Matrix Chernoff Inequality and Classical & Quantum Correlations

Stephen Diadamo

October 9, 2017

Contents

1 Matrix Chernoff Inequality

1.1 Introduction

The Chernoff inequality gives us bounds on probabilities for the tail distributions for a sum of independent random variables. Mathematically, for $X_1, X_2, ..., X_n$ independent random scalar variables such that $a \leq X_k \leq b$ for each k and $Y := \sum_k X_k$ with $\mu := \mathbb{E}Y$, for $\epsilon > 0$,

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

What we will investigate in this section is how this can be extended to matrix random variables. This is interesting because we can for example, investigate when a sum of random matrices is singular or not, a probabilistic lower bound on the norm, and further we will see an example in the next section of how we can use the Matrix Chernoff inequality to construct randomizing functions.

1.2 The Matrix Chernoff Inequality

Theorem 1.1. *(Matrix Chernoff Inequalities) Consider a finite sequence $\mathbf{X}_1, ..., \mathbf{X}_N$ of independent, random, Hermitian matrices all with dimension $d \times d$. Assume that*

$$0 \leq \lambda_{min}(\mathbf{X}_k) \text{ and } \lambda_{max}(\mathbf{X}_k) \leq L, \text{ for each index } k.$$

Introduce the random matrix

$$\mathbf{Y} \coloneqq \sum_k \mathbf{X}_k.$$

Define the minimum eigenvalue μ_{min} and the maximum eigenvalue μ_{max} of the expectation $\mathbb{E}\mathbf{Y}$:

$$\mu_{min} \coloneqq \lambda_{min}(\mathbb{E}\mathbf{Y}) \text{ and } \mu_{max} \coloneqq \lambda_{max}(\mathbb{E}\mathbf{Y}). \tag{1}$$

Then,

$$\mathbb{P}\{\lambda_{min}(\mathbf{Y}) \leq (1-\epsilon)\mu_{min}\} \leq d\left[\frac{e^{-\epsilon}}{(1-\epsilon)^{1-\epsilon}}\right]^{\mu_{min}/L} \quad \text{for } \epsilon \in [0,1), \text{ and} \tag{2}$$

$$\mathbb{P}\{\lambda_{max}(\mathbf{Y}) \geq (1+\epsilon)\mu_{max}\} \leq d\left[\frac{e^{\epsilon}}{(1+\epsilon)^{1+\epsilon}}\right]^{\mu_{max}/L} \quad \text{for } \epsilon \geq 0. \tag{3}$$

To prove this theorem, we firstly introduce and prove a useful lemma.

Lemma 1.2. (Bounds for Moment and Cumulant Generating functions) Suppose that $\mathbf{X}$ is a random Hermitian matrix such that $0 \leq \lambda_{\min}(\mathbf{X})$ and $\lambda_{\max}(\mathbf{X}) \leq L$, $L \in \mathbb{R}$. Then it holds, for $\theta \in \mathbb{R}$,

$$\Phi_{\mathbf{X}}(\theta) \coloneqq \mathbb{E}e^{\theta\mathbf{X}} \leq \exp\left(\frac{e^{\theta L}-1}{L} \cdot \mathbb{E}\mathbf{X}\right)$$

and

$$\Xi_{\mathbf{X}}(\theta) \coloneqq \log \mathbb{E}e^{\theta\mathbf{X}} \leq \frac{e^{\theta L}-1}{L} \cdot \mathbb{E}\mathbf{X}$$

Proof. Consider the function $f(x) \coloneqq e^{\theta x}$, which is convex on $[0, L], L \in \mathbb{R}$. We therefore have for $x \in [0, L]$,

$$f(x) \leq f(0) + \frac{f(L)-f(0)}{L} \cdot x,$$

and in particular, for $\theta \in \mathbb{R}$,

$$e^{\theta x} \leq 1 + \frac{e^{\theta L}-1}{L} \cdot x.$$

Let $\mathbf{X}$ be a random Hermitian matrix such that $0 \leq \lambda_{\min}(\mathbf{X})$ and $\lambda_{\max}(\mathbf{X}) \leq L$. We make use of the fact that for a Hermitian matrix $\mathbf{A}$ with eigenvalues contained in an interval $I \subseteq \mathbb{R}$, and two real valued functions f, g defined on I, such that for all $x \in I$, $f(x) \leq g(x)$, $f(\mathbf{A}) \leq g(\mathbf{A})$. With this, we have,

$$e^{\theta \mathbf{X}} \leq \mathbb{I} + \frac{e^{\theta L} - 1}{L} \cdot \mathbf{X}.$$

