On singularity of random ± 1 and $0/1$ matrices.

Alexander Litvak

University of Alberta

based on a joint work with

Konstantin Tikhomirov

to the memory of Yehoram (Yoram) Gordon and Nicole Tomczak-Jaegermann

Seville, 2022

Yehoram (Yoram) Gordon 02/07/1940 — 18/06/2022

Nicole Tomczak-Jaegermann 08/06/1945 — 17/06/2022

Yoram Gordon 1940–2022

Many results in Local Theory of Banach Spaces and Asymptotic Geometric Analysis.

Yoram Gordon 1940–2022

Many results in Local Theory of Banach Spaces and Asymptotic Geometric Analysis.

Celebrated mini-max theorem, Gaussian approach to Dvoretzky theorem, related results on Banach-Mazur distances, “low M^*-estimate”, diameters of sections ...

Yoram Gordon 1940–2022

Many results in Local Theory of Banach Spaces and Asymptotic Geometric Analysis.

Celebrated mini-max theorem, Gaussian approach to Dvoretzky theorem, related results on Banach-Mazur distances, “low M^*-estimate”, diameters of sections ...

Inverse Santaló inequality up to logarithm

Yoram Gordon 1940–2022

Many results in Local Theory of Banach Spaces and Asymptotic Geometric Analysis.

Celebrated mini-max theorem, Gaussian approach to Dvoretzky theorem, related results on Banach-Mazur distances, "low M^*-estimate", diameters of sections ...

Inverse Santaló inequality up to logarithm

Series of works with C. Schütt, E. Werner and myself on Orlicz norms of sequences of sandom variables, order statistics and applications to convex geometry.

Yoram Gordon 1940–2022

Many results in Local Theory of Banach Spaces and Asymptotic Geometric Analysis.

Celebrated mini-max theorem, Gaussian approach to Dvoretzky theorem, related results on Banach-Mazur distances, "low M^*-estimate", diameters of sections ...

Inverse Santaló inequality up to logarithm

Series of works with C. Schütt, E. Werner and myself on Orlicz norms of sequences of sandom variables, order statistics and applications to convex geometry.

Theorem (Gordon, L., Meyer, Pajor, 2004)

Let $K, L \subset \mathbb{R}^n$ be convex bodies such that $0 \in$ int L and K is in the maximal volume position in L. The $\exists\, z \in K$ such that denoting $K_z := K - z$ and $L_z := L - z$,

$$\exists m \leq n^2 + n, \quad \exists x_i \in K_z \cap L_z, \quad \exists y_i \in K_z^0 \cap L_z^0, \quad \langle x_i, y_i \rangle = 1, \quad \exists c_i > 0\, (i \leq m),$$

such that

$$I = \sum_{i=1}^{m} c_i\, x_i \otimes y_i \quad \text{and} \quad \sum_{i=1}^{m} c_i x_i = \sum_{i=1}^{m} c_i y_i = 0.$$

Yoram Gordon 1940–2022

Theorem (Gordon, L., Meyer, Pajor, 2004)

Let $K, L \subset \mathbb{R}^n$ be convex bodies such that $0 \in$ int L and K is of maximal volume position in L. The $\exists\, z \in K$ such that denoting $K_z := K - z$ and $L_z := L - z$,

$$\exists m \leq n^2 + n, \quad \exists x_i \in K_z \cap L_z, \quad \exists y_i \in K_z^0 \cap L_z^0, \quad \langle x_i, y_i \rangle = 1, \quad \exists c_i > 0 \, (i \leq m),$$

such that

$$I = \sum_{i=1}^{m} c_i \, x_i \otimes y_i \quad \text{and} \quad \sum_{i=1}^{m} c_i x_i = \sum_{i=1}^{m} c_i y_i = 0.$$

Yoram Gordon 1940–2022

Theorem (Gordon, L., Meyer, Pajor, 2004)

Let $K, L \subset \mathbb{R}^n$ be convex bodies such that $0 \in int\, L$ and K is of maximal volume position in L. The $\exists\, z \in K$ such that denoting $K_z := K - z$ and $L_z := L - z$,

$$\exists m \le n^2 + n, \quad \exists x_i \in K_z \cap L_z, \quad \exists y_i \in K_z^0 \cap L_z^0, \quad \langle x_i, y_i \rangle = 1, \quad \exists c_i > 0\,(i \le m),$$

such that

$$I = \sum_{i=1}^m c_i\, x_i \otimes y_i \quad \text{and} \quad \sum_{i=1}^m c_i x_i = \sum_{i=1}^m c_i y_i = 0.$$

One can show that in this case

$$K_z \subset L_z \subset -nK_z,$$

solving one of B. Grünbaum problems from his seminal 1963 paper.

Nicole Tomczak-Jaegermann 1945–2022

Many deep results in both, the classical infinite-dimensional Banach space theory and in its finite-dimensional counterpart, Asymptotic Geometric Analysis.

Nicole Tomczak-Jaegermann 1945–2022

Many deep results in both, the classical infinite-dimensional Banach space theory and in its finite-dimensional counterpart, Asymptotic Geometric Analysis.

Very influential book *“Banach-Mazur distances and finite-dimensional operator ideals*" and a survey (with P. Mankiewicz) *“Quotients of finite dimensional Banach spaces; random phenomena”* in Handbook of the geometry of Banach spaces.

Nicole Tomczak-Jaegermann 1945–2022

Many deep results in both, the classical infinite-dimensional Banach space theory and in its finite-dimensional counterpart, Asymptotic Geometric Analysis.

Very influential book *"Banach-Mazur distances and finite-dimensional operator ideals"* and a survey (with P. Mankiewicz) *"Quotients of finite dimensional Banach spaces; random phenomena"* in Handbook of the geometry of Banach spaces.

