An Invitation to Classical and Quantum Information Geometry

Attila Andai

September 4, 2017

These slides were presented at the following conference.

XXVI International Fall Workshop on Geometry and Physics

Universidade do Minho, Braga, Portugal

4-7 September 2017

The slides may contain minor errors and typos. Use at your own risk.

Classical information geometry

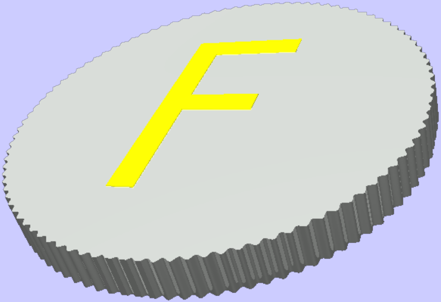
Quantum information geometry

1. Introduction to noncommutative information geometry
2. Preparations for Petz theorem
3. Means
4. Petz theorem
5. Operator monotone functions
6. Computing monotone metrics

Advanced topics

1. Relative entropy
2. Duality
3. About volume of the state space
4. Uncertainty relations

Classical information geometry



Information Geometry

Statistical model $\approx$ Parametric probability distribution

Information geometry $\approx$ Riemannian metric on statistical model

Statistical model

Definition

Statistical model: $\mathcal{S} = (X, \mathcal{B}(X), S, \Xi)$

1. $X \neq \emptyset$ set, $\mathcal{B}(X)$ σ algebra on X,
2. the elements of S are probability measures on $\mathcal{B}(X)$,
3. there exists a bijection $i : \Xi \to S \quad \vartheta \mapsto \mu_\vartheta$

Ξ: Parameter space

(This setting is too general.)

We make more assumptions.

1. $\exists n \in \mathbb{N}^+$: $\Xi \subseteq \mathbb{R}^n$, moreover Ξ connected open set. (*n-dimensional statistical model*)
2. If X is finite, then $\mathcal{B}(X) = \mathcal{P}(X)$.
3. If X is infinite, then $X \subseteq \mathbb{R}^m$, X connected open set, $\mathcal{B}(X)$ contains Borel sets and for every $\vartheta \in \Xi$ the probability distribution $\mu_\vartheta \in S$ has density function p_ϑ (with respect to the Lebesgue measure).
4. We refer to the elements of S as density functions and denote it by $p(x, \vartheta) = p_\vartheta(x)$.
5. Every function $p_\vartheta \in S$ has 1., 2., and 3. moment.

6. For every $x \in X$ the function

$$\Xi \to \mathbb{R} \quad \vartheta \mapsto p(x,\vartheta)$$

is smooth. We use the notation

$$\partial_i p(x,\vartheta) = \frac{\partial p(x,\vartheta)}{\partial \vartheta_i} \qquad i = 1,\dots,m.$$

7. We assume that

$$\int_X \partial_{i_1} \dots \partial_{i_k} p(x,\vartheta)\,\mathrm{d}\,x = \partial_{i_1} \dots \partial_{i_k} \int_X p(x,\vartheta)\,\mathrm{d}\,x = 0.$$

8. $\forall \vartheta \in \Xi$ and $\forall x \in X$: $p(x,\vartheta) > 0$

The statistical model is denoted by (X, S, Ξ).

Example (Discrete distribution)

$$X = \{0, 1, \ldots, n\}$$

$$\Xi = \left\{ (\vartheta_1, \ldots, \vartheta_n) \in \mathbb{R}^n \;\middle|\; \vartheta_i > 0, \; \sum_{k=1}^{n} \vartheta_k < 1 \right\}$$

$$p(x, \vartheta) = \begin{cases} \vartheta_x & \text{if} \quad 1 \leq x \leq n, \\ 1 - \sum_{k=1}^{n} \vartheta_k & \text{if} \quad x = 0. \end{cases}$$

The space of distributions:

$$\mathcal{P}_n = \left\{ (p_0, p_1, \ldots, p_n) \in]0, 1[^{n+1} \;\middle|\; \sum_{i=0}^{n} p_i = 1 \right\}.$$

Example (Normal distribution)

$$X = \mathbb{R}$$

$$\Xi = \mathbb{R} \times \mathbb{R}^+$$

$$p(x, \mu, \sigma) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right)$$

Fisher information matrix

For an n-dimensional statistical model (X, S, Ξ) the Fisher information is an $n \times n$ matrix for every parameter $\vartheta \in \Xi$.

Definition

Assume that (X, S, Ξ) is an n dimensional statistical model. For every point $\vartheta \in \Xi$ the *Fisher information matrix* is given by

$$g^{(\mathrm{F})}(\vartheta)_{ik} = \int_X \frac{1}{p(x,\vartheta)} (\partial_i p(x,\vartheta))(\partial_k p(x,\vartheta)) \,\mathrm{d}\, x.$$

The Fisher matrix denoted by $g^{(\mathrm{F})}(\vartheta)$.

We will use the following representations for Fisher matrix.

$$g^{(\mathrm{F})}(\vartheta)_{ik} = \int_X p(x,\vartheta)(\partial_i \log p(x,\vartheta))(\partial_k \log p(x,\vartheta)) \,\mathrm{d}\, x$$

$$g^{(\mathrm{F})}(\vartheta)_{ik} = 4 \int_X (\partial_i \sqrt{p(x,\vartheta)})(\partial_k \sqrt{p(x,\vartheta)}) \,\mathrm{d}\, x$$

Theorem

Assume that (X, S, Ξ) is an n dimensional statistical model. If the functions $(\partial_i p(\cdot, \vartheta))_{i=1,\ldots,n}$ are linearly independent at a point $\vartheta \in \Xi$ then the Fisher matrix $g^{(F)}(\vartheta)$ positive definite.

Proof.

For every $c \in \mathbb{R}^n$

$$\left\langle (c_1, \ldots, c_n), g^{(\mathrm{F})}(\vartheta)(c_1, \ldots, c_n) \right\rangle$$

$$= \int_X p(x, \vartheta) \left(\sum_{i=1}^{n} c_i \partial_i (\log p(x, \vartheta)) \right)^2 \, \mathrm{d}\, x \geq 0.$$

□

Induced statistical models

Assume that $(X, \mathcal{B}(X), S, \Xi)$ is a statistical model and

$$f : X \to Y \quad x \mapsto f(x)$$

is a surjective map.

Let us define $\mathcal{B}(Y) = \left\{ A \subseteq Y | \overset{-1}{f}(A) \in \mathcal{B}(X) \right\}$.

For every $\vartheta \in \Xi$, μ_ϑ is probability measure on X, with density function p_ϑ.

Now define $\tilde{\mu}_\vartheta$ as

$$\tilde{\mu}_\vartheta(A) = \mu_\vartheta \left(\overset{-1}{f}(A) \right) \quad \forall A \in \mathcal{B}(Y)$$

and denote its density function with $\tilde{p}_\vartheta$.

Define $\tilde{S}$ as $\{\tilde{\mu}_\vartheta | \vartheta \in \Xi\}$.

After these steps, we have an *induced statistical model*

$$(Y, \mathcal{B}(Y), \tilde{S}, \Xi).$$

Monotonicity of Fisher matrix

If we measure less precisely we can have less information.

Definition

Assume that (X, S, Ξ) is a statistical model and $f : X \to Y$ is a measurable surjective map. Let us define

$$r(\cdot,\cdot) : X \times \Xi \to \mathbb{R} \quad (x,\vartheta) \mapsto r(x,\vartheta) = \frac{p(x,\vartheta)}{\tilde{p}(f(x),\vartheta)}.$$

f *sufficient statistic* of S, if for every $x \in X$ the function

$$r(x,\cdot) : \Xi \to \mathbb{R} \quad \vartheta \mapsto r(x,\vartheta)$$

is constant.

Monotonicity of Fisher matrix

If we measure less precisely we can have less information.

Definition

Assume that (X, S, Ξ) is a statistical model and $f : X \to Y$ is a measurable surjective map. Let us define

$$r(\cdot,\cdot) : X \times \Xi \to \mathbb{R} \quad (x,\vartheta) \mapsto r(x,\vartheta) = \frac{p(x,\vartheta)}{\tilde{p}(f(x),\vartheta)}.$$

f *sufficient statistic* of S, if for every $x \in X$ the function

$$r(x,\cdot) : \Xi \to \mathbb{R} \quad \vartheta \mapsto r(x,\vartheta)$$

is constant.

Monotonicity of Fisher matrix

Theorem

Assume that (X, S, Ξ) *is a statistical model,* $f : X \to Y$ *is a measurable surjective map and* (Y, Q, Ξ) *is the induced statistical model. For every* $\vartheta \in \Xi$ *the Fisher information matrix in* S *is* $g_S^{(F)}(\vartheta)$ *and in* Q *is* $g_Q^{(F)}(\vartheta)$*. For every* $\vartheta \in \Xi$

$$g_Q^{(F)}(\vartheta) \leq g_S^{(F)}(\vartheta). \qquad (\star)$$

Information loss: $\Delta g(\vartheta) = g_S^{(F)}(\vartheta) - g_Q^{(F)}(\vartheta)$

$$\Delta g_{ik}(\vartheta) = \int_X p(x,\vartheta) \frac{\partial \ \log r(x,\vartheta)}{\partial \vartheta_i} \frac{\partial \ \log r(x,\vartheta)}{\partial \vartheta_k} \ dx$$

Equality holds in $(\star)$ *iff* f *sufficient statistic of* S*.*

Monotonicity under Markov kernel

Definition

Assume that $X \subseteq \mathbb{R}^n$ and $Y \subseteq \mathbb{R}^m$ are connect open sets. The map

$$\kappa : X \times Y \to \mathbb{R} \quad (x, y) \mapsto \kappa(y|x)$$

is *Markov kernel* or *transition probability* if $\forall x \in X$ and $\forall y \in Y$: $\kappa(y|x) \geq 0$, and $\forall x \in X$:

$$\int_Y \kappa(y|x) \, \mathrm{d}\, y = 1.$$

Theorem

Assume that (X, S, Ξ) *is a statistical model and*

$$\kappa : X \times Y \to \mathbb{R} \qquad (x, y) \mapsto \kappa(y|x)$$

is a Markov kernel. Define $\tilde{p}(y, \vartheta) = \int_X \kappa(y|x) p(x, \vartheta) \ d x$, *and denote the set of these distributions by* (Y, Q, Ξ). *Then for every* $\vartheta \in \Xi$ *we have*

$$g_Q^{(F)}(\vartheta) \leq g_S^{(F)}(\vartheta).$$

The information loss $\Delta g(\vartheta) = g_S^{(F)}(\vartheta) - g_Q^{(F)}(\vartheta)$ *is*

$$\Delta g_{ik}(\vartheta) = \int_X p(x, \vartheta) \frac{\partial \ \log r(x, \vartheta)}{\partial \vartheta_i} \frac{\partial \ \log r(x, \vartheta)}{\partial \vartheta_k} \ d x.$$

Cramer-Rao inequality

We consider the problem of estimating unknown parameter.

Assume that a data is randomly generated subject to a probability distribution which is unknown but is assumed to be in an n dimensional statistical model.

Assume that (X, S, Ξ) is a statistical model. The measurement is a map $\mathfrak{X} : X \to \mathbb{R}^m$. ($m = 1$ is the real valued measurement)

After k measurements we estimate the parameter ϑ with an *estimator*

$$\tilde{\vartheta} : (\mathbb{R}^m)^k \to \Xi \quad (x_1, \ldots, x_k) \mapsto \tilde{\vartheta}(x_1, \ldots, x_k).$$

Assume that we have independent measurements. The expected value of $\tilde{\vartheta}$ with respect to $p^{(k)}(x,\vartheta)$ is

$$E_\vartheta(\tilde{\vartheta}) = \int_{X^k} p^{(k)}(x,\vartheta)\tilde{\vartheta}(x) \ \mathrm{d}\,x.$$

The estimator $\tilde{\vartheta}$ is *unbiased* if for every $\vartheta \in \Xi$

$$E_\vartheta(\tilde{\vartheta}) = \vartheta.$$

The *variance* of the estimator is

$$V_\vartheta(\tilde{\vartheta})_{ij} = E_\vartheta\left((\tilde{\vartheta} - E_\vartheta(\tilde{\vartheta}))_i(\tilde{\vartheta} - E_\vartheta(\tilde{\vartheta}))_j\right) =$$

$$= \int_{X^k} p^{(k)}(x,\vartheta)(\tilde{\vartheta}(x) - E_\vartheta(\tilde{\vartheta}))_i(\tilde{\vartheta}(x) - E_\vartheta(\tilde{\vartheta}))_j \ \mathrm{d}\,x.$$

Theorem (*Cramer-Rao*)

Assume that (X, S, Ξ) *is a statistical model,* $k \in \mathbb{N}^+$, $g^{(F)}$ *is the Fisher information of* $(X^k, S^{(k)}, \Xi)$, $\tilde{\vartheta}$ *is an unbiased estimator of* ϑ *and* $V_{(\vartheta)}(\tilde{\vartheta})$ *its variance. For every* $\vartheta \in \Xi$ *we have*

$$V_\vartheta(\tilde{\vartheta}) \geq \left(g^{(F)}(\vartheta)\right)^{-1}.$$

Example (Cramer-Rao inequality)

Define $X = \{0,1\}$, $\Xi =]0,1[$ and S a set of functions

$$p : X \times \Xi \to \mathbb{R} \quad (x,\vartheta) \mapsto \begin{cases} 1-\vartheta & \text{if} \quad x = 0, \\ \vartheta & \text{if} \quad x = 1. \end{cases}$$

Then (X, S, Ξ) is a statistical model. Assume that we have independent measurements $x_1, \dots, x_k$. Consider the estimator for ϑ

$$\tilde{\vartheta} : X^k \to \Xi \quad (x_1, \dots, x_k) \mapsto \frac{1}{k}\sum_{i=1}^{k} x_i.$$

$\tilde{\vartheta}$ is unbiased

$$E_\vartheta(\tilde{\vartheta}) = \sum_{i=0}^{k} \binom{k}{i} \vartheta^{k-i}(1-\vartheta)^i \frac{k-i}{k} = \vartheta.$$

Example (Cramer-Rao inequality (cont.))

The variance of $\tilde{\vartheta}$ is

$$V_\vartheta(\tilde{\vartheta}) = \sum_{i=0}^{k} \binom{k}{i} \vartheta^{k-i}(1-\vartheta)^i \left(\frac{k-i}{k} - \vartheta\right)^2 = \frac{\vartheta(1-\vartheta)}{k}.$$

The Fisher information is $g_S(\vartheta) = \frac{1}{\vartheta(1-\vartheta)}$ for k measurements is $g^{(\mathrm{F})}(\vartheta) = k g_S(\vartheta)$.
The Cramer–Rao inequality in this setting is

$$\frac{\vartheta(1-\vartheta)}{k} \geq \frac{\vartheta(1-\vartheta)}{k}.$$

So $\tilde{\vartheta}$ has the least variance.

Fisher information of a density function

Consider a density function $f : \mathbb{R}^n \to \mathbb{R}$ and the shift as a parameter

$$\tilde{f} : \mathbb{R}^n \times \mathbb{R}^n \to \mathbb{R} \quad (x, y) \mapsto \tilde{f}(x, y) = f(x + y).$$

The Fisher information of $\tilde{f}$ is

$$g_{ik}(y) = \int_{\mathbb{R}^n} \frac{1}{\tilde{f}(x, y)} \frac{\partial \tilde{f}(x, y)}{\partial y_i} \frac{\partial \tilde{f}(x, y)}{\partial y_k} \, \mathrm{d}\, x.$$

It does not depend on y, reasonable to define

$$g_{ik} = \int_{\mathbb{R}^n} \frac{1}{p(x)} \frac{\partial p(x)}{\partial x_i} \frac{\partial p(x)}{\partial x_k} \, \mathrm{d}\, x$$

as Fisher information of f.

Entropy

Definition

The *entropy* of a density function $f : X \to \mathbb{R}$

$$S(f) = -\int_X f(x) \log f(x) \, \mathrm{d}x.$$

$(0 \log 0 = 0)$

Fisher information vs. Entropy

1. Fisher information is for family of distributions and for single distributions. Entropy is for single distributions.
2. Fisher information is strictly positive, entropy could be any real number.
3. There is maximum entropy principle and minimum Fisher information principle.

Fisher information vs. Entropy

1. Fisher information is for family of distributions and for single distributions. Entropy is for single distributions.
2. Fisher information is strictly positive, entropy could be any real number.
3. There is maximum entropy principle and minimum Fisher information principle.

Fisher information vs. Entropy

1. Fisher information is for family of distributions and for single distributions. Entropy is for single distributions.
2. Fisher information is strictly positive, entropy could be any real number.
3. There is maximum entropy principle and minimum Fisher information principle.

4. The Fisher information of the density function p with single variable is

$$g = 4 \int_{\mathbb{R}} \left(\frac{\mathrm{d}\sqrt{p(x)}}{\mathrm{d}\,x} \right)^2 \mathrm{d}\,x.$$

Fisher defined the *probability amplitude* $q(x) = \sqrt{p(x)}$.

He also studied the Lagrange density

$$\mathcal{L} = 4\big(q(x)'\big)^2$$

and gave information theoretical background of potential energy. Fisher studied complex probability amplitudes too and examined the Lagrange function with kinetic energy term in the form of

$$\mathcal{L}_{\mathrm{m}} = C\nabla\psi \times \nabla\psi^*.$$

(This was written down half year later in 1926 by Schrödinger for function ψ.)

4. The Fisher information of the density function p with single variable is

$$g = 4 \int_{\mathbb{R}} \left(\frac{\mathrm{d} \sqrt{p(x)}}{\mathrm{d}\, x} \right)^2 \mathrm{d}\, x.$$

Fisher defined the *probability amplitude* $q(x) = \sqrt{p(x)}$.
He also studied the Lagrange density

$$\mathcal{L} = 4\left(q(x)'\right)^2$$

and gave information theoretical background of potential energy. Fisher studied complex probability amplitudes too and examined the Lagrange function with kinetic energy term in the form of

$$\mathcal{L}_{\mathrm{m}} = C \nabla \psi \times \nabla \psi^*.$$

(This was written down half year later in 1926 by Schrödinger for function ψ.)

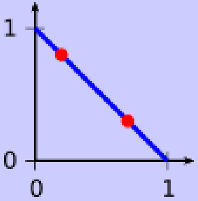
Distance of coins

What is the distance between coins $(p_1, 1-p_1)$ and $(p_2, 1-p_2)$?

In 1925 Fisher suggested the angle between vectors $(\sqrt{p_1}, \sqrt{1-p_1})$ and $(\sqrt{p_2}, \sqrt{1-p_2})$ by theoretical arguments.

Distance of coins

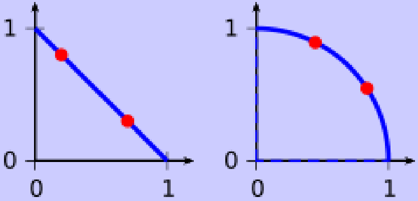
What is the distance between coins $(p_1, 1-p_1)$ and $(p_2, 1-p_2)$?



In 1925 Fisher suggested the angle between vectors $(\sqrt{p_1}, \sqrt{1-p_1})$ and $(\sqrt{p_2}, \sqrt{1-p_2})$ by theoretical arguments.

Distance of coins

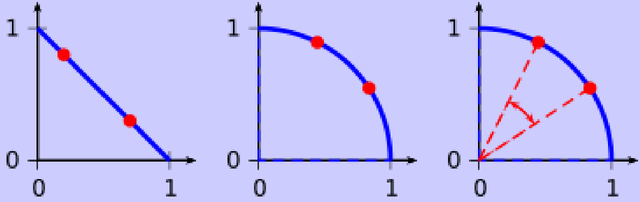
What is the distance between coins $(p_1, 1-p_1)$ and $(p_2, 1-p_2)$?



In 1925 Fisher suggested the angle between vectors $(\sqrt{p_1}, \sqrt{1-p_1})$ and $(\sqrt{p_2}, \sqrt{1-p_2})$ by theoretical arguments.

Distance of coins

What is the distance between coins $(p_1, 1-p_1)$ and $(p_2, 1-p_2)$?



In 1925 Fisher suggested the angle between vectors $(\sqrt{p_1}, \sqrt{1-p_1})$ and $(\sqrt{p_2}, \sqrt{1-p_2})$ by theoretical arguments.

The measurement based consideration is the following.
Assume that $p_1 < p_2$. If we can have n measurements then the uncertainty of measurements is the typical fluctuation

$$\Delta p = \sqrt{\frac{p(1-p)}{n}}.$$

The distributions $(p_1, 1-p_1)$ and $(p_2, 1-p_2)$ are said to be *distinguishable in n measurements* if

$$|p_1 - p_2| \geq \Delta p_1 + \Delta p_2.$$

Define $k(n, p_1, p_2)$ as the number of those probability distributions $(p_i, 1 - p_i)$ for which $p_1 < p_i < p_2$, $p_i < p_{i+1}$ and $(p_i, 1 - p_i)$ distinguishable in n measurements from $(p_{i+1}, 1 - p_{i+1})$. Let the distance be between $(p_1, 1 - p_1)$ and $(p_2, 1 - p_2)$

$$d(p_1, p_2) = \lim_{n\to\infty} \frac{k(n, p_1, p_2)}{\sqrt{n}}.$$

This gives us for distance $d(p_1, p_2)$

$$\int_{p_1}^{p_2} \frac{1}{\sqrt{p(1-p)}} \, \mathrm{d}\, p = \arccos\left(\sqrt{p_1 p_2} + \sqrt{(1-p_1)(1-p_2)}\right).$$

General contrast function

Definition

Let (X, S, Ξ) be a statistical model. A *general contrast function* is a function

$$D : S \times S \to \mathbb{R} \quad (p, q) \mapsto D(p, q)$$

if $\forall p, q \in S$: $D(p, q) \geq 0$ and $D(p, q) = 0$ iff $p = q$.

The *dual divergence* is given as $D^*(p, q) = D(q, p)$.

