Poincaré and Sobolev Inequalities for Matrix Φ-entropies, and an upper bound to Holevo Quantity

Beyond I.I.D. in Information Theory

Min-Hsiu Hsieh

QUANTUM COMPUTATION & INTELLIGENT SYSTEMS

Joint work with

Hao-Chung Cheng	National Taiwan University & QCIS, UTS

arXiv:1506.06801

Outlines

1. Characterizations of Matrix Φ-entropies
2. Functional Inequalities of Matrix Φ-entropies
3. Relationship with Holevo Quantity

Classical Φ-entropies

Classical Φ-Function Set $\mathcal{C}$

The class $\mathcal{C}$ contains each function $\Phi : [0,\infty) \to \mathbb{R}$ that is either affine or

- continuous and convex on $[0,\infty)$;
- twice differentiable on $(0,\infty)$;
- $\Phi'' > 0$ and $1/\Phi''$ is concave.

[1] Latała and Oleszkiewicz (2000), *Geometric Aspects of Functional Analysis*.

Definition (Classical Φ-entropies)

The classical Φ-entropy $H_\Phi(Z)$ for $\Phi \in \mathcal{C}$ and non-negative random variable Z is defined as:

$$H_\Phi(Z) = \mathbb{E}\Phi(Z) - \Phi(\mathbb{E}Z)$$

so that $\mathbb{E}|Z| < \infty$ and $\mathbb{E}|\Phi(Z)| < \infty$.

Example

Variance: $\Phi(x) = x^2$.
Entropy: $\Phi(x) = x \log x$.
f-divergence: $D_\Phi(\nu\|\mu) = H_\Phi(\frac{d\nu}{d\mu}) - \Phi(1)$.

Subadditivity

When $Z = f(X_1, \cdots, X_n)$, $X_1, \cdots, X_n$ are independent r.v.s, we say $H_\Phi(Z)$ is *subadditive*[2] if

$$H_\Phi(Z) \le \sum_{i=1}^{n} \mathbb{E} H_\Phi^{(i)}(Z),$$

where $H_\Phi^{(i)}(Z) = \mathbb{E}_i\Phi(Z) - \Phi(\mathbb{E}_i Z)$ is the conditional Φ-entropy, and $\mathbb{E}_i$ denotes conditional expectation conditioned on $X_{-i} \triangleq (X_1 \cdots, X_{i-1}, X_{i+1}, \cdots, X_n)$.

[2] Gross(1975). *American Journal of Mathematics*.

Applications

- $\star$ Poincaré: $\mathrm{Var}(f(X)) \le \mathbb{E}\left[\|\nabla f(X)\|^2\right]$.
- $\star$ Log Sobolev: $H_\Phi(f^2) \le C \cdot \mathcal{E}(f)$.
- $\star$ Bousquet[3]: Let $Z = \sup_{f\in\mathcal{F}} \sum_{i=1}^{n} f(X_i)$

$$\Pr\{Z \ge \mathbb{E}Z + t\} \le e^{-vh(t/v)}.$$

[3] Bousquet(2002), *Comptes Rendus Mathematique*.

Characterizations of Classical Φ-entropies

Theorem

The following statements are equivalent[1].

- **(a)** $\Phi \in \mathcal{C}$: Φ *is affine or* $\Phi'' > 0$ *and* $1/\Phi''$ *is concave;*
- **(b)** *Brègman divergence* $(u, v) \mapsto \Phi(u + v) - \Phi(u) - \Phi'(u)v$ *is convex;*
- **(c)** $(u, v) \mapsto (\Phi'(u + v) - \Phi'(u))v$ *is convex;*
- **(d)** $(u, v) \mapsto \Phi''(u)v^2$ *is convex;*
- **(e)** Φ *is affine or* $\Phi'' > 0$ *and* $\Phi''''\Phi'' \geq 2\Phi'''^2$;
- **(f)** $(u, v) \mapsto t\Phi(u) + (1 - t)\Phi(v) - \Phi(tu + (1 - t)v)$ *is convex for any* $0 \leq t \leq 1$;
- **(g)** $\mathbb{E}_1 H_\Phi(Z|X_1) \geq H_\Phi(\mathbb{E}_1 Z)$;
- **(h)** $\{H_\Phi(Z)\}_{\Phi \in \mathcal{C}}$ *forms a convex set;*
- **(i)** $H_\Phi(Z) = \sup_{T>0} (\mathbb{E}[(\Phi'(T) - \Phi'(\mathbb{E}T))(Z - T)] + H_\Phi(T))$;
- **(j)** *subadditivity:* $H_\Phi(Z) \leq \sum_{i=1}^n \mathbb{E} H_\Phi^{(i)}(Z)$.

[1] Chafaï(2006). ESAIM: Probability and Statistics.
[2] Boucheron, Lugosi, and Massart, Concentration Inequalities: A Nonasymptotic Theory of Independence. Oxford University Press, 2013.

Usefulness

- ⋆ (b)-(d) lead to entropic inequalities for $M/M/\infty$ queueing process[1].
- ⋆ (g)&(i) lead to subadditivity (j).
- ⋆ (j) leads to powerful concentration inequalities (Entropy Method).

Matrix Φ-entropies

Classical Φ-Function Set $\mathcal{C}$

The class $\mathcal{C}$ contains each function $\Phi : [0, \infty) \to \mathbb{R}$ that is either affine or

- continuous and convex on $[0, \infty)$;
- twice differentiable on $(0, \infty)$;
- $\Phi'' > 0$ and $1/\Phi''$ is concave.

[1] Latała and Oleszkiewicz (2000), *Geometric Aspects of Functional Analysis*.