Theorem (Adamczak, L., Pajor, Tomczak-Jaegermann, 2010)

Let $\varepsilon > 0$ and $X_1, \dots, X_N$ be i.i.d. random vectors, distributed according to an isotropic log-concave probability measure on $\mathbb{R}^n$. If $N \geq C_\varepsilon n$, then

$$\mathbb{P}\left(\left\| \frac{1}{N} \sum_{i=1}^{N} X_i \otimes X_i - I \right\| \leq \varepsilon \right) \geq 1 - e^{-c\sqrt{n}}.$$

Equivalently,

$$\mathbb{P}\left(\sup_{y \in S^{n-1}} \left| \frac{1}{N} \sum_{i=1}^{N} (\langle X_i, y\rangle^2 - \mathbb{E}\langle X_i, y\rangle^2) \right| \leq \varepsilon \right) \geq 1 - e^{-c\sqrt{n}}.$$

Random ± 1 matrices

An old problem: Let B be an $n \times n$ random matrix with i.i.d. ± 1 entries, that is,

$$B = \{\delta_{ij}\}_{i,j \leq n}, \qquad \delta_{ij} = \begin{cases} 1, & \text{with probability } 1/2 \\ -1, & \text{with probability } 1/2 \end{cases}$$

(such variables are often called Rademacher random variables).

Random ± 1 matrices

An old problem: Let B be an $n \times n$ random matrix with i.i.d. ± 1 entries, that is,

$$B = \{\delta_{ij}\}_{i,j \le n}, \qquad \delta_{ij} = \begin{cases} 1, & \text{with probability } 1/2 \\ -1, & \text{with probability } 1/2 \end{cases}$$

(such variables are often called Rademacher random variables).

Question. *What is*

$$P_n := \mathbb{P}\{B \textit{ is singular}\}?$$

Random ± 1 matrices

An old problem: Let B be an $n \times n$ random matrix with i.i.d. ± 1 entries, that is,

$$B = \{\delta_{ij}\}_{i,j \le n}, \qquad \delta_{ij} = \begin{cases} 1, & \text{with probability } 1/2 \\ -1, & \text{with probability } 1/2 \end{cases}$$

(such variables are often called Rademacher random variables).

Question. *What is*

$$P_n := \mathbb{P}\{B \textit{ is singular}\}?$$

Equivalently: Let $X_1, X_2, \ldots, X_n$ be independent random vectors uniformly distributed on the vertices of the n-dimensional cube $[-1, 1]^n$.

Random ± 1 matrices

An old problem: Let B be an $n \times n$ random matrix with i.i.d. ± 1 entries, that is,

$$B = \{\delta_{ij}\}_{i,j \le n}, \qquad \delta_{ij} = \begin{cases} 1, & \text{with probability } 1/2 \\ -1, & \text{with probability } 1/2 \end{cases}$$

(such variables are often called Rademacher random variables).

Question. *What is*

$$P_n := \mathbb{P}\{B \textit{ is singular}\}?$$

Equivalently: Let $X_1, X_2, \ldots, X_n$ be independent random vectors uniformly distributed on the vertices of the n-dimensional cube $[-1, 1]^n$.

Question. *What is the probability that the vectors are linearly dependent?*

The trivial lower bound:

$$P_n \geq \mathbb{P}\,\{\text{Two rows/columns of } B \text{ are equal up to a sign}\} \geq (1 - o_n(1))\, 2n^2\, 2^{-n}.$$

The trivial lower bound:

$$P_n \geq \mathbb{P}\{\text{Two rows/columns of } B \text{ are equal up to a sign}\} \geq (1 - o_n(1))\, 2n^2\, 2^{-n}.$$

A natural conjecture: *This is the main reason for singularity.*

Conjecture 1. $P_n = (1/2 + o_n(1))^n = 2^{-(1+o_n(1))n}.$

The trivial lower bound:

$$P_n \geq \mathbb{P}\{\text{Two rows/columns of } B \text{ are equal up to a sign}\} \geq (1 - o_n(1))\, 2n^2\, 2^{-n}.$$

A natural conjecture: *This is the main reason for singularity.*

Conjecture 1. $P_n = (1/2 + o_n(1))^n = 2^{-(1+o_n(1))n}.$

Conjecture 2. $P_n = (1+o_n(1))\, 2n^2\, 2^{-n}.$

Known results

Komlós (**67**): $P_n \to 0.$

Known results

Komlós (**67**): $P_n \to 0$.

Kahn, Komlós and Szemerédi (**95**): $P_n \le 0.999^n$.

Known results

Komlós (**67**): $P_n \to 0.$

Kahn, Komlós and Szemerédi (**95**): $P_n \leq 0.999^n.$

Tao–Vu (**07**): $P_n \leq (3/4 + o_n(1))^n.$

Known results

Komlós (**67**): $P_n \to 0.$

Kahn, Komlós and Szemerédi (**95**): $P_n \le 0.999^n.$

Tao–Vu (**07**): $P_n \le (3/4 + o_n(1))^n.$

Bourgain–Vu–P.M. Wood (**10**): $P_n \le (1/\sqrt{2} + o_n(1))^n.$

Known results

Komlós (**67**): $P_n \to 0$.

Kahn, Komlós and Szemerédi (**95**): $P_n \le 0.999^n$.

Tao–Vu (**07**): $P_n \le (3/4 + o_n(1))^n$.

Bourgain–Vu–P.M. Wood (**10**): $P_n \le (1/\sqrt{2} + o_n(1))^n$.

K. Tikhomirov (**20**): $P_n \le (1/2 + o_n(1))^n$ (this solves Conjecture 1).

Bernoulli random matrices

One can ask a similar question about Bernoulli 0/1 random matrices:

Bernoulli random matrices

One can ask a similar question about Bernoulli 0/1 random matrices:

Let $p \in (0, 1/2)$ and let B_p be an $n \times n$ random matrix with i.i.d. $0/1$ random variables taking value 1 with probability p, that is

$$B = \{\delta_{ij}\}_{i,j \le n}, \qquad \delta_{ij} = \begin{cases} 1, & \text{with probability } \; p \\ 0, & \text{with probability } \; 1-p \end{cases}$$

Bernoulli random matrices

One can ask a similar question about Bernoulli 0/1 random matrices:

Let $p \in (0, 1/2)$ and let B_p be an $n \times n$ random matrix with i.i.d. $0/1$ random variables taking value 1 with probability p, that is

$$B = \{\delta_{ij}\}_{i,j \leq n}, \qquad \delta_{ij} = \begin{cases} 1, & \text{with probability} \quad p \\ 0, & \text{with probability} \quad 1-p \end{cases}$$

Note that B_p can be viewed as the adjacency matrix of (directed) Erdős–Rényi graph — a random graph on n vertices whose edges appear independently of others with probability p.

