Matrix Concentration Inequalities

Afonso S. Bandeira
ETH Zürich

MoRAT 2026

PART I

Classical Matrix Concentration

Scalar concentration: Markov + independence

Let $H_1, \ldots, H_N$ be independent real random variables and

$$X = \sum_{k=1}^{N} H_k.$$

Scalar concentration: Markov + independence

Let $H_1, \ldots, H_N$ be independent real random variables and

$$X = \sum_{k=1}^{N} H_k.$$

For every $\theta > 0$,

$$\begin{aligned}
\mathbb{P}(X \geq t) &= \mathbb{P}\big(e^{\theta X} \geq e^{\theta t}\big) \\
&\leq e^{-\theta t}\, \mathbb{E}\, e^{\theta X} \\
&= e^{-\theta t}\, \mathbb{E} \prod_{k=1}^{N} e^{\theta H_k} = e^{-\theta t} \prod_{k=1}^{N} \mathbb{E}\, e^{\theta H_k}.
\end{aligned}$$

Scalar concentration: Markov $+$ independence

Let $H_1, \ldots, H_N$ be independent real random variables and

$$X = \sum_{k=1}^{N} H_k.$$

For every $\theta > 0$,

$$\begin{aligned} \mathbb{P}(X \geq t) &= \mathbb{P}\big(e^{\theta X} \geq e^{\theta t}\big) \\ &\leq e^{-\theta t}\, \mathbb{E}\, e^{\theta X} \\ &= e^{-\theta t}\, \mathbb{E} \prod_{k=1}^{N} e^{\theta H_k} = e^{-\theta t} \prod_{k=1}^{N} \mathbb{E}\, e^{\theta H_k}. \end{aligned}$$

$$\boxed{\mathbb{P}(X \geq t) \leq \inf_{\theta>0} \exp\left\{-\theta t + \sum_{k=1}^{N} \log \mathbb{E}\, e^{\theta H_k}\right\}.}$$

Scalar Bernstein

Assume

$$\mathbb{E}\, H_k = 0, \qquad |H_k| \leq R, \qquad \sigma^2 := \sum_{k=1}^{N} \mathbb{E}\, H_k^2.$$

Scalar Bernstein

Assume

$$\mathbb{E}\, H_k = 0, \qquad |H_k| \le R, \qquad \sigma^2 := \sum_{k=1}^{N} \mathbb{E}\, H_k^2.$$

For $\theta \ge 0$ and $x \le R$,

$$e^{\theta x} \le 1 + \theta x + g(\theta) x^2, \qquad g(\theta) := \frac{e^{\theta R} - \theta R - 1}{R^2},$$

so (recall $\log(1+t) \le t$)

$$\log \mathbb{E}\, e^{\theta H_k} \le g(\theta)\, \mathbb{E}\, H_k^2.$$

Scalar Bernstein

Assume

$$\mathbb{E}\, H_k = 0, \qquad |H_k| \le R, \qquad \sigma^2 := \sum_{k=1}^{N} \mathbb{E}\, H_k^2.$$

For $\theta \ge 0$ and $x \le R$,

$$e^{\theta x} \le 1 + \theta x + g(\theta) x^2, \qquad g(\theta) := \frac{e^{\theta R} - \theta R - 1}{R^2},$$

so (recall $\log(1+t) \le t$)

$$\log \mathbb{E}\, e^{\theta H_k} \le g(\theta)\, \mathbb{E}\, H_k^2.$$

optimize in θ, and use the same argument to $-X$:

$$\boxed{\mathbb{P}\big(\,|X| \ge t\big) \le 2 \exp\left\{ -\frac{t^2}{2\sigma^2 + \frac{2}{3} R t} \right\}.}$$

For matrices, the product step breaks

Let $H_1, \ldots, H_N$ be independent Hermitian matrices and $X = \sum_k H_k$.

For matrices, the product step breaks

Let $H_1, \ldots, H_N$ be independent Hermitian matrices and $X = \sum_k H_k$.

$$e^{\theta X} \neq \prod_k e^{\theta H_k} \qquad \text{unless the summands commute.}$$

For matrices, the product step breaks

Let $H_1, \ldots, H_N$ be independent Hermitian matrices and $X = \sum_k H_k$.

$$e^{\theta X} \neq \prod_k e^{\theta H_k} \qquad \text{unless the summands commute.}$$

Markov still works. For either sign $\eta \in \{\pm 1\}$,

$$\mathbb{P}(\lambda_{\max}(\eta X) \geq t) \leq \inf_{\theta > 0} e^{-\theta t}\, \mathbb{E}\, \mathrm{Tr} e^{\eta \theta X}.$$

Therefore

$$\mathbb{P}(\|X\| \geq t) \leq \sum_{\eta \in \{\pm 1\}} \inf_{\theta > 0} e^{-\theta t}\, \mathbb{E}\, \mathrm{Tr} e^{\eta \theta X}.$$

This replaces the "Chernoff trick".

How to go from $\exp$ of sum to product of $\exp$?

Theorem (Lieb, 1973)

For fixed Hermitian A, the map

$$M \longmapsto \mathrm{Tr}\exp\left(A + \log M\right)$$

is concave on positive definite matrices.

E. H. Lieb, *Convex trace functions and the Wigner–Yanase–Dyson conjecture*, 1973.

How to go from exp of sum to product of exp?

Theorem (Lieb, 1973)

For fixed Hermitian A, the map

$$M \longmapsto \operatorname{Tr} \exp\left(A + \log M\right)$$

is concave on positive definite matrices.

Jensen's inequality, with $M = e^H$, gives

$$\boxed{\mathop{\mathbb{E}}_{H} \operatorname{Tr} e^{A+H} \leq \operatorname{Tr} \exp\left(A + \log \mathbb{E}\, e^H\right).}$$

This is the non-commutative replacement for "exp of sum is product of exp"

E. H. Lieb, *Convex trace functions and the Wigner–Yanase–Dyson conjecture*, 1973.

Iterate Lieb, then control one summand

For independent Hermitian $H_1, \ldots, H_N$,

$$\boxed{\mathbb{E} \operatorname{Tr} \exp\left(\theta \sum_{k=1}^{N} H_k\right) \le \operatorname{Tr} \exp\left(\sum_{k=1}^{N} \log \mathbb{E}\, e^{\theta H_k}\right).}$$

Iterate Lieb, then control one summand

For independent Hermitian $H_1, \ldots, H_N$,

$$\boxed{\mathbb{E}\,\mathrm{Tr}\exp\left(\theta\sum_{k=1}^{N} H_k\right) \le \mathrm{Tr}\exp\left(\sum_{k=1}^{N}\log\mathbb{E}\,e^{\theta H_k}\right).}$$

If $\mathbb{E}\,H_k = 0$ and $\|H_k\| \le R$, an analogue to the the scalar argument

(+ operator monotonicity of the logarithm) yields

$$\boxed{\log\mathbb{E}\,e^{\pm\theta H_k} \preceq g(\theta)\;\mathbb{E}\,H_k^2, \qquad g(\theta) = \frac{e^{\theta R} - \theta R - 1}{R^2}.}$$

Iterate Lieb, then control one summand

For independent Hermitian $H_1, \ldots, H_N$,

$$\boxed{\mathbb{E}\,\mathrm{Tr}\exp\left(\theta\sum_{k=1}^{N} H_k\right) \le \mathrm{Tr}\exp\left(\sum_{k=1}^{N}\log\mathbb{E}\,e^{\theta H_k}\right).}$$

If $\mathbb{E}\,H_k = 0$ and $\|H_k\| \le R$, an analogue to the the scalar argument
(+ operator monotonicity of the logarithm) yields

$$\boxed{\log\mathbb{E}\,e^{\pm\theta H_k} \preceq g(\theta)\;\mathbb{E}\,H_k^2, \qquad g(\theta) = \frac{e^{\theta R} - \theta R - 1}{R^2}.}$$

Set $\sigma^2 = \left\|\sum_k \mathbb{E}\,H_k^2\right\|$. Then

$$\mathbb{E}\,\mathrm{Tr}e^{\pm\theta X} \le d\,e^{g(\theta)\sigma^2}.$$

Matrix Bernstein

Let $H_1, \ldots, H_N$ be independent centered Hermitian $d \times d$ matrices,

$$X = \sum_{k=1}^{N} H_k, \qquad \|H_k\| \le R, \qquad \sigma^2 = \left\| \sum_{k=1}^{N} \mathbb{E}\, H_k^2 \right\|.$$