Matrix Φ-Function Set $\mathcal{M}$

The class Φ_d, $d \in \mathbb{N}$, contains each function $\Phi : [0, \infty) \to \mathbb{R}$ that is either affine or continuous, convex, twice continuously differentiable and

- Define $\Psi(t) = \Phi'(t)$ for $t > 0$. The Fréchet derivative $\mathrm{D}\Psi$ of the standard matrix function $\Psi : \mathbb{M}_d^{++} \to \mathbb{M}_d^{\mathbf{sa}}$ is an invertible linear map on $\mathbb{M}_d^{++}$, and the map $\mathbf{A} \mapsto (\mathrm{D}\Psi[\mathbf{A}])^{-\mathbf{1}}$ is concave.

Define $\mathcal{M} \equiv \bigcap_{d=\mathbf{1}}^{\infty} \Phi_d$.

[2] Chen & Tropp (2014), *Electron. J. Probab.*.

Standard Matrix Function

Let $f : I \to \mathbb{R}$ be a real-valued function on an interval I of the real line. Suppose that $\mathbf{A} = \int_{\lambda \in \sigma(\mathbf{A})} \lambda \, \mathrm{d}\mathbf{E} \in \mathbb{M}^{\mathbf{sa}}(I)$. Then $f(\mathbf{A}) \triangleq \int_{\lambda \in \sigma(\mathbf{X})} f(\lambda) \, \mathrm{d}\mathbf{E}(\lambda)$.

Fréchet derivative

Let $\mathcal{U}, \mathcal{W}$ be real Banach spaces. The *Fréchet derivative* of a function $\mathcal{L} : \mathcal{U} \to \mathcal{W}$ at a point $\mathbf{X} \in \mathcal{U}$, if it exists, is a unique linear mapping $\mathrm{D}\mathcal{L}[\mathbf{X}] : \mathcal{U} \to \mathcal{W}$ such that

$$\frac{\|\mathcal{L}(\mathbf{X}+\mathbf{E}) - \mathcal{L}(\mathbf{X}) - \mathrm{D}\mathcal{L}[\mathbf{X}](\mathbf{E})\|_{\mathcal{W}}}{\|\mathbf{E}\|_{\mathcal{U}}} \to 0 \quad \text{as } \mathbf{E} \in \mathcal{U} \text{ and } \|\mathbf{E}\|_{\mathcal{U}} \to \mathbf{0},$$

QCIS

Matrix Φ-entropies

Definition (Classical Φ-entropies)

The classical Φ-entropy $H_\Phi(Z)$ for $\Phi \in \mathcal{C}$ and non-negative random variable Z is defined as:

$$H_\Phi(Z) = \mathbb{E}\Phi(Z) - \Phi(\mathbb{E}Z)$$

so that $\mathbb{E}|Z| < \infty$ and $\mathbb{E}|\Phi(Z)| < \infty$.

Definition (Matrix Φ-entropies)

The matrix Φ-entropy $H_\Phi(\mathbf{Z})$ for $\Phi \in \mathcal{M}$ and a random matrix $\mathbf{Z} \in \mathbb{M}_d^+$ is defined as:

$$H_\Phi(\mathbf{Z}) = \overline{\text{tr}}\left[\mathbb{E}\Phi(\mathbf{Z}) - \Phi(\mathbb{E}\mathbf{Z})\right]$$

where $\overline{\text{tr}}$ is the normalized trace function and $\mathbb{E}\|\mathbf{Z}\| < \infty$ and $\mathbb{E}\|\Phi(\mathbf{Z})\| < \infty$.

Random Matrix

Denote a probability space by a triple $(\Omega, \mathcal{F}, \mu)$, where Ω is the sample space, $\mathcal{F}$ is a σ-algebra of subsets of Ω, and μ is a probability measure on $(\Omega, \mathcal{F})$. Then a random matrix $\mathbf{X}_d$ is a measurable map from $(\Omega, \mathcal{F})$ to $\mathbb{M}_d(\mathbb{C})$.

QCIS

Characterizations of Matrix Φ-entropies

Theorem

The following statements are equivalent.

(a) $\Phi \in \mathcal{M}$: Φ *is affine or* $\mathrm{D}\Psi$ *is invertible and* $\mathbf{A} \mapsto (\mathrm{D}\Psi[\mathbf{A}])^{-1}$ *is operator concave;*

(b) *Matrix Brègman divergence:* $(\mathbf{A}, \mathbf{B}) \mapsto \mathrm{tr}[\Phi(\mathbf{A}+\mathbf{B}) - \Phi(\mathbf{A}) - \mathrm{D}\Phi[\mathbf{A}](\mathbf{B})]$ *is convex;*

(c) $(\mathbf{A}, \mathbf{B}) \mapsto \mathrm{tr}[\mathrm{D}\Phi[\mathbf{A}+\mathbf{B}](\mathbf{B}) - \mathrm{D}\Phi[\mathbf{A}](\mathbf{B})]$ *is convex;*

(d) $(\mathbf{A}, \mathbf{B}) \mapsto \mathrm{tr}[\mathrm{D}^2\Phi[\mathbf{A}](\mathbf{B}, \mathbf{B})]$ *is convex;*

(e) Φ *is affine or* $\Phi'' > 0$ *and*

$$\mathrm{tr}\left[\mathbf{h} \cdot (\mathrm{D}\Psi[\mathbf{A}])^{-1} \circ \mathrm{D}^3\Psi[\mathbf{A}]\left(\mathbf{k}, \mathbf{k}, (\mathrm{D}\Psi[\mathbf{A}])^{-1}(\mathbf{h})\right)\right]$$
$$\geq 2\mathrm{tr}\left[\mathbf{h} \cdot (\mathrm{D}\Psi[\mathbf{A}])^{-1} \circ \mathrm{D}^2\Psi[\mathbf{A}]\left(\mathbf{k}, (\mathrm{D}\Psi[\mathbf{A}])^{-1}\left(\mathrm{D}^2\Psi[\mathbf{A}]\left(\mathbf{k}, (\mathrm{D}\Psi[\mathbf{A}])^{-1}(\mathbf{h})\right)\right)\right)\right],$$

for each $\mathbf{A} \succeq \mathbf{0}$ *and* $\mathbf{h}, \mathbf{k} \in \mathbb{M}_d^{sa}$;