Bernoulli random matrices

One can ask a similar question about Bernoulli 0/1 random matrices:

Let $p \in (0, 1/2)$ and let B_p be an $n \times n$ random matrix with i.i.d. $0/1$ random variables taking value 1 with probability p, that is

$$B = \{\delta_{ij}\}_{i,j \leq n}, \qquad \delta_{ij} = \begin{cases} 1, & \text{with probability} \quad p \\ 0, & \text{with probability} \quad 1-p \end{cases}$$

Note that B_p can be viewed as the adjacency matrix of (directed) Erdős–Rényi graph — a random graph on n vertices whose edges appear independently of others with probability p.

Question: *What is*

$$P_n := \mathbb{P}\{B_p \textit{ is singular}\}?$$

Conjecture 3.

$$P_n = (1 + o_n(1))\mathbb{P}\left\{\exists \text{ a zero row or a zero column}\right\} = (1 + o_n(1))\, 2n(1-p)^n.$$

Geometrically the condition means that either

(i) there is a zero column or

(ii) there a *coordinate* hyperplane such that all columns belong to it.

Conjecture 3.

$$P_n = (1 + o_n(1))\mathbb{P}\left\{\exists \text{ a zero row or a zero column}\right\} = (1 + o_n(1))\, 2n(1-p)^n.$$

Geometrically the condition means that either

(i) there is a zero column or

(ii) there a *coordinate* hyperplane such that all columns belong to it.

Many works on different models of sparse matrices (with iid entries):
Götze–A. Tikhomirov, Costello–Vu, Basak–Rudelson, Rudelson–K. Tikhomirov, Tao–Vu, ...

Bernoulli random matrices

Basak–Rudelson (**17**): for $p = p(n) \geq (C \ln n)/n$

$$P_n \leq \exp(-cnp).$$

Bernoulli random matrices

Basak–Rudelson (**17**): for $p = p(n) \geq (C \ln n)/n$

$$P_n \leq \exp(-cnp).$$

K. Tikhomirov (**20**): for $p \in (0, 1/2]$ (note, p is independent of n)

$$P_n \leq (1 - p + o_n(1))^n.$$

Bernoulli random matrices

Basak–Rudelson (**17**): for $p = p(n) \geq (C \ln n)/n$

$$P_n \leq \exp(-cnp).$$

K. Tikhomirov (**20**): for $p \in (0, 1/2]$ (note, p is independent of n)

$$P_n \leq (1 - p + o_n(1))^n.$$

(compare with the conjectured bound $(1 + o_n(1))\, 2n(1-p)^n$).

Bernoulli random matrices

Basak–Rudelson (**17**): for $p = p(n) \geq (C \ln n)/n$

$$P_n \leq \exp(-cnp).$$

K. Tikhomirov (**20**): for $p \in (0, 1/2]$ (note, p is independent of n)

$$P_n \leq (1 - p + o_n(1))^n.$$

(compare with the conjectured bound $(1 + o_n(1))\, 2n(1-p)^n$).

As usual in such results corresponding bounds were given for the smallest singular value

$$s_n(M) = \inf_{\|x\|_2 = 1} \|Mx\|_2 = \inf\{\|M - T : \ell_2^n \to \ell_2^n\| \mid T \text{ is singular}\} = \frac{1}{\|M^{-1}\|}.$$

Solution of the conjecture

Conjecture 3. Let $p \in (0, 1/2)$ and let B_p be an $n \times n$ random matrix with i.i.d. $0/1$ random variables taking value 1 with probability p. Then

$$P_n = \mathbb{P}\{B_p \text{ is singular}\} = (1 + o_n(1))\, \mathbb{P}\{\exists \text{ a zero row or a zero column}\}.$$

Solution of the conjecture

Conjecture 3. Let $p \in (0, 1/2)$ and let B_p be an $n \times n$ random matrix with i.i.d. $0/1$ random variables taking value 1 with probability p. Then

$$P_n = \mathbb{P}\left\{B_p \text{ is singular}\right\} = (1 + o_n(1))\, \mathbb{P}\left\{\exists \text{ a zero row or a zero column}\right\}.$$

Remark. If $(pn - \ln n) \to \infty$ then

$$\mathbb{P}\left\{\exists \text{ a zero row or a zero column}\right\} = (2 + o_n(1))n(1-p)^n.$$

Solution of the conjecture

Conjecture 3. Let $p \in (0, 1/2)$ and let B_p be an $n \times n$ random matrix with i.i.d. $0/1$ random variables taking value 1 with probability p. Then

$$P_n = \mathbb{P}\{B_p \text{ is singular}\} = (1 + o_n(1))\,\mathbb{P}\{\exists \text{ a zero row or a zero column}\}.$$

Remark. If $(pn - \ln n) \to \infty$ then

$$\mathbb{P}\{\exists \text{ a zero row or a zero column}\} = (2 + o_n(1))n(1-p)^n.$$

Conjecture 3 has been recently verified (bounds on s_n were also provided):

Solution of the conjecture

Conjecture 3. Let $p \in (0, 1/2)$ and let B_p be an $n \times n$ random matrix with i.i.d. $0/1$ random variables taking value 1 with probability p. Then

$$P_n = \mathbb{P}\{B_p \text{ is singular}\} = (1 + o_n(1))\, \mathbb{P}\{\exists \text{ a zero row or a zero column}\}.$$

Remark. If $(pn - \ln n) \to \infty$ then

$$\mathbb{P}\{\exists \text{ a zero row or a zero column}\} = (2 + o_n(1))n(1-p)^n.$$

Conjecture 3 has been recently verified (bounds on s_n were also provided):

Basak–Rudelson (**21**): For $p = p(n) \le \frac{\ln n + o_n(\ln \ln n)}{n}$.

