On The Equivalence of Projections In Relative α-Entropy and Rényi Divergence

P. N. Karthik
Department of ECE
Indian Institute of Science
Bangalore, Karnataka 560012, India
Email: periyapatna@iisc.ac.in

Rajesh Sundaresan
Department of ECE,
Robert Bosch Centre for
Cyber-Physical Systems, and
Office of International Relations
Indian Institute of Science
Bangalore, Karnataka 560012, India
Email: rajeshs@iisc.ac.in

***Abstract*—The aim of this work is to establish that two recently published projection theorems, one dealing with a parametric generalization of relative entropy and another dealing with Rényi divergence, are equivalent under a correspondence on the space of probability measures. Further, we demonstrate that the associated "Pythagorean" theorems are equivalent under this correspondence. Finally, we apply Eguchi's method of obtaining Riemannian metrics from general divergence functions to show that the geometry arising from the above divergences are equivalent under the aforementioned correspondence.**

I. Introduction

Projections are often used in signal processing when one tries to find the best approximation in the signal space to a noisy observation. Such an exercise often entails a study of conditions on the signal space under which the approximation or projection can be uniquely determined. In this context, projection theorems provide sufficient conditions for the existence and uniqueness of projections. Such theorems fit into the following paradigm.

Consider a space $\mathbb{H}$ with a notion of a divergence $\mathscr{I}(P,Q)$ between any two points $P,Q \in \mathbb{H}$ that satisfies

$$\mathscr{I}(P,Q) \geq 0, \text{ with equality if and only if } P = Q. \quad (1)$$

The projection of a point Q onto a set $\mathbb{E} \subset \mathbb{H}$ is a member $P_* \in \mathbb{H}$ that satisfies

$$\mathscr{I}(P_*,Q) = \inf_{P \in \mathbb{E}} \mathscr{I}(P,Q), \quad (2)$$

and may be viewed as the best approximant of Q from the set $\mathbb{E}$.

(a) If $\mathbb{H}$ is a Hilbert space, $\mathscr{I}$ is the usual notion of distance $\langle P-Q, P-Q\rangle^{\frac{1}{2}}$ where $\langle\cdot,\cdot\rangle$ denotes the inner product, and if $\mathbb{E}$ is closed and convex, then a projection exists and is unique (see, for e.g., [1, Ch. 11, Th. 14]).

(b) If $\mathbb{H}$ is the space of probability measures on an abstract measure space, $\mathscr{I}$ is the relative entropy, and $\mathbb{E}$ is convex and closed with respect to the total variation metric, then a projection exists and is unique ([2, Th. 2.1]).

There are extensions in the latter context.

(c) In [3], $\mathbb{H}$ is the space of probability measures absolutely continuous with respect to some σ-finite measure μ and $\mathscr{I}$ is a parametric generalization of relative entropy, termed as relative α-entropy, and denoted $\mathscr{I}_\alpha$ for $\alpha > 0$, $\alpha \neq 1$; see Def. 2 later. If $\mathbb{E}$ is convex and its corresponding set of μ-densities is $L^\alpha(\mu)$-closed, then a projection exists and is unique ([3, Th. 8]).

(d) In [4], $\mathbb{H}$ is as in (c) and $\mathscr{I}$ is the Rényi divergence (see Def. 3 later) of order α, denoted D_α and defined for $\alpha > 0$, $\alpha \neq 1$. If $\mathbb{E}$ is α-convex (see Def. 6 later) and its corresponding set of μ-densities is $L^1(\mu)$-closed, then a projection exists and is unique ([4, Th. 1]).

From [3, Lemma 2.(c)], we know that the relative α-entropy between two probability measures is equal to the Rényi divergence of order $1/\alpha$ between the corresponding α-scaled measures (see Def. 1 later). This suggests that the hypotheses in (c) for the existence and uniqueness of projections for probability measures may be equivalent to those in (d) for the corresponding α-scaled measures, with α in (d) replaced by $1/\alpha$. In this paper, we explore this connection, and prove that this is indeed the case by establishing the equivalence between the hypotheses in items (c) and (d) above. For connections of these divergence measures to information theory and statistical estimation, we refer the reader to the introduction sections of [3] and [4].

When $\mathbb{H}$ is the space of probability measures, relative α-entropy satisfies a "Pythagorean property" that uniquely characterizes the projection [3], [5], [6]. Recently, van Erven and Harremöes [7] and [4] showed that an analogous property holds for Rényi divergence. The authors of [7] hinted at a plausible relation between their result with those of [5] and [6]. We argue in this paper that this is indeed the case, and show the equivalence between the Pythagorean theorems appearing in [3] and [7].

For a probability measure Q on a finite alphabet, [3] showed that when $\mathbb{E}$ is a linear family (see Def. 10 later), the projection is a member of the α-power-law family generated by Q (see Def. 8 later). Likewise, [4] showed that when $\mathbb{E}$ is an α-linear family (see Def. 11 later), the projection is a member of the α-exponential family generated by Q (see Def. 9 later). We prove that (i) $\mathbb{E}$ is linear iff $\mathbb{E}^{(\alpha)}$, the set of α-scaled measures associated with $\mathbb{E}$, is $(1/\alpha)$-linear, and (ii) the α-power-law family generated by Q is equivalent to the $(1/\alpha)$-exponential family generated by the α-scaled measure of Q.

