

## REMARKS ON THE RÉNYI ENTROPY OF A SUM OF IID RANDOM VARIABLES

BENJAMIN JAYE, GALYNA V. LIVSHYTS, GRIGORIS PAOURIS,  
AND PETER PIVOVAROV

ABSTRACT. In this note we study a conjecture of Madiman and Wang [MW] which predicted that the generalized Gaussian distribution minimizes the Rényi entropy of the sum of independent random variables. Through a variational analysis, we show that the generalized Gaussian fails to be a minimizer for the problem.

### 1. INTRODUCTION

For  $p > 1$ , the  $p$ -Rényi [Re] entropy of a (continuous) random vector  $X$  in  $\mathbb{R}^d$  distributed with density  $f$  is defined by

$$h_p(X) = -\frac{1}{p-1} \log \int_{\mathbb{R}^d} f(x)^p d\mu_d(x) = -\frac{1}{p-1} \log \|f\|_p^p,$$

where  $\mu_d$  denotes the  $d$ -dimensional Lebesgue measure. As  $p \rightarrow 1^+$ ,  $h_p(X)$  converges to the usual Shannon entropy

$$h(X) = -\int_{\mathbb{R}^d} f(x) \log f(x) d\mu_d(x)$$

(provided that the density of  $X$  is sufficiently regular to justify passage of the limit). For the entropy power  $N(X) = \exp(2h(X)/d)$ , the fundamental entropy power inequality (EPI) of Shannon [Sh] asserts that for independent random vectors  $X_1$  and  $X_2$ ,

$$N(X_1 + X_2) \geq N(Z_1 + Z_2),$$

where  $Z_1, Z_2$  are independent Gaussians satisfying  $N(X_i) = N(Z_i), i = 1, 2$ . A firm connection between the EPI,  $p$ -Rényi entropy and fundamental results like the Brunn-Minkowski and Young's convolution inequalities goes back to Dembo, Cover and Thomas [DCT]. See Principe [Pr] for more information about where the Rényi entropy arises; see also Bobkov, Marsiglietti [BM] for a related discussion.

Recently, there has been increasing interest in  $p$ -Rényi entropy inequalities. Interestingly, the following basic mathematical question is still open: *Over all random variables  $X$  with  $h_p(X)$  some fixed quantity, what are the minimizers of the entropy  $h_p(X + X')$ , where  $X'$  is an*

*independent copy of  $X$ ?* We learnt about this question from the papers of Madiman, Melbourne, Xu, and Wang [MW, MMX1], who studied unifying entropy power inequalities for the Rényi entropy, which, in the limit  $p \rightarrow 1^+$  recover the statement that, over all probability distributions with  $h(X)$  fixed,  $h(X + X')$  is minimized if (and only if)  $X$  is a Gaussian, see e.g. [DCT].

Several closely related questions have been recently addressed involving the  $p$ -Rényi entropy power  $N_p(X) = \exp(\frac{2}{d}h_p(X))$ . Bobkov and Chistyakov [BCh2] show that there is a constant  $c > 0$ , depending on  $d$  and  $p$ , such that  $N_p(\sum_{j=1}^n X_j) \geq c \sum_{j=1}^n N_p(X_j)$  for independent random vectors  $X_1, \dots, X_n$ . A sharper form of the constant was subsequently found by Ram and Sason [RS]. Bobkov and Marsiglietti [BM1] proved that  $N_p(X_1 + X_2)^\alpha \geq N_p(X_1)^\alpha + N_p(X_2)^\alpha$  for  $X_1, X_2$  independent Random vectors if  $\alpha \geq \frac{p+1}{2}$ . There has been considerable further recent success extending the EPI to the Rényi setting [BCh, Li, LiMM, MM, RS, Rioul2].

Following [LYZ, MW, MMX1], for  $\beta > 0$ , consider the *Generalized Gaussian*

$$G_{\beta,p}(x) = \alpha(1 - \beta|x|^2)_+^{1/(p-1)},$$

where  $\alpha$  is chosen so that  $\int_{\mathbb{R}^d} G_{\beta,p}(x) d\mu_d(x) = 1$ . The generalized Gaussian is the distribution with the smallest second moment with a given Rényi entropy, see work of Lutwak, Yang, and Zhang [LYZ], as well as earlier results of Costa, Hero, and Vignat [CHV]. Madiman and Wang made the following bold conjecture (Conjecture IV.3 in [MW]).

**Conjecture 1.1** (The Madiman-Wang Conjecture). *If  $X_j, j = 1, \dots, n$ , are independent random variables with densities  $f_j$ , and  $Z_j$  are independent random variables distributed with respect to  $G_{\beta_j,p}$  where  $\beta_j$  is chosen so that  $h_p(X_j) = h_p(Z_j)$ , then*

$$h_p(X_1 + \dots + X_n) \geq h_p(Z_1 + \dots + Z_n).$$

This conjecture has been confirmed in the case  $p = +\infty$ , see [MMX, MMX2].

In this note we will show that unfortunately this conjecture does not hold in the special case when  $d = 1, p = 2, n = 2$  and  $X_1$  and  $X_2$  are identically distributed, see Section 4. However, we do suspect that a minimizing distribution is a relatively small perturbation of the generalized Gaussian.