(f) $(\mathbf{A}, \mathbf{B}) \mapsto \mathrm{tr}[t\Phi(\mathbf{A}) + (1-t)\Phi(\mathbf{B}) - \Phi(t\mathbf{A} + (1-t)\mathbf{B})]$ *is convex for any* $0 \leq t \leq 1$;

(g) $\mathbb{E}_1 H_\Phi(\mathbf{Z}|\mathbf{X}_1) \geq H_\Phi(\mathbb{E}_1 \mathbf{Z})$;

(h) $\{H_\Phi(\mathbf{Z})\}_{\Phi \in \mathcal{M}}$ *forms a convex set of convex functions;*

(i) $H_\Phi(\mathbf{Z}) = \sup_{\mathbf{T} \succ \mathbf{0}} \left(\mathrm{tr}\mathbb{E}\left[(\Phi'(\mathbf{T}) - \Phi'(\mathbb{E}\mathbf{T}))(\mathbf{Z} - \mathbf{T})\right] + H_\Phi(\mathbf{T})\right)$;

(j) $H_\Phi(\mathbf{Z}) \leq \sum_{i=1}^n \mathbb{E} H_\Phi^{(i)}(\mathbf{Z})$.

QCIS

Characterizations of Matrix Φ-entropies

Theorem

The following statements are equivalent.

- **(b)** *Matrix Brègman divergence:* $(\mathbf{A}, \mathbf{B}) \mapsto [\Phi(\mathbf{A}+\mathbf{B}) - \Phi(\mathbf{A}) - \mathrm{D}\Phi[\mathbf{A}](\mathbf{B})]$ *is convex;*
- **(c)** $(\mathbf{A}, \mathbf{B}) \mapsto [\mathrm{D}\Phi[\mathbf{A}+\mathbf{B}](\mathbf{B}) - \mathrm{D}\Phi[\mathbf{A}](\mathbf{B})]$ *is convex;*
- **(d)** $(\mathbf{A}, \mathbf{B}) \mapsto [\mathrm{D}^2\Phi[\mathbf{A}](\mathbf{B}, \mathbf{B})]$ *is convex;*
- **(f)** $(\mathbf{A}, \mathbf{B}) \mapsto [t\Phi(\mathbf{A}) + (1-t)\Phi(\mathbf{B}) - \Phi(t\mathbf{A} + (1-t)\mathbf{B})]$ *is convex for any* $0 \leq t \leq 1$*;*
- **(g)** $\mathbb{E}_1 H_\Phi(\mathbf{Z}|\mathbf{X}_1) \geq H_\Phi(\mathbb{E}_1 \mathbf{Z})$*;*
- **(h)** $\{H_\Phi(\mathbf{Z})\}_{\Phi \in \mathcal{M}}$ *forms a convex set of convex functions;*
- **(i)** $H_\Phi(\mathbf{Z}) = \sup_{\mathbf{T} \succ \mathbf{0}} \left(\mathbb{E}\left[(\Phi'(\mathbf{T}) - \Phi'(\mathbb{E}\mathbf{T}))(\mathbf{Z} - \mathbf{T})\right] + H_\Phi(\mathbf{T})\right)$*;*
- **(j)** $H_\Phi(\mathbf{Z}) \leq \sum_{i=1}^n \mathbb{E} H_\Phi^{(i)}(\mathbf{Z})$.

$H_\Phi(\mathbf{Z}) = [\mathbb{E}\Phi(\mathbf{Z}) - \Phi(\mathbb{E}\mathbf{Z})]$

Known Results and Proofs

(a): Φ is affine or $\mathrm{D}\Psi$ is invertible and $\mathbf{A} \mapsto (\mathrm{D}\Psi[\mathbf{A}])^{-1}$ is operator concave

$\Downarrow$

(i): $H_\Phi(\mathbf{Z}) = \sup_{\mathbf{T} \succ \mathbf{0}} \left(\mathrm{tr}\mathbb{E}\left[(\Phi'(\mathbf{T}) - \Phi'(\mathbb{E}\mathbf{T}))(\mathbf{Z} - \mathbf{T})\right] + H_\Phi(\mathbf{T})\right)$

$\Downarrow$

(g): $\mathbb{E}_1 H_\Phi(\mathbf{Z}|\mathbf{X}_1) \geq H_\Phi(\mathbb{E}_1\mathbf{Z})$

$\Downarrow$

(j): $H_\Phi(\mathbf{Z}) \leq \sum_{i=1}^n \mathbb{E} H_\Phi^{(i)}(\mathbf{Z})$

[1] Chen & Tropp (2014), *Electron. J. Probab.*.

Known Results and Proofs

(a): Φ is affine or $\mathrm{D}\Psi$ is invertible and $\mathbf{A} \mapsto (\mathrm{D}\Psi[\mathbf{A}])^{-1}$ is operator concave

$\updownarrow$ [2]

(b): Matrix Brègman divergence: $(\mathbf{A}, \mathbf{B}) \mapsto \mathrm{tr}[\Phi(\mathbf{A}+\mathbf{B}) - \Phi(\mathbf{A}) - \mathrm{D}\Phi[\mathbf{A}](\mathbf{B})]$ is convex

[2] Pitrik & Virosztek (2015), *Letters in Mathematical Physics*.