Solution of the conjecture

Conjecture 3. Let $p \in (0, 1/2)$ and let B_p be an $n \times n$ random matrix with i.i.d. $0/1$ random variables taking value 1 with probability p. Then

$$P_n = \mathbb{P}\{B_p \text{ is singular}\} = (1 + o_n(1))\, \mathbb{P}\{\exists \text{ a zero row or a zero column}\}.$$

Remark. If $(pn - \ln n) \to \infty$ then

$$\mathbb{P}\{\exists \text{ a zero row or a zero column}\} = (2 + o_n(1))n(1-p)^n.$$

Conjecture 3 has been recently verified (bounds on s_n were also provided):

Basak–Rudelson (**21**): For $p = p(n) \le \frac{\ln n + o_n(\ln \ln n)}{n}$.

L.–K.Tikhomirov (**21**): For $\frac{C \ln n}{n} \le p = p(n) \le c$.

Solution of the conjecture

Conjecture 3. Let $p \in (0, 1/2)$ and let B_p be an $n \times n$ random matrix with i.i.d. $0/1$ random variables taking value 1 with probability p. Then

$$P_n = \mathbb{P}\{B_p \text{ is singular}\} = (1 + o_n(1))\, \mathbb{P}\{\exists \text{ a zero row or a zero column}\}.$$

Remark. If $(pn - \ln n) \to \infty$ then

$$\mathbb{P}\{\exists \text{ a zero row or a zero column}\} = (2 + o_n(1))n(1-p)^n.$$

Conjecture 3 has been recently verified (bounds on s_n were also provided):

Basak–Rudelson (**21**): For $p = p(n) \le \frac{\ln n + o_n(\ln \ln n)}{n}$.

L.–K.Tikhomirov (**21**): For $\frac{C \ln n}{n} \le p = p(n) \le c$.

Han Huang (**22**): For $\frac{\ln n}{n} \le p = p(n) \le \frac{C \ln n}{n}$.

Solution of the conjecture

Conjecture 3. Let $p \in (0, 1/2)$ and let B_p be an $n \times n$ random matrix with i.i.d. $0/1$ random variables taking value 1 with probability p. Then

$$P_n = \mathbb{P}\{B_p \text{ is singular}\} = (1 + o_n(1))\, \mathbb{P}\{\exists \text{ a zero row or a zero column}\}.$$

Remark. If $(pn - \ln n) \to \infty$ then

$$\mathbb{P}\{\exists \text{ a zero row or a zero column}\} = (2 + o_n(1))n(1-p)^n.$$

Conjecture 3 has been recently verified (bounds on s_n were also provided):

Basak–Rudelson (**21**): For $p = p(n) \le \frac{\ln n + o_n(\ln \ln n)}{n}$.

L.–K.Tikhomirov (**21**): For $\frac{C \ln n}{n} \le p = p(n) \le c$.

Han Huang (**22**): For $\frac{\ln n}{n} \le p = p(n) \le \frac{C \ln n}{n}$.

Jain–Sah–Sawhney (**22**): For $c \le p < 1/2$.

Jain–Sah–Sawhney result.

Let ξ be a non-constant real-valued random variable with finite support.
Let $M(\xi)$ be an $n \times n$ random matrix whose entries are independent copies of ξ.

Jain–Sah–Sawhney result.

Let ξ be a non-constant real-valued random variable with finite support.
Let $M(\xi)$ be an $n \times n$ random matrix whose entries are independent copies of ξ.

Consider events

$$\mathcal{E}_{sing} := \{M(\xi) \text{ is singular}\}, \quad \mathcal{E}_{zero} := \{\exists \text{ a zero row or a zero column}\}, \quad \text{and}$$
$$\mathcal{E}_{equal} := \{\exists \text{ two equal (up to a sign) rows or columns}\}.$$

Jain–Sah–Sawhney result.

Let ξ be a non-constant real-valued random variable with finite support.
Let $M(\xi)$ be an $n \times n$ random matrix whose entries are independent copies of ξ.

Consider events

$$\mathcal{E}_{sing} := \{M(\xi) \text{ is singular}\}, \quad \mathcal{E}_{zero} := \{\exists \text{ a zero row or a zero column}\}, \quad \text{and}$$
$$\mathcal{E}_{equal} := \{\exists \text{ two equal (up to a sign) rows or columns}\}.$$

If ξ is not uniform on its support, $\quad \mathbb{P}\left(\mathcal{E}_{sing}\right) = \mathbb{P}\left(\mathcal{E}_{zero}\right) + (1 + o_n(1))\,\mathbb{P}\left(\mathcal{E}_{equal}\right).$

Jain–Sah–Sawhney result.

Let ξ be a non-constant real-valued random variable with finite support.
Let $M(\xi)$ be an $n \times n$ random matrix whose entries are independent copies of ξ.

Consider events

$$\mathcal{E}_{sing} := \{M(\xi) \text{ is singular}\}, \quad \mathcal{E}_{zero} := \{\exists \text{ a zero row or a zero column}\}, \quad \text{and}$$
$$\mathcal{E}_{equal} := \{\exists \text{ two equal (up to a sign) rows or columns}\}.$$

If ξ is not uniform on its support, $\mathbb{P}\left(\mathcal{E}_{sing}\right) = \mathbb{P}\left(\mathcal{E}_{zero}\right) + (1 + o_n(1))\,\mathbb{P}\left(\mathcal{E}_{equal}\right).$

In particular, if ξ is Bernoulli random variable with $\mathbb{E}\xi = p \in (0, 1/2)$,

$$\mathbb{P}\{B_p \text{ is singular}\} = 2n(1-p)^n + (1 + o_n(1))n(n-1)(p^2 + (1-p)^2)^n$$

Jain–Sah–Sawhney result.

Let ξ be a non-constant real-valued random variable with finite support.
Let $M(\xi)$ be an $n \times n$ random matrix whose entries are independent copies of ξ.