Towards the study of the geometric structure of statistical models under general divergences, Eguchi [8] suggested a method of defining a Riemannian metric on statistical manifolds (see Def. 12 later) from a general divergence function. It is well known that Eguchi's method with relative entropy as the divergence results in a metric that is specified by the Fisher

978-1-5386-1224-8/18/$31.00 ©2018 IEEE

information matrix; see for example [9, Sec. 2.2 and Sec. 3.2]. We apply the same method to relative α-entropy and Rényi divergence, and show that under a suitable correspondence on the space of probability measures, the metrics specified in these cases are equivalent.

It is worthwhile to note that the q-exponential family of distributions studied by Amari and Ohara [10] bears a close resemblance with the α-exponential family considered in this paper, but for a normalization constant, with the parameter q in their paper corresponding to α of this paper. Furthermore, the Riemannian metric derived from Rényi divergence in Sec. III-D follows the procedure outlined in [10] for q-divergence. However, their definition of q-divergence is motivated by the construction of a divergence of Bregman's form starting from a convex function, while the divergence measures considered in this paper are motivated from estimation theory as described in the introduction sections of [3] and [4].

We now set up the basic notation and definitions in Section II, and present the main results in Section III. We conclude in Section IV.

II. Preliminaries

Let $(\mathbb{X}, \mathcal{X})$ be an abstract measure space, and let μ be any σ-finite measure on $(\mathbb{X}, \mathcal{X})$. Let P, Q be two probability measures absolutely continuous with respect to μ, denoted $P \ll \mu$, $Q \ll \mu$. Let $p = \frac{dP}{d\mu}$ and $q = \frac{dQ}{d\mu}$ denote the respective μ-densities. When $\mathbb{X}$ is finite, we take μ to be the counting measure. Consider $\alpha > 0$, $\alpha \neq 1$. Throughout this paper, we assume that p and q belong to the complete topological vector space $L^\alpha(\mu)$ defined by the metric

$$d(f,g) = \begin{cases} \left(\int |f-g|^\alpha d\mu\right)^{1/\alpha}, & \alpha > 1, \\ \int |f-g|^\alpha d\mu, & \alpha < 1. \end{cases} \tag{3}$$

We shall use the notation $||h|| = \left(\int h^\alpha d\mu\right)^{1/\alpha}$, even though $||\cdot||$, as defined, is not a norm when $\alpha < 1$. The dependence of $d(\cdot,\cdot)$ and $||\cdot||$ on α and μ is suppressed for convenience.

Definition 1 (α-scaled measure): Given a probability measure $P \ll \mu$ with μ-density p, its α-scaled measure $P^{(\alpha)}$ is the probability measure whose μ-density $p^{(\alpha)}$ is

$$p^{(\alpha)} := \frac{p^\alpha}{\int p^\alpha d\mu} = \left(\frac{p}{||p||}\right)^\alpha. \tag{4}$$

Definition 2 (Relative α-entropy): The relative α-entropy of P with respect to Q is defined as

$$\mathscr{I}_\alpha(P,Q) := \frac{\alpha}{1-\alpha} \log\left(\int \frac{p}{||p||}\left(\frac{q}{||q||}\right)^{\alpha-1} d\mu\right). \tag{5}$$

Definition 3 (Rényi divergence): The Rényi divergence of order α between P and Q is defined as

$$D_\alpha(P||Q) := \frac{1}{\alpha-1} \log\left(\int p^\alpha q^{1-\alpha} d\mu\right). \tag{6}$$

The relation between relative α-entropy and Rényi divergence is known to be [3, Lemma 2.(c)]

$$\mathscr{I}_\alpha(P,Q) = D_{1/\alpha}(P^{(\alpha)}||Q^{(\alpha)}), \tag{7}$$

where $P^{(\alpha)}$ and $Q^{(\alpha)}$ are the α-scaled measures of P and Q respectively. For a subset $\mathbb{E}$ of probability measures absolutely continuous with respect to μ, we denote the corresponding set of μ-densities by $\mathcal{E}$, i.e.,

$$\mathcal{E} := \left\{ p = \frac{dP}{d\mu} : P \in \mathbb{E} \right\}. \tag{8}$$

Also, we write $\mathbb{E}^{(\alpha)}$ for the set of α-scaled measures associated with the probability measures in $\mathbb{E}$, and $\mathcal{E}^{(\alpha)}$ for its corresponding set of μ-densities, i.e.,

$$\mathcal{E}^{(\alpha)} := \left\{ p^{(\alpha)} = \frac{p^\alpha}{\int p^\alpha d\mu} : p \in \mathcal{E} \right\}. \tag{9}$$

It thus follows that the probability measures in $\mathbb{E}^{(\alpha)}$ are absolutely continuous with respect to μ whenever those in $\mathbb{E}$ are. Further, the sets $\mathbb{E}$ and $\mathbb{E}^{(\alpha)}$ are in one-one correspondence, i.e., for each $P \in \mathbb{E}$ such that $\frac{dP}{d\mu} = p$, there exists a unique $P^{(\alpha)} \in \mathbb{E}^{(\alpha)}$ such that its μ-density $p^{(\alpha)}$ satisfies (4). Conversely, for each $P^{(\alpha)} \in \mathbb{E}^{(\alpha)}$ such that $\frac{dP^{(\alpha)}}{d\mu} = p^{(\alpha)}$, there exists a unique $P \in \mathbb{E}$ (upto equivalence with respect to μ) such that its μ-density p satisfies (4).

Definition 4 (The $p \longleftrightarrow p^{(\alpha)}$ correspondence): Given a probability measure $P \ll \mu$ with μ-density p, a function $p^{(\alpha)}$ is said to be in correspondence with p, denoted as $p \longleftrightarrow p^{(\alpha)}$, if it satisfies (4).

