On Projections of the Rényi Divergence on Generalized Convex Sets

M. Ashok Kumar
Department of Mathematics
Indian Institute of Technology
Indore 452020, India
Email: ashokm.shree@gmail.com

Igal Sason
Andrew and Erna Viterbi Faculty of Electrical Engineering
Technion-Israel Institute of Technology
Haifa 32000, Israel
Email: sason@ee.technion.ac.il

***Abstract*—Motivated by a recent result by van Erven and Harremoës, we study a forward projection problem for the Rényi divergence on a particular α-convex set, termed α-*linear family*. The solution to this problem yields a parametric family of probability measures which turns out to be an extension of the exponential family, and it is termed α-*exponential family*. An orthogonality relationship between the α-exponential and α-linear families is first established and is then used to transform the reverse projection on an α-exponential family into a forward projection on an α-linear family. The full paper version of this work is available on the arXiv at http://arxiv.org/abs/1512.02515.**[1]

Index Terms – α-convex set, relative entropy, variational distance, forward and reverse projections, Rényi divergence, exponential and linear families.

I. Introduction

Given a probability measure Q, and a set of probability measures $\mathcal{P}$ on an alphabet $\mathcal{A}$, a *forward projection* of Q on $\mathcal{P}$ is a $P^* \in \mathcal{P}$ which minimizes the relative entropy $D(P\|Q)$ subject to $P \in \mathcal{P}$. Forward projections appear predominantly in large deviations theory (see, e.g., [4, Chapter 11] in the context of Sanov's theorem and the conditional limit theorem). The forward projection of a generalization of the relative entropy on convex sets has been proposed by Sundaresan in [17] in the context of guessing under source uncertainty, and it was further studied in [12]. In this paper we consider forward projection of Rényi divergence on some generalized convex sets. A physical motivation for such a study stems from a maximum entropy problem proposed by Tsallis in statistical physics [18], [19] (for further details about the connection of this maximum entropy problem to forward projections of the Rényi divergence, the reader is referred to [11]).

The other problem of interest in this paper is the *reverse projection* where the minimization is over the second argument of the divergence measure. This problem is intimately related to the maximum-likelihood estimation and robust statistics. Suppose $X_1, \ldots, X_n$ are i.i.d. samples drawn according to a probability measure which is modeled by a parametric family of probability measures $\Pi = \{P_\theta \colon \theta \in \Theta\}$ where Θ is a parameter space, and all the members of Π are assumed to have a common finite support $\mathcal{A}$. The maximum-likelihood estimator of the given samples (if it exists) is the minimizer of $D(\hat{P}\|P_\theta)$ subject to $P_\theta \in \Pi$, where $\hat{P}$ is the empirical probability measure of the observed samples [9, Lemma 3.1]. The minimizing probability measure (if exists) is called the reverse projection of $\hat{P}$ on Π. Other divergences that have natural connection in statistical estimation problems include Hellinger divergence of order $\frac{1}{2}$ (see, e.g., [3]), Pearson's χ^2-divergence, and so on. All these belong to the more general family of Hellinger divergences of order $\alpha \in (0, \infty)$ (note that these divergences are, up to a positive scaling factor, equal to the power divergences introduced by Cressie and Read [5]); these divergences form a sub-family of f-divergences which were independently introduced by Ali and Silvey [1] and Csiszár [6]. The Hellinger divergences possess a very good robustness property when a significant fraction of the observed samples are outliers; the textbooks of Basu et al. [2] and Pardo [14] provide a coverage of the developments on the study of inference based on f-divergences. Since the Rényi divergence is a monotonically increasing function of the Hellinger divergence (see, e.g., [16, (1)]), minimizing a Hellinger divergence of order $\alpha \in (0, \infty)$ is equivalent to minimizing the Rényi divergence of the same order. This motivates the problem of studying reverse projections of the Rényi divergence in the context of robust statistics.