Let us consider some examples.

General contrast function

Definition

Let (X, S, Ξ) be a statistical model. A *general contrast function* is a function

$$D : S \times S \to \mathbb{R} \quad (p, q) \mapsto D(p, q)$$

if $\forall p, q \in S$: $D(p, q) \geq 0$ and $D(p, q) = 0$ iff $p = q$.

The *dual divergence* is given as $D^*(p, q) = D(q, p)$.

Let us consider some examples.

General contrast function

Definition

Let (X, S, Ξ) be a statistical model. A *general contrast function* is a function

$$D : S \times S \to \mathbb{R} \quad (p, q) \mapsto D(p, q)$$

if $\forall p, q \in S$: $D(p, q) \geq 0$ and $D(p, q) = 0$ iff $p = q$.

The *dual divergence* is given as $D^*(p, q) = D(q, p)$.

Let us consider some examples.

Kullback–Liebler	$D_{\mathrm{KL}}(p,q) = \int_X p(x) \log \frac{p(x)}{q(x)} \, \mathrm{d}\, x$
Hellinger	$D_{\mathrm{H}}(p,q) = \int_X \left(\sqrt{p(x)} - \sqrt{q(x)} \right)^2 \mathrm{d}\, x$
χ^2	$D_{\chi^2}(p,q) = \int_X p(x) \left[\left(\frac{p(x)}{q(x)} \right)^2 - 1 \right] \mathrm{d}\, x$
$\alpha \in]-1,1[$	$D_{\alpha}(p,q) = \frac{4}{1-\alpha^2} \left[1 - \int_X p(x)^{\frac{1-\alpha}{2}} q(x)^{\frac{1+\alpha}{2}} \, \mathrm{d}\, x \right]$
Harmonic	$D_{\mathrm{Ha}}(p,q) = 1 - \int_X \frac{2p(x)q(x)}{p(x)+q(x)} \, \mathrm{d}\, x$
Triangle	$D_{\Delta}(p,q) = \int_X \frac{(p(x)-q(x))^2}{p(x)+q(x)} \, \mathrm{d}\, x$

These distance like functions used in many areas of mathematics and applications.

For example $D_{\mathrm{KL}}(p, q)$:

- ⋆ is often called the information gain achieved if P is used instead of Q in the context of machine learning,
- ⋆ can be constructed as measuring the expected number of extra bits required to code samples from P using a code optimized for Q rather than the code optimized for P, in the context of coding theory.

Csiszár divergence

These quantities can be handled as a special cases of Csiszár divergence

Definition

Assume that $f : \mathbb{R}^+ \to \mathbb{R}$ is a strictly convex function and $f(1) = 0$. The *Csiszár divergence* is

$$D_f(p,q) = \int_X p(x) f\left(\frac{q(x)}{p(x)}\right) \, \mathrm{d}\, x.$$

For the function $f^\setminus(u) = uf(u^{-1})$ we have

$$D_f(p,q) = D_{f^\setminus}(q,p).$$

α-divergence

If $\alpha \in \mathbb{R}$ and

$$f_\alpha : \mathbb{R} \to \mathbb{R} \quad x \mapsto \begin{cases} \frac{4}{1-\alpha^2}\left(1 - x^{\frac{1+\alpha}{2}}\right) & \text{if} \quad \alpha \neq \pm 1 \\ x \log x & \text{if} \quad \alpha = 1 \\ -\log x & \text{if} \quad \alpha = -1 \end{cases}$$

then $D_{f_{-1}} = D_{\mathrm{KL}}$, $D_{f_0} = 2D_{\mathrm{H}}$ and in the $\alpha \neq \pm 1$ case $D_{f_\alpha} = D_\alpha$.

The Csiszár divergence D_f is monotone and jointly convex.

Theorem

For probability functions $p, q : X \to \mathbb{R}$ and Markov kernel $\kappa : X \times Y \to \mathbb{R}$ define $\tilde{p}(y) = \int_X \kappa(y|x) p(x)\, dx$ and $\tilde{q}(y) = \int_X \kappa(y|x) q(x)\, dx$. For the Csiszár divergences we have

$$D_f(\tilde{p}, \tilde{q}) \leq D_f(p, q).$$

Theorem

For density functions $p_1, p_2, q_1, q_2 : X \to \mathbb{R}$ and parameter $0 \leq \lambda_1 \leq 1$, $\lambda_2 = 1 - \lambda_1$

$$D_f(\lambda_1 p_1 + \lambda_2 p_2, \lambda_1 q_1 + \lambda_2 q_2) \leq \lambda_1 D_f(p_1, q_1) + \lambda_2 D_f(p_2, q_2)$$

holds.

The Csiszár divergence D_f is monotone and jointly convex.

Theorem

For probability functions $p, q : X \to \mathbb{R}$ *and Markov kernel* $\kappa : X \times Y \to \mathbb{R}$ *define* $\tilde{p}(y) = \int_X \kappa(y|x)p(x)\,dx$ *and* $\tilde{q}(y) = \int_X \kappa(y|x)q(x)\,dx$*. For the Csiszár divergences we have*

$$D_f(\tilde{p}, \tilde{q}) \leq D_f(p, q).$$

Theorem

For density functions $p_1, p_2, q_1, q_2 : X \to \mathbb{R}$ *and parameter* $0 \leq \lambda_1 \leq 1$, $\lambda_2 = 1 - \lambda_1$

$$D_f(\lambda_1 p_1 + \lambda_2 p_2, \lambda_1 q_1 + \lambda_2 q_2) \leq \lambda_1 D_f(p_1, q_1) + \lambda_2 D_f(p_2, q_2)$$

holds.

A general contrast function D (in some cases) has series expansion. From now assume that for every $\vartheta \in \Xi$ the function $y \mapsto D(p(x,\vartheta+y),p(x,\vartheta))$ has series expansion with respect to y.

$$D(p(x,\vartheta+y),p(x,\vartheta)) = \sum_{i,k=1}^{n} g_{ik}^{(D)}(p)\frac{y_i y_k}{2} + \sum_{i,j,k=1}^{n} h_{ijk}^{(D)}\frac{y_i y_j y_k}{6} + o(\|y\|^3)$$

Definition

We call D to *divergence* or *contrast function* if for every $\vartheta \in \Xi$ the function $D(p(x,\vartheta+y),p(x,\vartheta))$ has series expansion with respect to y and second order term $g_{ik}^{(D)}$ is positive definite.

Theorem

We have the following equalities for the series expansion of divergences.

$$g^{(D_{KL})} = g^{(F)} \qquad g^{(D_H)} = \frac{1}{2} g^{(F)} \qquad g^{(D_{\chi^2})} = 2 g^{(F)}$$

$$g^{(D_\alpha)} = g^{(F)} \qquad g^{(D_B)} = \frac{1}{4} g^{(F)} \qquad g^{(D_{Ha})} = \frac{1}{2} g^{(F)}$$

$$g^{(D_J)} = 2 g^{(F)} \qquad g^{(D_\Delta)} = g^{(F)} \qquad g^{(D_{LW})} = \frac{1}{4} g^{(F)}$$

$$g^{(D_f)} = f''(1) g^{(F)}$$

Differential geometry, Riemannian metric

Definition

$(M, \mathcal{A})$ is an n dimensional manifold if

1. M is a Hausdorff topological space with countable base,
2. $\mathcal{A}$ is countable and its elements are homeomorphisms $\phi_i : U_i \to V_i$, where $U_i \subseteq M$ and $V_i \subseteq \mathbb{R}^n$ are open sets,
3. for every pair of functions $\phi_i, \phi_j \in \mathcal{A}$ the map
$$\phi_i \circ \phi_j^{-1} : \phi_j(U_i \cap U_j) \to \phi_i(U_i \cap U_j)$$
is in C^∞,
4. every $x \in M$ point is contained in some U_i.

Assume that M is an n dimensional manifold and $p \in M$.

Denote by $\mathcal{F}_p$ the set of smooth functions defined in a neighbourhood of p.

A derivation is a map

$$D : \mathcal{F}_p \to \mathbb{R}$$

such that for every $a, b \in \mathbb{R}$ and functions $f, g \in \mathcal{F}$

$$D(af + bg) = aD(f) + bD(g) \qquad D(fg) = f(p)D(g) + D(f)g(p)$$

holds.
The set of derivations denoted by T_pM and called *tangent space*.

The *tangent bundle* is $TM = \bigcup_{p \in M} \{p\} \times T_p M$.

A *vector field* is a map

$$X : M \to \bigcup_{p \in M} T_p M \quad p \mapsto X(p)$$

if

1. for every $p \in M$: $X(p) \in T_p M$,
2. for every $p \in M$ and $f \in \mathcal{F}_p$ the function

$$Xf : \mathrm{Dom}(X) \cap \mathrm{Dom}(f) \to \mathbb{R} \quad p \mapsto X(p)f$$

is smooth.

The set of vector fields is denoted by $\mathcal{X}(M)$.

Definition

A map

$$g : M \to \bigcup_{p \in M} \mathrm{Lin}(T_pM \times T_pM, \mathbb{R})$$

is *Riemannian metric* if

1. for every $p \in M$ the map $g_p : T_pM \times T_pM \to \mathbb{R}$ is a scalar product,
2. for every vector field $X \in \mathcal{X}(M)$ the function

$$g(X, X) : M \to \mathbb{R} \quad p \mapsto g_p(X_p, X_p)$$

   is smooth.

The pair (M, g) is called *Riemannian geometry* or *Riemannian manifold*.

Assume that $p \in M$ and $\varphi : U \to \mathbb{R}^n$ is a local coordinate system around p. For every $f \in \mathcal{F}_p$ define $(i = 1, \ldots, n)$

$$\partial_i f = \frac{\partial (f \circ \varphi^{-1})}{\partial x_i}(\varphi(p)).$$

We consider $(\partial_1, \ldots, \partial_n)$ as a basis of $T_p M$. The Riemannian metric in this coordinate system can be described with the

$$g_{ij} = g(\partial_i, \partial_j).$$

matrix.

Covariant derivative

The map

$$\nabla : \mathcal{X}(M) \times \mathcal{X}(M) \to \mathcal{X}(M) \quad (X, Y) \mapsto \nabla_X Y$$

is a *covariant derivative* if

1. for every vector field $X, Y, Z \in \mathcal{X}(M)$

$$\nabla_{X+Y} Z = \nabla_X Z + \nabla_Y Z, \qquad \nabla_X (Y + Z) = \nabla_X Y + \nabla_X Z$$

2. for every vector field $X, Y \in \mathcal{X}(M)$ and function $f \in \mathcal{F}(M)$

$$\nabla_{fX} Y = f \nabla_X Y, \qquad \nabla_X (fY) = (Xf) Y + f \nabla_X Y.$$

Assume that $p \in M$ and $\varphi : U \to \mathbb{R}^n$ is a local coordinate system around p. The covariant derivative can be described by *Christoffel symbol of the first kind*

$$\Gamma_{ijk} = g(\nabla_{\partial_i}\partial_j, \partial_k)$$

and by *Christoffel symbol of the second kind*

$$\Gamma_{ij}^{\cdot\cdot k}\partial_k = \nabla_{\partial_i}\partial_j.$$

Levi-Civita covariant derivative

The pair (M, ∇) is called to be an *affine manifold*.
The affine manifold (M, ∇) called *torsion free* if $\Gamma_{ij}^{\cdot\cdot k} = \Gamma_{ji}^{\cdot\cdot k}$ holds in every local coordinate system.
The covariant derivative ∇ on a (M, g) Riemannian manifold called *Riemannian covariant derivative* if for every vector field $X, Y, Z \in \mathcal{X}(M)$

$$Xg(Y, Z) = g(\nabla_X Y, Z) + g(Y, \nabla_X Z).$$

The covariant derivative ∇ on a (M, g) Riemannian manifold called *Levi-Civita covariant derivative* if torsion free Riemannian covariant derivative.

Theorem

For every (M, g) Riemannian manifold there exists a unique Levi–Civita covariant derivative ∇, which can be expressed as

$$\Gamma_{ij}^{\cdot\cdot m} = g^{km}\frac{1}{2}(\partial_i g_{jk} + \partial_j g_{ik} - \partial_k g_{ij}).$$

in local coordinate systems.

Curvature

Definition

For an affine manifold (M, ∇) define the *curvature* as

$$R : \mathcal{X}(M) \times \mathcal{X}(M) \times \mathcal{X}(M) \to \mathcal{X}(M) \quad (X, Y, Z) \mapsto R(X, Y)Z$$

$$R(X, Y)Z = \nabla_X \nabla_Y Z - \nabla_Y \nabla_X Z - \nabla_{[X,Y]} Z.$$

The affine manifold (M, ∇) is *flat* if $R = 0$.

In a local coordinate system the curvature tensor can be handled by the

$$R(\partial_i, \partial_j)\partial_k = R_{ijk}^{\cdot\cdot\cdot l}\partial_l,$$

$$g(R(\partial_i, \partial_j)\partial_k, \partial_l) = R_{ijkl}$$

quantities.

The curvature tensor has symmetries

$$R_{ijkl} = -R_{jikl}, \quad R_{ijkl} = -R_{ijlk}, \quad R_{ijkl} = R_{klij}.$$

One can compute the curvature tensor as

$$R_{ijk}^{\cdot\cdot\cdot l} = \partial_i \Gamma_{jk}^{\cdot\cdot l} - \partial_j \Gamma_{ik}^{\cdot\cdot l} + \Gamma_{jk}^{\cdot\cdot m}\Gamma_{im}^{\cdot\cdot l} - \Gamma_{ik}^{\cdot\cdot m}\Gamma_{jm}^{\cdot\cdot l}.$$

Definition

For an (M, ∇) affine manifold with curvature R the function

$$\mathrm{Ric} : \mathcal{X}(M) \times \mathcal{X}(M) \to \mathcal{F}(M) \quad (X, Y) \mapsto \mathrm{Tr}\big(Z \mapsto R(Z, X)Y\big)$$

is called *Ricci curvature*.

In local coordinate system the matrix

$$\mathrm{Ric}_{ij} = \mathrm{Ric}(\partial_i, \partial_j)$$

can be computed as

$$\mathrm{Ric}_{jk} = R_{ijk}^{\cdot\cdot\cdot i}.$$

Length and volume

Assume that (M, g) is a Riemannian manifold and $\gamma :]a, b[\to M$ is a smooth curve. The *length* of the curve defined as

$$l_\gamma(a, b) = \int_a^b \sqrt{g(\dot{\gamma}(t), \dot{\gamma}(t))} \, \mathrm{d}\, t.$$

The *volume* of the set $U \subseteq \mathrm{Dom}(\phi)$

$$V(U) = \int_{\phi(U)} \sqrt{\det g}.$$

Geodesic line

A smooth curve $\gamma :]a, b[\to M$ is called to be a *geodesic line* if in local coordinate systems

$$\frac{\mathrm{d}^2 \gamma^k}{\mathrm{d}\, t^2} + \sum_{i,j=1}^{\dim M} (\Gamma_{ij}^{\cdot\cdot k} \circ \gamma) \frac{\mathrm{d}\, \gamma^i}{\mathrm{d}\, t} \frac{\mathrm{d}\, \gamma^j}{\mathrm{d}\, t} = 0$$

holds.

Information geometry basics

Consider a statistical model (X, S, Ξ).

The manifold $M = \Xi$, open connected subset of $\mathbb{R}^n$.

The Riemannian metric $g = g^{(\mathrm{F})}$ is the Fisher information.

We can compute the Levi–Civita covariant derivative or define new ones.

Information geometry basics

Consider a statistical model (X, S, Ξ).

The manifold $M = \Xi$, open connected subset of $\mathbb{R}^n$.

The Riemannian metric $g = g^{(\mathrm{F})}$ is the Fisher information.

We can compute the Levi–Civita covariant derivative or define new ones.

Information geometry basics

Consider a statistical model (X, S, Ξ).

The manifold $M = \Xi$, open connected subset of $\mathbb{R}^n$.

The Riemannian metric $g = g^{(\mathrm{F})}$ is the Fisher information.

We can compute the Levi–Civita covariant derivative or define new ones.

Information geometry basics

Consider a statistical model (X, S, Ξ).

The manifold $M = \Xi$, open connected subset of $\mathbb{R}^n$.

The Riemannian metric $g = g^{(\mathrm{F})}$ is the Fisher information.

We can compute the Levi–Civita covariant derivative or define new ones.

In 1945, Rao suggested to consider the Fisher information as Riemannian metric.

In 1975, Efron studied first the curvature of statistical manifolds.

In 1979, Ruppeiner claimed that thermodynamic systems can be represented by Riemannian geometry, and that statistical properties can be derived from the model. (For example he found connection between the behaviour of correlation functions and curvature at second order phase transitions.)

In 1999, Brody and Ritz studied the curvature of statistical model of Ising chains.

In 1945, Rao suggested to consider the Fisher information as Riemannian metric.

In 1975, Efron studied first the curvature of statistical manifolds.

In 1979, Ruppeiner claimed that thermodynamic systems can be represented by Riemannian geometry, and that statistical properties can be derived from the model. (For example he found connection between the behaviour of correlation functions and curvature at second order phase transitions.)

In 1999, Brody and Ritz studied the curvature of statistical model of Ising chains.

In 1945, Rao suggested to consider the Fisher information as Riemannian metric.

In 1975, Efron studied first the curvature of statistical manifolds.

In 1979, Ruppeiner claimed that thermodynamic systems can be represented by Riemannian geometry, and that statistical properties can be derived from the model. (For example he found connection between the behaviour of correlation functions and curvature at second order phase transitions.)

In 1999, Brody and Ritz studied the curvature of statistical model of Ising chains.

In 1945, Rao suggested to consider the Fisher information as Riemannian metric.

In 1975, Efron studied first the curvature of statistical manifolds.

In 1979, Ruppeiner claimed that thermodynamic systems can be represented by Riemannian geometry, and that statistical properties can be derived from the model. (For example he found connection between the behaviour of correlation functions and curvature at second order phase transitions.)

In 1999, Brody and Ritz studied the curvature of statistical model of Ising chains.

Alpha covariant derivatives

Definition

Consider the $\mathcal{P}_n$ set. For every $-1 \leq \alpha \leq 1$ define

$$\Gamma_{ijk}^{(\alpha)} = \sum_{l=0}^{n} p(l, \underline{\vartheta}) \Big(\partial_i \partial_j (\log p(l, \underline{\vartheta})) + \frac{1-\alpha}{2} (\partial_i \log p(l, \underline{\vartheta}))(\partial_j \log p(l, \underline{\vartheta}))(\partial_k \log p(l, \underline{\vartheta})) \Big),$$

which is called *α-covariant derivative*.

Theorem

The 0-covariant derivative is Levi-Civita covariant derivative.

Example (Geodesic line in $\mathcal{P}_1$)

In the space $(\mathcal{P}_1, \nabla)$ γ is geodesic line iff

$$\frac{\mathrm{d}^2 \gamma(t)}{\mathrm{d}\, t^2} - \frac{(1-2\gamma(t))}{2\gamma(t)(1-\gamma(t))} \left(\frac{\mathrm{d}\, \gamma(t)}{\mathrm{d}\, t} \right)^2 = 0.$$

The solution (with initial values $\gamma(0) = a$ and $\dot{\gamma}(0) = b$) is

$$\gamma(t) = \cos^2 \left(\frac{bt}{2\sqrt{a}\sqrt{1-a}} + \arccos \sqrt{a} \right).$$

Example (Normal distribution)

Let us define the base set $X = \mathbb{R}$, the parameter space $\Xi = \mathbb{R} \times \mathbb{R}^+$ and the elements of S as

$$p(x, \mu, \sigma) = \frac{1}{\sqrt{\pi}\sigma} \exp\left(-\frac{(x-\mu)^2}{\sigma^2}\right), \qquad (\mu, \sigma) \in \Xi.$$

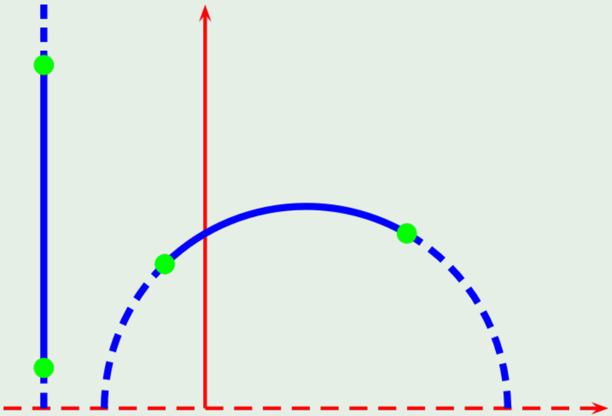
Using the coordinate system (μ, σ) the Fisher information of the statistical model (X, S, Ξ) is

$$(g_{ik}^{(\mathrm{F})}) = \begin{pmatrix} \frac{2}{\sigma^2} & 0 \\ 0 & \frac{2}{\sigma^2} \end{pmatrix}.$$

The pair $(\Xi, g^{(\mathrm{F})})$ is special Riemannian geometry, called *hyperbolic plane*.

Example (Normal distribution cont.)

The geodesic curves are those semicircles whose centre lies on the axis μ and the $\mu = constant$ half lines.



Example (Normal distribution cont.)

Consider the distributions given by parameters (μ_1, σ_1) and (μ_2, σ_2) $(\mu_1 \le \mu_2)$, where $\mu_1 \le \mu_2$. If $\mu_1 < \mu_2$ then define the parameters

$$R = \sqrt{\left(\frac{\mu_2 - \mu_1}{2}\right)^2 + \frac{\sigma_1^2 + \sigma_2^2}{2} + \left(\frac{\sigma_2^2 - \sigma_1^2}{2(\mu_2 - \mu_1)}\right)^2},$$

$$C = \frac{\mu_1 + \mu_2}{2} + \frac{\sigma_2^2 - \sigma_1^2}{2(\mu_2 - \mu_1)}.$$

The geodesic curve connecting the points (μ_1, σ_1) and (μ_2, σ_2) is the $(\mu - C)^2 + \sigma^2 = R^2$ semicircle $(\sigma > 0)$.