Matrix concentration in this form originates with Ahlswede–Winter, Oliveira, and Tropp. See Tropp'15 monograph

A somewhat different treatment is available in "Mathematics of Data Science" by B-Singer-Strohmer

Matrix Bernstein

Let $H_1, \ldots, H_N$ be independent centered Hermitian $d \times d$ matrices,

$$X = \sum_{k=1}^{N} H_k, \qquad \|H_k\| \le R, \qquad \sigma^2 = \left\| \sum_{k=1}^{N} \mathbb{E}\, H_k^2 \right\|.$$

The same scalar optimization gives, for every $t \ge 0$,

$$\boxed{\mathbb{P}\big(\|X\| \ge t\big) \le 2d \exp\left\{ -\frac{t^2}{2\sigma^2 + \frac{2}{3}Rt} \right\}.}$$

Matrix concentration in this form originates with Ahlswede–Winter, Oliveira, and Tropp. See Tropp'15 monograph

A somewhat different treatment is available in "Mathematics of Data Science" by B-Singer-Strohmer

Matrix Bernstein

Let $H_1, \ldots, H_N$ be independent centered Hermitian $d \times d$ matrices,

$$X = \sum_{k=1}^{N} H_k, \qquad \|H_k\| \le R, \qquad \sigma^2 = \left\| \sum_{k=1}^{N} \mathbb{E}\, H_k^2 \right\|.$$

The same scalar optimization gives, for every $t \ge 0$,

$$\boxed{\mathbb{P}(\, \|X\| \ge t) \le 2d \exp\left\{ -\frac{t^2}{2\sigma^2 + \frac{2}{3}Rt} \right\}}.$$

For rectangular sums, one can apply the theorem to

$$\mathcal{H}(Z) = \begin{bmatrix} 0 & Z \\ Z^* & 0 \end{bmatrix}, \qquad \|\mathcal{H}(Z)\| = \|Z\|\,.$$

Matrix concentration in this form originates with Ahlswede–Winter, Oliveira, and Tropp. See Tropp'15 monograph

A somewhat different treatment is available in "Mathematics of Data Science" by B-Singer-Strohmer

Matrix Bernstein

Let $H_1, \ldots, H_N$ be independent centered Hermitian $d \times d$ matrices,

$$X = \sum_{k=1}^{N} H_k, \qquad \|H_k\| \le R, \qquad \sigma^2 = \left\| \sum_{k=1}^{N} \mathbb{E}\, H_k^2 \right\|.$$

The same scalar optimization gives, for every $t \ge 0$,

$$\boxed{\mathbb{P}(\, \|X\| \ge t) \le 2d \exp\left\{ -\frac{t^2}{2\sigma^2 + \frac{2}{3} R t} \right\}}.$$

For rectangular sums, one can apply the theorem to

$$\mathcal{H}(Z) = \begin{bmatrix} 0 & Z \\ Z^* & 0 \end{bmatrix}, \qquad \|\mathcal{H}(Z)\| = \|Z\|\,.$$

Matrix Bernstein shows: $\mathbb{E}\, \|\mathbf{X}\| \lesssim \sigma \sqrt{\log(\mathbf{2d})} + \mathbf{R} \log(\mathbf{2d})$

(and can be combined with scalar concentration for tails)

Matrix concentration in this form originates with Ahlswede–Winter, Oliveira, and Tropp. See Tropp'15 monograph

A somewhat different treatment is available in "Mathematics of Data Science" by B-Singer-Strohmer

PART II

Structured OSE / OSI

Background: Martinsson–Tropp, *Acta Numerica* 2020.
OSI: Camaño–Epperly–Meyer–Tropp, 2025.

See also: Bandeira–Singer–Strohmer, *Mathematics of Data Science*, arXiv:2607.11938.

The matrix to control

Fix an arbitrary orthonormal basis of the subspace and a random sketch:

$$U \in \mathbb{R}^{n\times k}, \qquad U^\top U = I_k, \qquad \Omega \in \mathbb{R}^{n\times m}.$$

$$\boxed{\mathbb{R}^k} \xrightarrow{U} \boxed{\begin{matrix}\mathcal{U} = \operatorname{range}(U) \\ \subseteq \mathbb{R}^n\end{matrix}} \xrightarrow{\Omega^\top} \boxed{\mathbb{R}^m}$$

The matrix to control

Fix an arbitrary orthonormal basis of the subspace and a random sketch:

$$U \in \mathbb{R}^{n\times k}, \qquad U^\top U = I_k, \qquad \Omega \in \mathbb{R}^{n\times m}.$$

$$\boxed{\mathbb{R}^k} \xrightarrow{U} \boxed{\mathcal{U} = \mathrm{range}(U) \subseteq \mathbb{R}^n} \xrightarrow{\Omega^\top} \boxed{\mathbb{R}^m}$$

For every construction below, we study the same two matrices:

$$\boxed{Y := \Omega^\top U \in \mathbb{R}^{m\times k}, \qquad G := Y^\top Y = U^\top \Omega\Omega^\top U \in \mathbb{R}^{k\times k}.}$$

The matrix to control

Fix an arbitrary orthonormal basis of the subspace and a random sketch:

$$U \in \mathbb{R}^{n\times k}, \qquad U^\top U = I_k, \qquad \Omega \in \mathbb{R}^{n\times m}.$$

$$\boxed{\mathbb{R}^k} \xrightarrow{U} \boxed{\begin{array}{c}\mathcal{U} = \mathrm{range}(U) \\ \subseteq \mathbb{R}^n\end{array}} \xrightarrow{\Omega^\top} \boxed{\mathbb{R}^m}$$

For every construction below, we study the same two matrices:

$$\boxed{Y := \Omega^\top U \in \mathbb{R}^{m\times k}, \qquad G := Y^\top Y = U^\top \Omega\Omega^\top U \in \mathbb{R}^{k\times k}.}$$

(k, ε, δ)-OSE

With probability at least $1-\delta$,

$$(1-\varepsilon)I_k \preceq G \preceq (1+\varepsilon)I_k.$$

(k, ε, δ)-OSI

$$\mathbb{E}[\Omega\Omega^\top] = I_n, \qquad G \succeq (1-\varepsilon)I_k$$

with probability at least $1-\delta$.

Equivalently: two-sided norm preservation on $\mathcal{U}$.

Uniform row sampling: a very cheap embedding

$\widetilde{U} \in \mathbb{R}^{n \times k}$ with $\widetilde{U}^\top \widetilde{U} = I_k$

(we will take $\widetilde{U} = QU$ for orthogonal Q, but think of $Q = I$ for now)

Uniform row sampling: a very cheap embedding

$$\widetilde{U} \in \mathbb{R}^{n\times k} \text{ with } \widetilde{U}^\top \widetilde{U} = I_k$$

(we will take $\widetilde{U} = QU$ for orthogonal Q, but think of $Q = I$ for now)

Let $R \in \{0,1\}^{m\times n}$ sample rows $i_1, \ldots, i_m$ independently and uniformly. Define

$$\boxed{\Omega^\top := \sqrt{\frac{n}{m}} RQ, \qquad Y = \Omega^\top U = \sqrt{\frac{n}{m}} R\widetilde{U}.}$$

Uniform row sampling: a very cheap embedding

$$\widetilde{U} \in \mathbb{R}^{n\times k} \text{ with } \widetilde{U}^\top \widetilde{U} = I_k$$

(we will take $\widetilde{U} = QU$ for orthogonal Q, but think of $Q = I$ for now)

Let $R \in \{0,1\}^{m\times n}$ sample rows $i_1, \ldots, i_m$ independently and uniformly. Define

$$\boxed{\Omega^\top := \sqrt{\frac{n}{m}}\, RQ, \qquad Y = \Omega^\top U = \sqrt{\frac{n}{m}}\, R\widetilde{U}.}$$

Writing the rows as $\widetilde{u}_1^\top, \ldots, \widetilde{u}_n^\top$ gives

$$\boxed{G = Y^\top Y = \frac{n}{m}\sum_{t=1}^{m} \widetilde{u}_{i_t}\widetilde{u}_{i_t}^\top, \qquad \mathop{\mathbb{E}}_{R}[G] = I_k.}$$

