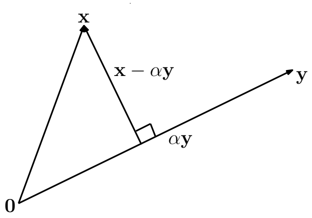
On the Equivalence of Projections in Relative α-Entropy and Rényi Divergence

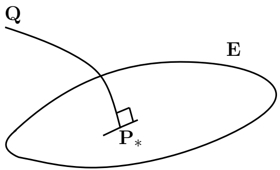
P. N. Karthik and Rajesh Sundaresan

Department of Electrical Communication Engineering

Indian Institute of Science, Bangalore

Projections





- Example: Given vectors $\mathbf{x}, \mathbf{y} \in \mathbb{R}^n$ of unit length, find

$$\alpha^* = \arg\min_{\alpha \in \mathbb{R}} ||\mathbf{x} - \alpha\mathbf{y}||_2^2$$

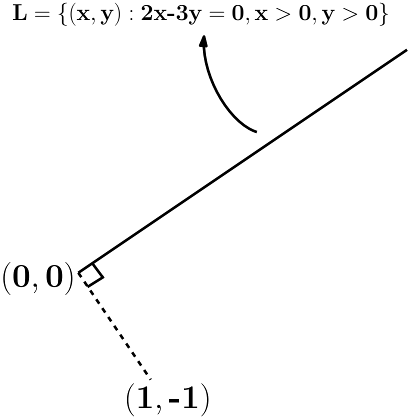
Answer: $\alpha^* = \langle \mathbf{x}, \mathbf{y} \rangle$

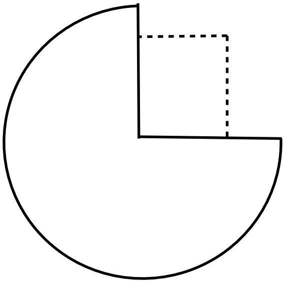
- More generally, given a point, find its *best* approximant from a set of points

Projections

- **Application:** In signal processing or communications when trying to find the best approximation to a noisy observation in the signal space
- It is not apriori clear that such an approximation can be found
- What about uniqueness of approximation?

Examples Where Existence or Uniqueness is Not Guaranteed





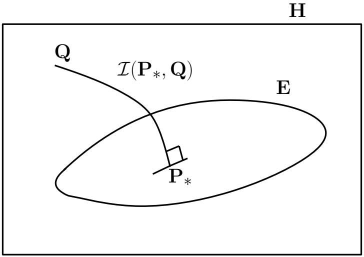
Approximant not in search space

Approximant not unique

Projections

- Additional information about the structure of the search space is needed to assert existence and/or uniqueness of projections
- Projection theorems provide this information

Projection Theorems

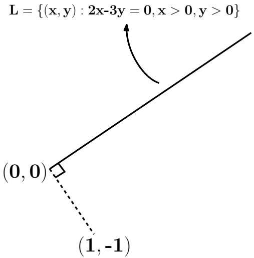


- Consider a set $\mathbb{H}$ equipped with the notion of a divergence $\mathcal{I}(P, Q)$ between any two points $P, Q \in \mathbb{H}$
- Projection of a point Q onto a set $\mathbb{E} \subset \mathbb{H}$ is a member $P_* \in \mathbb{H}$ that satisfies

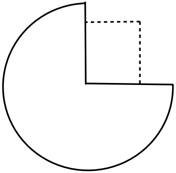
$$\mathcal{I}(P_*, Q) = \inf_{P \in \mathbb{E}} \mathcal{I}(P, Q)$$

- Question: What conditions on $\mathbb{E}$ ensure that
  - $P_* \in \mathbb{E}$ (existence), and
  - P_* is the only such point in $\mathbb{E}$ (uniqueness)?

