Spectral properties of random conjunction matrices

Mark Rudelson

Department of Mathematics
University of Michigan

joint work with Shiva Kasiwiswanathan,
Adam Smith, and Jon Ullman

Toronto 2010

Contingency tables

- Data base: a $d \times n$ matrix D with $\{0, 1\}$ entries.
  - n individual records;
  - d attributes of each individual.
- Contingency table: let $k < n$.
  For each subset $J \subset \{1, \dots, d\}$ of $|J| = k$ attributes record m_J – the percentage of the individual records having all attributes from J.

Conjunction matrix

- All attributes = Conjunction = product of $\{0,1\}$ variables
- Conjunction matrix: construct a $\binom{d}{k} \times n$ matrix $M^{(k)}$ as follows:
  for the set $J \subset \{1,\dots,d\}$, define the row $M_J^{(k)}$ as the entry-wise product of corresponding rows of D.

$$\begin{pmatrix} * & * & \dots & * \\ \delta_1 & \delta_2 & \dots & \delta_n \\ * & * & \dots & * \\ \varepsilon_1 & \varepsilon_2 & \dots & \varepsilon_n \\ * & * & \dots & * \\ \nu_1 & \nu_2 & \dots & \nu_n \\ * & * & \dots & * \end{pmatrix} \to (\delta_1 \cdot \varepsilon_1 \cdot \nu_1, \delta_2 \cdot \varepsilon_2 \cdot \nu_2, \dots, \delta_n \cdot \varepsilon_n \cdot \nu_n) =: \delta \odot \varepsilon \odot \nu.$$

  - $M^{(k)}$ is a $\{0,1\}$ matrix.
  - m_J is the percentage of 1-s in the row $M_J^{(k)}$.

Attribute non-privacy

Assume that the data base D contains $(d-1)$ publicly available attribute, and 1 sensitive one.

The privacy is violated for a data base D if knowing the contingency table, one can reconstruct coordinates of the sensitive vector

Attribute non-privacy

Assume that the data base D contains $(d-1)$ publicly available attribute, and 1 sensitive one.
The privacy is violated for a data base D if knowing the contingency table, one can reconstruct coordinates of the sensitive vector

- Denote $D = \binom{D_0}{x}$, where x is the sensitive vector. If D_0 is known, and the table of m_J is revealed, than one can form the conjunction matrix for D_0, and recover x_j solving a linear system.

Attribute non-privacy

Assume that the data base D contains $(d-1)$ publicly available attribute, and 1 sensitive one.
The privacy is violated for a data base D if knowing the contingency table, one can reconstruct coordinates of the sensitive vector

- Denote $D = \binom{D_0}{x}$, where x is the sensitive vector. If D_0 is known, and the table of m_J is revealed, than one can form the conjunction matrix for D_0, and recover x_j solving a linear system.
- Observation: m_J does not significantly depend on x.
- Privacy protection: release the contingency table with some noise.

Attribute non-privacy

Assume that the data base D contains $(d-1)$ publicly available attribute, and 1 sensitive one.
The privacy is violated for a data base D if knowing the noisy contingency table, one can reconstruct $(1-o(1))n$ coordinates of the sensitive vector with probability $(1-o(1))$.

- Denote $D = \binom{D_0}{x}$, where x is the sensitive vector. If D_0 is known, and the table of m_J is revealed, than one can form the conjunction matrix for D_0, and recover x_j solving a linear system.
- Observation: m_J does not significantly depend on x.
- Privacy protection: release the contingency table with some noise.

Attribute non-privacy

Assume that the data base D contains $(d-1)$ publicly available attribute, and 1 sensitive one.
The privacy is violated for a random data base D if knowing the noisy contingency table, one can reconstruct $(1-o(1))n$ coordinates of the sensitive vector with probability $(1-o(1))$.

- Denote $D = \binom{D_0}{x}$, where x is the sensitive vector. If D_0 is known, and the table of m_J is revealed, than one can form the conjunction matrix for D_0, and recover x_j solving a linear system.
- Observation: m_J does not significantly depend on x.
- Privacy protection: release the contingency table with some noise.
- Typical case: random data base.

Noise

- Let x be the private vector.
  The contingency table contains the vector $M^{(k)}x$.
- We release

$$y = M^{(k)}x + w, \qquad \text{where } w \text{ is the noise vector.}$$

Noise

- Let x be the private vector.
  The contingency table contains the vector $M^{(k)}x$.
- We release
  $$y = M^{(k)}x + w, \qquad \text{where } w \text{ is the noise vector.}$$
- The noise should be as small as possible to make the contingency table more reliable.
- The noise has to be big enough to protect the private vector.