Uniform row sampling: a very cheap embedding

$$\widetilde{U} \in \mathbb{R}^{n\times k} \text{ with } \widetilde{U}^\top \widetilde{U} = I_k$$

(we will take $\widetilde{U} = QU$ for orthogonal Q, but think of $Q = I$ for now)

Let $R \in \{0,1\}^{m\times n}$ sample rows $i_1, \ldots, i_m$ independently and uniformly. Define

$$\boxed{\Omega^\top := \sqrt{\frac{n}{m}}\, RQ, \qquad Y = \Omega^\top U = \sqrt{\frac{n}{m}}\, R\widetilde{U}.}$$

Writing the rows as $\widetilde{u}_1^\top, \ldots, \widetilde{u}_n^\top$ gives

$$\boxed{G = Y^\top Y = \frac{n}{m}\sum_{t=1}^{m} \widetilde{u}_{i_t}\widetilde{u}_{i_t}^\top, \qquad \mathop{\mathbb{E}}_{R}[G] = I_k.}$$

Uniform sampling will be governed by the **coherence**

$$\boxed{\mu(\widetilde{U}) := n \max_{1\le i\le n} \|\widetilde{u}_i\|_2^2, \qquad k \le \mu(\widetilde{U}) \le n.}$$

Matrix Bernstein: coherence controls row sampling

$$\boxed{G - I_k = \sum_{t=1}^{m} Z_t, \qquad Z_t := \frac{n}{m}\,\widetilde{u}_{i_t}\widetilde{u}_{i_t}^{\top} - \frac{1}{m} I_k, \qquad \mathop{\mathbb{E}}_{R}[Z_t] = 0.}$$

Matrix Bernstein: coherence controls row sampling

$$\boxed{G - I_k = \sum_{t=1}^{m} Z_t, \qquad Z_t := \frac{n}{m}\widetilde{u}_{i_t}\widetilde{u}_{i_t}^\top - \frac{1}{m} I_k, \qquad \mathop{\mathbb{E}}_{R}[Z_t] = 0.}$$

One can compute the Bernstein parameters:

$$\boxed{\|Z_t\| \le \frac{\mu(\widetilde{U})}{m}, \qquad \left\|\sum_{t=1}^{m} \mathop{\mathbb{E}}_{R} Z_t^2\right\| \le \frac{\mu(\widetilde{U})}{m}.}$$

Matrix Bernstein: coherence controls row sampling

$$\boxed{G - I_k = \sum_{t=1}^{m} Z_t, \qquad Z_t := \frac{n}{m}\,\widetilde{u}_{i_t}\widetilde{u}_{i_t}^\top - \frac{1}{m} I_k, \qquad \mathop{\mathbb{E}}_{R}[Z_t] = 0.}$$

One can compute the Bernstein parameters:

$$\boxed{\|Z_t\| \le \frac{\mu(\widetilde{U})}{m}, \qquad \left\| \sum_{t=1}^{m} \mathop{\mathbb{E}}_{R} Z_t^2 \right\| \le \frac{\mu(\widetilde{U})}{m}.}$$

For $0 < \varepsilon \le 1$, Matrix Bernstein yields

$$\boxed{\begin{gathered}\mathbb{P}_R(\|G - I_k\| \ge \varepsilon) \le 2k \exp\left\{-\frac{3m\varepsilon^2}{8\mu(\widetilde{U})}\right\}, \\ m \gtrsim \varepsilon^{-2}\mu(\widetilde{U})\log(k/\delta) \quad \Longrightarrow \quad \|G - I_k\| \le \varepsilon\end{gathered}}$$

Matrix Bernstein: coherence controls row sampling

$$\boxed{G - I_k = \sum_{t=1}^{m} Z_t, \qquad Z_t := \frac{n}{m}\widetilde{u}_{i_t}\widetilde{u}_{i_t}^\top - \frac{1}{m} I_k, \qquad \mathop{\mathbb{E}}_R [Z_t] = 0.}$$

One can compute the Bernstein parameters:

$$\boxed{\|Z_t\| \le \frac{\mu(\widetilde{U})}{m}, \qquad \left\| \sum_{t=1}^{m} \mathop{\mathbb{E}}_R Z_t^2 \right\| \le \frac{\mu(\widetilde{U})}{m}.}$$

For $0 < \varepsilon \le 1$, Matrix Bernstein yields

$$\boxed{\begin{gathered} \mathbb{P}_R(\|G - I_k\| \ge \varepsilon) \le 2k \exp\left\{ -\frac{3m\varepsilon^2}{8\mu(\widetilde{U})} \right\}, \\ m \gtrsim \varepsilon^{-2}\mu(\widetilde{U}) \log(k/\delta) \quad \Longrightarrow \quad \|G - I_k\| \le \varepsilon \end{gathered}}$$

Strategy: Find a fast, oblivious orthogonal Q that makes $\widetilde{U} = QU$ incoherent.

Orthogonality preserves: $\widetilde{U}^\top \widetilde{U} = U^\top U = I_k$.

SRHT: randomize, spread, sample

For $n = 2^q$, choose the fast orthogonal transformation $Q = HD$, where H is the normalized Walsh–Hadamard matrix and $D = \mathrm{diag}(\varepsilon_1, \ldots, \varepsilon_n)$:

$$\boxed{\Omega^\top = \sqrt{\frac{n}{m}}\, RHD, \qquad \widetilde{U} := HDU.}$$

Since H and D are orthogonal, $\widetilde{U}^\top \widetilde{U} = I_k$.

SRHT: randomize, spread, sample

For $n = 2^q$, choose the fast orthogonal transformation $Q = HD$, where H is the normalized Walsh–Hadamard matrix and $D = \mathrm{diag}(\varepsilon_1, \ldots, \varepsilon_n)$:

$$\boxed{\Omega^\top = \sqrt{\frac{n}{m}}\, RHD, \qquad \widetilde{U} := HDU.}$$

Since H and D are orthogonal, $\widetilde{U}^\top \widetilde{U} = I_k$.

$$\boxed{U} \xrightarrow{D} \boxed{DU} \xrightarrow{H} \boxed{\widetilde{U} = HDU} \xrightarrow{\sqrt{n/m}\,R} \boxed{Y = \Omega^\top U}$$

SRHT: randomize, spread, sample

For $n = 2^q$, choose the fast orthogonal transformation $Q = HD$, where H is the normalized Walsh–Hadamard matrix and $D = \operatorname{diag}(\varepsilon_1, \ldots, \varepsilon_n)$:

$$\boxed{\Omega^\top = \sqrt{\frac{n}{m}}\, RHD, \qquad \widetilde{U} := HDU.}$$

Since H and D are orthogonal, $\widetilde{U}^\top \widetilde{U} = I_k$.

$$\boxed{U} \xrightarrow{D} \boxed{DU} \xrightarrow{H} \boxed{\widetilde{U} = HDU} \xrightarrow{\sqrt{n/m}\, R} \boxed{Y = \Omega^\top U}$$

D: randomize

$$H(H^\top e_i) = e_i.$$

H alone has exceptional aligned directions.

H: spread

$$|\mathrm{supp}(x)|\; |\mathrm{supp}(Hx)| \geq n.$$

Uncertainty principle: a vector cannot be sparse in both bases.

R: compress

$$Y = \sqrt{\frac{n}{m}}\, R\widetilde{U}.$$

Bernstein applies once $\mu(\widetilde{U})$ is small.