Known Results and Proofs

(a): Φ is affine or $\mathrm{D}\Psi$ is invertible and $\mathbf{A} \mapsto (\mathrm{D}\Psi[\mathbf{A}])^{-1}$ is operator concave

$\updownarrow$ [3]

Convexity Lemma: $\mathbb{E}\langle \mathbf{X}, \mathrm{D}\Psi[\mathbf{A}](\mathbf{X})\rangle \geq \langle \mathbb{E}[\mathbf{X}], \mathrm{D}\Psi[\mathbb{E}\mathbf{A}](\mathbb{E}\,\mathbf{X})\rangle$

$\updownarrow$ $\mathrm{tr}(\mathrm{D}^2\Phi[\mathbf{A}](\mathbf{X},\mathbf{X})) = \langle \mathbf{X}, \mathrm{D}\Phi'[\mathbf{A}](\mathbf{X})\rangle$

(d): $(\mathbf{A},\mathbf{B}) \mapsto \mathrm{tr}[\mathrm{D}^2\Phi[\mathbf{A}](\mathbf{B},\mathbf{B})]$ is convex

[3] Hansen & Zhang (2014), arXiv: 1402.2118.

$(b) \Leftrightarrow (c) \Leftrightarrow (d)$

$$
\begin{aligned}
A_{\Phi}(\mathbf{u}, \mathbf{v}) &\triangleq \mathrm{tr}[\Phi(\mathbf{u}+\mathbf{v}) - \Phi(\mathbf{u}) - \mathrm{D}\Phi[\mathbf{u}](\mathbf{v})] \\
B_{\Phi}(\mathbf{u}, \mathbf{v}) &\triangleq \mathrm{tr}[\mathrm{D}\Phi[\mathbf{u}+\mathbf{v}](\mathbf{v}) - \mathrm{D}\Phi[\mathbf{u}](\mathbf{v})] \\
C_{\Phi}(\mathbf{u}, \mathbf{v}) &\triangleq \mathrm{tr}[\mathrm{D}^2\Phi[\mathbf{u}](\mathbf{v}, \mathbf{v})].
\end{aligned}
$$

$(b) \Leftrightarrow (c) \Leftrightarrow (d)$

$$
\begin{aligned}
A_\Phi(\mathbf{u},\mathbf{v}) &\triangleq \mathrm{tr}[\Phi(\mathbf{u}+\mathbf{v}) - \Phi(\mathbf{u}) - \mathrm{D}\Phi[\mathbf{u}](\mathbf{v})] \\
B_\Phi(\mathbf{u},\mathbf{v}) &\triangleq \mathrm{tr}[\mathrm{D}\Phi[\mathbf{u}+\mathbf{v}](\mathbf{v}) - \mathrm{D}\Phi[\mathbf{u}](\mathbf{v})] \\
C_\Phi(\mathbf{u},\mathbf{v}) &\triangleq \mathrm{tr}[\mathrm{D}^2\Phi[\mathbf{u}](\mathbf{v},\mathbf{v})].
\end{aligned}
$$

(b): $(\mathbf{u},\mathbf{v}) \mapsto A_\Phi(\mathbf{u},\mathbf{v})$ is convex

(c): $(\mathbf{u},\mathbf{v}) \mapsto B_\Phi(\mathbf{u},\mathbf{v})$ is convex

$$A_\Phi(\mathbf{u},\mathbf{v}) = \int_0^1 (1-s)\, C_\Phi(\mathbf{u}+s\mathbf{v},\mathbf{v})\mathrm{d}s$$

$$B_\Phi(\mathbf{u},\mathbf{v}) = \int_0^1 C_\Phi(\mathbf{u}+s\mathbf{v},\mathbf{v})\mathrm{d}s$$

(d): $(\mathbf{u},\mathbf{v}) \mapsto C_\Phi(\mathbf{u},\mathbf{v})$ is convex

$(b) \Leftrightarrow (c) \Leftrightarrow (d)$

$$A_\Phi(\mathbf{u},\mathbf{v}) \triangleq \mathrm{tr}[\Phi(\mathbf{u}+\mathbf{v}) - \Phi(\mathbf{u}) - \mathrm{D}\Phi[\mathbf{u}](\mathbf{v})]$$
$$B_\Phi(\mathbf{u},\mathbf{v}) \triangleq \mathrm{tr}[\mathrm{D}\Phi[\mathbf{u}+\mathbf{v}](\mathbf{v}) - \mathrm{D}\Phi[\mathbf{u}](\mathbf{v})]$$
$$C_\Phi(\mathbf{u},\mathbf{v}) \triangleq \mathrm{tr}[\mathrm{D}^2\Phi[\mathbf{u}](\mathbf{v},\mathbf{v})].$$

(b): $(\mathbf{u},\mathbf{v}) \mapsto A_\Phi(\mathbf{u},\mathbf{v})$ is convex

(c): $(\mathbf{u},\mathbf{v}) \mapsto B_\Phi(\mathbf{u},\mathbf{v})$ is convex

$A_\Phi(\mathbf{u},\varepsilon\mathbf{v}) = \frac{1}{2} C_\Phi(\mathbf{u},\mathbf{v})\varepsilon^2 + o(\varepsilon^2)$

$B_\Phi(\mathbf{u},\varepsilon\mathbf{v}) = C_\Phi(\mathbf{u},\mathbf{v})\varepsilon^2 + o(\varepsilon^2)$

(d): $(\mathbf{u},\mathbf{v}) \mapsto C_\Phi(\mathbf{u},\mathbf{v})$ is convex

Functional Inequalities

We proved two families of Matrix Concentration Inequalities:

- Poincaré Inequality:
- (Logarithmic) Sobolev Inequality:
  - Hypercontractive inequality
  - Influence functions
  - ....