It follows from Proposition 1.26 of the lecture script that expectation value respects semi-definite order, and so,

$$\mathbb{E} e^{\theta \mathbf{X}} \leq \mathbb{I} + \frac{e^{\theta L} - 1}{L} \cdot \mathbb{E}\mathbf{X}.$$

For functions $g(x) = 1 + x$ and $h(x) = e^x$, we use the same argument to find

$$\mathbb{I} + \frac{e^{\theta L} - 1}{L} \cdot \mathbb{E}\mathbf{X} \leq \exp\left(\frac{e^{\theta L} - 1}{L} \cdot \mathbb{E}\mathbf{X}\right).$$

We conclude with,

$$\mathbb{E} e^{\theta \mathbf{X}} \leq \exp\left(\frac{e^{\theta L} - 1}{L} \cdot \mathbb{E}\mathbf{X}\right),$$

which is the first inequality of the lemma. For the second inequality, we note that logarithm is matrix monotone increasing, and therefore preserves semidefinite order, that is,

$$\log \mathbb{E} e^{\theta \mathbf{X}} \leq \frac{e^{\theta L} - 1}{L} \cdot \mathbb{E}\mathbf{X}.$$

□

Proof. (Proof of Matrix Chernoff Bound) We prove each bound separately. For the finite sequence of independent, random, and Hermitian matrices with common dimension d, $\{\mathbf{X}_k\}_{k \geq 0}^N$, define $\mathbf{Y} \coloneqq \sum_k \mathbf{X}_k$ such that

$$0 \leq \lambda_{\min}(\mathbf{X}_k) \text{ and } \lambda_{\max}(\mathbf{X}_k) \leq L, \text{ for each index } k.$$

Define $g(\theta) \coloneqq (e^{\theta L} - 1)/L$. Starting from the Master Bound, we have for $t \in \mathbb{R}$,

$$\begin{aligned}
\mathbb{P}\{\lambda_{\max}(\mathbf{Y}) \geq t\} &\leq \inf_{\theta > 0} e^{-\theta t} \operatorname{tr} \exp\left(\sum_k \log \mathbb{E} e^{\theta \mathbf{X}_k}\right) \\
&\leq \inf_{\theta > 0} e^{-\theta t} \operatorname{tr} \exp\left(g(\theta) \cdot \sum_k \mathbb{E}\mathbf{X}_k\right) \\
&\leq \inf_{\theta > 0} e^{-\theta t} \left[d \cdot \lambda_{\max}(\exp\left(g(\theta) \cdot \mathbb{E}\mathbf{Y}\right))\right] \\
&= \inf_{\theta > 0} e^{-\theta t} \left[d \cdot \exp(g(\theta) \cdot \lambda_{\max}\left(\mathbb{E}\mathbf{Y}\right))\right] \\
&= \inf_{\theta > 0} e^{-\theta t} \left[d \cdot \exp(g(\theta) \cdot \mu_{\max})\right].
\end{aligned}$$

To complete the proof, we make a change of variables such that $t \mapsto (1 + \epsilon)\mu_{\max}, \epsilon \geq 0$, and note that the infimum is achieved at $\theta = L^{-1} \log(1 + \epsilon)$.

For the second bound, we again start from the Master Bound.

$$\begin{aligned}
\mathbb{P}\{\lambda_{\min}(\mathbf{Y}) \leq t\} &\leq \inf_{\theta < 0} e^{-\theta t} \operatorname{tr} \exp\left(\sum_k \log \mathbb{E} e^{\theta \mathbf{X}_k}\right) \\
&\leq \inf_{\theta < 0} e^{-\theta t} \left[d \cdot \lambda_{\max}(\exp\left(g(\theta) \cdot \mathbb{E}\mathbf{Y}\right))\right] \\
&= \inf_{\theta < 0} e^{-\theta t} \left[d \cdot \exp\left(\lambda_{\max}(g(\theta) \cdot \mathbb{E}\mathbf{Y})\right)\right] \\
&= \inf_{\theta < 0} e^{-\theta t} \left[d \cdot \exp\left(g(\theta) \cdot \lambda_{\min}(\mathbb{E}\mathbf{Y})\right)\right] \\
&= \inf_{\theta < 0} e^{-\theta t} \left[d \cdot \exp\left(g(\theta) \cdot \mu_{\min}\right)\right]
\end{aligned}$$