Example

Normal distribution cont. The geodesic distance between the points is the following.

1. If $(\mu_1 - \mu_2)^2 \leq \left|\sigma_2^2 - \sigma_1^2\right|$ then
$$d\Big((\mu_1, \sigma_1), (\mu_2, \sigma_2)\Big) = \sqrt{2}\left|\operatorname{arch}\frac{R}{\sigma_1} - \operatorname{arch}\frac{R}{\sigma_2}\right|.$$
2. If $(\mu_1 - \mu_2)^2 \geq |\sigma_1^2 - \sigma_2^2|$ then
$$d\Big((\mu_1, \sigma_1), (\mu_2, \sigma_2)\Big) = \sqrt{2}\left(\operatorname{arch}\frac{R}{\sigma_1} + \operatorname{arch}\frac{R}{\sigma_2}\right).$$
3. If $\mu_1 = \mu_2$ then $d\Big((\mu_1, \sigma_1), (\mu_2, \sigma_2)\Big) = \sqrt{2}\left|\log\frac{\sigma_1}{\sigma_2}\right|.$

Pull-back metric

Assume that $\varphi : M \to N$ is a smooth map between differentiable manifolds.
For every $p \in M$ we have maps

$$\varphi_1 : \mathcal{F}^N_{\varphi(p)} \to \mathcal{F}^M_p \qquad f \mapsto f \circ \varphi$$

and

$$\varphi_* : T_pM \to T_{\varphi(p)}N \qquad v \mapsto v \circ \varphi_1.$$

Definition

If (N, g) is a Riemannian manifold then we can define the *pull-back metric* on M as

$$g^M_p(x, y) = g^N_{\varphi(p)}(\varphi_*(x), \varphi_*(y)).$$

Theorem

The pull back metric of the euclidean metric by the map

$$\mathcal{P}_n \to \mathbb{R}^{n+1} \qquad (p_1, \dots, p_n) \mapsto \left(\sqrt{1 - \sum_{k=1}^{n} p_k}, \sqrt{p_1}, \dots, \sqrt{p_n} \right)$$

is the Fisher metric.

Theorem

The volume of the space $\mathcal{P}_n$ equals to the surface of the $n+1$ dimensional ball divided by 2^{n+1}, that is

$$V(\mathcal{P}_n) = \frac{\pi^{(n+1)/2}}{2^n \Gamma\left(\frac{n+1}{2}\right)}.$$

Uniqueness of Fisher metric

Theorem

Let us define $X_n = \{0, 1, \ldots, n\}$ $(n \in \mathbb{N}^+)$. Assume than for every n a Riemannian metric g_n is given on $\mathcal{P}_n$. For a $\kappa : X_n \times X_m \to \mathbb{R}$ transition probability denote by $\tilde{\kappa} : \mathcal{P}_n \to \mathcal{P}_m$. If for every transition probability $\kappa : X_n \times X_m \to \mathbb{R}$ for every point $p \in \mathcal{P}_n$ for every tangent vector $X \in T_p\mathcal{P}_n$

$$g_{\kappa(p)}(\tilde{\kappa}_*(X), \tilde{\kappa}_*(X)) \leq g_p(X, X)$$

holds then there exists a unique positive number c such that for every $n \in \mathbb{N}^+$ $g_n = c g_n^{(F)}$.

Duality on Riemannian manifolds

Definition

For an (M, g) Riemannian geometry the covariant derivatives ∇ and ∇^* are called *dual covariant derivatives* if for every vector field $X, Y, Z \in \mathcal{X}(M)$

$$Zg(X, Y) = g(\nabla_Z X, Y) + g(X, \nabla^*_Z Y)$$

holds. We call (M, g, ∇, ∇^*) *dual Riemannian geometry*.

Theorem

Consider a statistical model (X, S, Ξ) with Fisher metric g. For all $\alpha \in [-1, 1]$ the covariant derivatives $\nabla^{(\alpha)}$ and $\nabla^{(-\alpha)}$ are torsion free and dual.

Theorem

Assume that (M, g, ∇, ∇^*) torsion free dual geometry with curvatures R and R^*. In this case $R = 0$ iff $R^* = 0$.

In this case we call (M, g, ∇, ∇^*) *flat dual Riemannian geometry*.

From divergence to duality

Assume that M is an n dimensional manifold, $D : M \times M \to \mathbb{R}$ is a divergence, $\vartheta \in M$, ϕ is a local coordinate system in a neighbourhood of p. Consider the function

$$D^{(\vartheta,\phi)} : \mathbb{R}^n \to \mathbb{R} \qquad y \mapsto D(\vartheta, \phi^{-1}(\phi(\vartheta) + y))$$

and its series expansion

$$D^{(\vartheta,\phi)}(y) = \frac{1}{2} \sum_{i,k=1}^{n} g_{ik}^{(D)}(\vartheta) y_i y_k + \frac{1}{6} \sum_{i,j,k=1}^{n} h_{ijk}^{(D)}(\vartheta) y_i y_j y_k + o(\|y\|^3).$$

At every point $\vartheta \in M$ the matrix $g^{(D)}(\vartheta)$ is positive definite, so $(M, g^{(D)})$ is Riemannian geometry. From the third order term define

$$\Gamma_{ijk}^{(D)} = h_{ijk}^{(D)} - \partial_k g_{ij}^{(D)} \qquad i, j, k \in \{1, 2, \ldots, n\}.$$

Theorem (From divergence to duality)

Assume that M is an n dimensional manifold, D is a divergence on M and we have the induced quantities $g^{(D)}$, $\Gamma_{ijk}^{(D)}$ and $\Gamma_{ijk}^{(D^*)}$. In this case $\Gamma_{ijk}^{(D)}$ and $\Gamma_{ijk}^{(D^*)}$ can be considered as a Christoffel symbols of the first kind of torsion free covariant derivatives $\nabla^{(D)}$ and $\nabla^{(D^*)}$. Moreover $(M, g, \nabla^{(D)}, \nabla^{(D^*)})$ is a torsion free dual geometry.

Theorem

If (M, g, ∇, ∇^*) is a torsion free dual geometry then there exists a D divergence which induces the same duality.

From duality to divergence

Definition

If (M, ∇) is an affine manifold, $x \in M$ and ϕ and ϑ are local coordinate systems of a neighbourhood of x. We call ϕ to *affine coordinate system* if for all $1 \leq i, j \leq \dim M$

$$\nabla_{\partial_i} \partial_j = 0$$

holds and we call ϕ and ϑ *dual coordinate systems* if

$$g(x)(\partial_i^{(\vartheta)}, \partial_j^{(\eta)}) = \delta_{ij}.$$

Theorem (From duality to divergence)

Assume that (M, g, ∇, ∇^*) is a flat dual n dimensional geometry. Then every point $x \in M$ has a neighbourhood $U \subseteq M$ with dual coordinate systems ϑ and η. Assume that $U = M$.

1. *In this case there exists a function $\psi : M \to \mathbb{R}$ such that for every $1 \leq i \leq n$*
$$\partial_i^{(\vartheta)} \psi = \eta_i.$$
2. *For the function*
$$\phi : M \to \mathbb{R} \qquad x \mapsto \phi(x) = \sum_{i=1}^{n} \vartheta_i(x)\eta_i(x) - \psi(x)$$
we have
$$\partial_i^{(\eta)} \phi = \vartheta_i \quad 1 \leq i \leq n.$$

Theorem (From duality to divergence cont.)

3. *For every indices* $1 \leq i, j \leq n$

$$g(\partial_i^{(\vartheta)}, \partial_j^{(\vartheta)}) = \partial_i^{(\vartheta)} \partial_j^{(\vartheta)} \psi \qquad g(\partial_i^{(\eta)}, \partial_j^{(\eta)}) = \partial_i^{(\eta)} \partial_j^{(\eta)} \phi.$$

4. *The functions* ψ, ϕ *has extrema for every* $x \in M$

$$\phi(x) = \max_{y \in M} \left(\sum_{i=1}^{n} \vartheta_i(y) \eta_i(x) - \psi(y) \right)$$
$$\psi(x) = \max_{y \in M} \left(\sum_{i=1}^{n} \vartheta_i(x) \eta_i(y) - \phi(y) \right).$$

Theorem (From duality to divergence cont.)

5. *The functions ϕ and ψ are strictly convex functions of $(\eta_1, \ldots, \eta_n)$ and $(\vartheta_1, \ldots, \vartheta_n)$ respectively.*
6. *We have a canonical divergence $D : M \times M \to \mathbb{R}$*

$$D^{(g,\nabla)}(p, q) = \psi(p) + \phi(q) - \sum_{i=1}^{n} \vartheta^i(p)\eta^i(q).$$

Example (Duality for discrete distribution)

Base space is $X = \{0, 1, \ldots, n\}$ and the parameter space is $\Xi = \left\{ (p_1, \ldots, p_n) \in (\mathbb{R}^+)^n \,\middle|\, \sum_{k=1}^n p_k < 1 \right\}$. The Fisher metric is g.

The covariant derivatives $\nabla^{(-1)}$ and $\nabla^{(1)}$ are torsion free and $(\Xi, g, \nabla^{(1)}, \nabla^{(-1)})$ is flat dual geometry.

Let us define the following coordinate systems

$$\eta : \Xi \to \mathbb{R}^n \qquad p \mapsto \eta(p) = (p_1, \ldots, p_n)$$

$$\vartheta : \Xi \to \mathbb{R}^n \qquad p \mapsto \vartheta(p) = \left(\log \frac{p_1}{p_0}, \ldots, \log \frac{p_n}{p_0} \right),$$

where $p_0 = 1 - \sum_{k=1}^n p_k$.

Example (Duality for discrete distribution cont.)

The coordinate systems η and ϑ are affine for $(\Xi, \nabla^{(-1)})$ and $(\Xi, \nabla^{(1)})$.
($\nabla^{(1)}$ called *exponential covariant derivative* and $\nabla^{(-1)}$ called *mixture covariant derivative*.)
If we use the potential function

$$\psi : \Xi \to \mathbb{R} \qquad p \mapsto -\log p_0$$

then we have

$$\partial_i^{(\vartheta)} \psi(p) = \eta_i$$

Example (Duality for discrete distribution cont.)

The function ϕ is the following

$$\phi(p) = \sum_{i=0}^{n} p_i \log p_i = -S(p).$$

The canonical divergence of the $(\Xi, g, \nabla^{(1)}, \nabla^{(-1)})$ flat dual geometry is

$$D^{(g,\nabla)}(p,q) = \psi(p) + \phi(q) - \sum_{i=1}^{n} \vartheta_i(p)\eta_i(q)$$

$$= \sum_{i=0}^{n} q_i \log \frac{q_i}{p_i} = D_{\mathrm{KL}}(q,p).$$

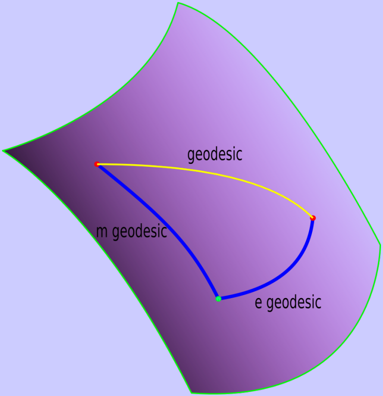
Pythagorean theorem

Theorem

Assume that (M, g, ∇, ∇^*) *is a flat dual geometry,* $a, b, c \in M$, γ_1 *is a* ∇ *geodesic curve connecting* a *and* b, γ_2 *is a* ∇^* *geodesic curve connecting* b *and* c *such that* $g(b)(\dot\gamma_1(b), \dot\gamma_2(b)) = 0$. *Then*

$$D^{(g,\nabla)}(a, c) = D^{(g,\nabla)}(a, b) + D^{(g,\nabla)}(b, c).$$

Pythagorean theorem



Projection theorem

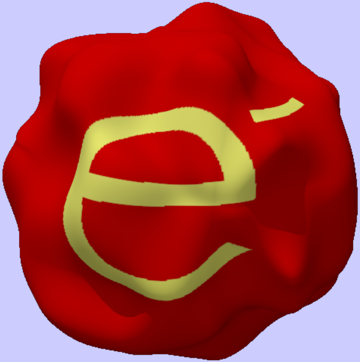
Theorem

Assume that (M, g, ∇, ∇^*) is a flat dual geometry, N is a submanifold of M and $x \in M \setminus N$. The point $y \in N$ is a critical point of the function

$$N \to \mathbb{R} \qquad y \mapsto D^{(g,\nabla)}(x, y)$$

iff the geodesic line between x and y is perpendicular to N.

Quantum mechanical setting



Quantum mechanical setting

In quantum setting we use n dimensional Hilbert space.

A self-adjoint, positive semidefinite trace one operator: *state*.

The set of states is called to be *state space*.

The interior of the state space is denoted by $\mathcal{M}_n^+$.

The extremal points of the state space: *pure states*.

A self-adjoint operator is called *observable*.

The *expected value* of an observable A in a state D is $\mathrm{Tr}(DA)$.

Example (2 dimensional Hilbert space (qubit))

Every state $D \in \mathcal{M}_2$ can be uniquely written in the form of

$$D = \frac{1}{2}\begin{pmatrix} 1+z & x+\mathrm{i}\,y \\ x-\mathrm{i}\,y & 1-z \end{pmatrix}. \qquad (\star\star)$$

For states we have

$$x^2 + y^2 + z^2 \leq 1$$

and for parameters $(x, y, z) \in \mathbb{R}^3$ equation $(\star\star)$ defines a state iff $x^2 + y^2 + z^2 \leq 1$.
Therefore the state space of a two dimensional quantum system can be identified with the closed unit ball in $\mathbb{R}^3$.
(x, y, z) are called to be *Stokes parameters*.

Entropy

The entropy of a state D can be defined as in the classical case

$$S(D) = -\operatorname{Tr} D \log D,$$

called *Neumann entropy*.
The entropy is a concave function.

Theorem

For every state $D_1, D_2 \in \mathcal{M}_n^+$ and parameter $\lambda \in [0,1]$

$$\lambda S(D_1) + (1-\lambda) S(D_2) \leq S(\lambda D_1 + (1-\lambda) D_2).$$

Riemannian metric on state space

We will refer to $\mathcal{M}_n^+$ as open convex subset of $\mathbb{R}^k$ with its canonical coordinate system. At a given point $D_0 \in \mathcal{M}_n^+$ we

identify the tangent space with $n \times n$ self-adjoint trace zero operators $\mathcal{M}_n$. For a given smooth function $f : \mathcal{M}_n^+ \to \mathbb{R}$ at a state $D_0 \in \mathcal{M}_n^+$ the effect of the tangent vector $X \in \mathcal{M}_n$ is

$$(Xf)(D_0) = \left. \frac{\mathrm{d}\, f(D_0 + tX)}{\mathrm{d}\, t} \right|_{t=0} .$$

We denote by $T_D\mathcal{M}_n^+$ the tangent space of $\mathcal{M}_n^+$ at a point $D \in \mathcal{M}_n^+$.

We can define Riemannian metrics on $\mathcal{M}_n^+$, for example

$$K_D(X, Y) = \mathrm{Tr}\, DXY \qquad D \in \mathcal{M}_n^+ \; X, Y \in T_M\mathcal{M}_n^+$$

is a Riemannian metric.

Problems with Fisher metric:
How to generalise equations like below?

$$g^{(\mathrm{F})}(\vartheta)_{ik} = \int_X p(x, \vartheta)(\partial_i \log p(x, \vartheta))(\partial_k \log p(x, \vartheta)) \;\mathrm{d}\, x$$

$$g^{(\mathrm{F})}(\vartheta)_{ik} = 4 \int_X (\partial_i \sqrt{p(x, \vartheta)})(\partial_k \sqrt{p(x, \vartheta)}) \;\mathrm{d}\, x$$

We can define Riemannian metrics on $\mathcal{M}_n^+$, for example

$$K_D(X,Y) = \operatorname{Tr} DXY \qquad D \in \mathcal{M}_n^+ \; X, Y \in T_M \mathcal{M}_n^+$$

is a Riemannian metric.

Problems with Fisher metric:
How to generalise equations like below?

$$g^{(\mathrm{F})}(\vartheta)_{ik} = \int_X p(x,\vartheta)(\partial_i \log p(x,\vartheta))(\partial_k \log p(x,\vartheta)) \,\mathrm{d}\, x$$

$$g^{(\mathrm{F})}(\vartheta)_{ik} = 4 \int_X (\partial_i \sqrt{p(x,\vartheta)})(\partial_k \sqrt{p(x,\vartheta)}) \,\mathrm{d}\, x$$

There was the concepts of *left* and *right logarithmic derivative*

$$\frac{\mathrm{d}\, D_\vartheta}{\mathrm{d}\,\vartheta} = D_\vartheta \times L_{r,\vartheta} \qquad \frac{\mathrm{d}\, D_\vartheta}{\mathrm{d}\,\vartheta} = L_{l,\vartheta} \times D_\vartheta .$$

The second derivative of the entropy generates a Riemannian metric too.

The pull back of the euclidean metric by

$$\mathcal{M}_n^+ \to \mathbb{R}^k \qquad D \mapsto \sqrt{D}$$

defines Riemannian metric too.

There was the concepts of *left* and *right logarithmic derivative*

$$\frac{\mathrm{d}\, D_\vartheta}{\mathrm{d}\, \vartheta} = D_\vartheta \times L_{r,\vartheta} \qquad \frac{\mathrm{d}\, D_\vartheta}{\mathrm{d}\, \vartheta} = L_{l,\vartheta} \times D_\vartheta .$$

The second derivative of the entropy generates a Riemannian metric too.

The pull back of the euclidean metric by

$$\mathcal{M}_n^+ \to \mathbb{R}^k \qquad D \mapsto \sqrt{D}$$

defines Riemannian metric too.

There was the concepts of *left* and *right logarithmic derivative*

$$\frac{\mathrm{d}\, D_\vartheta}{\mathrm{d}\,\vartheta} = D_\vartheta \times L_{r,\vartheta} \qquad \frac{\mathrm{d}\, D_\vartheta}{\mathrm{d}\,\vartheta} = L_{l,\vartheta} \times D_\vartheta .$$

The second derivative of the entropy generates a Riemannian metric too.

The pull back of the euclidean metric by

$$\mathcal{M}_n^+ \to \mathbb{R}^k \qquad D \mapsto \sqrt{D}$$

defines Riemannian metric too.

Extending some classical concept to quantum setting

Let us denote by M_n the space of $n \times n$ matrices and by $M_m(M_n)$ those $m \times m$ matrices whose elements are $n \times n$ matrices.

Definition

A linear map $T : M_n \to M_m$ is called *positive* if maps every positive operator to a positive operator.
A linear map $T : M_n \to M_m$ is called *completely positive* if for every $k \in \mathbb{N}$ the operator

$$T^{(k)} : M_k(M_n) \to M_k(M_m) \qquad [A_{ij}] \mapsto T^{(k)}([A_{ij}]) = [T(A_{ij})]$$

is positive.
We call a linear map $T : M_n \to M_m$ is called to be a *stochastic map* if completely positive and trace preserving.

Theorem

A linear map $T : M_n \to M_m$ is completely positive iff there exist operators $V_i : M_m \to M_n$ such that

$$T(A) = \sum_{i=1}^{N} V_i A V_i^* \qquad \forall A \in M_n.$$

The map T is trace preserving iff $\sum_{i=1}^{N} V_i V_i^* = I$.

Definition

Consider the family of Riemannian manifolds $(\mathcal{M}_n^+, K^{(n)})_{n\in\mathbb{N}}$. If for every $n, m \in \mathbb{N}$, stochastic map $T : M_n \to M_m$, state $D \in \mathcal{M}_n^+$ and tangent vector $X \in \mathcal{M}_n$

$$K^{(m)}_{T(D)}(T(X), T(X)) \leq K^{(n)}_D(X, X)$$

holds then we call $(\mathcal{M}_n^+, K^{(n)})_{n\in\mathbb{N}}$ a *family of monotone metrics*.

Consider a function $f:\mathbb{R}\to\mathbb{R}$ and a self-adjoint matrix X.
How to compute $f(X)$:
– $X \in \mathcal{M}_n^+$ can be diagonalized by some unitary matrix U, that is $X = UDU^*$.

$$f(X) := Uf(D)U^*$$

– X can be written as $X = \sum_{i=1}^{n} \lambda_i E_i$, where $(\lambda_i)_{i=1,\dots,n}$ are the eigenvalues and $(E_i)_{i=1,\dots,n}$ are the corresponding projections

$$f(X) = \sum_{i=1}^{n} f(\lambda_i) E_i.$$

Definition

A function $f:\mathbb{R}\to\mathbb{R}$ called *operator monotone* if for every $n\in\mathbb{N}$ and self-adjoint matrices $A, B \in M_n$ from $A \leq B$ follows $f(A) \leq f(B)$.

Consider a function $f : \mathbb{R} \to \mathbb{R}$ and a self-adjoint matrix X.
How to compute $f(X)$:
– $X \in \mathcal{M}_n^+$ can be diagonalized by some unitary matrix U, that is $X = UDU^*$.

$$f(X) := Uf(D)U^*$$

– X can be written as $X = \sum_{i=1}^{n} \lambda_i E_i$, where $(\lambda_i)_{i=1,\dots,n}$ are the eigenvalues and $(E_i)_{i=1,\dots,n}$ are the corresponding projections

$$f(X) = \sum_{i=1}^{n} f(\lambda_i) E_i.$$

Definition

A function $f : \mathbb{R} \to \mathbb{R}$ called *operator monotone* if for every $n \in \mathbb{N}$ and self-adjoint matrices $A, B \in M_n$ from $A \leq B$ follows $f(A) \leq f(B)$.

