The Annals of Probability
2023, Vol. 51, No. 5, 1603–1615
https://doi.org/10.1214/23-AOP1625

DECAY OF CONVOLVED DENSITIES VIA LAPLACE TRANSFORM

BY SERGEY G. BOBKOV[a]

School of Mathematics, University of Minnesota, [a]*bobkov@math.umn.edu*

Upper pointwise bounds are considered for convolution of bounded densities in terms of the associated Laplace and Legendre transforms. Applications of these bounds are illustrated in the central limit theorem with respect to the Rényi divergence.

1. Introduction. Given a random vector X in $\mathbb{R}^d$ with density p, we address the following question which often appears in a natural way: Under what conditions can one guarantee a certain decay of the density $p(x)$ at infinity? If p has a convolution structure, it turns out that an affirmative answer may be given under general moment-type conditions in terms of the associated Laplace transform. To give a precise statement, assume that

$$X = X_1 + \cdots + X_n \tag{1.1}$$

for some independent random vectors X_k in $\mathbb{R}^n$ with bounded densities p_k satisfying

$$p_k(x) \le M_k \quad \big(x \in \mathbb{R}^d, k = 1, \dots, n\big) \tag{1.2}$$

for some constants M_k (the number n does not need tend to infinity). Suppose that the convex function

$$V(t) = \log \mathbb{E} e^{\langle t, X\rangle}, \quad t \in \mathbb{R}^d,$$

is finite near zero, and introduce the corresponding Legendre transform

$$V^*(x) = \sup_{t \in \mathbb{R}^d} \big[\langle t, x\rangle - V(t)\big].$$

Here and elsewhere, we denote by $\langle \cdot, \cdot \rangle$ the canonical inner product in $\mathbb{R}^d$, and by $|\cdot|$ the Euclidean norm.

THEOREM 1.1. *Assume that X has mean zero. Under the conditions* (1.1)–(1.2) *with $n \ge 2$, the density p of X is continuous and satisfies*

$$p(x) \le M \exp\left\{-\frac{1}{2} V^*(x)\right\}, \quad x \in \mathbb{R}^d, \tag{1.3}$$

where the positive quantity M may be chosen as a function of $M_1, \dots, M_n$.

Under moment-type conditions on X_k, it is possible to derive upper bounds on the tail probabilities $\mathbb{P}\{|X_k| \ge r\}$ and $\mathbb{P}\{|X| \ge r\}$. However, in the case $n = 1$, even if the condition (1.2) is satisfied, it is not possible to get any information about the decay of $p(x)$ (cf. Remark 5.2 below). Thus, the convolution forces the density $p(x)$ to decay at a certain rate.

As we will see, the inequality (1.3) holds true with the geometric mean

$$M = (M_1 \cdots M_n)^{1/n}. \tag{1.4}$$

However, this choice may not reflect a correct asymptotic behavior of $p(x)$ with respect to the growing parameter n. The next variant based on harmonic-type means is more accurate.

Received March 2022; revised February 2023.
MSC2020 subject classifications. 60E, 60F.
Key words and phrases. Convolution, decay of densities.

THEOREM 1.2. *In the setting of Theorem 1.1, suppose that*

$$\sum_{j \neq k} M_j^{-\frac{2}{d}} \geq M_k^{-\frac{2}{d}} \tag{1.5}$$

for each $k \leq n$. Then the bound (1.3) holds true with

$$M = e^{\frac{d}{2}} \left(M_1^{-\frac{2}{d}} + \cdots + M_n^{-\frac{2}{d}}\right)^{-\frac{d}{2}}. \tag{1.6}$$

Let $M_k = M(X_k) = \|p_k\|_\infty$ be optimal in (1.2) and similarly $M = M(X)$. Applying the inequality (1.3) at the origin $x = 0$, we obtain that

$$M^{-\frac{2}{d}} \geq \frac{1}{e} \sum_{k=1}^{n} M_k^{-\frac{2}{d}}. \tag{1.7}$$

In the spirit of Shannon's entropy power inequality, this relation was derived in [5] (the constraint (1.5) is irrelevant); see, for example, [6, 11, 14] for further interesting information-theoretic developments in this direction involving the Rényi entropy functional. Using the central limit theorem, one can check that an asymptotic equality in (1.7) is attained when X_k's are uniformly distributed on the Euclidean ball in $\mathbb{R}^d$ with n and d growing to infinity. Hence, the factor $\frac{1}{e}$ is optimal, although it may be improved to $\frac{1}{2}$ in dimension $d = 1$.