Known Results



Approximant not in search space



Approximant not unique

- $\mathbb{H}$ = Hilbert space, $\mathcal{I}$ = metric induced by inner product. If $\mathbb{E}$ is closed and convex, then a projection exists and is unique
- $\mathbb{H}$ = space of probability measures on a measure space, $\mathcal{I}$ = relative entropy. If $\mathbb{E}$ is closed and convex, then a projection exists and is unique
- There are extensions in the latter context

Two Parametric Generalizations of Relative Entropy

- Fix $\alpha > 0$, $\alpha \neq 1$
- Given probability measures $P << \mu, Q << \mu$ with $p = \frac{dP}{d\mu}, q = \frac{dQ}{d\mu}$, we define *relative α-entropy* between P and Q as

$$\mathcal{I}_\alpha(P, Q) := \frac{\alpha}{1-\alpha} \log \left(\int \frac{p}{||p||} \left(\frac{q}{||q||} \right)^{\alpha - 1} d\mu \right)$$

- Similarly, we define *Rényi divergence of order α* between P and Q as

$$D_\alpha(P||Q) := \frac{1}{\alpha - 1} \log \left(\int p^\alpha q^{1-\alpha} d\mu \right).$$

- As $\alpha \to 1$,

$$\mathcal{I}_\alpha(P, Q), D_\alpha(P||Q) \to I(P||Q)$$

Fix $\alpha > 0$, $\alpha \neq 1$

Kumar and Sundaresan '15

- Let $Q << \mu$ be a probability measure, and $\mathbb{E}$ be a set of probability measures whose set of μ-densities is $\mathcal{E}$. Find P_* such that

$$\mathcal{I}_\alpha(P_*, Q) = \inf_{P \in \mathbb{E}} \mathcal{I}_\alpha(P, Q)$$

- Sufficient condition:

  $\mathbb{E}$ is convex

  $\mathcal{E}$ is closed in $L^\alpha(\mu)$

Kumar and Sason '16

- Let $Q << \mu$ be a probability measure, and $\mathbb{E}'$ be a set of probability measures whose set of μ-densities is $\mathcal{E}'$. Find P_* such that

$$D_\alpha(P_* || Q) = \inf_{P \in \mathbb{E}'} D_\alpha(P || Q)$$

- Sufficient condition:

  $\mathbb{E}'$ is α-convex

  $\mathcal{E}'$ is closed in $L^1(\mu)$

Convexity and α-Convexity

- **Convex:** A set $\mathcal{E}$ of densities is convex if:
  $\forall p_0, p_1 \in \mathcal{E}$ and $\forall \lambda \in (0,1)$,

  $$p_\lambda = \lambda p_1 + (1-\lambda) p_0$$

  belongs to $\mathcal{E}$
- **α-Convex:** A set $\mathcal{E}'$ of densities is α-convex if:
  $\forall p_0, p_1 \in \mathcal{E}'$ and $\forall \lambda \in (0,1)$,

  $$p_{\alpha,\lambda} \propto \left(\lambda \left(p_1\right)^\alpha + (1-\lambda) \left(p_0\right)^\alpha\right)^{1/\alpha}$$

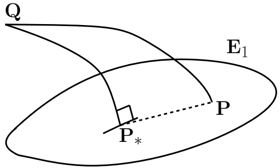
  belongs to $\mathcal{E}'$
- Convex $\equiv$ 1-convex

Equivalence of Projection Theorems

- Kumar and Sundaresan showed that $\mathscr{I}_\alpha$ (relative α-entropy) and D_α (Rényi divergence) are closely related
- This suggests that the hypotheses for existence and uniqueness of projections in these divergences may be equivalent to one another
- We explore this connection, and prove that this is indeed the case

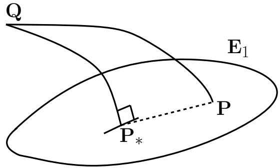
Pythagorean Property

$$\mathcal{I}_\alpha(\mathbf{P},\mathbf{Q}) = \mathcal{I}_\alpha(\mathbf{P},\mathbf{P}_*) + \mathcal{I}_\alpha(\mathbf{P}_*,\mathbf{Q})$$



