Bernstein inequality: non-commutative setting and dependence

Pierre Youssef
joint work with Marwa Banna and Florence Merlevède

LPMA, Université Paris-Diderot

December 1, 2015

Introduction and Motivation

- Let $(X_k)_k$ be a sequence of i.i.d. random variables with common mean μ and variance σ^2.

Introduction and Motivation

- Let $(X_k)_k$ be a sequence of i.i.d. random variables with common mean μ and variance σ^2.
- By the central limit theorem

$$\sqrt{n}\Big(\frac{1}{n}\sum_{k=1}^{n} X_k - \mu\Big) \xrightarrow[n\to\infty]{\mathcal{L}} \mathcal{N}(0,\sigma^2)\,.$$

Introduction and Motivation

- Let $(X_k)_k$be a sequence of i.i.d. random variables with common mean μ and variance σ^2.
- By the central limit theorem

$$\sqrt{n}\Big(\frac{1}{n}\sum_{k=1}^{n} X_k - \mu\Big) \xrightarrow[n\to\infty]{\mathcal{L}} \mathcal{N}(0,\sigma^2)\,.$$

- For any n,

$$\mathbb{P}\Big(\sum_{k=1}^{n} X_k - n\mu > x\Big) \leqslant \textbf{?}$$

- A long history:

Cramér, Benett, Hoeffding, Nagaev, Bernstein, . . .

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 $V_n := \frac{1}{n}\sum_{k=1}^{n}\sigma_k^2 = \frac{1}{n}\mathbb{E}\big(\sum_{k=1}^{n} X_k\big)^2.$

Matrix setting

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

Matrix setting

- Let $\mathbf{X}_1, \ldots, \mathbf{X}_n$ be $d \times d$ *centered* Hermitian random matrices.
- Bounded above: $\lambda_{\max}(\mathbf{X}_i) \leq 1$ for any $i \leq n$.

Matrix setting

- Let $\mathbf{X}_1, \ldots, \mathbf{X}_n$ be $d \times d$ *centered* Hermitian random matrices.
- Bounded above: $\lambda_{\max}(\mathbf{X}_i) \leq 1$ for any $i \leq n$.
- Goal: $\mathbb{P}\Big(\lambda_{\max}\big(\sum_{k=1}^{n} \mathbf{X}_k\big) \geqslant x\Big) \leqslant$ **?**

Matrix setting

- Let $\mathbf{X}_1, \ldots, \mathbf{X}_n$ be $d \times d$ *centered* Hermitian random matrices.
- Bounded above: $\lambda_{\max}(\mathbf{X}_i) \leq 1$ for any $i \leq n$.
- Goal: $\mathbb{P}\Big(\lambda_{\max}\big(\sum_{k=1}^{n} \mathbf{X}_k\big) \geqslant x\Big) \leqslant \mathbf{?}$
- What should the deviation depend on?

$$\mathbb{E}\lambda_{\max}\big(\sum_{k=1}^{n} \mathbf{X}_k\big)^2 \sim \mathbf{?}.$$

$$\mathbb{E}\lambda_{\max}\Big(\sum_{k=1}^{n}\mathbf{X}_k\Big)^2 = \mathbb{E}\Big\|\sum_{k=1}^{n}\mathbf{X}_k\Big\|^2 = \mathbb{E}\Big\|\sum_{k=1}^{n}\mathbf{X}_k - \mathbb{E}\mathbf{X}_k\Big\|^2$$

$$\mathbb{E}\lambda_{\max}\Big(\sum_{k=1}^{n}\mathbf{X}_k\Big)^2 = \mathbb{E}\Big\|\sum_{k=1}^{n}\mathbf{X}_k\Big\|^2 = \mathbb{E}\Big\|\sum_{k=1}^{n}\mathbf{X}_k - \mathbb{E}\mathbf{X}_k\Big\|^2$$

$$\leq \mathbb{E}\Big\|\sum_{k=1}^{n}\mathbf{X}_k - \mathbf{X}'_k\Big\|^2 \qquad (X'_k \text{ independent copy of } X_k)$$

$$\mathbb{E}\lambda_{\max}\Big(\sum_{k=1}^{n}\mathbf{X}_k\Big)^2 = \mathbb{E}\Big\|\sum_{k=1}^{n}\mathbf{X}_k\Big\|^2 = \mathbb{E}\Big\|\sum_{k=1}^{n}\mathbf{X}_k - \mathbb{E}\mathbf{X}_k\Big\|^2$$

$$\le \mathbb{E}\Big\|\sum_{k=1}^{n}\mathbf{X}_k - \mathbf{X}'_k\Big\|^2 \qquad (X'_k \text{ independent copy of } X_k)$$

$$\le \mathbb{E}\Big\|\sum_{k=1}^{n}\varepsilon_k(\mathbf{X}_k - \mathbf{X}'_k)\Big\|^2 \quad (\varepsilon_k \text{ independent Rademacher})$$

$$
\begin{aligned}
\mathbb{E}\lambda_{\max}\Big(\sum_{k=1}^{n}\mathbf{X}_k\Big)^2 &= \mathbb{E}\Big\|\sum_{k=1}^{n}\mathbf{X}_k\Big\|^2 = \mathbb{E}\Big\|\sum_{k=1}^{n}\mathbf{X}_k - \mathbb{E}\mathbf{X}_k\Big\|^2 \\
&\le \mathbb{E}\Big\|\sum_{k=1}^{n}\mathbf{X}_k - \mathbf{X}_k'\Big\|^2 \qquad (X_k' \text{ independent copy of } X_k) \\
&\le \mathbb{E}\Big\|\sum_{k=1}^{n}\varepsilon_k(\mathbf{X}_k - \mathbf{X}_k')\Big\|^2 \quad (\varepsilon_k \text{ independent Rademacher}) \\
&\le 4\mathbb{E}\Big\|\sum_{k=1}^{n}\varepsilon_k\mathbf{X}_k\Big\|^2
\end{aligned}
$$