On account of this definition, we have a one-one correspondence between $\mathcal{E}$ and $\mathcal{E}^{(\alpha)}$ whenever $\mathbb{E}$ and $\mathbb{E}^{(\alpha)}$ are in one-one correspondence.

Definition 5 ($(\alpha,\lambda)-$ mixture): Given two probability measures $P_0, P_1 \ll \mu$ and $\lambda \in (0,1)$, the (α,λ)-mixture of P_0 and P_1 is the probability measure $P_{\alpha,\lambda}$ whose μ-density $p_{\alpha,\lambda}$ is

$$p_{\alpha,\lambda} := \frac{(\lambda(p_1)^\alpha + (1-\lambda)(p_0)^\alpha)^{1/\alpha}}{\int (\lambda(p_1)^\alpha + (1-\lambda)(p_0)^\alpha)^{1/\alpha} d\mu}. \tag{10}$$

Note that $p_{\alpha,\lambda}$ is well-defined since

$$Z := \int (\lambda(p_1)^\alpha + (1-\lambda)(p_0)^\alpha)^{1/\alpha} d\mu \tag{11}$$

is always strictly positive and finite. Indeed, for $\lambda \in (0,1)$,

$$0 \leq (\lambda(p_1)^\alpha + (1-\lambda)(p_0)^\alpha)^{1/\alpha} \leq \max\{p_0, p_1\} \leq p_0 + p_1, \tag{12}$$

which implies that $0 \leq Z \leq 2$. The first inequality in (12) holds with equality if and only if $p_1 \equiv 0$ and $p_0 \equiv 0$. Hence, for any non-trivial densities p_0 and p_1, and hence for probability densities, we have $Z > 0$.

Definition 6 (α-convex set): A set $\mathbb{E}$ of probability measures is said to be α-convex if for any $P_0, P_1 \in \mathbb{E}$ and $\lambda \in (0,1)$, the (α,λ)-mixture of P_0 and P_1 belongs to $\mathbb{E}$.

Definition 7 (α-exponential function): The α-exponential function $e_\alpha : \mathbb{R} \cup \{\infty\} \to \mathbb{R}_+ \cup \{\infty\}$ is defined as

$$e_\alpha(u) = \begin{cases} (\max\{1 + (1-\alpha)u, 0\})^{\frac{1}{1-\alpha}}, & \alpha \neq 1, \\ \exp(u), & \alpha = 1. \end{cases} \tag{13}$$

Definition 8 (α-power-law family): Given a probability measure Q (with full support when $\alpha > 1$), $k \in \{1, 2, \ldots\}$ and $\Theta = \{\theta = (\theta_1, \ldots, \theta_k) : \theta_i \in \mathbb{R}\} \subset \mathbb{R}^k$, the α-power-law family generated by Q and functions $f_i : \mathbb{X} \to \mathbb{R}$, $1 \leq i \leq k$, is defined as the set of probability measures

$$\mathcal{Z}^{(\alpha)} = \{P_\theta : \theta \in \Theta\}, \tag{14}$$

where

$$P_\theta(x)^{-1} = M(\theta)\ e_\alpha\left(\frac{(Q(x))^{\alpha-1}-1}{1-\alpha} + \sum_{i=1}^{k}\theta_i f_i(x)\right) \quad (15)$$

for $x \in \mathbb{X}$ with $M(\theta)$ being the normalisation constant. Assuming that the argument of e_α is strictly positive, using (13) in (15) yields

$$P_\theta(x)^{-1} = M(\theta)\left((Q(x))^{\alpha-1} + (1-\alpha)\sum_{i=1}^{k}\theta_i f_i(x)\right)^{\frac{1}{1-\alpha}} \quad (16)$$

for $x \in \mathbb{X}$.

Definition 9 (α-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 α-exponential 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

$$\mathcal{Y}^{(\alpha)} = \{P_\theta : \theta \in \Theta\}, \quad (17)$$

where

$$P_\theta(x) = (N(\theta))^{-1}\left((Q(x))^{1-\alpha} + (1-\alpha)\sum_{i=1}^{k}\theta_i f_i(x)\right)^{\frac{1}{1-\alpha}} \quad (18)$$

for $x \in \mathbb{X}$, with $N(\theta)$ being the normalisation factor. The forms (16) and (18) will be used in Sec. III-C.

Definition 10 (Linear family): For any given functions $f_1,\ldots,f_k$ on a finite alphabet $\mathbb{X}$, the family probability measures defined by

$$\mathbb{L} := \left\{P : \sum_{x\in\mathbb{X}} f_i(x)P(x) = 0,\ 1 \le i \le k\right\}, \quad (19)$$

if nonempty, is called a linear family.

Definition 11 (α-linear family): For any given functions $f_1,\ldots,f_k$ on a finite alphabet $\mathbb{X}$, the family probability measures defined by

$$\mathbb{L}_{(\alpha)} := \left\{P : \sum_{x\in\mathbb{X}} f_i(x)(P(x))^\alpha = 0,\ 1 \le i \le k\right\}, \quad (20)$$

if nonempty, is called an α-linear family.

Clearly, if P is a member of $\mathbb{L}$, then $P^{(\alpha)}$ is a member of $\mathbb{L}_{(1/\alpha)}$. The converse likewise holds.

