

# Concentration of measure, free probability, and Markov chains

Sourav Chatterjee

# Motivational example

- ▶ Let  $M$  be an  $n \times n$  hermitian matrix with eigenvalues  $\lambda_1, \dots, \lambda_n$ .

# Motivational example

- ▶ Let  $M$  be an  $n \times n$  hermitian matrix with eigenvalues  $\lambda_1, \dots, \lambda_n$ .
- ▶ Let  $F_M(x) = \frac{1}{n} \sum_{i=1}^n \mathbb{I}_{\{\lambda_i \leq x\}}$  denote the empirical distribution function of  $M$ .

# Motivational example

- ▶ Let  $M$  be an  $n \times n$  hermitian matrix with eigenvalues  $\lambda_1, \dots, \lambda_n$ .
- ▶ Let  $F_M(x) = \frac{1}{n} \sum_{i=1}^n \mathbb{I}_{\{\lambda_i \leq x\}}$  denote the empirical distribution function of  $M$ .
- ▶ Let  $\rho_M$  denote the uniform distribution on the set of all hermitian matrices having the same spectrum as  $M$ .

# Example contd.

- ▶ Let  $M$  and  $N$  be two hermitian matrices of order  $n$ .

## Example contd.

- ▶ Let  $M$  and  $N$  be two hermitian matrices of order  $n$ .
- ▶ Let  $A \sim \rho_M$  and  $B \sim \rho_N$  be independent random matrices. Let  $H = A + B$ .

## Example contd.

- ▶ Let  $M$  and  $N$  be two hermitian matrices of order  $n$ .
- ▶ Let  $A \sim \rho_M$  and  $B \sim \rho_N$  be independent random matrices. Let  $H = A + B$ .
- ▶ Results of Gromov, Milman, Szarek give concentration bounds for  $\frac{1}{n} \sum_i f(\lambda_i^H)$  when  $f$  is smooth.

## Example contd.

- ▶ Let  $M$  and  $N$  be two hermitian matrices of order  $n$ .
- ▶ Let  $A \sim \rho_M$  and  $B \sim \rho_N$  be independent random matrices. Let  $H = A + B$ .
- ▶ Results of Gromov, Milman, Szarek give concentration bounds for  $\frac{1}{n} \sum_i f(\lambda_i^H)$  when  $f$  is smooth.
- ▶ What about the concentration of  $F_H(x)$ ?

## Theorem

Let  $M$  and  $N$  be two  $n \times n$  hermitian matrices. Suppose  $A \sim \rho_M$  and  $B \sim \rho_N$  are independent, and  $H = A + B$ . Then, for every  $x \in \mathbb{R}$ ,  $\text{Var}(F_H(x)) \leq \kappa n^{-1} \log n$ , and

$$\begin{aligned} & \mathbb{P}\{|F_H(x) - \mathbb{E}(F_H(x))| \geq t\} \\ & \leq 2 \exp\left(-\frac{nt^2}{2\kappa \log n}\right). \end{aligned}$$

Here  $\kappa$  is a numerical constant independent of everything else.

(Recall:  $\rho_M$  is the uniform distribution on all  $n \times n$  hermitian matrices having the same spectrum as  $M$ .)

- ▶ Gromov-Milman-Szarek give concentration under Hilbert-Schmidt distance.

- ▶ Gromov-Milman-Szarek give concentration under Hilbert-Schmidt distance.
- ▶  $H \mapsto F_H(x)$  is not Lipschitz under that metric.

- ▶ Gromov-Milman-Szarek give concentration under Hilbert-Schmidt distance.
- ▶  $H \mapsto F_H(x)$  is not Lipschitz under that metric.
- ▶ Idea # 1: Change the metric to make the map Lipschitz.

- ▶ Gromov-Milman-Szarek give concentration under Hilbert-Schmidt distance.
- ▶  $H \mapsto F_H(x)$  is not Lipschitz under that metric.
- ▶ Idea # 1: Change the metric to make the map Lipschitz. New metric:  $d(M, N) = \text{rank}(M - N)$ .