$$
\begin{aligned}
\mathbb{E}\lambda_{\max}\Big(\sum_{k=1}^{n}\mathbf{X}_k\Big)^2 &= \mathbb{E}\Big\|\sum_{k=1}^{n}\mathbf{X}_k\Big\|^2 = \mathbb{E}\Big\|\sum_{k=1}^{n}\mathbf{X}_k - \mathbb{E}\mathbf{X}_k\Big\|^2 \\
&\leq \mathbb{E}\Big\|\sum_{k=1}^{n}\mathbf{X}_k - \mathbf{X}'_k\Big\|^2 \qquad (X'_k \text{ independent copy of } X_k) \\
&\leq \mathbb{E}\Big\|\sum_{k=1}^{n}\varepsilon_k(\mathbf{X}_k - \mathbf{X}'_k)\Big\|^2 \quad (\varepsilon_k \text{ independent Rademacher}) \\
&\leq 4\mathbb{E}\Big\|\sum_{k=1}^{n}\varepsilon_k\mathbf{X}_k\Big\|^2 \\
&\sim \mathbb{E}\Big\|\sum_{k=1}^{n}\varepsilon_k\mathbf{X}_k\Big\|_{S_p}^2 \qquad (p \sim \log d)
\end{aligned}
$$

NC Khintchine inequality (Lust-Picard'86, LP-Pisier'91) $\Rightarrow$

$$\mathbb{E}_\varepsilon \Big\| \sum_{k=1}^{n} \varepsilon_k \mathbf{X}_k \Big\|_{S_p} \lesssim \sqrt{p} \, \Big\| \Big(\sum_{k=1}^{n} \mathbf{X}_k^2 \Big)^{\frac{1}{2}} \Big\|_{S_p}$$

NC Khintchine inequality (Lust-Picard'86, LP-Pisier'91) $\Rightarrow$

$$\mathbb{E}_\varepsilon \Big\| \sum_{k=1}^{n} \varepsilon_k \mathbf{X}_k \Big\|_{S_p} \lesssim \sqrt{p} \, \Big\| \Big(\sum_{k=1}^{n} \mathbf{X}_k^2 \Big)^{\frac{1}{2}} \Big\|_{S_p}$$

Using this, we get

$$\mathbb{E}\lambda_{\max}\Big(\sum_{k=1}^{n} \mathbf{X}_k \Big)^2 \lesssim \log d \, \mathbb{E} \Big\| \sum_{k=1}^{n} \mathbf{X}_k^2 \Big\|$$

$$\lesssim \log d \, \Big\| \sum_{k=1}^{n} \mathbb{E}\mathbf{X}_k^2 \Big\| + \log d \, \mathbb{E} \Big\| \sum_{k=1}^{n} \mathbf{X}_k^2 - \mathbb{E}\mathbf{X}_k^2 \Big\|$$

NC Khintchine inequality (Lust-Picard'86, LP-Pisier'91) $\Rightarrow$

$$\mathbb{E}_\varepsilon \Big\| \sum_{k=1}^n \varepsilon_k \mathbf{X}_k \Big\|_{S_p} \lesssim \sqrt{p} \Big\| \Big(\sum_{k=1}^n \mathbf{X}_k^2 \Big)^{\frac{1}{2}} \Big\|_{S_p}$$

Using this, we get

$$\mathbb{E}\lambda_{\max}\Big(\sum_{k=1}^n \mathbf{X}_k \Big)^2 \lesssim \log d \, \mathbb{E} \Big\| \sum_{k=1}^n \mathbf{X}_k^2 \Big\|$$

$$\lesssim \log d \, \Big\| \sum_{k=1}^n \mathbb{E}\mathbf{X}_k^2 \Big\| + \log d \, \mathbb{E} \Big\| \sum_{k=1}^n \mathbf{X}_k^2 - \mathbb{E}\mathbf{X}_k^2 \Big\|$$

Using again NC Khintchine and boundness of X_k $\Rightarrow$

$$\mathbb{E}\lambda_{\max}\Big(\sum_{k=1}^n \mathbf{X}_k \Big)^2 \lesssim \log d \, \Big\| \sum_{k=1}^n \mathbb{E}\mathbf{X}_k^2 \Big\|$$

Independent Matrix Case

Theorem (Ahlswede-Winter'02, Oliveira'11, Tropp'11)

For $\{\mathbf{X}_k\}_k$ of independent self-adjoint random matrices with dimension d. Assume that each matrix satisfies

$$\mathbb{E}\mathbf{X}_k = 0 \quad \text{and} \quad \lambda_{\max}(\mathbf{X}_k) \le M \text{ almost surely.}$$

Then for any $x > 0$,

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

where $\sigma^2 := \frac{1}{n}\|\sum_{k=1}^{n} \mathbb{E}\mathbf{X}_k^2\|$.

How can it be used?

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

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

In particular, $\mathbb{E}\|Y\|_2^2 = \mathbb{E}\mathrm{Tr}(YY^t) = d$.

How can it be used?

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

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

In particular, $\mathbb{E}\|Y\|_2^2 = \mathbb{E}\mathrm{Tr}(YY^t) = d$.

Suppose that $\|Y\|_2 \lesssim \sqrt{d}$ a.s.

How can it be used?

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

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

In particular, $\mathbb{E}\|Y\|_2^2 = \mathbb{E}\mathrm{Tr}(YY^t) = d$.

Suppose that $\|Y\|_2 \lesssim \sqrt{d}$ a.s.

Question: How many independent copies to approximate $\mathbb{E}YY^t$?

$Y_1, \ldots, Y_n$ independent copies of Y. Find minimal n such that

$$\left\|\frac{1}{n}\sum_{k=1}^{n} Y_k Y_k^t - Id\right\| \quad \text{is small.}$$

How can it be used?

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

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

In particular, $\mathbb{E}\|Y\|_2^2 = \mathbb{E}\mathrm{Tr}(YY^t) = d$.

Suppose that $\|Y\|_2 \lesssim \sqrt{d}$ a.s.

Question: How many independent copies to approximate $\mathbb{E}YY^t$?