Consider a function $f : \mathbb{R} \to \mathbb{R}$ and a self-adjoint matrix X.
How to compute $f(X)$:
– $X \in \mathcal{M}_n^+$ can be diagonalized by some unitary matrix U, that is $X = UDU^*$.

$$f(X) := Uf(D)U^*$$

– X can be written as $X = \sum_{i=1}^{n} \lambda_i E_i$, where $(\lambda_i)_{i=1,\dots,n}$ are the eigenvalues and $(E_i)_{i=1,\dots,n}$ are the corresponding projections

$$f(X) = \sum_{i=1}^{n} f(\lambda_i) E_i.$$

Definition

A function $f : \mathbb{R} \to \mathbb{R}$ called *operator monotone* if for every $n \in \mathbb{N}$ and self-adjoint matrices $A, B \in M_n$ from $A \leq B$ follows $f(A) \leq f(B)$.

Denote by $\mathrm{Lin}(M_n)$ the set of linear $A : M_n \to M_n$ maps and define the *Hilbert-Schmidt scalar product*

$$\langle \cdot, \cdot \rangle : \mathrm{Lin}(M_n) \times \mathrm{Lin}(M_n) \to \mathbb{C} \qquad (A, B) \mapsto \operatorname{Tr} A^* B.$$

For $D \in M_n$ define the *left* and the *right multiplication operators*

$$L_{n,D}(A) = DA \qquad R_{n,D}(A) = AD.$$

If $D \in \mathcal{M}_n^+$ then $L_{n,D}$ and $R_{n,D}$ are self-adjoint operator.

$$\langle L_{n,D}A, B\rangle = \langle DA, B\rangle = \operatorname{Tr}(DA)^* B = \operatorname{Tr} A^* D^* B =$$

$$= \operatorname{Tr} A^* DB = \langle A, DB\rangle = \langle A, L_{n,D}B\rangle$$

$$\langle R_{n,D}A, B\rangle = \langle AD, B\rangle = \operatorname{Tr}(AD)^* B = \operatorname{Tr} D^* A^* B =$$

$$= \operatorname{Tr} A^* BD = \langle A, BD\rangle = \langle A, R_{n,D}B\rangle$$

Basic property of means

What is a mean?

A function $M : \mathbb{R}^+ \times \mathbb{R}^+ \to \mathbb{R}^+$ is a *mean* if $(\forall x, y, x_0, y_0, t \in \mathbb{R}^+)$

Basic property of means

What is a mean?

A function $M : \mathbb{R}^+ \times \mathbb{R}^+ \to \mathbb{R}^+$ is a *mean* if $(\forall x, y, x_0, y_0, t \in \mathbb{R}^+)$

$M(x, x) = x$

Basic property of means

What is a mean?

A function $M : \mathbb{R}^+ \times \mathbb{R}^+ \to \mathbb{R}^+$ is a *mean* if $(\forall x, y, x_0, y_0, t \in \mathbb{R}^+)$

$$M(x,x) = x$$

$$M(x,y) = M(y,x)$$

Basic property of means

What is a mean?

A function $M : \mathbb{R}^+ \times \mathbb{R}^+ \to \mathbb{R}^+$ is a *mean* if $(\forall x, y, x_0, y_0, t \in \mathbb{R}^+)$

$$M(x, x) = x$$

$$M(x, y) = M(y, x)$$

$$x < y \;\Rightarrow\; x < M(x, y) < y$$

Basic property of means

What is a mean?

A function $M : \mathbb{R}^+ \times \mathbb{R}^+ \to \mathbb{R}^+$ is a *mean* if $(\forall x, y, x_0, y_0, t \in \mathbb{R}^+)$

$$M(x, x) = x$$

$$M(x, y) = M(y, x)$$

$$x < y \;\Rightarrow\; x < M(x, y) < y$$

$$x < x_0, y < y_0 \Rightarrow M(x, y) < M(x_0, y_0)$$

Basic property of means

What is a mean?

A function $M : \mathbb{R}^+ \times \mathbb{R}^+ \to \mathbb{R}^+$ is a *mean* if $(\forall x, y, x_0, y_0, t \in \mathbb{R}^+)$

$$M(x,x) = x$$

$$M(x,y) = M(y,x)$$

$$x < y \;\Rightarrow\; x < M(x,y) < y$$

$$x < x_0, y < y_0 \Rightarrow M(x,y) < M(x_0,y_0)$$

$M(x,y)$ is continuous

Basic property of means

What is a mean?

A function $M : \mathbb{R}^+ \times \mathbb{R}^+ \to \mathbb{R}^+$ is a *mean* if $(\forall x, y, x_0, y_0, t \in \mathbb{R}^+)$

$$M(x,x) = x$$

$$M(x,y) = M(y,x)$$

$$x < y \;\Rightarrow\; x < M(x,y) < y$$

$$x < x_0, y < y_0 \Rightarrow M(x,y) < M(x_0,y_0)$$

$M(x,y)$ is continuous

$$M(tx,ty) = tM(x,y)$$

Basic property of means

What is a mean?

A function $M : \mathbb{R}^+ \times \mathbb{R}^+ \to \mathbb{R}^+$ is a *mean* if $(\forall x, y, x_0, y_0, t \in \mathbb{R}^+)$

$$M(x, x) = x$$

$$M(x, y) = M(y, x)$$

$$x < y \;\Rightarrow\; x < M(x, y) < y$$

$$x < x_0, y < y_0 \Rightarrow M(x, y) < M(x_0, y_0)$$

$M(x, y)$ is continuous

$$M(tx, ty) = tM(x, y)$$

$$M(x, y) = xf\left(\frac{y}{x}\right)$$

Basic property of means

What is a mean?

A function $M : \mathbb{R}^+ \times \mathbb{R}^+ \to \mathbb{R}^+$ is a *mean* if $(\forall x, y, x_0, y_0, t \in \mathbb{R}^+)$

$M(x,x) = x$ $\qquad$ $f(1) = 1$

$M(x,y) = M(y,x)$

$x < y \;\Rightarrow\; x < M(x,y) < y$

$x < x_0, y < y_0 \Rightarrow M(x,y) < M(x_0,y_0)$

$M(x,y)$ is continuous

$M(tx,ty) = tM(x,y)$

$$M(x,y) = xf\left(\frac{y}{x}\right)$$

Basic property of means

What is a mean?

A function $M : \mathbb{R}^+ \times \mathbb{R}^+ \to \mathbb{R}^+$ is a *mean* if $(\forall x, y, x_0, y_0, t \in \mathbb{R}^+)$

$M(x,x) = x$ $\qquad f(1) = 1$

$M(x,y) = M(y,x)$ $\qquad f(t) = tf(t^{-1})$

$x < y \;\Rightarrow\; x < M(x,y) < y$

$x < x_0, y < y_0 \Rightarrow M(x,y) < M(x_0,y_0)$

$M(x,y)$ is continuous

$M(tx,ty) = tM(x,y)$

$$M(x,y) = xf\left(\frac{y}{x}\right)$$

Basic property of means

What is a mean?

A function $M : \mathbb{R}^+ \times \mathbb{R}^+ \to \mathbb{R}^+$ is a *mean* if $(\forall x, y, x_0, y_0, t \in \mathbb{R}^+)$

$M(x, x) = x$ $\quad f(1) = 1$

$M(x, y) = M(y, x)$ $\quad f(t) = tf(t^{-1})$

$x < y \;\Rightarrow\; x < M(x, y) < y$ $\quad f(\cdot > 1) > 1,\ f(0 < \cdot < 1) < 1$

$x < x_0, y < y_0 \Rightarrow M(x, y) < M(x_0, y_0)$

$M(x, y)$ is continuous

$M(tx, ty) = tM(x, y)$

$$M(x, y) = xf\left(\frac{y}{x}\right)$$

Basic property of means

What is a mean?

A function $M : \mathbb{R}^+ \times \mathbb{R}^+ \to \mathbb{R}^+$ is a *mean* if $(\forall x, y, x_0, y_0, t \in \mathbb{R}^+)$

$M(x,x) = x$ $\quad$ $f(1) = 1$

$M(x,y) = M(y,x)$ $\quad$ $f(t) = tf(t^{-1})$

$x < y \;\Rightarrow\; x < M(x,y) < y$ $\quad$ $f(\cdot > 1) > 1,\; f(0 < \cdot < 1) < 1$

$x < x_0, y < y_0 \Rightarrow M(x,y) < M(x_0,y_0)$ $\quad$ f increasing

$M(x,y)$ is continuous

$M(tx,ty) = tM(x,y)$

$$M(x,y) = xf\left(\frac{y}{x}\right)$$

Basic property of means

What is a mean?

A function $M : \mathbb{R}^+ \times \mathbb{R}^+ \to \mathbb{R}^+$ is a *mean* if $(\forall x, y, x_0, y_0, t \in \mathbb{R}^+)$

$M(x, x) = x$ $\quad$ $f(1) = 1$

$M(x, y) = M(y, x)$ $\quad$ $f(t) = tf(t^{-1})$

$x < y \;\Rightarrow\; x < M(x, y) < y$ $\quad$ $f(\cdot > 1) > 1,\ f(0 < \cdot < 1) < 1$

$x < x_0, y < y_0 \Rightarrow M(x, y) < M(x_0, y_0)$ $\quad$ f increasing

$M(x, y)$ is continuous $\quad$ f continuous

$M(tx, ty) = tM(x, y)$

$$M(x, y) = xf\left(\frac{y}{x}\right)$$

We have

$$\text{means} = \left\{ f \in C(\mathbb{R}^+, \mathbb{R}^+) \;\middle|\; \begin{array}{c} f \text{ increasing} \\ f(1) = 1 \\ \forall t \in \mathbb{R}^+ : \; f(t) = tf(t^{-1}) \end{array} \right\}$$

$$M(x, y) = xf\left(\frac{y}{x}\right)$$

We have

$$\text{means} = \left\{ f \in C(\mathbb{R}^+, \mathbb{R}^+) \;\middle|\; \begin{array}{c} f \text{ increasing} \\ f(1) = 1 \\ \forall t \in \mathbb{R}^+ : \; f(t) = tf(t^{-1}) \end{array} \right\}$$

$$M(x, y) = xf\left(\frac{y}{x}\right)$$

arithmetic mean: $f(t) = \dfrac{1+t}{2}$

geometric mean: $f(t) = \sqrt{t}$

logarithmic mean: $f(t) = \dfrac{t-1}{\log t}$

Means of matrices

Define means on $n \times n$, positive definite matrices $\mathcal{M}_n^+$:

$$X \in \mathcal{M}_n^+ \iff X = X^*, \begin{cases} \langle v, Xv \rangle > 0 \ \forall v \in \mathbb{C}^n \setminus \{0\} \\ \text{every eigenvalue of } X \text{ is positive} \end{cases}$$

We write $X \leq Y$ if $Y - X \in \mathcal{M}_n^+$.

Means of matrices

Define means on $n \times n$, positive definite matrices $\mathcal{M}_n^+$:

$$X \in \mathcal{M}_n^+ \iff X = X^*, \begin{cases} \langle v, Xv \rangle > 0 \ \forall v \in \mathbb{C}^n \setminus \{0\} \\ \text{every eigenvalue of } X \text{ is positive} \end{cases}$$

We write $X \leq Y$ if $Y - X \in \mathcal{M}_n^+$.

M is a *mean of matrices* if for every $X, Y \in \mathcal{M}_n^+$

– $X \leq X_0,\ Y \leq Y_0:\ M(X, Y) \leq M(X_0, Y_0)$

– $(X_n)_{n\in\mathbb{N}}$ and $(Y_n)_{n\in\mathbb{N}}$ are decreasing sequences ($X_{n+1} \leq X_n$, $Y_{n+1} \leq Y_n$) in $\mathcal{M}_n^+$ with limits X and Y then $M(X_n, Y_n)$ is decreasing and

$$\lim_{n\to\infty} M(X_n, Y_n) = M(X, Y)$$

– $T^* M(X, Y) T \leq M(T^* X T, T^* Y T)$ for all T

– $M(X, X) = X$

Theorem (*Kubo-Ando*)

If M is a matrix mean, then there exists an operator monotone function f with properties $f(t) = tf(t^{-1})$ and $f(1) = 1$ such that for every $X, Y \in \mathcal{M}_n^+$

$$M(X, Y) = X^{1/2} f(X^{-1/2} Y X^{-1/2}) X^{1/2}.$$

M is a *mean of matrices* if for every $X, Y \in \mathcal{M}_n^+$

– $X \leq X_0,\ Y \leq Y_0 :\ M(X, Y) \leq M(X_0, Y_0)$

– $(X_n)_{n\in\mathbb{N}}$ and $(Y_n)_{n\in\mathbb{N}}$ are decreasing sequences ($X_{n+1} \leq X_n$, $Y_{n+1} \leq Y_n$) in $\mathcal{M}_n^+$ with limits X and Y then $M(X_n, Y_n)$ is decreasing and

$$\lim_{n\to\infty} M(X_n, Y_n) = M(X, Y)$$

– $T^* M(X, Y) T \leq M(T^* X T, T^* Y T)$ for all T

– $M(X, X) = X$

Theorem (*Kubo-Ando*)

If M is a matrix mean, then there exists an operator monotone function f with properties $f(t) = tf(t^{-1})$ and $f(1) = 1$ such that for every $X, Y \in \mathcal{M}_n^+$

$$M(X, Y) = X^{1/2} f(X^{-1/2} Y X^{-1/2}) X^{1/2}.$$

M is a *mean of matrices* if for every $X, Y \in \mathcal{M}_n^+$

– $X \leq X_0,\ Y \leq Y_0 :\ M(X,Y) \leq M(X_0,Y_0)$

– $(X_n)_{n\in\mathbb{N}}$ and $(Y_n)_{n\in\mathbb{N}}$ are decreasing sequences ($X_{n+1} \leq X_n$, $Y_{n+1} \leq Y_n$) in $\mathcal{M}_n^+$ with limits X and Y then $M(X_n, Y_n)$ is decreasing and

$$\lim_{n\to\infty} M(X_n, Y_n) = M(X,Y)$$

– $T^* M(X,Y) T \leq M(T^*XT, T^*YT)$ for all T

– $M(X,X) = X$

Theorem (*Kubo-Ando*)

If M is a matrix mean, then there exists an operator monotone function f with properties $f(t) = tf(t^{-1})$ and $f(1) = 1$ such that for every $X, Y \in \mathcal{M}_n^+$

$$M(X,Y) = X^{1/2} f(X^{-1/2} Y X^{-1/2}) X^{1/2}.$$

M is a *mean of matrices* if for every $X, Y \in \mathcal{M}_n^+$

– $X \leq X_0,\ Y \leq Y_0:\ M(X,Y) \leq M(X_0,Y_0)$

– $(X_n)_{n\in\mathbb{N}}$ and $(Y_n)_{n\in\mathbb{N}}$ are decreasing sequences ($X_{n+1} \leq X_n$, $Y_{n+1} \leq Y_n$) in $\mathcal{M}_n^+$ with limits X and Y then $M(X_n, Y_n)$ is decreasing and

$$\lim_{n\to\infty} M(X_n, Y_n) = M(X, Y)$$

– $T^* M(X,Y) T \leq M(T^*XT, T^*YT)$ for all T

– $M(X,X) = X$

Theorem (*Kubo-Ando*)

If M is a matrix mean, then there exists an operator monotone function f with properties $f(t) = tf(t^{-1})$ and $f(1) = 1$ such that for every $X, Y \in \mathcal{M}_n^+$

$$M(X,Y) = X^{1/2} f(X^{-1/2} Y X^{-1/2}) X^{1/2}.$$

M is a *mean of matrices* if for every $X, Y \in \mathcal{M}_n^+$

– $X \leq X_0,\ \ Y \leq Y_0:\ \ M(X, Y) \leq M(X_0, Y_0)$

– $(X_n)_{n\in\mathbb{N}}$ and $(Y_n)_{n\in\mathbb{N}}$ are decreasing sequences ($X_{n+1} \leq X_n$, $Y_{n+1} \leq Y_n$) in $\mathcal{M}_n^+$ with limits X and Y then $M(X_n, Y_n)$ is decreasing and

$$\lim_{n\to\infty} M(X_n, Y_n) = M(X, Y)$$

– $T^*M(X, Y)T \leq M(T^*XT, T^*YT)$ for all T

– $M(X, X) = X$

Theorem (*Kubo-Ando*)

If M is a matrix mean, then there exists an operator monotone function f with properties $f(t) = tf(t^{-1})$ and $f(1) = 1$ such that for every $X, Y \in \mathcal{M}_n^+$

$$M(X, Y) = X^{1/2} f(X^{-1/2} Y X^{-1/2}) X^{1/2}.$$

M is a *mean of matrices* if for every $X, Y \in \mathcal{M}_n^+$

– $X \leq X_0,\ \ Y \leq Y_0 :\ \ M(X, Y) \leq M(X_0, Y_0)$

– $(X_n)_{n\in\mathbb{N}}$ and $(Y_n)_{n\in\mathbb{N}}$ are decreasing sequences ($X_{n+1} \leq X_n$, $Y_{n+1} \leq Y_n$) in $\mathcal{M}_n^+$ with limits X and Y then $M(X_n, Y_n)$ is decreasing and

$$\lim_{n\to\infty} M(X_n, Y_n) = M(X, Y)$$

– $T^* M(X, Y) T \leq M(T^* X T, T^* Y T)$ for all T

– $M(X, X) = X$

Theorem (*Kubo-Ando*)

If M is a matrix mean, then there exists an operator monotone function f with properties $f(t) = tf(t^{-1})$ and $f(1) = 1$ such that for every $X, Y \in \mathcal{M}_n^+$

$$M(X, Y) = X^{1/2} f(X^{-1/2} Y X^{-1/2}) X^{1/2}.$$

Looking for monotone metrics:

Looking for monotone metrics:
monotonicity:

$$g_{T(D)}(T(X), T(X)) \leq g_D(X, X) \; \forall D \in \mathcal{M}_n, \forall X \in \mathrm{T}_p\mathcal{M}_n \; ,$$

Looking for monotone metrics:
monotonicity:

$$g_{T(D)}(T(X), T(X)) \leq g_D(X, X) \ \forall D \in \mathcal{M}_n, \forall X \in \mathsf{T}_p\mathcal{M}_n \ ,$$

$g_D(X, Y) = \left\langle X, \mathbf{J}_D^{-1}(Y) \right\rangle = \mathrm{Tr}(X\mathbf{J}_D^{-1}(Y))$, where
$\mathbf{J}_D : \mathrm{Mat}(n, \mathbb{C}) \to \mathrm{Mat}(n, \mathbb{C})$ linear map.

Looking for monotone metrics:
monotonicity:

$$g_{T(D)}(T(X), T(X)) \leq g_D(X, X) \ \forall D \in \mathcal{M}_n, \forall X \in \mathrm{T}_p\mathcal{M}_n \ ,$$

$g_D(X, Y) = \langle X, \mathbf{J}_D^{-1}(Y) \rangle = \mathrm{Tr}(X\mathbf{J}_D^{-1}(Y))$, where
$\mathbf{J}_D : \mathrm{Mat}(n, \mathbb{C}) \to \mathrm{Mat}(n, \mathbb{C})$ linear map.

$$g_{T(D)}(T(X), T(X)) = \left\langle T(X), \mathbf{J}_{T(D)}^{-1}(T(X)) \right\rangle$$

Looking for monotone metrics:
monotonicity:

$$g_{T(D)}(T(X), T(X)) \leq g_D(X, X)\ \forall D \in \mathcal{M}_n, \forall X \in \mathsf{T}_p\mathcal{M}_n\ ,$$

$g_D(X, Y) = \langle X, \mathbf{J}_D^{-1}(Y) \rangle = \mathrm{Tr}(X\mathbf{J}_D^{-1}(Y))$, where
$\mathbf{J}_D$: $\mathrm{Mat}(n, \mathbb{C}) \to \mathrm{Mat}(n, \mathbb{C})$ linear map.

$$g_{T(D)}(T(X), T(X)) = \left\langle T(X), \mathbf{J}_{T(D)}^{-1}(T(X)) \right\rangle$$

$$g_{T(D)}(T(X), T(X)) = \left\langle X, T^* \mathbf{J}_{T(D)}^{-1} T(X) \right\rangle$$

Looking for monotone metrics:
monotonicity:

$$g_{T(D)}(T(X), T(X)) \leq g_D(X, X) \ \forall D \in \mathcal{M}_n, \forall X \in \mathrm{T}_p\mathcal{M}_n \ ,$$

$g_D(X, Y) = \left\langle X, \mathbf{J}_D^{-1}(Y) \right\rangle = \mathrm{Tr}(X\mathbf{J}_D^{-1}(Y))$, where
$\mathbf{J}_D : \mathrm{Mat}(n, \mathbb{C}) \to \mathrm{Mat}(n, \mathbb{C})$ linear map.

$$g_{T(D)}(T(X), T(X)) = \left\langle T(X), \mathbf{J}_{T(D)}^{-1}(T(X)) \right\rangle$$

$$g_{T(D)}(T(X), T(X)) = \left\langle X, T^*\mathbf{J}_{T(D)}^{-1} T(X) \right\rangle$$

$$g_D(X, X) = \left\langle X, \mathbf{J}_D^{-1}(X) \right\rangle = \left\langle X, T^*\mathbf{J}_D^{-1} T(X) \right\rangle$$

Looking for monotone metrics:
monotonicity:

$$g_{T(D)}(T(X), T(X)) \leq g_D(X, X) \ \forall D \in \mathcal{M}_n, \forall X \in \mathsf{T}_p\mathcal{M}_n \ ,$$

$g_D(X, Y) = \langle X, \mathbf{J}_D^{-1}(Y) \rangle = \mathrm{Tr}(X\mathbf{J}_D^{-1}(Y))$, where
$\mathbf{J}_D : \mathrm{Mat}(n, \mathbb{C}) \to \mathrm{Mat}(n, \mathbb{C})$ linear map.

$$g_{T(D)}(T(X), T(X)) = \left\langle T(X), \mathbf{J}_{T(D)}^{-1}(T(X)) \right\rangle$$

$$g_{T(D)}(T(X), T(X)) = \left\langle X, T^* \mathbf{J}_{T(D)}^{-1} T(X) \right\rangle$$

$$g_D(X, X) = \langle X, \mathbf{J}_D^{-1}(X) \rangle = \langle X, T^* \mathbf{J}_D^{-1} T(X) \rangle$$

monotonicity:

Looking for monotone metrics:
monotonicity:

$$g_{T(D)}(T(X), T(X)) \leq g_D(X, X) \; \forall D \in \mathcal{M}_n, \forall X \in \mathrm{T}_p\mathcal{M}_n \ ,$$

$g_D(X, Y) = \left\langle X, \mathbf{J}_D^{-1}(Y) \right\rangle = \mathrm{Tr}(X\mathbf{J}_D^{-1}(Y))$, where
$\mathbf{J}_D : \mathrm{Mat}(n, \mathbb{C}) \to \mathrm{Mat}(n, \mathbb{C})$ linear map.

$$g_{T(D)}(T(X), T(X)) = \left\langle T(X), \mathbf{J}_{T(D)}^{-1}(T(X)) \right\rangle$$

$$g_{T(D)}(T(X), T(X)) = \left\langle X, T^* \mathbf{J}_{T(D)}^{-1} T(X) \right\rangle$$

$$g_D(X, X) = \left\langle X, \mathbf{J}_D^{-1}(X) \right\rangle = \left\langle X, T^* \mathbf{J}_D^{-1} T(X) \right\rangle$$

monotonicity: $\quad T^* \mathbf{J}_{T(D)}^{-1} T \leq \mathbf{J}_D^{-1}$

Looking for monotone metrics:
monotonicity:

$$g_{T(D)}(T(X), T(X)) \leq g_D(X, X) \ \forall D \in \mathcal{M}_n, \forall X \in \mathrm{T}_p\mathcal{M}_n \ ,$$

$g_D(X, Y) = \left\langle X, \mathbf{J}_D^{-1}(Y) \right\rangle = \mathrm{Tr}(X\mathbf{J}_D^{-1}(Y))$, where
$\mathbf{J}_D : \mathrm{Mat}(n, \mathbb{C}) \to \mathrm{Mat}(n, \mathbb{C})$ linear map.

$$g_{T(D)}(T(X), T(X)) = \left\langle T(X), \mathbf{J}_{T(D)}^{-1}(T(X)) \right\rangle$$

$$g_{T(D)}(T(X), T(X)) = \left\langle X, T^* \mathbf{J}_{T(D)}^{-1} T(X) \right\rangle$$

$$g_D(X, X) = \left\langle X, \mathbf{J}_D^{-1}(X) \right\rangle = \left\langle X, T^* \mathbf{J}_D^{-1} T(X) \right\rangle$$

monotonicity: $\quad T^* \mathbf{J}_{T(D)}^{-1} T \leq \mathbf{J}_D^{-1}$

$$T \mathbf{J}_D T^* \leq \mathbf{J}_{T(D)}$$

What can $\mathbf{J}_D(X)$ be?