Again, we make a change of variables, such that $t \mapsto (1 - \epsilon)\mu_{\min}$ and the infimum will be found at $\theta = L^{-1} \log(1 - \epsilon)$. □

Remark 1.3. *The matrix Chernoff inequality can be slightly weakened (and also slightly reformulated) to give a clearer meaning to the theorem. For $\mathbf{X}_1, ..., \mathbf{X}_N$ i.i.d. random Hermitian matrices of dimension d, such that $0 \leq \mathbf{X}_i \leq \mathbb{I}$, $\lambda_{min}(\mathbb{E}\mathbf{X}_i) \geq \mu$, we have for $\mathbf{Y} \coloneqq \frac{1}{N}\sum \mathbf{X}_i$, with $\mu_{min} \coloneqq \lambda_{min}(\mathbb{E}\mathbf{Y}), \mu_{max} \coloneqq \lambda_{max}(\mathbb{E}\mathbf{Y})$, $\epsilon \in [0, 1)$*

$$\mathbb{P}(\lambda_{min}(\mathbf{Y}) \leq (1-\epsilon)\mu_{min}) \leq d \exp\left(-N\epsilon^2\mu/3\right)$$
$$\mathbb{P}(\lambda_{max}(\mathbf{Y}) \geq (1+\epsilon)\mu_{max}) \leq d \exp\left(-N\epsilon^2\mu/3\right)$$

Proof. First we rewrite the matrix Chernoff bound for $\mathbf{Y}$ averaged, i.e. $\mathbf{Y} \coloneqq \frac{1}{N}\sum_{i=1}^{N} \mathbf{X}_i$. With the assumptions made on each $\mathbf{X}_i$ in the statement, we have that

$$\begin{aligned}\lambda_{\min}(\mathbb{E}\mathbf{Y}) &= \lambda_{\min}\left(\mathbb{E}\left(\frac{1}{N}\sum_{i=1}^{N}\mathbf{X}_i\right)\right) \\ &= \frac{1}{N}\sum_{i=1}^{N}\lambda_{\min}(\mathbb{E}\mathbf{X}_i).\end{aligned}$$

Similarly,

$$\lambda_{\max}(\mathbb{E}\mathbf{Y}) = \frac{1}{N}\sum_{i=1}^{N}\lambda_{\max}(\mathbb{E}\mathbf{X}_i),$$

where linearity of the expectation value and spectral properties (i.e. commuting the factor $1/N$ with $\lambda_{\min}$ and $\lambda_{\max}$) are used. Defining $\mu_{\min} \coloneqq N\lambda_{\min}(\mathbb{E}\mathbf{Y})$ and $\mu_{\max} := N\lambda_{\max}(\mathbb{E}\mathbf{Y})$, the Chernoff inequality holds as stated previously.

We now show that the bounds in the remark follows from the Chernoff inequality. Let $\mathbf{X}_1, ..., \mathbf{X}_N$ be independent random Hermitian matrices of dimension d such that for each i, $0 \leq \mathbf{X}_i \leq \mathbb{I}$ and $\lambda_{\min}(\mathbb{E}\mathbf{X}_i) \geq \mu$. Note that for each i, $\lambda_{\max}(\mathbf{X}_i) \leq 1$ (i.e. $L = 1$). Define $\mathbf{Y} \coloneqq \frac{1}{N}\sum_{i=1}^{N}\mathbf{X}_i$. It holds that for $\mu_{\min} \coloneqq \lambda_{min}(\mathbb{E}\mathbf{Y})$,

$$\begin{aligned}\mu_{\min} &= \lambda_{\min}\left(\mathbb{E}\left(\frac{1}{N}\sum_{i=1}^{N}\mathbf{X}_i\right)\right) \\ &= \frac{1}{N}\sum_{i=1}^{N}\lambda_{\min}\left(\mathbb{E}\mathbf{X}_i\right) \\ &\geq \mu\end{aligned}$$

Now, for $\epsilon \in [0, 1)$, we have from the Chernoff inequality that,

$$\mathbb{P}(\lambda_{min}(\mathbf{Y}) \leq (1-\epsilon)\mu_{\min}) \leq d\left[\frac{e^{-\epsilon}}{(1-\epsilon)^{1-\epsilon}}\right]^{N\mu_{\min}}.$$