In a recent work [10, Theorem 14], van Erven and Harremoës proved a Pythagorean property for Rényi divergences of order $\alpha \in (0, \infty)$ on α-convex sets. By exploiting this property, we study forward projection of the Rényi divergence on an α-linear family. The form of forward projection suggests a parametric family of probability measures which turns out to be an extension of the exponential family, and it is termed an α-*exponential family*. We show an orthogonality relationship between the α-linear family and the α-exponential family. Using this orthogonality property one can transform a reverse projection problem on an α-exponential family into a forward projection problem on an α-convex family.

The following is an outline of the paper: Section II provides preliminary material; Section III identifies the form

[1]This work was done while the first author was a post-doctoral fellow at the Andrew and Erna Viterbi faculty of electrical engineering, Technion–Israel Institute of Technology, Haifa 32000, Israel.

of forward projections on α-linear families, and it proves convergence of an iterated process for forward projections. Section IV proves an orthogonality relationship between the α-linear and α-exponential families is established, and this latter property is used to find reverse projections on α-exponential families. The full paper version of this work is available at [11] which includes additional results and discussions, and all the proofs.

II. Preliminaries

Unless explicitly mentioned, it is assumed throughout the paper that probability measures are defined on a finite alphabet $\mathcal{A}$. Let $\mathcal{M}$ denote the set of all probability measures on $\mathcal{A}$. For $P \in \mathcal{M}$, let $\text{Supp}(P) := \{a \in \mathcal{A} \colon P(a) > 0\}$; for $\mathcal{P} \subseteq \mathcal{M}$, let $\text{Supp}(\mathcal{P})$ be the union of support of members of $\mathcal{P}$.

Definition 1 (Rényi divergence)**.** Let $\alpha \in (0,1) \cup (1,\infty)$. For $P, Q \in \mathcal{M}$, the *Rényi divergence* [15] of order α from P to Q is given by

$$D_\alpha(P\|Q) := \frac{1}{\alpha-1}\, \log\biggl(\sum_a P(a)^\alpha Q(a)^{1-\alpha}\biggr). \quad (1)$$

If $\alpha = 1$, then

$$D_1(P\|Q) := D(P\|Q), \quad (2)$$

which is the analytic extension of $D_\alpha(P\|Q)$ at $\alpha = 1$.

Definition 2 ((α,λ)-mixture [10])**.** Let $P_0, P_1 \in \mathcal{M}$, $\alpha \in (-\infty, 0) \cup (0,\infty)$, and let $\lambda \in (0,1)$. The (α,λ)*-mixture* of (P_0, P_1) is the probability measure $S_{0,1}$ defined by

$$S_{0,1}(a) := \frac{1}{Z}\Bigl[(1-\lambda)P_0(a)^\alpha + \lambda P_1(a)^\alpha\Bigr]^{\frac{1}{\alpha}}, \quad (3)$$

where Z is the normalizing constant in (3) such that

$$\sum_a S_{0,1}(a) = 1. \quad (4)$$

Here, for simplicity, we suppress the dependence of $S_{0,1}$ and Z on α, λ. Note that $S_{0,1}$ is well-defined as Z is always positive and finite.

Definition 3 (α-convex set)**.** Let $\alpha \in (-\infty,0)\cup(0,\infty)$. A set of probability measures $\mathcal{P}$ is said to be α*-convex* if, for every $P_0, P_1 \in \mathcal{P}$ and $\lambda \in (0,1)$, the (α,λ)-mixture $S_{0,1} \in \mathcal{P}$.

The following is a specific α-convex set which is of interest in this paper.