A Gaussian probe linearizes one row

Let $u_j^\top$ be row j of U. For every fixed row i,

$$\boxed{\widetilde{u}_i^\top = e_i^\top HDU = \frac{1}{\sqrt{n}} \sum_{j=1}^{n} \eta_j u_j^\top, \qquad \eta_j \overset{\text{iid}}{\sim} \{\pm 1\}.}$$

A Gaussian probe linearizes one row

Let $u_j^\top$ be row j of U. For every fixed row i,

$$\boxed{\widetilde{u}_i^\top = e_i^\top HDU = \frac{1}{\sqrt{n}} \sum_{j=1}^{n} \eta_j u_j^\top, \qquad \eta_j \overset{\text{iid}}{\sim} \{\pm 1\}.}$$

For $g \sim \mathcal{N}(0, I_k)$ and every $a \in \mathbb{R}^k$,

$$\boxed{\mathop{\mathbb{E}}_{g} e^{\langle g, a \rangle} = e^{\|a\|_2^2/2}.}$$

A Gaussian probe linearizes one row

Let $u_j^\top$ be row j of U. For every fixed row i,

$$\boxed{\widetilde{u}_i^\top = e_i^\top HDU = \frac{1}{\sqrt{n}} \sum_{j=1}^{n} \eta_j u_j^\top, \qquad \eta_j \overset{\text{iid}}{\sim} \{\pm 1\}.}$$

For $g \sim \mathcal{N}(0, I_k)$ and every $a \in \mathbb{R}^k$,

$$\boxed{\mathbb{E}_g e^{\langle g, a \rangle} = e^{\|a\|_2^2/2}.}$$

Taking $a = \sqrt{n/2}\,\widetilde{u}_i$ gives the extra linearization step

$$\boxed{e^{\frac{n}{4}\left\|\widetilde{u}_i\right\|_2^2} = \mathbb{E}_g e^{\sqrt{n/2}\left\langle g, \widetilde{u}_i \right\rangle} = \mathbb{E}_g \exp\left\{ \frac{1}{\sqrt{2}} \sum_{j=1}^{n} \eta_j \langle g, u_j \rangle \right\}.}$$

The signs now factor

Average the preceding identity over D. Independence of the signs gives

$$\mathop{\mathbb{E}}_{D} e^{\frac{n}{4}\left\|\widetilde{u_i}\right\|_2^2} = \mathop{\mathbb{E}}_{g} \prod_{j=1}^{n} \mathop{\mathbb{E}}_{\eta_j} e^{\eta_j \langle g, u_j\rangle/\sqrt{2}}$$

$$= \mathop{\mathbb{E}}_{g} \prod_{j=1}^{n} \cosh\left(\frac{\langle g, u_j\rangle}{\sqrt{2}}\right)$$

$$\leq \mathop{\mathbb{E}}_{g} \exp\left\{\frac{1}{4}\sum_{j=1}^{n} \langle g, u_j\rangle^2\right\} = \mathop{\mathbb{E}}_{g} e^{\|g\|_2^2/4} = 2^{k/2}.$$

Here $\cosh(s) \leq e^{s^2/2}$, $\sum_j u_j u_j^\top = I_k$, and $\|g\|_2^2 \sim \chi_k^2$.

The signs now factor

Average the preceding identity over D. Independence of the signs gives

$$\begin{aligned}
\mathop{\mathbb{E}}_{D} e^{\frac{n}{4}\left\|\widetilde{u_i}\right\|_2^2} &= \mathop{\mathbb{E}}_{g} \prod_{j=1}^{n} \mathop{\mathbb{E}}_{\eta_j} e^{\eta_j \langle g, u_j \rangle / \sqrt{2}} \\
&= \mathop{\mathbb{E}}_{g} \prod_{j=1}^{n} \cosh\left(\frac{\langle g, u_j \rangle}{\sqrt{2}} \right) \\
&\leq \mathop{\mathbb{E}}_{g} \exp\left\{ \frac{1}{4} \sum_{j=1}^{n} \langle g, u_j \rangle^2 \right\} = \mathop{\mathbb{E}}_{g} e^{\|g\|_2^2/4} = 2^{k/2}.
\end{aligned}$$

Here $\cosh(s) \leq e^{s^2/2}$, $\sum_j u_j u_j^\top = I_k$, and $\|g\|_2^2 \sim \chi_k^2$.

Markov gives

$$\mathbb{P}\left(n \,\|\widetilde{u}_i\|_2^2 \geq 2k + 4t \right) \leq e^{-k/2-t}\, 2^{k/2} \leq e^{-t}.$$

The signs now factor

Average the preceding identity over D. Independence of the signs gives

$$\begin{aligned}
\mathop{\mathbb{E}}_{D} e^{\frac{n}{4}\left\|\widetilde{u_i}\right\|_2^2} &= \mathop{\mathbb{E}}_{g} \prod_{j=1}^{n} \mathop{\mathbb{E}}_{\eta_j} e^{\eta_j \langle g, u_j\rangle/\sqrt{2}} \\
&= \mathop{\mathbb{E}}_{g} \prod_{j=1}^{n} \cosh\left(\frac{\langle g, u_j\rangle}{\sqrt{2}}\right) \\
&\le \mathop{\mathbb{E}}_{g} \exp\left\{\frac{1}{4}\sum_{j=1}^{n} \langle g, u_j\rangle^2\right\} = \mathop{\mathbb{E}}_{g} e^{\|g\|_2^2/4} = 2^{k/2}.
\end{aligned}$$

Here $\cosh(s) \le e^{s^2/2}$, $\sum_j u_j u_j^\top = I_k$, and $\|g\|_2^2 \sim \chi_k^2$.

Markov gives

$$\mathbb{P}\left(n\,\|\widetilde{u}_i\|_2^2 \ge 2k + 4t\right) \le e^{-k/2-t}\,2^{k/2} \le e^{-t}.$$

A union bound over the n rows therefore yields, with probability at least $1-\delta_0$,

$$\boxed{\mu(HDU) \le 2k + 4\log\frac{n}{\delta_0}.}$$

SRHT: combine incoherence and sampling

$$\boxed{\begin{array}{c} D, H \\ \mu(HDU) \lesssim k + \log(n/\delta) \end{array}} \longrightarrow \boxed{\begin{array}{c} \text{Coherence bound + Matrix Bernstein} \\ \|G - I_k\| \leq \varepsilon \end{array}}$$

SRHT: combine incoherence and sampling

$$\boxed{\begin{array}{c} D, H \\ \mu(HDU) \lesssim k + \log(n/\delta) \end{array}} \longrightarrow \boxed{\begin{array}{c} \text{Coherence bound + Matrix Bernstein} \\ \|G - I_k\| \leq \varepsilon \end{array}}$$

$$\boxed{m \ \gtrsim \ \varepsilon^{-2}\big(k + \log(n/\delta)\big)\log(k/\delta)}$$

suffices for

$$(1-\varepsilon)I_k \preceq G = U^\top \Omega \Omega^\top U \preceq (1+\varepsilon)I_k$$

with probability at least $1 - \delta$.

The SRHT is a (k, ε, δ)-OSE.

H can be replaced by other flat unitaries, like the Discrete Fourier Transform.

CountSketch: hash and sign

For every coordinate $j \in [n]$, choose independently

$$h(j) \sim \mathrm{Unif}[b], \qquad \xi_j \sim \mathrm{Unif}\{\pm 1\}.$$

The CountSketch matrix $S \in \mathbb{R}^{b \times n}$ is defined by

$$\boxed{Se_j = \xi_j e_{h(j)} \qquad \Longleftrightarrow \qquad S_{rj} = \xi_j \, \mathbf{1}_{\{h(j)=r\}}.}$$

CountSketch: hash and sign

For every coordinate $j \in [n]$, choose independently

$$h(j) \sim \mathrm{Unif}[b], \qquad \xi_j \sim \mathrm{Unif}\{\pm 1\}.$$

The CountSketch matrix $S \in \mathbb{R}^{b \times n}$ is defined by

$$\boxed{Se_j = \xi_j e_{h(j)} \qquad \Longleftrightarrow \qquad S_{rj} = \xi_j \mathbf{1}_{\{h(j)=r\}}.}$$

Every column has exactly one nonzero, and

$$(Sx)_r = \sum_{j:\, h(j)=r} \xi_j x_j.$$

Coordinates in the same bucket are (signed) added.

$$\begin{bmatrix} 0 & -1 & 0 & 0 & +1 \\ +1 & 0 & 0 & -1 & 0 \\ 0 & 0 & +1 & 0 & 0 \end{bmatrix}$$

one signed entry per column

CountSketch: hash and sign

For every coordinate $j \in [n]$, choose independently

$$h(j) \sim \mathrm{Unif}[b], \qquad \xi_j \sim \mathrm{Unif}\{\pm 1\}.$$

The CountSketch matrix $S \in \mathbb{R}^{b \times n}$ is defined by

$$\boxed{Se_j = \xi_j e_{h(j)} \qquad \Longleftrightarrow \qquad S_{rj} = \xi_j \mathbf{1}_{\{h(j)=r\}}.}$$

Every column has exactly one nonzero, and

$$(Sx)_r = \sum_{j:\, h(j)=r} \xi_j x_j.$$

$$\begin{bmatrix} 0 & -1 & 0 & 0 & +1 \\ +1 & 0 & 0 & -1 & 0 \\ 0 & 0 & +1 & 0 & 0 \end{bmatrix}$$

one signed entry per column

Coordinates in the same bucket are (signed) added.

$$\boxed{(S^\top S)_{ab} = \xi_a \xi_b \mathbf{1}_{\{h(a)=h(b)\}} \qquad \Longrightarrow \qquad \mathbb{E}[S^\top S] = I_n.}$$

SparseStack: stacked independent CountSketch matrices

Let $S_1, \ldots, S_\zeta \in \mathbb{R}^{b \times n}$ be independent CountSketch matrices and $m = b\zeta$:

$$\Omega^\top = \frac{1}{\sqrt{\zeta}} \begin{bmatrix} S_1 \\ \vdots \\ S_\zeta \end{bmatrix} \in \mathbb{R}^{m \times n}$$

ζ nonzeros per column.