Ignoring the factor d,

$$\begin{aligned}\left[\frac{e^{-\epsilon}}{(1-\epsilon)^{1-\epsilon}}\right]^{N\mu_{\min}} &= \frac{e^{-\epsilon N\mu_{\min}}}{(1-\epsilon)^{N\mu_{\min}(1-\epsilon)}} \\ &= \frac{e^{-\epsilon N\mu_{\min}}}{e^{\ln\left((1-\epsilon)^{N\mu_{\min}(1-\epsilon)}\right)}} \\ &= \exp\Big(-\epsilon N\mu_{\min} - \ln\Big((1-\epsilon)^{N\mu_{\min}(1-\epsilon)}\Big)\Big) \\ &= \exp(-\epsilon N\mu_{\min} - N\mu_{\min}(1-\epsilon)\ln(1-\epsilon)) \\ &= \exp(-N\mu_{\min}(\epsilon + (1-\epsilon)\ln(1-\epsilon))) \\ &\leq \exp\left(-N\mu_{\min}(\epsilon^2/3)\right) \\ &\leq \exp\left(-N\mu\epsilon^2/3\right)\end{aligned}$$

Where for $\epsilon \in [0, 1), \epsilon + (1-\epsilon)\ln(1-\epsilon) \geq \epsilon^2/3$ and $\mu_{\min} \geq \mu$ are used. The result is that,

$$\begin{aligned}\mathbb{P}(\lambda_{min}(\mathbf{Y}) \leq (1-\epsilon)\mu_{\min}) &\leq d\left[\frac{e^{-\epsilon}}{(1-\epsilon)^{1-\epsilon}}\right]^{N\mu_{\min}} \\ &\leq d\exp\left(-N\mu\epsilon^2/3\right)\end{aligned}$$

Similarly, from the Chernoff inequality,

$$\mathbb{P}(\lambda_{\max}(\mathbf{Y}) \geq (1+\epsilon)\mu_{\max}) \leq d\left[\frac{e^{\epsilon}}{(1+\epsilon)^{1+\epsilon}}\right]^{N\mu_{\max}}.$$

Again, ignoring the factor of d for now,

$$\begin{aligned}\left[\frac{e^{\epsilon}}{(1+\epsilon)^{1+\epsilon}}\right]^{N\mu_{\max}} &= \frac{e^{\epsilon N\mu_{\max}}}{e^{\ln\left((1+\epsilon)^{N\mu_{\max}(1+\epsilon)}\right)}} \\ &= \exp(-N\mu_{\max}(-\epsilon + (1+\epsilon)\ln(1+\epsilon))) \\ &\leq \exp\left(-N\mu_{\max}(\epsilon^2/3)\right) \\ &\leq \exp\left(-N\mu\epsilon^2/3\right)\end{aligned}$$

where for $\epsilon \in [0,1)$, $-\epsilon + (1+\epsilon)\ln(1+\epsilon) \geq \epsilon^2/3$ and $\mu_{\max} \geq \mu_{\min} \geq \mu$ so,

$$\begin{aligned}\mathbb{P}(\lambda_{\max}(\mathbf{Y}) \geq (1+\epsilon)\mu_{\max}) &\leq d\left[\frac{e^{\epsilon}}{(1+\epsilon)^{1+\epsilon}}\right]^{N\mu_{\max}} \\ &\leq d\exp\left(-N\mu\epsilon^2/3\right)\end{aligned}$$

□

2 Chernoff Bounds regarding Quantum and Classical Correlations in a Quantum State

2.1 Introduction

The aim of this section is to give an operational meaning to the total amount of correlation in a bipartite quantum state. We will firstly motivate this by taking a maximally entangled state, namely a Bell state, and noticing that the amount of noise one must add to remove all correlation in the system is larger than the amount of correlations one would expect the state to have. From this example, we will see that using the amount of noise one should add to a system to remove all correlation can be used as a measure of the correlation. This idea can be generalized and we introduce and prove a theorem which shows this. In the process, we will see how using the Chernoff Inequality allows us to construct a randomizing map which will remove correlations in (many copies) of an arbitrary state with arbitrary precision.

2.2 Explanatory Example

Consider a maximally entangled, two qubit, quantum state,

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

This state contains one bit of entanglement, this is, it can be used to send one bit of information via entanglement assistance, quantum teleportation for example, and it is therefore tempting to think that it also has one bit of correlation. This example will demonstrate there we can consider this state to have two bits of correlation, one of entanglement and one of secret classical correlation.

