Entropy Distance: New Quantum Phenomena

A. Knauf S. Weis

Department Mathematik
University of Erlangen

Information Geometry and its Applications III
Leipzig, 6th August 2010

Entropy Distance as an Information Functional

Definition

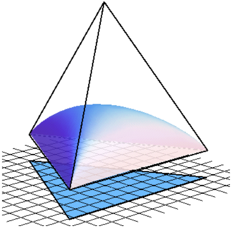
Entropy distance $d_{\mathcal{E}}$ is the relative entropy distance from an exponential family $\mathcal{E}$ in a finite-dimensional matrix algebra $\mathcal{A}$. Classical algebra $\mathcal{A} \cong \mathbb{C}^N$, quantum otherwise.

Overview of exponential families in statistics: Amari and Nagaoka, *Methods of information geometry* (2000).
Applications of entropy distance include

- MLE through the log-likelihood function (classical),
- the stochastic interaction measure of multi-information, this is the entropy distance from the independence model (classical & quantum).

Previous Work on $d_{\mathcal{E}}$ in Classical Probability Theory

Barndorff-Nielsen ('78),
Čencov ('82),
Ay ('02),
Csiszár and Matúš ('03, '05, '08)

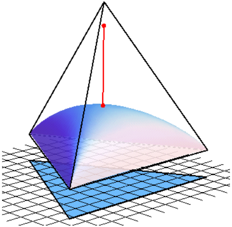


Rough Idea:

- Pythagorean theorem of relative entropy implies projection theorem along the normal space,
- mean value chart maps $\mathcal{E}$ to the mean value set (convex support),
- extension of $\mathcal{E}$ implies optimal projection theorem and optimal Pythagorean theorem.

Previous Work on $d_{\mathcal{E}}$ in Classical Probability Theory

Barndorff-Nielsen ('78),
Čencov ('82),
Ay ('02),
Csiszár and Matúš ('03, '05, '08)

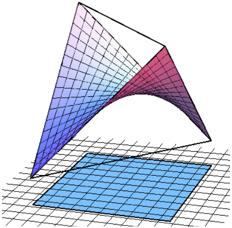


Rough Idea:

- Pythagorean theorem of relative entropy implies projection theorem along the normal space,
- mean value chart maps $\mathcal{E}$ to the mean value set (convex support),
- extension of $\mathcal{E}$ implies optimal projection theorem and optimal Pythagorean theorem.

Previous Work on $d_{\mathcal{E}}$ in Classical Probability Theory

Barndorff-Nielsen ('78),
Čencov ('82),
Ay ('02),
Csiszár and Matúš ('03, '05, '08)



Rough Idea:

- Pythagorean theorem of relative entropy implies projection theorem along the normal space,
- mean value chart maps $\mathcal{E}$ to the mean value set (convex support),
- extension of $\mathcal{E}$ implies optimal projection theorem and optimal Pythagorean theorem.

Previous Work on $d_{\mathcal{E}}$ in the Quantum Case

Classical extension of $\mathcal{E}$ is compact, the mean value set is a polytope and $d_{\mathcal{E}}$ is continuous. All wrong for quantum case!

Well-known in the quantum case:

- Pythagorean theorem implies (only in a restriction!) a projection theorem along the normal space, see Petz ('08),
- mean value chart and mean value set are known, see Wichmann ('63).

New in our work: Convex structure of mean value set includes non-exposed faces; extensions of $\mathcal{E}$; optimal projection theorem and Pythagorean theorem.