SparseStack: stacked independent CountSketch matrices

Let $S_1, \dots, S_\zeta \in \mathbb{R}^{b \times n}$ be independent CountSketch matrices and $m = b\zeta$:

$$\Omega^\top = \frac{1}{\sqrt{\zeta}} \begin{bmatrix} S_1 \\ \vdots \\ S_\zeta \end{bmatrix} \in \mathbb{R}^{m \times n}$$

ζ nonzeros per column.

$$Y = \Omega^\top U \qquad G = Y^\top Y = \frac{1}{\zeta} \sum_{\ell=1}^{\zeta} B_\ell$$

$$B_\ell = (S_\ell U)^\top (S_\ell U)$$

B_ℓ independent, $B_\ell \succeq 0$, $\mathbb{E}\, B_\ell = I_k$, $\mathbb{E}[\Omega\Omega^\top] = I_n$.

SparseStack: stacked independent CountSketch matrices

Let $S_1, \ldots, S_\zeta \in \mathbb{R}^{b \times n}$ be independent CountSketch matrices and $m = b\zeta$:

$$\Omega^\top = \frac{1}{\sqrt{\zeta}} \begin{bmatrix} S_1 \\ \vdots \\ S_\zeta \end{bmatrix} \in \mathbb{R}^{m \times n}$$

ζ nonzeros per column.

$$Y = \Omega^\top U \qquad G = Y^\top Y = \frac{1}{\zeta} \sum_{\ell=1}^{\zeta} B_\ell$$

$$B_\ell = (S_\ell U)^\top (S_\ell U)$$

B_ℓ independent, $B_\ell \succeq 0$, $\mathbb{E}\, B_\ell = I_k$, $\mathbb{E}[\Omega\Omega^\top] = I_n$.

collision variance $\lesssim \dfrac{k}{m}$

$$\lambda_{\max}\left(\frac{I_k - B_\ell}{\zeta}\right) \leq \frac{1}{\zeta}$$

One uses the one-sided version of Matrix Bernstein.

$$m \gtrsim \varepsilon^{-2} k \log(k/\delta), \qquad \zeta \gtrsim \varepsilon^{-1} \log(k/\delta) \quad \Longrightarrow \quad G \succeq (1-\varepsilon) I_k.$$

SparseStack is a (k, ε, δ)-OSI with probability at least $1 - \delta$.

PART III

Sharp Matrix Concentration

Non-Commutative Khintchine

$X \in \mathbb{R}^{d\times d}$ a symmetric matrix with jointly Gaussian entries $\quad \Leftrightarrow \quad X = \sum_{i=1}^{n} g_i A_i$

for $g_i \sim \mathcal{N}(0,1)$ and A_i's deterministic matrices

Non-Commutative Khintchine

$X \in \mathbb{R}^{d\times d}$ a symmetric matrix with jointly Gaussian entries $\quad \Leftrightarrow \quad X = \sum_{i=1}^{n} g_i A_i$

for $g_i \sim \mathcal{N}(0,1)$ and A_i's deterministic matrices

Theorem (Non-Commutative Khintchine; Lust-Piquard '86, L.-P. & Pisier '91)

$$\left(\mathbb{E}\,\mathrm{tr}\, X^{2p}\right)^{\frac{1}{2p}} \leq \sqrt{2p}\left(\mathrm{tr}\left(\mathbb{E}\, X^2\right)^p\right)^{\frac{1}{2p}}$$

Non-Commutative Khintchine

$X \in \mathbb{R}^{d\times d}$ a symmetric matrix with jointly Gaussian entries $\quad \Leftrightarrow \quad X = \sum_{i=1}^{n} g_i A_i$

for $g_i \sim \mathcal{N}(0,1)$ and A_i's deterministic matrices

Theorem (Non-Commutative Khintchine; Lust-Piquard '86, L.-P. & Pisier '91)

$$\left(\mathbb{E}\,\mathrm{tr}\, X^{2p}\right)^{\frac{1}{2p}} \le \sqrt{2p}\left(\mathrm{tr}\left(\mathbb{E}\, X^2\right)^p\right)^{\frac{1}{2p}}$$

Taking $p \sim \log(d)$: $\quad \mathbb{E}\,\|X\| \lesssim \sqrt{\log d}\cdot \sigma \quad$ where $\sigma^2 = \left\|\mathbb{E}\, X^2\right\| = \left\|\sum_{i=1}^{n} A_i^2\right\|$

Non-Commutative Khintchine

$X \in \mathbb{R}^{d\times d}$ a symmetric matrix with jointly Gaussian entries $\Leftrightarrow$ $X = \sum_{i=1}^{n} g_i A_i$

for $g_i \sim \mathcal{N}(0,1)$ and A_i's deterministic matrices

Theorem (Non-Commutative Khintchine; Lust-Piquard '86, L.-P. & Pisier '91)

$$\left(\mathbb{E}\,\mathrm{tr}\,X^{2p}\right)^{\frac{1}{2p}} \le \sqrt{2p}\left(\mathrm{tr}\left(\mathbb{E}\,X^2\right)^p\right)^{\frac{1}{2p}}$$

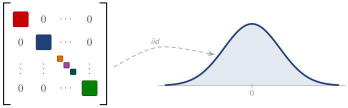
Taking $p \sim \log(d)$: $\mathbb{E}\,\|X\| \lesssim \sqrt{\log d}\cdot\sigma$ where $\sigma^2 = \left\|\mathbb{E}\,X^2\right\| = \left\|\sum_{i=1}^{n} A_i^2\right\|$

- **Central question:** Is the dimensional factor needed? If so, when?
- The same dimensional factor appeared in Matrix Bernstein in Part I:

$$\mathrm{Prob}\left(\|\sum H_k\| \ge t\right) \lesssim d\,\exp\left(-t^2/\cdots\right)$$

The role of non-commutativity

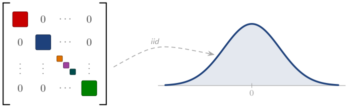
Diagonal Matrix: $D_{ii} \sim \mathcal{N}(0,1)$



$$\mathbb{E}\,\|X\| \approx \sqrt{2\log d} \qquad \sigma^2 = 1$$

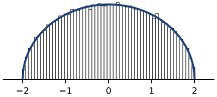
The role of non-commutativity

Diagonal Matrix: $D_{ii} \sim \mathcal{N}(0,1)$



$\mathbb{E}\,\|X\| \approx \sqrt{2\log d} \qquad \sigma^2 = 1$

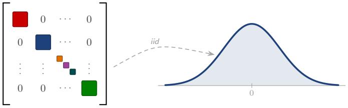
Wigner Matrix: $W_{ij} \sim \mathcal{N}(0,1/d)$



$\mathbb{E}\,\|X\| \approx 2 \qquad \sigma^2 = 1$

The role of non-commutativity

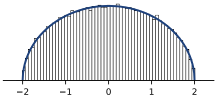
Diagonal Matrix: $D_{ii} \sim \mathcal{N}(0,1)$



$\mathbb{E}\,\|X\| \approx \sqrt{2\log d} \qquad \sigma^2 = 1$

Question by Joel Tropp, Oberwolfach 2014

Wigner Matrix: $W_{ij} \sim \mathcal{N}(0,1/d)$



$\mathbb{E}\,\|X\| \approx 2 \qquad \sigma^2 = 1$

Guiding principle:

Diagonal matrices **commute**
Wigner matrices do **not**

(Central Limit Theorem Vs Free CLT)

Proof of NCK (sketch)

