Optimal Transport to Rényi Entropies

Olivier Rioul(✉)

LTCI, Télécom ParisTech, Université Paris-Saclay, 75013 Paris, France
olivier.rioul@telecom-paristech.fr
http://perso.telecom-paristech.fr/rioul/

Abstract. Recently, an optimal transportation argument was proposed by the author to provide a simple proof of Shannon's entropy-power inequality. Interestingly, such a proof could have been given by Shannon himself in his 1948 seminal paper. In fact, by 1948 Shannon established all the ingredients necessary for the proof and the transport argument takes the form of a simple change of variables.

In this paper, the optimal transportation argument is extended to Rényi entropies in relation to Shannon's entropy-power inequality and to a reverse version involving a certain conditional entropy. The transportation argument turns out to coincide with Barthe's proof of sharp direct and reverse Young's convolutional inequalities and can be applied to derive recent Rényi entropy-power inequalities.

Keywords: Rényi entropy · Entropy-power inequality · Optimal transport

1 Introduction: A Proof that Shannon Missed

2016 was the Shannon Centenary which marked the life and influence of Claude E. Shannon on the 100th anniversary of his birth. On this occasion many scientific events were organized throughout the world in honor of his achievements—on top of which his 1948 seminal paper [1] which developed the mathematical foundations of communication. The French edition of the book re-edition of Shannon's paper [2] has recently been published.

Remarkably, Shannon's revolutionary work, in a single publication [1], established the *fully* formed field of information theory, with all insights and mathematical proofs, albeit in sketched form. There seems to be only one exception in which Shannon's proof turned out to be flawed: the celebrated *entropy-power inequality* (EPI).

The EPI can be described as follows. Letting $P(X) = \frac{1}{n}\mathbb{E}\{\|X\|^2\}$ be the average power of a random vector X taking values in $\mathbb{R}^n$, Shannon defined the *entropy-power* $N(X)$ as the *power* of a zero-mean white Gaussian random vector X^* having the same *entropy* as X. He argued [1, Sect. 21] that for continuous random vectors it is more convenient to work with the entropy-power $N(X)$ than with the differential entropy $h(X)$. By Shannon's formula [1, Sect. 20.6]

F. Nielsen and F. Barbaresco (Eds.): GSI 2017, LNCS 10589, pp. 1–8, 2017.
https://doi.org/10.1007/978-3-319-68445-1_17

$h(X^*) = \frac{n}{2}\log\big(2\pi e P(X^*)\big)$ for the entropy of the white Gaussian X^*, the closed-form expression of $N(X) = P(X^*)$ when $h(X^*) = h(X)$ is

$$N(X) = \frac{\exp\big(\frac{2}{n}h(X)\big)}{2\pi e} \tag{1}$$

which is essentially e to the *power* a multiple of the *entropy* of X, also recognized as the "entropy power" of X in this sense. Since the Gaussian maximizes entropy for a given power [1, Sect. 20.5]: $h(X) \le \frac{n}{2}\log\big(2\pi e P(X)\big)$, the entropy-power does not exceed the actual power: $N(X) \le P(X)$ with equality if and only if X is white Gaussian. The power of a *scaled* random vector is given by $P(aX) = a^2 P(X)$, and the same property holds for the entropy-power:

$$N(aX) = a^2 N(X) \tag{2}$$

thanks to the well-known scaling property of the entropy [1, Sect. 20.9]:

$$h(aX) = h(X) + n\log|a| \tag{3}$$

Now for any two *independent* continuous random vectors X and Y, the power of the sum equals the sum of the individual powers: $P(X+Y) = P(X) + P(Y)$ and clearly the same relation holds for the entropy-power in the case of white Gaussian vectors (or Gaussian vectors with proportional covariances). In general, however, the entropy-power of the sum exceeds the sum of the individual entropy-powers:

$$N(X+Y) \ge N(X) + N(Y) \tag{4}$$

where equality holds only if X and Y are Gaussian with proportional covariances. This is the celebrated entropy-power inequality (EPI) as stated by Shannon.

It is remarkable that Shannon had the intuition of this inequality since it turns out to be quite difficult to prove. Shannon's proof [1, Appendix 6] is an incomplete variational argument which shows that Gaussian densities yield a stationary point for $N(X+Y)$ with fixed $N(X)$ and $N(Y)$ but this does not exclude the possibility that the stationary point is not a global minimum.