Exponential Families

Definition

State space $\overline{\mathcal{S}} := \{\rho \in \mathcal{A} \mid \rho \geq 0, \mathrm{tr}(\rho) = 1\}$. Real vector space $\mathcal{A}_{\mathrm{sa}} \subset \mathcal{A}$ of self-adjoint matrices, Hilbert-Schmidt inner product. Analytic diffeo. $\exp_1 : \mathcal{A}_{\mathrm{sa}}/\mathbb{R}\mathbb{1} \to \mathcal{S} := \{\rho \in \overline{\mathcal{S}} | \rho^{-1}$ exists $\}$, $a \mapsto \frac{e^a}{\mathrm{tr}(e^a)}$, canonical chart $\ln_0 = \exp_1^{-1}$ to traceless matrices.
Exponential family $\mathcal{E} := \exp_1($ linear subspace of $\mathcal{A}_{\mathrm{sa}}$ $)$,
$V := \ln_0(\mathcal{E})$ tangent space, $V^\perp$ normal space,
orth. projection $\pi_V : \mathcal{A}_{\mathrm{sa}} \to V$, mean value set $\mathrm{mv}(V) := \pi_V(\overline{\mathcal{S}})$.

Theorem (Wichmann '63)

$\pi_V \circ \exp_1|_V : V \to \mathrm{Int}(\mathrm{mv}(V))$ *real analytic bijection.*

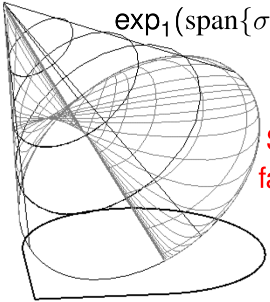
Two Examples

Definition (Mean value chart)

The analytic chart $\pi_V|_{\mathcal{E}}$ onto $\mathrm{Int}(\mathrm{mv}(V))$ is called **mean value chart**.

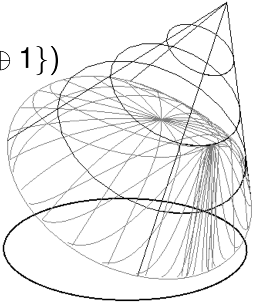
$\exp_1(\mathrm{span}\{\sigma_1 \oplus 1, \sigma_2 \oplus 1\})$

$\exp_1(\mathrm{span}\{\sigma_1 \oplus 0, \sigma_2 \oplus 1\})$



Swallow family

Staffelberg family



$\mathcal{A} := \mathrm{Mat}(2, \mathbb{C}) \oplus \mathbb{C}$

2D Mean Value Sets in $\mathcal{A} := \mathrm{Mat}(2, \mathbb{C}) \oplus \mathbb{C}$

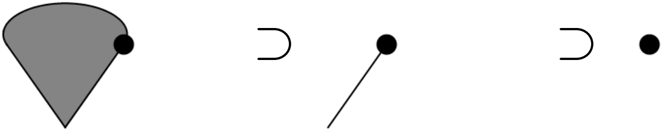
- $\overline{\mathcal{S}}$ is the 4D cone $\mathrm{conv}($ Bloch ball $\oplus\, 0, 0_2 \oplus 1)$,
- every 2D exponential family $\mathcal{E}$ of $\mathcal{A}$ is included in a 3D cone,
- modulo automorphism take $z := -\frac{1}{2}\mathbb{1}_2 \oplus 1$, $W := \mathrm{span}\{\sigma_1 \oplus 0, \sigma_2 \oplus 0, z\}$, $V \subset W$ and the cone $C := (\frac{1}{3}\mathbb{1} + W) \cap \overline{\mathcal{S}} = \overline{\exp_1(W)} \supset \mathcal{E}$,
- a complete orbit invariant of 2D planes V is the angle $\measuredangle(V, z) \in [0, \frac{\pi}{2}]$.

Exposed Faces, Poonems and Access Sequences

Definition (Rockafellar, Grünbaum, Csiszár and Matúš)

Let M be a convex set. The intersection F of M with a supporting hyperplane of M is called exposed face of M. $\emptyset$ and M are expsed faces by definition. In these cases we write $F \overset{\text{Ex}}{<} M$. A sequence $F = F_1 \overset{\text{Ex}}{<} \cdots \overset{\text{Ex}}{<} F_k \overset{\text{Ex}}{<} M$ is called access sequence and F is called poonem of M. A poonem, which is not an exposed face is called non-exposed face.

Concept of poonem is equivalent to the more popular concept of face.