$\left(\mathbb{E}\,\|X\|\right)^{2p} \leq d\,\mathbb{E}\,\mathrm{tr}\left(X^{2p}\right)$

Proof of NCK (sketch)

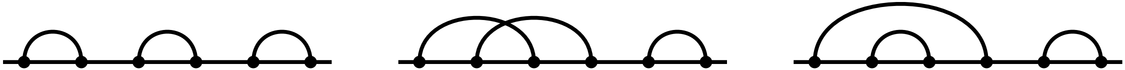
$X = \sum g_i A_i$

$$\left(\mathbb{E}\,\|X\|\right)^{2p} \le d\,\mathbb{E}\,\mathrm{tr}\left(X^{2p}\right) = d \sum_{i_1,\ldots,i_{2p}=1}^{n} \mathbb{E}[g_{i_1}\cdots g_{i_{2p}}]\cdot \mathrm{tr}\left(A_{i_1}\cdots A_{i_{2p}}\right)$$

Proof of NCK (sketch)

$$(\mathbb{E}\,\|X\|)^{2p} \le d\,\mathbb{E}\operatorname{tr}\left(X^{2p}\right) = d \sum_{i_1,\ldots,i_{2p}=1}^{n} \mathbb{E}[g_{i_1}\cdots g_{i_{2p}}]\cdot \operatorname{tr}\left(A_{i_1}\cdots A_{i_{2p}}\right)$$

(Wick's formula)
$$= d \sum_{\nu\in\mathbb{P}_2[2p]} \sum_{\substack{f:[2p]\to[n]\\ f\sim\nu}} \operatorname{tr}\left(A_{f(1)}\cdots A_{f(2p)}\right)$$

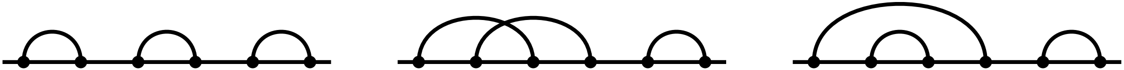


Proof of NCK (sketch)

$X = \sum g_i A_i$

$$(\mathbb{E}\,\|X\|)^{2p} \le d\,\mathbb{E}\,\mathrm{tr}\left(X^{2p}\right) = d \sum_{i_1,\dots,i_{2p}=1}^{n} \mathbb{E}[g_{i_1}\cdots g_{i_{2p}}]\cdot \mathrm{tr}\left(A_{i_1}\cdots A_{i_{2p}}\right)$$

(Wick's formula) $$= d \sum_{\nu\in\mathbb{P}_2[2p]} \sum_{\substack{f:[2p]\to[n]\\ f\sim\nu}} \mathrm{tr}\left(A_{f(1)}\cdots A_{f(2p)}\right)$$



"Commutative is worst-case" $$\le d\,|\#\mathbb{P}_2[2p]| \sum_{j_1,\dots,j_p=1}^{n} \mathrm{tr}\left[A_{j_1}^2\cdots A_{j_p}^2\right]$$ (Buchholz '01)

$$= d\,|\#\mathbb{P}_2[2p]|\ \mathrm{tr}\left(\sum_{j=1}^{n} A_j^2\right)^p \le d\,|\#\mathbb{P}_2[2p]|\ \sigma^{2p} = d(2p-1)!!\ \sigma^{2p}$$

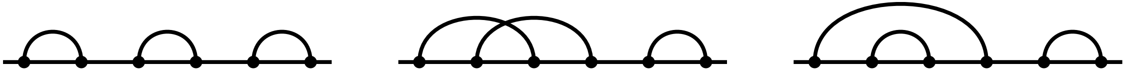
$$\Rightarrow \mathbb{E}\,\|X\| \lesssim d^{\frac{1}{2p}}\sqrt{p}\ \sigma \sim \sqrt{\log d}\ \sigma$$ for $p \sim \log d$

Proof of NCK (sketch)

$$(\mathbb{E}\,\|X\|)^{2p} \le d\,\mathbb{E}\,\mathrm{tr}\left(X^{2p}\right) = d \sum_{i_1,\ldots,i_{2p}=1}^{n} \mathbb{E}[g_{i_1}\cdots g_{i_{2p}}]\cdot \mathrm{tr}\left(A_{i_1}\cdots A_{i_{2p}}\right)$$

(Wick's formula)

$$= d \sum_{\nu\in\mathbb{P}_2[2p]} \sum_{\substack{f:[2p]\to[n]\\ f\sim\nu}} \mathrm{tr}\left(A_{f(1)}\cdots A_{f(2p)}\right)$$



"Commutative is worst-case"

$$\le d\,|\#\mathbb{P}_2[2p]| \sum_{j_1,\ldots,j_p=1}^{n} \mathrm{tr}\left[A_{j_1}^2\cdots A_{j_p}^2\right]$$

(Buchholz '01)

$$= d\,|\#\mathbb{P}_2[2p]|\;\mathrm{tr}\left(\sum_{j=1}^{n} A_j^2\right)^p \le d\,|\#\mathbb{P}_2[2p]|\;\sigma^{2p} = d(2p-1)!!\;\sigma^{2p}$$

$$\Rightarrow \mathbb{E}\,\|X\| \lesssim d^{\frac{1}{2p}}\sqrt{p}\,\sigma \sim \sqrt{\log d}\,\sigma$$

for $p \sim \log d$

Opportunity: Most partitions have crossings and may offer cancellations

$\mathbf{NC_2[2p]} \sim \mathbf{4^p}$ (Catalan #'s)

Improved NCK inequalities

Lust-Piquard & Pisier '86, '91: $\mathbb{E}\|X\| \lesssim \sqrt{\log d}\,\sigma$

B & van Handel '16 (independent entries): $\mathbb{E}\|X\| \lesssim \sigma + \sqrt{\log d}\,\sigma_*$

non-homogeneous refinement: Latała–van Handel–Youssef '18

Tropp '18: $\mathbb{E}\|X\| \lesssim (\log d)^{\frac{1}{4}}\,\sigma + \sqrt{\log d}\,\omega(X)$

A. S. Bandeira & R. van Handel, Annals of Probability 2016.

J. Tropp, Applied and Computational Harmonic Analysis, 2018

R. Latała, R. van Handel, P. Youssef, Inventiones Mathematicae, 2018

A. S. Bandeira & M. T. Boedihardjo & R. van Handel, Inventiones Mathematicae, 2023

Improved NCK inequalities

Lust-Piquard & Pisier '86, '91: $\mathbb{E}\,\|X\| \ \lesssim \ \sqrt{\log d}\ \sigma$

B & van Handel '16 (independent entries): $\mathbb{E}\,\|X\| \ \lesssim \ \sigma + \sqrt{\log d}\ \sigma_*$

non-homogeneous refinement: Latała–van Handel–Youssef '18

Tropp '18: $\mathbb{E}\,\|X\| \ \lesssim \ (\log d)^{\frac{1}{4}}\,\sigma + \sqrt{\log d}\ \omega(X)$

Intrinsic Freeness (B & Boedihardjo & van Handel, 2023)

$$\mathbb{E}\,\|X\| \ \lesssim \ \sigma \ + \ (\log d)^{\frac{3}{2}}\ v(X), \qquad\qquad v(X)^2 = \|\mathrm{Cov}(X)\|$$

for a Wigner matrix: $\sigma = 1, \ \ v \approx d^{-1/2}$; for the gaussian diagonal: $\sigma = 1, \ \ v = 1$

$v(X)$ tells them apart!