- ▶ Gromov-Milman-Szarek give concentration under Hilbert-Schmidt distance.
- ▶  $H \mapsto F_H(x)$  is not Lipschitz under that metric.
- ▶ Idea # 1: Change the metric to make the map Lipschitz. New metric:  $d(M, N) = \text{rank}(M - N)$ .
- ▶ Fact:

$$\|F_M - F_N\|_\infty \leq n^{-1} \text{rank}(M - N).$$

- ▶ Gromov-Milman-Szarek give concentration under Hilbert-Schmidt distance.
- ▶  $H \mapsto F_H(x)$  is not Lipschitz under that metric.
- ▶ Idea # 1: Change the metric to make the map Lipschitz. New metric:  $d(M, N) = \text{rank}(M - N)$ .

- ▶ Fact:

$$\|F_M - F_N\|_\infty \leq n^{-1} \text{rank}(M - N).$$

- ▶ Need: A measure concentration technique that can be adapted to arbitrary metrics.

- ▶ Idea # 2: Concentration inequalities from rates of convergence of Markov chains that take “small steps” w.r.t. a given metric.

- ▶ Idea # 2: Concentration inequalities from rates of convergence of Markov chains that take “small steps” w.r.t. a given metric.
- ▶ For rank distance, we have the “random reflection walk” on hermitian matrices.

- ▶ Idea # 2: Concentration inequalities from rates of convergence of Markov chains that take “small steps” w.r.t. a given metric.
- ▶ For rank distance, we have the “random reflection walk” on hermitian matrices.
  - ▶ At step  $i$ , let

$$U = I - (1 - e^{i\varphi})uu^*,$$

where

- ▶ Idea # 2: Concentration inequalities from rates of convergence of Markov chains that take “small steps” w.r.t. a given metric.
- ▶ For rank distance, we have the “random reflection walk” on hermitian matrices.
  - ▶ At step  $i$ , let

$$U = I - (1 - e^{i\varphi})uu^*,$$

where

- ▶  $u$  is chosen from the uniform distribution on the unit sphere in  $\mathbb{C}^n$ ,  
and

- ▶ Idea # 2: Concentration inequalities from rates of convergence of Markov chains that take “small steps” w.r.t. a given metric.
- ▶ For rank distance, we have the “random reflection walk” on hermitian matrices.
  - ▶ At step  $i$ , let

$$U = I - (1 - e^{i\varphi})uu^*,$$

where

- ▶  $u$  is chosen from the uniform distribution on the unit sphere in  $\mathbb{C}^n$ , and
- ▶  $\varphi$  has density proportional to  $(\sin(\varphi/2))^{n-1}$  on  $[0, 2\pi)$ .

- ▶ Idea # 2: Concentration inequalities from rates of convergence of Markov chains that take “small steps” w.r.t. a given metric.
- ▶ For rank distance, we have the “random reflection walk” on hermitian matrices.
  - ▶ At step  $i$ , let

$$U = I - (1 - e^{i\varphi})uu^*,$$

where

- ▶  $u$  is chosen from the uniform distribution on the unit sphere in  $\mathbb{C}^n$ , and
  - ▶  $\varphi$  has density proportional to  $(\sin(\varphi/2))^{n-1}$  on  $[0, 2\pi)$ .
- ▶ Let  $M_{i+1} = UM_iU^*$ .

- ▶ Idea # 2: Concentration inequalities from rates of convergence of Markov chains that take “small steps” w.r.t. a given metric.
- ▶ For rank distance, we have the “random reflection walk” on hermitian matrices.

- ▶ At step  $i$ , let

$$U = I - (1 - e^{i\varphi})uu^*,$$

where

- ▶  $u$  is chosen from the uniform distribution on the unit sphere in  $\mathbb{C}^n$ ,  
and
  - ▶  $\varphi$  has density proportional to  $(\sin(\varphi/2))^{n-1}$  on  $[0, 2\pi)$ .
  - ▶ Let  $M_{i+1} = UM_iU^*$ .
- ▶ Easy:  $\text{rank}(M_{i+1} - M_i) \leq 3$ .