The Pythagorean Theorem (Restricted Form)

Definition

The relative entropy of two states $\rho, \sigma \in \overline{\mathcal{S}}$ is $S(\rho, \sigma) := \infty$ unless $\mathrm{Im}(\sigma) \supset \mathrm{Im}(\rho)$ and then $S(\rho, \sigma) := \mathrm{tr}\rho(\ln \rho - \ln \sigma)$.

Theorem (Monograph Petz '08 for an overview)

If $\rho, \sigma, \tau \in \overline{\mathcal{S}}$ are states, σ and τ are invertible and $\rho - \sigma \perp \ln(\tau) - \ln(\sigma)$ holds, then we have $S(\rho, \sigma) + S(\sigma, \tau) = S(\rho, \tau)$.

- The Pythagorean theorem induces the projection $\pi_{\mathcal{E}} : \mathcal{E} + V^{\perp} \to \mathcal{E}$, $a \mapsto (a + V^{\perp}) \cap \mathcal{E}$ to an exponential family $\mathcal{E}$.
- For a state $\rho \in \mathcal{E} + V^{\perp}$ we have $d_{\mathcal{E}}(\rho) := \inf_{\sigma \in \mathcal{E}} S(\rho, \sigma) = S(\rho, \pi_{\mathcal{E}}(\rho))$.

Lattice Isomorphisms and Compressions

Assignment of an orth. projector $F \mapsto p^F$ to each face F of $\mathrm{mv}(V)$ through

- inverse projection $F \mapsto (F + V^\perp) \cap \overline{\mathcal{S}}$ lifts faces of $\mathrm{mv}(V)$ to faces of $\overline{\mathcal{S}}$,
- (face lattice of $\overline{\mathcal{S}}$) $\cong$ (projector lattice of $\mathcal{A}$).

Definition

Orth. projector p defines projection $\mathcal{A} \to p\mathcal{A}p := \{pap \mid a \in \mathcal{A}\}$ by $a \mapsto pap$. We denote by $\exp_1^p$ and $\ln^p$ the trace normalized exponential and the logarithm in $p\mathcal{A}p$.
A face F of $\mathrm{mv}(V)$ defines the compressed exponential family $\mathcal{E}^{p^F} := \exp_1^{p^F}(p^F V p^F)$.

The Geodesic Closure

Definition

The e-geodesic through $\theta \in \mathcal{A}_{\mathrm{sa}}$ in the direction of $v \in \mathcal{A}_{\mathrm{sa}}$ is the curve $\gamma \mapsto g_{\theta,v}(\lambda) = \exp_1(\theta + \lambda v) \subset \mathcal{S}$.

If p is the maximal projector of v (spectral projector of the largest eigenvalue), then $\mathcal{E}^p = \{\lim_{\lambda\to\infty} g_{\theta,v}(\lambda) \mid \theta \in V\}$.

Definition

The geodesic closure of $\mathcal{E}$ is $\mathrm{cl}_{\mathrm{geo}}(\mathcal{E}) := \bigcup_F \mathcal{E}^{p^F}$. Here F extends over the exposed faces ($\neq \emptyset$) of $\mathrm{mv}(V)$.

The geodesic closure $\mathrm{cl}_{\mathrm{geo}}(\mathcal{E})$ exceeds $\mathcal{E}$ by the limit points of e-geodesics in $\mathcal{E}$.

The rI-Closure

Definition (Csiszár and Matúš)

The rI-closure of $\mathcal{E}$ is $\mathrm{cl}_{\mathrm{rI}}(\mathcal{E}) := \{\rho \in \overline{\mathcal{S}} \mid d_{\mathcal{E}}(\rho) = 0\}$.