Consider two parties, Alice and Bob, who share one qubit of $|\Psi\rangle$ each. Suppose Alice wants to erase the entanglement between her and Bob's qubit. To do so, she applies at random, with equal probability, one of σ_z, the Pauli-Z operator, or $\mathbb{I}$. After doing this, the shared state becomes a mixture

$$\rho = \frac{1}{2}\left|\Psi^+\right\rangle\!\left\langle\Psi^+\right| + \frac{1}{2}\left|\Phi^-\right\rangle\!\left\langle\Phi^-\right|$$

where

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

but no longer contains entanglement, since we can write ρ as a product state,

$$\rho = \frac{1}{2}\left|0\right\rangle\!\left\langle 0\right|_A \otimes \left|0\right\rangle\!\left\langle 0\right|_B + \frac{1}{2}\left|1\right\rangle\!\left\langle 1\right|_A \otimes \left|1\right\rangle\!\left\langle 1\right| 1_B.$$

Although the entanglement has been removed, ρ still contains one bit of classical correlation. If Alice now wishes to erase this classical correlation, she can do so by randomly applying a bit flip operation σ_x or $\mathbb{I}$ with equal probability. The resulting state becomes,

$$\rho' = \frac{1}{2}\mathbb{I}_A \otimes \frac{1}{2}\mathbb{I}_B,$$

and so the qubits are completely independent of each other, and the correlations have been removed.

We can see that it was required to apply two bits of erasure in order to remove both quantum and classical correlation from the maximally entangled state $|\Phi^+\rangle$. With this, it can be said that $|\Phi^+\rangle$ contained two bits of correlation.

2.3 Definitions

We define some useful maps and terminology. Let A and B be Hilbert spaces of dimensions d_A and d_B respectively, such that $d_A, d_B < \infty$. Let $\rho \coloneqq \rho_{AB}$ be a bipartite state in $\mathcal{S}(A \otimes B)$.

Definition 2.1. (COLUR Maps) Let $\{\mathbf{U}_i \otimes \mathbf{V}_i\}_{i=1}^N, N \in \mathbb{N}$, where $\mathbf{U}_i \otimes \mathbf{V}_i$ is a local unitary map on $A \otimes B$ for all $i \in \{0, ..., N\}$. With values $\{p_i\}_{i=1}^N$ such that for all $i \in \{0, ..., N\}, p_i \leq 1$ and $\sum_i p_i = 1$, we use the ensemble $\{p_i, \mathbf{U}_i \otimes \mathbf{V}_i\}_{i=1}^N$, to construct a randomizing map $\mathbf{R} : \mathcal{S}(A \otimes B) \to \mathcal{S}(A \otimes B)$,

$$\mathbf{R} : \rho \mapsto \sum_{i=1}^{N} p_i (\mathbf{U}_i \otimes \mathbf{V}_i)\rho(\mathbf{U}_i \otimes \mathbf{V}_i)^\dagger.$$

Such maps that are completely positive and trace preserving (CPTP) on $A \otimes B$ are called coordinated local unitary randomizing (COLUR) maps. We use the notation $|\mathbf{R}| = \#\{p_i, \mathbf{U}_i \otimes \mathbf{V}_i\}_{i=1}^N = N$ to denote the size of the ensemble for $\mathbf{R}$.

A COLUR map, such that for all $i \in \{0, ..., N\}, \mathbf{V}_i = \mathbb{I}$, is called an A-LUR map. Similarly, a COLUR map with $\mathbf{U}_i = \mathbb{I}$ for all $i \in \{0, ..., N\}$ is called a B-LUR Map. COLUR maps which are a combination of an A-LUR followed by a B-LUR (or vice versa) are simply called LUR maps.