- Kumar and Sundaresan demonstrated that relative α-entropy satisfies a “Pythagorean property” that uniquely characterizes the projection

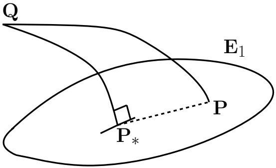
$$\mathbf{D}_\alpha(\mathbf{P}||\mathbf{Q}) = \mathbf{D}_\alpha(\mathbf{P}||\mathbf{P}_*) + \mathbf{D}_\alpha(\mathbf{P}_*||\mathbf{Q})$$



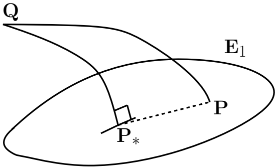
- van Erven and Harremöes showed that Pythagorean property holds for Rényi divergence
- This could be related to Kumar and Sundaresan’s result

Pythagorean Property

$$\mathcal{I}_\alpha(\mathbf{P},\mathbf{Q}) = \mathcal{I}_\alpha(\mathbf{P},\mathbf{P}_*) + \mathcal{I}_\alpha(\mathbf{P}_*,\mathbf{Q})$$



$$\mathbf{D}_\alpha(\mathbf{P}||\mathbf{Q}) = \mathbf{D}_\alpha(\mathbf{P}||\mathbf{P}_*) + \mathbf{D}_\alpha(\mathbf{P}_*||\mathbf{Q})$$



- We argue that this is indeed the case, and show the equivalence between the Pythagorean theorems for relative α-entropy and Rényi divergence

Two Families of Distributions

- The principle of (Shannon) entropy maximization is closely tied to that of minimizing relative entropy
- Minimizers of relative entropy are members of exponential families
- Minimizers of relative α-entropy and Rényi divergence are members of the following generalizations of exponential families:
  - Relative α-entropy: α-power law family
  - Rényi divergence: α-exponential family
- We show that the above generalized families are equivalent

Two Generalizations of Exponential Family

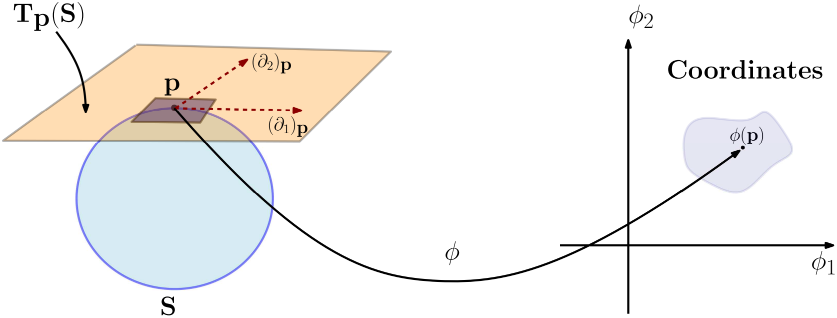
- Given a probability measure Q, $k \in \{1, 2, \ldots\}$ and $\Theta = \{\theta = (\theta_1, \ldots, \theta_k) : \theta_i \in \mathbb{R}\} \subset \mathbb{R}^k$, the *$\alpha$-power-law family* generated by Q and functions $f_i : \mathbb{X} \to \mathbb{R}$, $1 \le i \le k$, is defined as the set of probability measures $\mathcal{Z}^{(\alpha)} = \{P_\theta : \theta \in \Theta\}$, where

$$P_\theta(x)^{-1} \propto \left((Q(x))^{\alpha-1} + (1-\alpha)\sum_{i=1}^{k} \theta_i f_i(x)\right)^{\frac{1}{1-\alpha}} \quad \forall x \in \mathbb{X}$$