Definition 12 (Statistical manifold): A statistical manifold S is a parametric family of probability distributions on $\mathbb{X}$ (with full support) with a continuously varying parameter space. It is usually represented as

$$S = \{p_\phi : \phi = (\phi_1,\ldots,\phi_n) \in \Phi \subset \mathbb{R}^n\}. \quad (21)$$

Here Φ is the parameter space. In writing (21), we note that given any $\phi \in \Phi$, there exists a unique $p_\phi \in S$, and vice-versa.

The mapping $p \mapsto (\phi_1(p),\ldots,\phi_n(p))$ is called a *coordinate system* for S. The *tangent space* at a point p on a manifold S, denoted as $T_p(S)$, is a linear space that corresponds to the linearization of the manifold around p; the elements of the tangent space are called *tangent vectors*. For a coordinate system ϕ, we denote the basis vectors of a tangent space $T_p(S)$ by $(\partial_i)_p = (\partial/\partial\phi_i)_p$, $i = 1,\ldots,n$. A *(Riemannian) metric* at a point $p \in S$ is an inner product defined between any two tangent vectors at that point. Although it is convenient to identify a metric with a point on the manifold, it is conventional to identify it with the coordinate $\phi(p) = (\phi_1(p),\ldots,\phi_n(p))$ of p. A metric is completely characterized by a matrix whose entries are the inner products between the basis tangent vectors, i.e., it is characterized by the matrix

$$G(\phi) = [g_{i,j}(\phi)]_{i,j=1,\ldots,n}, \quad (22)$$

where $g_{i,j} = \langle\partial_i, \partial_j\rangle$.

Let S be a manifold with a coordinate system $\phi = (\phi_1,\ldots,\phi_n)$, and let $\mathscr{I}$ be a divergence function on $S \times S$. We shall use the notation $\mathscr{I}(p,q)$ to denote the divergence $\mathscr{I}(P,Q)$ between the probability measures P and Q. Eguchi [8] showed that there is a metric

$$G^{(\mathscr{I})}(\phi) = [g^{(\mathscr{I})}_{i,j}(\phi)]_{i,j=1,\ldots,n} \quad (23)$$

with

$$g^{(\mathscr{I})}_{i,j}(\phi) = -\frac{\partial}{\partial\phi_i}\frac{\partial}{\partial\phi'_j}\mathscr{I}(p_\phi, p_{\phi'})\Big|_{\phi'=\phi}, \quad (24)$$

where $\phi = (\phi_1,\ldots,\phi_n)$ and $\phi' = (\phi'_1,\ldots,\phi'_n)$. In Sec. III-D, we evaluate (24) for the individual cases when $\mathscr{I}$ is either relative α-entropy or Rényi divergence, and thereafter demonstrate an equivalence between the two metrics.

Note: For the material presented in Sec. III-C and Sec. III-D, we assume that $\mathbb{X}$ is a finite set. If $\mathbb{X}$ is not a finite set, it is not clear if the normalization factor $M(\theta)$ in (16) or $N(\theta)$ in (18) is finite.

III. Main Results

We begin this section with two important propositions that will be used to establish the results later.

Proposition 1: Fix $\alpha > 0$, $\alpha \ne 1$. 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.

Proof: See Appendix A. ∎

Proposition 2: Fix $\alpha > 0$, $\alpha \ne 1$. 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 $L^1(\mu)$.

Proof: See Appendix B. ∎

A. Equivalence of the Projection Problems

We now consider the following two projection problems appearing in the works of [3] and [4] respectively:

(A) Fix $\alpha > 0$, $\alpha \ne 1$. Let Q be any probability measure, $Q \ll \mu$, and $\mathbb{E}$ be a set of probability measures whose set of μ-densities is $\mathcal{E}$. Solve

$$\inf_{P\in\mathbb{E}} \mathscr{I}_\alpha(P,Q). \quad (25)$$

(B) Fix $\alpha > 0$, $\alpha \ne 1$. Let Q be any probability measure, $Q \ll \mu$, and $\mathbb{E}_1$ be a set of probability measures whose set of μ-densities is $\mathcal{E}_1$. Solve

$$\inf_{P\in\mathbb{E}_1} D_\alpha(P||Q). \quad (26)$$

Recall from Sec. I.(c) that a sufficient condition proposed in [3] for the existence and uniqueness of solution to (25) is that

$$\mathbb{E} \text{ is convex and } \mathcal{E} \text{ is closed in } L^\alpha(\mu), \quad (27)$$

and from Sec. I.(d) that a sufficient condition proposed in [4] for the existence and uniqueness of solution to (26) is that

$$\mathbb{E}_1 \text{ is } \alpha\text{-convex and } \mathcal{E}_1 \text{ is closed in } L^1(\mu). \tag{28}$$

We now demonstrate that, under the $p \longleftrightarrow p^{(\alpha)}$ correspondence, the problems (25) and (26) are equivalent.

Theorem 1: The minimization problem (25) for a given $\alpha > 0$, $\alpha \neq 1$, is equivalent to (26) with α replaced by $1/\alpha$ and $\mathbb{E}_1$ replaced by $\mathbb{E}^{(\alpha)}$, the set of α-scaled measures corresponding to $\mathbb{E}$. Moreover, the hypotheses in (27) and (28) are identical under the $p \longleftrightarrow p^{(\alpha)}$ correspondence.