Recovery and singular values

Singular value decomposition:

$M^{(k)} = P\Gamma Q$, where

- Q is an $n \times n$ isometry matrix;
- Γ is an $n \times n$ diagonal matrix of the singular values:

$$\Gamma = \operatorname{diag}\left(s_1(M^{(k)}), \dots, s_n(M^{(k)})\right).$$

- P is an $\binom{d}{k} \times n$ isometric embedding.

Recovery and singular values

Singular value decomposition:

$M^{(k)} = P\Gamma Q$, where

- Q is an $n \times n$ isometry matrix;
- Γ is an $n \times n$ diagonal matrix of the singular values:

$$\Gamma = \operatorname{diag}\big(s_1(M^{(k)}), \dots, s_n(M^{(k)})\big).$$

- P is an $\binom{d}{k} \times n$ isometric embedding.

Set $L = Q^T \Gamma^{-1} P^T$. Then $y = M^{(k)}x + w \quad \Rightarrow \quad x = Ly - Lw$.

$$\text{Hence,} \quad \|x - Ly\| \le \|L\| \cdot \|w\| \le \frac{1}{s_n\big(M^{(k)}\big)} \cdot \|w\|.$$

Recovery and singular values

Singular value decomposition:

$M^{(k)} = P\Gamma Q$, where

- Q is an $n \times n$ isometry matrix;
- Γ is an $n \times n$ diagonal matrix of the singular values:

$$\Gamma = \operatorname{diag}\big(s_1(M^{(k)}), \dots, s_n(M^{(k)})\big).$$

- P is an $\binom{d}{k} \times n$ isometric embedding.

Set $L = Q^T \Gamma^{-1} P^T$. Then $y = M^{(k)}x + w \quad \Rightarrow \quad x = Ly - Lw$.

$$\text{Hence,} \quad \|x - Ly\| \le \|L\| \cdot \|w\| \le \frac{1}{s_n(M^{(k)})} \cdot \|w\| .$$

The lower bound on the noise.

Assume that $\frac{1}{s_n(M^{(k)})} \cdot \|w\| = o(\sqrt{n})$ with high probability.
Then $(1 - o(1))n$ coordinates of this vector are of order $o(1)$.
Since x has $\{0, 1\}$ coordinates, most of the coordinates of x can be recovered by rounding.

First order contingency tables ⇒ random matrices

$$M^{(1)} = D \qquad d \geq n \qquad \text{hypothetical case}$$

Here D is a random matrix with i.i.d. bounded entries.

$$s_1(D) = \max_{x \in S^{n-1}} \|Dx\|, \qquad s_n(D) = \min_{x \in S^{n-1}} \|Dx\|.$$

Fact: $s_1(D) \leq C(\sqrt{d} + \sqrt{n})$ with probability very close to 1.

First order contingency tables ⇒ random matrices

$$M^{(1)} = D \qquad d \geq n \qquad \text{hypothetical case}$$

Here D is a random matrix with i.i.d. bounded entries.

$$s_1(D) = \max_{x \in S^{n-1}} \|Dx\|, \qquad s_n(D) = \min_{x \in S^{n-1}} \|Dx\|.$$

Fact: $s_1(D) \leq C(\sqrt{d} + \sqrt{n})$ with probability very close to 1.
General result [R, Vershynin, 2008]:

$$s_n(D) \geq c(\sqrt{d} - \sqrt{n-1}) \quad \text{with high probability.}$$

For $d \geq C'n$ this means $s_n(D) \geq c\sqrt{d}$.

If $d \geq C'n$, then the matrix D is nicely invertible (on its image)

Higher order conjunctions

If D is a $d \times n$ random matrix with independent entries, and $d \geq Cn$, then

$$s_n(D) \sim \sqrt{d} \quad \text{with high probability.}$$

Conjecture

If $k \geq 1$, M is the $\binom{d}{k} \times n$ conjunction matrix of a random matrix D, and $\binom{d}{k} \geq C(k)n$, then

$$s_n(D) \sim_k \sqrt{\binom{d}{k}} \qquad \text{with high probability}$$

Higher order conjunctions

If D is a $d \times n$ random matrix with independent entries, and $d \geq Cn$, then

$$s_n(D) \sim \sqrt{d} \quad \text{with high probability.}$$

Conjecture

If $k \geq 1$, M is the $\binom{d}{k} \times n$ conjunction matrix of a random matrix D, and $\binom{d}{k} \geq C(k)n$, then

$$s_n(D) \sim_k \sqrt{\binom{d}{k}} \sim_k d^{k/2} \quad \text{with high probability}$$

whenever $n \geq n(k)$.

n and d have to be big compare to k.