The first actual proof of the EPI occurred more than ten years later and was quite involved; subsequent proofs used either integration over a path of a continuous Gaussian perturbation or the sharp version of Young's inequality where the EPI is obtained as a limit (which precludes to settle the equality condition in this case). We refer to [3] for a comprehensive list of references and a detailed history.

Recently, an optimal transportation argument was proposed by the author [4,5] to provide a simple proof of the entropy-power inequality, including the equality condition. Interestingly, as we shall now demonstrate, such a proof, appropriately rephrased, could have been given by Shannon himself in his 1948 seminal paper. In fact, by 1948 Shannon established all the ingredients necessary for the proof. As in Shannon's paper [1], to simplify the presentation we assume, without loss of generality, that all considered random vectors have *zero mean* and we here restrict ourselves to real-valued random *variables* in one dimension $n = 1$.

The optimal transport argument takes the form of a simple change of variables: if e.g., X^* is Gaussian, then there exists a (possibly nonlinear) nondecreasing transformation T such that $T(X^*)$ is identically distributed as X—so that one would take $X = T(X^*)$ in what follows. Similarly if Y^* is Gaussian on can take $Y = U(Y^*)$. A detailed proof of this change of variable is given in [4,5] but this is easily seen as a generalization of the inverse c.d.f. method used e.g., for sampling random variables.

Theorem 1 (Shannon's Entropy-Power Inequality). *Let X, Y be independent zero-mean random variables with continuous densities. Then $N(X+Y) \geq N(X) + N(Y)$.*

Proof. The proof is in several steps, each being a direct consequence of Shannon's basic results established in [1].

1. We first proceed to prove the apparently more general inequality
$$N(aX + bY) \geq a^2 N(X) + b^2 N(Y) \tag{5}$$
for any real-valued coefficients a, b. By the scaling property of the entropy-power (2), this is in fact equivalent to the original EPI (4).
2. We can always assume that X and Y have the same entropy-power $N(X) = N(Y)$, or equivalently, have the same entropy $h(X) = h(Y)$. Otherwise, one could find constants c, d such that cX and dY have equal entropy-power (e.g., $c = \exp(-h(X))$ and $d = \exp(-h(Y))$) and applying (5) to cX and dY yields the general case, again thanks to the scaling property of the entropy-power.
3. Let X^*, Y^* be independent zero-mean Gaussian variables with the same entropy as X, Y. Since the entropies of X^* and Y^* are equal they have the same variance and are, therefore, identically distributed. Since equality holds in (5) for X^*, Y^*, we have $a^2N(X) + b^2N(Y) = a^2N(X^*) + b^2N(Y^*) = N(aX^* + bY^*)$ so that (5) is equivalent to $N(aX + bY) \geq N(aX^* + bY^*)$ or (taking the logarithm)
$$h(aX + bY) \geq h(aX^* + bY^*) \tag{6}$$
4. To prove (6) we may always assume the change of variables $X = T(X^*)$, $Y = U(Y^*)$ as explained above. One is led to prove that
$$h(aT(X^*) + bU(Y^*)) \geq h(aX^* + bY^*) \tag{7}$$
which is written only in terms of the Gaussian variables.
5. Since X^* and Y^* are i.i.d. Gaussian, the Gaussian variables $\tilde{X} = aX^* + bY^*$ and $\tilde{Y} = -bX^* + aY^*$ are uncorrelated and, therefore, independent. Letting $\Delta = a^2 + b^2$ we can write $X^* = (a\tilde{X} - b\tilde{Y})/\Delta$ and $Y^* = (b\tilde{X} + a\tilde{Y})/\Delta$. Since conditioning reduces entropy [1, Sect. 20.4],
$$\begin{aligned} h(aT(X^*) + bU(Y^*)) &= h(aT(\tfrac{a\tilde{X}-b\tilde{Y}}{\Delta}) + bU(\tfrac{b\tilde{X}+a\tilde{Y}}{\Delta})) \\ &\geq h(aT(\tfrac{a\tilde{X}-b\tilde{Y}}{\Delta}) + bU(\tfrac{b\tilde{X}+a\tilde{Y}}{\Delta})|\tilde{Y}) \end{aligned} \tag{8}$$