Proof: The proof is immediate from (7). However, we provide the details for completeness. The problem in (25) is

$$\inf_{P \in \mathbb{E}} \mathscr{I}_\alpha(P, Q). \tag{29}$$

Since (7) holds, under the $p \longleftrightarrow p^{(\alpha)}$ correspondence, the problem is equivalent to

$$\inf_{P^{(\alpha)} \in \mathbb{E}^{(\alpha)}} D_{1/\alpha}(P^{(\alpha)} || Q^{(\alpha)}), \tag{30}$$

which is (26), with $\mathbb{E}_1$ replaced by $\mathbb{E}^{(\alpha)}$ and α replaced by $1/\alpha$. Further, by Props. 1 and 2, the hypotheses in (27) and (28) are equivalent, with $\mathcal{E}_1$ replaced by $\mathcal{E}^{(\alpha)}$. ∎

B. Equivalence of the Pythagorean Theorems

We now argue the equivalence between the theorems on the "Pythagorean property" of relative α-entropy and Rényi divergence. The result [3, Th. $10.(a)$] establishes that if $\mathbb{E}$ is convex, then the projection P_* of Q onto $\mathbb{E}$, if it exists, satisfies

$$\mathscr{I}_\alpha(P, Q) \geq \mathscr{I}_\alpha(P, P_*) + \mathscr{I}_\alpha(P_*, Q) \text{ for all } P \in \mathbb{E}. \tag{31}$$

By virtue of $(1/\alpha)$-convexity of $\mathbb{E}^{(\alpha)}$ (Proposition 1) and (7), $P_*^{(\alpha)}$ is the $D_{1/\alpha}$-projection of $Q^{(\alpha)}$ onto $\mathbb{E}^{(\alpha)}$ and this projection satisfies

$$D_{1/\alpha}(P^{(\alpha)}, Q^{(\alpha)}) \geq D_{1/\alpha}(P^{(\alpha)}, P_*^{(\alpha)}) + D_{1/\alpha}(P_*^{(\alpha)}, Q^{(\alpha)}) \tag{32}$$

for all $P^{(\alpha)} \in \mathbb{E}^{(\alpha)}$. This recovers [7, Th. 14], as also [4, Prop. 1], with $1/\alpha$ replacing α.

C. Equivalence of α-Power-Law and α-Exponential Families

We now demonstrate that, under the $p \longleftrightarrow p^{(\alpha)}$ correspondence, the α-power-law family generated by a probability measure Q is equivalent to the α-exponential family generated by the α-scaled measure of Q.

Theorem 2: Let $\mathbb{X}$ be a finite alphabet. Fix $\alpha > 0$, $\alpha \neq 1$, $k \in \{1, 2, \ldots\}$, and $\Theta = \{\theta = (\theta_1, \ldots, \theta_k) : \theta_i \in \mathbb{R}\} \subset \mathbb{R}^k$. Let $f_i : \mathbb{X} \to \mathbb{R}$, $1 \leq i \leq k$, be specified. Given a probability measure Q, for every member of the α-power-law family generated by Q, $f_1, \ldots, f_k$ and Θ, its α-scaled measure is a member of the $(1/\alpha)$-exponential family generated by $Q^{(\alpha)}$, $f_1, \ldots, f_k$ and Θ', where Θ' is a scalar modification of Θ that depends on Q.

Proof: See Appendix C. ∎

D. Equivalence of the Riemannian Metrics for Relative α-Entropy and Rényi Divergence

It is well known from [9, Sec. 2.2 and Sec. 3.2] that when $\mathscr{I}(p, q) = I(p||q)$, the relative entropy between p and q, (24) can be written as

$$g_{i,j}^{(I)}(\phi) = E_{p_\phi}[\partial_i \log p_\phi, \partial_j \log p_\phi], \tag{33}$$

where E_{p_ϕ} denotes the expectation with respect to p_ϕ. The quantity in (33) is the (i, j) entry of the Fisher information matrix. Thus, with relative entropy as the divergence function, the Riemannian metric is the one specified by the Fisher information matrix.

On similar lines, when $\mathscr{I}(p, q) = D_\alpha(p||q)$, the Rényi divergence of order α between p and q, where $\alpha > 0$, $\alpha \neq 1$, using (6) in (24) for the finite alphabet setting results in the following set of equations:

$$\begin{aligned}
g_{i,j}^{(D_\alpha)}(\phi) &= -\frac{\partial}{\partial \phi_i}\frac{\partial}{\partial \phi_j'} D_\alpha(p_\phi || p_{\phi'}) \Big|_{\phi' = \phi} \\
&= -\frac{1}{\alpha - 1} \cdot \frac{\partial}{\partial \phi_i}\frac{\partial}{\partial \phi_j'} \log \left(\sum_{x \in \mathbb{X}} p_\phi(x)^\alpha p_{\phi'}(x)^{1-\alpha} \right) \Bigg|_{\phi' = \phi} \\
&= \frac{\partial}{\partial \phi_i} \left(\frac{\sum\limits_{x \in \mathbb{X}} p_\phi(x)^\alpha p_{\phi'}(x)^{-\alpha} \ \partial_j' p_{\phi'}(x)}{\sum\limits_{x \in \mathbb{X}} p_\phi(x)^\alpha p_{\phi'}(x)^{1-\alpha}} \right) \Bigg|_{\phi' = \phi} \\
&= \alpha \cdot E_{p_\phi}[\partial_i \log p_\phi \cdot \partial_j \log p_\phi] \\
&= \alpha \cdot g_{i,j}^{(I)}(\phi) \qquad (34) \\
&= g_{i,j}^{(\alpha I)}(\phi), \qquad (35)
\end{aligned}$$

where (34) follows from (33), and (35) follows by recognizing that (34) can be obtained from (24) by plugging $\mathscr{I}(p, q) = \alpha I(p, q)$. Our next result explores the consequences of the $p \longleftrightarrow p^{(\alpha)}$ correspondence.