Matrix Poincaré Inequality

Theorem (Matrix Poincaré Inequality)

Let $\mathbf{X} = (\mathbf{X}_1, \ldots, \mathbf{X}_n) \in (\mathbb{M}^{sa}_{d_1})^n$ *be an n-tuple independent random matrix taking values in the interval* $[\mathbf{0}, \mathbf{I}]$ *and let* $\mathcal{L} : (\mathbb{M}^{sa}_{d_1}([0,1]))^n \to \mathbb{M}^{sa}_{d_2}$ *be a separately convex function*[a] *with finite partial Fréchet derivatives. Then* $\mathcal{L}(\mathbf{X})$ *satisfies*

$$\mathrm{Var}\left(\mathcal{L}(\mathbf{X})\right) \le \sum_{i=1}^{n} \mathbb{E}\left[\|\mathrm{D}_{\mathbf{X}_i}\mathcal{L}[\mathbf{X}]\|_2^2 \right]$$

where $\|\mathrm{D}_{\mathbf{X}_i}\mathcal{L}[\mathbf{X}]\|_2$ *is the norm of the Fréchet derivative:*

$$\|\mathrm{D}\mathcal{L}[\mathbf{X}]\|_2 = \sup_{\mathbf{E}\neq 0} \frac{\|\mathrm{D}\mathcal{L}[[X](\mathbf{E})\|_2}{\|\mathbf{E}\|_2}$$

[a] The separate convexity means that: for $0 \le t \le 1$,

$$t\mathcal{L}(\mathbf{Y}) + (1-t)\mathcal{L}(\widetilde{\mathbf{Y}}^{(i)}) \preceq \mathcal{L}\left(t\mathbf{Y} + (1-t)\widetilde{\mathbf{Y}}^{(i)}\right)$$

where $\mathbf{Y} = (\mathbf{Y}_1, \ldots, \mathbf{Y}_n) \in (\mathbb{M}^{sa}_d)^n$, and $\widetilde{\mathbf{Y}}^{(i)} = (\mathbf{Y}_1, \ldots, \mathbf{Y}_{i-1}, \mathbf{Y}'_i, \mathbf{Y}_{i+1}, \ldots, \mathbf{Y}_n) \in (\mathbb{M}^{sa}_d)^n$.

Matrix Poincaré Inequality

Theorem (Matrix Poincaré Inequality)

Let $\mathbf{X} = (\mathbf{X}_1, \ldots, \mathbf{X}_n) \in (\mathbb{M}^{\text{sa}}_{d_1})^n$ *be an n-tuple independent random matrix taking values* ~~*in the interval* $[\mathbf{0}, \mathbf{I}]$~~ *and let* $\mathcal{L} : (\mathbb{M}^{\text{sa}}_{d_1})^n \to \mathbb{M}^{\text{sa}}_{d_2}$ *be* ~~*a separately convex function*~~ *with finite partial Fréchet derivatives. Then* $\mathcal{L}(\mathbf{X})$ *satisfies*

$$\text{Var}\left(\mathcal{L}(\mathbf{X})\right) \leq \sum_{i=1}^{n} \mathbb{E}\left[\|\mathrm{D}_{\mathbf{X}_i}\mathcal{L}[\mathbf{X}]\|_2^2 \right]$$

where $\|\mathrm{D}_{\mathbf{X}_i}\mathcal{L}[\mathbf{X}]\|_2$ *is the norm of the Fréchet derivative:*

$$\|\mathrm{D}\mathcal{L}[\mathbf{X}]\|_2 = \sup_{\mathbf{E}\neq 0} \frac{\|\mathrm{D}\mathcal{L}[[X](\mathbf{E})\|_2}{\|\mathbf{E}\|_2}$$

Matrix Gaussian Poincaré Inequality

Theorem (Matrix Gaussian Poincaré Inequality)

Let $\mathbf{X} = (\mathbf{X}_1, \ldots, \mathbf{X}_n) \in (\mathbb{M}^{\mathrm{sa}}_{d_1})^n$ *be an n-tuple independent random matrix taking values from Gaussian Unitary Ensemble and let* $\mathcal{L} : (\mathbb{M}^{\mathrm{sa}}_{d_1})^n \to \mathbb{M}^{\mathrm{sa}}_{d_2}$ *be any twice Fréchet differentiable function. Then* $\mathcal{L}(\mathbf{X})$ *satisfies*

$$\mathrm{Var}\left(\mathcal{L}(\mathbf{X})\right) \leq \sum_{i=1}^{n} \mathbb{E}\left[\|\mathrm{D}_{\mathbf{X}_i}\mathcal{L}[\mathbf{X}]\|_2^2 \right]$$

Matrix Poincaré Inequality

Classical Poincaré Inequality: $\mathrm{Var}(f(X)) \leq \mathbb{E}\left[\|\nabla f(X)\|^2\right]$

Matrix Poincaré Inequality: $\mathrm{Var}\left(\mathcal{L}(\mathbf{X})\right) \leq \sum_{i=1}^{n} \mathbb{E}\left[\,\|\mathrm{D}_{\mathbf{X}_i}\mathcal{L}[\mathbf{X}]\|_2^2\right],$

Matrix Poincaré Inequality

Matrix Poincaré Inequality: $\mathrm{Var}\left(\mathcal{L}(\mathbf{X})\right) \le \sum_{i=1}^{n} \mathbb{E}\left[\|\mathrm{D}_{\mathbf{X}_i}\mathcal{L}[\mathbf{X}]\|_2^2 \right]$,

Proof:

Matrix Efron-Stein Inequality[1]: $\mathrm{Var}\left(\mathcal{L}(\mathbf{X})\right) \le \sum_{i=1}^{n} \mathrm{tr}\mathbb{E}\left(\mathcal{L}\left(\mathbf{X}\right) - \mathcal{L}\left(\widetilde{\mathbf{X}}^{(i)}\right)\right)_+^2$,

⇓ separate operator convexity of $\mathcal{L}$

$$\le \sum_{i=1}^{n} \mathrm{tr}\mathbb{E}\left|\mathrm{D}_{\mathbf{X}_i}\mathcal{L}[\mathbf{X}]\left(\mathbf{X}_i - \mathbf{X}_i'\right)\right|^2$$

⇓ norm of Fréchet derivative

$$\le \sum_{i=1}^{n} \mathbb{E}\left[\|\mathrm{D}_{\mathbf{X}_i}\mathcal{L}[\mathbf{X}]\|_2^2 \right]$$

[1] Paulin, Mackey, and Tropp, "Efron-Stein inequalities for random matrices," arXiv: 1408.3470, 2014.