- E-geodesic asymptotics show $\mathrm{cl}_{\mathrm{geo}}(\mathcal{E}) \subset \mathrm{cl}_{\mathrm{rI}}(\mathcal{E})$.
- The proof idea to the following theorem is to concatenate e-geodesic asymptotics. For every poonem F in an access sequence of $\mathrm{mv}(V)$ we form the geodesic closure of the compressed exponential family $\mathcal{E}^{p^F}$ and take the union:

$$\mathrm{cl}_{\mathrm{rI}}(\mathcal{E}) = \bigcup_F \mathcal{E}^{p^F}.$$

- Equality $\mathrm{cl}_{\mathrm{geo}}(\mathcal{E}) = \mathrm{cl}_{\mathrm{rI}}(\mathcal{E})$ holds if and only if all faces of $\mathrm{mv}(V)$ are exposed. Examples: independence model, convex family or classical algebra.

The Optimal Projection Theorem

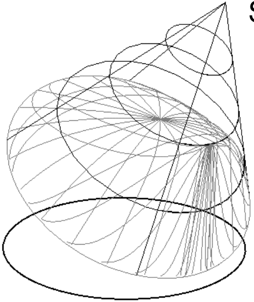
Theorem (Weis '09)

Let $\mathcal{E}$ be an exponential family with tangent space V.

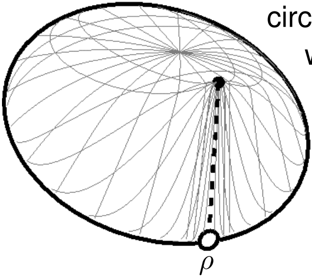
(1) *If $\rho \in \overline{\mathcal{S}}$, then $\rho + V^{\perp}$ intersects the reverse information closure $\mathrm{cl}_{\mathrm{rI}}(\mathcal{E})$ in a unique point denoted by $\pi_{\mathcal{E}}(\rho)$.*

(2) *If $\rho \in \overline{\mathcal{S}}$, then the relative entropy $S(\rho, \cdot)$ has a unique local minimum on $\mathrm{cl}_{\mathrm{rI}}(\mathcal{E})$ and*
$\min_{\sigma \in \mathrm{cl}_{\mathrm{rI}}(\mathcal{E})} S(\rho, \sigma) = S(\rho, \pi_{\mathcal{E}}(\rho)) = S_{\mathcal{E}}(\rho)$.

A Discontinuous Entropy Distance is Possible



Staffelberg family: Discontinuity on the base circle of the cone C where $d_{\mathcal{E}} \equiv 0$ except $d_{\mathcal{E}}(\rho) = \ln(2)$ for $\rho := \frac{1}{2}(\mathbb{1}_2 \oplus \sigma_2) \oplus 0$.

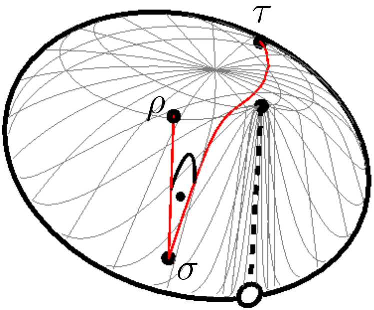


- The discontinuous $d_{\mathcal{E}}$ at $\varphi = \measuredangle(V, z) = \frac{\pi}{3}$ separates mean value sets with non-exposed faces from those without.
- According to the Pinsker-Csiszár inequality $\mathrm{cl}_{\mathrm{rI}}(\mathcal{E}) \subset \overline{\mathcal{E}}$ holds. Equality $\mathrm{cl}_{\mathrm{rI}}(\mathcal{E}) = \overline{\mathcal{E}} \iff d_{\mathcal{E}}$ is continuous. E.g. indep. model, convex family or (Ay '02) classical algebra.

The Optimal Pythagorean Theorem

Theorem (Weis '10)

Let $\mathcal{E}$ be an exponential family with tangent space V. If $\rho \in \overline{\mathcal{S}}$ and $\sigma, \tau \in \mathrm{cl}_{\mathrm{rI}}(\mathcal{E})$ are states such that $\sigma - \rho \perp V$, then $S(\rho, \sigma) + S(\sigma, \tau) = S(\rho, \tau)$ holds.