Consider events

$$\mathcal{E}_{sing} := \{M(\xi) \text{ is singular}\}, \quad \mathcal{E}_{zero} := \{\exists \text{ a zero row or a zero column}\}, \quad \text{and}$$
$$\mathcal{E}_{equal} := \{\exists \text{ two equal (up to a sign) rows or columns}\}.$$

If ξ is not uniform on its support, $\mathbb{P}\left(\mathcal{E}_{sing}\right) = \mathbb{P}\left(\mathcal{E}_{zero}\right) + (1 + o_n(1))\, \mathbb{P}\left(\mathcal{E}_{equal}\right).$

In particular, if ξ is Bernoulli random variable with $\mathbb{E}\xi = p \in (0, 1/2)$,

$$\mathbb{P}\{B_p \text{ is singular}\} = 2n(1-p)^n + (1 + o_n(1))n(n-1)(p^2 + (1-p)^2)^n$$

If ξ is uniform on its support, $\mathbb{P}\left(\mathcal{E}_{sing}\right) = (1 + o_n(1))^n\, \mathbb{P}\left(\mathcal{E}_{equal}\right).$

Jain–Sah–Sawhney result.

Let ξ be a non-constant real-valued random variable with finite support.
Let $M(\xi)$ be an $n \times n$ random matrix whose entries are independent copies of ξ.

Consider events

$$\mathcal{E}_{sing} := \{M(\xi) \text{ is singular}\}, \quad \mathcal{E}_{zero} := \{\exists \text{ a zero row or a zero column}\}, \quad \text{and}$$
$$\mathcal{E}_{equal} := \{\exists \text{ two equal (up to a sign) rows or columns}\}.$$

If ξ is not uniform on its support, $\quad \mathbb{P}\left(\mathcal{E}_{sing}\right) = \mathbb{P}\left(\mathcal{E}_{zero}\right) + (1 + o_n(1))\,\mathbb{P}\left(\mathcal{E}_{equal}\right).$

In particular, if ξ is Bernoulli random variable with $\mathbb{E}\xi = p \in (0, 1/2)$,

$$\mathbb{P}\{B_p \text{ is singular}\} = 2n(1-p)^n + (1 + o_n(1))n(n-1)(p^2 + (1-p)^2)^n$$

If ξ is uniform on its support, $\quad \mathbb{P}\left(\mathcal{E}_{sing}\right) = (1 + o_n(1))^n\,\mathbb{P}\left(\mathcal{E}_{equal}\right).$

This covers the case $p = 1/2$ as well as the case of Rademacher random variables ($\xi = \pm 1$ with probability 1/2).

Some ideas of the proof.

It is well-understood by now that to deal with the smallest singular value one needs to split S^{n-1} into several parts and to work separately on each part.

Some ideas of the proof.

It is well-understood by now that to deal with the smallest singular value one needs to split S^{n-1} into several parts and to work separately on each part.

This idea goes back to Kashin (**77**), where, in order obtain an orthogonal decomposition of ℓ_1^n, he split the sphere into two classes according to the ratio of ℓ_1^n and ℓ_2^n norms. In a similar context it was used by Schehtman (**04**).

Some ideas of the proof.

It is well-understood by now that to deal with the smallest singular value one needs to split S^{n-1} into several parts and to work separately on each part.

This idea goes back to Kashin (**77**), where, in order obtain an orthogonal decomposition of ℓ_1^n, he split the sphere into two classes according to the ratio of ℓ_1^n and ℓ_2^n norms. In a similar context it was used by Schehtman (**04**).

Since we want to provide a lower bound on the smallest singular value of a random matrix M, we need to show that $\|Mx\|_2$ is not very small for all $x \in S^{n-1}$. Usually it is done using the union bound — to prove a good probability bound for an individual vector x and then to find a good net in order to apply approximation. The main point is to have a good balance between the probability and the cardinality of a net.

Some ideas of the proof.

But vectors behave differently. Consider the following example, let $X = \{\varepsilon_i\}_{i=1}^n$ be a Rademacher random vector with ± 1 independent entries. Then

$$\langle X, e_1 + e_2 \rangle = \varepsilon_1 + \varepsilon_2 = 0 \quad \text{with probability} \quad 1/2.$$

On the other hand,

$$\langle X, \sum_{i=1}^n e_i \rangle = \sum_{i=1}^n \varepsilon_i = 0 \quad \text{with probability at most} \quad 1/\sqrt{n}$$

by the Erdős–Littlewood–Offord anti-concentration lemma.

Some ideas of the proof.

But vectors behave differently. Consider the following example, let $X = \{\varepsilon_i\}_{i=1}^n$ be a Rademacher random vector with ± 1 independent entries. Then

$$\langle X, e_1 + e_2 \rangle = \varepsilon_1 + \varepsilon_2 = 0 \quad \text{with probability} \quad 1/2.$$

On the other hand,

$$\langle X, \sum_{i=1}^n e_i \rangle = \sum_{i=1}^n \varepsilon_i = 0 \quad \text{with probability at most} \quad 1/\sqrt{n}$$

by the Erdős–Littlewood–Offord anti-concentration lemma.

Usually, it is hard to get good individual bounds for vectors of small support, so-called *sparse vectors*.

Some ideas of the proof.

But vectors behave differently. Consider the following example, let $X = \{\varepsilon_i\}_{i=1}^n$ be a Rademacher random vector with ± 1 independent entries. Then

$$\langle X, e_1 + e_2 \rangle = \varepsilon_1 + \varepsilon_2 = 0 \quad \text{with probability} \quad 1/2.$$

On the other hand,

$$\langle X, \sum_{i=1}^n e_i \rangle = \sum_{i=1}^n \varepsilon_i = 0 \quad \text{with probability at most} \quad 1/\sqrt{n}$$

by the Erdős–Littlewood–Offord anti-concentration lemma.

Usually, it is hard to get good individual bounds for vectors of small support, so-called *sparse vectors*. However, the set of such vectors is essentially of lower dimension, hence admit a very good net.

