Télécom ParisTech
CNRS

Inégalités de concentrations non commutatives dans un cadre de dépendance

Marwa Banna
avec F. Merlevède et P. Youssef

Journées MAS 2016
Grenoble

31 Août 2016

Bernstein inequality
Scalar independent case

Let $X_1, \ldots, X_n$ be independent random variables such that

$$\mathbb{E}X_k = 0\,, \quad \mathbb{E}X_k^2 = \sigma_k^2 \quad \text{and} \quad \sup_k |X_k| < 1 \text{ a.s.}$$

Bernstein inequality
Scalar independent case

Let $X_1, \ldots, X_n$ be independent random variables such that

$$\mathbb{E}X_k = 0\,, \quad \mathbb{E}X_k^2 = \sigma_k^2 \quad \text{and} \quad \sup_k |X_k| < 1 \text{ a.s.}$$

For any $x > 0$,

$$\mathbb{P}\Big(\sum_{k=1}^{n} X_k > x\Big) \leqslant \exp\Big(-\frac{x^2}{2\,nV_n + 2x}\Big)\,,$$

where $V_n = \frac{1}{n}\sum_{k=1}^{n} \sigma_k^2$.

Matrix Setting

Let $\mathbf{X}_1, \ldots, \mathbf{X}_n$ be $d \times d$ *centered* Hermitian random matrices.

What can be said about

$$\mathbb{P}\Big(\lambda_{\max}\big(\sum_{k=1}^{n} \mathbf{X}_k\big) \geqslant x\Big) \leqslant \mathbf{?}$$

Independent Matrix Case

Theorem (Ahlswede and Winter '02, Tropp '12)

Let $\mathbf{X}_1, \ldots, \mathbf{X}_n$ be $d \times d$ independent Hermitian random matrices. Assume that each matrix satisfies

$$\mathbb{E}\mathbf{X}_k = 0 \quad \textit{and} \quad \lambda_{\max}(\mathbf{X}_k) \leq 1 \textit{ a.s.}$$

Independent Matrix Case

Theorem (Ahlswede and Winter '02, Tropp '12)

Let $\mathbf{X}_1, \ldots, \mathbf{X}_n$ be $d \times d$ independent Hermitian random matrices. Assume that each matrix satisfies

$$\mathbb{E}\mathbf{X}_k = 0 \quad \textit{and} \quad \lambda_{\max}(\mathbf{X}_k) \leq 1 \textit{ a.s.}$$

Then for any $x > 0$,

$$\mathbb{P}\Big(\lambda_{\max}\big(\sum_{k=1}^{n} \mathbf{X}_k\big) \geq x\Big) \leq d \cdot \exp\Big(-\frac{x^2/2}{n\sigma^2 + x/3}\Big),$$

where $\sigma^2 := \frac{1}{n}\lambda_{\max}\big(\sum_{k=1}^{n} \mathbb{E}\mathbf{X}_k^2\big)$.

Applications?!

- Let $Y \in \mathbb{R}^d$ be an isotropic random vector i.e.

$$\mathbb{E}Y = 0 \quad \text{and} \quad \mathbb{E}YY^t = Id.$$

Applications?!

- Let $Y \in \mathbb{R}^d$ be an isotropic random vector i.e.

$$\mathbb{E}Y = 0 \quad \text{and} \quad \mathbb{E}YY^t = Id.$$

- Let $Y_1, \dots, Y_n$ be independent copies of Y.
- If d is fixed then by the law of large numbers

$$\lim_{n\to\infty} \frac{1}{n} \sum_{k=1}^{n} Y_k Y_k^T = Id \qquad \text{a.s.}$$

Applications?!

- Let $Y \in \mathbb{R}^d$ be an isotropic random vector i.e.

$$\mathbb{E}Y = 0 \quad \text{and} \quad \mathbb{E}YY^t = Id.$$

- Let $Y_1, \ldots, Y_n$ be independent copies of Y.
- If d is fixed then by the law of large numbers

$$\lim_{n \to \infty} \frac{1}{n} \sum_{k=1}^{n} Y_k Y_k^T = Id \qquad \text{a.s.}$$

Question: What is the minimal order of n so that

$$\|\frac{1}{n} \sum_{k=1}^{n} Y_k Y_k^t - Id\| < \epsilon?$$

Applications?!

- Set $\mathbf{X}_k = \frac{1}{n}(Y_k Y_k^t - Id)$.

Applications?!

- Set $\mathbf{X}_k = \frac{1}{n}(Y_k Y_k^t - Id)$.
- $\mathbb{E}\mathbf{X}_k = \mathbf{0}$.
- $\|Y_k\|_2 = \mathcal{O}(\sqrt{d}) \Rightarrow \|\mathbf{X}_k\| = \mathcal{O}(\frac{d}{n})$.

Applications?!

- Set $\mathbf{X}_k = \frac{1}{n}(Y_k Y_k^t - Id)$.
- $\mathbb{E}\mathbf{X}_k = \mathbf{0}$.
- $\|Y_k\|_2 = \mathcal{O}(\sqrt{d}) \Rightarrow \|\mathbf{X}_k\| = \mathcal{O}(\frac{d}{n})$.

Control: $\sigma^2 = \frac{1}{n}\lambda_{\max}\big(\sum_{k=1}^n \mathbb{E}\mathbf{X}_k^2\big)$

Applications?!

