On the Rényi Divergence and the Joint Range of Relative Entropies

Igal Sason
Department of Electrical Engineering
Technion, Haifa 32000, Israel
E-mail: sason@ee.technion.ac.il

***Abstract*—This paper starts with a study of the minimum of the Rényi divergence, of an arbitrary order $\alpha > 0$, subject to a fixed (or minimal) value of the total variation distance. Relying on the solution of this minimization problem, we determine the exact region of the points $\big(D(Q||P_1), D(Q||P_2)\big)$ where P_1 and P_2 are any probability distributions whose total variation distance is not below a fixed value, and the probability distribution Q is arbitrary (none of these three distributions is assumed to be fixed). It is further shown that all the points of this convex region are attained by a triple of 2-element probability distributions. As a byproduct of this characterization, we provide a geometric interpretation of the minimal Chernoff information subject to a minimal total variation distance. A full paper version, which includes more results and proofs, is available at http://arxiv.org/abs/1501.03616.**

1. Introduction

Rényi measures play an important role in various studies in information theory and statistical inference (see, e.g., [1], [3], [5], [6], [7], [8], [9], [16], [17], [18]).

This work starts with a study of the minimum of the Rényi divergence $D_\alpha(P\|Q)$, of an arbitrary order $\alpha > 0$, subject to a fixed (or minimal) value of the total variation distance between the probability distributions (PDs) P and Q. The solution of this minimization problem is obtained by adapting some arguments from [10] which considered the minimization of the relative entropy (a.k.a. Kullback-Leibler divergence) subject to a fixed value of the total variation distance. For orders $\alpha \in (0, 1)$, our analysis further relies on the Lagrange duality and a solution of the KKT equations where strong duality is first asserted for the studied problem. The use of Lagrange duality significantly simplifies the computational task of the studied minimization problem for $\alpha \in (0, 1)$, whose solution for this sub-interval of α is of special interest in the continuation of this work. The exact expression for the minimal Rényi divergence generalizes previous studies of the minimization of the relative entropy under the same constraint on the total variation distance (see [10], [12], [14]). The exact expression for this minimum is also compared with known Pinsker-type lower bounds on the Rényi divergence [13].

Relying on the solution of this minimization problem, this paper provides an exact characterization of the achievable region of $\big(D(Q||P_1),\, D(Q||P_2)\big)$ when the PDs P_1 and P_2 are any pair of PDs with a total variation distance of at least $\varepsilon \in (0, 1)$, and Q is any PD which is absolutely continuous w.r.t. P_1 and P_2. This problem is motivated by the significance of the relative entropy in a variety of fundamental problems in information theory and statistics. These include, e.g., the characterization of the gap of the lossless compression rate to the entropy of the source when there exists a mismatch between the assumed distribution of the compressor and the true distribution of the source; the relative entropy is also fundamental in the characterization of the best achievable error exponent for a Bayesian probability of error, being strongly related to the important information-theoretic measure of the Chernoff information.

The exact characterization of the considered achievable region also provides a geometric interpretation of the minimal Chernoff information subject to a minimal total variation distance. Every point in this region is shown to be achievable by a triple of 2-element PDs P_1, P_2 and Q, and their exact calculation is specified exactly by relying on the previous solved problem of the minimum of the Rényi divergence subject to a minimal total variation distance. The task of the numerical computation of this region is demonstrated to be very easy. Note that the considered problem is different from the characterization of joint ranges of points of f-divergences, which was studied in [11].

This paper is structured as follows: Section 2 solves the minimization problem for the Rényi divergence under a fixed total variation distance, and Section 3 provides an exact characterization of $\big(D(Q||P_1),\, D(Q||P_2)\big)$ under a constraint on the minimal total variation distance between P_1 and P_2. A full paper version, which includes more results and proofs, is available in [16].

We end this section by shortly introducing the definitions of the total variation distance and the Rényi divergence, to set the notation used in this work.

