MATRIX CHERNOFF INEQUALITY AND CLASSICAL & QUANTUM CORRELATIONS

Stephen Diadamo

October 6, 2017

Seminar: Applications of Matrix Techniques in Information and Signal Theory

Overview of the talk

1. Matrix Chernoff Inequality
2. Quantum and Classical Correlations
3. Motivating Example
4. Total Bipartite Correlations

Classical Chernoff Bounds: Let $X_1, X_2, ..., X_n$ be random variables such that $a \leq X_k \leq b$ for each k. Let $Y := \sum_k X_k$ and define $\mu := \mathbb{E}Y$. Then for $\epsilon > 0$,

$$\mathbb{P}(Y \geq (1+\epsilon)\mu) \leq \exp\left(\frac{2\epsilon^2\mu^2}{n(b-a)^2}\right)$$
$$\mathbb{P}(Y \leq (1-\epsilon)\mu) \leq \exp\left(\frac{\epsilon^2\mu^2}{n(b-a)^2}\right)$$

Matrix Chernoff Inequality

- Let $X_1, X_2, ..., X_n$ be a finite set of independent, random, positive-semidefinite matrices of dimension d
- We have that for each k, $\lambda_{\min}(X_k) \geq 0$, and X_k is Hermitian
- Assume for each k, $\lambda_{\max}(X_k) \leq L$
- Define $Y := \sum_k X_k$

Matrix Chernoff Inequality

- Let $X_1, X_2, ..., X_n$ be a finite set of independent, random, positive-semidefinite matrices of dimension d
- We have that for each k, $\lambda_{min}(X_k) \geq 0$, and X_k is Hermitian
- Assume for each k, $\lambda_{max}(X_k) \leq L$
- Define $Y := \sum_k X_k$

Goal: Study the expectation and tail behaviour of $\lambda_{min}(Y)$ and $\lambda_{max}(Y)$

Matrix Chernoff Inequality

- Let $X_1, X_2, ..., X_n$ be a finite set of independent, random, positive-semidefinite matrices of dimension d
- We have that for each k, $\lambda_{min}(X_k) \geq 0$, and X_k is Hermitian
- Assume for each k, $\lambda_{max}(X_k) \leq L$
- Define $Y := \sum_k X_k$

Goal: Study the expectation and tail behaviour of $\lambda_{min}(Y)$ and $\lambda_{max}(Y)$

Why?

- $\|Y\| \geq \lambda_{max}(Y)$
- $\lambda_{min}(Y)$ tell us if Y is singular

Theorem: Consider $X_1, X_2, ..., X_n$ and Y defined previously, then for

- $\mu_{\min} := \lambda_{\min}(\mathbb{E}Y)$
- $\mu_{\max} := \lambda_{\max}(\mathbb{E}Y)$

It holds:

- $\mathbb{P}\{\lambda_{\min}(Y) \leq (1-\epsilon)\mu_{\min}\} \leq d\left[\frac{e^{-\epsilon}}{(1-\epsilon)^{1-\epsilon}}\right]^{\mu_{\min}/L}$ for $\epsilon \in [0, 1)$
- $\mathbb{P}\{\lambda_{\max}(Y) \geq (1+\epsilon)\mu_{\max}\} \leq d\left[\frac{e^{\epsilon}}{(1+\epsilon)^{1+\epsilon}}\right]^{\mu_{\max}/L}$ for $\epsilon \geq 0$.