What can $\mathbf{J}_D(X)$ be?
"D can act on left $\varphi_1(D)X$ and on the right $X\varphi_1(D)$"

What can $\mathbf{J}_D(X)$ be?
"D can act on left $\varphi_1(D)X$ and on the right $X\varphi_1(D)$"
in general $\varphi_1(D)X\varphi_2(D)$ gives the idea:

$$\mathbf{J}_D(X) = M(L_D, R_D)(X).$$

Where $L_D(X) = DX$ and $R_D(X) = XD$.

What can $\mathbf{J}_D(X)$ be?
"D can act on left $\varphi_1(D)X$ and on the right $X\varphi_1(D)$"
in general $\varphi_1(D)X\varphi_2(D)$ gives the idea:

$$\mathbf{J}_D(X) = M(L_D, R_D)(X).$$

Where $L_D(X) = DX$ and $R_D(X) = XD$.
We have $M(L_D, R_D) = M(R_D, L_D)$

What can $\mathbf{J}_D(X)$ be?
"D can act on left $\varphi_1(D)X$ and on the right $X\varphi_1(D)$"
in general $\varphi_1(D)X\varphi_2(D)$ gives the idea:

$$\mathbf{J}_D(X) = M(L_D, R_D)(X).$$

Where $L_D(X) = DX$ and $R_D(X) = XD$.
We have $M(L_D, R_D) = M(R_D, L_D)$
and the monotonicity

$$T\mathbf{J}_D T^* \leq \mathbf{J}_{T(D)}$$

gives

$$TM(L_D, R_D)T^* \leq M(TL_D T^*, TR_D T^*).$$

What can $\mathbf{J}_D(X)$ be?
"D can act on left $\varphi_1(D)X$ and on the right $X\varphi_1(D)$"
in general $\varphi_1(D)X\varphi_2(D)$ gives the idea:

$$\mathbf{J}_D(X) = M(L_D, R_D)(X).$$

Where $L_D(X) = DX$ and $R_D(X) = XD$.
We have $M(L_D, R_D) = M(R_D, L_D)$
and the monotonicity

$$T\mathbf{J}_D T^* \leq \mathbf{J}_{T(D)}$$

gives

$$TM(L_D, R_D)T^* \leq M(TL_D T^*, TR_D T^*).$$

M is a mean!

Theorem (*Petz*)

Assume that for every $n \in \mathbb{N}$ *the pair* $(\mathcal{M}_n, g_n)$ *is a Riemannian-manifold. If for every stochastic map* T *the monotonicity*

$$g_{T(D)}(T(X), T(X)) \leq g_D(X, X) \ \forall D \in \mathcal{M}_n, \forall X \in T_p\mathcal{M}_n,$$

holds then there exists an operator monotone function $f : \mathbb{R}^+ \to \mathbb{R}$ *with the property* $f(x) = xf(x^{-1})$*, such that*

$$g_D(X, Y) = Tr\left(X \left(R_{n,D}^{\frac{1}{2}} f(L_{n,D} R_{n,D}^{-1}) R_{n,D}^{\frac{1}{2}} \right)^{-1} (Y) \right).$$

Classical case:

$$\mathcal{P}_n = \left\{ (p_1, \dots, p_n) \,\middle|\, 0 < p_i < 1, \sum_{i=1}^{n} p_i = 1 \right\}.$$

Theorem (*Cencov*) Assume that for every $n \in \mathbb{N}$ $(\mathcal{P}_n, g_n)$ is a Riemannian manifold. If for every transition probability $\kappa : X_n \times X_m \to \mathbb{R}$

$$g_{\tilde{\kappa}(p)}(\kappa^*(X), \kappa^*(X)) \le g_p(X, X) \qquad \forall p \in \Delta_{n-1}, \forall X \in \mathrm{T}_p\Delta_{n-1} \ ,$$

(*monotonicity*) holds, then the family of metrics $(g_n)_{n\in\mathbb{N}}$ unique up to a positive factor. given by the equation

$$g_D(X, Y) = \mathrm{Tr}\left(X \big(R_{n,D}^{\frac{1}{2}} f(L_{n,D} R_{n,D}^{-1}) R_{n,D}^{\frac{1}{2}}\big)^{-1}(Y) \right),$$

where $f : \mathbb{R}^+ \to \mathbb{R}^+$ is an operator monotone function such that $f(x) = x f(x^{-1})$ $(\forall x \in \mathbb{R}^+)$.

Quantum case:

$$\mathcal{M}_n = \left\{ D \in \mathrm{Mat}(n, \mathbb{C}) \,\middle|\, D = D^*, D > 0, \mathrm{Tr}\, D = 1 \right\}.$$

Theorem (*Cencov*) Assume that for every $n \in \mathbb{N}$ $(\mathcal{P}_n, g_n)$ is a Riemannian manifold. If for every transition probability $\kappa : X_n \times X_m \to \mathbb{R}$

$$g_{\tilde{\kappa}(p)}(\kappa^*(X), \kappa^*(X)) \le g_p(X, X) \qquad \forall p \in \Delta_{n-1}, \forall X \in \mathrm{T}_p\Delta_{n-1} \ ,$$

(*monotonicity*) holds, then the family of metrics $(g_n)_{n\in\mathbb{N}}$ unique up to a positive factor. given by the equation

$$g_D(X, Y) = \mathrm{Tr}\left(X \left(R_{n,D}^{\frac{1}{2}} f(L_{n,D} R_{n,D}^{-1}) R_{n,D}^{\frac{1}{2}} \right)^{-1} (Y) \right),$$

where $f : \mathbb{R}^+ \to \mathbb{R}^+$ is an operator monotone function such that $f(x) = x f(x^{-1})$ $(\forall x \in \mathbb{R}^+)$.

Quantum case:

$$\mathcal{M}_n = \left\{ D \in \mathrm{Mat}(n, \mathbb{C}) \,\middle|\, D = D^*, D > 0, \operatorname{Tr} D = 1 \right\}.$$

Assume that for every $n \in \mathbb{N}$ $(\mathcal{P}_n, g_n)$ is a Riemannian manifold. If for every transition probability $\kappa : X_n \times X_m \to \mathbb{R}$

$$g_{\tilde{\kappa}(p)}(\kappa^*(X), \kappa^*(X)) \leq g_p(X, X) \qquad \forall p \in \Delta_{n-1}, \forall X \in \mathrm{T}_p\Delta_{n-1} \ ,$$

(*monotonicity*) holds, then the family of metrics $(g_n)_{n\in\mathbb{N}}$ unique up to a positive factor. given by the equation

$$g_D(X,Y) = \operatorname{Tr}\left(X \left(R_{n,D}^{\frac{1}{2}} f(L_{n,D} R_{n,D}^{-1}) R_{n,D}^{\frac{1}{2}} \right)^{-1} (Y) \right),$$

where $f : \mathbb{R}^+ \to \mathbb{R}^+$ is an operator monotone function such that $f(x) = x f(x^{-1})$ $(\forall x \in \mathbb{R}^+)$.

Quantum case:

$$\mathcal{M}_n = \left\{ D \in \mathrm{Mat}(n, \mathbb{C}) \,\middle|\, D = D^*, D > 0, \operatorname{Tr} D = 1 \right\}.$$

Assume that for every $n \in \mathbb{N}$ $(\mathcal{M}_n, g_n)$ is a Riemannian manifold. If for every transition probability $\kappa : X_n \times X_m \to \mathbb{R}$

$$g_{\tilde{\kappa}(p)}(\kappa^*(X), \kappa^*(X)) \leq g_p(X, X) \qquad \forall p \in \Delta_{n-1}, \forall X \in \mathrm{T}_p\Delta_{n-1} \ ,$$

(*monotonicity*) holds, then the family of metrics $(g_n)_{n \in \mathbb{N}}$ unique up to a positive factor. given by the equation

$$g_D(X, Y) = \operatorname{Tr}\left(X \left(R_{n,D}^{\frac{1}{2}} f(L_{n,D} R_{n,D}^{-1}) R_{n,D}^{\frac{1}{2}} \right)^{-1} (Y) \right),$$

where $f : \mathbb{R}^+ \to \mathbb{R}^+$ is an operator monotone function such that $f(x) = x f(x^{-1})$ $(\forall x \in \mathbb{R}^+)$.

Quantum case:

$$\mathcal{M}_n = \left\{ D \in \mathrm{Mat}(n, \mathbb{C}) \,\middle|\, D = D^*, D > 0, \operatorname{Tr} D = 1 \right\}.$$

Assume that for every $n \in \mathbb{N}$ $(\mathcal{M}_n, g_n)$ is a Riemannian manifold. If for every stochastic map $T : \mathcal{M}_n^+ \to \mathcal{M}_n^+$

$$g_{\tilde{\kappa}(p)}(\kappa^*(X), \kappa^*(X)) \leq g_p(X, X) \qquad \forall p \in \Delta_{n-1}, \forall X \in \mathrm{T}_p\Delta_{n-1} \ ,$$

(*monotonicity*) holds, then the family of metrics $(g_n)_{n\in\mathbb{N}}$ unique up to a positive factor. given by the equation

$$g_D(X, Y) = \operatorname{Tr}\left(X \left(R_{n,D}^{\frac{1}{2}} f(L_{n,D} R_{n,D}^{-1}) R_{n,D}^{\frac{1}{2}} \right)^{-1} (Y) \right),$$

where $f : \mathbb{R}^+ \to \mathbb{R}^+$ is an operator monotone function such that $f(x) = x f(x^{-1})$ $(\forall x \in \mathbb{R}^+)$.

Quantum case:

$$\mathcal{M}_n = \left\{ D \in \mathrm{Mat}(n, \mathbb{C}) \,\middle|\, D = D^*, D > 0, \mathrm{Tr}\, D = 1 \right\}.$$

Assume that for every $n \in \mathbb{N}$ $(\mathcal{M}_n, g_n)$ is a Riemannian manifold. If for every stochastic map $T : \mathcal{M}_n^+ \to \mathcal{M}_n^+$

$$g_{T(D)}(T(X), T(X)) \leq g_D(X, X) \ \forall D \in \mathcal{M}_n, \forall X \in \mathrm{T}_D \mathcal{M}_n$$

(*monotonicity*) holds, then the family of metrics $(g_n)_{n \in \mathbb{N}}$ unique up to a positive factor. given by the equation

$$g_D(X, Y) = \mathrm{Tr}\left(X \left(R_{n,D}^{\frac{1}{2}} f(L_{n,D} R_{n,D}^{-1}) R_{n,D}^{\frac{1}{2}} \right)^{-1} (Y) \right),$$

where $f : \mathbb{R}^+ \to \mathbb{R}^+$ is an operator monotone function such that $f(x) = x f(x^{-1})$ $(\forall x \in \mathbb{R}^+)$.

Quantum case:

$$\mathcal{M}_n = \left\{ D \in \mathrm{Mat}(n, \mathbb{C}) \;\middle|\; D = D^*, D > 0, \operatorname{Tr} D = 1 \right\}.$$

Theorem (*Petz*) Assume that for every $n \in \mathbb{N}$ $(\mathcal{M}_n, g_n)$ is a Riemannian manifold. If for every stochastic map $T : \mathcal{M}_n^+ \to \mathcal{M}_n^+$

$$g_{T(D)}(T(X), T(X)) \leq g_D(X, X) \; \forall D \in \mathcal{M}_n, \forall X \in \mathrm{T}_D \mathcal{M}_n$$

(*monotonicity*) holds, then the family of metrics $(g_n)_{n \in \mathbb{N}}$ given by the equation

$$g_D(X, Y) = \operatorname{Tr}\left(X \left(R_{n,D}^{\frac{1}{2}} f(L_{n,D} R_{n,D}^{-1}) R_{n,D}^{\frac{1}{2}} \right)^{-1} (Y) \right),$$

where $f : \mathbb{R}^+ \to \mathbb{R}^+$ is an operator monotone function such that $f(x) = x f(x^{-1})$ $(\forall x \in \mathbb{R}^+)$.

Definition

Consider the Riemannian manifold $(\mathcal{M}_n^+, K^{(n)})$. The metric $K^{(n)}$ is called *monotone metric* if there exists an operator monotone function $f : \mathbb{R}^+ \to \mathbb{R}$ such that for every positive number x $f(x) = xf(x^{-1})$ and $K^{(n)}$ is generated by f.

$$g_D(X, Y) = \mathrm{Tr}\left(X \left(R_{n,D}^{\frac{1}{2}} f(L_{n,D} R_{n,D}^{-1}) R_{n,D}^{\frac{1}{2}} \right)^{-1} (Y) \right)$$

Properties of operator monotone functions

Definition

Assume that $f : \mathbb{R}_0^+ \to \mathbb{R}$ is an operator monotone function.
$f^\backslash(x) = xf(x^{-1})$ is called to *transpose of* f,
$f^\perp(x) = \frac{x}{f(x)}$ is called to *dual of* f.

f is *symmetric* if $f = f^\backslash$

f is *normalized* if $f(1) = 1$.

Theorem

If $f : \mathbb{R}_0^+ \to \mathbb{R}$ *is symmetric operator monotone, then its dual is symmetric and operator monotone too.*

Representations of operator monotone functions

Denote by $\mathcal{F}_{\mathbb{R}_0^+}$ the set of operator monotone functions defined on $\mathbb{R}_0^+$ and by $\mathcal{F}_{\mathbb{R}_0^+}^{(\mathrm{S},\mathrm{n})}$ the symmetric normalized ones.

Denote by $\mathcal{G}_I$ the set of positive Radon-measures on the interval $I \subseteq \mathbb{R}$.

A measure $\mu \in \mathcal{G}_I$ is said to be *normalized* if $\mu(I) = 1$.

Denote by $\mathcal{G}_I^{(\mathrm{n})}$ the set of normalized measures.

Theorem (*Löwner*)

There is a bijective correspondence

$$\phi : \mathcal{G}_{\mathbb{R}_0^+} \to \mathcal{F}_{\mathbb{R}_0^+} \qquad \mu \mapsto f_\mu$$

$$f_\mu(x) = \int_0^\infty \frac{x(1+t)}{x+t}\, d\,\mu(t).$$

Theorem

There is a bijective correspondence

$$\phi : \mathcal{G}_{[0,1]} \to \mathcal{F}_{\mathbb{R}_0^+} \qquad \mu \mapsto f_\mu$$

$$f_\mu(x) = \int_0^1 \frac{x}{(1-t)x+t} \; d\mu(t).$$

The function f_μ *is symmetric iff* $\mu([0,s]) = \mu([1-s,1])$ *holds for every* $0 \leq s \leq 1$.

Cencov-Morozova function

Definition

The function $c : \left(\mathbb{R}_0^+\right)^2 \to \mathbb{R}$ is called *Cencov–Morozova function* if there exists an $f \in \mathcal{F}_{\mathbb{R}_0^+}$ such that for every positive x, y

$$c(x, y) = \frac{1}{y f\left(\frac{x}{y}\right)}.$$

If $f : \mathbb{R}^+ \to \mathbb{R}^+$ is operator monotone then it is smooth, moreover it can be extended to a horizontal in the complex plane around the positive real axes.

So if f is operator monotone then for every $\rho \in \mathbb{R}$ we have

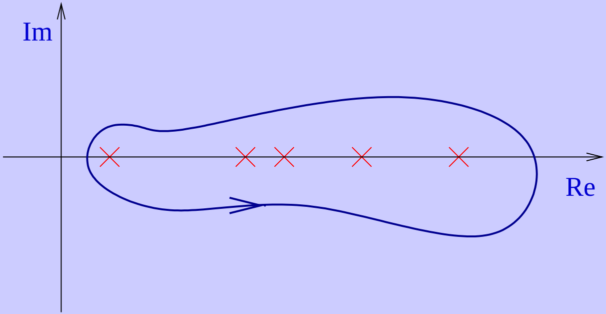
$$f(\rho) = \frac{1}{2\pi\,\mathrm{i}} \oint_\Gamma f(\xi)(\xi - \rho)^{-1}\ \mathrm{d}\,\xi$$

by Cauchy integral formula, where Γ is a smooth closed curve around ρ with counter-clockwise orientation.

The *Riesz–Dunford operator calculus* states that this can be done for operators too. If A is a self-adjoint operator then

$$f(A) = \frac{1}{2\pi\,\mathrm{i}} \oint_\Gamma f(\xi)(\xi\,\mathrm{I} - A)^{-1}\;\mathrm{d}\,\xi,$$

where the interior of Γ contains all the eigenvalues of A.



We have seen that for a state $D \in \mathcal{M}_n^+$ the multiplications $L_{n,D}$ and $R_{n,D}$ are self-adjoint operators, so we have

$$f(L_{n,D}) = \frac{1}{2\pi \mathrm{i}} \oint_\Gamma f(\xi)(\xi \mathrm{I} - L_{n,D})^{-1} \; \mathrm{d}\xi$$
$$f(R_{n,D}) = \frac{1}{2\pi \mathrm{i}} \oint_\Gamma f(\xi)(\xi \mathrm{I} - R_{n,D})^{-1} \; \mathrm{d}\xi.$$

This leads us to

$$f(L_{n,D})(X) = \frac{1}{2\pi \mathrm{i}} \oint_\Gamma f(\xi)(\xi \mathrm{I} - D)^{-1} X \; \mathrm{d}\xi$$

$$f(R_{n,D})(X) = \frac{1}{2\pi \mathrm{i}} \oint_\Gamma f(\xi) X (\xi \mathrm{I} - D)^{-1} \; \mathrm{d}\xi.$$

These expressions can be extended to multivariate case, such as

$$c(L_{n,D},R_{n,D})=\frac{1}{(2\pi\,\mathrm{i})^2}\oint\oint c(\xi,\eta)(\xi\,\mathrm{I}-L_{n,D})^{-1}(\eta\,\mathrm{I}-R_{n,D})^{-1}\,\mathrm{d}\,\xi\,\mathrm{d}\,\eta,$$

which effect can be computed as

$$c(L_{n,D},R_{n,D})(X)=\frac{1}{(2\pi\,\mathrm{i})^2}\oint\oint c(\xi,\eta)(\xi\,\mathrm{I}-D)^{-1}X(\eta\,\mathrm{I}-D)^{-1}\,\mathrm{d}\,\xi\,\mathrm{d}\,\eta.$$

Theorem

If $K^{(n)}$ is a monotone metric on $\mathcal{M}_n^+$ generated by an operator monotone function f then for every state $D \in \mathcal{M}_n^+$ and tangent vector $X, Y \in T_D\mathcal{M}_n^+$ we have

$$K_D^{(n)}(X, Y) = Tr\,\frac{1}{(2\pi\,\mathrm{i})^2}\oint\oint c(\xi, \eta) X (\xi I - D)^{-1} Y (\eta I - D)^{-1}\, d\xi\, d\eta.$$

Examples for functions in $\mathcal{F}_{\mathbb{R}_0^+}^{(\mathrm{S,n})}$.

$$f_{\mathrm{SM}}(x) = \frac{1+x}{2} \quad f_{\mathrm{LA}}(x) = \frac{2x}{1+x} \quad f_{\mathrm{KM}}(x) = \frac{x-1}{\log x}$$

$$f_{\mathrm{P1}}(x) = \frac{2x^{\alpha+1/2}}{1+x^{2\alpha}} \quad 0 \le \alpha \le 1/2$$

$$f_{\mathrm{P2}}(x) = \frac{\beta(1-\beta)(x-1)^2}{(x^\beta - 1)(x^{1-\beta} - 1)} \quad \beta \in [-1,2] \setminus \{0,1\},$$

$$\mathrm{WYD}(x) = \frac{1-\alpha^2}{4} \frac{(x-1)^2}{(1-x^{\frac{1-\alpha}{2}})(1-x^{\frac{1+\alpha}{2}})} \quad \alpha \in [-3,3] \setminus \{-1,1\}$$

$$f_{\mathrm{WY}}(x) = \frac{1}{4}(\sqrt{x}+1)^2$$

$$f_{\mathrm{P3}}(x) = \left(\frac{1+x^{\frac{1}{\nu}}}{2}\right)^{\nu} \qquad \nu \in [1,2]$$

Consider the matrix units E_{ij} ($(E_{ij})_{ab} = \delta_{ia}\delta_{jb}$) and matrices $F_{ij} = E_{ij} + E_{ji}$ and $H_{ij} = \mathrm{i}\, E_{ij} - \mathrm{i}\, E_{ji}$. (These form a basis of the tangent space.)

