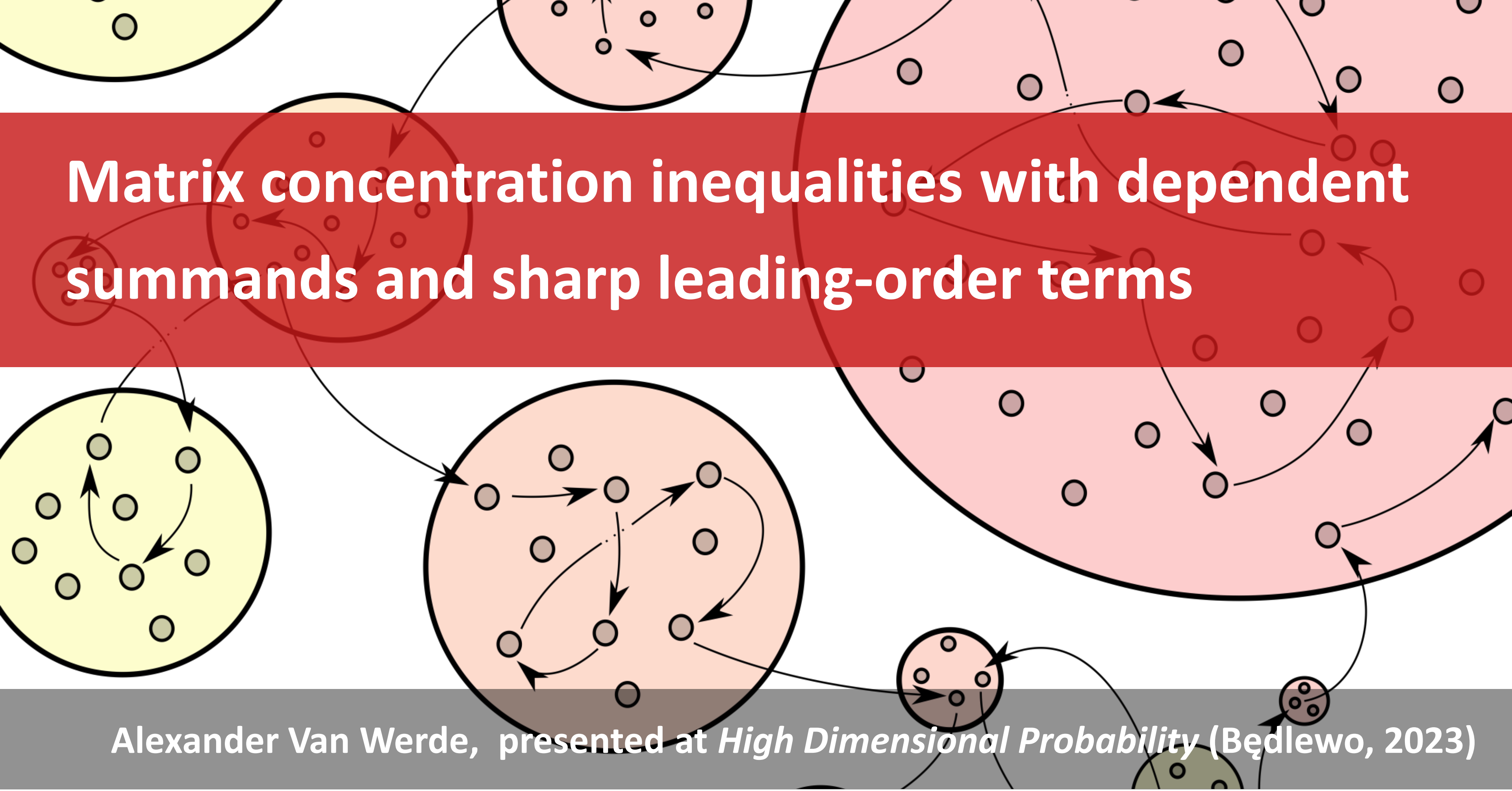


# Matrix concentration inequalities with dependent summands and sharp leading-order terms

Alexander Van Werde, presented at *High Dimensional Probability* (Będlewo, 2023)

Joint work with Jaron Sanders

TU/e EINDHOVEN UNIVERSITY OF TECHNOLOGY

# Part I

# Context & related work

# Matrix Bernstein inequality

**Theorem. (Tropp, Oliveira)**

Let $\boldsymbol{X}_1, \dots, \boldsymbol{X}_n$ be independent self-adjoint $d \times d$ matrices.
Assume $\mathbb{E}\boldsymbol{X}_i = 0$ and $\|\boldsymbol{X}_i\| \le R$. Denote $\boldsymbol{S} := \sum_{i=1}^{n} \boldsymbol{X}_i$.

Then, for some absolute $c > 0$, the operator norm satisfies

$$\mathbb{E}\|\boldsymbol{S}\| \le c\sqrt{\ln(2d)}\ \sigma + c\ln(2d)\ R$$

where $\sigma^2 := \|\,\mathbb{E}\boldsymbol{S}^2\|$.

# Matrix Bernstein inequality

**Theorem. (Tropp, Oliveira)**

Let $\boldsymbol{X}_1, \dots, \boldsymbol{X}_n$ be independent self-adjoint $d \times d$ matrices.
Assume $\mathbb{E}\boldsymbol{X}_i = 0$ and $\|\boldsymbol{X}_i\| \leq R$. Denote $\boldsymbol{S} := \sum_{i=1}^{n} \boldsymbol{X}_i$.

Then, for some absolute $c > 0$, the operator norm satisfies

$$\mathbb{E}\|\boldsymbol{S}\| \leq c\,\sqrt{\ln(2d)}\;\sigma + c\ln(2d)\;R$$

where $\sigma^2 := \|\,\mathbb{E}\boldsymbol{S}^2\|$.

# Matrix Bernstein inequality

**Theorem. (Tropp, Oliveira)**

Let $\boldsymbol{X}_1, \dots, \boldsymbol{X}_n$ be independent self-adjoint $d \times d$ matrices.
Assume $\mathbb{E}\boldsymbol{X}_i = 0$ and $\|\boldsymbol{X}_i\| \leq R$. Denote $\boldsymbol{S} := \sum_{i=1}^{n} \boldsymbol{X}_i$.

Then, for some absolute $c > 0$, the operator norm satisfies

$$\mathbb{E}\|\boldsymbol{S}\| \leq c\sqrt{\ln(2d)}\;\sigma + c\ln(2d)\;R$$

where $\sigma^2 := \|\,\mathbb{E}\boldsymbol{S}^2\|$.

# Matrix Bernstein inequality

**Theorem. (Tropp, Oliveira)**

Let $\boldsymbol{X}_1, \ldots, \boldsymbol{X}_n$ be independent self-adjoint $d \times d$ matrices.
Assume $\mathbb{E}\boldsymbol{X}_i = 0$ and $\|\boldsymbol{X}_i\| \leq R$. Denote $\boldsymbol{S} := \sum_{i=1}^{n} \boldsymbol{X}_i$.

Then, for some absolute $c > 0$, the operator norm satisfies

$$\mathbb{E}\|\boldsymbol{S}\| \leq c\sqrt{\ln(2d)}\;\sigma + c\ln(2d)\;R$$

where $\sigma^2 := \|\,\mathbb{E}\boldsymbol{S}^2\|$.

# Matrix Bernstein inequality

**Theorem. (Tropp, Oliveira)**

Let $\boldsymbol{X}_1, \dots, \boldsymbol{X}_n$ be independent self-adjoint $d \times d$ matrices.
Assume $\mathbb{E}\boldsymbol{X}_i = 0$ and $\|\boldsymbol{X}_i\| \le R$. Denote $\boldsymbol{S} := \sum_{i=1}^{n} \boldsymbol{X}_i$.

Then, for some absolute $c > 0$, the operator norm satisfies

$$\mathbb{E}\|\boldsymbol{S}\| \le c\sqrt{\ln(2d)}\ \sigma + c\ln(2d)\ R$$

where $\sigma^2 := \|\,\mathbb{E}\boldsymbol{S}^2\|$.

**Flexible:**

- Quantitative
- General-purpose

**Limitations:**

- Independence assumption
- $\sqrt{\ln(2d)}$ is often suboptimal

# Matrix Bernstein inequality

**Theorem. (Tropp, Oliveira)**

Let $\boldsymbol{X}_1, \dots, \boldsymbol{X}_n$ be independent self-adjoint $d \times d$ matrices.
Assume $\mathbb{E}\boldsymbol{X}_i = 0$ and $\|\boldsymbol{X}_i\| \leq R$. Denote $\boldsymbol{S} := \sum_{i=1}^{n} \boldsymbol{X}_i$.

Then, for some absolute $c > 0$, the operator norm satisfies

$$\mathbb{E}\|\boldsymbol{S}\| \leq c\sqrt{\ln(2d)}\,\sigma + c\ln(2d)\,R$$

where $\sigma^2 := \|\,\mathbb{E}\boldsymbol{S}^2\|$.

**Flexible:**

- Quantitative
- General-purpose

**Limitations:**

- Independence assumption
- $\sqrt{\ln(2d)}$ is often suboptimal

# Matrix Bernstein inequality

**Theorem. (Tropp, Oliveira)**

Let $\boldsymbol{X}_1, \dots, \boldsymbol{X}_n$ be independent self-adjoint $d \times d$ matrices.
Assume $\mathbb{E}\boldsymbol{X}_i = 0$ and $\|\boldsymbol{X}_i\| \leq R$. Denote $\boldsymbol{S} := \sum_{i=1}^{n} \boldsymbol{X}_i$.

Then, for some absolute $c > 0$, the operator norm satisfies

$$\mathbb{E}\|\boldsymbol{S}\| \leq c\,\sqrt{\ln(2d)}\;\sigma + c\ln(2d)\,R$$

where $\sigma^2 := \|\,\mathbb{E}\boldsymbol{S}^2\|$.

**Flexible:**

- Quantitative
- General-purpose

**Limitations:**

- Independence assumption
- $\sqrt{\ln(2d)}$ is often suboptimal

# Matrix Bernstein inequality

**Theorem. (Tropp, Oliveira)**

Let $\boldsymbol{X}_1, \ldots, \boldsymbol{X}_n$ be independent self-adjoint $d \times d$ matrices.
Assume $\mathbb{E}\boldsymbol{X}_i = 0$ and $\|\boldsymbol{X}_i\| \leq R$. Denote $\boldsymbol{S} := \sum_{i=1}^{n} \boldsymbol{X}_i$.

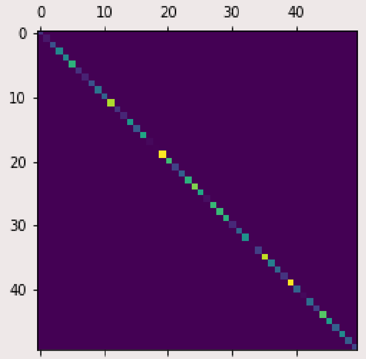
Then, for some absolute $c > 0$, the operator norm satisfies

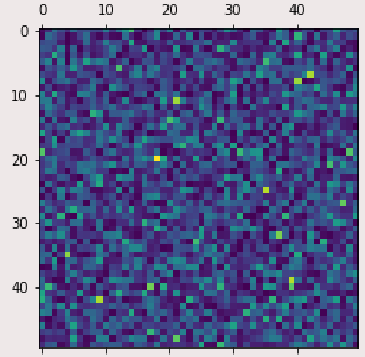
$$\mathbb{E}\|\boldsymbol{S}\| \leq c \sqrt{\ln(2d)}\ \sigma + c \ln(2d)\ R$$

where $\sigma^2 := \| \mathbb{E}\boldsymbol{S}^2 \|$.

**Flexible:**
- Quantitative
- General-purpose

**Limitations:**
- Independence assumption
- $\sqrt{\ln(2d)}$ is often suboptimal





# Matrix Bernstein inequality

**Theorem. (Tropp, Oliveira)**

Let $\boldsymbol{X}_1, \dots, \boldsymbol{X}_n$ be independent self-adjoint $d \times d$ matrices. Assume $\mathbb{E}\boldsymbol{X}_i = 0$ and $\|\boldsymbol{X}_i\| \le R$. Denote $\boldsymbol{S} := \sum_{i=1}^{n} \boldsymbol{X}_i$.

Then, for some absolute $c > 0$, the operator norm satisfies

$$\mathbb{E}\|\boldsymbol{S}\| \le c\,\sqrt{\ln(2d)}\;\sigma + c\ln(2d)\,R$$

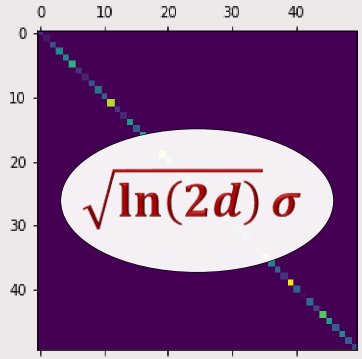
where $\sigma^2 := \|\,\mathbb{E}\boldsymbol{S}^2\|$.