Definition 4 (α-linear family)**.** Let $\alpha \in (-\infty, 0) \cup (0,\infty)$, and $f_1, \ldots, f_k$ be real-valued functions defined on $\mathcal{A}$. The α*-linear family* determined by $f_1, \ldots, f_k$ is defined to be the following parametric family of probability measures defined on $\mathcal{A}$:

$$\mathscr{L}_\alpha := \biggl\{P \in \mathcal{M} \colon P(a) = \Bigl[\sum_{i=1}^{k} \theta_i f_i(a)\Bigr]^{\frac{1}{\alpha}}, \quad \underline{\theta} \in \mathbb{R}^k\biggr\} \quad (5)$$

For typographical convenience, we have suppressed the dependence of $\mathscr{L}_\alpha$ in $f_1, \ldots, f_k$. It is easy to see that $\mathscr{L}_\alpha$ is an α-convex set. Without loss of generality, we shall assume that $f_1, \ldots, f_k$, as $|\mathcal{A}|$-dimensional vectors, are mutually orthogonal (otherwise, by the Gram-Schmidt procedure, these vectors can be orthogonalized without affecting the corresponding α-linear family in (5)). Let $\mathcal{F}$ be the subspace of $\mathbb{R}^{|\mathcal{A}|}$ spanned by $f_1, \ldots, f_k$, and let $\mathcal{F}^\perp$ denote the orthogonal complement of $\mathcal{F}$. Hence, there exist $f_{k+1}, \ldots, f_{|\mathcal{A}|}$ such that $f_1, \ldots, f_{|\mathcal{A}|}$ are mutually orthogonal as $|\mathcal{A}|$-dimensional vectors, and $\mathcal{F}^\perp = \text{Span}\{f_{k+1}, \ldots, f_{|\mathcal{A}|}\}$. Consequently, from (5), for $\alpha \in (0,1) \cup (1,\infty)$,

$$\mathscr{L}_\alpha = \biggl\{P \in \mathcal{M} \colon \sum_a P(a)^\alpha f_i(a) = 0, \ k+1 \le i \le |\mathcal{A}|\biggr\}. \quad (6)$$

From (6), it is clear that the set $\mathscr{L}_\alpha$ is closed. Since it is also bounded, it is compact.

III. Forward Projection on α-linear family

Let us first recall the Pythagorean property for a Rényi divergence on an α-convex set. As in the case of relative entropy [9] and relative α-entropy [13], the Pythagorean property is crucial in establishing orthogonality properties. In the sequel, we assume that Q is a probability measure with $\text{Supp}(Q) = \mathcal{A}$.

In a recent work, van Erven and Harremoës proved a Pythagorean inequality for Rényi divergences on α-convex sets under the assumption that the forward projection exists (see [10, Theorem 14]). Continuing this study, a sufficient condition for the existence of forward projection is proved in [11] for probability measures on a general alphabet. The following result extends the existence result in [8] for the forward projection of the relative entropy.

Theorem 1 (Existence of forward D_α-projection)**.** Let $\alpha \in (0,\infty)$, and let Q be an arbitrary probability measure defined on a set $\mathcal{A}$. Let $\mathcal{P}$ be an α-convex set of probability measures defined on $\mathcal{A}$, and assume that $\mathcal{P}$ is closed with respect to the total variation distance. If there exists $P \in \mathcal{P}$ such that $D_\alpha(P\|Q) < \infty$, then there exists a forward D_α-projection of Q on $\mathcal{P}$.

Proof: See [11]. For $\alpha \in (1,\infty)$, the proof relies on a new Apollonius theorem for the Hellinger divergence, and for $\alpha \in (0,1)$, the proof relies on the Banach-Alaoglu theorem from functional analysis. ∎

Proposition 1 (The Pythagorean property)**.** Let $\alpha \in (0,1) \cup (1,\infty)$, let $\mathcal{P} \subseteq \mathcal{M}$ be an α-convex set, and $Q \in \mathcal{M}$.