6. By the change of variable in the entropy [1, Sect. 20.8], for any transformation T, $h(T(X)) = h(X) + \mathbb{E}\log T'(X)$ where $T'(X) > 0$ is the jacobian of the transformation. Applying the transformation in $\tilde{X}$ for fixed $\tilde{Y}$ in the right-hand side of (8) we obtain

$$h(aT(\tfrac{a\tilde{X}-b\tilde{Y}}{\Delta}) + bU(\tfrac{b\tilde{X}+a\tilde{Y}}{\Delta})|\tilde{Y}) = h(\tilde{X}|\tilde{Y}) + \mathbb{E}\log\big(\tfrac{a^2}{\Delta}T'(\tfrac{a\tilde{X}-b\tilde{Y}}{\Delta}) + \tfrac{b^2}{\Delta}U'(\tfrac{b\tilde{X}+a\tilde{Y}}{\Delta})\big) \quad (9)$$

7. By the concavity of the logarithm,

$$\begin{aligned}\log\big(\tfrac{a^2}{\Delta}T'(\tfrac{a\tilde{X}-b\tilde{Y}}{\Delta}) + \tfrac{b^2}{\Delta}U'(\tfrac{b\tilde{X}+a\tilde{Y}}{\Delta})\big) &= \log\big(\tfrac{a^2}{\Delta}T'(X^*) + \tfrac{b^2}{\Delta}U'(Y^*)\big)\\ &\geq \tfrac{a^2}{\Delta}\log T'(X^*) + \tfrac{b^2}{\Delta}\log U'(Y^*) \quad (10)\end{aligned}$$

but again from change of variable in the entropy [1, Sect. 20.8], $\mathbb{E}\log T'(X^*) = h(T(X^*)) - h(X^*) = h(X) - h(X^*) = 0$ and similarly $\mathbb{E}\log U'(Y^*) = 0$. Thus the second term in the right-hand side of (9) is ≥ 0.

8. Since $\tilde{X}, \tilde{Y}$ are independent, one has [1, Sect. 20.2] $h(\tilde{X}|\tilde{Y}) = h(\tilde{X}) = h(aX^* + bY^*)$, which is the right-hand side of (7). Combining the established inequalities this proves the EPI. □

Remark 1. The case of equality can easily be settled by noting that equality holds in (10) only if $T'(X) = U'(Y)$ a.e., which since X and Y are independent implies that $T' = U'$ is constant, hence transformations T, U are linear and X, Y are Gaussian (see [4] for details).

Going back to the proof it is interesting to note that the only place where the gaussianity of X^*, Y^* is used is for the simplification $h(\tilde{X}|\tilde{Y}) = h(\tilde{X})$. If we drop this assumption we obtain the more general statement:

Corollary 1. *Let X, Y be independent zero-mean random variables with continuous densities, and similarly let X^*, Y^* be independent zero-mean random variables with continuous densities, all of equal entropies. Then for any real a, b,*

$$h(aX + bY) \geq h(aX^* + bY^*| - bX^* + aY^*) \quad (11)$$

If in addition we drop the assumption of equal entropies than letting $\lambda = a^2/\Delta$, $1 - \lambda = b^2/\Delta$ we obtain

Corollary 2. *Let X, Y be independent zero-mean random variables with continuous densities, and similarly let X^*, Y^* be independent zero-mean random variables with continuous densities. Then for any $0 < \lambda < 1$,*

$$\begin{aligned}&h(\sqrt{\lambda}X + \sqrt{1-\lambda}Y) - \lambda h(X) - (1-\lambda)h(Y) \quad (12)\\ &\geq h(\sqrt{\lambda}X^* + \sqrt{1-\lambda}Y^*| - \sqrt{1-\lambda}X^* + \sqrt{\lambda}Y^*) - \lambda h(X^*) - (1-\lambda)h(Y^*)\end{aligned}$$

In fact since the choice of (X, Y) and (X^*, Y^*) is arbitrary the latter inequality can be split into two inequalities [5], the EPI and a *reverse* EPI:

$$\begin{aligned}h(\sqrt{\lambda}X + \sqrt{1-\lambda}Y) &\geq \lambda h(X) + (1-\lambda)h(Y)\\ h(\sqrt{\lambda}X^* + \sqrt{1-\lambda}Y^*| - \sqrt{1-\lambda}X^* + \sqrt{\lambda}Y^*) &\leq \lambda h(X^*) + (1-\lambda)h(Y^*).\end{aligned} \quad (13)$$

2 Generalization to Rényi Entropies

We now extend the same argument to Rényi entropies.