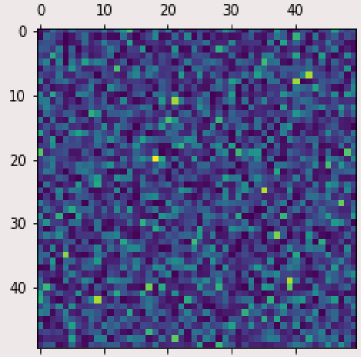
**Flexible:**

- Quantitative
- General-purpose

**Limitations:**

- Independence assumption
- $\sqrt{\ln(2d)}$ is often suboptimal





# Matrix Bernstein inequality

**Theorem. (Tropp, Oliveira)**

Let $\boldsymbol{X}_1, \ldots, \boldsymbol{X}_n$ be independent self-adjoint $d \times d$ matrices. Assume $\mathbb{E}\boldsymbol{X}_i = 0$ and $\|\boldsymbol{X}_i\| \leq R$. Denote $\boldsymbol{S} := \sum_{i=1}^{n} \boldsymbol{X}_i$.

Then, for some absolute $c > 0$, the operator norm satisfies

$$\mathbb{E}\|\boldsymbol{S}\| \leq c\,\sqrt{\ln(2d)}\;\sigma + c\ln(2d)\,R$$

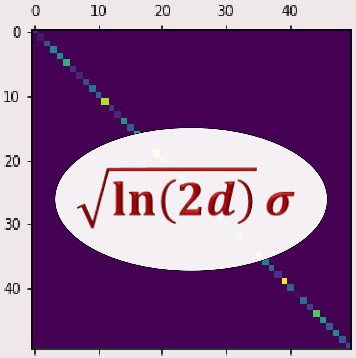
where $\sigma^2 := \|\,\mathbb{E}\boldsymbol{S}^2\|$.

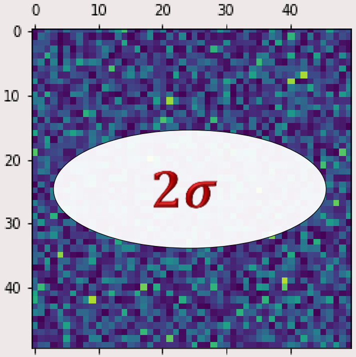
**Flexible:**

- Quantitative
- General-purpose

**Limitations:**

- Independence assumption
- $\sqrt{\ln(2d)}$ is often suboptimal





# Matrix Bernstein inequality

**Theorem. (Tropp, Oliveira)**

Let $\boldsymbol{X}_1, \dots, \boldsymbol{X}_n$ be independent self-adjoint $d \times d$ matrices. Assume $\mathbb{E}\boldsymbol{X}_i = 0$ and $\|\boldsymbol{X}_i\| \leq R$. Denote $\boldsymbol{S} := \sum_{i=1}^{n} \boldsymbol{X}_i$.

Then, for some absolute $c > 0$, the operator norm satisfies

$$\mathbb{E}\|\boldsymbol{S}\| \leq c\sqrt{\ln(2d)}\ \sigma + c\ln(2d)\ R$$

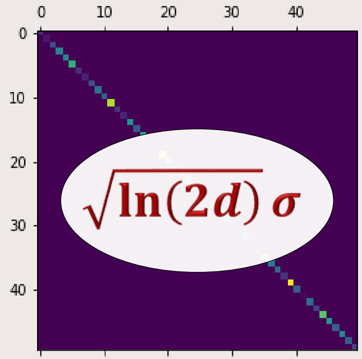
where $\sigma^2 := \|\ \mathbb{E}\boldsymbol{S}^2\|$.

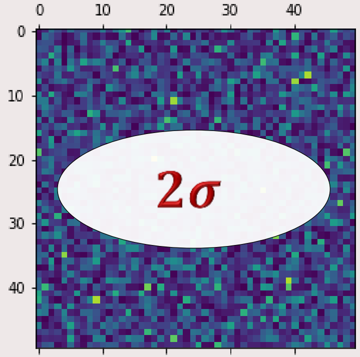
**Flexible:**

- Quantitative
- General-purpose

**Limitations:**

- Independence assumption
- $\sqrt{\ln(2d)}$ is often suboptimal





# Some related literature

## Matrix concentration with dependent summands

- **R.I. Oliveira (2009):**
  Matrix Bernstein for matrix martingales
- **J. Neeman, B. Shi, and R. Ward (2023):**
  Markovian matrix Bernstein.
- **M. Banna, F. Merlevède, and P. Youssef (2016):**
  $\beta$-mixing analogue of matrix Bernstein.

## When can $\sqrt{\ln(2d)}$ be removed?

- **J.A. Tropp (2018):**
  Introduced a difficult-to-compute but insightful parameter.
- **A.S. Bandeira, M.T. Boedihardjo, and R. van Handel (2021):**
  Breakthrough for Gaussian matrices: $v^2 := \|\mathrm{Cov}(\boldsymbol{S})\|$ and relation to free probability.
- **T. Brailovskaya and R. van Handel (2022):**
  Independent non-Gaussian summands.

# Some related literature

## Matrix concentration with dependent summands

- **R.I. Oliveira (2009):**
  Matrix Bernstein for matrix martingales
- **J. Neeman, B. Shi, and R. Ward (2023):**
  Markovian matrix Bernstein.
- **M. Banna, F. Merlevède, and P. Youssef (2016):**
  $\beta$-mixing analogue of matrix Bernstein.

## When can $\sqrt{\ln(2d)}$ be removed?

- **J.A. Tropp (2018):**
  Introduced a difficult-to-compute but insightful parameter.
- **A.S. Bandeira, M.T. Boedihardjo, and R. van Handel (2021):**
  Breakthrough for Gaussian matrices: $v^2 \coloneqq \|\mathrm{Cov}(\boldsymbol{S})\|$ and relation to free probability.
- **T. Brailovskaya and R. van Handel (2022):**
  Independent non-Gaussian summands.

# Some related literature

## Matrix concentration with dependent summands

- **R.I. Oliveira (2009):**
  Matrix Bernstein for matrix martingales
- **J. Neeman, B. Shi, and R. Ward (2023):**
  Markovian matrix Bernstein.
- **M. Banna, F. Merlevède, and P. Youssef (2016):**
  $\beta$-mixing analogue of matrix Bernstein.

## When can $\sqrt{\ln(2d)}$ be removed?

- **J.A. Tropp (2018):**
  Introduced a difficult-to-compute but insightful parameter.
- **A.S. Bandeira, M.T. Boedihardjo, and R. van Handel (2021):**
  Breakthrough for Gaussian matrices: $v^2 := \|\mathrm{Cov}(\boldsymbol{S})\|$ and relation to free probability.
- **T. Brailovskaya and R. van Handel (2022):**
  Independent non-Gaussian summands.

# Some related literature

## Matrix concentration with dependent summands

- **R.I. Oliveira (2009):**
  Matrix Bernstein for matrix martingales
- **J. Neeman, B. Shi, and R. Ward (2023):**
  Markovian matrix Bernstein.
- **M. Banna, F. Merlevède, and P. Youssef (2016):**
  $\beta$-mixing analogue of matrix Bernstein.

## When can $\sqrt{\ln(2d)}$ be removed?

- **J.A. Tropp (2018):**
  Introduced a difficult-to-compute but insightful parameter.
- **A.S. Bandeira, M.T. Boedihardjo, and R. van Handel (2021):**
  Breakthrough for Gaussian matrices: $v^2 := \|\mathrm{Cov}(\boldsymbol{S})\|$ and relation to free probability.
- **T. Brailovskaya and R. van Handel (2022):**
  Independent non-Gaussian summands.

# Some related literature

## Matrix concentration with dependent summands

- **R.I. Oliveira (2009):**
  Matrix Bernstein for matrix martingales
- **J. Neeman, B. Shi, and R. Ward (2023):**
  Markovian matrix Bernstein.
- **M. Banna, F. Merlevède, and P. Youssef (2016):**
  $\beta$-mixing analogue of matrix Bernstein.

## When can $\sqrt{\ln(2d)}$ be removed?

- **J.A. Tropp (2018):**
  Introduced a difficult-to-compute but insightful parameter.
- **A.S. Bandeira, M.T. Boedihardjo, and R. van Handel (2021):**
  Breakthrough for Gaussian matrices: $v^2 := \|\mathrm{Cov}(\boldsymbol{S})\|$ and relation to free probability.
- **T. Brailovskaya and R. van Handel (2022):**
  Independent non-Gaussian summands.

# Some related literature

## Matrix concentration with dependent summands

- **R.I. Oliveira (2009):**
  Matrix Bernstein for matrix martingales
- **J. Neeman, B. Shi, and R. Ward (2023):**
  Markovian matrix Bernstein.
- **M. Banna, F. Merlevède, and P. Youssef (2016):**
  $\beta$-mixing analogue of matrix Bernstein.

## When can $\sqrt{\ln(2d)}$ be removed?

- **J.A. Tropp (2018):**
  Introduced a difficult-to-compute but insightful parameter.
- **A.S. Bandeira, M.T. Boedihardjo, and R. van Handel (2021):**
  Breakthrough for Gaussian matrices: $v^2 \coloneqq \|\mathrm{Cov}(\boldsymbol{S})\|$ and relation to free probability.
- **T. Brailovskaya and R. van Handel (2022):**
  Independent non-Gaussian summands.

# Some related literature

## Matrix concentration with dependent summands

- **R.I. Oliveira (2009):**
  Matrix Bernstein for matrix martingales
- **J. Neeman, B. Shi, and R. Ward (2023):**
  Markovian matrix Bernstein.
- **M. Banna, F. Merlevède, and P. Youssef (2016):**
  $\beta$-mixing analogue of matrix Bernstein.

## When can $\sqrt{\ln(2d)}$ be removed?

- **J.A. Tropp (2018):**
  Introduced a difficult-to-compute but insightful parameter.
- **A.S. Bandeira, M.T. Boedihardjo, and R. van Handel (2021):**
  Breakthrough for Gaussian matrices: $v^2 \coloneqq \|\mathrm{Cov}(\boldsymbol{S})\|$ and relation to free probability.
- **T. Brailovskaya and R. van Handel (2022):**
  Independent non-Gaussian summands.

# Some related literature

## Matrix concentration with dependent summands

- **R.I. Oliveira (2009):**
  Matrix Bernstein for matrix martingales
- **J. Neeman, B. Shi, and R. Ward (2023):**
  Markovian matrix Bernstein.
- **M. Banna, F. Merlevède, and P. Youssef (2016):**
  $\beta$-mixing analogue of matrix Bernstein.

## When can $\sqrt{\ln(2d)}$ be removed?

- **J.A. Tropp (2018):**
  Introduced a difficult-to-compute but insightful parameter.
- **A.S. Bandeira, M.T. Boedihardjo, and R. van Handel (2021):**
  Breakthrough for Gaussian matrices: $v^2 := \|\mathrm{Cov}(\boldsymbol{S})\|$ and relation to free probability.
- **T. Brailovskaya and R. van Handel (2022):**
  Independent non-Gaussian summands.

# Part II

## Sharp matrix concentration

### for Markovian model & matrix series model

# Markovian model and parameters

**Markovian model.**

Consider a Markov chain $Z_1, \ldots, Z_n$.

Generate self-adjoint $d \times d$ matrices

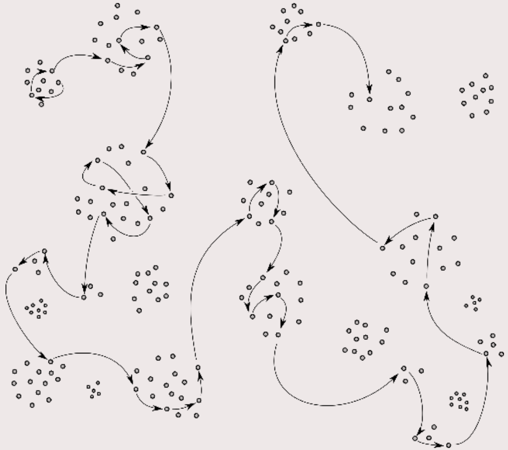
$$\boldsymbol{X}_i := f_i(Z_i) \quad \text{and} \quad \boldsymbol{S} := \sum_{i=1}^{n} \boldsymbol{X}_i$$

Further, for simplicity, also assume

- Finite state space
- Transition matrix $P$
  & equilibrium distribution $\pi$
- $\mathbb{E}\boldsymbol{X}_i = 0$

**Dependence parameter**

$$\Psi := \min\{\, t \geq 1 : \frac{|P_{x,y}^{t} - \pi_y|}{\pi_y} \leq \frac{1}{4} \ \forall x, y\}$$



# Markovian model and parameters

**Markovian model.**

Consider a Markov chain $Z_1, \ldots, Z_n$.

Generate self-adjoint $d \times d$ matrices

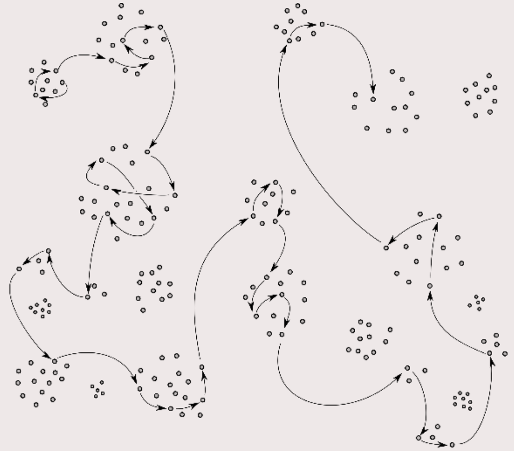
$\boldsymbol{X}_i := f_i(Z_i)$ and $\boldsymbol{S} := \sum_{i=1}^{n} \boldsymbol{X}_i$

Further, for simplicity, also assume

- Finite state space
- Transition matrix $P$
  & equilibrium distribution $\pi$
- $\mathbb{E}\boldsymbol{X}_i = 0$

**Dependence parameter**

$$\Psi := \min\{\, t \geq 1 \colon \frac{|P^t_{x,y} - \pi_y|}{\pi_y} \leq \frac{1}{4} \ \forall x, y\}$$



# Markovian model and parameters

**Markovian model.**

Consider a Markov chain $Z_1, \dots, Z_n$.