A. S. Bandeira & R. van Handel, Annals of Probability 2016.

J. Tropp, Applied and Computational Harmonic Analysis, 2018

R. Latała, R. van Handel, P. Youssef, Inventiones Mathematicae, 2018

A. S. Bandeira & M. T. Boedihardjo & R. van Handel, Inventiones Mathematicae, 2023

“Pictogram” of the proof

Free Probability: Voiculescu ’83, ’91
book: Nica & Speicher ’06

Create “freeness”

$$X^{(N)} = \sum_{k=1}^{n} A_k \otimes W_k^{(N)}$$

$W_k^{(N)}$: independent $N \times N$ Wigner

$N \to \infty$

strong convergence —
see R. van Handel’s ICM lecture

$$X_{\text{free}} = \sum_{k=1}^{n} A_k \otimes s_k$$

$s_1, \ldots, s_n$: free semicircular family

Gaussian interpolation

if $v(X) \ll \sigma$

$$\begin{bmatrix} X & & \\ & X' & \\ & & \ddots \end{bmatrix} \sim X$$

$$\sigma \;\le\; \|X_{\text{free}}\| \;\le\; 2\sigma$$

X_{free} satisfies Wick’s formula with $NC_2[2p]$ only:  ✓  $\to 0$

A. S. Bandeira & M. T. Boedihardjo & R. van Handel, Inventiones Mathematicae, 2023

Intrinsic Freeness: main theorem (two-sided)

Theorem (BBvH '23 (≈ trace moments; $\leq$)**; BCSvH '26** (regularity of X_{free}; $\geq$)**)**

Let $X = A_0 + \sum_{i=1}^n g_i A_i$ be a $d \times d$ random matrix with jointly gaussian entries. Then

$$\Big| \, \mathbb{E}\,\|X\| - \|X_{\mathrm{free}}\| \, \Big| \;\leq\; C\,\sqrt{\sigma(X)\,v(X)}\;(\log d)^{\frac{3}{4}}.$$

Moments, and the *whole spectrum* is captured: $\mathrm{sp}(X) \approx \mathrm{sp}(X_{\mathrm{free}})$ (in Hausdorff distance)

A. S. Bandeira & M. T. Boedihardjo & R. van Handel, Inventiones Mathematicae, 2023

A. S. Bandeira, G. Cipolloni, D. Schröder, R. van Handel, Communications of the AMS, 2026

T. Brailovskaya & R. van Handel, GAFA, 2024

D. Kunisky, arXiv, 2026

Intrinsic Freeness: main theorem (two-sided)

Theorem (BBvH '23 ($\approx$ trace moments; $\leq$); BCSvH '26 (regularity of X_{free}; $\geq$))

Let $X = A_0 + \sum_{i=1}^n g_i A_i$ be a $d \times d$ random matrix with jointly gaussian entries. Then

$$\Big| \mathbb{E}\,\|X\| - \|X_{\mathrm{free}}\| \Big| \;\leq\; C\,\sqrt{\sigma(X)\,v(X)}\;(\log d)^{\frac{3}{4}}.$$

Moments, and the *whole spectrum* is captured: $\mathrm{sp}(X) \approx \mathrm{sp}(X_{\mathrm{free}})$ (in Hausdorff distance)

- $\sigma(X) \leq \|X_{\mathrm{free}}\| \leq 2\sigma(X)$ (and $\|X_{\mathrm{free}}\|$ is given by Lehner's formula, an SDP — Kunisky '26)
- **Useful whenever** $v(X) \ll \sigma(X)/(\log d)^{\frac{3}{2}}$; **two-sided** $\Rightarrow$ sharp phase transitions
- **Universality** (Brailovskaya & van Handel '24) $\Rightarrow$ matrix Bernstein without d factor

---

A. S. Bandeira & M. T. Boedihardjo & R. van Handel, Inventiones Mathematicae, 2023

A. S. Bandeira, G. Cipolloni, D. Schröder, R. van Handel, Communications of the AMS, 2026

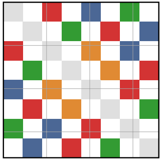
T. Brailovskaya & R. van Handel, GAFA, 2024

D. Kunisky, arXiv, 2026

Example: Glued entries / Patterned matrices

Split the $d \times d$ entries into sets (colors) $S_1, \ldots, S_n$

$$X_{i,j} = \frac{1}{\sqrt{d}} g_\ell \quad \text{if } (i,j) \in S_\ell, \qquad g_\ell \sim \mathcal{N}(0,1)$$

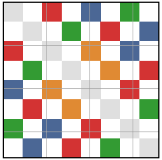


A. S. Bandeira & M. T. Boedihardjo & R. van Handel, Inventiones Mathematicae, 2023

Example: Glued entries / Patterned matrices

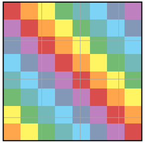
Split the $d \times d$ entries into sets (colors) $S_1, \ldots, S_n$

$$X_{i,j} = \frac{1}{\sqrt{d}} g_\ell \quad \text{if } (i,j) \in S_\ell, \qquad g_\ell \sim \mathcal{N}(0,1)$$



For which value of $\mathbf{K} = \max_\ell |\mathbf{S}_\ell|$ does X behave like a Wigner matrix?

Wigner matrix
$\mathbf{K = 1}$
$\mathbb{E}\,\|X\| \approx 2$



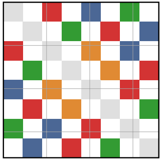
Circulant matrix, $\mathbf{K \sim d}$
diagonal in disguise:
$\mathbb{E}\,\|X\| \approx \sqrt{2 \log d}$

A. S. Bandeira & M. T. Boedihardjo & R. van Handel, Inventiones Mathematicae, 2023

Example: Glued entries / Patterned matrices

Split the $d \times d$ entries into sets (colors) $S_1, \dots, S_n$

$$X_{i,j} = \frac{1}{\sqrt{d}} g_\ell \quad \text{if } (i,j) \in S_\ell, \qquad g_\ell \sim \mathcal{N}(0,1)$$



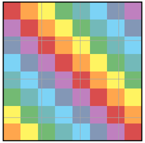
For which value of $\mathbf{K} = \max_\ell |\mathbf{S}_\ell|$ does X behave like a Wigner matrix?

Wigner matrix
$\mathbf{K} = \mathbf{1}$
$\mathbb{E}\,\|X\| \approx 2$

Answer [BBvH '23]: $\sigma(X) = 1$,

$$v(X)^2 \asymp \frac{K}{d}$$

Wigner behaviour whenever $\mathbf{K} \ll \frac{\mathbf{d}}{(\log \mathbf{d})^3}$



Circulant matrix, $\mathbf{K} \sim \mathbf{d}$
diagonal in disguise:
$\mathbb{E}\,\|X\| \approx \sqrt{2 \log d}$

A. S. Bandeira & M. T. Boedihardjo & R. van Handel, Inventiones Mathematicae, 2023

Two-sided bounds $\Rightarrow$ sharp phase transitions

"Signal-in-noise" Spiked model: $X(\lambda) = \lambda v v^\top + W$

$v \in S^{d-1}$, W Wigner

Detect "signal" $\lambda v v^\top$?   Spectral norm bounds give, at best, $\lambda_* \le 2$

J. Baik, G. Ben Arous, S. Péché, Annals of Probability, 2005

A. S. Wein & A. El Alaoui & C. Moore, FOCS 2019

A. S. Bandeira, G. Cipolloni, D. Schröder, R. van Handel, Communications of the AMS, 2026

Two-sided bounds ⇒ sharp phase transitions

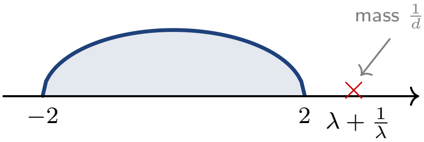
"Signal-in-noise" Spiked model: $X(\lambda) = \lambda vv^\top + W$

$v \in S^{d-1}$, W Wigner

Detect "signal" $\lambda vv^\top$?

Spectral norm bounds give, at best, $\lambda_* \leq 2$

$$\lambda_{\max}(X(\lambda)) \to \begin{cases} 2 & \lambda \leq 1 \\ \lambda + \frac{1}{\lambda} & \lambda \geq 1 \end{cases} \qquad \lambda_* = 1$$



BBP phase transition

Visible in X_{free}, thus now established for many RMT models!

J. Baik, G. Ben Arous, S. Péché, Annals of Probability, 2005

A. S. Wein & A. El Alaoui & C. Moore, FOCS 2019

A. S. Bandeira, G. Cipolloni, D. Schröder, R. van Handel, Communications of the AMS, 2026

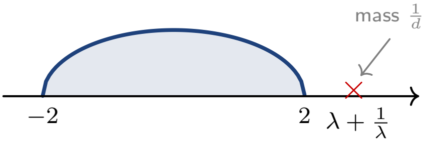
Two-sided bounds $\Rightarrow$ sharp phase transitions

"Signal-in-noise" Spiked model: $X(\lambda) = \lambda vv^\top + W$

$v \in S^{d-1}$, W Wigner

Detect "signal" $\lambda vv^\top$? Spectral norm bounds give, at best, $\lambda_* \leq 2$

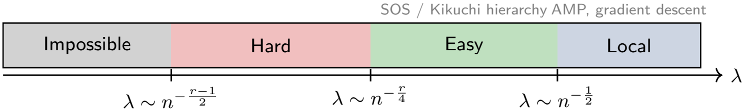
$$\lambda_{\max}(X(\lambda)) \to \begin{cases} 2 & \lambda \leq 1 \\ \lambda + \frac{1}{\lambda} & \lambda \geq 1 \end{cases} \qquad \lambda_* = 1$$



BBP phase transition

Visible in X_{free}, thus now established for many RMT models!