Matrix Gaussian Poincaré Inequality

Matrix Poincaré Inequality: $\mathrm{Var}\left(\mathcal{L}(\mathbf{X})\right) \leq \sum_{i=1}^{n} \mathbb{E}\left[\|\mathrm{D}_{\mathbf{X}_i}\mathcal{L}[\mathbf{X}]\|_2^2 \right]$,

Proof:

$$\mathbf{S}_m \triangleq \frac{1}{\sqrt{m}} \sum_{j=1}^{m} \epsilon_j \mathbf{Y}_j.$$

⇓

$$\mathrm{Var}\left(\mathcal{L}(\mathbf{S}_m)\right) \leq \frac{1}{4} \sum_{j=1}^{m} \mathrm{tr}\mathbb{E}\left[\left(\mathcal{L}\left(\mathbf{S}_m + \frac{1-\epsilon_j}{\sqrt{m}} \mathbf{Y}_j \right) - \mathcal{L}\left(\mathbf{S}_m - \frac{1+\epsilon_j}{\sqrt{m}} \mathbf{Y}_j \right) \right)^2 \right].$$

⇓ Taylor Expansion

$$\leq \frac{1}{4} \sum_{j=1}^{m} \mathrm{tr}\mathbb{E}\left[\left(\frac{2}{\sqrt{m}} \mathrm{D}\mathcal{L}[\mathbf{S}_m]\left(\mathbf{Y}_j\right) + o\left(\frac{1}{m}\right) \right)^2 \right]$$

⇓ Central Limit Theorem

$$\leq \mathrm{tr}\mathbb{E}\left[\|\mathrm{D}\mathcal{L}[\mathbf{X}]\|_2^2 \right]$$

Matrix Φ-Sobolev inequality

Theorem (Matrix Φ-Sobolev Inequalities for Symmetric Bernoulli Random Variables)

Let X be uniformly distributed over $\mathcal{X} \equiv \{0,1\}^n$ (an n-dimensional binary hypercube) and $\boldsymbol{f} : \mathcal{X} \to \mathbb{M}_d^+$ be an arbitrary matrix-valued function. Then for all $p \in (1,2)$, and $\Phi(u) = u^{2/p}$,

$$H_\Phi(\boldsymbol{f}^p) \leq (2-p)\mathcal{E}(\boldsymbol{f}) \cdot d^{1-2/p} + \mathrm{tr}\mathbb{E}[\boldsymbol{f}^2] \cdot (1 - d^{1-2/p}).$$

where

$$\mathcal{E}(\mathbf{f}) = \frac{1}{2}\mathrm{tr}\mathbb{E}\left[\sum_{i=1}^{n}\left(\mathbf{f}(X) - \mathbf{f}\left(\widetilde{X}^{(i)}\right)\right)^2\right].$$

Logarithmic Sobolev Inequality

Theorem (Matrix Φ-Sobolev Inequalities)

$H_\Phi(\boldsymbol{f}^p) \le (2-p)\mathcal{E}(\boldsymbol{f}) \cdot d^{1-2/p} + \mathrm{tr}\mathbb{E}[\boldsymbol{f}^2] \cdot (1 - d^{1-2/p})$. $[\Phi(u) = u^{2/p},\ p \in (1,2).]$

Logarithmic Sobolev Inequality

Theorem (Matrix Φ-Sobolev Inequalities)

$H_\Phi(\boldsymbol{f}^p) \leq (2-p)\mathcal{E}(\boldsymbol{f}) \cdot d^{1-2/p} + \mathrm{tr}\mathbb{E}[\boldsymbol{f}^2] \cdot (1 - d^{1-2/p})$. *[*$\Phi(u) = u^{2/p}$, $p \in (1,2)$*.]*

Corollary (Matrix Logarithmic Sobolev Inequalities)

When $p = 2$, $Ent(\mathbf{f}^2) \leq 2\mathcal{E}(\mathbf{f}) + \log(d) \cdot \mathrm{tr}\mathbb{E}\left[\mathbf{f}^2\right]$ *, where* $Ext{=}H_{x \log x}$.

Logarithmic Sobolev Inequality

Theorem (Matrix Φ-Sobolev Inequalities)

$H_\Phi(\boldsymbol{f}^p) \leq (2-p)\mathcal{E}(\boldsymbol{f}) \cdot d^{1-2/p} + \mathrm{tr}\mathbb{E}[\boldsymbol{f}^2] \cdot (1 - d^{1-2/p})$. *[$\Phi(u) = u^{2/p}$, $p \in (1,2)$.]*

Corollary (Matrix Logarithmic Sobolev Inequalities)

When $p = 2$, $Ent(\mathbf{f}^2) \leq 2\mathcal{E}(\mathbf{f}) + \log(d) \cdot \mathrm{tr}\mathbb{E}\left[\mathbf{f}^2\right]$, where $Ext{=}H_{x \log x}$.

Remark

In general, logarithmic Sobolev inequalities with constants $C > 0$, $D \geq 0$ are of the following form:

$$\mathrm{Ent}(f^2) \leq 2C\mathcal{E}(f) + D\mathbb{E}[f^2].$$

When $D = 0$, the logarithmic Sobolev inequality is called *tight*; otherwise, it is called *defective*. It is well known that the best constants of the classical logarithmic Sobolev inequalities for symmetric Bernoulli random variables and standard Gaussian random variables are $(C, D) = (1, 0)$. However, numerical simulation shows that examples $(d > 1)$ exist for matrix-valued functions so that: $\mathrm{Ent}(\mathbf{f}^2) > 2\mathcal{E}(\mathbf{f})$. Our result establishes the logarithmic Sobolev inequality with constant $(C, D) = (1, \log d)$.