Definition 1 (Total variation distance)**.** Let P and Q be two PDs defined on a measurable space $(\mathcal{X}, \mathcal{F})$. The total variation distance between P and Q is defined by

$$d_{\text{TV}}(P, Q) \triangleq \sup_{A \in \mathcal{F}} |P(A) - Q(A)| \tag{1}$$

which can be simplified in the discrete setting, where $\mathcal{X}$ is a countable set, to

$$d_{\text{TV}}(P, Q) = \frac{1}{2} \sum_x \big|P(x) - Q(x)\big| = \frac{||P - Q||_1}{2} \tag{2}$$

so, the total variation distance is equal to one-half the l_1-distance between P and Q. In the continuous setting, PDs are replaced by probability density functions, and the sum in (2) is replaced by an integral.

Definition 2 (Rényi divergence)**.** Let $\alpha \in [0, \infty) \setminus \{1\}$, and let P and Q be two discrete PDs. The Rényi divergence of order α of P from Q, which are both defined on a countable set $\mathcal{X}$, is

$$D_\alpha(P||Q) = \frac{1}{\alpha - 1} \log \sum_{x \in \mathcal{X}} P^\alpha(x)\, Q^{1-\alpha}(x) \tag{3}$$

with the convention that if $\alpha > 1$ and $Q(x) = 0$ then $P^\alpha(x)\, Q^{1-\alpha}(x)$ equals 0 or ∞ if $P(x) = 0$ or $P(x) > 0$, respectively. For $\alpha = 1$, the Rényi divergence is defined to be the relative entropy $D(P||Q) = \sum_{x \in \mathcal{X}} P(x) \log\left(\frac{P(x)}{Q(x)}\right)$.

If $D(P||Q) < \infty$, it can be verified by the use of L'Hôpital's rule that $D(P||Q) = \lim_{\alpha \to 1^-} D_\alpha(P||Q)$. Properties of the Rényi divergence are provided in [8].

2. The Minimum of the Rényi Divergence Subject to a Fixed Total Variation Distance

The task of minimizing an arbitrary *symmetric* f-divergence for a fixed total variation distance has been studied in [12], leading to a closed-form solution of this optimization problem. Although the Rényi divergence is not an f-divergence, it is a function of an f-divergence; however, this f-divergence is *asymmetric*, except for the case where $\alpha = \frac{1}{2}$, so the closed-form expression in [12] cannot be utilized to obtain a tight lower bound on the Rényi divergence subject to a fixed total variation distance.

In this section, we derive a tight lower bound on the Rényi divergence $D_\alpha(P_1||P_2)$ subject to a fixed total variation distance between P_1 and P_2. We further show that this lower bound is attained by a pair of 2-element PDs P_1 and P_2, and both distributions are obtained for a given order $\alpha \in (0, \infty)$ and a total variation distance $d_{\text{TV}}(P_1, P_2) = \varepsilon \in [0, 1)$ (note that if $\varepsilon = 1$ then $\text{Supp}(P_1) \cap \text{Supp}(P_2) = \emptyset$, and consequently $D_\alpha(P_1||P_2) = \infty$). For orders $\alpha \in (0, 1)$, the new tight lower bound is compared with existing Pinsker-type lower bounds on the Rényi divergence [13]. The special case where $\alpha = 1$, which is particularized to the minimization of the Kullback-Leibler divergence subject to a fixed total variation distance, has been studied extensively, and three equivalent forms of the solution to this optimization problem have been derived in [10], [12] and [14].

In [13, Corollaries 6 and 9], Gilardoni derived Pinsker-type lower bounds on the Rényi divergence of order $\alpha \in (0, 1)$ in terms of the total variation distance. Among these two bounds, the improved lower bound is

$$D_\alpha(P||Q) \geq 2\alpha\varepsilon^2 + \frac{4}{9}\,\alpha(1 + 5\alpha - 5\alpha^2)\varepsilon^4, \quad \forall\, \alpha \in (0, 1) \tag{4}$$

where $\varepsilon \triangleq d_{\text{TV}}(P, Q)$ denotes the total variation distance between P and Q. Note that in the limit where $\varepsilon \to 1$, this lower bound converges to a finite value that is at most $\frac{22}{9}$. This, however, is an artifact of the lower bound in light of the following simple observation:

Lemma 1.

$$\lim_{\varepsilon \to 1^-} \inf_{P,Q:\, d_{\text{TV}}(P,Q) = \varepsilon} D_\alpha(P||Q) = \infty, \quad \forall\, \alpha > 0. \tag{5}$$

Proof. See [16, Appendix I.A]. □

Lemma 1 motivates a study of the exact characterization of the infimum (or minimum) of the Rényi divergence for a fixed total variation distance. In the following, we derive a tight lower bound which is shown to be achievable by pairs of 2-element PDs for any fixed value $\varepsilon \in [0, 1)$ of the total variation distance.