Theorem

If the monotone metric $K^{(n),f}$ on $\mathcal{M}_n^+$ is generated by f then at a state $D \in \mathcal{M}_n^+$ in the form of $D = \sum_{k=1}^{n} \lambda_k E_{kk}$ we have

$$1 \leq i < j \leq n,\ 1 \leq k < l \leq n : \begin{cases} G_D(H_{ij}, H_{kl}) = \delta_{ik}\delta_{jl} 2c(\lambda_i, \lambda_j) \\ G_D(F_{ij}, F_{kl}) = \delta_{ik}\delta_{jl} 2c(\lambda_i, \lambda_j) \\ G_D(H_{ij}, F_{kl}) = 0, \end{cases}$$

$$1 \leq i < j \leq n,\ 1 \leq k \leq n : \quad G_D(H_{ij}, F_{kk}) = G(F_{ij}, F_{kk}) = 0,$$

$$1 \leq i \leq n,\ 1 \leq k \leq n : \quad G_D(F_{ii}, F_{kk}) = \delta_{ik} 4c(\lambda_i, \lambda_i).$$

Example (Smallest metric)

The metric $K_{\mathrm{SM}}^{(n)}$ generated by the function $f_{\mathrm{SM}}(x) = \frac{1+x}{2}$ is called *smallest metric* since $f_{\mathrm{SM}}(x)$ is maximal among functions in $\mathcal{F}_{\mathbb{R}_0^+}^{(\mathrm{S,n})}$ with respect to the pointwise order

$$f \underset{[0,1]}{\leq} g \iff f(x) \leq g(x) \qquad \forall x \in [0,1]\,.$$

The corresponding Chencov-Morozova function is

$$c_{SM}(x,y) = \frac{2}{x+y}.$$

Example (Smallest metric cont.)

The inner product of vectors X, Y can be written in the form of

$$K^{(n)}_{\mathrm{SM},D}(X,Y) = \mathrm{Tr}\, XZ,$$

where Z is the solution of the equation

$$DZ + ZD = 2Y.$$

The geodesic distance between states D_1 and D_2 according to this metric is

$$d_{\mathrm{SM}}(D_1, D_2) = \sqrt{2\left(1 - \mathrm{Tr}\left(D_1^{1/2} D_2 D_1^{1/2}\right)^{1/2}\right)}.$$

(1992 Uhlmann, studying Berry phase)

Example (Largest metric)

The metric $K^{(n)}_{\mathrm{LA}}$ generated by the function $f_{\mathrm{LA}}(x) = \frac{2x}{1+x}$ is called *largest metric* since $f_{\mathrm{SM}}(x)$ is maximal among functions in $\mathcal{F}^{(\mathrm{S,n})}_{\mathbb{R}^+_0}$. In this case the metric can be written in a simple form

$$K^{(n)}_{\mathrm{LA},D}(X,Y) = \mathrm{Tr}\, XD^{-1}Y.$$

Example (Kubo-Mori metric)

The metric generated by the function $f_{\mathrm{KM}}(x) = \dfrac{x-1}{\log x}$.

Its Cencov–Morozova function is

$$c_{\mathrm{KM}}(x,y) = \frac{\log x - \log y}{x-y} \ .$$

Using the integral representation

$$c_{\mathrm{KM}}(x,y) = \int_0^\infty (t+x)^{-1}(t+y)^{-1} \ \mathrm{d}\, t$$

we have for the metric

$$K^{(n)}_{\mathrm{KM},D}(X,Y) = \mathrm{Tr} \int_0^\infty X(t+D)^{-1} Y(t+D)^{-1} \ \mathrm{d}\, t.$$

(Linear response theory Fick, Sailer.)

A simple consequences of ordering

Theorem

For every $f \in \mathcal{F}_{\mathbb{R}_0^+}^{(S,n)}$ *we have*

$$f_{SM} \underset{[0,1]}{\geq} f \underset{[0,1]}{\geq} f_{LA}.$$

Theorem

Assume that $f \in \mathcal{F}_{\mathbb{R}_0^+}^{(S,n)}$. *For every state* $D \in \mathcal{M}_n^+$ *and tangent vector* $X \in T_D\mathcal{M}_n^+$ *we have*

$$K_{LA,D}^{(n)}(X,X) \geq K_D^{(n),f}(X,X) \geq K_{SM,D}^{(n)}(X,X).$$

We have a continuous path in $\mathcal{F}_{\mathbb{R}_0^+}^{(\mathrm{S},\mathrm{n})}$ from smallest to largest.

$$f_{\mathrm{SM}} = f_{\mathrm{P3}}^{(\nu=1)} \underset{[0,1]}{\geq} f_{\mathrm{P3}}^{(1\leq\nu\leq 2)} \underset{[0,1]}{\geq} f_{\mathrm{P3}}^{(\nu=2)} = f_{\mathrm{WY}}$$

$$f_{\mathrm{WY}} = f_{\mathrm{WYD}}^{(\alpha=0)} \underset{[0,1]}{\geq} f_{\mathrm{WYD}}^{(0\leq\alpha\leq 3)} \underset{[0,1]}{\geq} f_{\mathrm{WYD}}^{(\alpha=3)} = f_{\mathrm{LA}}$$

Monotone metric from entropy

Consider the integral representation of the log function

$$\log x = \int_0^\infty (1+t)^{-1} - (x+t)^{-1} \, \mathrm{d}\, t.$$

We have for the entropy

$$S(D) = \mathrm{Tr}\, D \int_0^\infty (D+t)^{-1} - (\mathrm{I}+t)^{-1} \, \mathrm{d}\, t.$$

The first derivative of the entropy is $\mathrm{d}\, S(D)(A) = -\mathrm{Tr}\, A \log D$.

The second derivative is

$$\mathrm{d}^2 S : \mathcal{M}_n^+ \to \mathrm{Lin}\Big(T\mathcal{M}_n, \mathrm{Lin}(T\mathcal{M}_n, \mathbb{R})\Big)$$

$$\mathrm{d}^2 S(D)(A)(B) = -\operatorname{Tr} \int_0^\infty (D+t)^{-1} A (D+t)^{-1} B \ \mathrm{d}\, t,$$

which is (-1) times the Kubo-Mori metric.

Monotone metric from euclidean metric

For the complex state space $\mathcal{M}_n^+$ denote by $S_1^{n^2-1}$ the unit ball in the euclidean space $\mathbb{R}^{n\times n}$ and consider the map

$$\phi : \mathcal{M}_n^+ \to S^{n^2-1} \qquad D \mapsto \sqrt{D}.$$

Using derivative of ϕ

$$(\mathrm{d}_D \,\phi)(A) = \left(L_D^{1/2} + R_D^{1/2}\right)^{-1}(A)$$

we can deduce that the pull back metric in this case is

$$\begin{aligned}(\phi^* g)(A,B) &= \langle (d_D\phi)(A), (d_D\phi)(B)\rangle \\ &= \mathrm{Tr}\, A(L_D^{1/2} + R_D^{1/2})^{-2}(B) \\ &= \frac{1}{4}\,\mathrm{Tr}\, A c_{\mathrm{WY}}(L_D, R_D)(B).\end{aligned}$$

So in this case easy to compute the geodesic distance between states D_1 and D_2

$$d_{\mathrm{WY}}(D_1, D_2) = 2 \arccos \mathrm{Tr}\, D_1^{1/2} D_2^{1/2}.$$

First relative entropy

The first version of relative entropy in quantum setting was given by Umegaki in 1962. He defined the relative entropy of states $D_1, D_2 \in \mathcal{M}_n^+$ as

$$S(D_1, D_2) = \operatorname{Tr} D_1(\log D_1 - \log D_2).$$

This relative entropy is called to *Umegaki relative entropy*.

Relative entropy from operator convex functions

Definition

A continuous function $f : \mathbb{R} \to \mathbb{R}$ is called *operator convex* if for every $n \in \mathbb{N}$ and $n \times n$ self-adjoint operator A, B and parameter $\lambda \in [0, 1]$

$$f(\lambda A + (1 - \lambda)B) \leq \lambda f(A) + (1 - \lambda)f(B)$$

holds.

The set of operator convex functions g with property $g(1) = 0$ defined on the interval $I \subseteq \mathbb{R}$ is denoted by $\mathcal{K}_I$.

Representation theorem for operator convex functions

Theorem

If $g : \mathbb{R}^+ \to \mathbb{R}$ *is an operator convex function then there exist parameters* $a \in \mathbb{R}$, $b, c \in \mathbb{R}_0^+$ *and a positive finite measure* μ_g *on the interval* $\mathbb{R}_0^+$ *such that*

$$g(x) = a(x-1) + b(x-1)^2 + c\frac{(x-1)^2}{x} + \int_0^\infty (x-1)^2 \frac{1+t}{x+t}\, d\mu_g(t).$$

For every parameter $a \in \mathbb{R}$, $b, c \in \mathbb{R}_0^+$ *and finite measure* μ *equation above defines an operator convex function.*

Definition (Petz)

If $g \in \mathcal{K}_{\mathbb{R}^+}$ then the function $H_g(\cdot,\cdot) : \mathcal{M}_n^+ \times \mathcal{M}_n^+ \to \mathbb{R}$

$$H_g(D_1, D_2) = \operatorname{Tr}\left(D_1^{1/2} g(L_{D_2} R_{D_1}^{-1}) D_1^{1/2}\right)$$

is called to *g-relative entropy*.

Theorem (Properties of g-relative entropy)

Assume that H is a g-relative entropy.

1. *Then for every state D_1, D_2: $H(D_1, D_2) \geq 0$, and $H(D_1, D_2) = 0$ iff $D_1 = D_2$.*
2. *H is jointly convex, that is for every state D_1, D_2, D_3, D_4 and parameter $\lambda \in [0, 1]$ we have*

$$H(\lambda D_1 + (1-\lambda) D_2, \lambda D_3 + (1-\lambda) D_4) \leq \lambda H(D_1, D_3) + (1-\lambda) H(D_2, D_4).$$

Theorem (Properties of g-relative entropy cont.)

3. *H is monotone: for every stochastic map $T : \mathcal{M}_n^+ \to \mathcal{M}_n^+$*

$$H(T(D_1), T(D_2)) \leq H(D_1, D_2) \qquad \forall D_1, D_2 \in \mathcal{M}_n^+.$$

4. *H is differentiable: for every state $D_1, D_2 \in \mathcal{M}_n^+$ and tangent vectors $A \in T_{D_1}\mathcal{M}_n^+$, $B \in T_{D_2}\mathcal{M}_n^+$ the map $\mathbb{R} \times \mathbb{R} \to \mathbb{R}$*

$$(x, y) \mapsto H(D_1 + xA, D_2 + yB)$$

is differentiable at the origin.

The quantity $H_g(D_1, D_2)$ depends mainly on $D_1 - D_2$.

Theorem

If $g \in \mathcal{K}_{\mathbb{R}^+}$ then for every state $D_1, D_2 \in \mathcal{M}_n^+$

$$H_g(D_1, D_2) = Tr\left((D_1 - D_2) R_{D_1}^{-1}\left(g(L_{D_2} R_{D_1}^{-1})(D_1 - D_2)\right)\right).$$

For an operator convex function g define its *transpose* as $g^{\setminus}(x) = xg(x^{-1})$, and *dual* as $g^{\perp}(x) = \frac{x}{g(x)}$.

g is said to be *symmetric* if $g^{\setminus} = g$.

g is said to be *normalised* if $g''(1) = 1$.

The effect of transpose is changing the arguments

$$H_g(D_1, D_2) = H_{g^{\setminus}}(D_2, D_1).$$

Theorem (Riemannian metric from relative entropy)

Assume that $g \in \mathcal{K}_{\mathbb{R}_0^+}$. *Then*
$K^{g,(n)} : \mathcal{M}_n^+ \to Lin(T\mathcal{M}_n \times T\mathcal{M}_n, \mathbb{R})$

$$K_D^{g,(n)}(X,Y) = -\frac{\partial^2}{\partial s \partial t} H_g(D+tX, D+sY)\Big|_{t=s=0}$$

is a Riemannian metric on $\mathcal{M}_n^+$.

Define an equivalence relation on $\mathcal{K}_{\mathbb{R}^+}$ as

$$f \sim g \iff f + f^{\backslash} = g + g^{\backslash}.$$

Theorem

The functions $g_1, g_2 \in \mathcal{K}_{\mathbb{R}^+}$ *generates the same metric iff* $g_1 \sim g_2$.

Relative entropy and monotone metrics

Theorem

The map $\phi : \mathcal{K}^S_{\mathbb{R}^+} \to \mathcal{F}^S_{\mathbb{R}^+_0}$

$$g(x) \mapsto \phi(g)(x) = \begin{cases} \dfrac{(x-1)^2}{g(x) + xg(x^{-1})} & \text{if} \quad x > 0,\ x \neq 1, \\ \dfrac{1}{g''(1)} & \text{if} \quad x = 1, \\ \dfrac{1}{\lim\limits_{x \to 0} g(x) + xg(x^{-1})} & \text{if} \quad x = 0 \end{cases}$$

is well-defined and

$$K^{g,(n)}_D(X,Y) = K^{(n),\phi(g)}_D(X,Y) \qquad \forall D \in \mathcal{M}^+_n \quad \forall X, Y \in T\mathcal{M}_n.$$

Relative entropy and monotone metrics

Theorem

The map $\epsilon : \mathcal{F}_{\mathbb{R}_0^+}^{(S)} \to \mathcal{K}_{\mathbb{R}^+}^{(S)}$

$$f(x) \mapsto \epsilon(f)(x) = \frac{(x-1)^2}{2f(x)}$$

is well-defined and $K^{(n),f} = K^{\epsilon(f),(n)}$ *holds.*

Relative entropy and monotone metrics

Combining these we have the following theorem.

Theorem

There is a simple bijective correspondence between

1. *the set of monotone metrics,*
2. $\mathcal{F}^{(S)}_{\mathbb{R}_0^+}$,
3. $\mathcal{K}^{(S)}_{\mathbb{R}^+}$.

Example (Smallest metric)

The corresponding operator monotone function is $f(x) = \frac{1+x}{2}$ and the generated operator convex function is

$$g(x) = \frac{(x-1)^2}{1+x}.$$

and the relative entropy

$$H_{\mathrm{SM}} : \mathcal{M}_n^+ \times \mathcal{M}_n^+ \to \mathbb{R} \qquad (D_1, D_2) \mapsto H_{\mathrm{SM}}(D_1, D_2)$$

$$H_{\mathrm{SM}}(D_1, D_2) = \mathrm{Tr}(D_1 - D_2)(L_{D_2} + R_{D_1})^{-1}(D_1 - D_2).$$

Bures relative entropy

Example (Largest metric)

The corresponding operator monotone function is $f(x) = \frac{2x}{1+x}$ and the generated operator convex function is

$$g(x) = (x-1)^2 \frac{1+x}{4x}$$

and the relative entropy is

$$H_{g_1}(D_1, D_2) = \frac{1}{2} \mathrm{Tr}(D_1 - D_2) D_1^{-1} (D_1 - D_2).$$

Quadratic relative entropy

Example (Kubo-Mori metric)

The corresponding operator monotone function is $f(x) = \frac{x-1}{\log x}$ and the generated operator convex function is

$$g(x) = \frac{x-1}{2} \log x$$

and the generated relative entropy is

$$H_{g_1}(D_1, D_2) = \mathrm{Tr}\, D_1 \Big(\log D_1 - \log D_2 \Big).$$

Umegaki relative entropy

Assume that $f \in \mathcal{F}^{(\mathrm{n})}_{\mathbb{R}^+_0}$ and $h \in \mathcal{K}^{\mathrm{n}}_{\mathbb{R}^+}$. We use the term *h is compatible with f* if for the function

$$g(x) = \frac{(x-1)^2}{2f(x)}$$

$h \sim g$ holds.
For a monotone metric $K^{(n),f}$ and a compatible function h we define a covariant derivative $\nabla^{f,h} : T\mathcal{M}_n \times T\mathcal{M}_n \to T\mathcal{M}_n$ as

$$K^{(n),f}_D(\nabla^{f,h}_X Y, Z) = -\frac{\partial^3}{\partial s \partial t \partial u} H_h(D + sX + tY, D + uZ)\Big|_{s,t,u=0},$$

where $X, Y, Z \in T_D\mathcal{M}^+_n$.
(Giblisco, Isola, Uhlmann, Dabrowksi, Jadczyk, Hübner)

Main theorem of duality

Theorem

For a function $f \in \mathcal{F}_{\mathbb{R}_0^+}^{(S,n)}$ *and a compatible function* $h \in \mathcal{K}_{\mathbb{R}^+}^{(n)}$ *the quadruplet* $(\mathcal{M}_n^+, K^{(n),f}, \nabla^{f,h}, \nabla^{f,h\backslash})$ *is torsion free dual geometry.*

A characterization of the Kubo-Mori metric

Theorem

If $(\mathcal{M}_n^+, g, \nabla^{(1)}, \nabla^{(-1)})$ *is a dual geometry for some Riemannian metric then* g *equals to Kubo-Mori metric* $g^{(KM)}$ *up to a positive multiplicative factor.*

Pythagorean theorem

Theorem

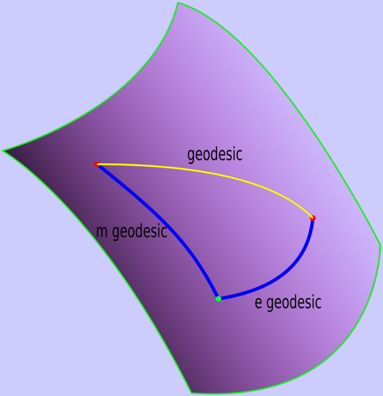
Consider states $D_1, D_2, D_3 \in \mathcal{M}_n^+$ *and* $\nabla^{(1)}$ *geodesic curve* γ_1 *connecting* D_1 *and* D_2 *and* $\nabla^{(-1)}$ *geodesic curve* γ_2 *connecting* D_2 *and* D_3. *If*

$$K^{(n)}_{KM,D_2}(\dot\gamma_1(D_2), \dot\gamma_2(D_2)) = 0 \ ,$$

holds then

$$H_{\log}(D_1, D_3) = H_{\log}(D_1, D_2) + H_{\log}(D_2, D_3).$$

Pythagorean theorem



Hilbert-Schmidt measure

The Hilbert-Schmidt measure on $\mathcal{M}_n^+$ is defined by the Euclidean metric

$$d(D_1, D_2) = \sqrt{\mathrm{Tr}(D_1 - D_2)^2}$$

We can consider $\mathcal{M}_n^+$ as a manifold with metric

$$g_D(X, Y) = \mathrm{Tr}(XY) \quad D \in \mathcal{M}_n^+ \quad X, Y \in T_D\mathcal{M}_n^+ \, .$$

Induces the flat, Euclidean geometry on the set of states.