(a) If P^* is a forward D_α-projection of Q on $\mathcal{P}$, then

$$D_\alpha(P\|Q) \ge D_\alpha(P\|P^*) + D_\alpha(P^*\|Q) \quad (7)$$

for all $P \in \mathcal{P}$. Furthermore, if $\alpha > 1$, then $\text{Supp}(P^*) = \text{Supp}(\mathcal{P})$.

(b) Conversely, if (7) is satisfied for some $P^* \in \mathcal{P}$, then P^* is the (unique) forward D_α-projection of Q on $\mathcal{P}$.

Proof: See [11]. ∎

Remark 1. The Pythagorean property (7) holds for probability measures defined on a general alphabet $\mathcal{A}$, as proved in [10, Theorem 14]. The novelty here is in the last assertion of (a), which extends the result for relative entropy in [9, Theorem 3.1], for which $\mathcal{A}$ needs to be finite.

Corollary 1. Let $\alpha \in (0, \infty)$. If a forward D_α-projection on an α-convex set exists, then it is unique.

Proof: For $\alpha = 1$, since an α-convex set is particularized to a convex set, this result is known in view of [9, p. 23]. Next, consider the case where $\alpha \neq 1$. Let P_1^* and P_2^* be forward D_α-projections of Q on an α-convex set $\mathcal{P}$. Applying Proposition 1 we have

$$D_\alpha(P_2^* \| Q) \geq D_\alpha(P_2^* \| P_1^*) + D_\alpha(P_1^* \| Q).$$

Since $D_\alpha(P_1^* \| Q) = D_\alpha(P_2^* \| Q)$, we must have $D_\alpha(P_2^* \| P_1^*) = 0$ which implies that $P_1^* = P_2^*$. ■

The last assertion in Proposition 1a) states that $\text{Supp}(P^*) = \text{Supp}(\mathcal{P})$ if $\alpha \in (1, \infty)$. The following counterexample, taken from [11], illustrates that this equality does not necessarily hold for $\alpha \in (0, 1)$.

Example 1. *Let* $\mathcal{A} = \{1, 2, 3, 4\}$, $\alpha = \frac{1}{2}$, $f \colon \mathcal{A} \to \mathbb{R}$ *be given by*

$$f(1) = 1, \; f(2) = -3, \; f(3) = -5, \; f(4) = -6 \tag{8}$$

and let $Q(a) = \frac{1}{4}$ *for all* $a \in \mathcal{A}$. *Consider the following* α*-linear family:*

$$\mathcal{P} := \Bigl\{ P \in \mathcal{M} \colon \sum_a P(a)^\alpha f(a) = 0 \Bigr\}. \tag{9}$$

Let

$$P^*(1) = \tfrac{9}{10}, \; P^*(2) = \tfrac{1}{10}, \; P^*(3) = 0, \; P^*(4) = 0. \tag{10}$$

It is easy to check that $P^* \in \mathcal{P}$. *Furthermore, setting* $\theta^* = \frac{1}{5}$ *and* $Z = \frac{2}{5}$ *yields that for* $a \in \{1, 2, 3\}$

$$P^*(a)^{1-\alpha} = Z^{\alpha-1} \Bigl[Q(a)^{1-\alpha} + (1-\alpha)\, f(a)\, \theta^* \Bigr], \tag{11}$$

and

$$P^*(4)^{1-\alpha} > Z^{\alpha-1} \Bigl[Q(4)^{1-\alpha} + (1-\alpha)\, f(4)\, \theta^* \Bigr]. \tag{12}$$

From (9), (11) *and* (12), *it follows that for every* $P \in \mathcal{P}$

$$\sum_{a \in \mathcal{A}} P(a)^\alpha P^*(a)^{1-\alpha} \geq Z^{\alpha-1} \sum_{a \in \mathcal{A}} P(a)^\alpha Q(a)^{1-\alpha}. \tag{13}$$

Furthermore, it can be also verified that

$$Z^{\alpha-1} \sum_{a \in \mathcal{A}} P^*(a)^\alpha Q(a)^{1-\alpha} = 1. \tag{14}$$