$Y_1, \ldots, Y_n$ independent copies of Y. Find minimal n such that

$$\left\| \frac{1}{n} \sum_{k=1}^{n} Y_k Y_k^t - Id \right\| \quad \text{is small.}$$

Set $\mathbf{X}_k = \frac{1}{n}(Y_k Y_k^t - Id)$. Need that $\left\| \sum_{k=1}^{n} \mathbf{X}_k \right\|$ is small.

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

- $\mathbf{X}_k = \frac{1}{n}(Y_k Y_k^t - Id)$ centered.
- Bounded: $\|Y_k\|_2 \lesssim \sqrt{d} \ \Rightarrow \ \|\mathbf{X}_k\| \lesssim \frac{d}{n}$.

- $\mathbf{X}_k = \frac{1}{n}(Y_k Y_k^t - Id)$ centered.
- Bounded: $\|Y_k\|_2 \lesssim \sqrt{d} \;\Rightarrow\; \|\mathbf{X}_k\| \lesssim \frac{d}{n}$.
- $\mathbb{E}\mathbf{X}_k^2 = \frac{1}{n^2}\left[\mathbb{E}\|Y_k\|_2^2 Y_k Y_k^t + Id - 2\mathbb{E} Y_k Y_k^t\right] \preceq \frac{d}{n^2} Id$.
- $n\sigma^2 = \|\sum_{k=1}^{n} \mathbb{E}\mathbf{X}_k^2\| \leq \frac{d}{n}$.

- $\mathbf{X}_k = \frac{1}{n}(Y_k Y_k^t - Id)$ centered.
- Bounded: $\|Y_k\|_2 \lesssim \sqrt{d} \;\Rightarrow\; \|\mathbf{X}_k\| \lesssim \frac{d}{n}$.
- $\mathbb{E}\mathbf{X}_k^2 = \frac{1}{n^2}\left[\mathbb{E}\|Y_k\|_2^2 Y_k Y_k^t + Id - 2\mathbb{E} Y_k Y_k^t\right] \preceq \frac{d}{n^2} Id$.
- $n\sigma^2 = \|\sum_{k=1}^{n} \mathbb{E}\mathbf{X}_k^2\| \leq \frac{d}{n}$.
- Bernstein $\Rightarrow \mathbb{P}\Big(\|\frac{1}{n}\sum_{k=1}^{n} Y_k Y_k^t - Id\| \geq \sqrt{\frac{d \log d}{n}}\Big) \lesssim \frac{1}{d}$.

- $\mathbf{X}_k = \frac{1}{n}(Y_k Y_k^t - Id)$ centered.
- Bounded: $\|Y_k\|_2 \lesssim \sqrt{d} \;\Rightarrow\; \|\mathbf{X}_k\| \lesssim \frac{d}{n}$.
- $\mathbb{E}\mathbf{X}_k^2 = \frac{1}{n^2}\left[\mathbb{E}\|Y_k\|_2^2 Y_k Y_k^t + Id - 2\mathbb{E} Y_k Y_k^t\right] \preceq \frac{d}{n^2} Id$.
- $n\sigma^2 = \|\sum_{k=1}^{n} \mathbb{E}\mathbf{X}_k^2\| \leq \dfrac{d}{n}$.
- Bernstein $\Rightarrow \mathbb{P}\Big(\|\dfrac{1}{n}\sum_{k=1}^{n} Y_k Y_k^t - Id\| \geq \sqrt{\dfrac{d \log d}{n}}\Big) \lesssim \dfrac{1}{d}$.

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

Rudelson'99: Approximation of identity compositions, reducing contact points of a convex body.

Dependent Matrix Case

- Let $\mathbf{X}_1, \ldots, \mathbf{X}_n$ be $d \times d$ *geometrically* β*-mixing* Hermitian random matrices.

Dependent Matrix Case

- Let $\mathbf{X}_1, \ldots, \mathbf{X}_n$ be $d \times d$ *geometrically β-mixing* Hermitian random matrices.
- 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

- Let $\mathbf{X}_1, \ldots, \mathbf{X}_n$ be $d \times d$ *geometrically β-mixing* Hermitian random matrices.
- 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 of $(\beta_k)_k$ associated with $(\mathbf{X}_k)_k$ is defined by

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

Dependent Matrix Case

- Let $\mathbf{X}_1, \ldots, \mathbf{X}_n$ be $d \times d$ *geometrically β-mixing* Hermitian random matrices.
- 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 of $(\beta_k)_k$ associated with $(\mathbf{X}_k)_k$ is defined by

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

- The term *geometrically* β-mixing means that

$$\beta_k \leq \mathrm{e}^{-ck}$$

  for some positive constant c.

Dependent Matrix Case

Theorem (Banna-Merlevède-Y'15)

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

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

Then for any $x > 0$,

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

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

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

Consequence

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

- $d \log^3 d \ll n$.

Consequence

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

- $d \log^3 d \ll n$.
- 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:

- $d \log^3 d \ll n$.
- 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.

Consequence

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

- $d \log^3 d \ll n$.
- $Y_i =$ 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 A be a $d \times n$ random matrix such that:

- $d \log^3 d \ll n$.
- 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.

Then with high probability

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

$\Rightarrow \frac{1}{\sqrt{n}} A$ is almost an isometry.

Matrix setting

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

What can be said about

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

Matrix setting

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

What can be said about

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

For any $t > 0$,

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

Matrix setting

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

What can be said about

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

For any $t > 0$,

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

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

Matrix setting

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

What can be said about

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

For any $t > 0$,

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

Matrix setting

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

What can be said about

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

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

True for any t, then

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

Matrix Chernoff Bound

True for any t, then

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

Aim: Give a suitable bound for

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

The Trace exponential operator function

- The Trace exponential is increasing and convex

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

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

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

The Trace exponential operator function

- The Trace exponential is increasing and convex

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

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

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

- Jensen's inequality for the Trace exponential function yields

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

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 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 (1965) inequality:

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

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 (1965) inequality:

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

This inequality fails for more than *two* matrices

$$\mathrm{Tr}\left(e^{\mathbf{A}+\mathbf{B}+\mathbf{C}}\right) \nleqslant \mathrm{Tr}\left(e^{\mathbf{A}} \cdot e^{\mathbf{B}} \cdot e^{\mathbf{C}}\right)$$