Proof of Matrix Sobolev Inequality

$$H_\Phi(\boldsymbol{f}^p) \leq (2-p)\mathcal{E}(\boldsymbol{f}) \cdot d^{1-2/p} + \mathrm{tr}\mathbb{E}[\boldsymbol{f}^2] \cdot (1 - d^{1-2/p}).$$

$$H_\Phi(\boldsymbol{f}^p) \leq \mathrm{tr}\mathbb{E}\left[\boldsymbol{f}^2\right] - \left(\mathbb{E}\left\|\boldsymbol{f}\right\|_{*p}^p\right)^{2/p}$$

Parseval's identity ⇓ Matrix Bonami-Beckner Inequality[1]

$$\leq \mathrm{tr}\left[\sum_{S\subseteq[n]} \widehat{\boldsymbol{f}}(S)^2\right] - \left(\sum_{S\subseteq[n]} (p-1)^{|S|} \left\|\widehat{\boldsymbol{f}}(S)\right\|_{*p}^2\right)$$

⇓

$$\leq \mathrm{tr}\left[\sum_{S\subseteq[n]} \left((2-p)|S|d^{1-2/p} + (1-d^{1-2/p})\right)\widehat{\boldsymbol{f}}(S)^2\right]$$

$\sum_{S\subseteq[n]} \mathrm{tr}\left[|S|\widehat{\boldsymbol{f}}(S)^2\right] = \mathcal{E}(\boldsymbol{f})$ ⇓ Parseval's identity

$$= (2-p)\mathcal{E}(\boldsymbol{f}) \cdot d^{1-2/p} + \mathrm{tr}\mathbb{E}[\boldsymbol{f}^2] \cdot (1 - d^{1-2/p})$$

[1] Ben-Aroya, Regev, and de Wolf (2008), FOCS.

Proof of Matrix Logarithmic Sobolev Inequalities

$$\mathrm{Ent}(\mathbf{f}^2) \leq 2\mathcal{E}(\mathbf{f}) + \log d \cdot \mathrm{tr}\mathbb{E}\left[\mathbf{f}^2\right]$$

$$H_\Phi(\boldsymbol{f}^p) \leq (2-p)\mathcal{E}(\boldsymbol{f}) \cdot d^{1-2/p} + \mathrm{tr}\mathbb{E}[\boldsymbol{f}^2] \cdot (1 - d^{1-2/p})$$

⇓

$$\lim_{p\to 2^-} \frac{H_\Phi(\mathbf{f}^p)}{2-p} \leq \mathcal{E}(\boldsymbol{f}) + \lim_{p\to 2^-} \frac{\mathrm{tr}\mathbb{E}\left[\boldsymbol{f}^2\right]\cdot(1-d^{1-2/p})}{2-p}$$

⇓

$$\frac{1}{2}\mathrm{Ext}(\mathbf{f}^2) \leq \mathcal{E}(\boldsymbol{f}) + \frac{\log d}{2}\mathrm{tr}\mathbb{E}\left[\boldsymbol{f}^2\right]$$

$$\lim_{p\to 2^-} \frac{\mathbb{E}\left[\mathbf{Z}^2\right] - \left(\mathbb{E}\left[\mathbf{Z}^p\right]\right)^{2/p}}{2-p} = \frac{1}{2}\mathbb{E}\left[\mathbf{Z}^2 \log \mathbf{Z}^2\right] - \frac{1}{2}\mathbb{E}\left[\mathbf{Z}^2\right] \cdot \log \mathbb{E}\left[\mathbf{Z}^2\right].$$

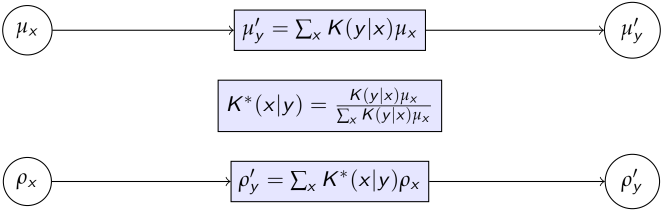
Holevo Quantity

If we consider a random matrix ρ_X to be an ensemble of classical-quantum (c-q) states $(\mu, \nu) \triangleq \{(\mu(x), \rho_x)\}_{x \in \mathcal{X}}$ whose average $\mathbb{E}\rho_X \equiv \bar{\rho}$, then its Φ-entropy is related to the *Holevo quantity* of $\{(\mu(x), \rho_x)\}_{x \in \mathcal{X}}$:

$$
\begin{aligned}
d \cdot H_\Phi(\rho_X) &\equiv \sum_{x \in \mathcal{X}} \mu(x) \mathrm{tr}\left[\rho_x \log \rho_x\right] - \mathrm{tr}\left[\bar{\rho} \log \bar{\rho}\right] \\
&= \sum_{x \in \mathcal{X}} \mu(x) \cdot S\left(\rho_x \| \bar{\rho}\right) \\
&=: \chi(\mu, \nu),
\end{aligned}
$$

where $S(\rho \| \sigma) \triangleq \mathrm{tr}\rho(\log \rho - \log \sigma)$ is the quantum relative entropy.

Classical evolution of a quantum ensemble



Let $\eta_\Phi(\mu, K) \triangleq \sup_{\{\rho_x\}} \frac{\chi(\{\mu'_y, \rho'_y\})}{\chi(\{\mu_x, \rho_x\})}$.

Theorem

Let (X, Y) be a random pair with probability law $\mu \otimes K$. Then $\eta_\Phi(\mu, K) \leq c$ if and only if the inequality $\chi(\{\mu_x, \rho_x\}) \leq \frac{1}{1-c}\mathbb{E}_Y\left[\chi(\{K^*(x|y), \rho_x\}|Y = y)\right]$

Open Questions

1. Is the logarithmic Sobolev inequality with constant $(C, D) = (1, \log d)$ optimal?
2. Could we find more applications in quantum information theory?