- Set $\mathbf{X}_k = \frac{1}{n}(Y_k Y_k^t - Id)$.
- $\mathbb{E}\mathbf{X}_k = \mathbf{0}$.
- $\|Y_k\|_2 = \mathcal{O}(\sqrt{d}) \Rightarrow \|\mathbf{X}_k\| = \mathcal{O}(\frac{d}{n})$.

$$\textbf{Control: } \sigma^2 = \frac{1}{n}\lambda_{\max}\Big(\sum_{k=1}^n \mathbb{E}\mathbf{X}_k^2\Big)$$

- $\mathbb{E}\mathbf{X}_k^2 = \frac{1}{n^2}\mathbb{E}\big[\|Y_k\|_2^2 Y_k Y_k^t + Id - 2 Y_k Y_k^t\big] \precsim \frac{d}{n^2} Id$.
- $\sigma^2 = \mathcal{O}(\frac{d}{n^2})$.

Applications?!

- Set $\mathbf{X}_k = \frac{1}{n}(Y_k Y_k^t - Id)$.
- $\mathbb{E}\mathbf{X}_k = \mathbf{0}$.
- $\|Y_k\|_2 = \mathcal{O}(\sqrt{d}) \ \Rightarrow \ \|\mathbf{X}_k\| = \mathcal{O}(\frac{d}{n})$.

$$\textbf{Control: } \sigma^2 = \frac{1}{n}\lambda_{\max}\big(\sum_{k=1}^n \mathbb{E}\mathbf{X}_k^2\big)$$

- $\mathbb{E}\mathbf{X}_k^2 = \frac{1}{n^2}\mathbb{E}\big[\|Y_k\|_2^2 Y_k Y_k^t + Id - 2Y_k Y_k^t\big] \precsim \frac{d}{n^2} Id$.
- $\sigma^2 = \mathcal{O}(\frac{d}{n^2})$.

Bernstein's inequality then yields

$$\mathbb{P}\Big(\|\frac{1}{n}\sum_{k=1}^n Y_k Y_k^t - Id\| \geq \sqrt{\frac{d\log d}{n}}\Big) = \mathcal{O}(d^{-1}).$$

Applications?!

- Set $\mathbf{X}_k = \frac{1}{n}(Y_k Y_k^t - Id)$.
- $\mathbb{E}\mathbf{X}_k = \mathbf{0}$.
- $\|Y_k\|_2 = \mathcal{O}(\sqrt{d}) \Rightarrow \|\mathbf{X}_k\| = \mathcal{O}(\frac{d}{n})$.

$$\textbf{Control: } \sigma^2 = \frac{1}{n}\lambda_{\max}\big(\sum_{k=1}^n \mathbb{E}\mathbf{X}_k^2\big)$$

- $\mathbb{E}\mathbf{X}_k^2 = \frac{1}{n^2}\mathbb{E}\big[\|Y_k\|_2^2 Y_k Y_k^t + Id - 2Y_k Y_k^t\big] \precsim \frac{d}{n^2} Id$.
- $\sigma^2 = \mathcal{O}(\frac{d}{n^2})$.

Bernstein's inequality then yields

$$\mathbb{P}\Big(\|\frac{1}{n}\sum_{k=1}^n Y_k Y_k^t - Id\| \geq \sqrt{\frac{d\log d}{n}}\Big) = \mathcal{O}(d^{-1}).$$

Conclusion: $n \sim d\log d$ copies are sufficient.

Dependent Matrix Case

- The β-mixing coefficient between two σ-algebras $\mathcal{A}$ and $\mathcal{B}$ is defined by

$$\beta(\mathcal{A}, \mathcal{B}) = \frac{1}{2} \sup \left\{ \sum_{i \in I} \sum_{j \in J} \left| \mathbb{P}(A_i \cap B_j) - \mathbb{P}(A_i)\mathbb{P}(B_j) \right| \right\},$$

where the supremum is taken over all finite partitions $(A_i)_{i \in I}$ and $(B_j)_{j \in J}$ that are respectively $\mathcal{A}$ and $\mathcal{B}$ measurable.

Dependent Matrix Case

- The β-mixing coefficient between two σ-algebras $\mathcal{A}$ and $\mathcal{B}$ is defined by

$$\beta(\mathcal{A}, \mathcal{B}) = \frac{1}{2} \sup \left\{ \sum_{i \in I} \sum_{j \in J} \left| \mathbb{P}(A_i \cap B_j) - \mathbb{P}(A_i)\mathbb{P}(B_j) \right| \right\},$$

where the supremum is taken over all finite partitions $(A_i)_{i \in I}$ and $(B_j)_{j \in J}$ that are respectively $\mathcal{A}$ and $\mathcal{B}$ measurable.
- The sequence $(\beta_k)_k$ associated with $(\mathbf{X}_k)_k$ is defined by

$$\beta_k := \sup_j \beta(\sigma(\mathbf{X}_i\,,\ i \leq j), \sigma(\mathbf{X}_i\,,\ i \geq j + k))$$