- ▶ Idea # 2: Concentration inequalities from rates of convergence of Markov chains that take “small steps” w.r.t. a given metric.
- ▶ For rank distance, we have the “random reflection walk” on hermitian matrices.

- ▶ At step  $i$ , let

$$U = I - (1 - e^{i\varphi})uu^*,$$

where

- ▶  $u$  is chosen from the uniform distribution on the unit sphere in  $\mathbb{C}^n$ , and
  - ▶  $\varphi$  has density proportional to  $(\sin(\varphi/2))^{n-1}$  on  $[0, 2\pi)$ .
  - ▶ Let  $M_{i+1} = UM_iU^*$ .
- ▶ Easy:  $\text{rank}(M_{i+1} - M_i) \leq 3$ .
  - ▶ Difficult: Convergence to stationary measure  $\rho_{M_1}$  in  $Cn \log n$  steps (Porod '96).

# General theorem about Haar measures

- ▶ Let  $G$  be a compact topological group.

# General theorem about Haar measures

- ▶ Let  $G$  be a compact topological group.
- ▶ Let  $X$  be a Haar-distributed random variable on  $G$ .

# General theorem about Haar measures

- ▶ Let  $G$  be a compact topological group.
- ▶ Let  $X$  be a Haar-distributed random variable on  $G$ .
- ▶ Let  $Y$  be another  $G$ -valued r.v. having the following properties:

# General theorem about Haar measures

- ▶ Let  $G$  be a compact topological group.
- ▶ Let  $X$  be a Haar-distributed random variable on  $G$ .
- ▶ Let  $Y$  be another  $G$ -valued r.v. having the following properties:
  1.  $Y^{-1}$  has the same distribution as  $Y$ .

# General theorem about Haar measures

- ▶ Let  $G$  be a compact topological group.
- ▶ Let  $X$  be a Haar-distributed random variable on  $G$ .
- ▶ Let  $Y$  be another  $G$ -valued r.v. having the following properties:
  1.  $Y^{-1}$  has the same distribution as  $Y$ .
  2. For any  $x \in G$ ,  $xYx^{-1}$  has the same distribution as  $Y$ .

# General theorem about Haar measures

- ▶ Let  $G$  be a compact topological group.
- ▶ Let  $X$  be a Haar-distributed random variable on  $G$ .
- ▶ Let  $Y$  be another  $G$ -valued r.v. having the following properties:
  1.  $Y^{-1}$  has the same distribution as  $Y$ .
  2. For any  $x \in G$ ,  $xYx^{-1}$  has the same distribution as  $Y$ .
- ▶ Let  $Y_1, Y_2, \dots$ , be i.i.d. copies of  $Y$ . Suppose  $a$  and  $b$  are two positive constants such that

$$d_{TV}(Y_1 Y_2 \cdots Y_k, X) \leq a e^{-bk}$$

for every  $k$ , where  $d_{TV}$  is the total variation distance.

# General theorem about Haar measures

- ▶ Let  $G$  be a compact topological group.
- ▶ Let  $X$  be a Haar-distributed random variable on  $G$ .
- ▶ Let  $Y$  be another  $G$ -valued r.v. having the following properties:
  1.  $Y^{-1}$  has the same distribution as  $Y$ .
  2. For any  $x \in G$ ,  $xYx^{-1}$  has the same distribution as  $Y$ .
- ▶ Let  $Y_1, Y_2, \dots$ , be i.i.d. copies of  $Y$ . Suppose  $a$  and  $b$  are two positive constants such that

$$d_{TV}(Y_1 Y_2 \cdots Y_k, X) \leq ae^{-bk}$$

for every  $k$ , where  $d_{TV}$  is the total variation distance.

(Think of  $X$  as a Haar-distributed unitary matrix and  $Y$  as a random reflection. Porod '96 gives  $a \sim n$  and  $b \sim 1/n$ .)