Thank you for your attention!

Exponential Decay of Matrix Φ-Entropies on Markov Semigroups

- The Markov semigroup $\mathsf{P}_t : \mathbb{M}^{\mathsf{sa}} \to \mathbb{M}^{\mathsf{sa}}$, $t \geq 0$:
  $\mathsf{P}_t\mathbf{f}(x) \triangleq \int \mathbf{f}(y)\,\mathsf{P}_t(x, \mathrm{d}y)$.
  $\mathsf{P}_t \circ \mathsf{P}_s = \mathsf{P}_{s+t}, \quad \mathsf{P}_t\mathbf{I} = \mathbf{I}, \quad \mathbf{f} \succeq \mathbf{0} \Rightarrow \mathsf{P}_t\mathbf{f} \succeq \mathbf{0}$.
- Infinitesimal generator: $\mathsf{L}(\mathbf{f}) \triangleq \lim_{t\to 0^+} \frac{1}{t}(\mathsf{P}_t\mathbf{f} - \mathbf{f})$
- The *carré du champ* operator: $\Gamma(\mathbf{f},\mathbf{f}) \triangleq \frac{1}{2}\left(\mathsf{L}(\mathbf{f}^2) - \mathbf{f}\mathsf{L}(\mathbf{f}) - \mathsf{L}(\mathbf{f})\mathbf{f}\right)$,
  $\Gamma(\mathbf{f},\mathbf{g}) = \Gamma(\mathbf{g},\mathbf{f}) \triangleq \frac{1}{2}\left(\Gamma(\mathbf{f}+\mathbf{g},\mathbf{f}+\mathbf{g}) - \Gamma(\mathbf{f},\mathbf{f}) - \Gamma(\mathbf{g},\mathbf{g})\right)$
- *Dirichlet form*:
  $\mathcal{E}(\mathbf{f},\mathbf{g}) \triangleq \int \Gamma(\mathbf{f},\mathbf{g})\,\mathrm{d}\mu = -\frac{1}{4}\int \mathbf{f}\mathsf{L}(\mathbf{g}) + \mathbf{g}\mathsf{L}(\mathbf{f}) + \mathsf{L}(\mathbf{f})\mathbf{g} + \mathsf{L}(\mathbf{g})\mathbf{f}\,\mathrm{d}\mu$.

Theorem (Exponential Decay of Variances for Bernoulli Random Variables)

Given a Markov Triple $(\{0,1\}^n, \Gamma, \mu_{n,p})$ *of a Markovian jump process, one has*

$$\mathsf{Var}\left(\mathsf{P}_t\mathbf{f}\right) \leq \mathrm{e}^{-2t} \cdot \mathsf{Var}\left(\mathbf{f}\right),$$

for any matrix-valued function $\mathbf{f} : \{0,1\}^n \to \mathbb{M}_d^{sa}$.

Exponential Decay of Matrix Φ-Entropies on Markov Semigroups

- The Markov semigroup $\mathsf{P}_t : \mathbb{M}^{\text{sa}} \to \mathbb{M}^{\text{sa}}$, $t \geq 0$:
  $\mathsf{P}_t \mathbf{f}(x) \triangleq \int \mathbf{f}(y)\, \mathsf{P}_t(x, \mathrm{d}y)$.
  $\mathsf{P}_t \circ \mathsf{P}_s = \mathsf{P}_{s+t}, \quad \mathsf{P}_t \mathbf{I} = \mathbf{I}, \quad \mathbf{f} \succeq \mathbf{0} \Rightarrow \mathsf{P}_t \mathbf{f} \succeq \mathbf{0}$.
- Infinitesimal generator: $\mathsf{L}(\mathbf{f}) \triangleq \lim_{t \to 0^+} \frac{1}{t}(\mathsf{P}_t \mathbf{f} - \mathbf{f})$
- The *carré du champ* operator: $\Gamma(\mathbf{f}, \mathbf{f}) \triangleq \frac{1}{2}\left(\mathsf{L}(\mathbf{f}^2) - \mathbf{f}\mathsf{L}(\mathbf{f}) - \mathsf{L}(\mathbf{f})\mathbf{f}\right)$,
  $\Gamma(\mathbf{f}, \mathbf{g}) = \Gamma(\mathbf{g}, \mathbf{f}) \triangleq \frac{1}{2}\left(\Gamma(\mathbf{f}+\mathbf{g}, \mathbf{f}+\mathbf{g}) - \Gamma(\mathbf{f}, \mathbf{f}) - \Gamma(\mathbf{g}, \mathbf{g})\right)$
- *Dirichlet form*:
  $\mathcal{E}(\mathbf{f}, \mathbf{g}) \triangleq \int \Gamma(\mathbf{f}, \mathbf{g})\, \mathrm{d}\mu = -\frac{1}{4} \int \mathbf{f}\mathsf{L}(\mathbf{g}) + \mathbf{g}\mathsf{L}(\mathbf{f}) + \mathsf{L}(\mathbf{f})\mathbf{g} + \mathsf{L}(\mathbf{g})\mathbf{f}\, \mathrm{d}\mu$.

Theorem (Exponential Decay of Entropies for Bernoulli Random Variables)

Given a Markov Triple $(\{0,1\}^n, \Gamma, \mu_{n,p})$ *of a Markovian jump process, one has*

$$\mathit{Ent}(\mathsf{P}_t \mathbf{f}) \leq \mathrm{e}^{-t} \cdot \mathit{Ent}(\mathbf{f}),$$

for any matrix-valued function $\mathbf{f} : \{0,1\}^n \to \mathbb{M}_d^{sa}$.