Throughout this note we only consider the case where  $X_1, \dots, X_n$  are independent copies of a random variable  $X$  with density  $f$ . The question of finding the minimizer of  $h_p(X_1 + \dots + X_n)$  with  $h_p(X)$  fixed can then be rephrased as a constrained maximization problem,

which we introduce in Section 2. Subsequently, in Section 3 we take the first variation of this maximization problem. We have not been able to develop a satisfactory theory of the associated Euler-Lagrange equation (3.2), but we show in Section 4 that the generalized Gaussian is not a solution to (3.2), and so fails to be a maximizer of the extremal problem. We conclude the paper with some elementary remarks and speculation.

**Acknowledgement.** The first named author is supported by NSF DMS-1830128, DMS-1800015 and NSF CAREER DMS-1847301. The second named author is supported by the NSF CAREER DMS-1753260. The third named author is supported by the NSF DMS-1812240. The fourth named author is supported by the NSF DMS-1612936. The work was partially supported by the National Science Foundation under Grant No. DMS-1440140 while the authors were in residence at the Mathematical Sciences Research Institute in Berkeley, California, during the Fall 2017 semester.

The authors are especially grateful to the reviewers for valuable comments and suggestions, which helped improve the paper and clarify the exposition.

## 2. THE CONSTRAINED MAXIMIZATION PROBLEM

Denote by  $\mathcal{C}_n(f)$  the  $(n - 1)$ -fold convolution of a given function  $f$  with itself, that is,  $\mathcal{C}_n(f) = f * f * \cdots * f$ , where there are  $n$  factors of  $f$  (and  $n - 1$  convolutions). Then  $\mathcal{C}_1(f) = f$ . It will be convenient to set  $\mathcal{C}_0(f) = \delta_0$ , the Dirac delta measure, so that  $g * \mathcal{C}_0(f) = g$  for any measurable function  $g$ .

Throughout the text, we fix  $M > 0$ ,  $n \in \mathbb{N}$  and  $p \in (1, \infty)$ . We set

$$\mathcal{F} = \{f \in L^1(\mathbb{R}^d) \cap L^p(\mathbb{R}^d), f \geq 0, \|f\|_p^p = M, \|f\|_1 = 1\}$$

and consider the extremal problem

$$(2.1) \quad \begin{cases} \text{Maximize } \mathcal{I}(f) \stackrel{\text{def}}{=} \int_{\mathbb{R}^d} [\mathcal{C}_n(f)(x)]^p d\mu_d(x) \\ \text{subject to } f \in \mathcal{F}. \end{cases}$$

Put

$$(2.2) \quad \Lambda = \Lambda(p, M) = \sup\{\mathcal{I}(f) : f \in \mathcal{F}\}.$$

We begin with a simple scaling lemma, which we will use often in what follows.

**Lemma 2.1.** *Suppose that  $f \in L^1(\mathbb{R}^d) \cap L^p(\mathbb{R}^d)$  is non-negative, and  $\|f\|_1 > 0$ . The function*

$$\tilde{f} = \frac{1}{\lambda^d \|f\|_1} f\left(\frac{\cdot}{\lambda}\right), \text{ with } \lambda = \left(\frac{\|f\|_p^p}{M \|f\|_1^p}\right)^{\frac{1}{d(p-1)}},$$

belongs to  $\mathcal{F}$ , and

$$\mathcal{I}(\tilde{f}) = \frac{M}{\|f\|_p^p} \frac{1}{\|f\|_1^{p(n-1)}} \mathcal{I}(f).$$

*Proof.* Observe that, for any  $r \in [1, \infty)$ ,

$$\|\tilde{f}\|_r^r = \frac{1}{\lambda^{d(r-1)} \|f\|_1^r} \|f\|_r^r.$$

Plugging in  $r = 1$  and  $r = p$  (and recalling the definition of  $\lambda$ ) we see that  $\tilde{f} \in \mathcal{F}$ . Next, observe that

$$\mathcal{C}_n(\tilde{f})(x) = \frac{1}{\|f\|_1^n \lambda^d} \mathcal{C}_n(f)\left(\frac{x}{\lambda}\right) \text{ for any } x \in \mathbb{R}^d.$$

Whence,

$$\mathcal{I}(\tilde{f}) = \frac{1}{\lambda^{d(p-1)} \|f\|_1^{pn}} \mathcal{I}(f),$$

and the proof is complete by recalling the definition of  $\lambda$ .  $\square$

We next prove that (2.1) has a maximizer. A radial function  $f$  on  $\mathbb{R}^d$  is called decreasing if  $f(y) \leq f(x)$  whenever  $|y| \geq |x|$ .

**Proposition 2.2.** *The problem (2.1) has a lower-semicontinuous, radially decreasing, maximizer  $Q$ .*

*Proof.* First observe that for any measurable function  $f$ , iterating Riesz's rearrangement inequality [LL, Theorem 3.7] yields  $\mathcal{I}(f) \leq \mathcal{I}(f^*)$ , where  $f^*$  is the symmetric rearrangement of  $f$ ; see [B, Section 3.4] for related multiple convolution rearrangement inequalities and their equality cases. Also, notice that if  $f \in \mathcal{F}$ , then  $f^* \in \mathcal{F}$ .