Definition 2.2. (ϵ-decorrelates) We say that a COLUR map $\mathbf{R}$ ϵ-decorrelates a state $\rho \in \mathcal{S}(A \otimes B)$ if there is a product state $\omega_A \otimes \omega_B \in \mathcal{S}(A \otimes B)$ such that,

$$\|\mathbf{R}(\rho) - \omega_A \otimes \omega_B\|_1 \leq \epsilon \tag{4}$$

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

Definition 2.3. (Entropy Exchange) For a CPTP map $\mathbf{R}$ acting on the Hilbert space A, and a purification of state $\rho_A \in \mathcal{S}(A)$, $\rho_{ZA} = |\psi\rangle\langle\psi|$, pure in the Hilbert space $Z \otimes A$, with Z the reference system, we define the entropy exchange as,

$$S_e(\mathbf{R}, \rho_A) \coloneqq S((\mathbb{I}_Z \otimes \mathbf{R})\, |\psi\rangle\langle\psi|). \tag{5}$$

2.4 Total Bipartite Correlations

Theorem 2.4. *The total correlation in a bipartite state ρ_{AB}, as measured by the asymptotically minimal amount of local noise one has to add to turn it into a product state, is I(A:B) = $S(\rho_A)+S(\rho_B)-S(\rho_{AB})$. More formally,*

$$\sup_{\epsilon>0} \liminf_{n\to\infty} \frac{1}{n} \min\{S_e(\mathbf{R}, \rho^{\otimes n}) \mid \mathbf{R} \in COL(A \otimes B) \text{ and } \epsilon\text{-decorrelates } \rho^{\otimes n}\} =$$

$$\sup_{\epsilon>0} \limsup_{n\to\infty} \frac{1}{n} \min\{\log N \mid \mathbf{R} \in A\text{-}LUR(A \otimes B), |R| = N, \text{ and } \epsilon\text{-decorrelates } \rho^{\otimes n}\} = I(A{:}B)$$

Proposition 2.5. *For finite dimension Hilbert spaces A and B and large enough $n \in \mathbb{N}$, consider a COLUR map, $\mathbf{R} : \mathcal{S}(A^{\otimes n} \otimes B^{\otimes n}) \to \mathcal{S}(A^{\otimes n} \otimes B^{\otimes n})$,*

$$\mathbf{R} : \rho^{\otimes n} \mapsto \sum_{i=1}^{N} p_i (\mathbf{U}_i \otimes \mathbf{V}_i) \rho^{\otimes n} (\mathbf{U}_i \otimes \mathbf{V}_i)^\dagger, \tag{6}$$

which $\epsilon-$decorrelates the state $\rho^{\otimes n} \in \mathcal{S}(A^{\otimes n} \otimes B^{\otimes n})$. Then the entropy exchange of $\mathbf{R}$ relative to $\rho^{\otimes n}$ is lower bounded by

$$S_e(\mathbf{R}, \rho^{\otimes n}) \geq n[I(A:B) - O(\epsilon)]. \tag{7}$$

Proof. Since $\mathbf{R}$ acts locally on each of $A^{\otimes n}$ and $B^{\otimes n}$, we can define $R_A \coloneqq \mathrm{Tr}_B \mathbf{R}(\rho^{\otimes n}) = \sum_{i=1}^{|R|} p_i \mathbf{U}_i \rho_A^{\otimes n} \mathbf{U}_i^\dagger$ and similarly $R_B \coloneqq \mathrm{Tr}_A \mathbf{R}(\rho^{\otimes n}) = \sum_{i=1}^{|R|} p_i \mathbf{V}_i \rho_B^{\otimes n} \mathbf{V}_i^\dagger$. By the concavity of von Neumann entropy,

$$S(R_A) \geq nS(\rho_A), \text{ and } S(R_B) \geq nS(\rho_B) \tag{8}$$

Since $\mathbf{R}$ is assumed to ϵ-decorrelate $\rho^{\otimes n}$, we have that there is a product state $\omega_A \otimes \omega_B$ with

$$\|\mathbf{R}(\rho^{\otimes n}) - \omega_A \otimes \omega_B\|_1 \leq \epsilon$$

and therefore,

$$\|R_A - \omega_A\|_1 \leq \|\mathbf{R}(\rho^{\otimes n}) - \omega_A \otimes \omega_B\|_1 \leq \epsilon$$

and by the same argument,

$$\|R_B - \omega_B\|_1 \leq \epsilon.$$

By the triangle inequality,

$$\|R_A \otimes R_B - \omega_A \otimes \omega_B\|_1 \leq 2\epsilon \Rightarrow \|\mathbf{R}(\rho^{\otimes n}) - R_A \otimes R_B\|_1 \leq 3\epsilon.$$