The invariant volume measure is

$$\rho(D) = \sqrt{\det g_D} = 1 \ .$$

(Which is the most simple prior on $\mathcal{M}_n^+$.) The volume of the state space is

$$\text{Volume} = \int\limits_{\mathcal{M}_n^+} 1 \ \mathrm{d}\, D \ ,$$

where

$$\mathrm{d}\, D = \mathrm{d}\, a_{11} \,\mathrm{d}\, a_{12} \ldots \mathrm{d}\, a_{22} \,\mathrm{d}\, a_{23} \ldots \mathrm{d}\, a_{n-1,n} \ .$$

The invariant volume measure is

$$\rho(D) = \sqrt{\det g_D} = 1 \ .$$

(Which is the most simple prior on $\mathcal{M}_n^+$.) The volume of the state space is

$$\text{Volume} = \int\limits_{\mathcal{M}_n^+} 1 \ \mathrm{d}\, D \ ,$$

where

$$\mathrm{d}\, D = \mathrm{d}\, a_{11} \, \mathrm{d}\, a_{12} \ldots \mathrm{d}\, a_{22} \, \mathrm{d}\, a_{23} \ldots \mathrm{d}\, a_{n-1,n} \ .$$

Some notations:

$$A_4 = \begin{pmatrix} a_{11} & a_{12} & a_{13} & a_{14} \\ a_{12}^* & a_{22} & a_{23} & a_{24} \\ a_{13}^* & a_{23}^* & a_{33} & a_{34} \\ a_{14}^* & a_{24}^* & a_{34}^* & a_{44} \end{pmatrix}$$

Some notations:

$$A_4 = \begin{pmatrix} a_{11} & a_{12} & a_{13} & a_{14} \\ a_{12}^* & a_{22} & a_{23} & a_{24} \\ a_{13}^* & a_{23}^* & a_{33} & a_{34} \\ a_{14}^* & a_{24}^* & a_{34}^* & a_{44} \end{pmatrix} \qquad A_1$$

Some notations:

$$A_4 = \begin{pmatrix} a_{11} & a_{12} & a_{13} & a_{14} \\ a_{12}^* & a_{22} & a_{23} & a_{24} \\ a_{13}^* & a_{23}^* & a_{33} & a_{34} \\ a_{14}^* & a_{24}^* & a_{34}^* & a_{44} \end{pmatrix} \qquad A_2$$

Some notations:

$$A_4 = \begin{pmatrix} a_{11} & a_{12} & a_{13} & a_{14} \\ a_{12}^* & a_{22} & a_{23} & a_{24} \\ a_{13}^* & a_{23}^* & a_{33} & a_{34} \\ a_{14}^* & a_{24}^* & a_{34}^* & a_{44} \end{pmatrix} \qquad A_3$$

Some notations:

$$A_4 = \begin{pmatrix} a_{11} & a_{12} & a_{13} & a_{14} \\ a_{12}^* & a_{22} & a_{23} & a_{24} \\ a_{13}^* & a_{23}^* & a_{33} & a_{34} \\ a_{14}^* & a_{24}^* & a_{34}^* & a_{44} \end{pmatrix} \qquad A_3$$

$$T_n := \det(A_n) \times (A_n)^{-1} \qquad \det T_n = (\det A_n)^{n-1}$$

Some notations:

$$A_4 = \begin{pmatrix} a_{11} & a_{12} & a_{13} & a_{14} \\ a_{12}^* & a_{22} & a_{23} & a_{24} \\ a_{13}^* & a_{23}^* & a_{33} & a_{34} \\ a_{14}^* & a_{24}^* & a_{34}^* & a_{44} \end{pmatrix} \qquad A_3$$

$$T_n := \det(A_n) \times (A_n)^{-1} \qquad \det T_n = (\det A_n)^{n-1}$$

$$A_4 = \begin{pmatrix} a_{11} & a_{12} & a_{13} & a_{14} \\ a_{12}^* & a_{22} & a_{23} & a_{24} \\ a_{13}^* & a_{23}^* & a_{33} & a_{34} \\ a_{14}^* & a_{24}^* & a_{34}^* & a_{44} \end{pmatrix} \qquad \underline{x}_1, \underline{x}_2, \underline{x}_3$$

Some notations:

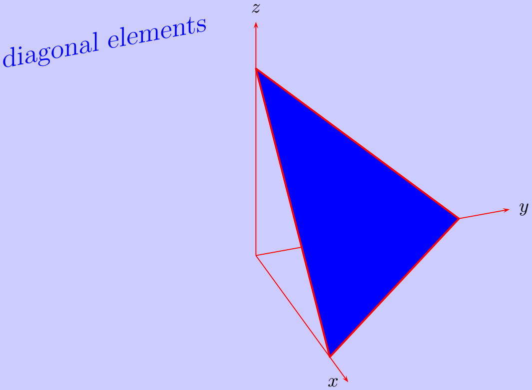
$$A_4 = \begin{pmatrix} a_{11} & a_{12} & a_{13} & a_{14} \\ a_{12}^* & a_{22} & a_{23} & a_{24} \\ a_{13}^* & a_{23}^* & a_{33} & a_{34} \\ a_{14}^* & a_{24}^* & a_{34}^* & a_{44} \end{pmatrix} \qquad A_3$$

$$T_n := \det(A_n) \times (A_n)^{-1} \qquad \det T_n = (\det A_n)^{n-1}$$

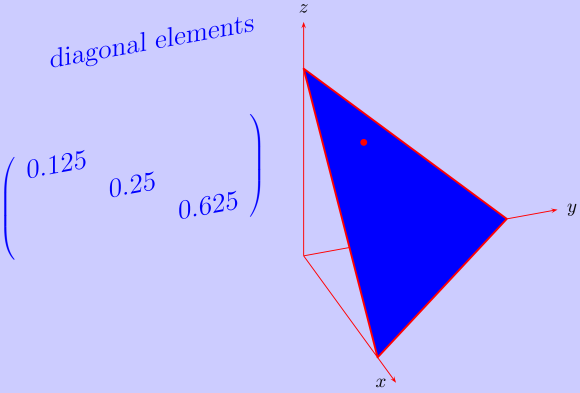
$$A_4 = \begin{pmatrix} a_{11} & a_{12} & a_{13} & a_{14} \\ a_{12}^* & a_{22} & a_{23} & a_{24} \\ a_{13}^* & a_{23}^* & a_{33} & a_{34} \\ a_{14}^* & a_{24}^* & a_{34}^* & a_{44} \end{pmatrix} \qquad \underline{x}_1, \underline{x}_2, \underline{x}_3$$

Lemma: $\det A_n = a_{nn}(\det A_{n-1}) - \left\langle \underline{x}_{n-1}, T_{n-1}\underline{x}_{n-1} \right\rangle.$

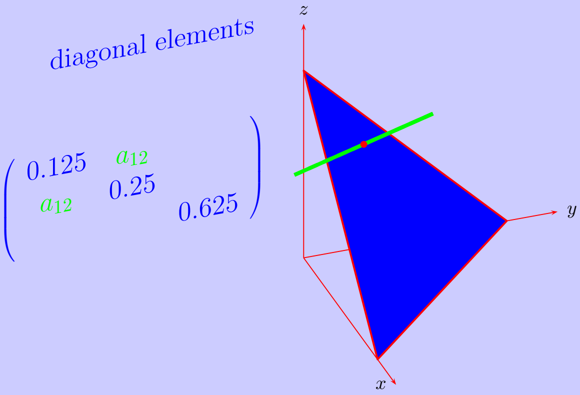
Decomposition of the state space: 3×3 real case:



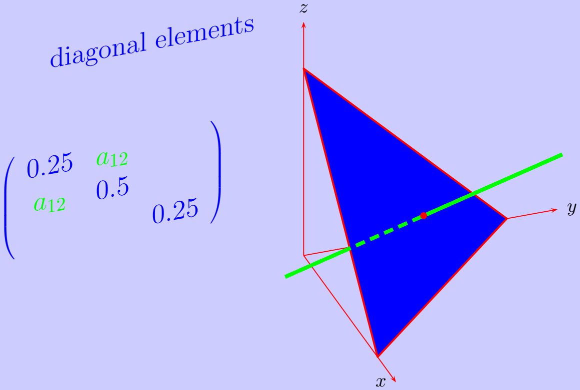
Decomposition of the state space: 3×3 real case:



Decomposition of the state space: 3×3 real case:



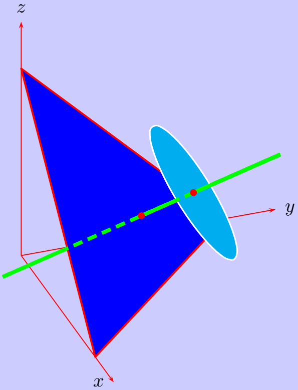
Decomposition of the state space: 3×3 real case:



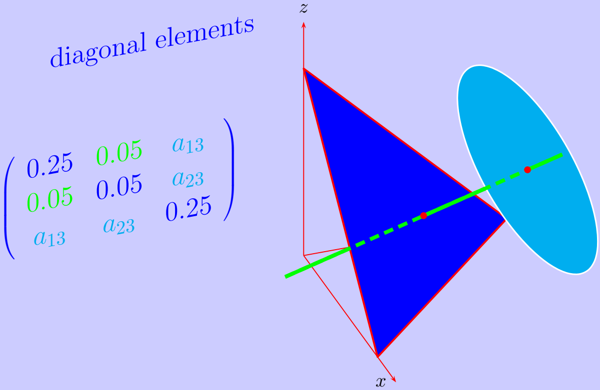
Decomposition of the state space: 3×3 real case:

diagonal elements

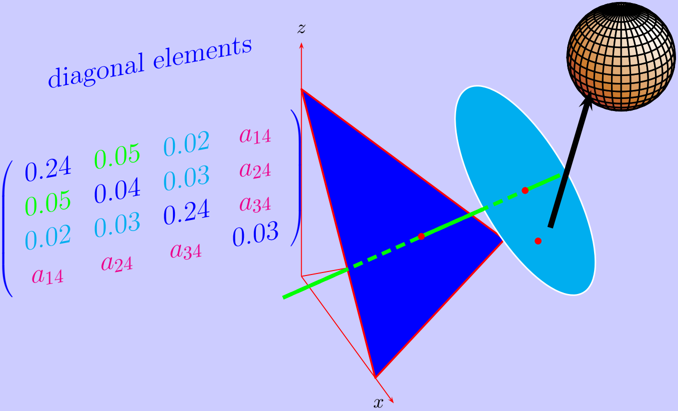
$$\begin{pmatrix} 0.25 & 0.1 & a_{13} \\ 0.1 & 0.5 & a_{23} \\ a_{13} & a_{23} & 0.25 \end{pmatrix}$$



Decomposition of the state space: 3×3 real case:



Decomposition of the state space: 4×4 real case:



Theorem

For every $n \in \mathbb{N}$ *the volume of the state space* $\mathcal{M}_n^+$ *is*

$$V(\mathcal{M}_n^+) = \frac{\pi^{dn(n-1)/4}}{\Gamma\left(d\frac{n(n-1)}{2} + n\right)} \prod_{i=1}^{n-1} \Gamma\left(\frac{id}{2} + 1\right)$$

and the integral of the function $\det^\alpha$ *with respect to the normalized Hilbert–Schmidt measure is*

$$\int_{\mathcal{M}_n^+} \det{}^\alpha = \frac{\Gamma\left(\frac{dn(n-1)}{2} + n\right)}{\Gamma\left(\frac{dn(n-1)}{2} + n + n\alpha\right)} \prod_{i=1}^{n} \frac{\Gamma\left(d\frac{i-1}{2} + 1 + \alpha\right)}{\Gamma\left(d\frac{i-1}{2} + 1\right)}.$$

In the space of qubits we use the Stokes parametrization

$$D = \frac{1}{2}\begin{pmatrix} 1+x & y+\mathrm{i}\,z \\ y+\mathrm{i}\,z & 1-x \end{pmatrix}.$$

$\mathcal{M}_2$ can be identified with the unit ball in $\mathbb{R}^3$ and $\mathbb{R}^2$.
The Riemannian metric $g^{(f)}$ in this coordinate system is

$$g_f(x,y,z) = \frac{1}{2}\begin{pmatrix} \frac{1}{2\lambda_1\lambda_2} & 0 & 0 \\ 0 & \frac{1}{\lambda_1 f\left(\frac{\lambda_2}{\lambda 1}\right)} & 0 \\ 0 & 0 & \frac{1}{\lambda_1 f\left(\frac{\lambda_2}{\lambda 1}\right)} \end{pmatrix}$$

$$g_f(x,y) = \frac{1}{2}\begin{pmatrix} \frac{1}{2\lambda_1\lambda_2} & 0 \\ 0 & \frac{1}{\lambda_1 f\left(\frac{\lambda_2}{\lambda 1}\right)} \end{pmatrix}.$$

The volume is an integral on the unit ball, which can be expressed as

$$V\left(\mathcal{M}_2^{(\mathbb{C})}\right) = 2\pi \int_0^1 \left(\frac{1-t}{1+t}\right)^2 \frac{1}{\sqrt{t} f(t)} \,\mathrm{d}\, t$$

$$V\left(\mathcal{M}_2^{(\mathbb{R})}\right) = \sqrt{2}\pi \int_0^1 \frac{1-t}{1+t} \frac{1}{\sqrt{t+t^2}\sqrt{f(t)}} \,\mathrm{d}\, t.$$

The volume of the state space with monotone metric is unknown.

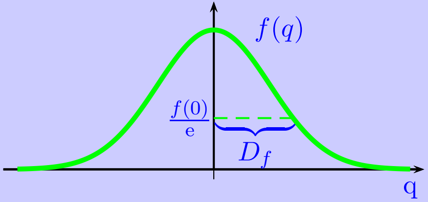
Some operator monotone functions and the corresponding volumes.

$f(x)$:	$V\left(\mathcal{M}_2^{(\mathbb{C})}\right)$:	$V\left(\mathcal{M}_2^{(\mathbb{R})}\right)$:
$\frac{1+x}{2}$	π^2	2π
$\frac{2x}{1+x}$	∞	∞
$\frac{x-1}{\log x}$	$2\pi^2$	~ 8.298
$\sqrt{x}$	∞	4π
$(\sqrt{x}+1)^2/4$	$4\pi(\pi-2)$	$4\pi(2-\sqrt{2})$
$\frac{2\sqrt{x}(x-1)}{(1+x)\log x}$	∞	~ 19.986

Brief history of uncertainty relations

1927, Heisenberg: not possible to measure the position and moment at a same time. (Idea, not a theorem.)

Heisenberg studied Gauss distributions ($f(q)$), where "uncertainty" was the width of D_f.



If $\mathcal{F}(f)$ denotes the Fourier transform of f then the first equation for uncertainty was

$$D_f D_{\mathcal{F}(f)} = \text{constant}.$$

1927, Kennard: For observables A, B if $[A, B] = -\mathrm{i}$ then

$$\mathrm{Var}_D(A)\,\mathrm{Var}_D(B) \geq \frac{1}{4},$$

where $\mathrm{Var}_D(A) = \mathrm{Tr}(DA^2) - (\mathrm{Tr}(DA))^2$.

1929, Robertson: For all observables A, B

$$\mathrm{Var}_D(A)\,\mathrm{Var}_D(B) \geq \frac{1}{4}\left|\mathrm{Tr}(D\,[A, B])\right|^2.$$

1927, Kennard: For observables A, B if $[A, B] = -\mathrm{i}$ then

$$\mathrm{Var}_D(A)\,\mathrm{Var}_D(B) \geq \frac{1}{4},$$

where $\mathrm{Var}_D(A) = \mathrm{Tr}(DA^2) - (\mathrm{Tr}(DA))^2$.

1929, Robertson: For all observables A, B

$$\mathrm{Var}_D(A)\,\mathrm{Var}_D(B) \geq \frac{1}{4}\left|\mathrm{Tr}(D\,[A, B])\right|^2.$$

1930, Schrödinger: For all observables A, B

$$\mathrm{Var}_D(A)\,\mathrm{Var}_D(B) - \mathrm{Cov}_D(A,B)^2 \geq \frac{1}{4}\left|\mathrm{Tr}(D\,[A,B])\right|^2,$$

where

$$\mathrm{Cov}_D(A,B) = \frac{1}{2}\Big(\mathrm{Tr}(DAB) + \mathrm{Tr}(DBA)\Big) - \mathrm{Tr}(DA)\,\mathrm{Tr}(DB).$$

Or in a bit different form:

$$\det\begin{pmatrix} \mathrm{Cov}_D(A,A) & \mathrm{Cov}_D(A,B) \\ \mathrm{Cov}_D(B,A) & \mathrm{Cov}_D(B,B) \end{pmatrix} \geq$$

$$\geq \det\left[-\frac{\mathrm{i}}{2}\begin{pmatrix} \mathrm{Tr}(D\,[A,A]) & \mathrm{Tr}(D\,[A,B]) \\ \mathrm{Tr}(D\,[B,A]) & \mathrm{Tr}(D\,[B,B]) \end{pmatrix}\right].$$

1930, Schrödinger: For all observables A, B

$$\mathrm{Var}_D(A)\,\mathrm{Var}_D(B) - \mathrm{Cov}_D(A,B)^2 \geq \frac{1}{4}\left|\mathrm{Tr}(D\,[A,B])\right|^2 ,$$

where

$$\mathrm{Cov}_D(A,B) = \frac{1}{2}\Big(\mathrm{Tr}(DAB) + \mathrm{Tr}(DBA)\Big) - \mathrm{Tr}(DA)\,\mathrm{Tr}(DB).$$

Or in a bit different form:

$$\det\begin{pmatrix} \mathrm{Cov}_D(A,A) & \mathrm{Cov}_D(A,B) \\ \mathrm{Cov}_D(B,A) & \mathrm{Cov}_D(B,B) \end{pmatrix} \geq$$

$$\geq \det\left[-\frac{\mathrm{i}}{2}\begin{pmatrix} \mathrm{Tr}(D\,[A,A]) & \mathrm{Tr}(D\,[A,B]) \\ \mathrm{Tr}(D\,[B,A]) & \mathrm{Tr}(D\,[B,B]) \end{pmatrix}\right].$$

1934, Robertson: For finite set of observables $(A_i)_{i \in I}$

$$\det\left(\left[\mathrm{Cov}_D(A_h, A_j)\right]_{h,j \in I}\right) \geq \det\left(\left[-\frac{\mathrm{i}}{2}\,\mathrm{Tr}(D\,[A_h, A_j])\right]_{h,j \in I}\right).$$

$\sim$2000–, Furuichi, Gibilisco, Hansen, Imparato, Isola, Kosaki, Kuriyama, Luo, Petz, Yanagi, Q. Zhang, Z. Zhang

1934, Robertson: For finite set of observables $(A_i)_{i\in I}$

$$\det\left(\left[\mathrm{Cov}_D(A_h, A_j)\right]_{h,j\in I}\right) \geq \det\left(\left[-\frac{\mathrm{i}}{2}\,\mathrm{Tr}(D\,[A_h, A_j])\right]_{h,j\in I}\right).$$

∼2000–, Furuichi, Gibilisco, Hansen, Imparato, Isola, Kosaki, Kuriyama, Luo, Petz, Yanagi, Q. Zhang, Z. Zhang

New concepts

For observables A, B, state $D \in \mathcal{M}_n^+$ and operator monotone function f:

$$\mathrm{Cov}_D(A,B) = \frac{1}{2}\left(\mathrm{Tr}(DAB) + \mathrm{Tr}(DBA)\right) - \mathrm{Tr}(DA)\,\mathrm{Tr}(DB)$$

$$\mathrm{Cov}_D^f(A,B) = \langle A, B\rangle_{D,f} \qquad \text{(2002, Petz)}$$

$$\mathrm{qCov}_{D,f}^{as}(A,B) = \frac{f(0)}{2}\left\langle \mathrm{i}\,[D,A]\,,\mathrm{i}\,[D,B]\right\rangle_{D,f}$$

$$\mathrm{qCov}_{D,f}^{s}(A,B) = \frac{f(0)}{2}\left\langle \{D,A\}\,,\{D,B\}\right\rangle_{D,f}\,,$$

where $[.,.]$ is the commutator and $\{.,.\}$ is the anticommutator.

For an observable A and state D define $A_0 = A - \mathrm{Tr}(DA)I$, then $\mathrm{Tr}\,DA_0 = 0$.

New concepts

For observables A, B, state $D \in \mathcal{M}_n^+$ and operator monotone function f:

$$\mathrm{Cov}_D(A,B) = \frac{1}{2}\left(\mathrm{Tr}(DAB) + \mathrm{Tr}(DBA)\right) - \mathrm{Tr}(DA)\,\mathrm{Tr}(DB)$$

$$\mathrm{Cov}_D^f(A,B) = \langle A, B\rangle_{D,f} \qquad \text{(2002, Petz)}$$

$$\mathrm{qCov}_{D,f}^{as}(A,B) = \frac{f(0)}{2}\left\langle \mathrm{i}\,[D,A]\,, \mathrm{i}\,[D,B]\right\rangle_{D,f}$$

$$\mathrm{qCov}_{D,f}^{s}(A,B) = \frac{f(0)}{2}\left\langle \{D,A\}\,, \{D,B\}\right\rangle_{D,f}\,,$$

where $[.,.]$ is the commutator and $\{.,.\}$ is the anticommutator.

For an observable A and state D define $A_0 = A - \mathrm{Tr}(DA)I$, then $\mathrm{Tr}\, DA_0 = 0$.

For observables $(A^{(k)})_{k=1,\dots,N}$ with zero mean at a state D define

$$\begin{aligned} [\mathrm{Cov}_D]_{ij} &= \mathrm{Cov}_D(A^{(i)}, A^{(j)}) \\ \left[\mathrm{Cov}_D^f\right]_{ij} &= \mathrm{Cov}_D^f(A^{(i)}, A^{(j)}) \\ \left[\mathrm{qCov}_{D,f}^{as}\right]_{ij} &= \mathrm{qCov}_{D,f}^{as}(A^{(i)}, A^{(j)}) \\ \left[\mathrm{qCov}_{D,f}^{s}\right]_{ij} &= \mathrm{qCov}_{D,f}^{s}(A^{(i)}, A^{(j)}). \end{aligned}$$

2006, Gibilisco: Conjecture: $\det(\mathrm{Cov}_D) \geq \det(\mathrm{qCov}_{D,f}^{as})$.

2008, Andai: The conjecture is true.

For observables $(A^{(k)})_{k=1,\dots,N}$ with zero mean at a state D define

$$\begin{aligned}
[\mathrm{Cov}_D]_{ij} &= \mathrm{Cov}_D(A^{(i)}, A^{(j)}) \\
\left[\mathrm{Cov}_D^f\right]_{ij} &= \mathrm{Cov}_D^f(A^{(i)}, A^{(j)}) \\
\left[\mathrm{qCov}_{D,f}^{as}\right]_{ij} &= \mathrm{qCov}_{D,f}^{as}(A^{(i)}, A^{(j)}) \\
\left[\mathrm{qCov}_{D,f}^{s}\right]_{ij} &= \mathrm{qCov}_{D,f}^{s}(A^{(i)}, A^{(j)}).
\end{aligned}$$

2006, Gibilisco: Conjecture: $\det(\mathrm{Cov}_D) \geq \det(\mathrm{qCov}_{D,f}^{as})$.

2008, Andai: The conjecture is true.

For observables $(A^{(k)})_{k=1,\dots,N}$ with zero mean at a state D define

$$\begin{aligned}
[\mathrm{Cov}_D]_{ij} &= \mathrm{Cov}_D(A^{(i)}, A^{(j)}) \\
\left[\mathrm{Cov}_D^f\right]_{ij} &= \mathrm{Cov}_D^f(A^{(i)}, A^{(j)}) \\
\left[\mathrm{qCov}_{D,f}^{as}\right]_{ij} &= \mathrm{qCov}_{D,f}^{as}(A^{(i)}, A^{(j)}) \\
\left[\mathrm{qCov}_{D,f}^{s}\right]_{ij} &= \mathrm{qCov}_{D,f}^{s}(A^{(i)}, A^{(j)}).
\end{aligned}$$

2006, Gibilisco: Conjecture: $\det(\mathrm{Cov}_D) \geq \det(\mathrm{qCov}_{D,f}^{as})$.