- Similarly, the *α-exponential family* generated by Q is defined as the set of probability measures $\mathcal{Y}^{(\alpha)} = \{P_\theta : \theta \in \Theta\}$, where

$$P_\theta(x) \propto \left((Q(x))^{1-\alpha} + (1-\alpha)\sum_{i=1}^{k} \theta_i f_i(x)\right)^{\frac{1}{1-\alpha}} \quad \forall x \in \mathbb{X}$$

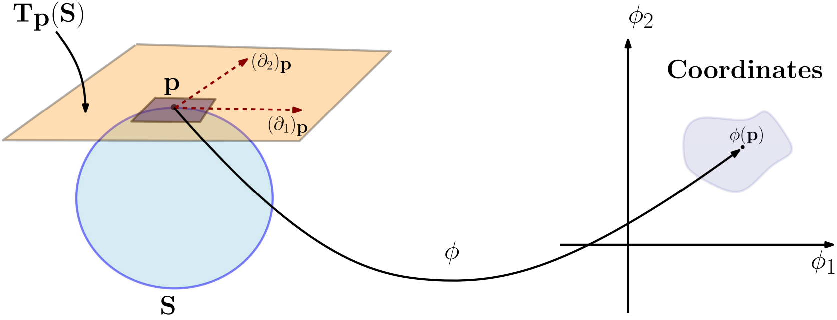
Riemannian Metrics From Divergences



$$\mathbf{g}_{ij}(\mathbf{p}) := \langle (\partial_i)_\mathbf{p}, (\partial_j)_\mathbf{p} \rangle_\mathbf{p}$$

- Eguchi suggested a method of defining a Riemannian metric on the manifold of probability distributions, starting from a divergence

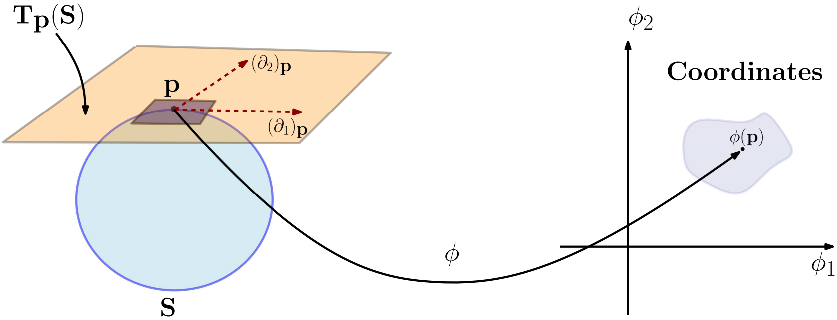
Riemannian Metrics From Divergences



$$\mathbf{g}_{ij}(\mathbf{p}) := \langle (\partial_i)_\mathbf{p}, (\partial_j)_\mathbf{p} \rangle_\mathbf{p}$$

- It is well known that Eguchi's method with relative entropy as the divergence results in the Fisher metric

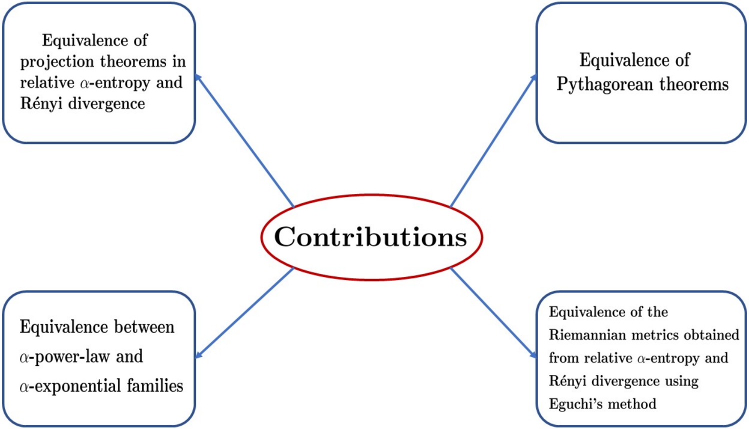
Riemannian Metrics From Divergences



$$\mathbf{g}_{ij}(\mathbf{p}) := \left\langle (\partial_i)_\mathbf{p}, (\partial_j)_\mathbf{p} \right\rangle_\mathbf{p}$$