Dependent Matrix Case

- The β-mixing coefficient between two σ-algebras $\mathcal{A}$ and $\mathcal{B}$ is defined by

$$\beta(\mathcal{A}, \mathcal{B}) = \frac{1}{2} \sup \left\{ \sum_{i \in I} \sum_{j \in J} \left| \mathbb{P}(A_i \cap B_j) - \mathbb{P}(A_i)\mathbb{P}(B_j) \right| \right\},$$

  where the supremum is taken over all finite partitions $(A_i)_{i \in I}$ and $(B_j)_{j \in J}$ that are respectively $\mathcal{A}$ and $\mathcal{B}$ measurable.
- The sequence $(\beta_k)_k$ associated with $(\mathbf{X}_k)_k$ is defined by

$$\beta_k := \sup_j \beta(\sigma(\mathbf{X}_i \,,\, i \leq j), \sigma(\mathbf{X}_i \,,\, i \geq j + k))$$

- $\beta_k \leq \mathrm{e}^{-ck}$ for some positive constant c.

Dependent Matrix Case

Theorem (B., Merlevède and Youssef '15)

Let $(\mathbf{X}_k)_{k\geq 1}$ be a family of geometrically β-mixing random matrices of dimension d.

Dependent Matrix Case

Theorem (B., Merlevède and Youssef '15)

Let $(\mathbf{X}_k)_{k\geq 1}$ *be a family of geometrically* β*-mixing random matrices of dimension* d. *Assume that*

$$\mathbb{E}(\mathbf{X}_k) = \mathbf{0} \quad \textit{and} \quad \lambda_{\max}(\mathbf{X}_k) \leq 1 \quad \textit{a.s.}$$

Dependent Matrix Case

Theorem (B., Merlevède and Youssef '15)

Let $(\mathbf{X}_k)_{k\geq 1}$ be a family of geometrically β-mixing random matrices of dimension d. Assume that

$$\mathbb{E}(\mathbf{X}_k) = \mathbf{0} \quad \text{and} \quad \lambda_{\max}(\mathbf{X}_k) \leq 1 \quad \text{a.s.}$$

Then for any $x > 0$,

$$\mathbb{P}\Big(\lambda_{\max}\Big(\sum_{k=1}^{n} \mathbf{X}_k\Big) \geq x\Big) \leq d \exp\Big(-\frac{Cx^2}{nv^2 + xc^{-1}(\log n)^2}\Big),$$

where C is a universal constant and v^2 is given by

$$v^2 = \sup_{J\subseteq\{1,\dots,n\}} \frac{1}{\mathrm{Card} J} \lambda_{\max}\Big(\sum_{k,\ell\in J} \mathrm{Cov}(\mathbf{X}_k, \mathbf{X}_\ell)\Big).$$

Consequence

Let **A** be a $d \times n$ random matrix such that:

- Y_i= Columns of A are isotropic i.e. $\mathbb{E} Y_i = 0$ and $\mathbb{E} Y_i Y_i^t = Id$.

Consequence

Let **A** be a $d \times n$ random matrix such that:

- Y_i= Columns of A are isotropic i.e. $\mathbb{E}Y_i = 0$ and $\mathbb{E}Y_iY_i^t = Id$.
- $\|Y_i\|_2 \leq \sqrt{d}$ a.s.

Consequence

Let **A** be a $d \times n$ random matrix such that:

- Y_i= Columns of A are isotropic i.e. $\mathbb{E} Y_i = 0$ and $\mathbb{E} Y_i Y_i^t = Id$.
- $\|Y_i\|_2 \leq \sqrt{d}$ a.s.
- $(Y_i)_{i \leq n}$ are geometrically β-mixing.

Consequence

Let $\mathbf{A}$ be a $d \times n$ random matrix such that:

- Y_i= Columns of A are isotropic i.e. $\mathbb{E} Y_i = 0$ and $\mathbb{E} Y_i Y_i^t = Id$.
- $\|Y_i\|_2 \leq \sqrt{d} \quad$ a.s.
- $(Y_i)_{i \leq n}$ are geometrically β-mixing.

Recall that $\mathbf{AA}^T = \sum_{k=1}^n Y_k Y_k^T$ and $\lambda_k(\mathbf{AA}^T) = s_k^2(\mathbf{A})$.

Consequence

Let **A** be a $d \times n$ random matrix such that:

- Y_i= Columns of A are isotropic i.e. $\mathbb{E} Y_i = 0$ and $\mathbb{E} Y_i Y_i^t = Id$.
- $\|Y_i\|_2 \leq \sqrt{d}$ a.s.
- $(Y_i)_{i \leq n}$ are geometrically β-mixing.

Recall that $\mathbf{A}\mathbf{A}^T = \sum_{k=1}^n Y_k Y_k^T$ and $\lambda_k(\mathbf{A}\mathbf{A}^T) = s_k^2(\mathbf{A})$.

Then with high probability

$$1 - \sqrt{d \log^3 d/n} \leq s_{\min}\Big(\frac{\mathbf{A}}{\sqrt{n}}\Big) \leq s_{\max}\Big(\frac{\mathbf{A}}{\sqrt{n}}\Big) \leq 1 + \sqrt{d \log^3 d/n}.$$

Consequence

Let **A** be a $d \times n$ random matrix such that:

- Y_i= Columns of A are isotropic i.e. $\mathbb{E}Y_i = 0$ and $\mathbb{E}Y_iY_i^t = Id$.
- $\|Y_i\|_2 \leq \sqrt{d}$ a.s.
- $(Y_i)_{i \leq n}$ are geometrically β-mixing.