Construction of Cantor-type sets

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

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

Construction of Cantor-type sets

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

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

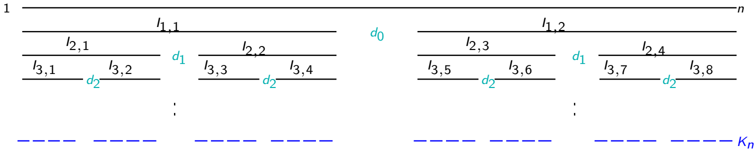


Figure: Construction of the Cantor-type set K_n

Construction of Cantor-type sets

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

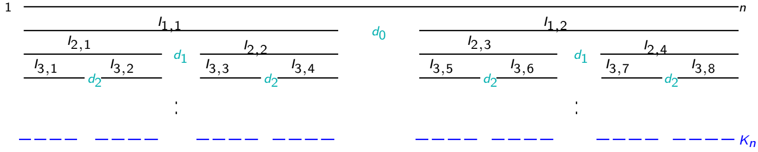


Figure: Construction of the Cantor-type set K_n

Aim: Control the Laplace transform on the Cantor set K_n

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

Gathering the upper bounds

Lemma

Let $\mathbb{U}_1, \mathbb{U}_2, \ldots$ *be* $d \times d$ *Hermitian r.m. Assume that there exist* $(\sigma_k)_{k\geq 1}$ *and* $(\kappa_k)_{k\geq 1}$ *such that for any* t *in* $[0, 1/\kappa_i[$,

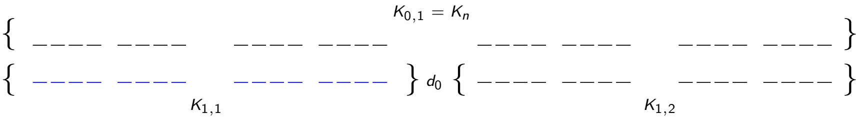
$$\log \mathbb{E}\mathrm{Tr}\left(\mathrm{e}^{t\,\mathbb{U}_i}\right) \leq C_d + \frac{(\sigma_i t)^2}{1 - \kappa_i t}\,.$$

Then, for any m *and* $t \in [0, 1/(\kappa_1 + \kappa_2 + \cdots + \kappa_m)[$,

$$\log \mathbb{E}\mathrm{Tr}\left(\mathrm{e}^{t\sum_{k=1}^{m}\mathbb{U}_k}\right) \leq C_d + \frac{(\sigma t)^2}{1 - \kappa t},$$

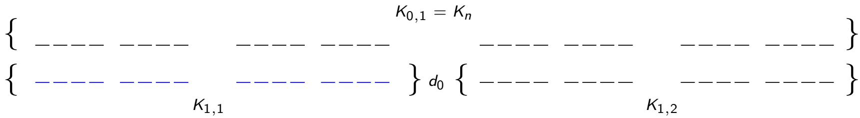
where $\sigma = \sigma_1 + \cdots + \sigma_m$ *and* $\kappa = \kappa_1 + \cdots + \kappa_m$.

Berbee Coupling



$$\sum_{k \in K_n} \mathbf{X}_k = \sum_{k \in K_{1,1}} \mathbf{X}_k + \sum_{k \in K_{1,2}} \mathbf{X}_k := \mathbf{S}_{1,1} + \mathbf{S}_{1,2}$$

Berbee Coupling

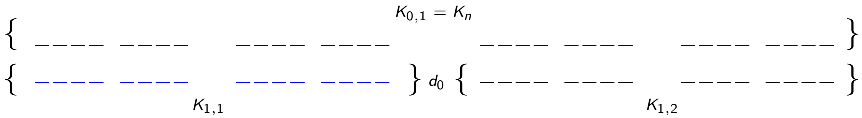


$$\sum_{k\in K_n} \mathbf{X}_k = \sum_{k\in K_{1,1}} \mathbf{X}_k + \sum_{k\in K_{1,2}} \mathbf{X}_k := \mathbf{S}_{1,1} + \mathbf{S}_{1,2}$$

Aim: Relpace up to a *small error*

$$\mathbb{E}\mathrm{Tr}\exp(t\mathbf{S}_{1,1} + t\mathbf{S}_{1,2}) \quad \text{by} \quad \mathbb{E}\mathrm{Tr}\exp(t\mathbf{S}_{1,1} + t\mathbf{S}^*_{1,2})$$

Berbee Coupling



$$\sum_{k \in K_n} \mathbf{X}_k = \sum_{k \in K_{1,1}} \mathbf{X}_k + \sum_{k \in K_{1,2}} \mathbf{X}_k := \mathbf{S}_{1,1} + \mathbf{S}_{1,2}$$

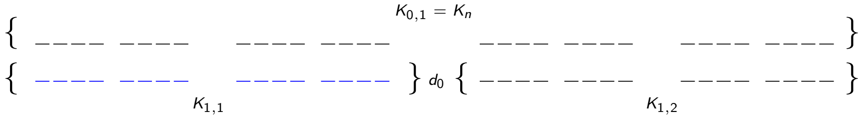
Aim: Relpace up to a *small error*

$$\mathbb{E}\mathrm{Tr}\exp(t\mathbf{S}_{1,1} + t\mathbf{S}_{1,2}) \quad \text{by} \quad \mathbb{E}\mathrm{Tr}\exp(t\mathbf{S}_{1,1} + t\mathbf{S}^*_{1,2})$$

where

- $\mathbf{S}^*_{1,2}$ has the same distribution as $\mathbf{S}_{1,2}$
- $\mathbf{S}^*_{1,2}$ is independent of $\mathbf{S}_{1,1}$

Coupling



$$\sum_{k \in K_n} \mathbf{X}_k = \sum_{k \in K_{1,1}} \mathbf{X}_k + \sum_{k \in K_{1,2}} \mathbf{X}_k := \mathbf{S}_{1,1} + \mathbf{S}_{1,2}$$

By Berbee's coupling, one can construct matrices $(\mathbf{X}_k^*)_{k \in K_{1,2}}$ such that