Maximization of Entropy Distance

Previous work e.g. by Ay, Knauf, Matúš

Definition

If $\rho \in \overline{\mathcal{S}}$ then the support projector of ρ is the ortho. projector $s(\rho) \in \mathcal{A}$ with the image $\mathrm{Im}(s(\rho)) = \mathrm{Im}(\rho)$. For $a \in \mathcal{A}_{\mathrm{sa}}$ denote the free energy $F(a) := \ln \mathrm{tr} e^a$ and use F^p for free energy in $p\mathcal{A}p$.

Theorem (Knauf and Weis '10)

Let $\rho \in \overline{\mathcal{S}}$, $p := s(\rho)$ and $q := s(\pi_{\mathcal{E}}(\rho))$. For every traceless self-adjoint matrix $u \in p\mathcal{A}p$ we have the directional derivative $\mathrm{D}|_\rho d_{\mathcal{E}}(u) = \langle u, \ln^p(\rho) - \ln^q \circ \pi_{\mathcal{E}}(\rho)\rangle$.
If ρ is a local maximizer of $d_{\mathcal{E}}$ then for $\theta := \ln^q \circ \pi_{\mathcal{E}}(\rho)$ we have $\rho = \exp_1^p(p\theta p)$ and $d_{\mathcal{E}}(\rho) = F^q(\theta) - F^p(p\theta p)$.

Maximum Entropy (Wichmann '63, Ingarden et al. '97)

Let $U \subset \mathcal{A}_{\mathrm{sa}}$ be a vector space and let $u \in U$.

Definition

The von Neumann entropy of a state $\rho \in \overline{\mathcal{S}}$ is $S(\rho) := -\mathrm{tr}\rho \ln \rho$. The constraint set of (U, u) is $C_{U,u} := \{\rho \in \overline{\mathcal{S}} \mid \pi_U(\rho) = u\}$. Let $\mathcal{E} := \exp_1(U)$.

Theorem (Weis '09)

We have $\mathrm{argmax}_{\rho \in C_{U,u}} S(\rho) = C_{U,u} \cap \mathrm{cl}_{\mathrm{rI}}(\mathcal{E})$.

Idea: replace $\max_\rho S(\rho) - \ln \mathrm{tr}\mathbb{1} = \max_\rho -S(\rho, \frac{\mathbb{1}}{\mathrm{tr}\mathbb{1}})$ by $\min_\rho S(\rho, \frac{\mathbb{1}}{\mathrm{tr}\mathbb{1}})$ and use Pythagorean theorem $\min_\rho [S(\rho, \pi_\mathcal{E}(\rho)) + S(\pi_\mathcal{E}(\rho), \frac{\mathbb{1}}{\mathrm{tr}\mathbb{1}})] = S(\pi_\mathcal{E}(\rho), \frac{\mathbb{1}}{\mathrm{tr}\mathbb{1}})$.

Coordinates for Maximum Entropy Ensembles

We fix $a_1, \dots, a_k \in \mathcal{A}_{\mathrm{sa}}$ (observables) and put $V := \mathrm{span}\{a_1, \dots, a_k, \mathbb{1}\} \cap \{\mathrm{tr}(\cdot) = 0\}$. Then we choose a tuple of mean values $(\xi_i)_{i=1}^k \in \{\{\mathrm{tr}(a_i\rho)\}_{i=1}^k \mid \rho \in \overline{\mathcal{S}}\} \cong \mathrm{mv}(V)$.

Theorem (Weis '10)

There exists a unique face G of $\mathrm{mv}(V)$ and unique coefficients $\beta_1, \dots, \beta_k \in \mathbb{R}$ such that for $a(\beta) := -\sum_{i=1}^k \beta_i p^G a_i p^G$ and for $j = 1, \dots, k$ we have

$$-\frac{\partial}{\partial \beta_j} F(a(\beta)) = \xi_j \langle p^G, \exp_1(a(\beta)) \rangle.$$

The maximizer of von Neumann entropy S with mean $(\xi_i)_{i=1}^k$ is $\rho := \exp_1^{p^G}(a(\beta))$ and $S(\rho) = \sum_{i=1}^k \beta_i \xi_i + F^{p^G}(a(\beta))$.