Assembling (13) *and* (14) *yields*

$$\sum_{a \in \mathcal{A}} P(a)^\alpha P^*(a)^{1-\alpha} \geq \frac{\sum_{a \in \mathcal{A}} P(a)^\alpha Q(a)^{1-\alpha}}{\sum_{a \in \mathcal{A}} P^*(a)^\alpha Q(a)^{1-\alpha}}, \tag{15}$$

which is equivalent to (7). *Hence, Proposition 1b) implies that* P^* *is the forward* D_α*-projection of* Q *on* $\mathcal{P}$. *Note, however, that Supp*$(P^*) \neq$ *Supp*$(\mathcal{P})$*; to this end, from* (9), *it can be verified numerically that*

$$P = (0.984688, \, 0.00568298, \, 0.0041797, \, 0.00544902) \in \mathcal{P} \tag{16}$$

which implies that Supp$(P^*) = \{1, 2\} \subset \mathcal{A}$ *whereas Supp*$(\mathcal{P}) = \mathcal{A}$.

We shall now focus our attention on forward D_α-projections on α-linear families.

Theorem 2 (Pythagorean equality)**.** Let $\alpha > 1$, and let P^* be the forward D_α-projection of Q on an α-linear family $\mathscr{L}_\alpha$. Then, P^* satisfies (7) with equality, i.e.,

$$D_\alpha(P \| Q) = D_\alpha(P \| P^*) + D_\alpha(P^* \| Q), \quad \forall \, P \in \mathscr{L}_\alpha. \tag{17}$$

Proof: See [11]. ■

In [8, Theorem 3.2], Csiszár proposed a convergent iterative process for finding the forward projection for the relative entropy on a finite intersection of linear families. This result is generalized in this work for the Rényi divergence of order $\alpha \in (0, \infty)$ on a finite intersection of α-linear families.

Theorem 3 (Iterative projections)**.** Let $\alpha \in (1, \infty)$. Suppose that $\mathscr{L}_\alpha^{(1)}, \ldots, \mathscr{L}_\alpha^{(m)}$ are α-linear families, and let

$$\mathcal{P} := \bigcap_{n=1}^{m} \mathscr{L}_\alpha^{(n)} \tag{18}$$

where $\mathcal{P}$ is assumed to be a non-empty set. Let $P_0 = Q$, and let P_n for $n \in \mathbb{N}$ be the forward D_α-projection of P_{n-1} on $\mathscr{L}_\alpha^{(i_n)}$ with $i_n = n \bmod (m)$. Then, $P_n \to P^*$ (a pointwise convergence) as we let $n \to \infty$.

Proof: See [11]. ■

We next identify the form of the forward D_α-projection on an α-linear family.

Theorem 4 (Forward projection on an α-linear family)**.** Let $\alpha \in (0,1) \cup (1, \infty)$, and let P^* be the forward D_α-projection of Q on an α-linear family $\mathscr{L}_\alpha$. Suppose that

$$\text{Supp}(P^*) = \text{Supp}(\mathscr{L}_\alpha) = \mathcal{A}. \tag{19}$$

Then,

(a) P^* satisfies (17).
(b) There exists $\theta^* = (\theta_{k+1}^*, \ldots, \theta_{|\mathcal{A}|}^*) \in \mathbb{R}^{|\mathcal{A}|-k}$ such that, for all $a \in \mathcal{A}$,

$$P^*(a) = Z(\theta^*)^{-1} \Biggl[Q(a)^{1-\alpha} + (1-\alpha) \sum_{i=k+1}^{|\mathcal{A}|} \theta_i^* f_i(a) \Biggr]^{\frac{1}{1-\alpha}} \tag{20}$$

where $Z(\theta^*)$ is a normalizing constant in (20).