Take non-negative functions  $f_j \in \mathcal{F}$  such that  $\Lambda = \lim_{j \rightarrow \infty} \mathcal{I}(f_j)$  (recall  $\Lambda$  from (2.2)). By replacing  $f_j$  with its symmetric rearrangement, we may assume that  $f_j$  are radial and decreasing. Passing to a subsequence if necessary, we may in addition assume that  $f_j \rightarrow f$  weakly in  $L^p(\mathbb{R}^d)$ . Consequently,  $f$  is radial, decreasing,  $f \geq 0$ , and  $\|f\|_p^p \leq M$ . (To see this, observe that the set of radial decreasing nonnegative functions with norm at most  $M^{1/p}$  is a closed convex set in  $L^p(\mathbb{R}^d)$ , so by Mazur's Lemma, see e.g. [LL, Theorem 2.13], this set is weakly closed.)

By modifying  $f$  on a set of measure zero if necessary, we may assume that  $f$  is lower semi-continuous<sup>1</sup>.

**Claim 2.3.** As  $j \rightarrow \infty$ ,  $f_j \rightarrow f$   $\mu_d$ -almost everywhere.

*Proof.* For  $r > 0$ , define  $v_j(r) = f_j(x)$  and  $v(r) = f(x)$  whenever  $|x| = r$ . Then since  $f_j$  converges weakly to  $f$  in  $L^p(\mathbb{R}^d)$ , we have that whenever  $I$  is a closed interval of finite Lebesgue measure in  $(0, \infty)$ ,

$$\lim_{j \rightarrow \infty} \int_I v_j(s) d\mu_1(s) = \int_I v(s) d\mu_1(s).$$

Insofar as the function  $v$  is non-decreasing, it has at most countably many points of discontinuity. If  $r > 0$  is a point of continuity of  $v$ , and  $I_k = [r - 2^{-k}, r]$ , then

$$v(r) = \lim_{k \rightarrow \infty} \frac{1}{2^{-k}} \int_{I_k} v(s) d\mu_1(s) = \lim_{k \rightarrow \infty} \lim_{j \rightarrow \infty} \frac{1}{2^{-k}} \int_{I_k} v_j(s) d\mu_1(s)$$

but since  $v_j$  is decreasing we have that  $v_j(s) \geq v_j(r)$  for  $s \in I_k$ . Thus

$$v(r) \geq \limsup_{j \rightarrow \infty} v_j(r).$$

Arguing similarly with intervals whose left end-point is  $r$ , we also have that

$$v(r) \leq \liminf_{j \rightarrow \infty} v_j(r).$$

Thus  $\lim_{j \rightarrow \infty} v_j = v$  at every point of continuity of  $v$ . If  $E$  is a countable set in  $(0, \infty)$ , then  $E \times \mathbb{S}^{d-1}$  is a Lebesgue null set in  $\mathbb{R}^d$ , so the claim follows.  $\square$

Notice that, as a consequence of this claim, Fatou's Lemma ensures that  $\|f\|_1 \leq 1$ . Our next claim is

**Claim 2.4.** If  $1 < q < p$ , then  $f_j \rightarrow f$  strongly in  $L^q(\mathbb{R}^d)$  as  $j \rightarrow \infty$ .

The proof of this claim is a variant of the Vitali convergence theorem (see e.g. Theorem 9.1.6 of [Ros]), but observe that it does not necessarily hold if one was to remove the radially decreasing property of the functions  $f_j$  (just consider a sequence of translates of a fixed function).

*Proof.* Fix  $\varepsilon > 0, \delta > 0$ . Insofar as the functions  $f_j$  and  $f$  are radially decreasing,

$$\bigcup_j \{|f_j| \geq \frac{\delta}{2}\} \cup \{|f| \geq \frac{\delta}{2}\} \subset B,$$

<sup>1</sup>If  $f$  is discontinuous at  $x \in \mathbb{R}^d$ , then define  $f(x) = \sup_{|y| > |x|} f(y)$  (i.e. the one-sided radial limit from the right). Then  $\{f > \lambda\}$  is open for every  $\lambda > 0$ .

where  $B$  is the closed ball centered at 0 of radius  $(\frac{2}{\mu_d(B(0,1))\delta})^{1/d}$ . (Otherwise we would have  $\|f_j\|_1 > 1$  for some  $j$ , or  $\|f\|_1 > 1$ .)

On  $\mathbb{R}^d \setminus B$ , we have  $|f_j| < \delta/2$  for every  $j$ , and  $|f| < \delta/2$ , whence

$$\int_{\mathbb{R}^d \setminus B} |f_j(x) - f(x)|^q d\mu_d(x) \leq \delta^{q-1} (\|f_j\|_1 + \|f\|_1) \leq 2\delta^{q-1} < \frac{\varepsilon}{3}$$

provided  $\delta > 0$  is chosen sufficiently small.