Recall that $\mathbf{A}\mathbf{A}^T = \sum_{k=1}^n Y_k Y_k^T$ and $\lambda_k(\mathbf{A}\mathbf{A}^T) = s_k^2(\mathbf{A})$.

Then with high probability

$$1 - \sqrt{d \log^3 d / n} \leq s_{\min}\left(\frac{\mathbf{A}}{\sqrt{n}}\right) \leq s_{\max}\left(\frac{\mathbf{A}}{\sqrt{n}}\right) \leq 1 + \sqrt{d \log^3 d / n}.$$

If $d \log^3 d \ll n$ then $\frac{1}{\sqrt{n}}\mathbf{A}$ is "almost" an isometry.

Matrix Chernoff Bound

Ahlswede and Winter (2002) prove the following *matrix* Chernoff bound:

$$\mathbb{P}\Big(\lambda_{\max}\big(\sum_{k=1}^{n}\mathbf{X}_k\big) \geqslant x\Big) \leqslant \inf_{t>0}\Big\{\mathrm{e}^{-tx}\cdot\mathbb{E}\mathrm{Tr}\exp\big(t\sum_{i=1}^{n}\mathbf{X}_i\big)\Big\}$$

Matrix Chernoff Bound

Ahlswede and Winter (2002) prove the following *matrix* Chernoff bound:

$$\mathbb{P}\Big(\lambda_{\max}\big(\sum_{k=1}^{n}\mathbf{X}_k\big) \geqslant x\Big) \leqslant \inf_{t>0}\Big\{e^{-tx}\cdot \mathbb{E}\mathrm{Tr}\exp\big(t\sum_{i=1}^{n}\mathbf{X}_i\big)\Big\}$$

For any $t > 0$,

$$\mathbb{P}\Big(\lambda_{\max}\big(\sum_{k=1}^{n}\mathbf{X}_k\big) \geqslant x\Big) = \mathbb{P}\Big(e^{\lambda_{\max}(t\sum_{k=1}^{n}\mathbf{X}_k)} \geqslant e^{tx}\Big)$$

Matrix Chernoff Bound

Ahlswede and Winter (2002) prove the following *matrix* Chernoff bound:

$$\mathbb{P}\Big(\lambda_{\max}\big(\sum_{k=1}^{n}\mathbf{X}_k\big) \geqslant x\Big) \leqslant \inf_{t>0}\Big\{\mathrm{e}^{-tx}\cdot \mathbb{E}\mathrm{Tr}\exp\big(t\sum_{i=1}^{n}\mathbf{X}_i\big)\Big\}$$

For any $t > 0$,

$$\begin{aligned}\mathbb{P}\Big(\lambda_{\max}\big(\sum_{k=1}^{n}\mathbf{X}_k\big) \geqslant x\Big) &= \mathbb{P}\Big(\mathrm{e}^{\lambda_{\max}(t\sum_{k=1}^{n}\mathbf{X}_k)} \geqslant \mathrm{e}^{tx}\Big)\\ &\leqslant \mathrm{e}^{-tx}\cdot \mathbb{E}\mathrm{e}^{\lambda_{\max}\left(t\sum_{k=1}^{n}\mathbf{X}_k\right)}\end{aligned}$$

Matrix Chernoff Bound

Ahlswede and Winter (2002) prove the following *matrix* Chernoff bound:

$$\mathbb{P}\Big(\lambda_{\max}\big(\sum_{k=1}^{n}\mathbf{X}_k\big) \geqslant x\Big) \leqslant \inf_{t>0}\Big\{e^{-tx}\cdot\mathbb{E}\mathrm{Tr}\exp\big(t\sum_{i=1}^{n}\mathbf{X}_i\big)\Big\}$$

For any $t > 0$,

$$\begin{aligned}\mathbb{P}\Big(\lambda_{\max}\big(\sum_{k=1}^{n}\mathbf{X}_k\big) \geqslant x\Big) &= \mathbb{P}\Big(e^{\lambda_{\max}(t\sum_{k=1}^{n}\mathbf{X}_k)} \geqslant e^{tx}\Big) \\ &\leqslant e^{-tx}\cdot\mathbb{E}\lambda_{\max}\big(e^{t\sum_{k=1}^{n}\mathbf{X}_k}\big)\end{aligned}$$

Matrix Chernoff Bound

Ahlswede and Winter (2002) prove the following *matrix* Chernoff bound:

$$\mathbb{P}\Big(\lambda_{\max}\big(\sum_{k=1}^{n}\mathbf{X}_k\big) \geqslant x\Big) \leqslant \inf_{t>0}\Big\{\mathrm{e}^{-tx}\cdot \mathbb{E}\mathrm{Tr}\exp\big(t\sum_{i=1}^{n}\mathbf{X}_i\big)\Big\}$$