Generate self-adjoint $d \times d$ matrices

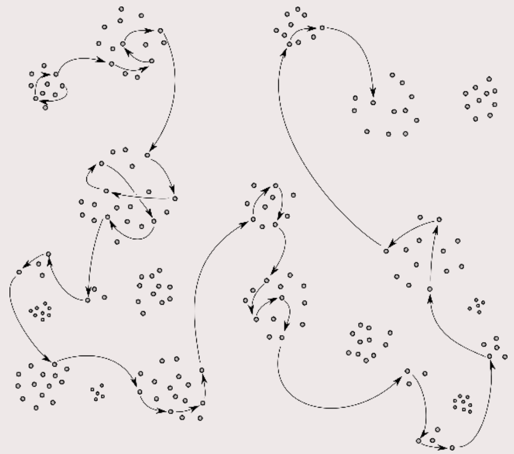
$\boldsymbol{X}_i := f_i(Z_i)$ and $\boldsymbol{S} := \sum_{i=1}^{n} \boldsymbol{X}_i$

Further, for simplicity, also assume

- Finite state space
- Transition matrix $P$
  & equilibrium distribution $\pi$
- $\mathbb{E}\boldsymbol{X}_i = 0$

**Dependence parameter**

$$\Psi := \min\{\, t \geq 1 : \frac{|P_{x,y}^t - \pi_y|}{\pi_y} \leq \frac{1}{4} \;\forall x, y\}$$



# Markovian model and parameters

**Markovian model.**

Consider a Markov chain $Z_1, \dots, Z_n$.

Generate self-adjoint $d \times d$ matrices

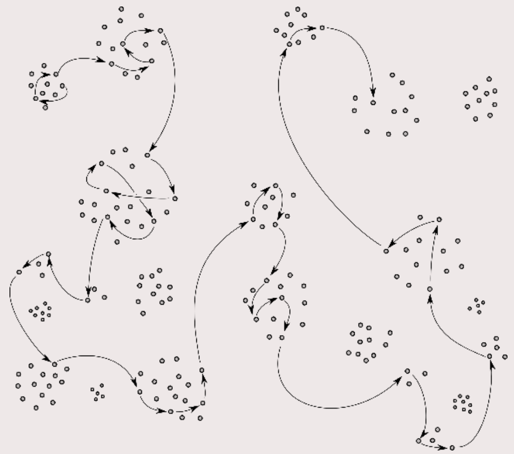
$$\boldsymbol{X}_i := f_i(Z_i) \quad \text{and} \quad \boldsymbol{S} := \textstyle\sum_{i=1}^{n} \boldsymbol{X}_i$$

Further, for simplicity, also assume

- Finite state space
- Transition matrix $P$
  & equilibrium distribution $\pi$
- $\mathbb{E}\boldsymbol{X}_i = 0$

**Dependence parameter**

$$\Psi := \min\{\, t \geq 1 : \frac{|P^t_{x,y} - \pi_y|}{\pi_y} \leq \frac{1}{4} \;\forall x, y\}$$



# Markovian model and parameters

**Markovian model.**

Consider a Markov chain $Z_1, \dots, Z_n$.

Generate self-adjoint $d \times d$ matrices

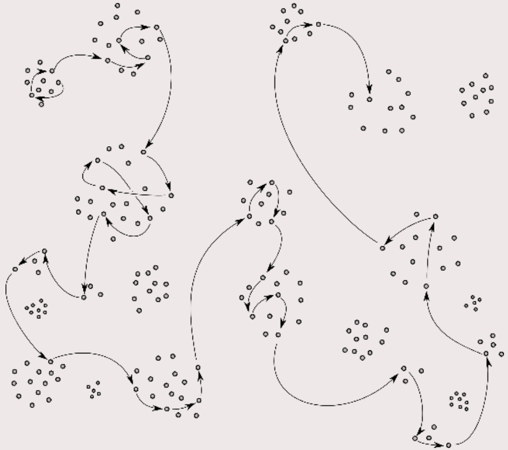
$\boldsymbol{X}_i \coloneqq f_i(Z_i)$ and $\boldsymbol{S} \coloneqq \sum_{i=1}^{n} \boldsymbol{X}_i$

Further, for simplicity, also assume

- Finite state space
- Transition matrix $P$
  & equilibrium distribution $\pi$
- $\mathbb{E}\boldsymbol{X}_i = 0$

**Dependence parameter**

$$\Psi \coloneqq \min\{\, t \geq 1 : \frac{|P_{x,y}^{t} - \pi_y|}{\pi_y} \leq \frac{1}{4}\ \forall x, y\}$$



# Markovian model and parameters

**Markovian model.**

Consider a Markov chain $Z_1, \dots, Z_n$.

Generate self-adjoint $d \times d$ matrices

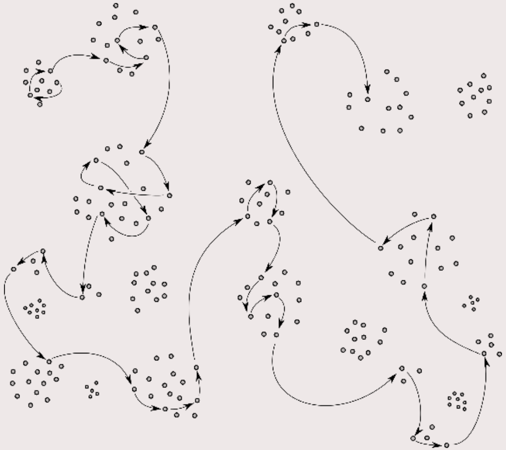
$$\boldsymbol{X}_i := f_i(Z_i) \quad \text{and} \quad \boldsymbol{S} := \textstyle\sum_{i=1}^{n} \boldsymbol{X}_i$$

Further, for simplicity, also assume

- Finite state space
- Transition matrix $P$
  & equilibrium distribution $\pi$
- $\mathbb{E}\boldsymbol{X}_i = 0$

**Dependence parameter**

$$\Psi := \min\{\, t \geq 1 : \frac{|P^t_{x,y} - \pi_y|}{\pi_y} \leq \frac{1}{4} \ \forall x, y\}$$



# Markovian model and parameters

**Markovian model.**

Consider a Markov chain $Z_1, \dots, Z_n$.

Generate self-adjoint $d \times d$ matrices

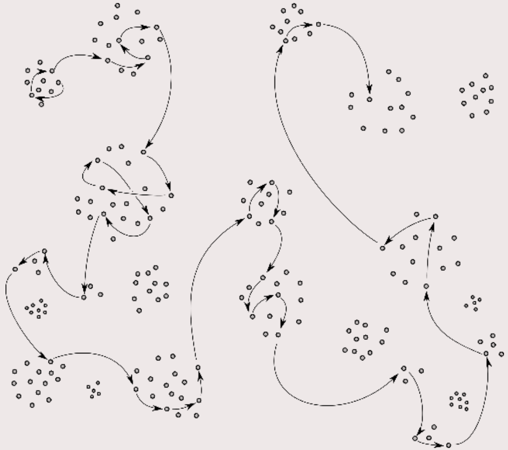
$$\boldsymbol{X}_i := f_i(Z_i) \quad \text{and} \quad \boldsymbol{S} := \textstyle\sum_{i=1}^{n} \boldsymbol{X}_i$$

Further, for simplicity, also assume

- Finite state space
- Transition matrix $P$
  & equilibrium distribution $\pi$
- $\mathbb{E}\boldsymbol{X}_i = 0$

**Dependence parameter**

$$\Psi := \min\{\, t \geq 1 : \frac{|P^t_{x,y} - \pi_y|}{\pi_y} \leq \frac{1}{4} \ \forall x, y\}$$



# Markovian model and parameters

**Markovian model.**

Consider a Markov chain $Z_1, \dots, Z_n$.

Generate self-adjoint $d \times d$ matrices

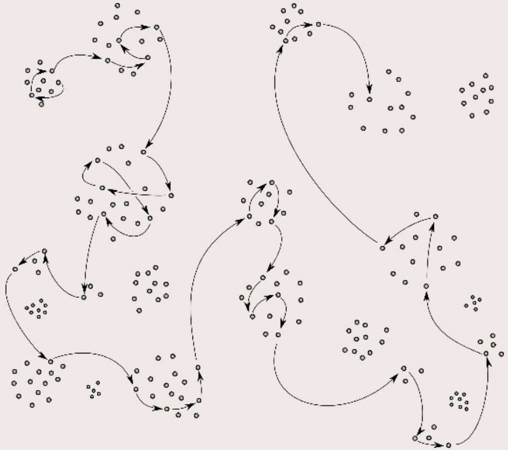
$$\boldsymbol{X}_i := f_i(Z_i) \quad \text{and} \quad \boldsymbol{S} := \sum_{i=1}^{n} \boldsymbol{X}_i$$

Further, for simplicity, also assume

- Finite state space
- Transition matrix $P$
  & equilibrium distribution $\pi$
- $\mathbb{E}\boldsymbol{X}_i = 0$

**Dependence parameter**

$$\Psi := \min\{\, t \geq 1 : \frac{|P_{x,y}^t - \pi_y|}{\pi_y} \leq \frac{1}{4} \ \forall x, y\}$$



# Sharp matrix concentration in Markovian model

**Main result.**

There exists $c > 0$ such that for $0 < \delta \leq 1$

$$\mathbb{E}\|\boldsymbol{S}\| \leq (1+\delta)\|\boldsymbol{S}_{\mathrm{free}}\| + c\mathcal{E}_{d,\delta}$$

where

$$\begin{aligned}\mathcal{E}_{d,\delta} := {} & v^{1/2}\sigma^{1/2}\log_{1+\delta}(2d)^{3/4} \\ & + R^{1/3}\Psi^{2/3}\varsigma^{2/3}\log_{1+\delta}(2d)^{2/3} \\ & + R\,\Psi\log_{1+\delta}(2d)\end{aligned}$$

**Dependence parameter**

$$\Psi := \min\{\, t \geq 1 \colon \frac{|P^t_{x,y} - \pi_y|}{\pi_y} \leq \frac{1}{4} \ \forall x, y\}$$

**Free-probabilistic quantity**

$$\|\boldsymbol{S}_{\mathrm{free}}\| = \inf_{\boldsymbol{W} \succ \boldsymbol{0}} \lambda_{\max}\left(\boldsymbol{W}^{-1} + \mathbb{E}[\boldsymbol{S}\boldsymbol{W}\boldsymbol{S}]\right).$$

**Matrix parameters**

$$\varsigma^2 := \left\|\mathbb{E}\left[\textstyle\sum_{i=1}^n \boldsymbol{X}_i^2\right]\right\|, \quad \sigma^2 := \|\mathbb{E}[\boldsymbol{S}^2]\|$$

$$v^2 := \|\mathrm{Cov}(\boldsymbol{S})\|, \quad R \geq \|\boldsymbol{X}_i\|.$$

# Sharp matrix concentration in Markovian model

**Main result.**

There exists $c > 0$ such that for $0 < \delta \leq 1$

$$\mathbb{E}\|\boldsymbol{S}\| \leq (1+\delta)\|\boldsymbol{S}_{\text{free}}\| + c\mathcal{E}_{d,\delta}$$

where

$$\mathcal{E}_{d,\delta} := v^{1/2}\sigma^{1/2}\log_{1+\delta}(2d)^{3/4}$$
$$+R^{1/3}\Psi^{2/3}\varsigma^{2/3}\log_{1+\delta}(2d)^{2/3}$$
$$+\ R\,\Psi\log_{1+\delta}(2d)$$

**Dependence parameter**

$$\Psi := \min\{\, t \geq 1 : \frac{|P^t_{x,y} - \pi_y|}{\pi_y} \leq \frac{1}{4}\ \forall x, y\}$$

**Free-probabilistic quantity**

$$\|\boldsymbol{S}_{\text{free}}\| = \inf_{\boldsymbol{W} \succ \boldsymbol{0}} \lambda_{\max}\left(\boldsymbol{W}^{-1} + \mathbb{E}[\boldsymbol{S}\boldsymbol{W}\boldsymbol{S}]\right).$$

**Matrix parameters**

$$\varsigma^2 := \left\|\mathbb{E}\left[\textstyle\sum_{i=1}^n \boldsymbol{X}_i^2\right]\right\|, \quad \sigma^2 := \|\mathbb{E}[\boldsymbol{S}^2]\|$$

$$v^2 := \|\mathrm{Cov}(\boldsymbol{S})\|, \quad R \geq \|\boldsymbol{X}_i\|.$$

# Sharp matrix concentration in Markovian model

**Main result.**

There exists $c > 0$ such that for $0 < \delta \le 1$

$$\mathbb{E}\|\boldsymbol{S}\| \le (1+\delta)\|\boldsymbol{S}_{\text{free}}\| + c\,\mathcal{E}_{d,\delta}$$

where

$$\mathcal{E}_{d,\delta} := v^{1/2}\sigma^{1/2}\log_{1+\delta}(2d)^{3/4}$$
$$+R^{1/3}\Psi^{2/3}\varsigma^{2/3}\log_{1+\delta}(2d)^{2/3}$$
$$+\ R\,\Psi\,\log_{1+\delta}(2d)$$