- $\mathbf{S}_{1,2}^*$ has the same distribution as $\mathbf{S}_{1,2}$
- $\mathbf{S}_{1,2}^*$ is independent of $\mathbf{S}_{1,1}$
- $\mathbb{P}(\mathbf{S}_{1,2}^* \neq \mathbf{S}_{1,2}) \leqslant \beta_{d_0}$,

where

$$\mathbf{S}_{1,2}^* = \sum_{k \in K_{1,2}} \mathbf{X}_k^*$$

Breaking the dependence structure

$$\mathbb{E}\mathrm{Tr}\left(\mathrm{e}^{t\sum_{k\in K_n}\mathbf{X}_k}\right) = \mathbb{E}\mathrm{Tr}\left(\mathrm{e}^{t\mathbf{S}_{1,1}+t\mathbf{S}_{1,2}}\right)$$

$$= \mathbb{E}\mathrm{Tr}\left(\mathrm{e}^{t\mathbf{S}_{1,1}+t\mathbf{S}_{1,2}}\mathbf{1}_{\mathbf{S}^*_{1,2}=\mathbf{S}_{1,2}}\right) + \mathbb{E}\mathrm{Tr}\left(\mathrm{e}^{t\mathbf{S}_{1,1}+t\mathbf{S}_{1,2}}\mathbf{1}_{\mathbf{S}^*_{1,2}\neq\mathbf{S}_{1,2}}\right)$$

Breaking the dependence structure

$$\mathbb{E}\mathrm{Tr}\left(\mathrm{e}^{t\sum_{k\in K_n}\mathbf{X}_k}\right) = \mathbb{E}\mathrm{Tr}\left(\mathrm{e}^{t\mathbf{S}_{1,1}+t\mathbf{S}_{1,2}}\right)$$

$$= \mathbb{E}\mathrm{Tr}\left(\mathrm{e}^{t\mathbf{S}_{1,1}+t\mathbf{S}_{1,2}}\mathbf{1}_{\mathbf{S}^*_{1,2}=\mathbf{S}_{1,2}}\right) + \mathbb{E}\mathrm{Tr}\left(\mathrm{e}^{t\mathbf{S}_{1,1}+t\mathbf{S}_{1,2}}\mathbf{1}_{\mathbf{S}^*_{1,2}\neq\mathbf{S}_{1,2}}\right)$$

$$\leq \mathbb{E}\mathrm{Tr}\left(\mathrm{e}^{t\mathbf{S}_{1,1}+t\mathbf{S}^*_{1,2}}\right)$$

Breaking the dependence structure

$$\mathbb{E}\mathrm{Tr}\left(\mathrm{e}^{t\sum_{k\in K_n}\mathbf{X}_k}\right) = \mathbb{E}\mathrm{Tr}\left(\mathrm{e}^{t\mathbf{S}_{1,1}+t\mathbf{S}_{1,2}}\right)$$

$$= \mathbb{E}\mathrm{Tr}\left(\mathrm{e}^{t\mathbf{S}_{1,1}+t\mathbf{S}_{1,2}}\mathbf{1}_{\mathbf{S}^*_{1,2}=\mathbf{S}_{1,2}}\right) + \mathbb{E}\mathrm{Tr}\left(\mathrm{e}^{t\mathbf{S}_{1,1}+t\mathbf{S}_{1,2}}\mathbf{1}_{\mathbf{S}^*_{1,2}\neq\mathbf{S}_{1,2}}\right)$$

$$\leq \mathbb{E}\mathrm{Tr}\left(\mathrm{e}^{t\mathbf{S}_{1,1}+t\mathbf{S}^*_{1,2}}\right) + \underbrace{\mathbb{E}\mathrm{Tr}\left(\mathrm{e}^{t\mathbf{S}_{1,1}+t\mathbf{S}_{1,2}}\mathbf{1}_{\mathbf{S}^*_{1,2}\neq\mathbf{S}_{1,2}}\right)}_{\text{Control of this term}}$$

Breaking the dependence structure

Aim: $\mathbb{E}\mathrm{Tr}\left(\mathrm{e}^{t\mathbf{S}_{1,1}+t\mathbf{S}_{1,2}}\mathbf{1}_{\mathbf{S}^*_{1,2}\neq\mathbf{S}_{1,2}}\right) \leq$ **?**

$$\mathrm{e}^{t\mathbf{S}_{1,1}+t\mathbf{S}_{1,2}} \preceq \lambda_{\max}\left(\mathrm{e}^{t\mathbf{S}_{1,1}+t\mathbf{S}_{1,2}}\right)\cdot\mathbb{I}_d$$

Breaking the dependence structure

Aim: $\mathbb{E}\mathrm{Tr}\left(\mathrm{e}^{t\mathbf{S}_{1,1}+t\mathbf{S}_{1,2}}\mathbf{1}_{\mathbf{S}^*_{1,2}\neq\mathbf{S}_{1,2}}\right) \leq$ **?**

$$\mathrm{e}^{t\mathbf{S}_{1,1}+t\mathbf{S}_{1,2}} \preceq \lambda_{\max}\left(\mathrm{e}^{t\mathbf{S}_{1,1}+t\mathbf{S}_{1,2}}\right)\cdot\mathbb{I}_d$$

Breaking the dependence structure

Aim: $\mathbb{E}\mathrm{Tr}\left(\mathrm{e}^{t\mathbf{S}_{1,1}+t\mathbf{S}_{1,2}}\mathbf{1}_{\mathbf{S}^*_{1,2}\neq\mathbf{S}_{1,2}}\right) \leq$ **?**

$$\mathrm{e}^{t\mathbf{S}_{1,1}+t\mathbf{S}_{1,2}} \preceq \lambda_{\max}\left(\mathrm{e}^{t\mathbf{S}_{1,1}+t\mathbf{S}_{1,2}}\right)\cdot\mathbb{I}_d = \mathrm{e}^{\lambda_{\max}(t\mathbf{S}_{1,1}+t\mathbf{S}_{1,2})}\cdot\mathbb{I}_d$$