For any $t > 0$,

$$\begin{aligned}\mathbb{P}\Big(\lambda_{\max}\big(\sum_{k=1}^{n}\mathbf{X}_k\big) \geqslant x\Big) &= \mathbb{P}\Big(\mathrm{e}^{\lambda_{\max}(t\sum_{k=1}^{n}\mathbf{X}_k)} \geqslant \mathrm{e}^{tx}\Big)\\ &\leqslant \mathrm{e}^{-tx}\cdot\mathbb{E}\lambda_{\max}\big(\mathrm{e}^{t\sum_{k=1}^{n}\mathbf{X}_k}\big)\\ &\leqslant \mathrm{e}^{-tx}\cdot\mathbb{E}\mathrm{Tr}\big(\mathrm{e}^{t\sum_{k=1}^{n}\mathbf{X}_k}\big)\end{aligned}$$

Matrix Chernoff Bound

Ahlswede and Winter (2002) prove the following *matrix* Chernoff bound:

$$\mathbb{P}\Big(\lambda_{\max}\big(\sum_{k=1}^{n}\mathbf{X}_k\big) \geqslant x\Big) \leqslant \inf_{t>0}\Big\{e^{-tx}\cdot \mathbb{E}\mathrm{Tr}\exp\big(t\sum_{i=1}^{n}\mathbf{X}_i\big)\Big\}$$

For any $t > 0$,

$$\begin{aligned}\mathbb{P}\Big(\lambda_{\max}\big(\sum_{k=1}^{n}\mathbf{X}_k\big) \geqslant x\Big) &= \mathbb{P}\Big(e^{\lambda_{\max}(t\sum_{k=1}^{n}\mathbf{X}_k)} \geqslant e^{tx}\Big)\\ &\leqslant e^{-tx}\cdot \mathbb{E}\lambda_{\max}\big(e^{t\sum_{k=1}^{n}\mathbf{X}_k}\big)\\ &\leqslant e^{-tx}\cdot \mathbb{E}\mathrm{Tr}\big(e^{t\sum_{k=1}^{n}\mathbf{X}_k}\big)\end{aligned}$$

Aim: Give a suitable bound for

$$L_n(t) := \mathbb{E}\,\mathrm{Tr}\exp\big(t\sum_{i=1}^{n}\mathbf{X}_i\big)$$

On the matrix exponential

- The matrix exponential is given by

$$\mathbf{A} \mapsto e^{\mathbf{A}} := Id + \sum_{k=1}^{\infty} \frac{\mathbf{A}^k}{k!}$$

On the matrix exponential

- The matrix exponential is given by

$$\mathbf{A} \mapsto e^{\mathbf{A}} := Id + \sum_{k=1}^{\infty} \frac{\mathbf{A}^k}{k!}$$

- The matrix exponential is positive definite but it is NOT increasing

$$\mathbf{A} \preceq \mathbf{B} \quad \not\Rightarrow \quad \mathbf{A}^k \preceq \mathbf{B}^k$$

On the matrix exponential

- The matrix exponential is given by

$$\mathbf{A} \mapsto e^{\mathbf{A}} := Id + \sum_{k=1}^{\infty} \frac{\mathbf{A}^k}{k!}$$

- The matrix exponential is positive definite but it is NOT increasing

$$\mathbf{A} \preceq \mathbf{B} \quad \not\Rightarrow \quad \mathbf{A}^k \preceq \mathbf{B}^k$$

NOR convex

$$\Big(\frac{1}{2}\mathbf{A} + \frac{1}{2}\mathbf{B}\Big)^k \not\preceq \frac{1}{2}\mathbf{A}^k + \frac{1}{2}\mathbf{B}^k \quad \text{for } k > 2\,.$$

On the Trace exponential

- The Trace exponential is increasing and convex

$$\mathbf{A} \preceq \mathbf{B} \quad \Longrightarrow \quad \operatorname{Tr}\exp(\mathbf{A}) \leq \operatorname{Tr}\exp(\mathbf{B})$$

and for any $t \in [0, 1]$,

$$\operatorname{Tr}\exp(t\mathbf{A} + (1-t)\mathbf{B}) \leq t \operatorname{Tr}\exp(\mathbf{A}) + (1-t)\operatorname{Tr}\exp(\mathbf{B})\,.$$

On the Trace exponential

- The Trace exponential is increasing and convex

$$\mathbf{A} \preceq \mathbf{B} \quad \Longrightarrow \quad \mathrm{Tr}\exp(\mathbf{A}) \le \mathrm{Tr}\exp(\mathbf{B})$$

  and for any $t \in [0, 1]$,

$$\mathrm{Tr}\exp(t\mathbf{A} + (1-t)\mathbf{B}) \le t\,\mathrm{Tr}\exp(\mathbf{A}) + (1-t)\mathrm{Tr}\exp(\mathbf{B}) .$$

- Jensen's inequality for the Trace exponential function yields

$$\mathrm{Tr}\exp(\mathbb{E}\mathbf{A}) \le \mathbb{E}\mathrm{Tr}\exp(\mathbf{A}) .$$

The Golden-Thompson Inequality

The matrix exponential does not convert sums into products!

$$e^{\mathbf{A}+\mathbf{B}} \neq e^{\mathbf{A}} \cdot e^{\mathbf{B}}$$

The Golden-Thompson Inequality

The matrix exponential does not convert sums into products!

$$\mathrm{e}^{\mathbf{A}+\mathbf{B}} \neq \mathrm{e}^{\mathbf{A}} \cdot \mathrm{e}^{\mathbf{B}}$$