Now fix  $\varkappa > 0$ . Observe that,

$$\int_{B \cap \{|f_j - f| < \varkappa\}} |f_j(x) - f(x)|^q d\mu_d(x) \leq \mu_d(B) \varkappa^q < \frac{\varepsilon}{3}$$

if  $\varkappa$  is chosen sufficiently small. On the other hand, since  $B$  has finite measure, one can invoke continuity of measure from above, thus we have that  $f_j \rightarrow f$  in measure on  $B$  as  $j \rightarrow \infty$ . From the inequalities

$$\begin{aligned} \int_{B \cap \{|f_j - f| \geq \varkappa\}} |f_j(x) - f(x)|^q d\mu_d(x) &\leq \mu_d(B \cap \{|f_j - f| \geq \varkappa\})^{1-q/p} \|f_j - f\|_p^q \\ &\leq 2^q M^{q/p} \mu_d(B \cap \{|f_j - f| \geq \varkappa\})^{1-q/p}, \end{aligned}$$

we infer that there exists  $N \in \mathbb{N}$  such that

$$\int_{B \cap \{|f_j - f| \geq \varkappa\}} |f_j(x) - f(x)|^q d\mu_d(x) < \frac{\varepsilon}{3} \text{ for all } j \geq N.$$

Bringing these estimates together, it follows that  $\|f_j - f\|_q^q < \varepsilon$  for every  $j \geq N$ .  $\square$

Our next goal is to use this claim in order to show that  $\mathcal{I}(f) = \Lambda$ . To this end, observe that repeated application of Young's convolution inequality [LL] yields that, for any  $n$ -tuple of functions  $g_1, \dots, g_n$ ,

$$(2.3) \quad \left( \int_{\mathbb{R}^d} |g_1 * g_2 * \dots * g_n(x)|^p d\mu_d(x) \right)^{1/p} \leq \prod_{j=1}^n \|g_j\|_{(np)'} ,$$

where  $p' = p/(p-1)$  is the Hölder conjugate of  $p$ , so  $(np)'' = \frac{np}{np-p+1}$ . Since  $n > 1$ ,  $(np)'' \in (1, p)$ .

To apply this inequality, first use Minkowski's inequality to observe that,

$$|\mathcal{I}(g_1)^{1/p} - \mathcal{I}(g_2)^{1/p}| \leq \left( \int_{\mathbb{R}^d} |\mathcal{C}_n(g_1)(x) - \mathcal{C}_n(g_2)(x)|^p d\mu_d(x) \right)^{1/p},$$

but,

$$\mathcal{C}_n(g_1) - \mathcal{C}_n(g_2) = \sum_{k=0}^{n-1} \mathcal{C}_k(g_1) * (g_1 - g_2) * \mathcal{C}_{n-k-1}(g_2),$$

and hence

$$|\mathcal{I}(g_1)^{1/p} - \mathcal{I}(g_2)^{1/p}| \leq \sum_{k=0}^{n-1} \left( \int_{\mathbb{R}^d} |\mathcal{C}_k(g_1) * (g_1 - g_2) * \mathcal{C}_{n-k-1}(g_2)(x)|^p d\mu_d(x) \right)^{1/p}.$$

Appealing to (2.3) now yields,

$$|\mathcal{I}(g_1)^{1/p} - \mathcal{I}(g_2)^{1/p}| \leq \sum_{k=0}^{n-1} \|g_1\|_{(np)'}^k \|g_2\|_{(np)'}^{n-k-1} \|g_1 - g_2\|_{(np)'}$$

Returning to our sequence  $f_j$ , it is a consequence of Hölder's inequality that  $\|f_j\|_{(np)'} \leq \|f_j\|_1^\theta \|f_j\|_p^{1-\theta}$  with some  $\theta \in (0, 1)$  depending on  $n$  and  $p$ , so  $\|f_j\|_{(np)'} \leq C(M, n, p)$  (and the same inequality holds with  $f_j$  replaced by  $f$ ). Whence there is a constant  $C'(n, p, M)$  such that

$$|\mathcal{I}(f_j)^{1/p} - \mathcal{I}(f)^{1/p}| \leq C'(n, p, M) \|f_j - f\|_{(np)'}$$
 for every  $j$ .

Since  $(np)' \in (1, p)$ , Claim 2.4 yields that  $f_j \rightarrow f$  in  $L^{(np)'}$  as  $j \rightarrow \infty$ . Hence  $\mathcal{I}(f) = \lim_{j \rightarrow \infty} \mathcal{I}(f_j) = \Lambda$ . (It follows that  $f$  is not identically zero.)

It remains to show that  $f \in \mathcal{F}$ . To this end, we apply Lemma 2.1: Consider the function

$$\tilde{f} = \frac{1}{\|f\|_1 \lambda^d} f\left(\frac{\cdot}{\lambda}\right), \text{ with } \lambda = \left(\frac{\|f\|_p^p}{M\|f\|_1^p}\right)^{\frac{1}{d(p-1)}}.$$

Then  $\tilde{f} \in \mathcal{F}$  and  $\mathcal{I}(\tilde{f}) = \frac{M}{\|f\|_p^p} \frac{1}{\|f\|_1^{p(n-1)}} \Lambda$ . Consequently, if  $\|f\|_p^p < M$  or  $\|f\|_1 < 1$ , then  $\mathcal{I}(\tilde{f}) > \Lambda$ , which is absurd. Thus  $f \in \mathcal{F}$  and the proof of the proposition is complete.  $\square$

### 3. THE FIRST VARIATION

With the existence of a maximizer proved, we now wish to analyze it analytically.