For $\alpha > 0$, let

$$g_\alpha(\varepsilon) \triangleq \inf_{P_1, P_2:\, d_{\text{TV}}(P_1, P_2) = \varepsilon} D_\alpha(P_1||P_2), \quad \forall\, \varepsilon \in [0, 1). \tag{6}$$

Since $g_\alpha(\varepsilon)$ is monotonic non-decreasing in $\varepsilon \in [0, 1)$, it can be expressed for $\alpha \in (0, \infty)$ as

$$g_\alpha(\varepsilon) = \inf_{P_1, P_2:\, d_{\text{TV}}(P_1, P_2) \geq \varepsilon} D_\alpha(P_1||P_2), \quad \forall\, \varepsilon \in [0, 1).$$

Remark 1. For $\alpha \in [0, 1]$, since $D_\alpha(P||Q)$ is jointly convex in (P, Q), the same arguments in [10] yield that g_α is a convex function, and the infimum in (6) is a minimum.

In the following, we provide an expression for the function g_α in (6). Following [10, Section 2] that characterizes the minimum of the Kullback-Leibler divergence in terms of the total variation distance, we first extend their argument to obtain the following lemma:

Lemma 2. For an arbitrary $\alpha > 0$, there is no loss of generality by restricting the minimization of the Rényi divergence of order α, subject to a fixed total variation distance, to pairs of 2-element PDs.

Proof. See [16, Appendix I.B]. It relies on the data processing inequality for the Rényi divergence of any order $\alpha > 0$ (see [8, Theorem 9]). □

The following proposition provides an expression for g_α for an arbitrary positive α.

Proposition 1. Let $\alpha > 0$ and $\varepsilon \in [0, 1)$. The function g_α in (6) satisfies

$$g_\alpha(\varepsilon) = \min_{p, q \in [0,1]:\, |p-q| \geq \varepsilon} d_\alpha(p\|q) \tag{7}$$

where

$$d_\alpha(p\|q) \triangleq \frac{\log\Big(p^\alpha q^{1-\alpha} + (1-p)^\alpha (1-q)^{1-\alpha}\Big)}{\alpha - 1} \tag{8}$$

is the Rényi divergence $D_\alpha(P||Q)$ from the 2-element PD $P = (p, 1-p)$ to $Q = (q, 1-q)$.

Proof. Eq. (7) follows from Lemma 2 where $D_\alpha(P_1||P_2)$ is minimized over all pairs of 2-element PDs $P_1 = (p, 1-p)$, $P_2 = (q, 1-q)$ with $|p - q| = d_{\text{TV}}(P_1, P_2) \geq \varepsilon$. □

Corollary 1. For $\alpha = \frac{1}{2}$ and $\alpha = 2$, g_α admits the following closed-form expressions:

$$g_{\frac{1}{2}}(\varepsilon) = -\log(1-\varepsilon^2), \tag{9}$$

and

$$g_2(\varepsilon) = \begin{cases} \log(1+4\varepsilon^2), & \text{if } \varepsilon \in \left[0, \frac{1}{2}\right], \\ \log\left(\frac{1}{1-\varepsilon}\right), & \text{if } \varepsilon \in \left(\frac{1}{2}, 1\right). \end{cases} \tag{10}$$

Furthermore, for $\alpha \in (0,1)$ and $\varepsilon \in [0,1)$, we have

$$g_\alpha(\varepsilon) = \left(\frac{\alpha}{1-\alpha}\right) g_{1-\alpha}(\varepsilon), \tag{11}$$

and

$$g_\alpha(\varepsilon) \geq c_1(\alpha) \log\left(\frac{1}{1-\varepsilon}\right) + c_2(\alpha), \tag{12}$$

where $c_1(\alpha) \triangleq \min\left\{1, \frac{\alpha}{1-\alpha}\right\}$, and $c_2(\alpha) \triangleq -\frac{\log(2)}{1-\alpha}$.