Lemma: For $X \in \{X_k\}$ as above and $\theta \in \mathbb{R}$, $\theta > 0$, it holds:

- Mgf: $\Phi_X(\theta) := \mathbb{E}e^{\theta X} \leq \exp\left(\frac{e^{\theta L}-1}{L} \cdot \mathbb{E}X\right)$
- Cgf: $\Xi_X(\theta) := \log \mathbb{E}e^{\theta X} \leq \frac{e^{\theta L}-1}{L} \cdot \mathbb{E}X$

Remark: For $X_1, ..., X_N$ iid random Hermitian matrices of dimension d, such that $0 \leq X_i \leq \mathbb{I}$, $\lambda_{\min}(\mathbb{E}X_i) \geq \mu$, we have for $Y := \frac{1}{N}\sum X_i$, and $\epsilon \in [0, 1]$

$$\mathbb{P}(\lambda_{\min}(Y) \leq (1-\epsilon)\mu) \leq d \exp\left(-N\epsilon^2\mu/2\right)$$
$$\mathbb{P}(\lambda_{\max}(Y) \geq (1+\epsilon)\mu) \leq d \exp\left(-N\epsilon^2\mu/2\right)$$

Correlation: Given two systems, A and B, how much information about system A can be predicted given system B?

Correlation: Given two systems, A and B, how much information about system A can be predicted given system B?

Quantum and Classical Correlations:

- Classical Correlations: Correlations governed by laws of classical mechanics

Correlation: Given two systems, A and B, how much information about system A can be predicted given system B?

Quantum and Classical Correlations:

- Classical Correlations: Correlations governed by laws of classical mechanics
- Quantum Correlations: Correlations governed by laws of quantum mechanics

Correlation: Given two systems, A and B, how much information about system A can be predicted given system B?

Quantum and Classical Correlations:

- Classical Correlations: Correlations governed by laws of classical mechanics
- Quantum Correlations: Correlations governed by laws of quantum mechanics

Goal: Find an operational meaning for total correlation of a bipartite quantum system

For two sites A and B, we consider the amount of “work” required to erase all correlations between a maximally entangled state

For two sites A and B, we consider the amount of "work" required to erase all correlations between a maximally entangled state

- **Consider**: $|\Phi^+\rangle = \frac{1}{\sqrt{2}}\left(|0\rangle_A |0\rangle_B + |1\rangle_A |1\rangle_B\right)$

For two sites A and B, we consider the amount of “work” required to erase all correlations between a maximally entangled state

- **Consider**: $|\Phi^+\rangle = \frac{1}{\sqrt{2}}\left(|0\rangle_A |0\rangle_B + |1\rangle_A |1\rangle_B\right)$
- **Question**: What is the total correlation of this state?

For two sites A and B, we consider the amount of "work" required to erase all correlations between a maximally entangled state

- **Consider**: $|\Phi^+\rangle = \frac{1}{\sqrt{2}} \left(|0\rangle_A |0\rangle_B + |1\rangle_A |1\rangle_B \right)$
- **Question**: What is the total correlation of this state?
- **Claim**: This state contains two bits of correlation, 1 bit of quantum correlation (entanglement) and 1 bit of classical correlation

Step One: Erase the quantum correlation

Step One: Erase the quantum correlation

- From site A, 1 bit of randomness is applied, that is, $\mathbb{I}$ or σ_z is applied with equal probability
- $|\Phi^+\rangle \to \rho = \frac{1}{2}|\Phi^+\rangle\langle\Phi^+| + \frac{1}{2}|\Phi^-\rangle\langle\Phi^-|$

Step One: Erase the quantum correlation

- From site A, 1 bit of randomness is applied, that is, $\mathbb{I}$ or σ_z is applied with equal probability
- $|\Phi^+\rangle \to \rho = \frac{1}{2}|\Phi^+\rangle\langle\Phi^+| + \frac{1}{2}|\Phi^-\rangle\langle\Phi^-|$

Step Two: Erase the classical correlation

Step One: Erase the quantum correlation

- From site A, 1 bit of randomness is applied, that is, $\mathbb{I}$ or σ_z is applied with equal probability
- $|\Phi^+\rangle \to \rho = \frac{1}{2}|\Phi^+\rangle\langle\Phi^+| + \frac{1}{2}|\Phi^-\rangle\langle\Phi^-|$

Step Two: Erase the classical correlation

- From site A, another 1 bit of randomness is applied, that is, $\mathbb{I}$ or σ_x is applied with equal probability
- $\rho \to \frac{1}{2}\mathbb{I}_A \otimes \frac{1}{2}\mathbb{I}_B$

Theorem: The total correlation, measured by the amount of local erasure, for a bipartite state ρ_{AB} is the quantum mutual information

$$I(A:B) = S(\rho_A) + S(\rho_B) - S(\rho_{AB})$$

Let A and B be finite dimensional Hilbert spaces.