E.g.: **Tensor PCA** $(Y = \lambda x^{\otimes r} + Z)$ via **Kikuchi matrices** (WEAM '19): $\mathbb{E}(M - \mathbb{E}\,M)^2 = d_\ell I$



$\Rightarrow$ exact spectral threshold $(\frac{1}{2}r \leq \ell < \frac{3}{4}r)$ $\lambda_* = \frac{1}{\sqrt{d_\ell}} \sim_r n^{-r/4}$

J. Baik, G. Ben Arous, S. Péché, Annals of Probability, 2005

A. S. Wein & A. El Alaoui & C. Moore, FOCS 2019

A. S. Bandeira, G. Cipolloni, D. Schröder, R. van Handel, Communications of the AMS, 2026

Matrix Chaos: Matrices with polynomial valued entries.

For degree q=2: $X = \sum_{i,j} g_i g_j A_{ij}$

A. S. Bandeira, K. Lucca, P. Nizić-Nikolac, R. van Handel, STOC 2025

Matrix Chaos: Matrices with polynomial valued entries.

For degree q=2: $X = \sum_{i,j} g_i g_j A_{ij}$

Matrix Concentration Inequalities can be iterated:

σ in NCK *is* a flattening

$$\sigma\left(\sum_i g_i A_i\right) = \max\left\{\left\|\sum_i A_i^\top A_i\right\|^{1/2}, \left\|\sum_i A_i A_i^\top\right\|^{1/2}\right\} = \max\left\{\left\|\begin{matrix} A_1 \\ \vdots \\ A_n \end{matrix}\right\|, \|A_1 \ \cdots \ A_n\|\right\}.$$

and so are other matrix concentration parameters

A. S. Bandeira, K. Lucca, P. Nizić-Nikolac, R. van Handel, STOC 2025

Matrix Chaos: Matrices with polynomial valued entries.

For degree q=2: $X = \sum_{i,j} g_i g_j A_{ij}$

Matrix Concentration Inequalities can be iterated:

σ in NCK *is* a flattening

$$\sigma\left(\sum_i g_i A_i\right) = \max\left\{\left\|\sum_i A_i^\top A_i\right\|^{1/2}, \left\|\sum_i A_i A_i^\top\right\|^{1/2}\right\} = \max\left\{\left\|\begin{matrix} A_1 \\ \vdots \\ A_n \end{matrix}\right\|, \|A_1 \ \cdots \ A_n\|\right\}.$$

and so are other matrix concentration parameters

degree q, NCK iterated: $\mathbb{E}\,\|Y\| \lesssim_q (\log d)^{\frac{q}{2}}\,\sigma(\mathcal{A})$

v is *also* a flattening $\Rightarrow$ intrinsic freeness iterates too:

$\mathbb{E}\,\|Y\| \lesssim_q \sigma(\mathcal{A}) + (\mathrm{polylog}\, d)\, v(\mathcal{A})$

A. S. Bandeira, K. Lucca, P. Nizić-Nikolac, R. van Handel, STOC 2025

Matrix Chaos: Matrices with polynomial valued entries.

For degree q=2: $X = \sum_{i,j} g_i g_j A_{ij}$

Matrix Concentration Inequalities can be iterated:

σ in NCK *is* a flattening

$$\sigma\left(\sum_i g_i A_i\right) = \max\left\{\left\|\sum_i A_i^\top A_i\right\|^{1/2}, \left\|\sum_i A_i A_i^\top\right\|^{1/2}\right\} = \max\left\{\left\|\begin{matrix} A_1 \\ \vdots \\ A_n \end{matrix}\right\|, \left\|A_1 \ \cdots \ A_n\right\|\right\}.$$

and so are other matrix concentration parameters

degree q, NCK iterated: $\mathbb{E}\|Y\| \lesssim_q (\log d)^{\frac{q}{2}} \sigma(\mathcal{A})$

v is *also* a flattening $\Rightarrow$ intrinsic freeness iterates too:

$\mathbb{E}\|Y\| \lesssim_q \sigma(\mathcal{A}) + (\mathrm{polylog}\, d)\, v(\mathcal{A})$

- Can be used to analyse: SOS for Tensor PCA (cf. Hopkins–Shi–Steurer '15), Khatri–Rao (cf. Rudelson '11), quantum expanders (cf. Lancien–Youssef '23), Matrix Product States, Graph Matrices, partial circulants, ...

Talk to Kevin Lucca to know more

A. S. Bandeira, K. Lucca, P. Nizić-Nikolac, R. van Handel, STOC 2025

Matrix Spencer

Conjecture (Matrix Spencer)

$\exists_{C\geq 0}$ s.t. $\forall_{n\in\mathbb{N}}$ and for any $n\times n$ matrices $A_1,\ldots,A_n$ all satisfying $\|A_k\|\leq 1$, we have

$$\min_{\varepsilon\in\{\pm 1\}^n}\left\|\sum_{k=1}^n \varepsilon_k A_k\right\|\leq C\sqrt{n}.$$

If "very non-commutative" holds by concentration/CLT, if commutative holds by entropy/pigeonhole

N. Bansal & H. Jiang & R. Meka, STOC 2023

A. S. Bandeira, D. Kunisky, D. G. Mixon, X. Zeng, ACHA 2024

A. S. Bandeira & H. Bölcskei, arXiv:2606.12181

E. Akbas & S. Sra, arXiv:2606.16005

Matrix Spencer

Conjecture (Matrix Spencer)

$\exists_{C\geq 0}$ s.t. $\forall_{n\in\mathbb{N}}$ and for any $n\times n$ matrices $A_1,\ldots,A_n$ all satisfying $\|A_k\|\leq 1$, we have

$$\min_{\varepsilon\in\{\pm1\}^n}\left\|\sum_{k=1}^{n}\varepsilon_k A_k\right\|\leq C\sqrt{n}.$$

If "very non-commutative" holds by concentration/CLT, if commutative holds by entropy/pigeonhole

Progress using Intrinsic Freeness:

- **Bansal, Jiang & Meka '23** resolved the Conjecture when $\mathrm{rank}(A_k)\leq\frac{n}{\log(n)^3}$,

  by using [BBvH '23] to show $\mathrm{Prob}\left(\left\|\sum g_k A_k\right\|\leq C\sqrt{n}\right)\geq\exp(-cn)$

- The **group case** (A_k's are the regular representation of a finite group):

  [B., Kunisky, Mixon, Zeng '24], **[B. & Bölcskei '26]**, [Akbas, Sra '26]

N. Bansal & H. Jiang & R. Meka, STOC 2023

A. S. Bandeira, D. Kunisky, D. G. Mixon, X. Zeng, ACHA 2024

A. S. Bandeira & H. Bölcskei, arXiv:2606.12181

E. Akbas & S. Sra, arXiv:2606.16005

Thank You

and here are two more ($\pm$ unrelated) open problems:

- **Tensors and geometric proofs of NCK?** [Conjecture 16 @ `randomstrasse101.math.ethz.ch`]
  For symmetric r-tensors $T_1, \ldots, T_n$, is $\mathbb{E}\left\|\sum_i g_i T_i\right\| \lesssim_r \mathrm{polylog}(d)(\sum_i \|T_i\|^2)^{1/2}$?

- **Sampling from the SK model?** [Problem 15 @ `randomstrasse101.math.ethz.ch`]
  Does Glauber dynamics mix in polynomial time up to the replica-symmetry-breaking threshold $\beta_* = 1$?

Book: *Mathematics of Data Science* arXiv:2607.11938
Survey: *Random Matrices, Intrinsic Freeness, and Sharp Non-Asymptotic Inequalities* arXiv:2510.01021
Positions at ETH Zürich: `people.math.ethz.ch/~abandeira/positions.html`
Open problems: `randomstrasse101.math.ethz.ch`

`afonsobandeira.com`