Proof. See [16, Appendix II]. □

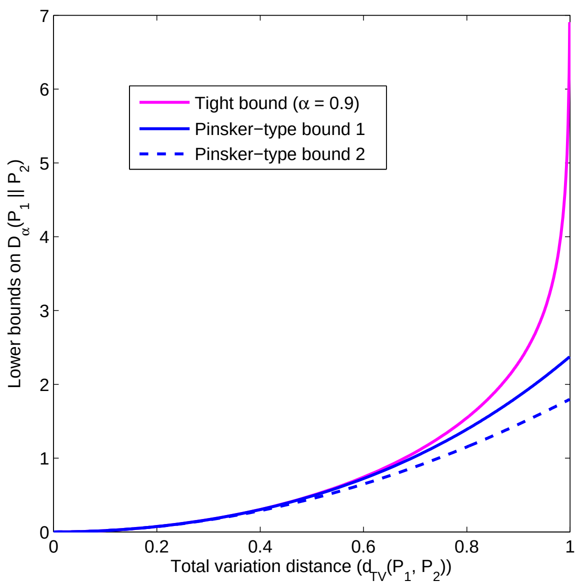

Fig. 1. A plot of the minimum of the Rényi divergence $D_\alpha(P_1||P_2)$ of order $\alpha = 0.90$ subject to a fixed total variation distance between P_1 and P_2. This tight lower bound is compared with the two Pinsker-type lower bounds in [13, Corollaries 6 and 9].

An alternative simplified form for the optimization problem in Proposition 1 is provided in the following for orders $\alpha \in (0,1)$. Hence, Proposition 1 applies to every $\alpha > 0$, whereas the following is restricted to $\alpha \in (0,1)$.

Proposition 2. Let $\alpha \in (0,1)$ and $\varepsilon \in (0,1)$ denote, respectively, the order of the Rényi divergence and the fixed value of the total variation distance. A solution of the minimization problem for g_α in Proposition 1 is obtained by calculating the objective function of (7) while $p = q+\varepsilon$, and $q \in (0, 1-\varepsilon)$ is the unique solution of the equation

$$f_{\alpha,\varepsilon}(q) = \frac{1-\alpha}{\alpha} \tag{13}$$

where

$$f_{\alpha,\varepsilon}(q) \triangleq \frac{\left(1-\frac{\varepsilon}{1-q}\right)^{\alpha-1} - \left(1+\frac{\varepsilon}{q}\right)^{\alpha-1}}{\left(1+\frac{\varepsilon}{q}\right)^{\alpha} - \left(1-\frac{\varepsilon}{1-q}\right)^{\alpha}}, \quad \forall q \in (0, 1-\varepsilon). \tag{14}$$

is a strictly monotonic increasing, positive and continuous function, and

$$\lim_{q\to 0^+} f_{\alpha,\varepsilon}(q) = 0, \qquad \lim_{q\to(1-\varepsilon)^-} f_{\alpha,\varepsilon}(q) = +\infty. \tag{15}$$

Proof. See [16, Appendix III]. The proof relies on Lagrange duality and KKT conditions, while strong duality is first asserted by verifying the satisfiability of Slater's condition (see [2, Chapter 5]). □

Remark 2. Since $f_{\alpha,\varepsilon}\colon (0, 1-\varepsilon) \to (0,\infty)$ and this function is strictly monotonic increasing, the task of numerically solving equation (13) and finding its unique solution is easy.

Remark 3. It should be noted that, in light of Remark 2, the running time of our computer program for a numerical calculation of $g_\alpha(\varepsilon)$ with Proposition 2, for $\alpha \in (0,1)$, has been considerably reduced (by a factor of 100) in comparison to its direct computation with Proposition 1 (note, however, that Proposition 1 applies to every $\alpha \in (0,\infty)$). This significant reduction in the computational complexity is very helpful in Section 3. A high-precision computation of $g_\alpha(\varepsilon)$ with Proposition 2 requires about 1 msec on a standard PC.