Proof: See [11]. ■

Remark 2. In view of Example 1, the assumption in (19) does not necessarily hold for $\alpha \in (0, 1)$. However, due to

Proposition 1, this assumption necessarily holds for every $\alpha \in (1, \infty)$.

For $\alpha \in (0, \infty)$, the forward D_α-projection on an α-linear family $\mathscr{L}_\alpha$ motivates the definition of the following parametric family of probability measures. Let $Q \in \mathcal{M}$, and let

$$\mathscr{E}_\alpha = \Bigl\{ P \in \mathcal{M} \colon P(a) = \frac{1}{Z(\theta)} \Bigl[Q(a)^{1-\alpha} + (1-\alpha) \sum_{i=k+1}^{|\mathcal{A}|} \theta_i f_i(a) \Bigr]^{\frac{1}{1-\alpha}}, \quad \theta = (\theta_{k+1}, \ldots, \theta_{|\mathcal{A}|}) \in \mathbb{R}^{|\mathcal{A}|-k} \Bigr\}. \tag{21}$$

It is easy to see that $\mathscr{E}_\alpha$ is an $(1-\alpha)$-convex set. Also, the family $\mathscr{E}_\alpha$ generalizes the *exponential family* in [9, p. 24]:

$$\mathscr{E} = \Biggl\{ P \in \mathcal{M} \colon P(a) = Z(\theta)^{-1} Q(a) \exp\Biggl(\sum_{i=k+1}^{|\mathcal{A}|} \theta_i f_i(a) \Biggr), \quad \theta = (\theta_{k+1}, \ldots, \theta_{|\mathcal{A}|}) \in \mathbb{R}^{|\mathcal{A}|-k} \Biggr\}. \tag{22}$$

This extension is demonstrated in [11]. We shall call the family $\mathscr{E}_\alpha$ an *α-exponential family*.[2] It is easy to see that the reference measure Q in the definition of $\mathscr{E}_\alpha$ is always a member of $\mathscr{E}_\alpha$. As in the case of the exponential family, the α-exponential family $\mathscr{E}_\alpha$ also depends on the reference measure Q only in a loose manner. Any other member of $\mathscr{E}_\alpha$ could very well play the role of Q in defining the family. The proof is very similar to the one for the α-power-law family in [13, Proposition 22]. It should also be noted that, for $\alpha \in (1, \infty)$, all members of $\mathscr{E}_\alpha$ have the same support (i.e., same as the support of Q).

IV. Orthogonality of α-Linear and α-Exponential Families

We first make precise the notion of orthogonality between two sets of probability measures with respect to D_α ($\alpha > 0$).

Definition 5 (Orthogonal sets of probability measures)**.** Let $\alpha \in (0,1) \cup (1, \infty)$, and let $\mathcal{P}$ and $\mathcal{Q}$ be sets of probability measures. We say that $\mathcal{P}$ *is α-orthogonal to $\mathcal{Q}$ at P^** if the following hold:

(i) $\mathcal{P} \cap \mathcal{Q} = \{P^*\}$
(ii) $D_\alpha(P \| Q) = D_\alpha(P \| P^*) + D_\alpha(P^* \| Q)$ for every $P \in \mathcal{P}$ and $Q \in \mathcal{Q}$.

Note that, when $\alpha = 1$, this refers to the orthogonality between the linear and exponential families, which is essentially [9, Corollary 3.1].

We are now ready to state the second main result in [11] namely, the orthogonality between $\mathscr{L}_\alpha$ and $\mathscr{E}_\alpha$.

[2] We emphasize that the *α-power-law family* proposed in [13, Definition 8] is different extension of the exponential family $\mathscr{E}$.