**Dependence parameter**

$$\Psi := \min\{\, t \ge 1 : \frac{|P^t_{x,y} - \pi_y|}{\pi_y} \le \frac{1}{4}\ \forall x, y\}$$

**Free-probabilistic quantity** (Lehner)

$$\|\boldsymbol{S}_{\text{free}}\| = \inf_{\boldsymbol{W} \succ \boldsymbol{0}} \lambda_{\max}\left(\boldsymbol{W}^{-1} + \mathbb{E}[\boldsymbol{S}\boldsymbol{W}\boldsymbol{S}]\right).$$

**Matrix parameters**

$$\varsigma^2 := \left\|\mathbb{E}\left[\textstyle\sum_{i=1}^n \boldsymbol{X}_i^2\right]\right\|, \quad \sigma^2 := \|\mathbb{E}[\boldsymbol{S}^2]\|$$

$$v^2 := \|\mathrm{Cov}(\boldsymbol{S})\|, \quad R \ge \|\boldsymbol{X}_i\|.$$

# Sharp matrix concentration in Markovian model

**Main result.**

There exists $c > 0$ such that for $0 < \delta \leq 1$

$$\mathbb{E}\|\boldsymbol{S}\| \leq (1+\delta)\|\boldsymbol{S}_{\mathrm{free}}\| + c\mathcal{E}_{d,\delta}$$

where

$$\mathcal{E}_{d,\delta} := v^{1/2}\sigma^{1/2}\log_{1+\delta}(2d)^{3/4}$$
$$+R^{1/3}\Psi^{2/3}\varsigma^{2/3}\log_{1+\delta}(2d)^{2/3}$$
$$+\ R\,\Psi\log_{1+\delta}(2d)$$

**Dependence parameter**

$$\Psi := \min\{\, t \geq 1 \colon \frac{|P_{x,y}^{t} - \pi_y|}{\pi_y} \leq \frac{1}{4}\ \forall x, y\}$$

**Free-probabilistic quantity**

$$\|\boldsymbol{S}_{\mathrm{free}}\| = \inf_{\boldsymbol{W}\succ\boldsymbol{0}} \lambda_{\max}\left(\boldsymbol{W}^{-1} + \mathbb{E}[\boldsymbol{S}\boldsymbol{W}\boldsymbol{S}]\right).$$

**Matrix parameters**

$$\varsigma^2 := \left\|\mathbb{E}\left[\textstyle\sum_{i=1}^{n}\boldsymbol{X}_i^2\right]\right\|, \quad \sigma^2 := \|\mathbb{E}[\boldsymbol{S}^2]\|$$

$$v^2 := \|\mathrm{Cov}(\boldsymbol{S})\|, \quad R \geq \|\boldsymbol{X}_i\|.$$

# Sharp matrix concentration in Markovian model

**Main result.**

There exists $c > 0$ such that for $0 < \delta \leq 1$

$$\mathbb{E}\|\boldsymbol{S}\| \leq (1+\delta)\|\boldsymbol{S}_{\mathrm{free}}\| + c\mathcal{E}_{d,\delta}$$

where

$$\begin{aligned}\mathcal{E}_{d,\delta} := {} & v^{1/2}\sigma^{1/2}\log_{1+\delta}(2d)^{3/4} \\ & + R^{1/3}\Psi^{2/3}\varsigma^{2/3}\log_{1+\delta}(2d)^{2/3} \\ & + R\,\Psi\log_{1+\delta}(2d)\end{aligned}$$

**Dependence parameter**

$$\Psi := \min\left\{ t \geq 1 : \frac{|P^t_{x,y} - \pi_y|}{\pi_y} \leq \frac{1}{4}\ \forall x, y\right\}$$

**Free-probabilistic quantity**

$$\|\boldsymbol{S}_{\mathrm{free}}\| = \inf_{\boldsymbol{W} \succ \boldsymbol{0}} \lambda_{\max}\left(\boldsymbol{W}^{-1} + \mathbb{E}[\boldsymbol{SWS}]\right).$$

**Matrix parameters**

$$\varsigma^2 := \left\|\mathbb{E}\left[\textstyle\sum_{i=1}^n \boldsymbol{X}_i^2\right]\right\|, \quad \sigma^2 := \|\mathbb{E}[\boldsymbol{S}^2]\|$$

$$v^2 := \|\mathrm{Cov}(\boldsymbol{S})\|, \quad R \geq \|\boldsymbol{X}_i\|.$$

# Informal summary

**Main result. (Informally)**

We provide explicit $\varepsilon > 0$ such that

$$\mathbb{E}\|\boldsymbol{S}\| \leq \|\boldsymbol{S}_{\mathrm{free}}\| + \varepsilon \ln(2d)$$

# Informal summary

**Main result. (Informally)**

We provide explicit $\varepsilon > 0$ such that

$$\mathbb{E}\|\boldsymbol{S}\| \leq \|\boldsymbol{S}_{\text{free}}\| + \varepsilon \ln(2d)$$

# Informal summary

**Main result. (Informally)**

We provide explicit $\varepsilon > 0$ such that

$$\mathbb{E}\|\boldsymbol{S}\| \leq \|\boldsymbol{S}_{\text{free}}\| + \varepsilon \ln(2d)$$

**Removal of $\sqrt{\ln(2d)}$**

Due to Pisier,

$$\sigma \leq \|\boldsymbol{S}_{\text{free}}\| \leq 2\sigma$$

**Asymptotically sharp**

In many relevant models

$$\lim_{d\to\infty} \|\boldsymbol{S}\| = \|\boldsymbol{S}_{\text{free}}\|$$

# Informal summary

**Main result. (Informally)**

We provide explicit $\varepsilon > 0$ such that

$$\mathbb{E}\|\boldsymbol{S}\| \leq \|\boldsymbol{S}_{\mathrm{free}}\| + \varepsilon \ln(2d)$$

**Removal of $\sqrt{\ln(2d)}$**

Due to Pisier,

$$\sigma \leq \|\boldsymbol{S}_{\mathrm{free}}\| \leq 2\sigma$$

**Asymptotically sharp**

In many relevant models

$$\lim_{d\to\infty} \|\boldsymbol{S}\| = \|\boldsymbol{S}_{\mathrm{free}}\|$$

# Informal summary

**Main result. (Informally)**

We provide explicit $\varepsilon > 0$ such that

$$\mathbb{E}\|\boldsymbol{S}\| \leq \|\boldsymbol{S}_{\text{free}}\| + \varepsilon \ln(2d)$$

**Removal of $\sqrt{\ln(2d)}$**

Due to Pisier,

$$\sigma \leq \|\boldsymbol{S}_{\text{free}}\| \leq 2\sigma$$

**Asymptotically sharp**

In many relevant models

$$\lim_{d\to\infty} \|\boldsymbol{S}\| = \|\boldsymbol{S}_{\text{free}}\|$$

**Matrix series model.**

We show a similar result for a *matrix series model*.

There, summands are of the form $\boldsymbol{X}_i := Y_i \, \boldsymbol{A}_i$ for deterministic $\boldsymbol{A}_i$.

# Informal summary

**Main result. (Informally)**

We provide explicit $\varepsilon > 0$ such that

$$\mathbb{E}\|\boldsymbol{S}\| \leq \|\boldsymbol{S}_{\text{free}}\| + \varepsilon \ln(2d)$$

**Removal of $\sqrt{\ln(2d)}$**

Due to Pisier,

$$\sigma \leq \|\boldsymbol{S}_{\text{free}}\| \leq 2\sigma$$

**Asymptotically sharp**

In many relevant models

$$\lim_{d\to\infty} \|\boldsymbol{S}\| = \|\boldsymbol{S}_{\text{free}}\|$$

**Matrix series model.**

We show a similar result for a *matrix series model*.

There, summands are of the form $\boldsymbol{X}_i := Y_i \boldsymbol{A}_i$ for deterministic $\boldsymbol{A}_i$.

# Part III

# Proof sketch

# Proof sketch.

**Goal.** $\mathbb{E}[\mathrm{tr}\, \mathbf{S}^{2p}] \approx \mathbb{E}[\mathrm{tr}\, \mathbf{G}^{2p}]$ where $\boldsymbol{G}$ is a *Gaussian model* for **S**

**Interpolate:** For $t \in [0,1]$ set

$$\boldsymbol{S}(t) := \sqrt{t}\boldsymbol{S} + \sqrt{1-t}\boldsymbol{G}$$

and show that

$$\frac{\mathrm{d}}{\mathrm{d}t}\mathbb{E}[\,\mathrm{tr}\,\boldsymbol{S}(t)^{2p}] \le \epsilon(t)$$

# Proof sketch.

**Goal.** $\mathbb{E}[\operatorname{tr} \mathbf{S}^{2p}] \approx \mathbb{E}[\operatorname{tr} \mathbf{G}^{2p}]$ where $\boldsymbol{G}$ is a *Gaussian model* for **S**

**Interpolate:** For $t \in [0,1]$ set

$$\boldsymbol{S}(t) := \sqrt{t}\boldsymbol{S} + \sqrt{1-t}\boldsymbol{G}$$

and show that

$$\frac{\mathrm{d}}{\mathrm{d}t}\mathbb{E}[\operatorname{tr} \boldsymbol{S}(t)^{2p}] \leq \epsilon(t)$$

# Proof sketch.

**Goal.** $\mathbb{E}[\operatorname{tr} \boldsymbol{S}^{2p}] \approx \mathbb{E}[\operatorname{tr} \mathbf{G}^{2p}]$ where $\boldsymbol{G}$ is a *Gaussian model* for **S**

**Interpolate:** For $t \in [0,1]$ set

$$\boldsymbol{S}(t) \coloneqq \sqrt{t}\boldsymbol{S} + \sqrt{1-t}\boldsymbol{G}$$

and show that

$$\frac{\mathrm{d}}{\mathrm{d}t}\mathbb{E}[\operatorname{tr} \boldsymbol{S}(t)^{2p}] \leq \epsilon(t)$$

# Proof sketch.

**Goal.** $\mathbb{E}[\operatorname{tr} \mathbf{S}^{2p}] \approx \mathbb{E}[\operatorname{tr} \mathbf{G}^{2p}]$ where $\boldsymbol{G}$ is a *Gaussian model* for **S**

**Interpolate:** For $t \in [0,1]$ set

$$\boldsymbol{S}(t) := \sqrt{t}\boldsymbol{S} + \sqrt{1-t}\boldsymbol{G}$$

and show that

$$\frac{\mathrm{d}}{\mathrm{d}t}\mathbb{E}[\operatorname{tr} \boldsymbol{S}(t)^{2p}] \leq \epsilon(t)$$

# Proof sketch.

**Goal.** $\mathbb{E}[\mathrm{tr}\,\mathbf{S}^{2p}] \approx \mathbb{E}[\mathrm{tr}\,\mathbf{G}^{2p}]$ where $\boldsymbol{G}$ is a *Gaussian model* for **S**

**Interpolate:** For $t \in [0,1]$ set

$$\boldsymbol{S}(t) := \sqrt{t}\boldsymbol{S} + \sqrt{1-t}\boldsymbol{G}$$

and show that

$$\frac{\mathrm{d}}{\mathrm{d}t}\mathbb{E}[\,\mathrm{tr}\,\boldsymbol{S}(t)^{2p}] \leq \epsilon(t)$$

**Brailovskaya & van Handel (independent setting)**

- Expansion with *classical cumulants.*
- Independence-implies-vanishing property reduces combinatorics

We lack independence!

# Proof sketch.

**Goal.** $\mathbb{E}[\operatorname{tr} \mathbf{S}^{2p}] \approx \mathbb{E}[\operatorname{tr} \mathbf{G}^{2p}]$ where $\boldsymbol{G}$ is a *Gaussian model* for **S**

**Interpolate:** For $t \in [0,1]$ set

$$\boldsymbol{S}(t) := \sqrt{t}\boldsymbol{S} + \sqrt{1-t}\boldsymbol{G}$$

and show that

$$\frac{\mathrm{d}}{\mathrm{d}t}\mathbb{E}[\,\operatorname{tr} \boldsymbol{S}(t)^{2p}] \leq \epsilon(t)$$

**Brailovskaya & van Handel (independent setting)**

- Expansion with *classical cumulants.*
- Independence-implies-vanishing property reduces combinatorics

We lack independence!

# Proof sketch.

**Goal.** $\mathbb{E}[\operatorname{tr} \mathbf{S}^{2p}] \approx \mathbb{E}[\operatorname{tr} \mathbf{G}^{2p}]$ where $\boldsymbol{G}$ is a *Gaussian model* for **S**

**Interpolate:** For $t \in [0,1]$ set

$$\boldsymbol{S}(t) := \sqrt{t}\boldsymbol{S} + \sqrt{1-t}\boldsymbol{G}$$

and show that

$$\frac{\mathrm{d}}{\mathrm{d}t}\mathbb{E}[\operatorname{tr} \boldsymbol{S}(t)^{2p}] \leq \epsilon(t)$$

**Brailovskaya & van Handel (independent setting)**

- Expansion with *classical cumulants.*
- Independence-implies-vanishing property reduces combinatorics

We lack independence!