Higher order conjunctions

If D is a $d \times n$ random matrix with independent entries, and $d \geq Cn$, then

$$s_n(D) \sim \sqrt{d} \quad \text{with high probability.}$$

Conjecture

If $k \geq 1$, M is the $\binom{d}{k} \times n$ conjunction matrix of a random matrix D, and $\binom{d}{k} \geq C(k)n$, then

$$s_n(D) \sim_k \sqrt{\binom{d}{k}} \sim_k d^{k/2} \quad \text{with high probability}$$

whenever $n \geq n(k)$.

n and d have to be big compare to k.

Numerical experiments

If $k = 2$, and $\binom{d}{2} \geq 4n$, then $\quad s_n(D) \sim d \quad$ with high probability

Random data base

- Attributes of different individuals are independent and identically distributed.

Random data base

- Attributes of different individuals are independent and identically distributed. Not very realistic.

Random data base

- Attributes of different individuals are independent and identically distributed. Not very realistic.
- Each attribute has its own distribution.
- Each attribute is random.
- The individual records are independent.

Random data base

- Attributes of different individuals are independent and identically distributed. Not very realistic.
- Each attribute has its own distribution.
- Each attribute is random.
- The individual records are independent.

Random data base

Let $0 < p_1 < p_2 < 1$. Let D be a $\{0, 1\}$ random matrix with independent entries. Assume that

$$\mathbb{P}\left(d_{j,k} = 1\right) = \delta_k,$$

where $p_1 \leq \delta_k \leq p_2$.

δ_k is the probability of k-th attribute.

Higher order conjunctions

Conjecture (still open)

If $k \geq 1$, M is the $\binom{d}{K} \times n$ conjunction matrix of a random data base D, and $\binom{d}{K} \geq C(K)n$, then

$$s_n(D) \sim_k d^{K/2} \quad \text{with high probability, whenever } n \geq n(K)$$

Higher order conjunctions

Conjecture (still open)

If $k \geq 1$, M is the $\binom{d}{K} \times n$ conjunction matrix of a random data base D, and $\binom{d}{K} \geq C(K)n$, then

$$s_n(D) \sim_k d^{K/2} \quad \text{with high probability, whenever } n \geq n(K)$$

Theorem

let D be an $d \times n$ random data base. Let M be the K-conjunction matrix of D.

$$\textit{If} \qquad n \leq \frac{c'}{\log^{c(K)} d} \cdot d^K, \quad \textit{then}$$

Higher order conjunctions

Conjecture (still open)

If $k \geq 1$, M is the $\binom{d}{K} \times n$ conjunction matrix of a random data base D, and $\binom{d}{K} \geq C(K)n$, then

$$s_n(D) \sim_k d^{K/2} \quad \text{with high probability, whenever } n \geq n(K)$$

Theorem

let D be an $d \times n$ random data base. Let M be the K-conjunction matrix of D.

$$\textit{If} \qquad n \leq \frac{c'}{\log^{c(K)} d} \cdot d^K, \quad \textit{then}$$

$$\mathbb{P}\left(s_n(M) \leq C_K \frac{d^{K/2}}{\log^{c_K} n}\right) \leq \exp\left(-C'_K \frac{d}{\log^{c'_K} n}\right), \quad \textit{provided that } n \textit{ is big enough.}$$

Conjunctions of order 2

Iterated logarithm: $\log^{(q)}, q \in \mathbb{N}$.

1. $\log^{(1)} x = \max(\log x, 1)$.
2. $\log^{(k+1)} x = \max(\log \log^{(k)} x, 1)$.

Conjunctions of order 2

Iterated logarithm: $\log^{(q)}, q \in \mathbb{N}$.

1. $\log^{(1)} x = \max(\log x, 1)$.
2. $\log^{(k+1)} x = \max(\log \log^{(k)} x, 1)$.

Theorem ($k = 2$)

let D be an $d \times n$ random data base. Let M be the 2-conjunction matrix of D.

$$\textit{If} \qquad n \le \frac{c'}{\log^{(q)} d} \cdot d^2,$$

Conjunctions of order 2

Iterated logarithm: $\log^{(q)}, q \in \mathbb{N}$.

1. $\log^{(1)} x = \max(\log x, 1)$.
2. $\log^{(k+1)} x = \max(\log \log^{(k)} x, 1)$.