# General Theorem (contd.)

## Theorem

Suppose  $f : G \rightarrow \mathbb{R}$  is bounded and measurable. Let  $\|f\|_\infty = \sup_{x \in G} |f(x)|$  and

$$\|f\|_Y := \sup_{x \in G} [\mathbb{E}(f(x) - f(Y_x))^2]^{1/2}.$$

Let  $A$  and  $B$  be two numbers such that  $\|f\|_\infty \leq A$  and  $\|f\|_Y \leq B$ , and  $a, b$  as above. Let

$$C = \frac{B^2}{b} \left[ \left( \log \frac{4aA}{B} \right)^+ + \frac{b(2 - e^{-b})}{1 - e^{-b}} \right].$$

Then  $\text{Var}(f(X)) \leq C/2$ , and for any  $t \geq 0$ ,

$$\mathbb{P}\{|f(X) - \mathbb{E}f(X)| \geq t\} \leq 2e^{-t^2/C}.$$

- ▶ In our example,  $X$  is a Haar-distributed unitary matrix, and  $f(X) = F_H(x)$ , where

- ▶ In our example,  $X$  is a Haar-distributed unitary matrix, and  $f(X) = F_H(x)$ , where  $x \in \mathbb{R}$  is fixed, and

$$H = XMN^* + N,$$

$M$  and  $N$  being fixed hermitian matrices.

- ▶ In our example,  $X$  is a Haar-distributed unitary matrix, and  $f(X) = F_H(x)$ , where  $x \in \mathbb{R}$  is fixed, and

$$H = XMX^* + N,$$

$M$  and  $N$  being fixed hermitian matrices.

- ▶ We have  $A = 1$ ,  $B \sim 1/n$ ,  $a \sim n$ , and  $b \sim 1/n$ . Plugging in, we get our result.

- ▶ In our example,  $X$  is a Haar-distributed unitary matrix, and  $f(X) = F_H(x)$ , where  $x \in \mathbb{R}$  is fixed, and

$$H = XMX^* + N,$$

$M$  and  $N$  being fixed hermitian matrices.

- ▶ We have  $A = 1$ ,  $B \sim 1/n$ ,  $a \sim n$ , and  $b \sim 1/n$ . Plugging in, we get our result.
- ▶ Also applicable to products of random matrices.

- ▶ In our example,  $X$  is a Haar-distributed unitary matrix, and  $f(X) = F_H(x)$ , where  $x \in \mathbb{R}$  is fixed, and

$$H = XMX^* + N,$$

$M$  and  $N$  being fixed hermitian matrices.

- ▶ We have  $A = 1$ ,  $B \sim 1/n$ ,  $a \sim n$ , and  $b \sim 1/n$ . Plugging in, we get our result.
- ▶ Also applicable to products of random matrices.
- ▶ Applies to other groups, e.g. symmetric groups.

- ▶ In our example,  $X$  is a Haar-distributed unitary matrix, and  $f(X) = F_H(x)$ , where  $x \in \mathbb{R}$  is fixed, and

$$H = XMX^* + N,$$

$M$  and  $N$  being fixed hermitian matrices.

- ▶ We have  $A = 1$ ,  $B \sim 1/n$ ,  $a \sim n$ , and  $b \sim 1/n$ . Plugging in, we get our result.
- ▶ Also applicable to products of random matrices.
- ▶ Applies to other groups, e.g. symmetric groups.
- ▶ Can go beyond Haar measures and groups. Will not discuss today.

- ▶ Let  $X_0 = X$ ,  $X'_0 = XY_0$ . Construct

$$X_0 \rightarrow X_1 \rightarrow X_2 \rightarrow \dots \text{ and } X'_0 \rightarrow X'_1 \rightarrow X'_2 \rightarrow \dots$$

# Idea of proof

- ▶ Let  $X_0 = X$ ,  $X'_0 = XY_0$ . Construct

$$X_0 \rightarrow X_1 \rightarrow X_2 \rightarrow \dots \text{ and } X'_0 \rightarrow X'_1 \rightarrow X'_2 \rightarrow \dots$$

as  $X_{i+1} = X_i Y_{i+1}$  and  $X'_{i+1} = X'_i Y'_{i+1}$ , where  $Y'_{i+1} = Y_0^{-1} Y_{i+1} Y_0$ .