# Proof sketch.

**Goal.** $\mathbb{E}[\operatorname{tr} \mathbf{S}^{2p}] \approx \mathbb{E}[\operatorname{tr} \mathbf{G}^{2p}]$ where $\boldsymbol{G}$ is a *Gaussian model* for **S**

**Interpolate:** For $t \in [0,1]$ set

$$\boldsymbol{S}(t) := \sqrt{t}\boldsymbol{S} + \sqrt{1-t}\boldsymbol{G}$$

and show that

$$\frac{\mathrm{d}}{\mathrm{d}t}\mathbb{E}[\operatorname{tr} \boldsymbol{S}(t)^{2p}] \leq \epsilon(t)$$

**Brailovskaya & van Handel (independent setting)**

- Expansion with *classical cumulants.*
- Independence-implies-vanishing property reduces combinatorics

We lack independence!

# Proof sketch.

**Goal.** $\mathbb{E}[\operatorname{tr} \mathbf{S}^{2p}] \approx \mathbb{E}[\operatorname{tr} \mathbf{G}^{2p}]$ where $\boldsymbol{G}$ is a *Gaussian model* for **S**

**Interpolate:** For $t \in [0,1]$ set

$$\boldsymbol{S}(t) := \sqrt{t}\boldsymbol{S} + \sqrt{1-t}\boldsymbol{G}$$

and show that

$$\frac{\mathrm{d}}{\mathrm{d}t}\mathbb{E}[\operatorname{tr} \boldsymbol{S}(t)^{2p}] \leq \epsilon(t)$$

**Brailovskaya & van Handel (independent setting)**

- Expansion with *classical cumulants.*
- Independence-implies-vanishing property reduces combinatorics

We lack independence!

**Direct modification using classical cumulants gives suboptimal results…**

# Proof sketch.

**Goal.** Show that $\frac{\mathrm{d}}{\mathrm{d}t}\mathbb{E}[\operatorname{tr} \boldsymbol{S}(t)^{2p}] \leq \epsilon(t)$

**New approach:**

- Boolean cumulants interact with Markovian structure, e.g.,
$$b\big(g_1(Z_1), g_2(Z_2)\big) = \int\int g_1(z_1) g_2(z_2)\, \mathrm{d}\,\mathbb{P}_{Z_1}\big(\mathrm{d}\mathbb{P}_{Z_2|Z_1=z_1} - \mathrm{d}\,\mathbb{P}_{Z_2}\big).$$
- Classical-to-Boolean relation due to Arizmendi et al. (2015)
- Change-of-measure: encode suppression in random variables $\Delta^{(\rho,\alpha,i)}$

# Proof sketch.

**Goal.** Show that $\frac{\mathrm{d}}{\mathrm{d}t}\mathbb{E}[\operatorname{tr}\boldsymbol{S}(t)^{2p}] \leq \epsilon(t)$

**New approach:**

- Boolean cumulants interact with Markovian structure, e.g.,

$$b\big(g_1(Z_1), g_2(Z_2)\big) = \int\int g_1(z_1)g_2(z_2)\,\mathrm{d}\,\mathbb{P}_{Z_1}\big(\mathrm{d}\mathbb{P}_{Z_2|Z_1=z_1} - \mathrm{d}\,\mathbb{P}_{Z_2}\big).$$

- Classical-to-Boolean relation due to Arizmendi et al. (2015)
- Change-of-measure: encode suppression in random variables $\Delta^{(\rho,\alpha,i)}$

# Proof sketch.

**Goal.** Show that $\frac{\mathrm{d}}{\mathrm{d}t}\mathbb{E}[\operatorname{tr} \boldsymbol{S}(t)^{2p}] \leq \epsilon(t)$

**New approach:**

- Boolean cumulants interact with Markovian structure, e.g.,
$$b\big(g_1(Z_1), g_2(Z_2)\big) = \int \int g_1(z_1) g_2(z_2)\, \mathrm{d}\, \mathbb{P}_{Z_1} \big(\mathrm{d}\mathbb{P}_{Z_2|Z_1=z_1} - \mathrm{d}\, \mathbb{P}_{Z_2}\big).$$
- Classical-to-Boolean relation due to Arizmendi et al. (2015)
- Change-of-measure: encode suppression in random variables $\Delta^{(\rho,\alpha,i)}$

# Proof sketch.

**Goal.** Show that $\frac{\mathrm{d}}{\mathrm{d}t}\mathbb{E}[\,\mathrm{tr}\,\boldsymbol{S}(t)^{2p}] \leq \epsilon(t)$

**New approach:**

- Boolean cumulants interact with Markovian structure, e.g.,
$$b\big(g_1(Z_1), g_2(Z_2)\big) = \int\int g_1(z_1) g_2(z_2)\,\mathrm{d}\,\mathbb{P}_{Z_1}\big(\mathrm{d}\mathbb{P}_{Z_2|Z_1=z_1} - \mathrm{d}\,\mathbb{P}_{Z_2}\big).$$
- Classical-to-Boolean relation due to Arizmendi et al. (2015)
- Change-of-measure: encode suppression in random variables $\Delta^{(\rho,\alpha,i)}$

# Proof sketch.

**Goal.** Show that $\frac{\mathrm{d}}{\mathrm{d}t}\mathbb{E}[\,\mathrm{tr}\,\boldsymbol{S}(t)^{2p}] \le \epsilon(t)$

**New approach:**

- Boolean cumulants interact with Markovian structure, e.g.,
$$b\big(g_1(Z_1), g_2(Z_2)\big) = \int\int g_1(z_1)g_2(z_2)\,\mathrm{d}\,\mathbb{P}_{Z_1}\big(\mathrm{d}\mathbb{P}_{Z_2|Z_1=z_1} - \mathrm{d}\,\mathbb{P}_{Z_2}\big).$$
- Classical-to-Boolean relation due to Arizmendi et al. (2015)
- Change-of-measure: encode suppression in random variables $\Delta^{(\rho,\alpha,i)}$

# Proof sketch.

**Goal.** Show that $\frac{\mathrm{d}}{\mathrm{d}t}\mathbb{E}[\,\mathrm{tr}\,\boldsymbol{S}(t)^{2p}] \leq \epsilon(t)$

**New approach:**

- Boolean cumulants interact with Markovian structure, e.g.,

$$b\big(g_1(Z_1), g_2(Z_2)\big) = \int\int g_1(z_1)g_2(z_2)\,\mathrm{d}\,\mathbb{P}_{Z_1}\big(\mathrm{d}\mathbb{P}_{Z_2|Z_1=z_1} - \mathrm{d}\,\mathbb{P}_{Z_2}\big).$$

- Classical-to-Boolean relation due to Arizmendi et al. (2015)
- Change-of-measure: encode suppression in random variables $\Delta^{(\rho,\alpha,i)}$

**Main technical results:**

1) $$\frac{\mathrm{d}}{\mathrm{d}t}\mathbb{E}[\,\mathrm{tr}\,\boldsymbol{S}(t)^{2p}] = \sum_{k=3}^{\infty}\frac{t^{k/2}}{2(k-1)!}\sum_{(\rho,\alpha,i)}\mathbb{E}\Big[\,\Delta^{(\rho,\alpha,i)}\;\partial_{\boldsymbol{X}_{i,1}^{(\rho,\alpha,i)}}\cdots\partial_{\boldsymbol{X}_{i,k}^{(\rho,\alpha,i)}}\;\mathrm{tr}\,\boldsymbol{S}(t)^{2p}\,\Big]$$

# Proof sketch.

**Goal.** Show that $\frac{\mathrm{d}}{\mathrm{d}t}\mathbb{E}[\,\mathrm{tr}\,\boldsymbol{S}(t)^{2p}] \leq \epsilon(t)$

**New approach:**

- Boolean cumulants interact with Markovian structure, e.g.,

$$b\big(g_1(Z_1), g_2(Z_2)\big) = \int\int g_1(z_1)g_2(z_2)\,\mathrm{d}\,\mathbb{P}_{Z_1}\big(\mathrm{d}\mathbb{P}_{Z_2|Z_1=z_1} - \mathrm{d}\,\mathbb{P}_{Z_2}\big).$$

- Classical-to-Boolean relation due to Arizmendi et al. (2015)
- Change-of-measure: encode suppression in random variables $\Delta^{(\rho,\alpha,i)}$

**Main technical results:**

1) $$\frac{\mathrm{d}}{\mathrm{d}t}\mathbb{E}[\,\mathrm{tr}\,\boldsymbol{S}(t)^{2p}] = \sum_{k=3}^{\infty}\frac{t^{k/2}}{2(k-1)!}\sum_{(\rho,\alpha,i)}\mathbb{E}\left[\,\Delta^{(\rho,\alpha,i)}\;\partial_{\boldsymbol{X}_{i,1}^{(\rho,\alpha,i)}}\cdots\partial_{\boldsymbol{X}_{i,k}^{(\rho,\alpha,i)}}\;\mathrm{tr}\,\boldsymbol{S}(t)^{2p}\,\right]$$

# Proof sketch.

**Goal.** Show that $\frac{\mathrm{d}}{\mathrm{d}t}\mathbb{E}[\operatorname{tr} \boldsymbol{S}(t)^{2p}] \leq \epsilon(t)$

**New approach:**

- Boolean cumulants interact with Markovian structure, e.g.,

$$b\big(g_1(Z_1), g_2(Z_2)\big) = \int\int g_1(z_1) g_2(z_2)\, \mathrm{d}\,\mathbb{P}_{Z_1}\big(\mathrm{d}\mathbb{P}_{Z_2|Z_1=z_1} - \mathrm{d}\,\mathbb{P}_{Z_2}\big).$$

- Classical-to-Boolean relation due to Arizmendi et al. (2015)
- Change-of-measure: encode suppression in random variables $\Delta^{(\rho,\alpha,i)}$

**Main technical results:**

1) $$\frac{\mathrm{d}}{\mathrm{d}t}\mathbb{E}[\operatorname{tr} \boldsymbol{S}(t)^{2p}] = \sum_{k=3}^{\infty} \frac{t^{k/2}}{2(k-1)!} \sum_{(\rho,\alpha,i)} \mathbb{E}\left[ \Delta^{(\rho,\alpha,i)}\ \partial_{\boldsymbol{X}_{i,1}^{(\rho,\alpha,i)}} \cdots \partial_{\boldsymbol{X}_{i,k}^{(\rho,\alpha,i)}} \operatorname{tr} \boldsymbol{S}(t)^{2p} \right]$$

# Proof sketch.

**Goal.** Show that $\frac{\mathrm{d}}{\mathrm{d}t}\mathbb{E}[\operatorname{tr}\boldsymbol{S}(t)^{2p}] \leq \epsilon(t)$

**New approach:**

- Boolean cumulants interact with Markovian structure, e.g.,
  $$b\big(g_1(Z_1), g_2(Z_2)\big) = \int\int g_1(z_1)g_2(z_2)\,\mathrm{d}\,\mathbb{P}_{Z_1}\big(\mathrm{d}\mathbb{P}_{Z_2|Z_1=z_1} - \mathrm{d}\,\mathbb{P}_{Z_2}\big).$$
- Classical-to-Boolean relation due to Arizmendi et al. (2015)
- Change-of-measure: encode suppression in random variables $\Delta^{(\rho,\alpha,i)}$

**Main technical results:**

1) $\frac{\mathrm{d}}{\mathrm{d}t}\mathbb{E}[\operatorname{tr}\boldsymbol{S}(t)^{2p}] = \sum_{k=3}^{\infty}\frac{t^{k/2}}{2(k-1)!}\sum_{(\rho,\alpha,i)}\mathbb{E}\Big[\Delta^{(\rho,\alpha,i)}\ \partial_{\boldsymbol{X}_{i,1}^{(\rho,\alpha,i)}}\cdots\partial_{\boldsymbol{X}_{i,k}^{(\rho,\alpha,i)}}\operatorname{tr}\boldsymbol{S}(t)^{2p}\Big]$

2) $\sum_{(\rho,\alpha,i)}\big\|\Delta^{(\rho,\alpha,i)}\big\|_{L^\infty} \leq k!\,c^k\Psi^k$

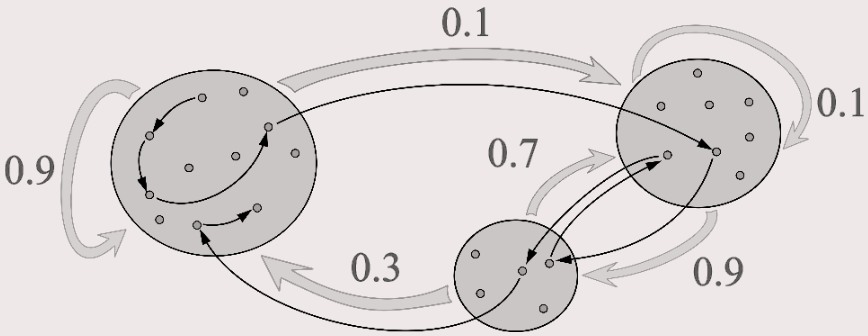
# Part IV

# Illustrative applications

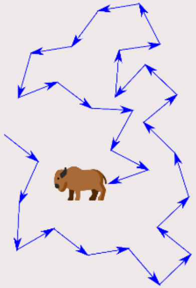
# Literature: Block Markov chains

Spectral clustering in block Markov chains

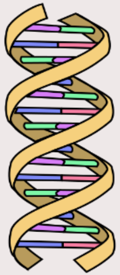
- **J. Sanders, A. Proutière, S.-Y. Yun (2020):**
  Spectral clustering algorithm
- **J. Sanders, A. Senen–Cerda (2023)**
  Asymptotic estimate $c\sigma$ for non-explicit $c$.
- **A. Van Werde, A. Senen–Cerda, G. Kosmella, J. Sanders (2023)**
  Real-world data