Theorem ($k = 2$)

let D be an $d \times n$ random data base. Let M be the 2-conjunction matrix of D.

$$\textit{If} \qquad n \le \frac{c'}{\log^{(q)} d} \cdot d^2,$$

$$\textit{then} \quad \mathbb{P}\left(s_n(M) \le c^q d\right) \le e^{-cd}, \quad \textit{provided that } n \textit{ is big enough.}$$

ε-net argument for matrices with independent entries.

$$s_n(D) = \min_{x \in S^{n-1}} \|Dx\| .$$

1. Individual estimate: $\mathbb{P}(\|Dy\| < t)$ is small for a fixed $y \in S^{n-1}$.

2. Discretization: Find a small ε-net $\mathcal{N} \subset S^{n-1}$ and use the union bound.

3. Approximation:

ε-net argument for matrices with independent entries.

$$s_n(D) = \min_{x \in S^{n-1}} \|Dx\| .$$

1. Individual estimate: $\mathbb{P}(\|Dy\| < t)$ is small for a fixed $y \in S^{n-1}$.
   1. Each coordinate of Dy is a linear combination of independent random variables.
   2. Small ball probability: $\mathbb{P}\left(|(Dy)_j| < \mu\right) < \nu$ for some $\mu, \nu < 1$.
   3. Rows are independent
   $$\mathbb{P}\left(\|Dy\| < \mu' \sqrt{d}\right) \leq \eta^d .$$
2. Discretization: Find a small ε-net $\mathcal{N} \subset S^{n-1}$ and use the union bound.
3. Approximation:

ε-net argument for matrices with independent entries.

$$s_n(D) = \min_{x \in S^{n-1}} \|Dx\| .$$

1. Individual estimate: $\mathbb{P}(\|Dy\| < t)$ is small for a fixed $y \in S^{n-1}$.
   1. Each coordinate of Dy is a linear combination of independent random variables.
   2. Small ball probability: $\mathbb{P}\left(|(Dy)_j| < \mu\right) < \nu$ for some $\mu, \nu < 1$.
   3. Rows are independent

$$\mathbb{P}(\|Dy\| < \mu' \sqrt{d}) \le \eta^d.$$

2. Discretization: Find a small ε-net $\mathcal{N} \subset S^{n-1}$ and use the union bound.

$$\text{Volumetric estimate :} \qquad |\mathcal{N}| \le (3/\varepsilon)^n.$$

Union bound:

$$\mathbb{P}(\exists y \in \mathcal{N} \ \ \|Dy\| < \mu' \sqrt{d}) \le \eta^d \cdot (3/\varepsilon)^n.$$

Since $d \gg n$, this probability is very small.

3. Approximation:

ε-net argument for matrices with independent entries.

$$s_n(D) = \min_{x \in S^{n-1}} \|Dx\| .$$

1. Individual estimate: $\mathbb{P}(\|Dy\| < t)$ is small for a fixed $y \in S^{n-1}$.
   1. Each coordinate of Dy is a linear combination of independent random variables.
   2. Small ball probability: $\mathbb{P}\left(|(Dy)_j| < \mu\right) < \nu$ for some $\mu, \nu < 1$.
   3. Rows are independent

$$\mathbb{P}(\|Dy\| < \mu' \sqrt{d}) \leq \eta^d .$$

2. Discretization: Find a small ε-net $\mathcal{N} \subset S^{n-1}$ and use the union bound.

$$\text{Volumetric estimate} : \qquad |\mathcal{N}| \leq (3/\varepsilon)^n .$$

Union bound:

$$\mathbb{P}(\exists y \in \mathcal{N} \;\; \|Dy\| < \mu' \sqrt{d}) \leq \eta^d \cdot (3/\varepsilon)^n .$$

Since $d \gg n$, this probability is very small.

3. Approximation: assume that $\|Dy\| \geq \mu' \sqrt{d}$ for all $y \in \mathcal{N}$.
   Then $\|Dx\| \geq \mu' \sqrt{d}/2$ for all $x \in S^{n-1}$.

ε-net argument: a failing attempt.

1. Individual estimate: $\mathbb{P}\left(\|My\| < t\right)$ is small for a fixed $y \in S^{n-1}$.

2. Discretization (geometry): Find a small ε-net $\mathcal{N} \subset S^{n-1}$.

3. Approximation

ε-net argument: a failing attempt.

1. Individual estimate: $\mathbb{P}\left(\|My\| < t\right)$ is small for a fixed $y \in S^{n-1}$.
   1. Each coordinate of Dx is a linear combination of independent random variables.
   2. Small ball probability: $\mathbb{P}\left(|(My)_j| < \mu\right) < \nu$ for some $\mu, \nu < 1$.
   3. The coordinates of Mx are dependent $\Rightarrow$ one cannot bound $\mathbb{P}\left(\|My\| < \mu' d\right)$
2. Discretization (geometry): Find a small ε-net $\mathcal{N} \subset S^{n-1}$.

3. Approximation

Obstacles:

- Insufficient independence.

ε-net argument: a failing attempt.