General Properties of the Entropy Distance $d_{\mathcal{E}}$

- Entropy distance $d_{\mathcal{E}}(\rho) = S(\rho, \pi_{\mathcal{E}}(\rho))$ of a state ρ given by projection $\pi_{\mathcal{E}} : \overline{\mathcal{S}} \to \mathrm{cl}_{\mathrm{rI}}(\mathcal{E})$ along the normal space $V^{\perp}$.
- Optimal Pythagorean theorem, for states $\rho \in \overline{\mathcal{S}}$ and $\tau \in \mathrm{cl}_{\mathrm{rI}}(\mathcal{E})$ we have $S(\rho, \tau) = d_{\mathcal{E}}(\rho) + S(\pi_{\mathcal{E}}(\rho), \tau)$.
- Geo. closure $\mathrm{cl}_{\mathrm{geo}}(\mathcal{E}) \subset$ rI-closure $\mathrm{cl}_{\mathrm{rI}}(\mathcal{E}) \subset$ topo. closure $\overline{\mathcal{E}}$,
  - $\mathrm{cl}_{\mathrm{geo}}(\mathcal{E}) = \mathrm{cl}_{\mathrm{rI}}(\mathcal{E}) \iff$ no non-exposed faces,
  - $\mathrm{cl}_{\mathrm{rI}}(\mathcal{E}) = \overline{\mathcal{E}} \iff$ entropy distance is continuous.

General Properties of the Entropy Distance $d_{\mathcal{E}}$

- Entropy distance $d_{\mathcal{E}}(\rho) = S(\rho, \pi_{\mathcal{E}}(\rho))$ of a state ρ given by projection $\pi_{\mathcal{E}} : \overline{\mathcal{S}} \to \mathrm{cl}_{\mathrm{rI}}(\mathcal{E})$ along the normal space $V^{\perp}$.
- Optimal Pythagorean theorem, for states $\rho \in \overline{\mathcal{S}}$ and $\tau \in \mathrm{cl}_{\mathrm{rI}}(\mathcal{E})$ we have $S(\rho, \tau) = d_{\mathcal{E}}(\rho) + S(\pi_{\mathcal{E}}(\rho), \tau)$.
- Geo. closure $\mathrm{cl}_{\mathrm{geo}}(\mathcal{E}) \subset$ rI-closure $\mathrm{cl}_{\mathrm{rI}}(\mathcal{E}) \subset$ topo. closure $\overline{\mathcal{E}}$,
  - $\mathrm{cl}_{\mathrm{geo}}(\mathcal{E}) = \mathrm{cl}_{\mathrm{rI}}(\mathcal{E}) \iff$ no non-exposed faces,
  - $\mathrm{cl}_{\mathrm{rI}}(\mathcal{E}) = \overline{\mathcal{E}} \iff$ entropy distance is continuous.

General Properties of the Entropy Distance $d_{\mathcal{E}}$

- Entropy distance $d_{\mathcal{E}}(\rho) \;=\; S(\rho, \pi_{\mathcal{E}}(\rho))$ of a state ρ given by projection $\pi_{\mathcal{E}} : \overline{\mathcal{S}} \to \mathrm{cl}_{\mathrm{rI}}(\mathcal{E})$ along the normal space $V^{\perp}$.
- Optimal Pythagorean theorem, for states $\rho \in \overline{\mathcal{S}}$ and $\tau \in \mathrm{cl}_{\mathrm{rI}}(\mathcal{E})$ we have $S(\rho, \tau) \;=\; d_{\mathcal{E}}(\rho) + S(\pi_{\mathcal{E}}(\rho), \tau)$.
- Geo. closure $\mathrm{cl}_{\mathrm{geo}}(\mathcal{E}) \subset$ rI-closure $\mathrm{cl}_{\mathrm{rI}}(\mathcal{E}) \subset$ topo. closure $\overline{\mathcal{E}}$,
  - $\mathrm{cl}_{\mathrm{geo}}(\mathcal{E}) = \mathrm{cl}_{\mathrm{rI}}(\mathcal{E}) \iff$ no non-exposed faces,
  - $\mathrm{cl}_{\mathrm{rI}}(\mathcal{E}) = \overline{\mathcal{E}} \iff$ entropy distance is continuous.