3. The Achievable Region of $(D(Q||P_1), D(Q||P_2))$ for Arbitrary Q, P_1, P_2 Subject to a Minimal Total Variation Distance Between P_1 and P_2

In this section, we address the following question:

Question 1. What is the achievable region of all the points $\big(D(Q||P_1),\, D(Q||P_2)\big)$ when P_1 and P_2 are arbitrary PDs whose total variation distance is at least $\varepsilon \in (0,1)$, and Q is any PD that is absolutely continuous w.r.t. P_1, P_2 ?

The present section provides an exact characterization of this achievable region by relying on the results of Section 2, and the following lemma:

Lemma 3. Let P_1 and P_2 be mutually absolutely continuous probability measures, and let Q be a third probability measure such that $Q \ll P_1$. Then, for an arbitrary $\alpha > 0$,

$$D_\alpha(P_1||P_2) = D(Q||P_2) + \frac{\alpha}{1-\alpha}\cdot D(Q||P_1) + \frac{1}{\alpha-1}\cdot D(Q||Q_\alpha) \tag{16}$$

where Q_α is given by

$$Q_\alpha(x) \triangleq \frac{P_1^\alpha(x)\, P_2^{1-\alpha}(x)}{\sum_u P_1^\alpha(u)\, P_2^{1-\alpha}(u)}, \qquad \forall x \in \text{Supp}(P_1). \tag{17}$$

As a corollary of Lemma 3, the following tight inequality holds, which is attributed to Shayevitz (see [18, Section IV.B.8]). It will be useful for the continuation of this section, jointly with the results in Section 2.

Corollary 2. If $\alpha \in (0,1)$ then

$$\frac{\alpha}{1-\alpha} \cdot D(Q||P_1) + D(Q||P_2) \geq D_\alpha(P_1||P_2) \quad (18)$$

with equality if and only if $Q = Q_\alpha$. For $\alpha > 1$, inequality (18) is reversed with the same necessary and sufficient condition for an equality.

Remark 4. Corollary 2 with the optimizing PD Q_α in (17) strengthens [17, Eq. (6)] in the sense that it was stated there that, for $\alpha > 1$,

$$D_\alpha(P_1||P_2) = \max_{Q \ll P_1} \left\{ D(Q||P_2) + \frac{\alpha}{\alpha - 1} \cdot D(Q||P_1) \right\} \quad (19)$$

where the max is replaced by min for $\alpha \in (0,1)$. Equality (19) was proved in [17] by the method of types, and the optimizing PD $Q = Q_\alpha$ was stated in [18, Section IV.B.8]. The identity in Lemma 3 leads directly to the maximizing/ minimizing PD $Q = Q_\alpha$ (due to the non-negativity of the relative entropy). The knowledge of the maximizing PD in (17) is necessary for the characterization of the achievable region studied in this section.

The region that includes all the achievable points of $\big(D(Q||P_1), D(Q||P_2)\big)$ is determined as follows: let $d_{\text{TV}}(P_1, P_2) \geq \varepsilon$ for a fixed $\varepsilon \in (0,1)$, and let $\alpha \in (0,1)$ be chosen arbitrarily. By the tight lower bound in Section 2, we have

$$D_\alpha(P_1||P_2) \geq g_\alpha(\varepsilon) \quad (20)$$

where g_α is expressed in (7) or by the efficient algorithm in Proposition 2. For $\alpha \in (0,1)$ and for a fixed value of $\varepsilon \in (0,1)$, let $p = p_\alpha$ and $q = q_\alpha$ in $(0,1)$ be set to achieve the global minimum in (7) (note that, without loss of generality, one can assume that $p \geq q$ since if (p,q) achieves the minimum in (7) then also $(1-p, 1-q)$ achieves the same minimum). Consequently, the lower bound in (20) is attained by the pair of 2-element PDs (see Lemma 2)

$$P_1 = (p_\alpha, 1 - p_\alpha), \quad P_2 = (q_\alpha, 1 - q_\alpha). \quad (21)$$

From Corollary 2 and Eqs. (20) and (21), it follows that for every $\alpha \in (0,1)$

$$g_\alpha(\varepsilon) \leq D(Q||P_2) + \frac{\alpha}{1-\alpha} \cdot D(Q||P_1) \quad (22)$$

where equality in (22) holds if P_1 and P_2 are the 2-element PDs in (21), and Q is the respective PD in (17). Hence, there exists a triple of 2-element PDs P_1, P_2, Q that satisfy (22) with equality, and they are easy to calculate for every $\alpha \in (0,1)$ and $\varepsilon \in (0,1)$.