To introduce the Euler-Lagrange equation associated to (2.1) it will be convenient to define, for a function  $f$ ,  $\mathcal{T}(f)(x) = f(-x)$ . Observe that, if  $f, g, h$  are non-negative measurable functions,

$$(3.1) \quad \int_{\mathbb{R}^d} f(x)(\mathcal{T}(g) * h)(x) d\mu_d(x) = \int_{\mathbb{R}^d} (f * g)(x)h(x) d\mu_d(x).$$

**Proposition 3.1.** *A lower-semicontinuous function  $Q \in \mathcal{F}$  is a maximizer of the problem (2.1) if and only if*

$$(3.2) \quad [\mathcal{T}(\mathcal{C}_{n-1}(Q))] * [\mathcal{C}_n(Q)]^{p-1} = \frac{\Lambda}{Mn} Q^{p-1} + \frac{\Lambda(n-1)}{n} \text{ on } \{Q > 0\}.$$

**Remark 3.2.** Observe that if  $Q$  is radially decreasing, then  $\mathcal{C}_{n-1}(Q)$  is again radially decreasing for any  $n \in \mathbb{N}$ , so  $\mathcal{T}(\mathcal{C}_{n-1}(Q)) = \mathcal{C}_{n-1}(Q)$  in this case.

*Proof.* The sufficiency is easy to show. Integrating both sides of (3.2) against  $Q$ , and recalling that  $Q \in \mathcal{F}$ , yields

$$\int_{\mathbb{R}^d} Q(x) \cdot ([\mathcal{T}(\mathcal{C}_{n-1}(Q))] * [\mathcal{C}_n(Q)](x))^{p-1} d\mu_d(x) = \Lambda.$$

But using Tonelli's theorem and (3.1), the left hand side is equal to  $\int_{\mathbb{R}^d} (\mathcal{C}_n(Q)(x))^p d\mu_d(x) = \mathcal{I}(Q)$ .

Conversely, consider a bounded function  $\varphi$  compactly supported in the open set  $\{Q > 0\}$ . Since  $Q$  is lower-semicontinuous,  $\inf_{\text{supp}(\varphi)} Q > 0$ . Therefore, (insofar as  $\varphi$  is bounded) there exists a constant  $C > 0$  such that

$$(3.3) \quad |\varphi| \leq CQ \text{ on } \mathbb{R}^d,$$

so in particular, there exists  $t_0 > 0$  such that for  $|t| \leq t_0$  it follows that  $Q_t \stackrel{\text{def}}{=} Q + t\varphi$  is non-negative. In the notation of Lemma 2.1 with  $f = Q_t$ , we consider the function

$$\tilde{Q}_t = \frac{1}{\lambda^d} \frac{(Q + t\varphi)\left(\frac{\cdot}{\lambda}\right)}{\|Q + t\varphi\|_1},$$

with the corresponding  $\lambda > 0$  satisfying  $\|\tilde{Q}_t\|_p^p = \|Q\|_p^p = M$ . Of course we also have  $\int_{\mathbb{R}^d} \tilde{Q}_t(x) d\mu_d(x) = 1$  regardless of  $\lambda$  for  $|t| < t_0$ . We conclude that  $\tilde{Q}_t$  belongs to  $\mathcal{F}$ , and therefore

$$(3.4) \quad \mathcal{I}(\tilde{Q}_t) \leq \mathcal{I}(Q) = \Lambda, \text{ for all } |t| < t_0.$$

Moreover, as in Lemma 2.1,

$$(3.5) \quad \mathcal{I}(\tilde{Q}_t) = \frac{1}{\lambda^{d(p-1)} \|Q + t\varphi\|_1^{np}} \int_{\mathbb{R}^d} [\mathcal{C}_n(Q + t\varphi)(x)]^p d\mu_d(x).$$

For  $|t| < t_0$ , we calculate, using commutativity and associativity of the convolution operator,

$$\frac{d}{dt} \mathcal{C}_n(Q + t\varphi)^p = pn[\varphi * \mathcal{C}_{n-1}(Q + t\varphi)][\mathcal{C}_n(Q + t\varphi)]^{p-1},$$

and

$$(3.6) \quad \begin{aligned} \frac{d^2}{dt^2} \mathcal{C}_n(Q + t\varphi)^p &= pn(n-1)\varphi * \varphi * \mathcal{C}_{n-2}(Q + t\varphi)[\mathcal{C}_n(Q + t\varphi)]^{p-1} \\ &\quad + n^2 p(p-1)[\varphi * \mathcal{C}_{n-1}(Q + t\varphi)]^2 [\mathcal{C}_n(Q + t\varphi)]^{p-2}. \end{aligned}$$



Crudely employing the bound (3.3) in (3.6), we infer that there is a constant  $C > 0$ , depending on  $n, p$  and  $t_0$ , such that for all  $|t| < t_0$ ,

$$\left| \frac{d^2}{dt^2} \mathcal{C}_n(Q + t\varphi)^p \right| \leq C \mathcal{C}_n(Q)^p.$$

Whence, the second order Taylor formula yields that

$$(3.7) \quad \left| \mathcal{C}_n(Q + t\varphi)^p - \mathcal{C}_n(Q)^p - npt[\varphi * \mathcal{C}_{n-1}(Q)][\mathcal{C}_n(Q)]^{p-1} \right| \leq Ct^2 \mathcal{C}_n(Q)^p,$$

for  $|t| < t_0$ . Integrating the pointwise inequality (3.7) yields

$$(3.8) \quad \int_{\mathbb{R}^d} [\mathcal{C}_n(Q + t\varphi)(x)]^p d\mu_d(x) = \Lambda + npt \int_{\mathbb{R}^d} [\varphi * \mathcal{C}_{n-1}(Q)(x)][\mathcal{C}_n(Q)(x)]^{p-1} d\mu_d(x) + O(t^2)$$

as  $t \rightarrow 0$ .