Theorem 3: Consider a finite alphabet $\mathbb{X}$ and fix $\alpha > 0$, $\alpha \neq 1$. Let S be a statistical manifold equipped with a coordinate system $\phi = (\phi_1, \ldots, \phi_n)$, and let $S^{(\alpha)}$ denote the statistical manifold of the corresponding α-scaled measures. Then, for every $p \in S$, the Riemannian metric specified by relative α-entropy on $T_p(S)$ is equivalent to that specified by Rényi divergence of order $1/\alpha$ on $T_{p^{(\alpha)}}(S^{(\alpha)})$.

Proof: The proof is immediate from (7). ∎

By virtue of (35) and the above theorem, we can also conclude that for every $p \in S$, the Riemannian metric specified by relative α-entropy on $T_p(S)$ is equivalent to that specified by $\alpha^{-1} I$ on $T_{p^{(\alpha)}}(S^{(\alpha)})$, where I is the relative entropy.

IV. Conclusions

This paper studied two recently published projection theorems, one in relative α-entropy and the other in Rényi divergence. It was shown that the sufficient conditions for the existence and uniqueness of projections in relative α-entropy are equivalent to those for Rényi divergence under the $p \longleftrightarrow p^{(\alpha)}$ correspondence relation on the space of probability measures. Specifically, it was shown that a set $\mathbb{E}$ of probability measures is convex if and only if the corresponding set of α-scaled measures $\mathbb{E}^{(\alpha)}$ is $(1/\alpha)$-convex, and a set $\mathcal{E}$ of μ-densities is closed in the topological space $L^\alpha(\mu)$ if and only if the corresponding set of α-scaled densities $\mathcal{E}^{(\alpha)}$ is closed in $L^1(\mu)$.

Next, the Pythagorean theorems under the aforementioned divergences were shown to be equivalent under the same correspondence. For the case when the underlying alphabet $\mathbb{X}$ is finite, it was shown that the α-power law family of probability distributions is equivalent to the $(1/\alpha)$-exponential family of the corresponding α-scaled distributions. Finally, in the finite alphabet setting, the Riemannian metric specified by relative α-entropy on the statistical manifold was shown to be equivalent to that specified by Rényi divergence of order $1/\alpha$ on the corresponding manifold of α-scaled measures. Thus, owing to the $p \longleftrightarrow p^{(\alpha)}$correspondence, several independently established parallel results can now be viewed through a single lens. We leave for the future the exploration of the relations between the dual affine connections arising from the aforementioned divergences, and the connections thereof to the Pythagorean theorems stated in this paper.

V. Acknowledgment

This work was supported by the Robert Bosch Centre for Cyber-Physical Systems at the Indian Institute of Science and in part by the Science and Engineering Research Board, Department of Science and Technology [grant no. EMR/2016/002503].

Appendix

A. Proof of Proposition 1

We begin with the "only if" part. Suppose that $\mathbb{E}$ is a convex set of probability measures. Let $P_0^{(\alpha)}, P_1^{(\alpha)} \in \mathbb{E}^{(\alpha)}$ and $\lambda \in (0,1)$ be arbitrary. Let $P^{(\alpha)}_{\frac{1}{\alpha},\lambda}$ denote the $\left(\frac{1}{\alpha},\lambda\right)$-mixture of $P_0^{(\alpha)}$ and $P_1^{(\alpha)}$. Then, we need to show that $dP^{(\alpha)}_{\frac{1}{\alpha},\lambda}/d\mu = p^{(\alpha)}_{\frac{1}{\alpha},\lambda} \in \mathcal{E}^{(\alpha)}$. Using (10), we have

$$p^{(\alpha)}_{\frac{1}{\alpha},\lambda} \propto \left(\lambda(p_1^{(\alpha)})^{1/\alpha} + (1-\lambda)(p_0^{(\alpha)})^{1/\alpha}\right)^{\alpha} \tag{36}$$

$$\propto \left(\frac{\lambda}{||p_1||}p_1 + \frac{(1-\lambda)}{||p_0||}p_0\right)^{\alpha} \tag{37}$$

$$\propto (\lambda' p_1 + (1-\lambda')p_0)^{\alpha}, \tag{38}$$

where (37) follows from the application of (4) to $p_1^{(\alpha)}$ and $p_0^{(\alpha)}$ and (38) follows by setting

$$\lambda' = \frac{\frac{\lambda}{||p_1||}}{\frac{\lambda}{||p_1||} + \frac{1-\lambda}{||p_0||}} \tag{39}$$

and then absorbing the scaling in the normalisation constant. We now recognize that

$$\lambda' p_1 + (1-\lambda')p_0 = \frac{d\left(\lambda' P_1 + (1-\lambda')P_0\right)}{d\mu}, \tag{40}$$

and since $\mathbb{E}$ is convex by assumption, we have $\lambda' P_1 + (1-\lambda')P_0 \in \mathbb{E}$, which implies, by (8), that $\lambda' p_1 + (1-\lambda')p_0 \in \mathcal{E}$. Using this and the fact that (38) implies $p^{(\alpha)}_{\frac{1}{\alpha},\lambda} \longleftrightarrow (\lambda' p_1 + (1-\lambda')p_0)$, we conclude that $p^{(\alpha)}_{\frac{1}{\alpha},\lambda} \in \mathcal{E}^{(\alpha)}$, hence completing the proof of the "only if" part.