The Golden-Thompson (1965) inequality:

$$\mathrm{Tr}\big(\mathrm{e}^{\mathbf{A}+\mathbf{B}}\big) \leqslant \mathrm{Tr}\big(\mathrm{e}^{\mathbf{A}} \cdot \mathrm{e}^{\mathbf{B}}\big)$$

The Golden-Thompson Inequality

The matrix exponential does not convert sums into products!

$$\mathrm{e}^{\mathbf{A}+\mathbf{B}} \neq \mathrm{e}^{\mathbf{A}} \cdot \mathrm{e}^{\mathbf{B}}$$

The Golden-Thompson (1965) inequality:

$$\mathrm{Tr}\big(\mathrm{e}^{\mathbf{A}+\mathbf{B}}\big) \leqslant \mathrm{Tr}\big(\mathrm{e}^{\mathbf{A}} \cdot \mathrm{e}^{\mathbf{B}}\big)$$

This inequality fails for more than *two* matrices

$$\mathrm{Tr}\big(\mathrm{e}^{\mathbf{A}+\mathbf{B}+\mathbf{C}}\big) \not\leqslant \mathrm{Tr}\big(\mathrm{e}^{\mathbf{A}} \cdot \mathrm{e}^{\mathbf{B}} \cdot \mathrm{e}^{\mathbf{C}}\big)$$

Independent Matrix Case

Alshwede-Winter's approach

Let $\mathbf{X}_1, \ldots, \mathbf{X}_n$ be $d \times d$ mutually *independent centered* Hermitian random matrices.

Independent Matrix Case

Alshwede-Winter's approach

Let $\mathbf{X}_1, \ldots, \mathbf{X}_n$ be $d \times d$ mutually *independent centered* Hermitian random matrices.

Aim: Control $L_n(t) = \mathbb{E}\,\mathrm{Tr}\exp\left(t\sum_{k=1}^{n}\mathbf{X}_k\right)$.

Independent Matrix Case
Alshwede-Winter's approach

Let $\mathbf{X}_1, \dots, \mathbf{X}_n$ be $d \times d$ mutually *independent centered* Hermitian random matrices.

Aim: Control $L_n(t) = \mathbb{E}\,\mathrm{Tr}\exp\big(t\sum_{k=1}^n \mathbf{X}_k\big)$.

$$\mathbb{E}\mathrm{Tr}\exp\big(t\sum_{k=1}^{n}\mathbf{X}_k\big) \le \mathbb{E}\mathrm{Tr}\big(\mathrm{e}^{t\mathbf{X}_n}\cdot \mathrm{e}^{t\sum_{k=1}^{n-1}\mathbf{X}_k}\big) \qquad \text{Golden-Thompson}$$

Independent Matrix Case
Alshwede-Winter's approach

Let $\mathbf{X}_1, \ldots, \mathbf{X}_n$ be $d \times d$ mutually *independent centered* Hermitian random matrices.

Aim: Control $L_n(t) = \mathbb{E}\,\mathrm{Tr}\exp\left(t\sum_{k=1}^n \mathbf{X}_k\right)$.

$$\mathbb{E}\mathrm{Tr}\exp\left(t\sum_{k=1}^{n}\mathbf{X}_k\right) \leq \mathbb{E}\mathrm{Tr}\left(\mathrm{e}^{t\mathbf{X}_n}\cdot \mathrm{e}^{t\sum_{k=1}^{n-1}\mathbf{X}_k}\right) \qquad \text{Golden-Thompson}$$

$$= \mathrm{Tr}\Big(\mathbb{E}(\mathrm{e}^{t\mathbf{X}_n})\cdot \mathbb{E}\big(\mathrm{e}^{t\sum_{k=1}^{n-1}\mathbf{X}_k}\big)\Big) \quad \text{Independence}$$

Independent Matrix Case
Alshwede-Winter's approach

Let $\mathbf{X}_1, \ldots, \mathbf{X}_n$ be $d \times d$ mutually *independent centered* Hermitian random matrices.

Aim: Control $L_n(t) = \mathbb{E}\,\mathrm{Tr}\exp\left(t\sum_{k=1}^n \mathbf{X}_k\right)$.

$$\mathbb{E}\mathrm{Tr}\exp\left(t\sum_{k=1}^n \mathbf{X}_k\right) \leq \mathbb{E}\mathrm{Tr}\left(\mathrm{e}^{t\mathbf{X}_n}\cdot \mathrm{e}^{t\sum_{k=1}^{n-1}\mathbf{X}_k}\right) \qquad \text{Golden-Thompson}$$

$$= \mathrm{Tr}\left(\mathbb{E}(\mathrm{e}^{t\mathbf{X}_n})\cdot \mathbb{E}\left(\mathrm{e}^{t\sum_{k=1}^{n-1}\mathbf{X}_k}\right)\right) \quad \text{Independence}$$

$$\leq \lambda_{\max}(\mathbb{E}\mathrm{e}^{t\mathbf{X}_n})\cdot \mathbb{E}\mathrm{Tr}\left(\mathrm{e}^{t\sum_{k=1}^{n-1}\mathbf{X}_k}\right).$$

The last inequality follows from:

$$\mathbb{E}(\mathrm{e}^{t\mathbf{X}_n}) \preceq \lambda_{\max}(\mathbb{E}(\mathrm{e}^{t\mathbf{X}_n}))\cdot Id\,.$$

Independent Matrix Case
Alshwede-Winter's approach

Let $\mathbf{X}_1, \ldots, \mathbf{X}_n$ be $d \times d$ mutually *independent centered* Hermitian random matrices.