Definition 1 (Hölder Conjugate). *Let $p > 0$, its Hölder conjugate is p' such that $\frac{1}{p} + \frac{1}{p'} = 1$. We write $p' = \infty$ if $p = 1$; note that p' can be negative if $p < 1$.*

Definition 2 (Rényi Entropy). *The Rényi entropy of order p of a random vector X with density $f \in L^p(\mathbb{R}^n)$ is defined by*

$$h_p(X) = -p' \log \|f\|_p = \frac{1}{1-p} \log \int_{\mathbb{R}^n} f^p. \tag{14}$$

As is well known, $h_p(X)$ is non-increasing in p and we recover Shannon's entropy by letting $p \to 1$ from above or below: $h(X) = \lim_{p \to 1} h_p(X)$. We also make the following definitions.

Definition 3 (Power Transformation). *Given a random vector X with density $f \in L^\alpha$, we define X_α as the random vector with density*

$$f_\alpha = \frac{f^\alpha}{\int f^\alpha}. \tag{15}$$

Definition 4 (Young's Triple). *A* Young triple *(p, q, r) consists of three positive real numbers such that p', q', r' are of the same sign and*

$$\frac{1}{p'} + \frac{1}{q'} = \frac{1}{r'}. \tag{16}$$

The triple rate *λ associated to (p, q, r) is the ratio of $1/p'$ in $1/r'$:*

$$\lambda = \frac{1/p'}{1/r'} = \frac{r'}{p'} \qquad 1 - \lambda = \frac{1/q'}{1/r'} = \frac{r'}{q'}. \tag{17}$$

In other words $1/p + 1/q = 1 + 1/r$ as in the classical Young's inequality. If all p', q', r' are > 0 then $p, q, r > 1$; otherwise $p', q', r' < 0$ and $p, q, r < 1$. Thus we always have $0 < \lambda < 1$.

Definition 5 (Dual Young's Triple). *A Young triple (p^*, q^*, r^*) (with rate λ^*) is* dual *to (p, q, r) if it satisfies $r^* = \frac{1}{r}$ and $\lambda^* = 1 - \lambda$.*

From the definition we have $p, q, r > 1 \iff p^*, q^*, r^* < 1$ and *vice versa*. Since $\frac{1}{p^{*\prime}} = \lambda^* \frac{1}{r^{*\prime}} = \frac{1/r' - 1/p'}{1/r'}(1 - r) = \frac{1/r - 1/p}{1/r}$ and similarly for $q^{*\prime}$, the definition fully determines (p^*, q^*, r^*) as

$$(p^* = \frac{p}{r}, q^* = \frac{q}{r}, r^* = \frac{1}{r}) \tag{18}$$

We observe from the definition that the dual of (p^*, q^*, r^*) is the original triple (p, q, r).

We can now state the following

Theorem 2. *Let X, Y be independent zero-mean random variables with continuous densities, and similarly let X^*, Y^* be independent zero-mean random variables with continuous densities. Then for any Young's triple (p, q, r) with dual (p^*, q^*, r^*),*

$$h_r(\sqrt{\lambda}X_{1/p} + \sqrt{1-\lambda}Y_{1/q}) - \lambda h_p(X_{1/p}) - (1-\lambda)h_q(Y_{1/q}) \tag{19}$$
$$\geq \lambda^* h_{p^*}(X^*_{1/p^*}) + (1-\lambda^*)h_{q^*}(Y^*_{1/q^*}) - h_{r^*}(-\sqrt{\lambda^*}X^*_{1/p^*} + \sqrt{1-\lambda^*}Y^*_{1/q^*})$$

Proof. The proof uses the same transportation argument $X = T(X^*)$, $Y = U(Y^*)$ as above, combined with an application of Hölder's inequality. It is omitted due to lack of space (but see Sect. 3.2 below). □

Remark 2. In (19) terms like $h_p(X_{1/p})$ may be simplified since

$$h_p(X_{1/p}) = \frac{1}{1-p}\log\frac{\int f}{(\int f^{1/p})^p} = \frac{1}{1-1/p}\log\int f^{1/p} = h_{1/p}(X). \tag{20}$$

The above form was chosen to stress the similarity with (12).