- ▶ Let  $X_0 = X$ ,  $X'_0 = XY_0$ . Construct

$$X_0 \rightarrow X_1 \rightarrow X_2 \rightarrow \dots \text{ and } X'_0 \rightarrow X'_1 \rightarrow X'_2 \rightarrow \dots$$

as  $X_{i+1} = X_i Y_{i+1}$  and  $X'_{i+1} = X'_i Y'_{i+1}$ , where  $Y'_{i+1} = Y_0^{-1} Y_{i+1} Y_0$ .

- ▶ Key step:

$$\mathbb{E}(g(X)f(X)) = \frac{1}{2} \sum_{i=0}^{\infty} \mathbb{E}[(g(X_0) - g(X'_0))(f(X_i) - f(X'_i))]$$

for any  $g$  and  $f$ .

# Idea of proof

- ▶ Let  $X_0 = X$ ,  $X'_0 = XY_0$ . Construct

$$X_0 \rightarrow X_1 \rightarrow X_2 \rightarrow \dots \text{ and } X'_0 \rightarrow X'_1 \rightarrow X'_2 \rightarrow \dots$$

as  $X_{i+1} = X_i Y_{i+1}$  and  $X'_{i+1} = X'_i Y'_{i+1}$ , where  $Y'_{i+1} = Y_0^{-1} Y_{i+1} Y_0$ .

- ▶ Key step:

$$\mathbb{E}(g(X)f(X)) = \frac{1}{2} \sum_{i=0}^{\infty} \mathbb{E}[(g(X_0) - g(X'_0))(f(X_i) - f(X'_i))]$$

for any  $g$  and  $f$ .

- ▶ Note that  $X'_i = X_i Y_0$  for all  $i$ .

# Idea of proof

- ▶ Let  $X_0 = X$ ,  $X'_0 = XY_0$ . Construct

$$X_0 \rightarrow X_1 \rightarrow X_2 \rightarrow \dots \text{ and } X'_0 \rightarrow X'_1 \rightarrow X'_2 \rightarrow \dots$$

as  $X_{i+1} = X_i Y_{i+1}$  and  $X'_{i+1} = X'_i Y'_{i+1}$ , where  $Y'_{i+1} = Y_0^{-1} Y_{i+1} Y_0$ .

- ▶ Key step:

$$\mathbb{E}(g(X)f(X)) = \frac{1}{2} \sum_{i=0}^{\infty} \mathbb{E}[(g(X_0) - g(X'_0))(f(X_i) - f(X'_i))]$$

for any  $g$  and  $f$ .

- ▶ Note that  $X'_i = X_i Y_0$  for all  $i$ . Thus,  $|\mathbb{E}(g(X)f(X))| \leq O(B^2\tau)$ , where  $B \geq$  the “size” of  $g(X) - g(XY)$  and  $f(X) - f(XY)$ , and  $\tau =$  mixing time of the chain.

# Idea of proof

- ▶ Let  $X_0 = X$ ,  $X'_0 = XY_0$ . Construct

$$X_0 \rightarrow X_1 \rightarrow X_2 \rightarrow \dots \text{ and } X'_0 \rightarrow X'_1 \rightarrow X'_2 \rightarrow \dots$$

as  $X_{i+1} = X_i Y_{i+1}$  and  $X'_{i+1} = X'_i Y'_{i+1}$ , where  $Y'_{i+1} = Y_0^{-1} Y_{i+1} Y_0$ .

- ▶ Key step:

$$\mathbb{E}(g(X)f(X)) = \frac{1}{2} \sum_{i=0}^{\infty} \mathbb{E}[(g(X_0) - g(X'_0))(f(X_i) - f(X'_i))]$$

for any  $g$  and  $f$ .