1. Individual estimate: $\mathbb{P}\left(\|My\| < t\right)$ is small for a fixed $y \in S^{n-1}$.
   1. Each coordinate of Dx is a linear combination of independent random variables.
   2. Small ball probability: $\mathbb{P}\left(|(My)_j| < \mu\right) < \nu$ for some $\mu, \nu < 1$.
   3. The coordinates of Mx are dependent $\Rightarrow$ one cannot bound $\mathbb{P}\left(\|My\| < \mu' d\right)$
2. Discretization (geometry): Find a small ε-net $\mathcal{N} \subset S^{n-1}$.

$$\text{Volumetric estimate}: \qquad |\mathcal{N}| \le (3/\varepsilon)^n.$$

   Union bound:

$$\mathbb{P}\left(\exists y \in \mathcal{N} \ \ \|My\| < \mu' \sqrt{d}\right) \le \eta^d \cdot (3/\varepsilon)^n$$

3. Approximation

Obstacles:

- Insufficient independence.

ε-net argument: a failing attempt.

1. Individual estimate: $\mathbb{P}\left(\|My\| < t\right)$ is small for a fixed $y \in S^{n-1}$.
    1. Each coordinate of Dx is a linear combination of independent random variables.
    2. Small ball probability: $\mathbb{P}\left(|(My)_j| < \mu\right) < \nu$ for some $\mu, \nu < 1$.
    3. The coordinates of Mx are dependent $\Rightarrow$ one cannot bound $\mathbb{P}\left(\|My\| < \mu' d\right)$
2. Discretization (geometry): Find a small ε-net $\mathcal{N} \subset S^{n-1}$.

$$\text{Volumetric estimate :} \qquad |\mathcal{N}| \leq (3/\varepsilon)^n.$$

Union bound:

$$\mathbb{P}\left(\exists y \in \mathcal{N} \;\; \|My\| < \mu' \sqrt{d}\right) \leq \eta^d \cdot (3/\varepsilon)^n \gg 1$$

- If $n \gg d$ this is too big.
- Volumetric estimate cannot be significantly improved.

3. Approximation ???

Obstacles:

- Insufficient independence.
- Insufficient randomness.

Strategy of the proof

1. Decoupling = boosting the independence.
2. Decomposition of the sphere.
3. Balancing the small ball probability and the complexity of the set for each part separately.

Decoupling = boosting the independence.

The coordinates of Mx are dependent

- Let M' be a matrix consisting of a part of the rows of M.
  Then $s_n(M') \leq s_n(M)$.
- Divide $\{1, \ldots, d\}$ in two parts I, J of approximately equal size.
- Consider the matrix M' with rows $d_i \odot d_j$, where $i \in I, \ j \in J$.
  M corresponds to a complete graph, M' corresponds to a complete bipartite graph

$$M = \begin{pmatrix} * & * & \dots & * \\ * & * & \dots & * \\ * & * & \dots & * \\ * & * & \dots & * \\ * & * & \dots & * \\ * & * & \dots & * \end{pmatrix}$$

Decoupling = boosting the independence.

The coordinates of Mx are dependent

- Let M' be a matrix consisting of a part of the rows of M.
  Then $s_n(M') \le s_n(M)$.
- Divide $\{1, \dots, d\}$ in two parts I, J of approximately equal size.
- Consider the matrix M' with rows $d_i \odot d_j$, where $i \in I,\ j \in J$.
  M corresponds to a complete graph, M' corresponds to a complete bipartite graph

$$M = \begin{pmatrix} * & * & \dots & * \\ * & * & \dots & * \\ * & * & \dots & * \\ * & * & \dots & * \\ * & * & \dots & * \\ * & * & \dots & * \end{pmatrix} \to \begin{pmatrix} * \cdot * & * \cdot * & \dots & * \cdot * \\ * \cdot * & * \cdot * & \dots & * \cdot * \\ * \cdot * & * \cdot * & \dots & * \cdot * \\ * & * & \dots & * \\ * & * & \dots & * \\ * & * & \dots & * \\ * & * & \dots & * \\ * & * & \dots & * \\ * & * & \dots & * \end{pmatrix}$$

Decoupling = boosting the independence.

The coordinates of Mx are dependent

- Let M' be a matrix consisting of a part of the rows of M.
  Then $s_n(M') \leq s_n(M)$.
- Divide $\{1, \dots, d\}$ in two parts I, J of approximately equal size.
- Consider the matrix M' with rows $d_i \odot d_j$, where $i \in I,\ j \in J$.
  M corresponds to a complete graph, M' corresponds to a complete bipartite graph

$$M = \begin{pmatrix} * & * & \dots & * \\ * & * & \dots & * \\ * & * & \dots & * \\ * & * & \dots & * \\ * & * & \dots & * \\ * & * & \dots & * \end{pmatrix} \to \begin{pmatrix} * & * & \dots & * \\ * & * & \dots & * \\ * & * & \dots & * \\ * \cdot * & * \cdot * & \dots & * \cdot * \\ * \cdot * & * \cdot * & \dots & * \cdot * \\ * \cdot * & * \cdot * & \dots & * \cdot * \\ * & * & \dots & * \\ * & * & \dots & * \\ * & * & \dots & * \end{pmatrix}$$

Decoupling = boosting the independence.