Theorem 5 (Orthogonality of $\mathscr{L}_\alpha$ and $\mathscr{E}_\alpha$)**.** Let $\alpha \in (1, \infty)$, let $\mathscr{L}_\alpha$ and $\mathscr{E}_\alpha$ be given by (5) and (21), respectively, and let P^* be the forward D_α-projection of Q on $\mathscr{L}_\alpha$. Then,

(a) $\mathscr{L}_\alpha$ is α-orthogonal to $\mathrm{cl}(\mathscr{E}_\alpha)$ at P^*.
(b) If $\mathrm{Supp}(\mathscr{L}_\alpha) = \mathcal{A}$, then $\mathscr{L}_\alpha$ is α-orthogonal to $\mathscr{E}_\alpha$ at P^*.

Proof: See [11]. ∎

Remark 3. In view of Example 1, if $\alpha \in (0,1)$, then $\mathrm{Supp}(P^*)$ is not necessarily equal to $\mathrm{Supp}(\mathscr{L}_\alpha)$; this is consistent with Theorem 5 which is stated only for $\alpha \in (1, \infty)$. Nevertheless, in view of the proof of Theorem 2, the following holds for $\alpha \in (0,1)$: if $\mathrm{Supp}(P^*) = \mathrm{Supp}(\mathscr{L}_\alpha) = \mathcal{A}$, then $\mathscr{L}_\alpha$ is α-orthogonal to $\mathscr{E}_\alpha$ at P^*.

In [11], Theorem 5 and Remark 3 are applied to find a reverse projection on an α-exponential family. Before we proceed, we now make precise the definition of a reverse D_α-projection.

Definition 6 (Reverse D_α-projection)**.** *Let $P \in \mathcal{M}$, $\mathcal{Q} \subseteq \mathcal{M}$, and $\alpha > 0$. If there exists $Q^* \in \mathcal{Q}$ which attains the global minimum of $D_\alpha(P \| Q)$ over all $Q \in \mathcal{Q}$ and $D_\alpha(P \| Q^*) < \infty$, then Q^* is said to be a* reverse D_α-projection *of P on $\mathcal{Q}$.*

Theorem 6. Let $\alpha \in (0,1) \cup (1, \infty)$, and let $\mathscr{E}_\alpha$ be an α-exponential family determined by $Q, f_{k+1}, \ldots, f_{|\mathcal{A}|}$. Let $X_1, \ldots, X_n$ be i.i.d. samples drawn at random according to a probability measure in $\mathscr{E}_\alpha$. Let $\hat{P}_n$ be the empirical probability measure of $X_1, \ldots, X_n$, and let P_n^* be the forward D_α-projection of Q on the α-linear family

$$\mathscr{L}_\alpha^{(n)} := \Bigl\{ P \in \mathcal{M} \colon \sum_a P(a) \hat{f}_i(a) = 0, \; k+1 \le i \le |\mathcal{A}| \Bigr\}, \tag{23}$$

where

$$\hat{f}_i(a) := f_i(a) - \hat{\eta}_i^{(n)} Q(a)^{1-\alpha}, \quad \forall\, a \in \mathcal{A} \tag{24}$$

with

$$\hat{\eta}_i^{(n)} := \frac{\sum_a \hat{P}_n(a) f_i(a)}{\sum_a \hat{P}_n(a)^\alpha Q(a)^{1-\alpha}}, \quad i \in \{k+1, \ldots, |\mathcal{A}|\}. \tag{25}$$

Then, the following hold:

(a) If $\mathrm{Supp}(P_n^*) = \mathrm{Supp}(\mathscr{L}_\alpha^{(n)}) = \mathcal{A}$, then P_n^* is the reverse D_α-projection of $\hat{P}_n$ on $\mathscr{E}_\alpha$.
(b) For $\alpha \in (1, \infty)$, if $\mathrm{Supp}(\mathscr{L}_\alpha^{(n)}) \neq \mathcal{A}$, then the reverse D_α-projection of $\hat{P}_n$ on $\mathscr{E}_\alpha$ does not exist. Nevertheless, P_n^* is the reverse D_α-projection of $\hat{P}_n$ on $\mathrm{cl}(\mathscr{E}_\alpha)$.