Aim: Control $L_n(t) = \mathbb{E}\,\mathrm{Tr}\exp\left(t\sum_{k=1}^n \mathbf{X}_k\right)$.

$$\mathbb{E}\mathrm{Tr}\exp\left(t\sum_{k=1}^{n}\mathbf{X}_k\right) \le \mathbb{E}\mathrm{Tr}\left(\mathrm{e}^{t\mathbf{X}_n}\cdot \mathrm{e}^{t\sum_{k=1}^{n-1}\mathbf{X}_k}\right) \qquad \text{Golden-Thompson}$$

$$= \mathrm{Tr}\Big(\mathbb{E}(\mathrm{e}^{t\mathbf{X}_n})\cdot\mathbb{E}\left(\mathrm{e}^{t\sum_{k=1}^{n-1}\mathbf{X}_k}\right)\Big) \quad \text{Independence}$$

$$\le \lambda_{\max}(\mathbb{E}\mathrm{e}^{t\mathbf{X}_n})\cdot\mathbb{E}\mathrm{Tr}\left(\mathrm{e}^{t\sum_{k=1}^{n-1}\mathbf{X}_k}\right).$$

Iterating this procedure:

$$\mathbb{E}\mathrm{Tr}\exp\left(t\sum_{k=1}^{n}\mathbf{X}_k\right) \leqslant d\prod_{k=1}^{n}\lambda_{\max}(\mathbb{E}\mathrm{e}^{t\mathbf{X}_k}).$$

Construction of Cantor-type sets

$\mathbf{X}_1, \ldots, \mathbf{X}_n$ geometrically β-mixing.

Aim: Bound $\mathbb{E}\mathrm{Tr} \exp\left(t \sum_{k=1}^{n} \mathbf{X}_k\right)$.

Construction of Cantor-type sets

$\mathbf{X}_1, \ldots, \mathbf{X}_n$ geometrically β-mixing.

Aim: Bound $\mathbb{E}\mathrm{Tr}\exp\left(t\sum_{k=1}^{n}\mathbf{X}_k\right)$.

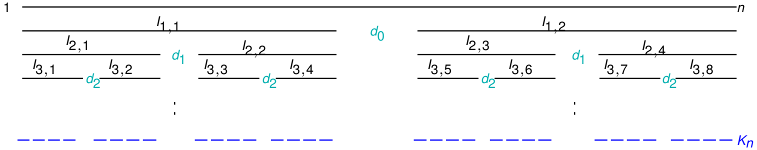


Figure: Construction of the Cantor-type set K_n

Construction of Cantor-type sets

$\mathbf{X}_1, \ldots, \mathbf{X}_n$ geometrically β-mixing.

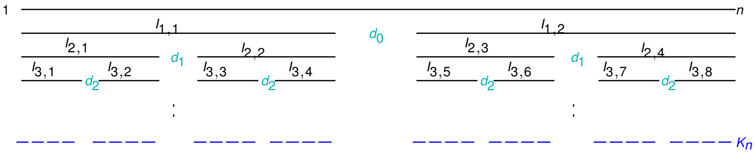


Figure: Construction of the Cantor-type set K_n

Aim: Control the Laplace transform on the Cantor set K_n

$$\mathbb{E}\mathrm{Tr}\exp\Big(t\sum_{k\in K_n}\mathbf{X}_k\Big) \le \mathbf{?}$$

Dependent matrix case
Control of the Laplace transform

- Control:

$$\mathbb{E}\mathrm{Tr}\left(e^{t\mathbf{S}_1+t\mathbf{S}_2+t\mathbf{S}_3+t\mathbf{S}_4}\right) \leq \textbf{?}$$

- Scalar Setting:

$$\mathbb{E}(e^{tS_1+tS_2+tS_3+tS_4}) \leq \mathbb{E}(e^{tS_1+tS_2}) \cdot \mathbb{E}(e^{tS_3+tS_4})(1+\epsilon(d_{2,3}))$$

Dependent matrix case
Control of the Laplace transform

- Control:

$$\mathbb{E}\mathrm{Tr}\left(e^{t\mathbf{S}_1+t\mathbf{S}_2+t\mathbf{S}_3+t\mathbf{S}_4}\right) \leq \mathbf{?}$$

- Scalar Setting:

$$\mathbb{E}(e^{tS_1+tS_2+tS_3+tS_4}) \leq \mathbb{E}(e^{tS_1+tS_2}) \cdot \mathbb{E}(e^{tS_3+tS_4})(1+\epsilon(d_{2,3}))$$

- Matrix Setting:

$$\mathbb{E}\mathrm{Tr}(e^{t\mathbf{S}_1+t\mathbf{S}_2+t\mathbf{S}_3+t\mathbf{S}_4}) \leq \mathrm{Tr}\Big(\mathbb{E}(e^{t\mathbf{S}_1+t\mathbf{S}_2}) \cdot \mathbb{E}(e^{t\mathbf{S}_3+t\mathbf{S}_4})\Big)(1+\epsilon(d_{2,3}))$$