Remark 5. Similarly to (22), since $d_{\text{TV}}(P_1, P_2) = d_{\text{TV}}(P_2, P_1)$, it follows from (22) that

$$g_\alpha(\varepsilon) \leq D(Q||P_1) + \frac{\alpha}{1-\alpha} \cdot D(Q||P_2). \quad (23)$$

By multiplying both sides of inequality (23) by $\frac{1-\alpha}{\alpha}$ and relying on the skew-symmetry property in (11), it follows that (23) is equivalent to

$$g_{1-\alpha}(\varepsilon) \leq D(Q||P_2) + \frac{1-\alpha}{\alpha} \cdot D(Q||P_1)$$

which is inequality (22) when $\alpha \in (0,1)$ is replaced by $1-\alpha$. Hence, since (22) holds for every $\alpha \in (0,1)$, there is no additional information in (23).

Proposition 3. The intersection of the half spaces that are given in (22), where the parameter α varies continuously in (0,1), determines the joint range of $\big(D(Q||P_1), D(Q||P_2)\big)$ addressed in Question 1. Furthermore, all the points in this region are obtained by triples of 2-element PDs P_1, P_2, Q.

Proof. The boundary of this region is determined by letting α increase continuously in (0,1), and by drawing the following straight lines in the plane of $\big(D(Q||P_1), D(Q||P_2)\big)$:

$$D(Q||P_2) + \frac{\alpha}{1-\alpha} \cdot D(Q||P_1) = g_\alpha(\varepsilon), \quad \forall\, \alpha \in (0,1). \quad (24)$$

Once the boundary of this region is determined by the envelope of all the straight lines in (24), every point on the boundary of this region is a tangent point to one of the straight lines in (24). Furthermore, the triple of 2-element PDs P_1, P_2 and Q that achieves an arbitrary point on the boundary of this region is determined as follows:

- Find the slope s of the tangent line ($s < 0$), and determine $\alpha \in (0,1)$ such that $-\frac{\alpha}{1-\alpha} = s$ (see (24)). This gives that $\alpha = -\frac{s}{1-s}$.
- Determine the 2-element PDs $P_1 = (p, 1-p)$, $P_2 = (q, 1-q)$ such that $D_\alpha(P_1||P_2) = g_\alpha(\varepsilon)$ (see Prop. 2).
- Calculate the respective PD $Q = Q_\alpha$ in (17).

Every point on the plane $\big(D(Q||P_1), D(Q||P_2)\big)$, which is below the envelope of all the straight lines in (24) (i.e., the colored regions in Fig. 2) is not achievable by any triple of PDs P_1, P_2 and Q with $d_{\text{TV}}(P_1, P_2) \geq \varepsilon$. This is because every such a point violates at least one of the inequality constraints in (22). On the other hand, every point which is above this envelope is achievable by a triple of 2-element PDs P_1, P_2, Q. To verify the last claim, first note that it has been demonstrated to hold for all the points on the boundary. Furthermore, based on the set of inequalities in (22) for $\alpha \in (0,1)$ and $\varepsilon \in [0,1)$, choose an arbitrary interior point in the convex region which is above the envelope. Note that $g_\alpha(\cdot)$ is a strictly monotonic increasing and continuous function in $(0,1)$; it also tends to infinity as we let ε tend to 1 (see Lemma 1). This implies that the achievable region of $\big(D(Q||P_1), D(Q||P_2)\big)$, subject to the constraint where $D(P_1||P_2) \geq \varepsilon$, shrinks continuously as the value of $\varepsilon \in (0,1)$ is increased, and it therefore lies on the boundary of the respective achievable region for some $\varepsilon' > \varepsilon$. One can find, accordingly, the 2-element PDs P_1, P_2 and Q in a similar way to the 3-item procedure which was stated earlier in this proof where ε is replaced by ε'. This therefore shows that all points on the boundary of this region, as well as all the interior points to the right of this boundary (i.e., the points above the envelope of all the straight lines in (24)) are achievable by 2-element PDs; furthermore, none of the points below this envelope is achievable. $\square$

As it is shown in Fig. 2, the boundaries of these achievable regions become less curvy as $\varepsilon \to 1$.