Animal movement



DNA



Text

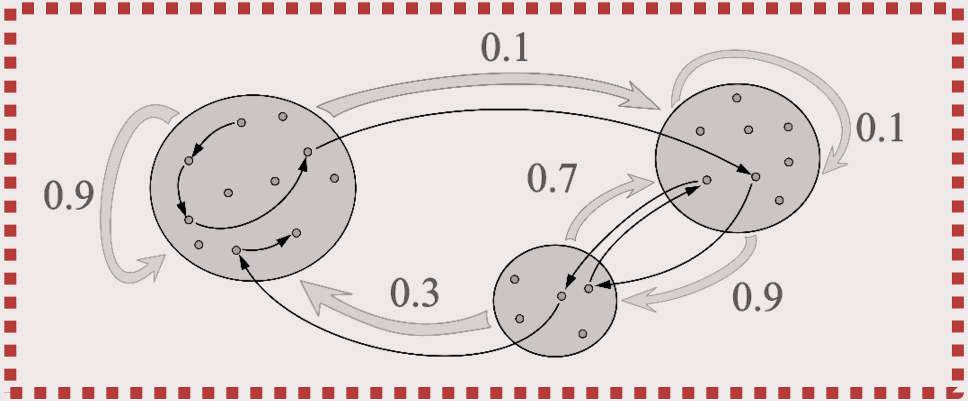
Lorem ipsum dolor sit amet, consectetur adipiscing elit. Maecenas in turpis sit amet risus ultricies tincidunt vitae accumsan enim. Pellentesque ac efficitur enim. Praesent dui magna, mattis at bibendum eu, fringilla eu neque. Nulla vel accumsan arcu. Quisque sagittis, massa quis maximus vestibulum, sapien ligula venenatis ex, eu elementum diam ante a neque. Nunc egestas malesuada interdum. Morbi eget sem in neque pretium vehicula. Duis nulla turpis, efficitur sed lacus at, auctor tempor lorem. Maecenas pharetra et erat non viverra. Morbi pellentesque velit nec risus imperdiet, ac fermentum lectus aliquam. Maecenas rutrum nisi ut turpis ultricies, id placerat quam fringilla. Donec non interdum urna. Nam justo mi, malesuada eu sollicitudin vitae, euismod vitae tortor.

Stock market

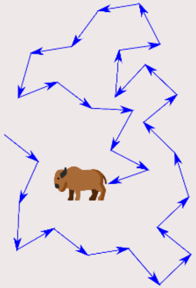
# Literature: Block Markov chains

Spectral clustering in block Markov chains

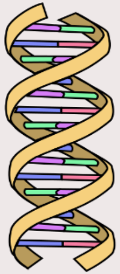
- **J. Sanders, A. Proutière, S.-Y. Yun (2020):**
  Spectral clustering algorithm
- **J. Sanders, A. Senen–Cerda (2023)**
  Asymptotic estimate $c\sigma$ for non-explicit $c$.
- **A. Van Werde, A. Senen–Cerda, G. Kosmella, J. Sanders (2023)**
  Real-world data



Animal movement



DNA



Text

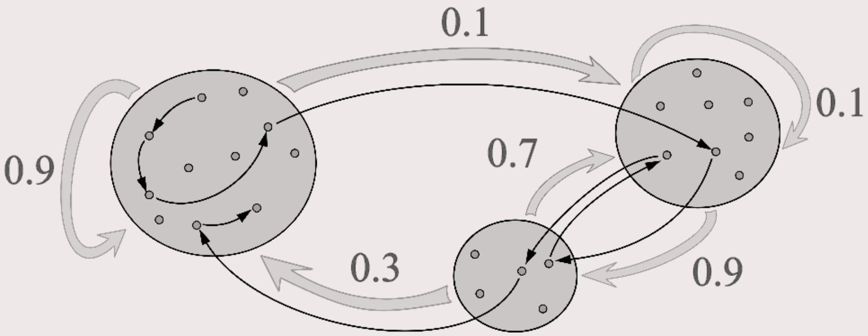
Lorem ipsum dolor sit amet, consectetur adipiscing elit. Maecenas in turpis sit amet risus ultricies tincidunt vitae accumsan enim. Pellentesque ac efficitur enim. Praesent dui magna, mattis at bibendum eu, fringilla eu neque. Nulla vel accumsan arcu. Quisque sagittis, massa quis maximus vestibulum, sapien ligula venenatis ex, eu elementum diam ante a neque. Nunc egestas malesuada interdum. Morbi eget sem in neque pretium vehicula. Duis nulla turpis, efficitur sed lacus at, auctor tempor lorem. Maecenas pharetra et erat non viverra. Morbi pellentesque velit nec risus imperdiet, ac fermentum lectus aliquam. Maecenas rutrum nisi ut turpis ultricies, id placerat quam fringilla. Donec non interdum urna. Nam justo mi, malesuada eu sollicitudin vitae, euismod vitae tortor.

Stock market

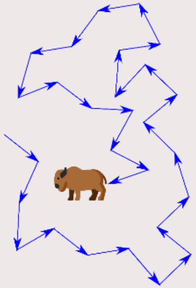
# Literature: Block Markov chains

Spectral clustering in block Markov chains

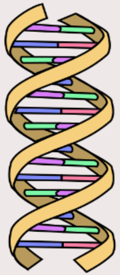
- **J. Sanders, A. Proutière, S.-Y. Yun (2020):**
  Spectral clustering algorithm
- **J. Sanders, A. Senen–Cerda (2023)**
  Asymptotic estimate $c\sigma$ for non-explicit $c$.
- **A. Van Werde, A. Senen–Cerda, G. Kosmella, J. Sanders (2023)**
  Real-world data



Animal movement



DNA



Text

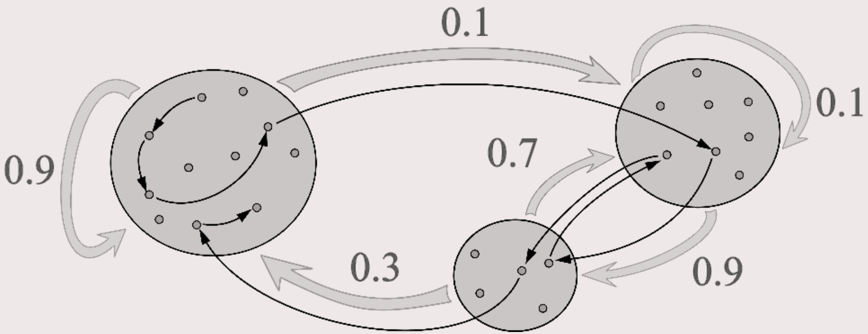
Lorem ipsum dolor sit amet, consectetur adipiscing elit. Maecenas in turpis sit amet risus ultricies tincidunt vitae accumsan enim. Pellentesque ac efficitur enim. Praesent dui magna, mattis at bibendum eu, fringilla eu neque. Nulla vel accumsan arcu. Quisque sagittis, massa quis maximus vestibulum, sapien ligula venenatis ex, eu elementum diam ante a neque. Nunc egestas malesuada interdum. Morbi eget sem in neque pretium vehicula. Duis nulla turpis, efficitur sed lacus at, auctor tempor lorem. Maecenas pharetra et erat non viverra. Morbi pellentesque velit nec risus imperdiet, ac fermentum lectus aliquam. Maecenas rutrum nisi ut turpis ultricies, id placerat quam fringilla. Donec non interdum urna. Nam justo mi, malesuada eu sollicitudin vitae, euismod vitae tortor.

Stock market

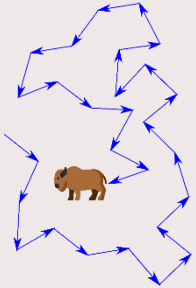
# Literature: Block Markov chains

Spectral clustering in block Markov chains

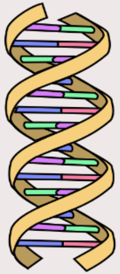
- **J. Sanders, A. Proutière, S.-Y. Yun (2020):**
  Spectral clustering algorithm
- **J. Sanders, A. Senen–Cerda (2023)**
  Asymptotic estimate $c\sigma$ for non-explicit $c$.
- **A. Van Werde, A. Senen–Cerda, G. Kosmella, J. Sanders (2023)**
  Real-world data



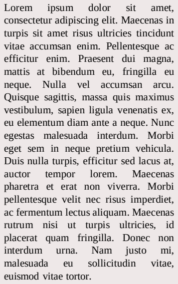
Animal movement



DNA



Text



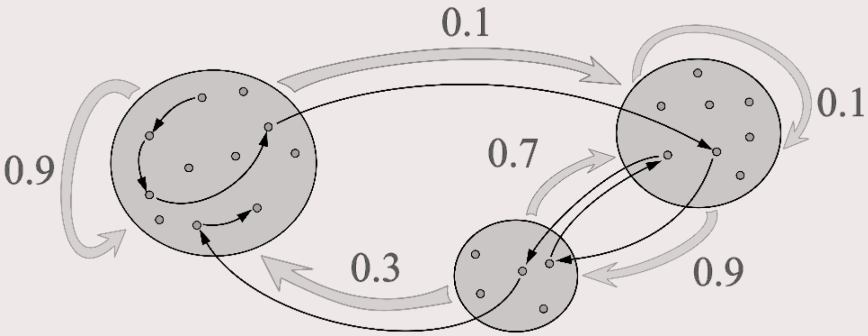
Lorem ipsum dolor sit amet, consectetur adipiscing elit. Maecenas in turpis sit amet risus ultricies tincidunt vitae accumsan enim. Pellentesque ac efficitur enim. Praesent dui magna, mattis at bibendum eu, fringilla eu neque. Nulla vel accumsan arcu. Quisque sagittis, massa quis maximus vestibulum, sapien ligula venenatis ex, eu elementum diam ante a neque. Nunc egestas malesuada interdum. Morbi eget sem in neque pretium vehicula. Duis nulla turpis, efficitur sed lacus at, auctor tempor lorem. Maecenas pharetra et erat non viverra. Morbi pellentesque velit nec risus imperdiet, ac fermentum lectus aliquam. Maecenas rutrum nisi ut turpis ultricies, id placerat quam fringilla. Donec non interdum urna. Nam justo mi, malesuada eu sollicitudin vitae, euismod vitae tortor.

Stock market

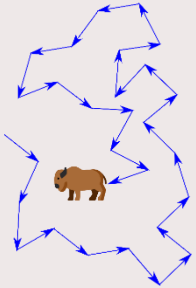
# Literature: Block Markov chains

Spectral clustering in block Markov chains

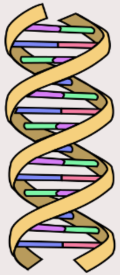
- **J. Sanders, A. Proutière, S.-Y. Yun (2020):**
  Spectral clustering algorithm
- **J. Sanders, A. Senen–Cerda (2023)**
  Asymptotic estimate $c\sigma$ for non-explicit $c$.
- **A. Van Werde, A. Senen–Cerda, G. Kosmella, J. Sanders (2023)**
  Real-world data



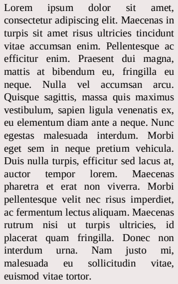
Animal movement



DNA



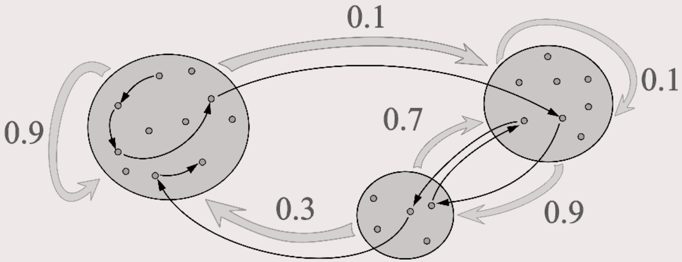
Text



Lorem ipsum dolor sit amet, consectetur adipiscing elit. Maecenas in turpis sit amet risus ultricies tincidunt vitae accumsan enim. Pellentesque ac efficitur enim. Praesent dui magna, mattis at bibendum eu, fringilla eu neque. Nulla vel accumsan arcu. Quisque sagittis, massa quis maximus vestibulum, sapien ligula venenatis ex, eu elementum diam ante a neque. Nunc egestas malesuada interdum. Morbi eget sem in neque pretium vehicula. Duis nulla turpis, efficitur sed lacus at, auctor tempor lorem. Maecenas pharetra et erat non viverra. Morbi pellentesque velit nec risus imperdiet, ac fermentum lectus aliquam. Maecenas rutrum nisi ut turpis ultricies, id placerat quam fringilla. Donec non interdum urna. Nam justo mi, malesuada eu sollicitudin vitae, euismod vitae tortor.