Breaking the dependence structure

Aim: $\mathbb{E}\mathrm{Tr}\left(\mathrm{e}^{t\mathbf{S}_{1,1}+t\mathbf{S}_{1,2}}\mathbf{1}_{\mathbf{S}^*_{1,2}\neq\mathbf{S}_{1,2}}\right) \leq$ **?**

$$\begin{aligned}\mathrm{e}^{t\mathbf{S}_{1,1}+t\mathbf{S}_{1,2}} &\preceq \lambda_{\max}\left(\mathrm{e}^{t\mathbf{S}_{1,1}+t\mathbf{S}_{1,2}}\right)\cdot\mathbb{I}_d = \mathrm{e}^{\lambda_{\max}(t\mathbf{S}_{1,1}+t\mathbf{S}_{1,2})}\cdot\mathbb{I}_d \\ &\preceq \mathrm{e}^{t|K_n|}\cdot\mathbb{I}_d\end{aligned}$$

since $\lambda_{\max}(\mathbf{X}_k) \leq 1$ a.s. and $|K_n| = |K_{1,1}| + |K_{1,2}|$.

Breaking the dependence structure

Aim: $\mathbb{E}\mathrm{Tr}\left(\mathrm{e}^{t\mathbf{S}_{1,1}+t\mathbf{S}_{1,2}}\mathbf{1}_{\mathbf{S}^*_{1,2}\neq\mathbf{S}_{1,2}}\right) \leq \textbf{?}$

$$\mathrm{e}^{t\mathbf{S}_{1,1}+t\mathbf{S}_{1,2}} \preceq \lambda_{\max}\left(\mathrm{e}^{t\mathbf{S}_{1,1}+t\mathbf{S}_{1,2}}\right)\cdot\mathbb{I}_d = \mathrm{e}^{\lambda_{\max}(t\mathbf{S}_{1,1}+t\mathbf{S}_{1,2})}\cdot\mathbb{I}_d$$
$$\preceq \mathrm{e}^{t|K_n|}\cdot\mathbb{I}_d$$

since $\lambda_{\max}(\mathbf{X}_k) \leq 1$ a.s. and $|K_n| = |K_{1,1}| + |K_{1,2}|$.

Then

$$\mathbb{E}\mathrm{Tr}\left(\mathrm{e}^{t\mathbf{S}_{1,1}+t\mathbf{S}_{1,2}}\mathbf{1}_{\mathbf{S}^*_{1,2}\neq\mathbf{S}_{1,2}}\right) \leq \mathrm{e}^{t|K_n|}\,\mathbb{E}(\mathbf{1}_{\mathbf{S}^*_{1,2}\neq\mathbf{S}_{1,2}})\cdot\mathrm{Tr}(\mathbb{I}_d)$$

Breaking the dependence structure

Aim: $\mathbb{E}\mathrm{Tr}\left(\mathrm{e}^{t\mathbf{S}_{1,1}+t\mathbf{S}_{1,2}}\mathbf{1}_{\mathbf{S}^*_{1,2}\neq\mathbf{S}_{1,2}}\right) \leq$ **?**

$$\begin{aligned}\mathrm{e}^{t\mathbf{S}_{1,1}+t\mathbf{S}_{1,2}} &\preceq \lambda_{\max}\left(\mathrm{e}^{t\mathbf{S}_{1,1}+t\mathbf{S}_{1,2}}\right)\cdot\mathbb{I}_d = \mathrm{e}^{\lambda_{\max}(t\mathbf{S}_{1,1}+t\mathbf{S}_{1,2})}\cdot\mathbb{I}_d \\ &\preceq \mathrm{e}^{t|K_n|}\cdot\mathbb{I}_d\end{aligned}$$

since $\lambda_{\max}(\mathbf{X}_k)\leq 1$ a.s. and $|K_n| = |K_{1,1}| + |K_{1,2}|$.

Then

$$\mathbb{E}\mathrm{Tr}\left(\mathrm{e}^{t\mathbf{S}_{1,1}+t\mathbf{S}_{1,2}}\mathbf{1}_{\mathbf{S}^*_{1,2}\neq\mathbf{S}_{1,2}}\right) \leq \mathrm{e}^{t|K_n|}\,\mathbb{P}(\mathbf{S}^*_{1,2}\neq\mathbf{S}_{1,2})\cdot\mathrm{Tr}(\mathbb{I}_d)$$

Breaking the dependence structure

Aim: $\mathbb{E}\mathrm{Tr}\left(\mathrm{e}^{t\mathbf{S}_{1,1}+t\mathbf{S}_{1,2}}\mathbf{1}_{\mathbf{S}^*_{1,2}\neq\mathbf{S}_{1,2}}\right) \leq \mathbf{?}$

$$\begin{aligned}\mathrm{e}^{t\mathbf{S}_{1,1}+t\mathbf{S}_{1,2}} &\preceq \lambda_{\max}\left(\mathrm{e}^{t\mathbf{S}_{1,1}+t\mathbf{S}_{1,2}}\right)\cdot\mathbb{I}_d = \mathrm{e}^{\lambda_{\max}(t\mathbf{S}_{1,1}+t\mathbf{S}_{1,2})}\cdot\mathbb{I}_d \\ &\preceq \mathrm{e}^{t|K_n|}\cdot\mathbb{I}_d\end{aligned}$$

since $\lambda_{\max}(\mathbf{X}_k) \leq 1$ a.s. and $|K_n| = |K_{1,1}| + |K_{1,2}|$.