Proof: See [11]. ∎

Acknowledgment

This work has been supported by the Israeli Science Foundation (ISF) under Grant 12/12.

References

[1] S. M. Ali and S. D. Silvey, "A general class of coefficients of divergence of one distribution from another," *Journal of the Royal Statistics Society*, series B, vol. 28, no. 1, pp. 131–142, 1966.

[2] A. Basu, H. Shioya and C. Park, "Statistical Inference: The Minimum Distance Approach," *Chapman & Hall/ CRC Monographs on Statistics and Applied Probability*, vol. 120, CRC Press, Boca Raton, Florida, USA, June 2011.

[3] R. Beran, "Minimum Hellinger distance estimates for parametric models," *Annals of Statistics*, vol. 5, no. 3, pp. 445–463, 1977.

[4] T. M. Cover and J. A. Thomas, *Elements of Information Theory*, John Wiley and Sons, second edition, 2006.

[5] N. Cressie and T. R. C. Read, "Multinomial Goodness-of-fit Tests," *Journal of the Royal Statistical Society B*, no. 46, pp. 440–464, 1984.

[6] I. Csiszár, "Eine Informationstheoretische Ungleichung und ihre Anwendung auf den Bewis der Ergodizität von Markhoffschen Ketten," *Publ. Math. Inst. Hungar. Acad. Sci.*, vol. 8, pp. 85–108, January 1963.

[7] I. Csiszár, "Information-type measures of difference of probability distributions and indirect observations," *Studia Scientiarum Mathematicarum Hungarica*, vol. 2, pp. 299–318, 1967.

[8] I. Csiszár, "I-divergence geometry of probability distributions and minimization problems," *Annals of Probability*, vol. 3, pp. 146–158, 1975.

[9] I. Csiszár and P. C. Shields, "Information Theory and Statistics: A Tutorial," *Foundations and Trends in Communications and Information Theory*, vol. 1, no. 4, pp. 417–528, 2004.

[10] T. van Erven and P. Harremoës, "Rényi divergence and Kullback-Leibler divergence," *IEEE Trans. on Information Theory*, vol. 60, no. 7, pp. 3797–3820, July 2014.

[11] M. A. Kumar and I. Sason, "Projection theorems for the Rényi divergence on α-convex sets," preprint, http://arxiv.org/abs/1512.02515.

[12] M. A. Kumar and R. Sundaresan, "Minimization problems based on α-relative entropy I: Forward Projection," *IEEE Trans. on Information Theory*, vol. 61, no. 9, pp. 5063–5080, September 2015.

[13] M. A. Kumar and R. Sundaresan, "Minimization problems based on α-relative entropy II: Reverse Projection," *IEEE Trans. on Information Theory*, vol. 61, no. 9, pp. 5081–5095, September 2015.

[14] L. Pardo, "Statistical Inference based on Divergences," *CRC/Chapman–Hall*, 2006.

[15] A. Rényi, "On measures of entropy and information," *Proceedings of the 4th Berekely Symposium on Probability Theory and Mathematical Statistics*, pp. 547–561, Berekeley, California, USA, 1961.

[16] I. Sason, "On the Rényi divergence, joint range of relative entropies, and a channel coding theorem," *IEEE Trans. on Information Theory*, vol. 62, no. 1, pp. 23–34, January 2016.

[17] R. Sundaresan, "A measure of discrimination and its geometric properties," *Proccedings of the IEEE 2002 International Symposium on Information Theory*, p. 264, Laussane, Switzerland, June 2002.

[18] C. Tsallis, "Possible generalization of Bolzmann-Gibbs statistics," *Journal of Statistical Physics*, vol. 52, no. 1-2, 1988.

[19] C. Tsallis, R. S. Mendes and A. R. Plastino, "The role of constraints within generalized nonextensive statistics," *Physica A*, vol. 261, pp. 534–554, 1998.