Definition: COLUR Maps
For $N \in \mathbb{N}$, let the ensemble $\{p_i, U_i \otimes V_i\}_{i=1}^{N}$ be such that for each i, $U_i \otimes V_i$ is a unitary on $A \otimes B$. A randomizing map $R : \mathcal{S}(A \otimes B) \to \mathcal{S}(A \otimes B)$,

$$R : \rho \mapsto \sum_{i=1}^{N} p_i (U_i \otimes V_i) \rho (U_i \otimes V_i)^\dagger$$

is called a coordinated local unitary randomizing (COLUR) map (COLUR) if it is CPTP. If for each i, $U_i = \mathbb{I}$, then R is called a B-LUR map. Similarly, if $V_i = \mathbb{I}$ for each i, then R is called A-LUR.

Definition: ϵ-decorrelates

A COLUR map R is said to ϵ-decorrelate a state $\rho \in \mathcal{S}(\mathsf{A} \otimes \mathsf{B})$ if there is a product state $\omega_\mathsf{A} \otimes \omega_\mathsf{B} \in \mathcal{S}(\mathsf{A} \otimes \mathsf{B})$ such that,

$$\|\mathsf{R}(\rho) - \omega_\mathsf{A} \otimes \omega_\mathsf{B}\|_1 \leq \epsilon$$

where $\| \cdot \|_1$ is the trace norm of an operator.

Definition: Entropy Exchange

For a COLUR map R and a purification of state $\rho \in \mathcal{S}(\mathsf{A} \otimes \mathsf{B})$, $\rho_{\mathsf{p}} = |\psi\rangle\langle\psi|_{\mathsf{ZAB}}$, with reference Hilbert space Z, we define the entropy exchange as,

$$\mathsf{S}_{\mathsf{e}}(\mathsf{R}, \rho_{\mathsf{A}}) := \mathsf{S}((\mathbb{I}_{\mathsf{Z}} \otimes \mathsf{R})\, |\psi\rangle\langle\psi|_{\mathsf{ZAB}})$$

Proposition 1: For any COLUR map R that $\epsilon-$decorrelates the state $\rho^{\otimes n} \in \mathcal{S}(A^{\otimes n} \otimes B^{\otimes n})$, the entropy exchange of R relative to $\rho^{\otimes n}$ has the lower bound

$$S_e(R, \rho^{\otimes n}) \geq n[I(A : B) - O(\epsilon)],$$

Proposition 1: For any COLUR map R that $\epsilon-$decorrelates the state $\rho^{\otimes n} \in \mathcal{S}(A^{\otimes n} \otimes B^{\otimes n})$, the entropy exchange of R relative to $\rho^{\otimes n}$ has the lower bound

$$S_e(\mathsf{R}, \rho^{\otimes n}) \geq n[I(A:B) - O(\epsilon)],$$

Proof Idea:

- Using the locality of R, separate $\rho^{\otimes n}$ into two parts $R_A := \mathrm{Tr}_B \mathsf{R}(\rho^{\otimes n})$ and $R_B = \mathrm{Tr}_A \mathsf{R}(\rho^{\otimes n})$
- Using concavity of von Neumann entropy and Fannes inequality, $S(R_A) + S(R_B) - S(\mathsf{R}(\rho^{\otimes n})) \leq O(\epsilon)$
- Introduce a purification of $\rho = \mathrm{Tr}_Z(|\psi\rangle\langle\psi|)$, $\psi := |\psi\rangle\langle\psi|$, and conclude using Araki-Lieb (triangle) inequality $S_e(\mathsf{R}, \rho^{\otimes n}) = S(\mathbb{I}_Z^{\otimes n} \otimes \mathsf{R})(\psi^{\otimes n}) \geq n[I(A:B) - O(\epsilon)]$