Then

$$\begin{aligned}\mathbb{E}\mathrm{Tr}\left(\mathrm{e}^{t\mathbf{S}_{1,1}+t\mathbf{S}_{1,2}}\mathbf{1}_{\mathbf{S}^*_{1,2}\neq\mathbf{S}_{1,2}}\right) &\leq \mathrm{e}^{t|K_n|}\,\mathbb{P}(\mathbf{S}^*_{1,2}\neq\mathbf{S}_{1,2})\cdot\mathrm{Tr}(\mathbb{I}_d) \\ &\leq \beta_{d_0}\mathrm{e}^{t|K_n|}\cdot\mathrm{Tr}(\mathbb{I}_d)\end{aligned}$$

Breaking the dependence structure

Aim: $\mathbb{E}\mathrm{Tr}\left(\mathrm{e}^{t\mathbf{S}_{1,1}+t\mathbf{S}_{1,2}}\mathbf{1}_{\mathbf{S}^*_{1,2}\neq\mathbf{S}_{1,2}}\right) \leq \mathbf{?}$

$$\mathrm{e}^{t\mathbf{S}_{1,1}+t\mathbf{S}_{1,2}} \preceq \lambda_{\max}\left(\mathrm{e}^{t\mathbf{S}_{1,1}+t\mathbf{S}_{1,2}}\right)\cdot\mathbb{I}_d = \mathrm{e}^{\lambda_{\max}(t\mathbf{S}_{1,1}+t\mathbf{S}_{1,2})}\cdot\mathbb{I}_d$$
$$\preceq \mathrm{e}^{t|K_n|}\cdot\mathbb{I}_d$$

since $\lambda_{\max}(\mathbf{X}_k) \leq 1$ a.s. and $|K_n| = |K_{1,1}| + |K_{1,2}|$.

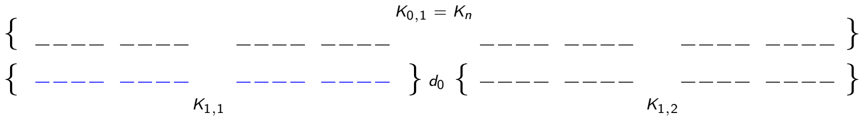
Then

$$\mathbb{E}\mathrm{Tr}\left(\mathrm{e}^{t\mathbf{S}_{1,1}+t\mathbf{S}_{1,2}}\mathbf{1}_{\mathbf{S}^*_{1,2}\neq\mathbf{S}_{1,2}}\right) \leq \mathrm{e}^{t|K_n|}\,\mathbb{P}(\mathbf{S}^*_{1,2}\neq\mathbf{S}_{1,2})\cdot\mathrm{Tr}(\mathbb{I}_d)$$
$$\leq \beta_{d_0}\mathrm{e}^{t|K_n|}\cdot\mathrm{Tr}(\mathbb{I}_d)$$

As $\mathbb{E}(\mathbf{S}_{1,1}) = \mathbb{E}(\mathbf{S}^*_{1,2}) = \mathbf{0}$ then

$$\mathrm{Tr}(\mathbb{I}_d) = \mathrm{Tr}\mathrm{e}^{t\mathbb{E}(\mathbf{S}_{1,1}+\mathbf{S}^*_{1,2})} \leq \mathbb{E}\mathrm{Tr}\mathrm{e}^{t\mathbf{S}_{1,1}+t\mathbf{S}^*_{1,2}}$$

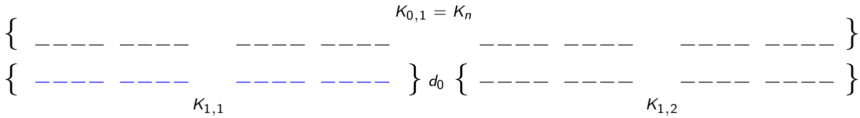
Breaking the dependence structure



$$\mathbf{S}_{1,1} = \sum_{k \in K_{1,1}} \mathbf{X}_k \quad \text{and} \quad \mathbf{S}^*_{1,2} = \sum_{k \in K_{1,2}} \mathbf{X}^*_k$$

$$\mathbb{E}\mathrm{Tr} \exp\left(t \sum_{k \in K_n} \mathbf{X}_k\right) \leqslant \left(1 + \beta_{d_0}\, \mathrm{e}^{t\,|K_{0,1}|}\right) \cdot \mathbb{E}\mathrm{Tr} \exp\left(t\mathbf{S}_{1,1} + t\mathbf{S}^*_{1,2}\right)$$

Breaking the dependence structure

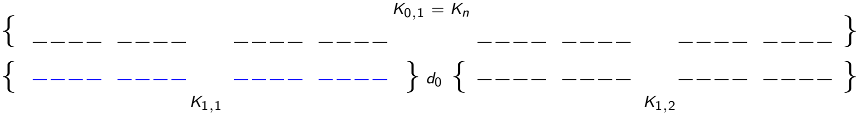


$$\mathbf{S}_{1,1} = \sum_{k \in K_{1,1}} \mathbf{X}_k \quad \text{and} \quad \mathbf{S}^*_{1,2} = \sum_{k \in K_{1,2}} \mathbf{X}^*_k$$

$$\mathbb{E}\mathrm{Tr}\exp\left(t \sum_{k \in K_n} \mathbf{X}_k\right) \leqslant \left(1 + \beta_{d_0}\, \mathrm{e}^{t\,|K_{0,1}|}\right) \cdot \mathbb{E}\mathrm{Tr}\exp\left(t\mathbf{S}_{1,1} + t\mathbf{S}^*_{1,2}\right)$$

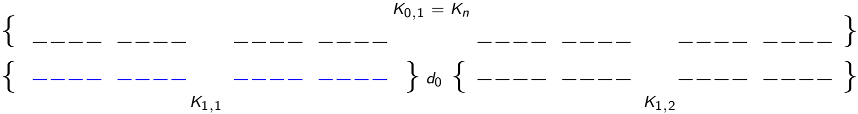
The Cantor set K_n should be constructed in a way that β_0 compensates the cardinal of $K_{0,1}$!!