We now proceed to prove the "if part". Suppose that $\mathbb{E}^{(\alpha)}$ is $(1/\alpha)$-convex. We need to show that for any $P_0, P_1 \in \mathbb{E}$ and $\lambda \in (0,1)$, we have $\lambda P_1 + (1-\lambda)P_0 \in \mathbb{E}$. By definition, $P_0^{(\alpha)}, P_1^{(\alpha)} \in \mathbb{E}^{(\alpha)}$. Let $p_0^{(\alpha)} = dP_0^{(\alpha)}/d\mu$, $p_1^{(\alpha)} = dP_1^{(\alpha)}/d\mu$. Set

$$\lambda'' = \frac{\frac{\lambda}{||p_0||}}{\frac{\lambda}{||p_0||} + \frac{1-\lambda}{||p_1||}}. \tag{41}$$

Noting that $\lambda'' \in (0,1)$ and $\mathbb{E}^{(\alpha)}$ is $(1/\alpha)$-convex, the $\left(\frac{1}{\alpha},\lambda''\right)$-mixture of $P_0^{(\alpha)}$ and $P_1^{(\alpha)}$ belongs to $\mathbb{E}^{(\alpha)}$. This implies that

$$\begin{aligned} p^{(\alpha)}_{\frac{1}{\alpha},\lambda''} &\propto \left(\lambda''(p_1^{(\alpha)})^{1/\alpha} + (1-\lambda'')(p_0^{(\alpha)})^{1/\alpha}\right)^{\alpha} \\ &\propto \left(\lambda'' \frac{p_1}{||p_1||} + (1-\lambda'')\frac{p_0}{||p_0||}\right)^{\alpha} \\ &\propto (\lambda p_1 + (1-\lambda)p_0)^{\alpha} \end{aligned} \tag{42}$$

belongs to $\mathcal{E}^{(\alpha)}$, where (42) follows by plugging in (41) for λ''. Since (42) implies $p^{(\alpha)}_{\frac{1}{\alpha},\lambda''} \longleftrightarrow (\lambda p_1 + (1-\lambda)p_0)$, we conclude that $(\lambda p_1 + (1-\lambda)p_0) \in \mathcal{E}$, which implies, by (8), that $\lambda P_1 + (1-\lambda)P_0 \in \mathbb{E}$, hence completing the proof of the "if" part.

B. Proof of Proposition 2

The arguments we present here are already in [3], but not in an isolated form. We bring them out here to establish the centrality of the correspondence.

We prove the forward and backward directions in order.

(i) $\Longrightarrow$: Let $\mathcal{E}$ be closed in $L^{\alpha}(\mu)$. Let $p^{(\alpha)}$ be any limit point of $\mathcal{E}^{(\alpha)}$. If $p^{(\alpha)} \in \mathcal{E}^{(\alpha)}$, then there is nothing to prove. So, suppose that $p^{(\alpha)} \notin \mathcal{E}^{(\alpha)}$. Then, there exists a sequence $\{p_n^{(\alpha)}\} \subset \mathcal{E}^{(\alpha)}$ such that $p_n^{(\alpha)} \to p^{(\alpha)}$ in $L^1(\mu)$, i.e.,

$$\lim_{n\to\infty} \int |p_n^{(\alpha)} - p^{(\alpha)}| d\mu = 0. \tag{43}$$

It follows that $\int p_n^{(\alpha)} d\mu \to \int p^{(\alpha)} d\mu$, and since $\int p_n^{(\alpha)} d\mu = 1$ for all n, we must have $\int p^{(\alpha)} d\mu = 1$.
From the $L^1(\mu)$ convergence in (43), it follows that $p_n^{(\alpha)} \to p^{(\alpha)}$ in $[\mu]$-measure. We now demonstrate that the μ-density proportional to $(p^{(\alpha)})^{1/\alpha}$ is in $\mathcal{E}$, thereby establishing that $p^{(\alpha)} \in \mathcal{E}^{(\alpha)}$ and hence the fact that $\mathcal{E}^{(\alpha)}$ is closed.
In view of the convergence in $[\mu]$-measure and the upper bound

$$|(p_n^{(\alpha)})^{1/\alpha} - (p^{(\alpha)})^{1/\alpha}|^{\alpha} \le 2^{\alpha}(p_n^{(\alpha)} + p^{(\alpha)}), \tag{44}$$

we can apply the generalized version of the dominated convergence theorem (see [11], Ch. 2, Ex. 20) to get

$$\frac{p_n}{||p_n||} = (p_n^{(\alpha)})^{1/\alpha} \longrightarrow (p^{(\alpha)})^{1/\alpha} \text{ in } L^{\alpha}(\mu). \tag{45}$$

We now claim that

$$||p_n|| \text{ is bounded.} \tag{46}$$

Suppose not; then working on a subsequence if needed, we have $||p_n|| := M_n \to \infty$. As $\int p_n d\mu = 1$, given any $\epsilon > 0$,

$$\begin{aligned} \mu\left(p_n^{(\alpha)} > \epsilon\right) &= \mu\left(p_n > \epsilon^{1/\alpha} M_n\right) \\ &\le \frac{1}{\epsilon^{1/\alpha} M_n} \longrightarrow 0 \text{ as } n \to \infty, \end{aligned} \tag{47}$$

and hence $p_n^{(\alpha)} \to 0$ in $[\mu]$-measure, or $p^{(\alpha)} = 0$ except on a set of $[\mu]$-measure 0 (i.e., $p^{(\alpha)} = 0$ a.e.$[\mu]$). But this is a contradiction since $\int p^{(\alpha)} d\mu = 1$. Hence, (46) holds, and we can pick a subsequence of the sequence $||p_n||$ that