- We apply Eguchi's method to relative α-entropy and Rényi divergence, and demonstrate the equivalence of the resulting Riemannian metrics

Summary of Contributions



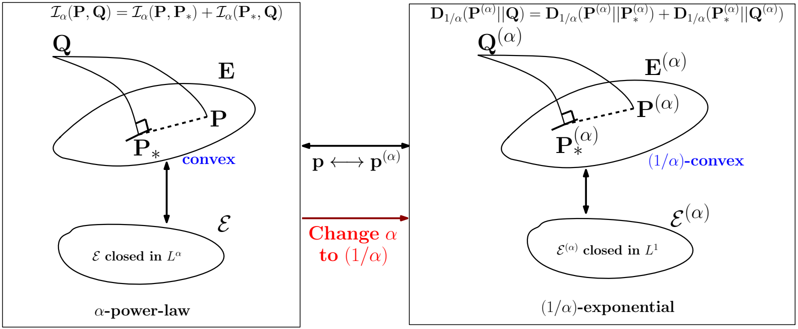
A Correspondence Relation on Probability Measures

- Given $P << \mu$ with $p = \frac{dP}{d\mu}$, we define its α-*scaled measure* $P^{(\alpha)}$ as a probability measure whose μ-density $p^{(\alpha)}$ satisfies

$$p^{(\alpha)} := \frac{p^\alpha}{\int p^\alpha d\mu}$$

- Two functions p and $p^{(\alpha)}$ related as above are said to be in one-one correspondence, denoted as $p \longleftrightarrow p^{(\alpha)}$

The Main Picture



The world of relative α-entropy

The world of Rényi divergence

Two Propositions

Proposition (1)

Fix $\alpha > 0$, $\alpha \neq 1$. A set $\mathbb{E}$ of probability measures absolutely continuous with respect to μ is convex if and only if the corresponding set of α-scaled measures $\mathbb{E}^{(\alpha)}$ is $(1/\alpha)$-convex

Proposition (2)

Fix $\alpha > 0$, $\alpha \neq 1$. Let $\mathbb{E}$ be a set of probability measures and let $\mathbb{E}^{(\alpha)}$ be the corresponding set of α-scaled measures. Let $\mathcal{E}$ and $\mathcal{E}^{(\alpha)}$ be the set of μ-densities associated with the probability measures in $\mathbb{E}$ and $\mathbb{E}^{(\alpha)}$ respectively. Then, $\mathcal{E}$ is closed in $L^{\alpha}(\mu)$ if and only if $\mathcal{E}^{(\alpha)}$ is closed in $L^{1}(\mu)$.

References

- M. A. Kumar and R. Sundaresan, "Minimization Problems Based on Relative α-Entropy I: Forward Projection," IEEE Transactions on Information Theory, vol. 61, no. 9, pp. 5063–5080, September 2015
- M. A. Kumar and I. Sason, "Projection Theorems for the Rényi Divergence on α-convex Sets," IEEE Transactions on Information Theory, vol. 62, no. 9, pp. 4924–4935, September 2016
- T. V. Erven and P. Harremoës, "Rényi Divergence and Kullback-Leibler Divergence," IEEE Trans. on Information Theory, vol. 60, no. 7, pp. 3797–3820, July 2014
- S. Eguchi, "Geometry of Minimum Contrast," Hiroshima Math. J, vol. 22, no. 3, pp. 631–647, 1992
- S. I. Amari and H. Nagaoka, "Methods of Information Geometry," Amer- ican Mathematical Soc., 2007, vol. 191

Thank you