Breaking the dependence structure



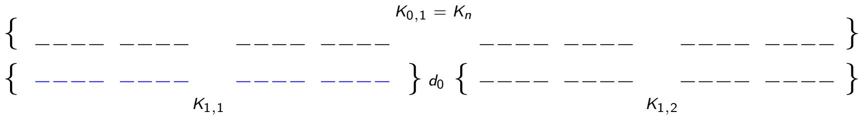
$$\mathbb{E}\mathrm{Tr}\left(\mathrm{e}^{t\mathbf{S}_{1,1}+t\mathbf{S}^*_{1,2}}\right) = \mathbb{E}\mathrm{Tr}\left(\mathrm{e}^{t\mathbf{S}_{2,1}+t\mathbf{S}_{2,2}+t\mathbf{S}^*_{1,2}}\right)$$

Breaking the dependence structure



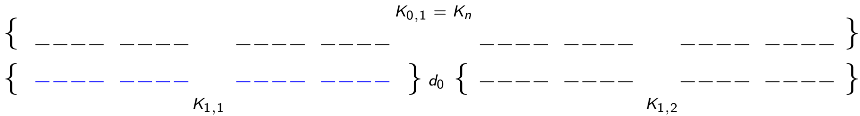
$$
\begin{aligned}
\mathbb{E}\mathrm{Tr}\left(\mathrm{e}^{t\mathbf{S}_{1,1}+t\mathbf{S}^*_{1,2}}\right) &= \mathbb{E}\mathrm{Tr}\left(\mathrm{e}^{t\mathbf{S}_{2,1}+t\mathbf{S}_{2,2}+t\mathbf{S}^*_{1,2}}\right) \\
&\leq \left(1+\beta_{d_1}\mathrm{e}^{t\,|K_{1,1}|}\right)\cdot\mathbb{E}\mathrm{Tr}\left(\mathrm{e}^{t\mathbf{S}_{2,1}+t\mathbf{S}'_{2,2}+t\mathbf{S}^*_{1,2}}\right)
\end{aligned}
$$

Breaking the dependence structure



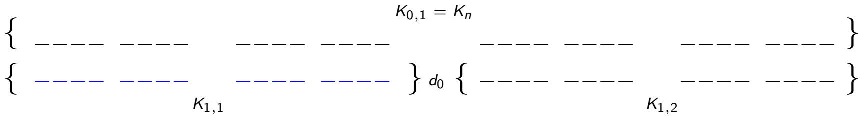
$$\begin{aligned}
\mathbb{E}\mathrm{Tr}\left(\mathrm{e}^{t\mathbf{S}_{1,1}+t\mathbf{S}^*_{1,2}}\right) &= \mathbb{E}\mathrm{Tr}\left(\mathrm{e}^{t\mathbf{S}_{2,1}+t\mathbf{S}_{2,2}+t\mathbf{S}^*_{1,2}}\right) \\
&\leq \left(1+\beta_{d_1}\mathrm{e}^{t\,|K_{1,1}|}\right)\cdot\mathbb{E}\mathrm{Tr}\left(\mathrm{e}^{t\mathbf{S}_{2,1}+t\mathbf{S}'_{2,2}+t\mathbf{S}^*_{1,2}}\right) \\
&\leq \left(1+\beta_{d_1}\mathrm{e}^{t\,|K_{1,1}|}\right)\cdot\mathbb{E}\mathrm{Tr}\left(\mathrm{e}^{t\mathbf{S}_{2,1}+t\mathbf{S}'_{2,2}+t\mathbf{S}^*_{2,3}+t\mathbf{S}^*_{2,4}}\right)
\end{aligned}$$

Breaking the dependence structure



$$
\begin{aligned}
\mathbb{E}\mathrm{Tr}\left(\mathrm{e}^{t\mathbf{S}_{1,1}+t\mathbf{S}^*_{1,2}}\right) &= \mathbb{E}\mathrm{Tr}\left(\mathrm{e}^{t\mathbf{S}_{2,1}+t\mathbf{S}_{2,2}+t\mathbf{S}^*_{1,2}}\right) \\
&\leq \left(1+\beta_{d_1}\mathrm{e}^{t\,|K_{1,1}|}\right)\cdot\mathbb{E}\mathrm{Tr}\left(\mathrm{e}^{t\mathbf{S}_{2,1}+t\mathbf{S}'_{2,2}+t\mathbf{S}^*_{1,2}}\right) \\
&\leq \left(1+\beta_{d_1}\mathrm{e}^{t\,|K_{1,1}|}\right)\cdot\mathbb{E}\mathrm{Tr}\left(\mathrm{e}^{t\mathbf{S}_{2,1}+t\mathbf{S}'_{2,2}+t\mathbf{S}^*_{2,3}+t\mathbf{S}^*_{2,4}}\right) \\
&\leq \left(1+\beta_{d_1}\mathrm{e}^{t\,|K_{1,1}|}\right)^2\cdot\mathbb{E}\mathrm{Tr}\left(\mathrm{e}^{t\mathbf{S}_{2,1}+t\mathbf{S}'_{2,2}+t\mathbf{S}^*_{2,3}+t\mathbf{S}^{**}_{2,4}}\right)
\end{aligned}
$$

Breaking the dependence structure



$$
\begin{aligned}
\mathbb{E}\mathrm{Tr}\left(\mathrm{e}^{t\mathbf{S}_{1,1}+t\mathbf{S}^*_{1,2}}\right) &= \mathbb{E}\mathrm{Tr}\left(\mathrm{e}^{t\mathbf{S}_{2,1}+t\mathbf{S}_{2,2}+t\mathbf{S}^*_{1,2}}\right) \\
&\leq \left(1+\beta_{d_1}\mathrm{e}^{t\,|K_{1,1}|}\right)\cdot \mathbb{E}\mathrm{Tr}\left(\mathrm{e}^{t\mathbf{S}_{2,1}+t\mathbf{S}'_{2,2}+t\mathbf{S}^*_{1,2}}\right) \\
&\leq \left(1+\beta_{d_1}\mathrm{e}^{t\,|K_{1,1}|}\right)\cdot \mathbb{E}\mathrm{Tr}\left(\mathrm{e}^{t\mathbf{S}_{2,1}+t\mathbf{S}'_{2,2}+t\mathbf{S}^*_{2,3}+t\mathbf{S}^*_{2,4}}\right) \\
&\leq \left(1+\beta_{d_1}\mathrm{e}^{t\,|K_{1,1}|}\right)^2\cdot \mathbb{E}\mathrm{Tr}\left(\mathrm{e}^{t\mathbf{S}_{2,1}+t\mathbf{S}'_{2,2}+t\mathbf{S}^*_{2,3}+t\mathbf{S}^{**}_{2,4}}\right)
\end{aligned}
$$

Continue procedure then use independent case.

Merci de votre attention!