this polynomial. Make the change of variables $s = \varrho t^2$ to see that

$$\left| \sum_{k=1}^{p} c_k \varrho^k t^{2k} \right| \leq e^p F(\varrho t^2) \leq e^p F(\varrho) \qquad \text{when } |t| \leq 1.$$

The second inequality follows from the monotonicity of $F$. The polynomial on the left-hand side has degree $2p$ in the variable $t$, so Proposition 14 results in

$$|c_k| \varrho^k \leq e^{3p} F(\varrho). \qquad \text{for } k = 1, 2, \ldots, p.$$

From here, it also follows that $|c_k| \leq e^{3p}$ by taking $\varrho = 1$.

Finally, we directly evaluate the polynomial at $\delta$ using the facts we have uncovered. For an arbitrary value of $\lambda$ in $(0,1)$, we have

$$F(\delta) \leq \left| \sum\nolimits_{k=1}^{d} c_k \delta^k \right|$$

$$\leq \sum\nolimits_{k=1}^{\lfloor \lambda p \rfloor} |c_k| + \sum\nolimits_{k=1+\lfloor \lambda p \rfloor}^{p} |c_k| \, \delta^k$$

$$\leq \mathrm{e}^{3p} F(\varrho) \sum_{k=1}^{\lfloor \lambda p \rfloor} \varrho^{-k} + p\mathrm{e}^{3p}\delta^{\lambda p}$$

$$\leq 2\mathrm{e}^{3p}\varrho^{-\lambda p} F(\varrho) + p\mathrm{e}^{3p}\delta^{\lambda p}$$

since $\varrho \leq 0.5$. Since $x \mapsto x^{1/p}$ is subadditive, we conclude that

$$F(\delta)^{1/p} \leq 2^{1/p}\mathrm{e}^3 \cdot \varrho^{-\lambda} F(\varrho)^{1/p} + p^{1/p}\mathrm{e}^3 \cdot \delta^{\lambda}.$$

A numerical calculation shows that both the leading terms are less than 30, irrespective of $p$.    $\square$

## Acknowledgments

## References

[And79]  J. Anderson. Extreme points in sets of positive linear maps on $B(H)$. *J. Functional Anal.*, 31:195–217, 1979.

[BT87]  J. Bourgain and L. Tzafriri. Invertibility of "large" submatrices with applications to the geometry of Banach spaces and harmonic analysis. *Israel J. Math*, 57(2):137–224, 1987.

[BT91]  J. Bourgain and L. Tzafriri. On a problem of Kadison and Singer. *J. reine angew. Math.*, 420:1–43, 1991.

[Buc01]  A. Buchholz. Operator Khintchine inequality in non-commutative probability. *Math. Annalen*, 319:1–16, 2001.

[CR06]  E. J. Candès and J. Romberg. Quantitative robust uncertainty principles and optimally sparse decompositions. *Foundations of Comput. Math*, 2006. To appear.

[CT06]  P. G. Casazza and J. C. Tremain. The Kadison–Singer problem in mathematics and engineering. *Proc. Natl. Acad. Sci.*, 103(7):2032–2039, Feb. 2006.

[Haa82]  U. Haagerup. The best constants in the Khintchine inequality. *Studia Math.*, 70:231–283, 1982.

[JS68]  K. Jogdeo and S. M. Samuels. Monotone convergence of binomial probabilities and generalization of Ramanujan's equation. *Ann. Math. Stat.*, 39:1191–1195, 1968.

[LP86]  F. Lust-Picquard. Inégalités de Khintchine dans $C_p$ $(1 < p < \infty)$. *Comptes Rendus Acad. Sci. Paris, Série I*, 303(7):289–292, 1986.

[LT91]  M. Ledoux and M. Talagrand. *Probability in Banach Spaces: Isoperimetry and Processes*. Springer, 1991.

[Rud99]  M. Rudelson. Random vectors in the isotropic position. *J. Functional Anal.*, 164:60–72, 1999.

[RV07]  M. Rudelson and R. Vershynin. Sampling from large matrices: An approach through geometric functional analysis. To appear, *J. Assoc. Comput. Mach.*, 2007.

[Tim63]  A. F. Timan. *Theory of approximation of functions of a real variable*. Pergamon, 1963.