General Properties of the Entropy Distance $d_{\mathcal{E}}$

- Entropy distance $d_{\mathcal{E}}(\rho) = S(\rho, \pi_{\mathcal{E}}(\rho))$ of a state ρ given by projection $\pi_{\mathcal{E}} : \overline{\mathcal{S}} \to \mathrm{cl}_{\mathrm{rI}}(\mathcal{E})$ along the normal space $V^{\perp}$.
- Optimal Pythagorean theorem, for states $\rho \in \overline{\mathcal{S}}$ and $\tau \in \mathrm{cl}_{\mathrm{rI}}(\mathcal{E})$ we have $S(\rho, \tau) = d_{\mathcal{E}}(\rho) + S(\pi_{\mathcal{E}}(\rho), \tau)$.
- Geo. closure $\mathrm{cl}_{\mathrm{geo}}(\mathcal{E}) \subset$ rI-closure $\mathrm{cl}_{\mathrm{rI}}(\mathcal{E}) \subset$ topo. closure $\overline{\mathcal{E}}$,
  - $\mathrm{cl}_{\mathrm{geo}}(\mathcal{E}) = \mathrm{cl}_{\mathrm{rI}}(\mathcal{E}) \iff$ no non-exposed faces,
  - $\mathrm{cl}_{\mathrm{rI}}(\mathcal{E}) = \overline{\mathcal{E}} \iff$ entropy distance is continuous.

General Properties of the Entropy Distance $d_{\mathcal{E}}$

- Entropy distance $d_{\mathcal{E}}(\rho) = S(\rho, \pi_{\mathcal{E}}(\rho))$ of a state ρ given by projection $\pi_{\mathcal{E}} : \overline{\mathcal{S}} \to \mathrm{cl}_{\mathrm{rI}}(\mathcal{E})$ along the normal space $V^{\perp}$.
- Optimal Pythagorean theorem, for states $\rho \in \overline{\mathcal{S}}$ and $\tau \in \mathrm{cl}_{\mathrm{rI}}(\mathcal{E})$ we have $S(\rho, \tau) = d_{\mathcal{E}}(\rho) + S(\pi_{\mathcal{E}}(\rho), \tau)$.
- Geo. closure $\mathrm{cl}_{\mathrm{geo}}(\mathcal{E}) \subset$ rI-closure $\mathrm{cl}_{\mathrm{rI}}(\mathcal{E}) \subset$ topo. closure $\overline{\mathcal{E}}$,
  - $\mathrm{cl}_{\mathrm{geo}}(\mathcal{E}) = \mathrm{cl}_{\mathrm{rI}}(\mathcal{E}) \iff$ no non-exposed faces,
  - $\mathrm{cl}_{\mathrm{rI}}(\mathcal{E}) = \overline{\mathcal{E}} \iff$ entropy distance is continuous.

Open Questions

- What can be said about special families, e.g. independence model, Boltzmann machines?
- Where are the discontinuities of entropy distance?
- What are the one-sided directional derivatives?
- Are there connections to the geometry of entanglement?

Evidence

- A. Knauf and S. Weis.
  Entropy Distance: New Quantum Phenomena.
  arXiv:1007.5464
- S. Weis.
  The Pythagorean Theorem of Relative Entropy.
  arXiv:1003.5671
- S. Weis.
  Exponential Families with Incompatible Statistics and Their Entropy Distance, PhD thesis, Erlangen (2009).
  www.opus.ub.uni-erlangen.de/opus/volltexte/2010/1580

Thanks!

The Staffelberg Mountain