Using Fannes inequality, we get,

$$S(R_A) + S(R_B) - S(\mathbf{R}(\rho^{\otimes n})) \leq O(\epsilon)$$

Using (8), we obtain,

$$S(\mathbf{R}(\rho^{\otimes n})) \geq n[S(\rho_A) + S(\rho_B) - O(\epsilon)].$$

Introducing a purified state $\psi \coloneqq |\psi\rangle\langle\psi|$ with the reference system Z such that $\rho = \mathrm{Tr}_Z \psi$, where ψ is pure in $Z \otimes A \otimes B$. Using the fact that $\mathbf{R}$ is a map on $A^{\otimes n} \otimes B^{\otimes n}$, we construct $\Omega \coloneqq (\mathbb{I}_Z^{\otimes n} \otimes \mathbf{R})(\psi^{\otimes n})$ on $Z^{\otimes n} \otimes A^{\otimes n} \otimes B^{\otimes n}$. By definition of entropy exchange and with the use of Araki-Lieb inequality,

$$\begin{aligned} S_e(\mathbf{R}, \rho^{\otimes n}) &= S(\Omega) \\ &\geq S(\mathrm{Tr}_{Z^{\otimes n}} \Omega) - S(\mathrm{Tr}_{A^{\otimes n} \otimes B^{\otimes n}} \Omega) \\ &= S(\mathbf{R}(\rho^{\otimes n})) - S(\rho^{\otimes n}) \\ &\geq n[S(\rho_A) + S(\rho_B) - S(\rho) - O(\epsilon)] \\ &= n[I(A:B) - O(\epsilon)] \end{aligned}$$

□

Proposition 2.6. *For any state ρ on $A \otimes B$ and $\epsilon \in (0,1)$, there exists, for all sufficiently large $n \in \mathbb{N}$, an A-LUR map, $\mathbf{R} : \mathcal{S}(A^{\otimes n} \otimes B^{\otimes n}) \to \mathcal{S}(A^{\otimes n} \otimes B^{\otimes n})$,*

$$\mathbf{R} : \rho^{\otimes n} \mapsto \frac{1}{N} \sum_{i=1}^{N} (\mathbf{U}_i \otimes \mathbb{I}) \rho^{\otimes n} (\mathbf{U}_i \otimes \mathbb{I})^\dagger \tag{9}$$

which ϵ-decorrelates $\rho^{\otimes n}$, and

$$\log N \leq n[I(A:B) + O(\epsilon)] \tag{10}$$

Proof. For a large enough n, we can apply the theory of typicality. We restrict the state $\rho^{\otimes n}$ to its typical subspace of dimension D via a projection $\mathbf{\Pi}$, and further restrict the bipartite state to the local typical subspaces of $\rho_A^{\otimes n}$ and $\rho_B^{\otimes n}$ via projections $\mathbf{\Pi}_A$ and $\mathbf{\Pi}_B$. We define

$$\hat{\rho} \coloneqq (\mathbf{\Pi}_A \otimes \mathbf{\Pi}_B) \mathbf{\Pi} \rho^{\otimes n} \mathbf{\Pi} (\mathbf{\Pi}_A \otimes \mathbf{\Pi}_B)$$

and by definition of typical subspace projectors and using the "gentle measurement" lemma we have with this construction,

$$\|\hat{\rho} - \rho^{\otimes n}\|_1 \leq 5\sqrt{\epsilon}. \tag{11}$$

We make further restrictions to $\hat{\rho}$ which will be useful for the construction of $\mathbf{R}$. We define an ϵ-typical projector $\mathbf{\Pi}'_B$ on the subspace where $\mathrm{Tr}_A(\hat{\rho}) \geq \epsilon / D_B$, where D_B is the dimension of the restricted typical subspace for $\hat{\rho}$ on B, and define

$$\tilde{\rho} \coloneqq (\mathbb{I} \otimes \mathbf{\Pi}'_B) \hat{\rho} (\mathbb{I} \otimes \mathbf{\Pi}'_B)$$

and it follows from (11) and using gentle measurement again that

$$\|\tilde{\rho} - \rho^{\otimes n}\|_1 \leq 8\sqrt{\epsilon}.$$

Note that we have defined the projections up to now such that

$$\omega'_B \coloneqq \mathrm{Tr}_A \tilde{\rho} \geq \frac{\epsilon}{D_B} \mathbf{\Pi}'_B.$$