The coordinates of Mx are dependent

- Let M' be a matrix consisting of a part of the rows of M.
  Then $s_n(M') \leq s_n(M)$.
- Divide $\{1, \dots, d\}$ in two parts I, J of approximately equal size.
- Consider the matrix M' with rows $d_i \odot d_j$, where $i \in I,\ j \in J$.
  M corresponds to a complete graph, M' corresponds to a complete bipartite graph

$$M = \begin{pmatrix} * & * & \dots & * \\ * & * & \dots & * \\ * & * & \dots & * \\ * & * & \dots & * \\ * & * & \dots & * \\ * & * & \dots & * \end{pmatrix} \to \begin{pmatrix} * & * & \dots & * \\ * & * & \dots & * \\ * & * & \dots & * \\ * & * & \dots & * \\ * & * & \dots & * \\ * & * & \dots & * \\ * \cdot * & * \cdot * & \dots & * \cdot * \\ * \cdot * & * \cdot * & \dots & * \cdot * \\ * \cdot * & * \cdot * & \dots & * \cdot * \end{pmatrix}$$

Decoupling = boosting the independence.

The coordinates of Mx are dependent

- Condition on $d_j,\ j \in J$.
  The matrix M' consists of $d/2$ blocks, which are essentially independent.
- Improvement: more independence.
- Complication: independent entries $\Rightarrow$ independent blocks.

$$M = \begin{pmatrix} * & * & \dots & * \\ * & * & \dots & * \\ * & * & \dots & * \\ * & * & \dots & * \\ * & * & \dots & * \\ * & * & \dots & * \end{pmatrix} \rightarrow \begin{pmatrix} * & * & \dots & * \\ * & * & \dots & * \\ * & * & \dots & * \\ * & * & \dots & * \\ * & * & \dots & * \\ * & * & \dots & * \\ * \cdot * & * \cdot * & \dots & * \cdot * \\ * \cdot * & * \cdot * & \dots & * \cdot * \\ * \cdot * & * \cdot * & \dots & * \cdot * \end{pmatrix}$$

Small ball probability and conditioning: two sides of one coin

- $M' = \Pi_1 \odot \Pi_2$, where Π_1, Π_2 are independent random matrices.
- We need to bound $\mathbb{P}\left(\|(\Pi_1 \odot \Pi_2)x\| \text{ is small}\right)$.

Small ball probability and conditioning: two sides of one coin

- $M' = \Pi_1 \odot \Pi_2$, where Π_1, Π_2 are independent random matrices.
- We need to bound $\mathbb{P}\left(\|(\Pi_1 \odot \Pi_2)x\| \text{ is small}\right)$.

1. Condition on Π_2.

2. Condition on Π_1.

Small ball probability and conditioning: two sides of one coin

- $M' = \Pi_1 \odot \Pi_2$, where Π_1, Π_2 are independent random matrices.
- We need to bound $\mathbb{P}\left(\|(\Pi_1 \odot \Pi_2)x\| \text{ is small}\right)$.

1. Condition on Π_2.
   - $(\Pi_1 \odot \Pi_2)x$ consists of d blocks $\Pi_1 y_j$, $j = 1, \dots, d$, where y_j is the row product of the j-th row of Π_2 and x.
   - $\|y_j\| \geq c\,\|x\|$ with high probability $\Rightarrow$ $\|\Pi_1 y_j\| \gtrsim \sqrt{d}$
     $\Rightarrow$ $\|(\Pi_1 \odot \Pi_2)x\| \gtrsim d$ with high probability.

2. Condition on Π_1.

Small ball probability and conditioning: two sides of one coin

- $M' = \Pi_1 \odot \Pi_2$, where Π_1, Π_2 are independent random matrices.
- We need to bound $\mathbb{P}\left(\|(\Pi_1 \odot \Pi_2)x\| \text{ is small}\right)$.

1. Condition on Π_2.
   - $(\Pi_1 \odot \Pi_2)x$ consists of d blocks $\Pi_1 y_j$, $j = 1, \ldots, d$, where y_j is the row product of the j-th row of Π_2 and x.
   - $\|y_j\| \geq c\,\|x\|$ with high probability $\Rightarrow$ $\|\Pi_1 y_j\| \gtrsim \sqrt{d}$
     $\Rightarrow$ $\|(\Pi_1 \odot \Pi_2)x\| \gtrsim d$ with high probability.