Remark 3. The inequality (19) is invariant by duality, in the sense that if we permute the roles of all variables (p, q, r, λ, X, Y) and starred variables $(p^*, q^*, r^*, \lambda^*, X^*, Y^*)$ we obtain the exact same inequality.

Remark 4. The case of equality can be determined as in the proof of Theorem 1: this is the case where $T' = U'$ is constant, hence transformations T, U are linear. Hence equality holds in (19) if and only if there exists a constant $c > 0$ such that X has the same distribution as cX^* and Y has the same distribution as cY^*.

3 Some Applications

3.1 Back to Shannon's Entropy-Power Inequality

There is a striking similarity between Theorem 2 and Corollary 2. In fact for fixed $\lambda = 1 - \lambda^*$, we can let $p, q, r \to 1$ from above (or below) so that $p^*, q^*, r^* \to 1$ from below (or above) to obtain

$$h(\sqrt{\lambda}X + \sqrt{1-\lambda}Y) - \lambda h(X) - (1-\lambda)h(Y) \tag{21}$$
$$\geq (1-\lambda)h(X^*) + \lambda h(Y^*) - h(-\sqrt{1-\lambda}X^* + \sqrt{\lambda}Y^*).$$

This is exactly (12) in Corollary 2 because the right-hand side can be rewritten as

$$\begin{aligned}
&(1-\lambda)h(X^*) + \lambda h(Y^*) - h(-\sqrt{1-\lambda}X^* + \sqrt{\lambda}Y^*)\\
&= h(X^*) + h(Y^*) - h(-\sqrt{1-\lambda}X^* + \sqrt{\lambda}Y^*) - \lambda h(X^*) - (1-\lambda)h(Y^*)\\
&= h(\sqrt{\lambda}X^* + \sqrt{1-\lambda}Y^*, -\sqrt{1-\lambda}X^* + \sqrt{\lambda}Y^*)\\
&\qquad - h(-\sqrt{1-\lambda}X^* + \sqrt{\lambda}Y^*) - \lambda h(X^*) - (1-\lambda)h(Y^*) \qquad (22)\\
&= h(\sqrt{\lambda}X^* + \sqrt{1-\lambda}Y^* | -\sqrt{1-\lambda}X^* + \sqrt{\lambda}Y^*) - \lambda h(X^*) - (1-\lambda)h(Y^*) \qquad (23)
\end{aligned}$$

where (22) holds because the entropy is invariant by rotation.

Thus, Theorem 2 implies the classical Shannon's entropy-power inequality. It is the natural generalization to Rényi entropies using optimal transport arguments.

Remark 5. The above calculation (22)–(23) also shows that the EPI and the "reverse EPI" (13) are in fact equivalent, as already noted in [5]. This is due to the fact that Theorem 2 is invariant by duality (Remark 3).

3.2 Relation to Sharp Young Direct and Reverse Inequalities

To simplify the presentation we stay with one-dimensional random variables. As in Corollary 2, since the choice of (X, Y) and (X^*, Y^*) is arbitrary, (19) can be simplified. If we let $X_{1/p}, Y_{1/q}$ be i.i.d. centered Gaussian, $\sqrt{\lambda}X_{1/p} + \sqrt{1-\lambda}Y_{1/q}$ also has the same Gaussian distribution, and since the Rényi entropy of a Gaussian variable $X \sim \mathcal{N}(m, \sigma^2)$ is easily found to be

$$h_p(X) = \frac{p' \log p}{2p} + \log \sqrt{2\pi\sigma^2}, \tag{24}$$

the l.h.s. of (19) is equal to $\frac{r'}{2}(\frac{\log r}{r} - \frac{\log p}{p} - \frac{\log q}{q})$. By the equality case (Remark 4) this expression is also the value taken by the r.h.s. of (19) when $X^*_{1/p^*}, Y^*_{1/q^*}$ are i.i.d. Gaussian (this can also be checked directly from the above definition of the dual Young's triple). Therefore, the expression $\frac{r'}{2}(\frac{\log r}{r} - \frac{\log p}{p} - \frac{\log q}{q})$ can be inserted between the two sides of (19) in Theorem 2. In other words, (19) is split into two equivalent inequalities which can be rewritten as