Some ideas of the proof.

But vectors behave differently. Consider the following example, let $X = \{\varepsilon_i\}_{i=1}^n$ be a Rademacher random vector with ± 1 independent entries. Then

$$\langle X, e_1 + e_2 \rangle = \varepsilon_1 + \varepsilon_2 = 0 \quad \text{with probability} \quad 1/2.$$

On the other hand,

$$\langle X, \sum_{i=1}^n e_i \rangle = \sum_{i=1}^n \varepsilon_i = 0 \quad \text{with probability at most} \quad 1/\sqrt{n}$$

by the Erdős–Littlewood–Offord anti-concentration lemma.

Usually, it is hard to get good individual bounds for vectors of small support, so-called *sparse vectors*. However, the set of such vectors is essentially of lower dimension, hence admit a very good net. This leads to splitting the sphere into *compressible* vectors – those closed to sparse, and *incompressible* vectors – the rest.

Some ideas of the proof.

But vectors behave differently. Consider the following example, let $X = \{\varepsilon_i\}_{i=1}^n$ be a Rademacher random vector with ± 1 independent entries. Then

$$\langle X, e_1 + e_2 \rangle = \varepsilon_1 + \varepsilon_2 = 0 \quad \text{with probability} \quad 1/2.$$

On the other hand,

$$\langle X, \sum_{i=1}^n e_i \rangle = \sum_{i=1}^n \varepsilon_i = 0 \quad \text{with probability at most} \quad 1/\sqrt{n}$$

by the Erdős–Littlewood–Offord anti-concentration lemma.

Usually, it is hard to get good individual bounds for vectors of small support, so-called *sparse vectors*. However, the set of such vectors is essentially of lower dimension, hence admit a very good net. This leads to splitting the sphere into *compressible* vectors – those closed to sparse, and *incompressible* vectors – the rest. For compressible vectors we have a net of small cardinality, therefore relatively poor individual probability bounds work, while incompressible vectors are well spread and therefore have very good anti-concentration properties.

Some ideas of the proof.

But vectors behave differently. Consider the following example, let $X = \{\varepsilon_i\}_{i=1}^n$ be a Rademacher random vector with ± 1 independent entries. Then

$$\langle X, e_1 + e_2 \rangle = \varepsilon_1 + \varepsilon_2 = 0 \quad \text{with probability} \quad 1/2.$$

On the other hand,

$$\langle X, \sum_{i=1}^n e_i \rangle = \sum_{i=1}^n \varepsilon_i = 0 \quad \text{with probability at most} \quad 1/\sqrt{n}$$

by the Erdős–Littlewood–Offord anti-concentration lemma.

Usually, it is hard to get good individual bounds for vectors of small support, so-called *sparse vectors*. However, the set of such vectors is essentially of lower dimension, hence admit a very good net. This leads to splitting the sphere into *compressible* vectors – those closed to sparse, and *incompressible* vectors – the rest. For compressible vectors we have a net of small cardinality, therefore relatively poor individual probability bounds work, while incompressible vectors are well spread and therefore have very good anti-concentration properties. This approach was first used in L–Pajor–Rudelson–Tomczak-Jaegermann (**05**) for rectangular matrices and was later developed in series of works by Rudelson–Vershynin.

Some ideas of the proof.

For Bernoulli $0/1$ matrices the splitting of the sphere is much more involved.

Some ideas of the proof.

For Bernoulli $0/1$ matrices the splitting of the sphere is much more involved.

An additional problem is caused by constant vectors. Indeed, while properly normalized centered random matrices (say, with entries ± 1) have norm of order $\sqrt{n}$, the norm $\|B_p\| \approx pn$.

Some ideas of the proof.

For Bernoulli $0/1$ matrices the splitting of the sphere is much more involved.

An additional problem is caused by constant vectors. Indeed, while properly normalized centered random matrices (say, with entries ± 1) have norm of order $\sqrt{n}$, the norm $\|B_p\| \approx pn$. Fortunately, this large norm is only in the direction of $\mathbf{1} = (1, 1, ..., 1)$. On the subspace orthogonal to $\mathbf{1}$ the norm is of the order $\sqrt{pn}$.

Some ideas of the proof.

For Bernoulli $0/1$ matrices the splitting of the sphere is much more involved.

An additional problem is caused by constant vectors. Indeed, while properly normalized centered random matrices (say, with entries ± 1) have norm of order $\sqrt{n}$, the norm $\|B_p\| \approx pn$. Fortunately, this large norm is only in the direction of $\mathbf{1} = (1, 1, ..., 1)$. On the subspace orthogonal to $\mathbf{1}$ the norm is of the order $\sqrt{pn}$.

So-called *structured* vectors, corresponding to incompressible vectors in other models, are vectors of 3 types.

Some ideas of the proof.

For Bernoulli $0/1$ matrices the splitting of the sphere is much more involved.

An additional problem is caused by constant vectors. Indeed, while properly normalized centered random matrices (say, with entries ± 1) have norm of order $\sqrt{n}$, the norm $\|B_p\| \approx pn$. Fortunately, this large norm is only in the direction of $\mathbf{1} = (1, 1, ..., 1)$. On the subspace orthogonal to $\mathbf{1}$ the norm is of the order $\sqrt{pn}$.

So-called *structured* vectors, corresponding to incompressible vectors in other models, are vectors of 3 types.

First, *almost constant vectors* — vectors having many almost equal coordinates. In other words, they are compressible vectors shifted by constants vectors.

Some ideas of the proof.

For Bernoulli $0/1$ matrices the splitting of the sphere is much more involved.

An additional problem is caused by constant vectors. Indeed, while properly normalized centered random matrices (say, with entries ± 1) have norm of order $\sqrt{n}$, the norm $\|B_p\| \approx pn$. Fortunately, this large norm is only in the direction of $\mathbf{1} = (1, 1, ..., 1)$. On the subspace orthogonal to $\mathbf{1}$ the norm is of the order $\sqrt{pn}$.

So-called *structured* vectors, corresponding to incompressible vectors in other models, are vectors of 3 types.