Now, recalling the definition of  $\lambda$ , we calculate

$$(3.9) \quad \begin{aligned} \lambda^{d(p-1)} \|Q + t\varphi\|_1^{np} &= \frac{\|Q + t\varphi\|_p^p}{M} \|Q + t\varphi\|_1^{(n-1)p} \\ &= \left( 1 + \frac{pt}{M} \int_{\mathbb{R}^d} \varphi(x) Q(x)^{p-1} d\mu_d(x) + O(t^2) \right) \\ &\quad \cdot \left( 1 + t(n-1)p \int_{\mathbb{R}^d} \varphi(x) d\mu_d(x) + O(t^2) \right), \end{aligned}$$

where in the expansion of  $\|Q + t\varphi\|_p^p$  we have again used the inequality (3.3) to obtain the  $O(t^2)$  term.

Plugging the two expansions (3.9) and (3.8) into (3.5) yields that, as  $t \rightarrow 0$ ,

$$\begin{aligned} \mathcal{I}(\tilde{Q}_t) &= \Lambda + pt \left\{ n \int_{\mathbb{R}^d} [\varphi * \mathcal{C}_{n-1}(Q)(x)][\mathcal{C}_n(Q)(x)]^{p-1} d\mu_d(x) \right. \\ &\quad \left. - \frac{\Lambda}{M} \int_{\mathbb{R}^d} \varphi(x) Q^{p-1}(x) d\mu_d(x) - (n-1)\Lambda \int_{\mathbb{R}^d} \varphi(x) d\mu_d(x) \right\} + O(t^2). \end{aligned}$$

From (3.4) it follows that  $\lim_{t \rightarrow 0} \frac{\mathcal{I}(\tilde{Q}_t) - \mathcal{I}(Q)}{t} = 0$ , so the second term in the prior expansion must vanish, that is,

$$\int_{\mathbb{R}^d} \varphi(x) \left\{ n[\mathcal{T}(\mathcal{C}_{n-1}(Q))] * [\mathcal{C}_n(Q)]^{p-1}(x) - \frac{\Lambda}{M} Q^{p-1} - (n-1)\Lambda \right\} d\mu_d(x) = 0,$$

where (3.1) has been used. Since  $\varphi$  was any bounded function compactly supported in  $\{Q > 0\}$ , we conclude that (3.2) holds.  $\square$

4. ON THE MADIMAN-WANG CONJECTURE

**Proposition 4.1.** *The generalized Gaussian is not the extremizer for problem (2.1).*

*Proof.* Consider the simplest case  $d = 1$ ,  $p = 2$ , and  $n = 2$ . We shall show that the function  $G(x) = \alpha(1 - |x|^2)_+$  does not satisfy the equation

$$(4.1) \quad \mathcal{C}_3(f) = af + b \text{ on } [-1, 1] \text{ with } a, b > 0,$$

and so no function of the form  $\frac{c}{\lambda}G(\frac{\cdot}{\lambda})$ , with  $c, \lambda > 0$ , satisfies (3.2), for any value of  $\Lambda$  (recall Remark 3.2). In fact, we shall show that  $\mathcal{C}_3(G) = G * G * G$  is not a quadratic polynomial near 0.

For this, observe:

$$G'' = 2\alpha(\delta_{-1} - \chi_{[-1,1]} + \delta_1).$$

Thus,  $(G * G * G)'''''' = (G'' * G'' * G'')$  is the threefold convolution of the above measure. The threefold convolution of  $-2\chi_{[-1,1]}$  equals  $-8(3 - |x|^2)_+$  on  $[-1, 1]$ , and no other term in the convolution  $G'' * G'' * G''$  is quadratic in  $|x|$ . Therefore,  $G * G * G$  has non-vanishing sixth derivative at 0, but  $a + bG$  does have vanishing sixth derivative at 0.  $\square$

**Remark 4.2.** Moreover, for any dimension  $d$ , the random vector  $X$  in  $\mathbb{R}^d$  with i.i.d. coordinates  $X_i$ , each distributed according to the generalized Gaussian density, does not constitute the extremizer for this problem. Indeed, in this case  $h_p(X) = dh_p(X_i)$ , and it remains to use Proposition 4.1. Therefore, a random vector with i.i.d. coordinates which are generalized Gaussians is not an extremal case for this question.

5. ANY RADIALLY DECREASING SOLUTION OF (3.2) IS COMPACTLY SUPPORTED

In this section, we discuss the following

**Proposition 5.1.** *Decreasing radial solutions of (3.2) are compactly supported.*

*Proof.* Suppose that  $Q \in \mathcal{F}$  solves (3.2) and  $Q$  is not compactly supported. Since  $Q$  is non-negative and radially decreasing, its support is  $\mathbb{R}^d$ .