$$h_r(\sqrt{\lambda}X + \sqrt{1-\lambda}Y) - \lambda h_p(X) - (1-\lambda)h_q(Y) \geq \frac{r'}{2}(\frac{\log r}{r} - \frac{\log p}{p} - \frac{\log q}{q}) \tag{25}$$

with equality if and only if X and Y are i.i.d. Gaussian. Plugging the definition (14) of Rényi entropies and dividing by r' (which can be positive of negative), it is easily found [5] that (25) yields the optimal Young's direct and reverse inequalities:

$$\sqrt{\frac{r^{1/r}}{|r'|^{1/r'}}}\|f * g\|_r \leq \sqrt{\frac{p^{1/p}}{|p'|^{1/p'}}}\|f\|_p \cdot \sqrt{\frac{q^{1/q}}{|q'|^{1/q'}}}\|g\|_q. \tag{26}$$

for $p, q, r > 1$ ($r' > 0$) and the reverse inequality for $0 < p, q, r < 1$ ($r' < 0$), where f and g denote the densities of $\sqrt{\lambda}X$ and $\sqrt{1-\lambda}Y$. Equality holds if and only if $X/\sqrt{p'}$ and $Y/\sqrt{q'}$ are i.i.d. Gaussian. In fact, a closer look at (19) shows that it coincide with Barthe's transportation proof of sharp Young's inequalities [6, Lemma 1] which uses the same change of variables $X = T(X^*)$, $Y = U(Y^*)$ as above.

3.3 Rényi Entropy-Power Inequalities

Again to simplify the presentation we stay with two one-dimensional independent random variables X, Y. By analogy with the entropy-power (1), the Rényi entropy-power of order p is defined by

$$N_p(X) = \frac{\exp\big(\frac{2}{n}h_p(X)\big)}{2\pi e} \tag{27}$$

We have the following characterization which is an immediate generalization of the classical case $r = c = 1$:

Lemma 1. *Let $r > 0$, $c > 0$. The Renyi entropy-power inequality*

$$N_r(X+Y) \geq c\big(N_r(X) + N_r(Y)\big) \tag{28}$$

is equivalent to

$$h_r(\sqrt{\lambda}X + \sqrt{1-\lambda}Y) - \lambda h_r(X) - (1-\lambda)h_r(Y) \geq \frac{n}{2}\log c \quad \big(\forall \lambda \in (0,1)\big). \tag{29}$$

Now suppose $p^*, q^*, r^* < 1$ so that $r > 1$ is greater than p and q. Since $h_p(X)$ is non-increasing in p, one has $h_p(X) \geq h_r(X)$ and $h_q(Y) \geq h_r(Y)$, hence Theorem 2 in the form (25) implies (29) for any $\lambda \in (0,1)$ provided that $\frac{1}{2}\log c$ is taken as the minimum of the r.h.s. of (25) taken over all p, q such that $1/p + 1/q = 1 + 1/r$.

The method can easily be generalized to more than two independent random variables. In this way we obtain the recent Renyi entropy-power inequalities obtained by Bobkov and Chistyakov [7] and by Ram and Sason [8].

References

1. Shannon, C.E.: A mathematical theory of communication. Bell Syst. Tech. J. **27**, 379–423, 623–656 (1948)
2. Shannon, C.E., Weaver, W.: La théorie mathématique de la communication. Cassini, Paris (2017)
3. Rioul, O.: Information theoretic proofs of entropy power inequalities. IEEE Trans. Inf. Theory **57**(1), 33–55 (2011)
4. Rioul, O.: Yet another proof of the entropy power inequality. IEEE Trans. Inf. Theory **63**(6), 3595–3599 (2017)
5. Rioul, O.: Optimal transportation to the entropy-power inequality. In: IEEE Information Theory and Applications Workshop (ITA 2017), San Diego, USA, February 2017
6. Barthe, F.: Optimal Young's inequality and its converse: a simple proof. GAFA Geom. Funct. Anal. **8**(2), 234–242 (1998)
7. Bobkov, S.G., Chistyakov, G.P.: Entropy power inequality for the Rényi entropy. IEEE Trans. Inf. Theory **61**(2), 708–714 (2015)
8. Ram, E., Sason, I.: On Rényi entropy power inequalities. IEEE Trans. Inf. Theory **62**(12), 6800–6815 (2016)