Stock market

# Application: Block Markov chains

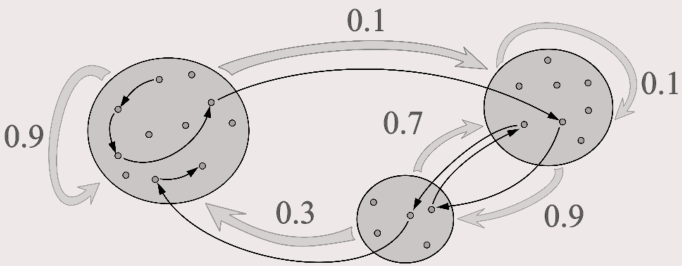


**Definition.** Fix $K \geq 1$, a stochastic matrix $p \in [0,1]^{K\times K}$, and partition $\mathcal{V}_1, \ldots, \mathcal{V}_K$ of $\{1, \ldots, d\}$.

A *Block Markov chain* has transition matrix

$$P_{i,j} = \frac{p_{x,y}}{\#\mathcal{V}_y} \qquad \forall i \in \mathcal{V}_x, j \in \mathcal{V}_y$$

# Application: Block Markov chains



**Definition.** Fix $K \geq 1$, a stochastic matrix $p \in [0,1]^{K\times K}$, and partition $\mathcal{V}_1, \ldots, \mathcal{V}_K$ of $\{1, \ldots, d\}$.

A *Block Markov chain* has transition matrix

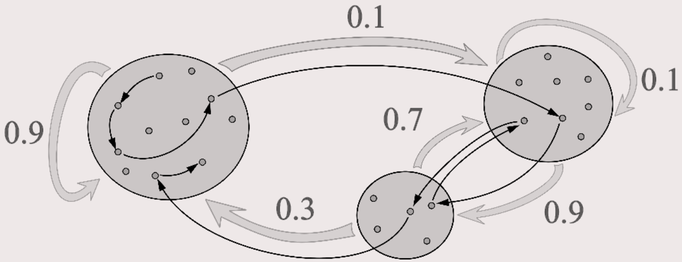
$$P_{i,j} = \frac{p_{x,y}}{\#\mathcal{V}_y} \qquad \forall i \in \mathcal{V}_x, j \in \mathcal{V}_y$$

# Application: Block Markov chains

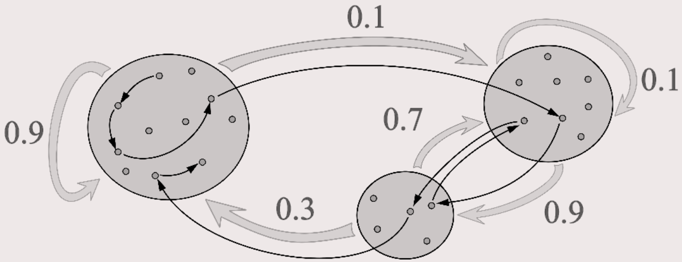
**Definition.** Fix $K \geq 1$, a stochastic matrix $p \in [0,1]^{K\times K}$, and partition $\mathcal{V}_1, \dots, \mathcal{V}_K$ of $\{1, \dots, d\}$.

A *Block Markov chain* has transition matrix

$$P_{i,j} = \frac{p_{x,y}}{\#\mathcal{V}_y} \qquad \forall i \in \mathcal{V}_x, j \in \mathcal{V}_y$$



# Application: Block Markov chains



**Definition.** Fix $K \geq 1$, a stochastic matrix $p \in [0,1]^{K\times K}$, and partition $\mathcal{V}_1, \dots, \mathcal{V}_K$ of $\{1, \dots, d\}$.

A *Block Markov chain* has transition matrix

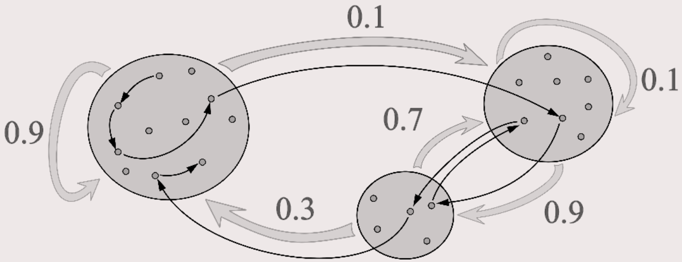
$$P_{i,j} = \frac{p_{x,y}}{\#\mathcal{V}_y} \qquad \forall i \in \mathcal{V}_x, j \in \mathcal{V}_y$$

# Application: Block Markov chains

**Definition.** Fix $K \geq 1$, a stochastic matrix $p \in [0,1]^{K\times K}$, and partition $\mathcal{V}_1, \ldots, \mathcal{V}_K$ of $\{1, \ldots, d\}$.

A *Block Markov chain* has transition matrix

$$P_{i,j} = \frac{p_{x,y}}{\#\mathcal{V}_y} \qquad \forall i \in \mathcal{V}_x, j \in \mathcal{V}_y$$



# Application: Block Markov chains

**Definition.** Fix $K \geq 1$, a stochastic matrix $p \in [0,1]^{K\times K}$, and partition $\mathcal{V}_1, \ldots, \mathcal{V}_K$ of $\{1, \ldots, d\}$.

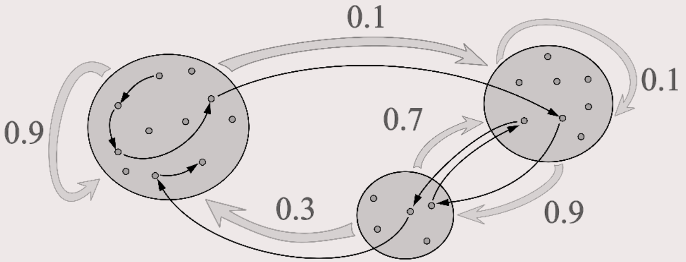
A *Block Markov chain* has transition matrix

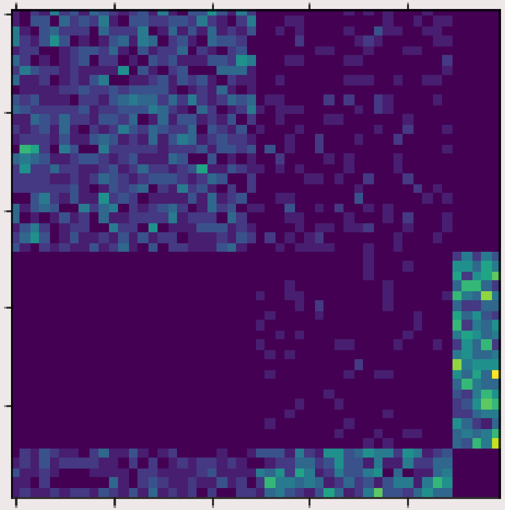
$$P_{i,j} = \frac{p_{x,y}}{\#\mathcal{V}_y} \qquad \forall i \in \mathcal{V}_x, j \in \mathcal{V}_y$$

**Definition.** Given a sample path $Z_1, \ldots, Z_n$. *Sample frequency matrix* $\boldsymbol{N} \in \mathbb{Z}^{d\times d}$:

$$\boldsymbol{N}_{i,j} := \#\{\text{transitions i} \rightarrow \text{j}\}$$

$$= \sum_{t=1}^{n-1} 1\{Z_t = i, Z_{t+1} = j\}$$





# Application: Block Markov chains

**Definition.** Fix $K \geq 1$, a stochastic matrix $p \in [0,1]^{K\times K}$, and partition $\mathcal{V}_1, \dots, \mathcal{V}_K$ of $\{1, \dots, d\}$.

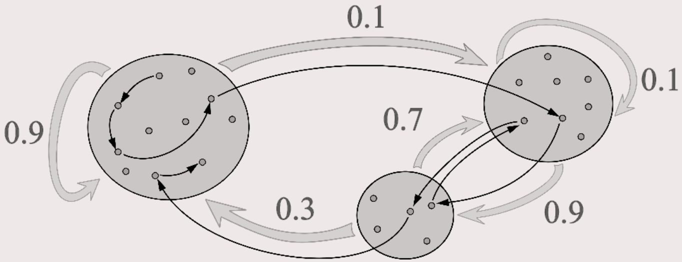
A *Block Markov chain* has transition matrix

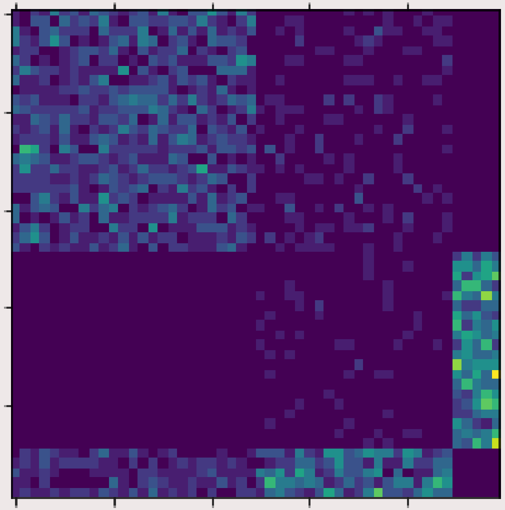
$$P_{i,j} = \frac{p_{x,y}}{\#\mathcal{V}_y} \qquad \forall i \in \mathcal{V}_x, j \in \mathcal{V}_y$$

**Definition.** Given a sample path $Z_1, \dots, Z_n$. *Sample frequency matrix* $\boldsymbol{N} \in \mathbb{Z}^{d\times d}$:

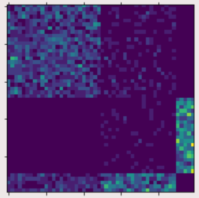
$$\boldsymbol{N}_{i,j} \coloneqq \#\{\text{transitions i} \to \text{j}\}$$

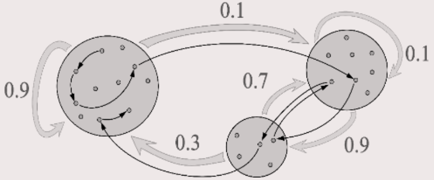
$$= \sum_{t=1}^{n-1} 1\{Z_t = i, Z_{t+1} = j\}$$





# Application: Block Markov chains





Let $\boldsymbol{M} := \sqrt{d/n}\,(\boldsymbol{N} - \mathbb{E}\boldsymbol{N})$

**Corollary.**

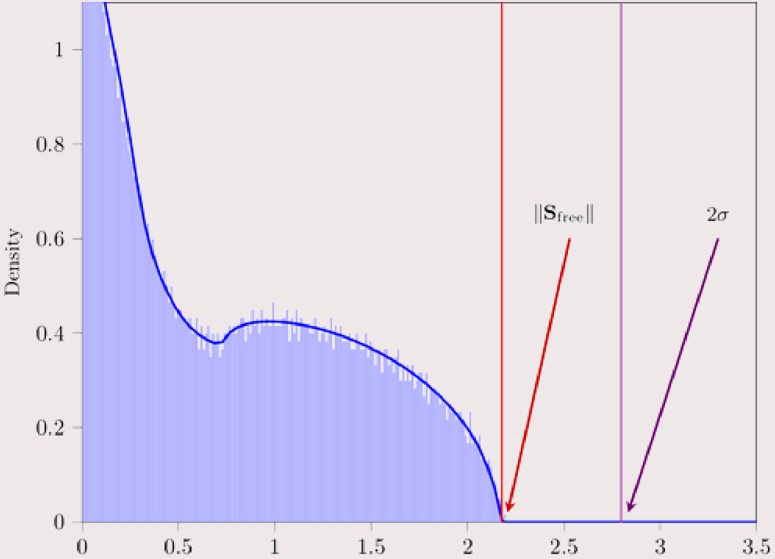
Suppose $n \gg d\ln(d)^4$ as $d \to \infty$, then

$$\|\boldsymbol{M}\| \to \mathcal{M} \quad \text{in probability}$$

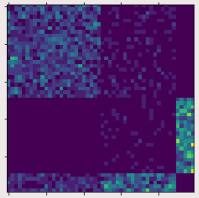
where

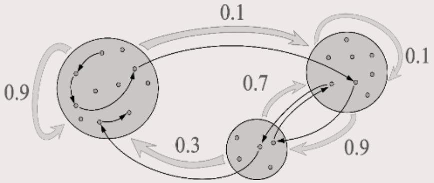
$$\mathcal{M} := \inf_{x \in \mathbb{R}^{2K}_{>0}} \max_{i=1,..,2K} \left\{ \frac{1}{x_i} + \sum_{j=1}^{n} c_{ij}\, x_j \right\}$$

with explicit coefficients $c_{ij}$.