2. Condition on Π_1.
   - Discarding a set of small probability, we may assume that Π_1 is typical.
   - $\|(\Pi_1 \odot \Pi_2)x\|$ is highly concentrated around its mean
     $\Rightarrow$ $\|(\Pi_1 \odot \Pi_2)x\| \gtrsim d$ with high probability.

Small ball probability and conditioning: two sides of one coin

- $M' = \Pi_1 \odot \Pi_2$, where Π_1, Π_2 are independent random matrices.
- We need to bound $\mathbb{P}\left(\|(\Pi_1 \odot \Pi_2)x\| \text{ is small}\right)$.

1. Condition on Π_2.
   - $(\Pi_1 \odot \Pi_2)x$ consists of d blocks $\Pi_1 y_j$, $j = 1, \ldots, d$, where y_j is the row product of the j-th row of Π_2 and x.
   - $\|y_j\| \geq c\,\|x\|$ with high probability $\Rightarrow$ $\|\Pi_1 y_j\| \gtrsim \sqrt{d}$
     $\Rightarrow$ $\|(\Pi_1 \odot \Pi_2)x\| \gtrsim d$ with high probability.

2. Condition on Π_1.
   - Discarding a set of small probability, we may assume that Π_1 is typical.
   - $\|(\Pi_1 \odot \Pi_2)x\|$ is highly concentrated around its mean
     $\Rightarrow$ $\|(\Pi_1 \odot \Pi_2)x\| \gtrsim d$ with high probability.

Which strategy is better?

Small ball probability and conditioning: two sides of one coin

- $M' = \Pi_1 \odot \Pi_2$, where Π_1, Π_2 are independent random matrices.
- We need to bound $\mathbb{P}\left(\|(\Pi_1 \odot \Pi_2)x\| \text{ is small}\right)$.

1. Condition on Π_2.
   - $(\Pi_1 \odot \Pi_2)x$ consists of d blocks $\Pi_1 y_j$, $j = 1, \dots, d$, where y_j is the row product of the j-th row of Π_2 and x.
   - $\|y_j\| \geq c\,\|x\|$ with high probability $\Rightarrow$ $\|\Pi_1 y_j\| \gtrsim \sqrt{d}$
     $\Rightarrow$ $\|(\Pi_1 \odot \Pi_2)x\| \gtrsim d$ with high probability.
   - Works when $|\operatorname{supp} x| \ll d$.
2. Condition on Π_1.
   - Discarding a set of small probability, we may assume that Π_1 is typical.
   - $\|(\Pi_1 \odot \Pi_2)x\|$ is highly concentrated around its mean
     $\Rightarrow$ $\|(\Pi_1 \odot \Pi_2)x\| \gtrsim d$ with high probability.
   - Works if x has many commensurate coordinates.

Which strategy is better?

Analysis on one block: random sums of random vectors

Condition on the matrix Π_1.
Let B be one of the blocks:

$$B = \varepsilon \odot \Pi_1 = \left[\varepsilon_1 \cdot Y_1, \varepsilon_2 \cdot Y_2, \dots, \varepsilon_n \cdot Y_N\right]$$

- Y_j is a column of Π_1 (fixed after conditioning)
- $\varepsilon_1, \dots, \varepsilon_n$ are independent ± 1 random variables.

Individual estimate: we have to bound $\|Bx\|$ below.

Analysis on one block: random sums of random vectors

Condition on the matrix Π_1.
Let B be one of the blocks:

$$B = \varepsilon \odot \Pi_1 = \left[\varepsilon_1 \cdot Y_1, \varepsilon_2 \cdot Y_2, \ldots, \varepsilon_n \cdot Y_N\right]$$

- Y_j is a column of Π_1 (fixed after conditioning)
- $\varepsilon_1, \ldots, \varepsilon_n$ are independent ± 1 random variables.

Individual estimate: we have to bound $\|Bx\|$ below.
Trick: interchanging the roles of ε and x:

$$Bx = \left[\varepsilon_1 Y_1, \varepsilon_2 Y_2, \ldots, \varepsilon_n Y_n\right] \cdot \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix}$$

Analysis on one block: random sums of random vectors

Condition on the matrix Π_1.
Let B be one of the blocks:

$$B = \varepsilon \odot \Pi_1 = \left[\varepsilon_1 \cdot Y_1, \varepsilon_2 \cdot Y_2, \ldots, \varepsilon_n \cdot Y_N\right]$$

- Y_j is a column of Π_1 (fixed after conditioning)
- $\varepsilon_1, \ldots, \varepsilon_n$ are independent ± 1 random variables.