First, *almost constant vectors* — vectors having many almost equal coordinates. In other words, they are compressible vectors shifted by constants vectors.

Second, vectors such that after removing k largest coordinates with $\frac{1}{p} \leq k \leq \frac{n}{\ln^2(pn)}$, have a good comparison of ℓ_2- and ℓ_∞-norms. Then Rogosin anti-concentration bounds provide a good result (such bounds say that an inner product of a random vector with a flat vector can't concentrate around a number).

Some ideas of the proof.

Third, vectors having a big "jump" between certain coordinates. For such vectors technique developed in L.–Lytova–K.Tikhomirov–Tomczak-Jaegermann–Youssef papers on random regular matrices can be applied.

(A $0/1$ matrix is regular if the sums of 1 in all columns and in all rows are the same — it is the adjacency matrix of a regular directed graph).

Some ideas of the proof.

Then the remaining vectors are contained in the class of *gradual non-constant vectors*, that is, vectors after certain normalization satisfying for some parameters r, δ, h and some increasing function G,

1. $x^*_{rn} = 1$

2. $x^*_i \leq G(n/i)$

3. If $\{y_i\}_i$ is a non-increasing rearrangement of $\{x_i\}_i$ then $y_{\delta n} - y_{n-\delta n} \geq h$.

Some ideas of the proof.

Then the remaining vectors are contained in the class of *gradual non-constant vectors*, that is, vectors after certain normalization satisfying for some parameters r, δ, h and some increasing function G,

1. $x^*_{rn} = 1$

2. $x^*_i \le G(n/i)$

3. If $\{y_i\}_i$ is a non-increasing rearrangement of $\{x_i\}_i$ then $y_{\delta n} - y_{n-\delta n} \ge h$.

To work with this class we partially follow Rudelson–Vershynin scheme.

Some ideas of the proof.

Then the remaining vectors are contained in the class of *gradual non-constant vectors*, that is, vectors after certain normalization satisfying for some parameters r, δ, h and some increasing function G,

1. $x^*_{rn} = 1$
2. $x^*_i \leq G(n/i)$
3. If $\{y_i\}_i$ is a non-increasing rearrangement of $\{x_i\}_i$ then $y_{\delta n} - y_{n-\delta n} \geq h$.

To work with this class we partially follow Rudelson–Vershynin scheme.

First, one reduces estimating the smallest singular value to estimating distances between a column X_i to the span of remaining columns, say H_i, $i \leq n$.

Some ideas of the proof.

Then the remaining vectors are contained in the class of *gradual non-constant vectors*, that is, vectors after certain normalization satisfying for some parameters r, δ, h and some increasing function G,

1. $x^*_{rn} = 1$
2. $x^*_i \leq G(n/i)$
3. If $\{y_i\}_i$ is a non-increasing rearrangement of $\{x_i\}_i$ then $y_{\delta n} - y_{n-\delta n} \geq h$.

To work with this class we partially follow Rudelson–Vershynin scheme.

First, one reduces estimating the smallest singular value to estimating distances between a column X_i to the span of remaining columns, say H_i, $i \leq n$.

This distance is a projection on a (random) normal vector to H_i.

Some ideas of the proof.

Then the remaining vectors are contained in the class of *gradual non-constant vectors*, that is, vectors after certain normalization satisfying for some parameters r, δ, h and some increasing function G,

1. $x^*_{rn} = 1$

2. $x^*_i \leq G(n/i)$

3. If $\{y_i\}_i$ is a non-increasing rearrangement of $\{x_i\}_i$ then $y_{\delta n} - y_{n-\delta n} \geq h$.

To work with this class we partially follow Rudelson–Vershynin scheme.

First, one reduces estimating the smallest singular value to estimating distances between a column X_i to the span of remaining columns, say H_i, $i \leq n$.

This distance is a projection on a (random) normal vector to H_i.

Thus, we have an inner product of X_i and the normal (note that they are independent).

Some ideas of the proof.

Then the remaining vectors are contained in the class of *gradual non-constant vectors*, that is, vectors after certain normalization satisfying for some parameters r, δ, h and some increasing function G,

1. $x^*_{rn} = 1$

2. $x^*_i \le G(n/i)$

3. If $\{y_i\}_i$ is a non-increasing rearrangement of $\{x_i\}_i$ then $y_{\delta n} - y_{n-\delta n} \ge h$.

To work with this class we partially follow Rudelson–Vershynin scheme.

First, one reduces estimating the smallest singular value to estimating distances between a column X_i to the span of remaining columns, say H_i, $i \le n$.

This distance is a projection on a (random) normal vector to H_i.

Thus, we have an inner product of X_i and the normal (note that they are independent).

Then we apply an anti-concentration property (such a property says that an inner product of a random vector with a flat vector can't concentrate around a number).

Some ideas of the proof.

Then the remaining vectors are contained in the class of *gradual non-constant vectors*, that is, vectors after certain normalization satisfying for some parameters r, δ, h and some increasing function G,

1. $x^*_{rn} = 1$
2. $x^*_i \le G(n/i)$
3. If $\{y_i\}_i$ is a non-increasing rearrangement of $\{x_i\}_i$ then $y_{\delta n} - y_{n-\delta n} \ge h$.

To work with this class we partially follow Rudelson–Vershynin scheme.

First, one reduces estimating the smallest singular value to estimating distances between a column X_i to the span of remaining columns, say H_i, $i \le n$.

This distance is a projection on a (random) normal vector to H_i.

Thus, we have an inner product of X_i and the normal (note that they are independent).

Then we apply an anti-concentration property (such a property says that an inner product of a random vector with a flat vector can't concentrate around a number).

To make this scheme work, Rudelson–Vershynin introduced LCD (*least common denominator*), which, in a sense, measures how close a proportional coordinate projection of a vector to the properly rescaled integer lattice. They also had to develope Littlewood–Offord theory.

Some ideas of the proof.

In our case both, the LCD, and the known anti-concentration results are not strong enough, so we need to develop new tools.

Some ideas of the proof.