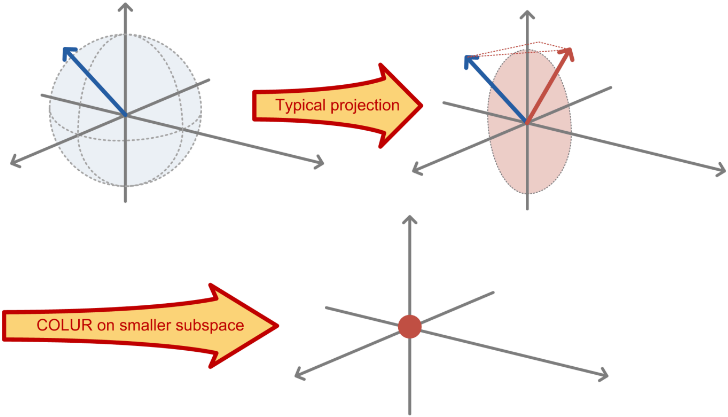
Proposition 2: For any state $\rho \in \mathcal{S}(A \otimes B)$ and $\epsilon > 0$, there exists for sufficiently large $n \in \mathbb{N}$, an A-LUR map,
$R : \mathcal{S}(A^{\otimes n} \otimes B^{\otimes n}) \to \mathcal{S}(A^{\otimes n} \otimes B^{\otimes n})$,

$$R : \rho^{\otimes n} \mapsto \frac{1}{N} \sum_{i=1}^{N} (U_i \otimes \mathbb{I}) \rho^{\otimes n} (U_i \otimes \mathbb{I})^{\dagger}$$

which ϵ-decorrelates $\rho^{\otimes n}$ such that

$$\log N \leq n[I(A : B) + O(\epsilon)]$$

Proof Idea



Theorem (restated more mathematically): The total correlations in a bipartite state ρ_{AB} measured in terms of how much noise needs to be added to turn ρ_{AB} into a product state is given by

$$\sup_{\epsilon>0} \liminf_{n\to\infty} \frac{1}{n} \min\{S_e(R, \rho^{\otimes n}) \,|R,\ \epsilon\text{-decorr } \rho^{\otimes n},\ \text{COLUR}\}$$

$$= \sup_{\epsilon>0} \limsup_{n\to\infty} \frac{1}{n} \min\{\log N \mid R,\ \epsilon\text{-decorr } \rho^{\otimes n},\ \text{A-LUR}\}$$

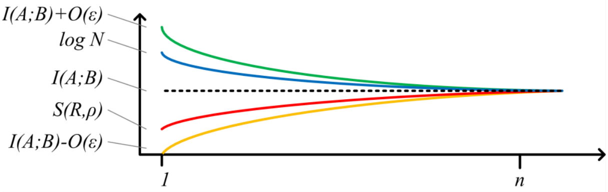
$$= I(A{:}B)$$

Theorem (restated more mathematically): The total correlations in a bipartite state ρ_{AB} measured in terms of how much noise needs to be added to turn ρ_{AB} into a product state is given by

$$\sup_{\epsilon>0} \liminf_{n\to\infty} \frac{1}{n} \min\{S_e(R, \rho^{\otimes n}) \mid R,\ \epsilon\text{-decorr } \rho^{\otimes n},\ \text{COLUR}\}$$

$$= \sup_{\epsilon>0} \limsup_{n\to\infty} \frac{1}{n} \min\{\log N \mid R,\ \epsilon\text{-decorr } \rho^{\otimes n},\ \text{A-LUR}\}$$

$$= I(A{:}B)$$



References

1. J. Tropp: An Introduction to Matrix Concentration Inequalities, Chapter 5
2. http://www.markwilde.com/teaching/qinfo-mcgill/presentations/2-JanFlorjanczyk.pdf, Jan Florjanczyk
3. Quantum, classical, and total amount of correlations in a quantum state, Groisman, Popescu, Winter