Dependent matrix case
Control of the Laplace transform

- Control:

$$\mathbb{E}\mathrm{Tr}\left(e^{t\mathbf{S}_1+t\mathbf{S}_2+t\mathbf{S}_3+t\mathbf{S}_4}\right) \le \textbf{?}$$

- Scalar Setting:

$$\mathbb{E}(e^{tS_1+tS_2+tS_3+tS_4}) \le \mathbb{E}(e^{tS_1+tS_2}) \cdot \mathbb{E}(e^{tS_3+tS_4})(1+\epsilon(d_{2,3}))$$

- Matrix Setting:

$$\mathbb{E}\mathrm{Tr}(e^{t\mathbf{S}_1+t\mathbf{S}_2+t\mathbf{S}_3+t\mathbf{S}_4}) \le \mathrm{Tr}\Big(\mathbb{E}(e^{t\mathbf{S}_1+t\mathbf{S}_2}) \cdot \mathbb{E}(e^{t\mathbf{S}_3+t\mathbf{S}_4})\Big)(1+\epsilon(d_{2,3}))$$

We shall prove instead:

$$\mathbb{E}\mathrm{Tr}\left(\mathrm{e}^{t\mathbf{S}_1+t\mathbf{S}_2+t\mathbf{S}_3+t\mathbf{S}_4}\right) \le \mathbb{E}\mathrm{Tr}\left(\mathrm{e}^{t\mathbf{S}_1^*+t\mathbf{S}_2^*+t\mathbf{S}_3^*+t\mathbf{S}_4^*}\right)(1+\epsilon(d_{1,2}, d_{2,3}, d_{3,4}))$$

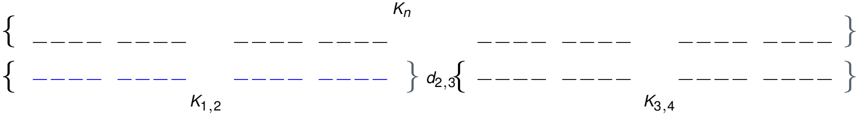
Thank you for your attention!

[1] Ahlswede, R. and Winter, A. (2002). Strong converse for identification via quantum channels. *IEEE Trans. Inform. Theory*, **48**, no. 3, 569-579.

[2] Banna M., Merlevède F. and Youssef P. (2015). Bernstein type inequality for a class of dependent random matrices. *In revision: Random Matrices: Theory and Applications*

[3] Merlevède F., Peligrad M. and Rio E. (2009). Bernstein inequality and moderate deviations under strong mixing conditions. *High dimensional probability V: the Luminy volume, 273–292, Inst. Math. Stat. Collect., 5, Inst. Math. Statist., Beachwood, OH*

[4] Tropp, J. A. (2012). User-friendly tail bounds for sums of random matrices. *Found. Comput. Math.*, **12**, no. 4, 389-434.



$$\sum_{k\in K_1} \mathbf{X}_k + \sum_{k\in K_2} \mathbf{X}_k + \sum_{k\in K_3} \mathbf{X}_k + \sum_{k\in K_4} \mathbf{X}_k = \mathbf{S}_1 + \mathbf{S}_2 + \mathbf{S}_3 + \mathbf{S}_4 \,.$$

Step 1: Berbee's coupling allow the construction of $(\mathbf{X}_k^*)_{k\in K_3\cup K_4}$ such that

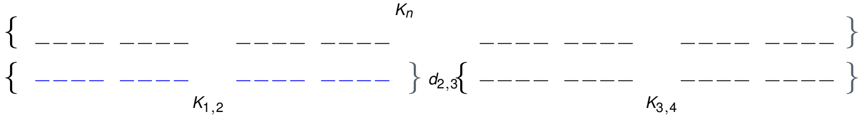
- $\mathbf{S}_3^* + \mathbf{S}_4^*$ has the same distribution as $\mathbf{S}_3 + \mathbf{S}_4$
- $\mathbf{S}_3^* + \mathbf{S}_4^*$ is independent of $\mathbf{S}_1 + \mathbf{S}_2$
- $\mathbb{P}\big(\mathbf{S}_3^* + \mathbf{S}_4^* \neq \mathbf{S}_3 + \mathbf{S}_4\big) \leqslant \beta_{d_{2,3}}$.

where

$$\mathbf{S}_3^* = \sum_{k\in K_3} \mathbf{X}_k^* \qquad \text{and} \qquad \mathbf{S}_4^* = \sum_{k\in K_4} \mathbf{X}_k^*$$

Dependent matrix case

Berbee's Coupling



$$\sum_{k\in K_1}\mathbf{X}_k+\sum_{k\in K_2}\mathbf{X}_k+\sum_{k\in K_3}\mathbf{X}_k+\sum_{k\in K_4}\mathbf{X}_k=\mathbf{S}_1+\mathbf{S}_2+\mathbf{S}_3+\mathbf{S}_4\,.$$

This allows us to replace up to an error depending on $d_{2,3}$ and $|K_3|+|K_4|$

$\mathbb{E}\,\mathrm{Tr}\exp(t\mathbf{S}_1+t\mathbf{S}_2+t\mathbf{S}_3+t\mathbf{S}_4)$ by $\mathbb{E}\,\mathrm{Tr}\exp(t\mathbf{S}_1+t\mathbf{S}_2+t\mathbf{S}_3^*+t\mathbf{S}_4^*)$