Geometric Interpretation of the Chernoff information

We consider in the following the point in Fig. 2 which is specified, in the plane of $\big(D(Q||P_1), D(Q||P_2)\big)$, by the intersection of the straight line $D(Q||P_1) = D(Q||P_2)$ with the boundary of the achievable region for a fixed value of $\varepsilon \in (0,1)$. Based on the above explanation (see the third item after equation (24)), this intersection point satisfies the equality $D(Q_\alpha||P_1) = D(Q_\alpha||P_2)$ for some $\alpha \in (0,1)$, 2-element PDs P_1, P_2 with $d_{\text{TV}}(P_1,P_2) = \varepsilon$, and Q_α in (17). The two equal coordinates of this intersection point are therefore equal to the Chernoff information $C(P_1,P_2)$ (see [4, Section 11.9]). Due to the symmetry of the achievable region w.r.t. the line $D(Q||P_1) = D(Q||P_2)$ (this symmetry follows from the symmetry of the total variation distance where $d_{\text{TV}}(P_1,P_2) = d_{\text{TV}}(P_2,P_1)$), the slope of the tangent line to the boundary at this intersection point is $s = -1$ (see Fig. 2). This yields that $\alpha = -\frac{s}{1-s} = \frac{1}{2}$, and from Corollary 1 we have $g_\alpha(\varepsilon) = -\log(1-\varepsilon^2)$ for $\varepsilon \in [0,1)$. Hence, from (24) with $\alpha = \frac{1}{2}$, the equal coordinates of this intersection point are $D(Q||P_1) = D(Q||P_2) = -\frac{1}{2}\log(1-\varepsilon^2)$. Based on [15, Proposition 2], this value is equal to the minimum of the Chernoff information subject to a fixed total variation distance $\varepsilon \in [0,1)$. In the following, we also calculate the three PDs P_1, P_2 and Q that achieve this intersection point. Eq. (7) with $\alpha = \frac{1}{2}$ gives that

$$-2\log\big(\sqrt{pq} + \sqrt{(1-p)(1-q)}\big) = -\log(1-\varepsilon^2)$$

subject to the constraints $p, q \in [0,1]$ and $|p-q| \geq \varepsilon$. A possible solution of this equation is $p = \frac{1+\varepsilon}{2}$ and $q = \frac{1-\varepsilon}{2}$, so the respective 2-element PDs are given by $P_1 = \left(\frac{1+\varepsilon}{2}, \frac{1-\varepsilon}{2}\right)$, $P_2 = \left(\frac{1-\varepsilon}{2}, \frac{1+\varepsilon}{2}\right)$ and, from (17), $Q = \left(\frac{1}{2}, \frac{1}{2}\right)$. As a byproduct of the characterization of this achievable region, we provide a geometric interpretation of the minimal Chernoff information subject to a minimal total variation distance.

The straight line $D(Q||P_1) = D(Q||P_2)$, in the plane of Fig. 2, intersects the boundaries of the respective regions at points whose coordinates are equal to the minimum Chernoff information for the fixed total variation distance (ε). The equal coordinates of each of these 4 intersection points, referring to $\varepsilon = 0.50, 0.70, 0.90, 0.99$, are $-\frac{1}{2}\log(1-\varepsilon^2) = 0.144, 0.337, 0.830, 1.959$ nats, respectively (see Fig. 2).

Acknowledgment: This work has been supported by the Israeli Science Foundation (ISF), grant number 12/12.