Individual estimate: we have to bound $\|Bx\|$ below.
Trick: interchanging the roles of ε and x:

$$Bx = \left[\varepsilon_1 Y_1, \varepsilon_2 Y_2, \ldots, \varepsilon_n Y_n\right] \cdot \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix} = \left[x_1 Y_1, x_2 Y_2, \ldots, x_n Y_n\right] \cdot \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_n \end{bmatrix}$$

This is a sum of vectors $x_1 Y_1, x_2 Y_2, \ldots, x_n Y_n$ with random signs.

Random sums of vectors

Let $V_1, V_2, \ldots, V_n \in \mathbb{R}^d$ be fixed vectors.
We want to show that

$$\mathbb{P}\left(\left\| \sum_{j=1}^{n} \varepsilon_j V_j \right\| \leq ? \right) \leq ??$$

Parallelogram equality:

$$\left\| \sum_{j=1}^{n} \varepsilon_j V_j \right\|^2 = \sum_{j=1}^{n} \|V_j\|^2$$

We need estimate with high probability.

Random sums of vectors

Let $V_1, V_2, \ldots, V_n \in \mathbb{R}^d$ be fixed vectors.
We want to show that

$$\mathbb{P}\left(\left\|\sum_{j=1}^{n} \varepsilon_j V_j\right\| \le ?\right) \le ??$$

Parallelogram equality:

$$\left\|\sum_{j=1}^{n} \varepsilon_j V_j\right\|^2 = \sum_{j=1}^{n} \|V_j\|^2$$

We need estimate with high probability.
Impossible in general: if $V_1 = V_2$, and $V_3 = \ldots = V_n = 0$, then

$$\mathbb{P}\left(\left\|\sum_{j=1}^{n} \varepsilon_j V_j\right\| = 0\right) = 1/2.$$

Concentration of measure

Let $V_1, V_2, \dots, V_n \in \mathbb{R}^d$ be fixed vectors.
We want to show that

$$\mathbb{P}\left(\left\| \sum_{j=1}^n \varepsilon_j V_j \right\| \le \left(\sum_{j=1}^n \|V_j\|^2 \right)^{1/2} \right) \le ??$$

1. View $F(\varepsilon_1, \dots, \varepsilon_n) = \left\| \sum_{j=1}^n \varepsilon_j V_j \right\|$ as a function on $\mathbb{R}^n$, and on the discrete cube $\{-1, 1\}^n$ simultaneously.
2. Talagrand's measure concentration theorem:
   Every convex Lipschitz function $F : \mathbb{R}^n \to \mathbb{R}$ is close to a constant on the discrete cube with high probability.
   (How close depends on the Lipschitz constant of F)

Lipschitz constant of F is the norm of the matrix $[V_1, V_2, \dots, V_n]$.

To get a meaningful estimate, we need $\| [V_1, V_2, \dots, V_n] \| \ll \left(\sum_{j=1}^n \|V_j\|^2 \right)^{1/2}$.

Stratification of the sphere

When $\left\| \left[V_1, V_2, \ldots, V_n \right] \right\| \ll \left(\sum_{j=1}^{n} \|V_j\|^2 \right)^{1/2}$?

- If $V_1 = V_2 = \ldots = V_n$, then "$=$".
- If $\|V_1\| \gg \|V_j\|$ for all $j > 1$, then "$\approx$".
- We need V_j to be independent vectors of commensurate norms.

Stratification of the sphere

When $\left\| [V_1, V_2, \ldots, V_n] \right\| \ll \left(\sum_{j=1}^{n} \|V_j\|^2 \right)^{1/2}$?

- If $V_1 = V_2 = \ldots = V_n$, then "$=$".
- If $\|V_1\| \gg \|V_j\|$ for all $j > 1$, then "$\approx$".
- We need V_j to be independent vectors of commensurate norms.

$$[V_1, V_2, \ldots, V_n] = [x_1 Y_1, x_2 Y_2, \ldots, x_n Y_n].$$

- Independence – yes
- commensurate norms – not for any x

Stratification of the sphere

When $\left\| \left[V_1, V_2, \ldots, V_n \right] \right\| \ll \left(\sum_{j=1}^{n} \|V_j\|^2 \right)^{1/2}$?

- If $V_1 = V_2 = \ldots = V_n$, then "$=$".
- If $\|V_1\| \gg \|V_j\|$ for all $j > 1$, then "$\approx$".
- We need V_j to be independent vectors of commensurate norms.

$$\left[V_1, V_2, \ldots, V_n \right] = \left[x_1 Y_1, x_2 Y_2, \ldots, x_n Y_n \right].$$

- Independence – yes
- commensurate norms – not for any x
- x is far from a coordinate subspace of a small dimension (incompressible) $\Rightarrow$ many commensurate coordinates.