The term  $G = \mathcal{C}_{n-1}(Q) * (\mathcal{C}_n(Q))^{p-1}$  on the left hand side of (3.2) belongs to  $L^r$ , where  $r = \max(1, 1/(p - 1))$ . Indeed, if  $p \geq 2$  then  $\int_{\mathbb{R}^d} G(x)d\mu_d(x) = \int_{\mathbb{R}^d} [\mathcal{C}_n(Q)(x)]^{p-1}d\mu_d(x)$  (recall that  $Q \geq 0$  with

$\int_{\mathbb{R}^d} Q(x) d\mu_d(x) = 1$ ), but  $\int_{\mathbb{R}^d} \mathcal{C}_n(Q)(x) d\mu_d(x) = 1$  and

$$\int_{\mathbb{R}^d} \mathcal{C}_n(Q)^p(x) d\mu_d(x) = \Lambda < \infty,$$

so  $\mathcal{C}_n(Q) \in L^{p-1}(\mathbb{R}^d)$ . If  $1 < p < 2$ , then  $t \mapsto t^{1/(p-1)}$  is convex, so by Jensen's inequality,  $G^{1/(p-1)} \leq \mathcal{C}_{n-1}(Q) * (\mathcal{C}_n(Q)^{p-1})^{1/(p-1)} = \mathcal{C}_{2n-1}(Q)$ , whence  $\|G\|_{1/(p-1)} \leq 1$  in this case.

On the other hand, the right hand side of (3.2) belongs to  $L^r$  only if  $\Lambda = 0$ , which is absurd, since  $\mathcal{F}$  certainly contains non-zero functions.  $\square$

## 6. REMARKS

In this section we make some remarks that suggest that although the generalized Gaussian is not an optimal distribution for the problem (2.1), a reasonably small perturbation of the generalized Gaussian could well be.

Beginning with  $f_0(x) = \mathbf{1}_{[-1,1]}$ , consider the following iteration for  $j \geq 1$

$$f_j(x) = \frac{\mathcal{C}_3(f_{j-1})(x) - \mathcal{C}_3(f_{j-1})(1)}{\mathcal{C}_3(f_{j-1})(0) - \mathcal{C}_3(f_{j-1})(1)}.$$

Numerically, this iteration converges pointwise to a solution of the equation (4.1) for some  $a, b > 0$  satisfying the constraints  $f(0) = 1$  and  $f(1) = 0$  (so the support of  $f$  is  $[-1, 1]$ ). The resulting function  $f$  can then be re-scaled via the transformation  $\frac{c}{\lambda} f(\frac{\cdot}{\lambda})$  ( $c, \lambda > 0$ ) to have any given positive integral and  $L^2$ -norm. We do not know if the solution of  $\mathcal{C}_3(f) = af + b$  is unique (modulo natural invariants in the problem), so we cannot say that this function  $f$  corresponds to a solution of the constrained maximization problem (2.1).

We provide the graphs of  $f_1, f_2, f_3$  and  $f_4$  (see Figure 1 below), and the algebraic expressions for  $f_1, f_2$  and  $f_3$  on  $[-1, 1]$ .

$$f_1(x) = 1 - x^2, \quad f_2(x) = 1 - \frac{6x^2}{5} + \frac{x^4}{5}$$

$$f_3(x) = 1 - \frac{62325x^2}{50521} + \frac{12810x^4}{50521} - \frac{1050x^6}{50521} + \frac{45x^8}{50521} - \frac{x^{10}}{50521}.$$

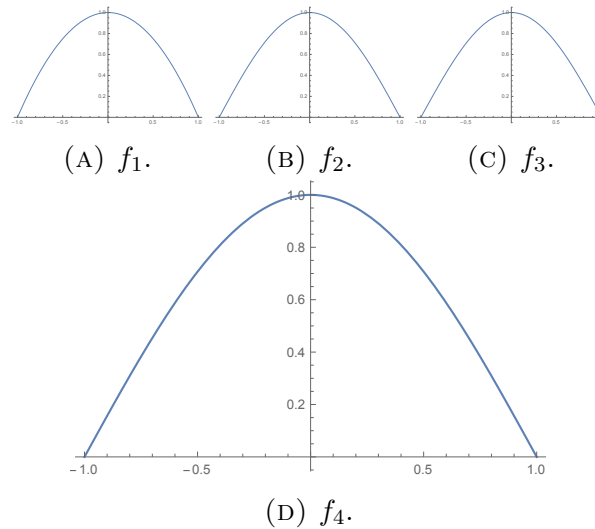


FIGURE 1. The graphs of  $f_1, \dots, f_4$  on  $[-1, 1]$ .