# Application: Block Markov chains





Let $\boldsymbol{M} := \sqrt{d/n}\,(\boldsymbol{N} - \mathbb{E}\boldsymbol{N})$

**Corollary.**

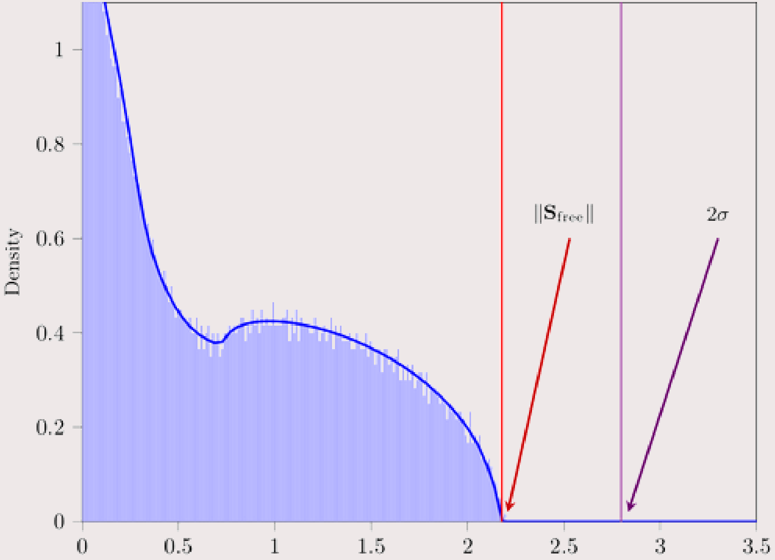
Suppose $n \gg d\ln(d)^4$ as $d \to \infty$, then

$$\|\boldsymbol{M}\| \to \mathcal{M} \quad \text{in probability}$$

where

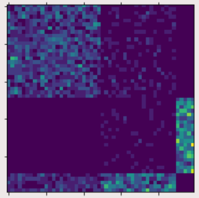
$$\mathcal{M} := \inf_{x\in\mathbb{R}^{2K}_{>0}} \max_{i=1,..,2K} \left\{ \frac{1}{x_i} + \sum_{j=1}^{n} c_{ij}\, x_j \right\}$$

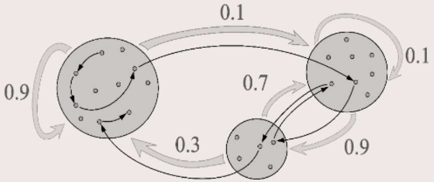
with explicit coefficients $c_{ij}$.



# Application: Block Markov chains

Let $\boldsymbol{M} := \sqrt{d/n}\,(\boldsymbol{N} - \mathbb{E}\boldsymbol{N})$





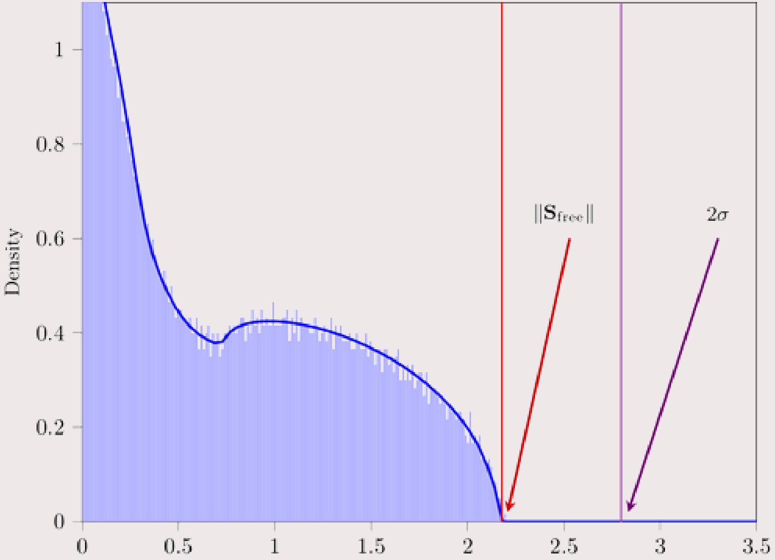
**Corollary.**
Suppose $n \gg d\ln(d)^4$ as $d \to \infty$, then

$$\|\boldsymbol{M}\| \to \mathcal{M} \quad \text{in probability}$$

where

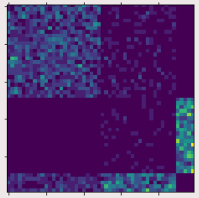
$$\mathcal{M} := \inf_{x\in\mathbb{R}^{2K}_{>0}} \max_{i=1,..,2K} \left\{ \frac{1}{x_i} + \sum_{j=1}^{n} c_{ij}\, x_j \right\}$$

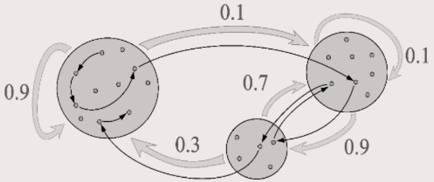
with explicit coefficients $c_{ij}$.



# Application: Block Markov chains

Let $\boldsymbol{M} := \sqrt{d/n}\,(\boldsymbol{N} - \mathbb{E}\boldsymbol{N})$





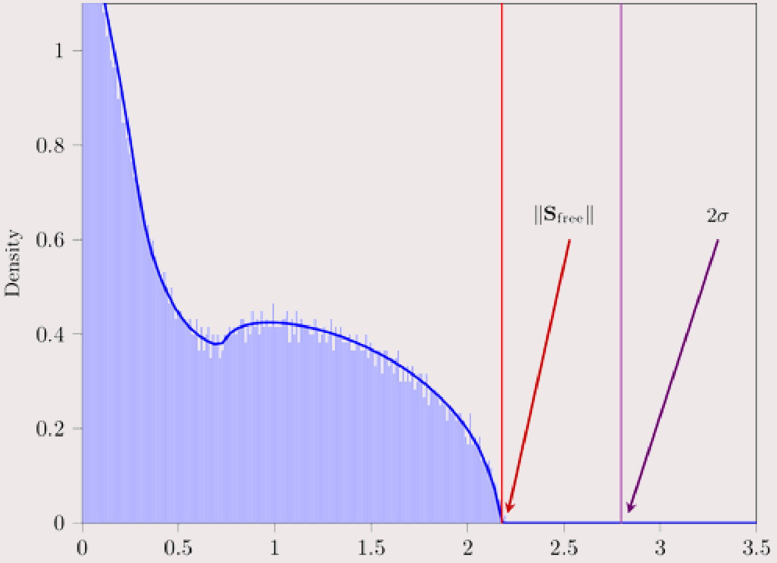
**Corollary.**
Suppose $n \gg d \ln(d)^4$ as $d \to \infty$, then

$$\|\boldsymbol{M}\| \to \mathcal{M} \quad \text{in probability}$$

where

$$\mathcal{M} := \inf_{x \in \mathbb{R}^{2K}_{>0}} \max_{i=1,..,2K} \left\{ \frac{1}{x_i} + \sum_{j=1}^{n} c_{ij}\, x_j \right\}$$

with explicit coefficients $c_{ij}$.



# Applications of matrix series model

**Heavy-tailed independent entries**

Let $\boldsymbol{S}$ be a Wigner matrix with independent sub-Weibull entries.

Then, we show that

$$\|\boldsymbol{S}\| \leq 2\sigma + \varepsilon \text{ with high probability.}$$

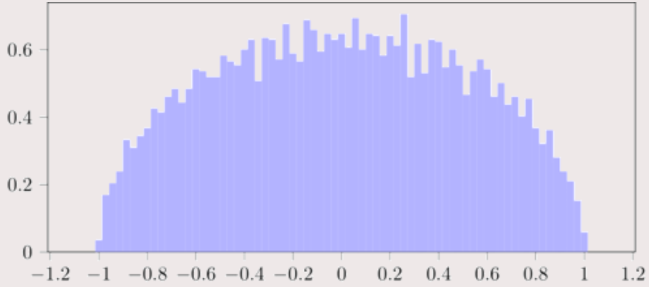
We provide explicit tail bounds.

This is a quantitative non-asymptotic variant on the Bai-Yin theorem.

**Spectra of graphs with dependent edges**

Let $\boldsymbol{A}$ be the adjacency matrix of an Erdős–Rényi random graph with $m$ edges and $d$ nodes.

We show convergence of empirical eigenvalue distributions when $m = \omega(d)$.



# Applications of matrix series model

**Heavy-tailed independent entries**

Let $\boldsymbol{S}$ be a Wigner matrix with independent sub-Weibull entries.

Then, we show that

$$\|\boldsymbol{S}\| \leq 2\sigma + \varepsilon \text{ with high probability.}$$

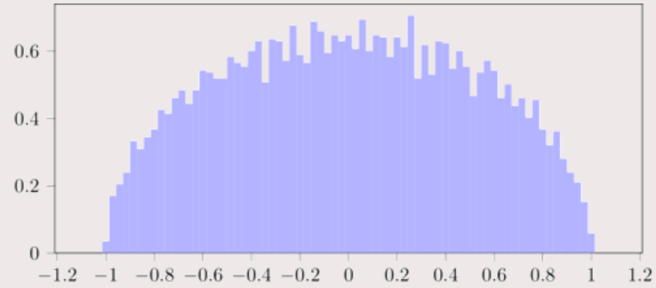
We provide explicit tail bounds.

This is a quantitative non-asymptotic variant on the Bai-Yin theorem.

**Spectra of graphs with dependent edges**

Let $\boldsymbol{A}$ be the adjacency matrix of an Erdős–Rényi random graph with $m$ edges and $d$ nodes.

We show convergence of empirical eigenvalue distributions when $m = \omega(d)$.



# Applications of matrix series model

**Heavy-tailed independent entries**

Let $\boldsymbol{S}$ be a Wigner matrix with independent sub-Weibull entries.

Then, we show that

$$\|\boldsymbol{S}\| \leq 2\sigma + \varepsilon \text{ with high probability.}$$

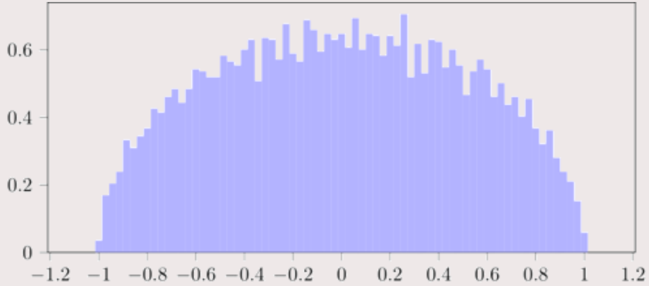
We provide explicit tail bounds.

This is a quantitative non-asymptotic variant on the Bai-Yin theorem.

**Spectra of graphs with dependent edges**

Let $\boldsymbol{A}$ be the adjacency matrix of an Erdős–Rényi random graph with $m$ edges and $d$ nodes.

We show convergence of empirical eigenvalue distributions when $m = \omega(d)$.



# Summary

We establish ***sharp*** matrix concentration inequalities for sum of ***dependent*** random matrices.

Boolean cumulants play a crucial role in the proof. Surprisingly, classical cumulants are insufficient for sharp estimates.

Sharp concentration in applications, such as block Markov chains, would be inaccessible by previous nonasymptotic dependent results.

# Summary

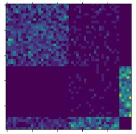
We establish ***sharp*** matrix concentration inequalities for sum of ***dependent*** random matrices.

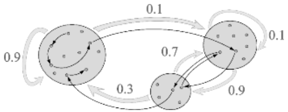
Boolean cumulants play a crucial role in the proof. Surprisingly, classical cumulants are insufficient for sharp estimates.

Sharp concentration in applications, such as block Markov chains, would be inaccessible by previous nonasymptotic dependent results.

**Appears on arXiv soon!**

# Thank you!





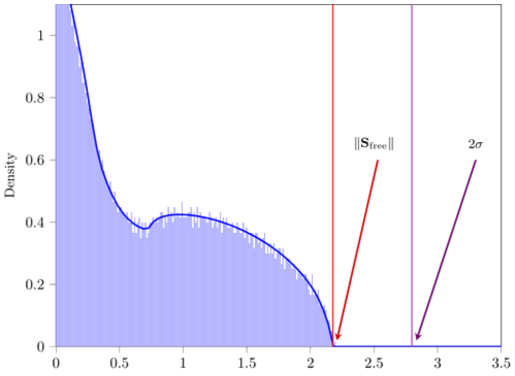
**Main result for Markovian model.**

There exists $c > 0$ such that for any $0 < \delta \leq 1$

$$\mathbb{E}\|\boldsymbol{S}\| \leq (1+\delta)\|\boldsymbol{S}_{\text{free}}\| + c\mathcal{E}_{d,\delta}$$

where

$$\mathcal{E}_{d,\delta} \coloneqq v^{1/2}\sigma^{1/2}\log_{1+\delta}(2d)^{3/4} + R^{1/3}\Psi^{2/3}\varsigma^{2/3}\log_{1+\delta}(2d)^{2/3} + R\Psi\log_{1+\delta}(2d)$$



**Dependence parameter**

$$\Psi \coloneqq \min\{\, t \geq 1 \colon \frac{|P^t_{x,y} - \pi_y|}{\pi_y} \leq \frac{1}{4} \}$$

**Matrix parameters**

$$\varsigma^2 \coloneqq \left\|\mathbb{E}\left[\textstyle\sum_{i=1}^n \boldsymbol{X}_i^2\right]\right\|, \quad \sigma^2 \coloneqq \|\mathbb{E}[\boldsymbol{S}^2]\|$$

$$v^2 \coloneqq \|\mathrm{Cov}(\boldsymbol{S})\|, \quad R \geq \|\boldsymbol{X}_i\|$$

**Free-probabilistic quantity**

$$\|\boldsymbol{S}_{\text{free}}\| = \inf_{\boldsymbol{W}\succ\boldsymbol{0}} \lambda_{\max}\left(\boldsymbol{W}^{-1} + \mathbb{E}[\boldsymbol{S}\boldsymbol{W}\boldsymbol{S}]\right).$$

a.van.werde@tue.nl