Take any ensemble of unitaries $\mathbf{U} \coloneqq \{p(d\mathbf{U}_i), \mathbf{U}_i\}$, such that for any state $\sigma \in \mathcal{S}(\mathbf{\Pi}_A \rho_A^{\otimes n})$, that is, a state in the typical subspace of $\rho_A^{\otimes n}$, where it holds that,

$$\int_{\mathbf{U}} p(d\mathbf{U}_i) \mathbf{U}_i \sigma \mathbf{U}_i^\dagger = \frac{1}{D_A} \mathbf{\Pi}_A =: \omega_A,$$

with D_A the dimension of the typical subspace of $\rho_A^{\otimes n}$. It is clear that,

$$\int_{\mathbf{U}} p(d\mathbf{U}_i) (\mathbf{U}_i \otimes \mathbb{I}) \tilde{\rho} (\mathbf{U}_i^\dagger \otimes \mathbb{I}) = \omega_A \otimes \omega'_B.$$

We define a random variable from $\mathbf{U}$ as follows

$$\mathbf{X} \coloneqq D(\mathbf{U} \otimes \mathbb{I})\tilde{\rho}(\mathbf{U}^\dagger \otimes \mathbb{I}),$$

then,

$$\mathbb{E}\mathbf{X} = D\omega_A \otimes \omega'_B \geq \epsilon 2^{-n[I(A:B)+O(\epsilon)]}\mathbf{\Pi}_A \otimes \mathbf{\Pi}'_B.$$

Take $\mathbf{X}_1, ..., \mathbf{X}_N$ as i.i.d. realizations of $\mathbf{X}$, then for each realization $\mathbf{X}_i$, by construction, $0 \leq \mathbf{X}_i \leq \mathbb{I}_{A^{\otimes n} \otimes B^{\otimes n}}$. Define $\mathbf{Y} \coloneqq \frac{1}{N}\sum \mathbf{X}_i$, and by the weaker version of the matrix Chernoff inequality, we have for $\mu \coloneqq \epsilon 2^{-n[I(A:B)+O(\epsilon)]}$

$$\mathbb{P}((1-\epsilon)\mu_{\min} \not\leq \lambda_{\min}(\mathbf{Y}) \leq \lambda_{\max}(\mathbf{Y}) \not\leq (1+\epsilon)\mu_{\max}) \leq 2d_A^n d_B^n \exp\bigl(-N\epsilon^2\mu/3\bigr)$$

With $N = 2^{n[I(A:B)+O(\epsilon)]}$, this bound can be made less than 1, which implies that there exists a finite set of unitaries $\mathbf{U}_1, ..., \mathbf{U}_N$ such that

$$(1-\epsilon)\omega_A \otimes \omega'_B \leq \frac{1}{N}\sum_{i=1}^{N}(\mathbf{U}_i \otimes \mathbb{I})\tilde{p}(\mathbf{U}_i \otimes \mathbb{I}) \leq (1+\epsilon)\omega_A \otimes \omega'_B$$

Further,

$$\left\| \frac{1}{N}\sum_{i=1}^{N}(\mathbf{U}_i \otimes \mathbb{I})\rho^{\otimes n}(\mathbf{U}_i \otimes \mathbb{I})^\dagger - \omega_A \otimes \omega'_B \right\|_1 \leq O(\epsilon).$$

By defining the state $\omega_B \coloneqq \omega'_B/\mathrm{Tr}(\omega'_B)$,

$$\left\| \frac{1}{N}\sum_{i=1}^{N}(\mathbf{U}_i \otimes \mathbb{I})\rho^{\otimes n}(\mathbf{U}_i \otimes \mathbb{I})^\dagger - \omega_A \otimes \omega_B \right\|_1 \leq O(\epsilon)$$

which implies there is an A-LUR map $\mathbf{R}$ which ϵ-decorrelates $\rho^{\otimes n}$. With this choice of N, we also have that $\log N \leq n[I(A:B) + O(\epsilon)]$. □

By applying the two propositions, we immediately get the theorem statement. In summary, we have shown that for a bipartite state $\rho \in \mathcal{S}(A \otimes B)$, the amount of noise needed to erase all correlations from the system is given by the mutual information $I(A:B)$.

References

[1] J. Tropp: *An Introduction to Matrix Concentration Inequalities*, Chapter 5

[2] B. Groisman et al. *Quantum, classical, and total amount of correlations in a quantum state*, Phys. Rev. A 72, 032317 (2005)

[3] Presentation: http://www.markwilde.com/teaching/qinfo-mcgill/presentations/2-JanFlorjanczyk.pdf, Jan Florjanczyk