- ▶ Note that  $X'_i = X_i Y_0$  for all  $i$ . Thus,  $|\mathbb{E}(g(X)f(X))| \leq O(B^2\tau)$ , where  $B \geq$  the “size” of  $g(X) - g(XY)$  and  $f(X) - f(XY)$ , and  $\tau =$  mixing time of the chain.
- ▶ For variance bound, take  $g = f$ . For concentration, take  $g = e^{\theta f}$ .

# Summary

- ▶ Given a random variable  $X$  on some space, and a function  $f$ , we want to get tail bounds for  $f(X)$ .

# Summary

- ▶ Given a random variable  $X$  on some space, and a function  $f$ , we want to get tail bounds for  $f(X)$ .
  - ▶ Step 1: Choose a metric  $d$  so that  $f$  is Lipschitz w.r.t.  $d$ .

# Summary

- ▶ Given a random variable  $X$  on some space, and a function  $f$ , we want to get tail bounds for  $f(X)$ .
  - ▶ Step 1: Choose a metric  $d$  so that  $f$  is Lipschitz w.r.t.  $d$ .
  - ▶ Step 2: Find a Markov chain whose stationary distribution is the law of  $X$ , and which takes small steps w.r.t.  $d$ .

# Summary

- ▶ Given a random variable  $X$  on some space, and a function  $f$ , we want to get tail bounds for  $f(X)$ .
  - ▶ Step 1: Choose a metric  $d$  so that  $f$  is Lipschitz w.r.t.  $d$ .
  - ▶ Step 2: Find a Markov chain whose stationary distribution is the law of  $X$ , and which takes small steps w.r.t.  $d$ .
  - ▶ Step 3: Under some further conditions, compute concentration inequalities for  $f(X)$  using the rate of convergence of the Markov chain.

# Summary

- ▶ Given a random variable  $X$  on some space, and a function  $f$ , we want to get tail bounds for  $f(X)$ .
  - ▶ Step 1: Choose a metric  $d$  so that  $f$  is Lipschitz w.r.t.  $d$ .
  - ▶ Step 2: Find a Markov chain whose stationary distribution is the law of  $X$ , and which takes small steps w.r.t.  $d$ .
  - ▶ Step 3: Under some further conditions, compute concentration inequalities for  $f(X)$  using the rate of convergence of the Markov chain.
- ▶ For Haar measures, take random walks that are “constant on conjugacy classes”, like the random reflections walk.

# Summary

- ▶ Given a random variable  $X$  on some space, and a function  $f$ , we want to get tail bounds for  $f(X)$ .
  - ▶ Step 1: Choose a metric  $d$  so that  $f$  is Lipschitz w.r.t.  $d$ .
  - ▶ Step 2: Find a Markov chain whose stationary distribution is the law of  $X$ , and which takes small steps w.r.t.  $d$ .
  - ▶ Step 3: Under some further conditions, compute concentration inequalities for  $f(X)$  using the rate of convergence of the Markov chain.
- ▶ For Haar measures, take random walks that are “constant on conjugacy classes”, like the random reflections walk.
- ▶ Applications to random matrices, random permutations, etc.

# Summary

- ▶ Given a random variable  $X$  on some space, and a function  $f$ , we want to get tail bounds for  $f(X)$ .
  - ▶ Step 1: Choose a metric  $d$  so that  $f$  is Lipschitz w.r.t.  $d$ .
  - ▶ Step 2: Find a Markov chain whose stationary distribution is the law of  $X$ , and which takes small steps w.r.t.  $d$ .
  - ▶ Step 3: Under some further conditions, compute concentration inequalities for  $f(X)$  using the rate of convergence of the Markov chain.
- ▶ For Haar measures, take random walks that are “constant on conjugacy classes”, like the random reflections walk.
- ▶ Applications to random matrices, random permutations, etc.
- ▶ Special thanks to: Persi Diaconis (my thesis advisor) and Yuval Peres (my unofficial postdoc mentor).