converges to some $c > 0$. Reindex and work on this subsequence to get $p_n \to c(p^{(\alpha)})^{1/\alpha}$ in $L^\alpha(\mu)$. The closedness of $\mathcal{E}$ implies that the limiting function $c(p^{(\alpha)})^{1/\alpha} = q$ for some $q \in \mathcal{E}$. Since we also have $\int p^{(\alpha)} d\mu = 1$, it follows that $c = ||q||$ and $p^{(\alpha)} = (q/||q||)^\alpha$. Thus, we have $p^{(\alpha)} \longleftrightarrow q$, which implies that $p^{(\alpha)} \in \mathcal{E}^{(\alpha)}$. This completes the proof of one direction.

(ii) $\Longleftarrow$: Suppose that $\mathcal{E}^{(\alpha)}$ is closed in $L^1(\mu)$. Let p be any arbitrary limit point of $\mathcal{E}$. Following the arguments as before, if $p \in \mathcal{E}$, then there is nothing to prove. So, suppose that $p \notin \mathcal{E}$. Then, there exists a sequence $\{p_n\} \subset \mathcal{E}$ such that $p_n \to p$ in $L^\alpha(\mu)$, i.e.,

$$\lim_{n\to\infty} \int |p_n - p|^\alpha d\mu = 0. \tag{48}$$

This also implies that $||p_n|| \to ||p|| > 0$, and since $|p_n^\alpha - p^\alpha| \le p_n^\alpha + p^\alpha$, the generalized version of the dominated convergence theorem ([11], Ch. 2, Ex. 20) yields

$$p_n^{(\alpha)} = (p_n/||p_n||)^\alpha \longrightarrow (p/||p||)^\alpha \text{ in } L^1(\mu). \tag{49}$$

The closedness of $\mathcal{E}^{(\alpha)}$ in $L^1(\mu)$ implies that the limiting function $(p/||p||)^\alpha = p^{(\alpha)}$ for some $p^{(\alpha)} \in \mathcal{E}^{(\alpha)}$. This implies that $p \longleftrightarrow p^{(\alpha)}$, and thus the fact that $p \in \mathcal{E}$, thereby demonstrating that $\mathcal{E}$ is closed in $L^\alpha(\mu)$.

C. Proof of Theorem 2

Suppose that P_θ. $\theta \in \Theta$, is a member of the α-power-law family generated by Q. According to (16), for any $x \in \mathbb{X}$, we have

$$P_\theta(x) \propto \left((Q(x))^{\alpha-1} + (1-\alpha) \sum_{i=1}^{k} \theta_i f_i(x) \right)^{-\frac{1}{1-\alpha}}. \tag{50}$$

From this, we get

$$\begin{aligned} P_\theta^{(\alpha)}(x) &\propto (P_\theta(x))^\alpha \\ &\propto \left((Q(x))^{\alpha-1} + (1-\alpha) \sum_{i=1}^{k} \theta_i f_i(x) \right)^{-\frac{\alpha}{1-\alpha}} && (51) \\ &\propto \left(\left(\frac{Q(x)}{||Q||} \right)^{\alpha-1} + (1-\alpha) \sum_{i=1}^{k} \frac{\theta_i}{||Q||^{\alpha-1}} f_i(x) \right)^{-\frac{\alpha}{1-\alpha}} && (52) \\ &\propto \left((Q^{(\alpha)}(x))^{1-\frac{1}{\alpha}} + \left(1 - \frac{1}{\alpha}\right) \sum_{i=1}^{k} \theta_i' f_i(x) \right)^{\frac{1}{1-\frac{1}{\alpha}}}, && (53) \end{aligned}$$

where (52) follows by multiplying the scale factor $||Q||^\alpha$ and (53) follows by setting $\theta_i' := \frac{(-\alpha)\theta_i}{||Q||^{\alpha-1}}$, $1 \le i \le k$. We recognize that (53) is of the form (18), with α replaced by $1/\alpha$, $Q(x)$ replaced by $Q^{(\alpha)}(x)$, and θ_i replaced by θ_i'. This completes the proof.

REFERENCES

[1] R. Bhatia, *Notes on Functional Analysis*. Hindustan Book Agency, 2009.

[2] I. Csiszár, "I-divergence Geometry of Probability Distributions and Minimization Problems," *Annals of Probability*, vol. 3, pp. 146–158, 1975.

[3] 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.

[4] 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.

[5] R. Sundaresan, "A Measure of Discrimination and its Geometric Properties," in *Information Theory, 2002. Proceedings. 2002 IEEE International Symposium on*. IEEE, 2002, p. 264.

[6] ——, "Guessing Under Source Uncertainty," *IEEE Trans. on Information Theory*, vol. 53, no. 1, pp. 269–287, January 2007.

[7] 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.

[8] S. Eguchi, "Geometry of Minimum Contrast," *Hiroshima Math. J*, vol. 22, no. 3, pp. 631–647, 1992.

[9] S. I. Amari and H. Nagaoka, *Methods of Information Geometry*. American Mathematical Soc., 2007, vol. 191.

[10] S. I. Amari and A. Ohara, "Geometry of q-Exponential Family of Probability Distributions," *Entropy*, vol. 13, no. 6, pp. 1170–1185, 2011.

[11] G. B. Folland, *Real Analysis: Modern Techniques and Their Applications*, 2nd ed. New York, NY, USA: Wiley, 1999.