2008, Andai: The conjecture is true.

Up to date results

Theorem (2016, Lovas, Andai)

$$\det(\mathrm{Cov}_D) \geq \det(\mathrm{qCov}^{s}_{D,f}) \geq \det(\mathrm{qCov}^{as}_{D,f}).$$

$$2f(0)\,\mathrm{Cov}_D^{f_{RLD}}(A_0,B_0) \leq \mathrm{qCov}^{s}_{D,f}(A_0,B_0) - \mathrm{qCov}^{as}_{D,f}(A_0,B_0) \leq \mathrm{Cov}_D^{f_{RLD}}(A_0,B_0)$$

$$\det(\mathrm{qCov}^{s}_{D,f}) - \det(\mathrm{qCov}^{as}_{D,f}) \geq (2f(0))^N \det(\mathrm{Cov}_D^{f_{RLD}})$$

2017, Lovas, Andai: Further extensions of symmetric and antisymmetric covariant derivatives and simplified proof for the original Robertson inequality

2018: ???

Up to date results

Theorem (2016, Lovas, Andai)

$$\det(\mathrm{Cov}_D) \geq \det(\mathrm{qCov}^{s}_{D,f}) \geq \det(\mathrm{qCov}^{as}_{D,f}).$$

$$\begin{aligned} 2f(0)\,\mathrm{Cov}_D^{f_{RLD}}(A_0, B_0) &\leq \mathrm{qCov}^{s}_{D,f}(A_0, B_0) - \mathrm{qCov}^{as}_{D,f}(A_0, B_0) \\ &\leq \mathrm{Cov}_D^{f_{RLD}}(A_0, B_0) \end{aligned}$$

$$\det(\mathrm{qCov}^{s}_{D,f}) - \det(\mathrm{qCov}^{as}_{D,f}) \geq (2f(0))^N \det(\mathrm{Cov}_D^{f_{RLD}})$$

2017, Lovas, Andai: Further extensions of symmetric and antisymmetric covariant derivatives and simplified proof for the original Robertson inequality

2018: ???

Up to date results

Theorem (2016, Lovas, Andai)

$$\det(\mathrm{Cov}_D) \geq \det(\mathrm{qCov}^{s}_{D,f}) \geq \det(\mathrm{qCov}^{as}_{D,f}).$$

$$\begin{aligned} 2f(0)\,\mathrm{Cov}_D^{f_{RLD}}(A_0, B_0) & \\ &\leq \mathrm{qCov}^{s}_{D,f}(A_0, B_0) - \mathrm{qCov}^{as}_{D,f}(A_0, B_0) \\ &\leq \mathrm{Cov}_D^{f_{RLD}}(A_0, B_0) \end{aligned}$$

$$\det(\mathrm{qCov}^{s}_{D,f}) - \det(\mathrm{qCov}^{as}_{D,f}) \geq (2f(0))^N \det(\mathrm{Cov}_D^{f_{RLD}})$$

2017, Lovas, Andai: Further extensions of symmetric and antisymmetric covariant derivatives and simplified proof for the original Robertson inequality
2018: ???

References I

[1] P. M. Alberti és A. Uhlmann.
Stochasticity and partial order, volume 18 of *Mathematische Monographien [Mathematical Monographs]*.
VEB Deutscher Verlag der Wissenschaften, Berlin, 1981.
Doubly stochastic maps and unitary mixing.

[2] S. Amari.
Differential-geometrical methods in statistics, volume 28 of *Lecture Notes in Statistics*.
Springer-Verlag, New York, 1985.

[3] S. Amari és H. Nagaoka.
Methods of information geometry, volume 191 of *Translations of Mathematical Monographs*.
American Mathematical Society, Providence, RI, 2000.
Translated from the 1993 Japanese original by Daishi Harada.

[4] S.-I. Amari, O. E. Barndorff-Nielsen, R. E. Kass, S. L. Lauritzen, és C. R. Rao.
Differential geometry in statistical inference.
Institute of Mathematical Statistics Lecture Notes—Monograph Series, 10. Institute of Mathematical Statistics, Hayward, CA, 1987.

[5] T. Ando.
Concavity of certain maps on positive definite matrices and applications to Hadamard products.
Linear Algebra Appl., 26:203–241, 1979.

[6] W. B. Arveson.
Subalgebras of C^*-algebras.
Acta Math., 123:141–224, 1969.

[7] R. Balian, Y. Alhassid, és H. Reinhardt.
Dissipation in many-body systems: a geometric approach based on information theory.
Phys. Rep., 131(1-2):1–146, 1986.

[8] M. B. Bassat.
f-entropies, probability of error and feature selection.
Inform. Control, 39:227–242, 1978.

References II

[9] V. P. Belavkin és P. Staszewski.
C^*-algebraic generalization of relative entropy and entropy.
Ann. Inst. H. Poincaré Sect. A (N.S.), 37(1):51–58, 1982.

[10] R. Bhatia.
Matrix analysis, volume 169 of *Graduate Texts in Mathematics*.
Springer-Verlag, New York, 1997.

[11] A. Bhattacharyya.
On a measure of divergence between two statistical populations defined by their probability distributions.
Bull. Calcutta Math. Soc., 35:99–109, 1943.

[12] F. J. Bloore.
Geometrical description of the convex sets of states for systems with spin- $1/2$ and spin-1.
J. Phys. A, 9(12):2059–2067, 1976.

[13] J. Bognár, F. Göndőcs, L. Kászonyi, A. Kováts, Gy. Michaletzky, T. Móri, Á. Somogyi, L. Szeidl, és J. G. Székely.
Matematikai statisztika.
Nemzeti Tankönyvkiadó, Budapest, 1995.
Szerkesztette J. Mogyoródi és Gy. Michaletzky.

[14] M. Bolla, I. Csiszár, I. Gaudi, O. Gulyás, B. Hajtman, A. Kun, T. Lengyel, G. Michaletzky, L. Rejtő, T. Rudas, G. Székely, L. Telegdi, és G. Tusnády.
Többváltozós statisztikai analízis.
Műszaki Könyvkiadó, Budapest, 1986.
Szerkesztette T. Móri és Székely, G.

[15] D. C. Brody és A. Ritz.
Geometric phase transitions.
cond-mat/9903168, 1999.

[16] N. N. Čencov.
Statistical decision rules and optimal inference, volume 53 of *Translations of Mathematical Monographs*.
American Mathematical Society, Providence, R.I., 1982.
Translation from the Russian edited by Lev J. Leifman.

References III

[17] Man Duen Choi.
Completely positive linear maps on complex matrices.
Linear Algebra and Appl., 10:285–290, 1975.

[18] J. Conway.
A Course in Functional Analysis.
Graduate Texts in Mathematics. Springer Verlag, 1990.
Second Edition.

[19] I. Csiszár.
Information-type measures of difference of probability distributions and indirect observations.
Studia Sci. Math. Hungar., 2:299–318, 1967.

[20] I. Csiszár.
On topology properties of f-divergences.
Studia Sci. Math. Hungar., 2:329–339, 1967.

[21] I. Csiszár.
I-divergence geometry of probability distributions and minimization problems.
Ann. Probability, 3:146–158, 1975.

[22] I. Csiszár és G. Tusnády.
Information geometry and alternating minimization procedures.
Statist. Decisions, (suppl. 1):205–237, 1984.
Recent results in estimation theory and related topics.

[23] L. Dabrowski és A. Jadczyk.
Quantum statistical holonomy.
J. Phys. A, 22(15):3167–3170, 1989.

[24] C. Davis.
Notions generalizing convexity for functions defined on spaces of matrices.
In *Proc. Sympos. Pure Math., Vol. VII*, pages 187–201. Amer. Math. Soc., Providence, R.I., 1963.

References IV

[25] J. Dittmann.
On the curvature of monotone metrics and a conjecture concerning the Kubo-Mori metric.
Linear Algebra Appl., 315(1-3):83–112, 2000.

[26] J. Dittmann és A. Uhlmann.
Connections and metrics respecting purification of quantum states.
J. Math. Phys., 40(7):3246–3267, 1999.

[27] B. Efron.
Defining the curvature of a statistical problem (with applications to second order efficiency).
Ann. Statist., 3(6):1189–1242, 1975.
With a discussion by C. R. Rao, Don A. Pierce, D. R. Cox, D. V. Lindley, Lucien LeCam, J. K. Ghosh, J. Pfanzagl, Neils Keiding, A. P. Dawid, Jim Reeds and with a reply by the author.

[28] S. Eguchi.
Second order efficiency of minimum contrast estimators in a curved exponential family.
Ann. Statist., 11(3):793–803, 1983.

[29] S. Eguchi.
A characterization of second order efficiency in a curved exponential family.
Ann. Inst. Statist. Math., 36(2):199–206, 1984.

[30] S. Eguchi.
A geometric look at nuisance parameter effect of local powers in testing hypothesis.
Ann. Inst. Statist. Math., 43(2):245–260, 1991.

[31] E. Fick és G. Sauermann.
The quantum statistics of dynamic processes, volume 86 of *Springer Series in Solid-State Sciences*.
Springer-Verlag, Berlin, 1990.
Translated from the German by W. D. Brewer.

[32] R. A. Fisher.
On the mathematical foundations of theoretical statistics.
Phil. Trans. R. Soc., A, 222:309–368, 1922.

References V

[33] R. A. Fisher.
The statistical utilization of multiple measurements.
Annals of Eugenics, 8:376–386, 1938.

[34] R. A. Fisher.
Statistical methods and scientific inference.
2nd ed., revised. Hafner Publishing Company, New York, 1959.

[35] B. R. Frieden.
Fisher information, disorder, and the equilibrium distributions of physics.
Phys. Rev. A (3), 41(8):4265–4276, 1990.

[36] A. Fujiwara és H. Nagaoka.
Quantum Fisher metric and estimation for pure state models.
Phys. Lett. A, 201(2-3):119–124, 1995.

[37] A. Fujiwara és H. Nagaoka.
An estimation theoretical characterization of coherent states.
J. Math. Phys., 40(9):4227–4239, 1999.

[38] P. Gibilisco és T. Isola.
Connections on statistical manifolds of density operators by geometry of noncommutative L^p-spaces.
Infin. Dimens. Anal. Quantum Probab. Relat. Top., 2(1):169–178, 1999.

[39] P. Gibilisco és T. Isola.
A characterisation of Wigner-Yanase skew information among statistically monotone metrics.
Infin. Dimens. Anal. Quantum Probab. Relat. Top., 4(4):553–557, 2001.

[40] P. Gibilisco és T. Isola.
On characterization of dual statistically monotone metrics.
math.PR/0303059, 2003.
Probability Theory.

References VI

[41] P. Gibilisco és T. Isola.
Wigner-Yanase information on quantum state space: the geometric approach.
math.PR/0304170, 2003.
Probability Theory.

[42] P. Gibilisco és G. Pistone.
Connections on non-parametric statistical manifolds by Orlicz space geometry.
Infin. Dimens. Anal. Quantum Probab. Relat. Top., 1(2):325–347, 1998.

[43] P. Giblisco és T. Isola.
Monotone metrics on statistical manifolds of denity matrices by geometry of non-commutative l^2-spaces.
???, pages 129–140, 2001.

[44] M. R. Grasselli és R. F. Streater.
On the uniqueness of the Chentsov metric in quantum information geometry.
Infin. Dimens. Anal. Quantum Probab. Relat. Top., 4(2):173–182, 2001.

[45] F. Hansen és G. K. Pedersen.
Jensen's inequality for operators and Löwner's theorem.
Math. Ann., 258(3):229–241, 1981/82.

[46] H. Hasegawa.
Dual geometry of the Wigner–Yanase–Dyson information content.

[47] H. Hasegawa.
α-divergence of the noncommutative information geometry.
In *Proceedings of the XXV Symposium on Mathematical Physics (Toruń, 1992)*, volume 33, pages 87–93, 1993.

[48] H. Hasegawa.
Non-commutative extension of the information geometry.
In *Quantum communications and measurement (Nottingham, 1994)*, pages 327–337. Plenum, New York, 1995.

References VII

[49] M. Hayashi.
Asymptotic estimation theory for a finite-dimensional pure state model.
J. Phys. A, 31(20):4633–4655, 1998.

[50] M. Hayashi.
Corrigenda: "Asymptotic estimation theory for a finite-dimensional pure state model".
J. Phys. A, 31(41):8405, 1998.

[51] E. Hellinger.
Neue Begründung der Theorie quadratischer Formen von unendlich vielen Veränderlichen.
J. für reine and Angew. Math., 36:210–271, 1909.

[52] F. Hiai és D. Petz.
The semicircle law, free random variables and entropy, volume 77 of *Mathematical Surveys and Monographs*.
American Mathematical Society, Providence, RI, 2000.

[53] M. Hübner.
Computation of Uhlmann's parallel transport for density matrices and the Bures metric on three-dimensional Hilbert space.
Phys. Lett. A, 179(4-5):226–230, 1993.

[54] R. S. Ingarden, H. Janyszek, A. Kossakowski, és T. Kawaguchi.
Information geometry of quantum statistical systems.
Tensor (N.S.), 37(1):105–111, 1982.

[55] H. Jeffreys.
An invariant form for the prior probability in estimation problems.
Proc. Roy. Soc. London. Ser. A., 186:453–461, 1946.

[56] A. Jenčová.
Dualistic properties of the manifold of quantum states.
In *Disordered and complex systems (London, 2000)*, volume 553 of *AIP Conf. Proc.*, pages 147–152. Amer. Inst. Phys., Melville, NY, 2001.

References VIII

[57] A. Jenčová.
Geometry of quantum states: dual connections and divergence functions.
Rep. Math. Phys., 47(1):121–138, 2001.

[58] A. Jenčová.
Quantum information geometry and standard purification.
J. Math. Phys., 43(5):2187–2201, 2002.

[59] A. Jenčová.
Flat connections and Wigner-Yanase-Dyson metrics.
math-ph/0307057, 2003.
Mathematical Physics.

[60] K. Kraus.
States, effects, and operations, volume 190 of *Lecture Notes in Physics*.
Springer-Verlag, Berlin, 1983.
Fundamental notions of quantum theory, Lecture notes edited by A. Böhm, J. D. Dollard and W. H. Wootters.

[61] F. Kubo és T. Ando.
Means of positive linear operators.
Math. Ann., 246(3):205–224, 1979/80.

[62] Leibler R. A. Kullback, S.
On information and sufficiency.
Ann. Math. Statistics, 22:79–86, 1951.

[63] S. Kullback.
Information theory and statistics.
Dover Publications Inc., Mineola, NY, 1997.
Reprint of the second (1968) edition.

[64] A. Lesniewski és M. B. Ruskai.
Monotone Riemannian metrics and relative entropy on noncommutative probability spaces.
J. Math. Phys., 40(11):5702–5724, 1999.

References IX

[65] J. Lin és S. K. M. Wong.
A new directed divergence measure and its characterization.
Int. J. General Systems, 17:73–81, 1990.

[66] G. Lindblad.
Entropy, information and quantum measurements.
Comm. Math. Phys., 33:305–322, 1973.

[67] G. Lindblad.
Expectations and entropy inequalities for finite quantum systems.
Comm. Math. Phys., 39:111–119, 1974.

[68] G. Lindblad.
Completely positive maps and entropy inequalities.
Comm. Math. Phys., 40:147–151, 1975.

[69] K. Löwner.
Über monotone Matrixfunctionen.
Math. Z., 38:177–216, 1934.

[70] T. Matumoto.
Any statistical manifold has a contrast function—on the C^3-functions taking the minimum at the diagonal of the product manifold.
Hiroshima Math. J., 23(2):327–332, 1993.

[71] M. Mei.
The theory of genetic distance and evaluation of human races.
Japan J. Human Genetics, 23:341–369, 1978.

[72] E. A. Morozova és N. N. Chentsov.
Markov invariant geometry on state manifolds.
In *Current problems in mathematics. Newest results, Vol. 36 (Russian)*, Itogi Nauki i Tekhniki, pages 69–102, 187. Akad. Nauk SSSR Vsesoyuz. Inst. Nauchn. i Tekhn. Inform., Moscow, 1989.
Translated in J. Soviet Math. **56** (1991), no. 5, 2648–2669.

References X

[73] H. Nagaoka.
On the parameter estimation problem for quantum statistical models.
In *Proceedings of 12th Symposium on Information Theory ans Its Applications*, pages 577–582. Society of Information Theory ans Its Applications in Japan, 1989.

[74] H. Nagaoka és S. Amari.
Differential geometry of smooth families of probability distributions.
Technical Report 82-7, Dept. of Math. Eng. and Instr. Phys., Univ. of Tokyo, 1982.

[75] J. von Neumann.
Thermodynamik quantenmechanischer gesamtheiten.
Gött. Nachr., pages 273–291, 1927.

[76] M. Ohya és D. Petz.
Quantum entropy and its use.
Texts and Monographs in Physics. Springer-Verlag, Berlin, 1993.

[77] M. Ohya és D. Petz.
Notes on quantum entropy.
Studia Sci. Math. Hungar., 31(4):423–430, 1996.

[78] D. Petz.
Quasi-entropies for finite quantum systems.
Rep. Math. Phys., 23(1):57–65, 1986.

[79] D. Petz.
On certain properties of the relative entropy of states of operator algebras.
Math. Z., 206(3):351–361, 1991.

[80] D. Petz.
Geometry of canonical correlation on the state space of a quantum system.
J. Math. Phys., 35(2):780–795, 1994.

References XI

[81] D. Petz.
Monotone metrics on matrix spaces.
Linear Algebra Appl., 244:81–96, 1996.

[82] D. Petz.
Information-geometry of quantum states.
In *Quantum probability communications*, QP-PQ, X, pages 135–157. World Sci. Publishing, River Edge, NJ, 1998.

[83] D. Petz.
Covariance and Fisher information in quantum mechanics.
J. Phys. A, 35(4):929–939, 2002.

[84] D. Petz.
Monotonicity of quantum relative entropy revisited.
Rev. Math. Phys., 15(1):79–91, 2003.

[85] D. Petz és H. Hasegawa.
On the Riemannian metric of α-entropies of density matrices.
Lett. Math. Phys., 38(2):221–225, 1996.

[86] D. Petz és M. B. Ruskai.
Contraction of generalized relative entropy under stochastic mappings on matrices.
Infin. Dimens. Anal. Quantum Probab. Relat. Top., 1(1):83–89, 1998.

[87] D. Petz és G. Tóth.
The Bogoliubov inner product in quantum statistics.
Lett. Math. Phys., 27(3):205–216, 1993.

[88] E. C. Pielou.
Ecological diversity.
Wiley, New York, 1975.

[89] C. R. Rao.
Information and accuracy atainable in the estimation of statistical parameters.
Bulletin of the Calcutta Mathematical Society, 37:81–91, 1945.

References XII

[90] C. R. Rao.
Diversity and dissimilarity coefficients: a unified approach.
Theoretic Population Biology, 21:24–43, 1982.

[91] G. Ruppeiner.
Riemannian geometry in thermodynamic fluctuation theory.
Rev. Modern Phys., 67(3):605–659, 1995.

[92] M. B. Ruskai.
Beyond strong subadditivity? Improved bounds on the contraction of generalized relative entropy.
Rev. Math. Phys., 6(5A):1147–1161, 1994.
With an appendix on applications to logarithmic Sobolev inequalities, Special issue dedicated to Elliott H. Lieb.

[93] K. Sailer.
Nemegyensúlyi statisztikus fizika.
Kossuth Lajos Tudományegyetem, Debrecen, 1994.
Speciális előadások.

[94] W. F. Stinespring.
Positive functions on C^*-algebras.
Proc. Amer. Math. Soc., 6:211–216, 1955.

[95] R. F. Streater.
Statistical dynamics and information geometry.
In *Geometry and nature (Madeira, 1995)*, volume 203 of *Contemp. Math.*, pages 117–131. Amer. Math. Soc., Providence, RI, 1997.

[96] T. Tanaka.
Information geometry of mean-field approximation.
In *Advanced mean field methods (Birmingham, 1999)*, Neural Inf. Process. Ser., pages 259–273. MIT Press, Cambridge, MA, 2001.

[97] I. J. Taneja.
Generalised information measures and their applications.
preprint.
http://www.mtm.ufsc.br/~taneja/bhtml/bhtml.html.

References XIII

[98] F. Topsoe.
Some inequalities for information divergence and related measures of discrimination.
Res. Rep. Coll., RGMIA, 2(1):85–98, 1999.

[99] A. Uhlmann.
Relative entropy and the Wigner-Yanase-Dyson-Lieb concavity in an interpolation theory.
Comm. Math. Phys., 54(1):21–32, 1977.

[100] A. Uhlmann.
Parallel transport and "quantum holonomy" along density operators.
Rep. Math. Phys., 24(2):229–240, 1986.

[101] A. Uhlmann.
Density operators as an arena for differential geometry.
In *Proceedings of the XXV Symposium on Mathematical Physics (Toruń, 1992)*, volume 33, pages 253–263, 1993.

[102] A. Uhlmann.
Geometric phases and related structures.
In *Proceedings of the XXVII Symposium on Mathematical Physics (Toruń, 1994)*, volume 36, pages 461–481, 1995.

[103] A. Uhlmann.
Spheres and hemispheres as quantum state spaces.
J. Geom. Phys., 18(1):76–92, 1996.

[104] H. Umegaki.
Conditional expectation in an operator algebra. IV. Entropy and information.
Kōdai Math. Sem. Rep., 14:59–85, 1962.

[105] E. P. Wigner és Mutsuo M. Yanase.
Information contents of distributions.
Proc. Nat. Acad. Sci. U.S.A., 49:910–918, 1963.

[106] W. K. Wooters.
Statistical distance and Hilbert space.
Physical Review D, 23:357–362, 1981.

Thank you for your attention!