In our case both, the LCD, and the known anti-concentration results are not strong enough, so we need to develop new tools.

First idea is to pass from a Bernoulli random vector, which may have many zeros, to a random $0/1$ vector with prescribed number of ones, say, with m ones, where m is of the order pn. Note that pn is an average number of ones in a Bernoulli vector.

Some ideas of the proof.

In our case both, the LCD, and the known anti-concentration results are not strong enough, so we need to develop new tools.

First idea is to pass from a Bernoulli random vector, which may have many zeros, to a random $0/1$ vector with prescribed number of ones, say, with m ones, where m is of the order pn. Note that pn is an average number of ones in a Bernoulli vector.

Second idea is to substitute LCD with another, more appropriate estimator.

Some ideas of the proof.

In our case both, the LCD, and the known anti-concentration results are not strong enough, so we need to develop new tools.

First idea is to pass from a Bernoulli random vector, which may have many zeros, to a random $0/1$ vector with prescribed number of ones, say, with m ones, where m is of the order pn. Note that pn is an average number of ones in a Bernoulli vector.

Second idea is to substitute LCD with another, more appropriate estimator.

Next we have to prove a Littlewood–Offord type anti-concentration property for this new estimator.

Some ideas of the proof.

In our case both, the LCD, and the known anti-concentration results are not strong enough, so we need to develop new tools.

First idea is to pass from a Bernoulli random vector, which may have many zeros, to a random $0/1$ vector with prescribed number of ones, say, with m ones, where m is of the order pn. Note that pn is an average number of ones in a Bernoulli vector.

Second idea is to substitute LCD with another, more appropriate estimator.

Next we have to prove a Littlewood–Offord type anti-concentration property for this new estimator.

In particular, we also extend the Littlewood–Offord theory to the case of dependent random variables (in our case – the coordinates of a vector with fixed number of ones).

Degree of unstructuredness

Recall, Lévy concentration function is $\mathcal{L}(\xi, t) := \max_{\lambda \in \mathbb{R}} \mathbb{P}\left(|\xi - \lambda| < t\right).$

Degree of unstructuredness

Recall, Lévy concentration function is $\mathcal{L}(\xi, t) := \max_{\lambda \in \mathbb{R}} \mathbb{P}(|\xi - \lambda| < t)$.

Esseen Lemma (**66**):

$$\mathcal{L}\Big(\sum_{i=1}^{m} \xi_i, \tau\Big) \le C' \int_{-1}^{1} \prod_{i=1}^{m} |\mathbb{E} \exp(2\pi \mathbf{i} \xi_i s/\tau)| \, ds.$$

Degree of unstructuredness

Recall, Lévy concentration function is $\mathcal{L}(\xi, t) := \max_{\lambda \in \mathbb{R}} \mathbb{P}\left(|\xi - \lambda| < t\right).$

Esseen Lemma (**66**):

$$\mathcal{L}\Big(\sum_{i=1}^{m} \xi_i, \tau\Big) \leq C' \int_{-1}^{1} \prod_{i=1}^{m} |\mathbb{E} \exp(2\pi \mathbf{i} \xi_i s/\tau)| \, ds.$$

For a finite integer subset S, let $\eta[S]$ denotes a r.v. uniformly distributed on S.

Degree of unstructuredness

Recall, Lévy concentration function is $\mathcal{L}(\xi, t) := \max_{\lambda \in \mathbb{R}} \mathbb{P}(|\xi - \lambda| < t)$.

Esseen Lemma (**66**):

$$\mathcal{L}\Big(\sum_{i=1}^{m} \xi_i, \tau\Big) \leq C' \int_{-1}^{1} \prod_{i=1}^{m} |\mathbb{E}\exp(2\pi \mathbf{i} \xi_i s/\tau)|\, ds.$$

For a finite integer subset S, let $\eta[S]$ denotes a r.v. uniformly distributed on S. Then

$$\mathbf{UD}_n(v, m, K) := \sup\left\{t > 0 : \frac{1}{N} \sum_{(S_1,\ldots,S_m)} \int_{-t}^{t} \prod_{i=1}^{m} \left|\mathbb{E}\exp\left(2\pi \mathbf{i}\, v_{\eta[S_i]}\, m^{-1/2} s\right)\right| ds \leq K\right\},$$

where the sum is taken over all sequences $(S_i)_{i=1}^{m}$ of disjoint subsets $S_1, \ldots, S_m \subset [n]$, each of cardinality $\lfloor n/m \rfloor$, N is the number of such sequences, $K \geq 1$ is a parameter.

Degree of unstructuredness

Recall, Lévy concentration function is $\mathcal{L}(\xi, t) := \max_{\lambda\in\mathbb{R}} \mathbb{P}\left(|\xi - \lambda| < t\right).$

Esseen Lemma (**66**):

$$\mathcal{L}\Big(\sum_{i=1}^{m} \xi_i, \tau\Big) \le C' \int_{-1}^{1} \prod_{i=1}^{m} |\mathbb{E}\exp(2\pi \mathbf{i} \xi_i s/\tau)|\, ds.$$

For a finite integer subset S, let $\eta[S]$ denotes a r.v. uniformly distributed on S. Then

$$\mathbf{UD}_n(v, m, K) := \sup\Big\{t > 0 : \frac{1}{N} \sum_{(S_1,\dots,S_m)} \int_{-t}^{t} \prod_{i=1}^{m} \Big|\mathbb{E}\exp\left(2\pi\mathbf{i}\, v_{\eta[S_i]}\, m^{-1/2} s\right)\Big|\, ds \le K\Big\},$$

where the sum is taken over all sequences $(S_i)_{i=1}^m$ of disjoint subsets $S_1, \dots, S_m \subset [n]$, each of cardinality $\lfloor n/m \rfloor$, N is the number of such sequences, $K \ge 1$ is a parameter. We prove that

$$\mathcal{L}\Big(\sum_{i=1}^{n} v_i X_i, \sqrt{m}\, t\Big) \le C\left(t + 1/\mathbf{UD}_n(v, m, K)\right) \quad \text{for all } t > 0.$$