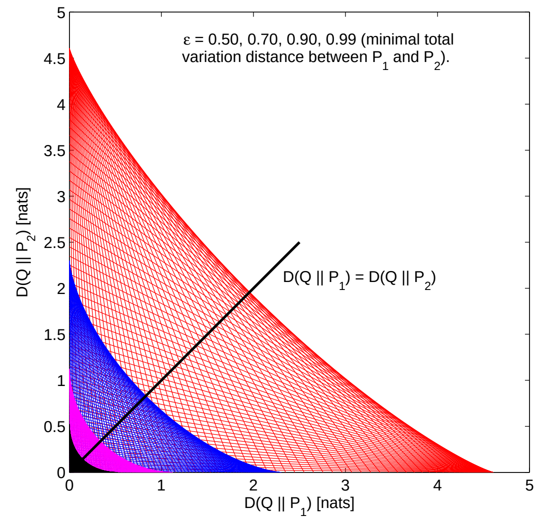

Fig. 2. The 4 achievable regions of $\big(D(Q||P_1), D(Q||P_2)\big)$ subject to a total variation distance between P_1 and P_2 of at least $\varepsilon = 0.50, 0.70, 0.90, 0.99$. The respective achievable region for a fixed ε is above the envelope of all straight lines in (24), shrinking as ε is increased.

References

[1] F. Alajaji, P. N. Chen and Z. Rached, "Csiszár's cutoff rates for the general hypothesis testing problem," *IEEE Trans. on Information Theory*, vol. 50, no. 4, pp. 663–678, April 2004.

[2] S. Boyd and L. Vanderberghe, *Convex Optimization*, Cambridge University Press, 2004.

[3] C. Bunte and A. Lapidoth, "Encoding tasks and Rényi entropy," *IEEE Trans. on Information Theory*, vol. 60, no. 9, pp. 5065–5076, September 2014.

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

[5] T. A. Courtade and S. Verdú, "Cumulant generating function of codeword lengths in optimal lossless compression," *Proceedings of the 2014 IEEE International Symposium on Information Theory*, pp. 2494–2498, Honolulu, HI, USA, July 2014.

[6] T. A. Courtade and S. Verdú, "Variable-length lossy compression and channel coding: non-asymptotic converses via cumulant generating functions," *Proceedings of the 2014 IEEE International Symposium on Information Theory*, pp. 2499–2503, Honolulu, HI, USA, July 2014.

[7] I. Csiszár, "Generalized cutoff rates and Rényi information measures," *IEEE Trans. on Information Theory*, vol. 41, no. 1, pp. 26–34, January 1995.

[8] 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.

[9] T. van Erven and P. Harremoës, "Rényi divergence and majorization," *Proceedings of the 2010 IEEE International Symposium on Information Theory*, pp. 1335–1339, Austin, Texas, USA, June 2010.

[10] A. A. Fedotov, P. Harremoës and F. Topsøe, "Refinements of Pinsker's inequality," *IEEE Trans. on Information Theory*, vol. 49, no. 6, pp. 1491–1498, June 2003.

[11] P. Harremoës and I. Vajda, "On pairs of f-divergences and their joint range," *IEEE Trans. on Information Theory*, vol. 57, no. 6, pp. 3230–3235, June 2011.

[12] G. L. Gilardoni, "On the minimum f-divergence for given total variation," *Comptes Rendus Mathematique*, vol. 343, no. 11–12, pp. 763–766, 2006.

[13] G. L. Gilardoni, "On Pinsker's and Vajda's type inequalities for Csiszár's f-divergences," *IEEE Trans. on Information Theory*, vol. 56, no. 11, pp. 5377–5386, November 2010.

[14] M. D. Reid and R. C. Williamson, "Information, divergence and risk for binary experiments," *Journal of Machine Learning Research*, vol. 12, pp. 731–817, March 2011.

[15] I. Sason, "Tight bounds for symmetric divergence measures and a refined bound for lossless source coding," *IEEE Trans. on Information Theory*, vol. 61, no. 2, pp. 701–707, February 2015.

[16] I. Sason, "On the Rényi divergence and a related problem on the joint range of relative entropies," March 2015. [Online]. Available at `http://arxiv.org/abs/1501.03616`.

[17] O. Shayevitz, "On Rényi measures and hypothesis testing," *Proceedings of the 2011 IEEE International Symposium on Information Theory*, pp. 800–804, Saint Petersburg, Russia, August 2011.

[18] O. Shayevitz, "A note on a characterization of Rényi measures, and its relation to composite hypothesis testing," December 2010. [Online]. Available at `http://arxiv.org/abs/1012.4401`.