## REFERENCES

- [BCh] S. G. Bobkov, G. P. Chistyakov, *Bounds for the maximum of the density of the sum of independent random variables*, Zap. Nauchn. Sem. S.-Peterburg. Otdel. Mat. Inst. Steklov. (POMI), 408(Veroyatnost i Statistika. 18):62-73, 324, 2012.
- [BCh2] S. G. Bobkov, G. P. Chistyakov, *Entropy power inequality for the Rényi entropy*, IEEE Trans. Inform. Theory, 61(2):708-714, February 2015.
- [BM1] S. Bobkov, A. Marsiglietti, *Variants of the entropy power inequality*, IEEE Transactions of Information Theory, 63(12):7747-7752, (2017).
- [BM] S. Bobkov, A. Marsiglietti, *Asymptotic behavior of Rényi entropy in the central limit theorem*, submitted, arXiv:1802.10212.
- [B] A. Burchard, *Cases of equality in the Riesz rearrangement inequality*. Thesis (Ph.D.) Georgia Institute of Technology. 1994. 94 pp.
- [CHV] J. Costa, A. Hero, and C. Vignat, *On solutions to multivariate maximum alpha-entropy problems*, Lecture Notes in Computer Science, vol. 2683, no.EMMCVPR 2003, Lisbon, 7-9 July 2003, pp. 211-228, 2003.
- [DCT] A. Dembo, T. M. Cover and J. A. Thomas, *Information theoretic inequalities*, IEEE Transactions on Information Theory, vol. 37, no. 6, pp. 1501–1518, Nov. 1991.
- [Li] J. Li, *Rényi entropy power inequality and a reverse*, Studia Mathematica, 242 (2018) 303-319.
- [LiMM] J. Li, A. Marsiglietti, J. Melbourne, *Further investigations of Rényi entropy power inequalities and an entropic characterization of  $s$ -concave densities.*, arXiv preprint arXiv:1901.10616, 2019.
- [LL] E. Lieb and M. Loss *Analysis, 2nd Edition*, AMS Graduate Studies in Mathematics, 2002.

- [LYZ] E. Lutwak, D. Yang and G. Zhang, *Moment-Entropy Inequalities for a Random Vector*, IEEE Transactions on Information Theory, vol. 53, no. 4, pp. 1603–1607, April 2007.
- [MMX] M. Madiman, J. Melbourne, P. Xu, *Rogozin’s convolution inequality for locally compact groups*, arXiv preprint arXiv:1705.00642, 2017.
- [MMX1] M. Madiman, J. Melbourne and P. Xu, *Forward and reverse entropy power inequalities in convex geometry*. Convexity and Concentration, E. Carlen, M. Madiman, E. Werner (eds.), The IMA Volumes in Mathematics and its Applications, Vol. 161, pp. 427–485, Springer, 2017.
- [MW] M. Madiman and L. Wang, *Beyond the entropy power inequality, via rearrangements*. IEEE Transactions on Information Theory, vol. 60, no. 9, pp. 5116–5137, Sept. 2014.
- [MM] A. Marsiglietti, J. Melbourne, *On the entropy power inequality for the Rényi entropy of order  $[0, 1]$* , IEEE Transactions on Information Theory, 65(3):1387–1396, 2018.
- [Pr] J. C. Principe, *Information theoretic learning*, Springer-Verlag New York 2010.
- [RS] E. Ram, I. Sason, *On Rényi entropy power inequalities*, IEEE Transactions on Information Theory, 62(12):6800–6815, 2016.
- [Re] A. Rényi. *On Measures of Entropy and Information*. Proceedings of the Fourth Berkeley Symposium on Mathematical Statistics and Probability, Volume 1: Contributions to the Theory of Statistics, 547–561, University of California Press, Berkeley, Calif., 1961.
- [Rioul] O. Rioul, *Information theoretic proofs of entropy power inequalities*, IEEE Transactions on Information Theory, 57(1):33–55, 2010.
- [Rioul-1] O. Rioul, *Rényi entropy power inequalities via normal transport and rotation*, Entropy, 20(9):641, 2018.
- [Rioul2] O. Rioul, *Transportation proofs of Rényi entropy power inequalities*, arXiv preprint arXiv:1902.06120, 2019.
- [Ros] J. Rosenthal, *A first look at rigorous probability theory (Second ed.)* Hackensack, NJ: World Scientific Publishing Co, 2006.
- [Sh] C.E. Shannon. *A mathematical theory of communication*. Bell System Tech. J., 27:379–423, 623–656, 1948.
- [WM] L. Wang, M. Madiman. *Beyond the entropy power inequality, via rearrangements*, IEEE Transactions on Information Theory, 60(9):5116–5137, 2014.
- [MMX2] P. Xu, J. Melbourne, M. Madiman, *Infinity Entropy Power Inequalities*, Proc. IEEE Intl. Symp. Inform. Theory., 2017.
- [WM] R. Zamir, M. Feder, *A generalization of the entropy power inequality with applications*, IEEE Trans. Inform. Theory, 39(5):1723–1728, 1993.

SCHOOL OF MATHEMATICAL SCIENCES, CLEMSON UNIVERSITY  
*E-mail address:* `bjaye@clemson.edu`

DEPARTMENT OF MATHEMATICS, GEORGIA TECH  
*E-mail address:* `glivshyts6@math.gatech.edu`

DEPARTMENT OF MATHEMATICS, TEXAS A&M  
*E-mail address:* `grigoris@math.tamu.edu`

DEPARTMENT OF MATHEMATICS, UNIVERSITY OF MISSOURI  
*E-mail address:* `pivovarovp@missouri.edu`