To illustrate the advantage of (1.6) over (1.4), consider the weighted sums

$$Z_n = a_1 X_1 + \cdots + a_n X_n$$

of independent and identically distributed random vectors $X_1, \dots, X_n$ in $\mathbb{R}^d$ with coefficients $a_k \in \mathbb{R}$ such that $a_1^2 + \cdots + a_n^2 = 1$. Applying (1.7) to the random vectors $a_k X_k$ and using the homogeneity $M(a\xi) = a^{-d} M(\xi)$ of the maximum-of-density functional, we get an upper bound

$$M(Z_n) \leq e^{d/2} M_1$$

for the maximum of the density $q_n(x)$ of Z_n. A remarkable feature of this bound is that it does not not involve n and the weights a_k.

A similar phenomenon holds true when dealing with nonuniform bounds on $q_n(x)$. For example, in the sub-Gaussian case with the Laplace transform satisfying

$$\mathbb{E} e^{\langle t, X_1 \rangle} \leq \exp\left\{\frac{1}{2}\sigma^2 |t|^2\right\}, \quad t \in \mathbb{R}^d,$$

Theorem 1.2 yields a sub-Gaussian pointwise bound

$$q_n(x) \leq e^{d/2} M_1 \exp\left\{-\frac{1}{4\sigma^2}|x|^2\right\} \tag{1.8}$$

under the condition $\max_{k \leq n} a_k^2 \leq \frac{1}{2}$, to meet the requirement (1.5).

Theorems 1.1–1.2 are proved in Sections 2–3. In Sections 4–5, they are clarified for the class of subexponential and sub-Gaussian distributions. In particular, it will be shown that, with respect to the space variable x, the right-hand side of (1.8) may slightly be improved for a growing number of summands. This will be illustrated in the application to the central limit theorem by means of the Rényi divergence (Section 6). We will conclude with short remarks on Bernoulli convolutions (Section 7).

2. Preliminaries. With independent summands X_k in (1.1), we associate the log-Laplace transforms

$$V_k(t) = \log \mathbb{E} e^{\langle t, X_k \rangle} \quad \bigl(t \in \mathbb{R}^d, \quad k = 1, \dots, n\bigr).$$

Given a collection of real numbers $r_k \geq 2$ such that

$$\frac{1}{r_1} + \cdots + \frac{1}{r_n} = 1, \tag{2.1}$$

define the convex function

$$W(t) = \sum_{k=1}^{n} \frac{1}{r_k'} V_k\bigl(r_k' t\bigr), \quad r_k' = \frac{r_k}{r_k - 1}, \tag{2.2}$$

together with its Legendre transform $W^*(x) = \sup_{t \in \mathbb{R}^d} [\langle t, x \rangle - W(t)]$. Put

$$A_r = \bigl(r'\bigr)^{1/r'} r^{-1/r}, \quad r > 1. \tag{2.3}$$

As a preliminary step toward Theorems 1.1–1.2, we first derive the following.

LEMMA 2.1. *Under the conditions* (1.1)–(1.2) *with $n \geq 2$, the density p of the random vector X is continuous and satisfies*

$$p(x) \leq M \exp\bigl\{-W^*(x)\bigr\}, \quad x \in \mathbb{R}^d, \tag{2.4}$$

where

$$M = \prod_{k=1}^{n} A_{r_k}^{d/2} M_k^{1/r_k}. \tag{2.5}$$

PROOF. Introduce the characteristic functions

$$f_k(t) = \mathbb{E} e^{i\langle t, X_k \rangle} = \int_{\mathbb{R}^n} e^{i\langle t, x \rangle} p_k(x)\, dx, \quad t \in \mathbb{R}^d$$

so that the sum X has the characteristic function

$$f(t) = f_1(t) \cdots f_n(t) = \int_{\mathbb{R}^d} e^{i\langle t, x \rangle} p(x)\, dx.$$

Using the assumption (1.2) and applying the Hausdorff–Young inequality, we have that, for any $r \geq 2$ with conjugate $r' = \frac{r}{r-1}$,

$$\begin{aligned} \frac{1}{(2\pi)^d} \|f_k\|_r^r = \frac{1}{(2\pi)^d} \int_{\mathbb{R}^d} \bigl|f_k(t)\bigr|^r dx &\leq \left(\int_{\mathbb{R}^d} p_k(x)^{r'} dx \right)^{r/r'} \\ &\leq \left(\int_{\mathbb{R}^d} M_k^{r'-1} p_k(x)\, dx \right)^{r/r'} = M_k < \infty. \end{aligned}$$

Hence, by Hölder's inequality

$$\int_{\mathbb{R}^n} \bigl|f(t)\bigr|\, dx \leq \|f_1\|_{r_1} \cdots \|f_d\|_{r_d} < \infty$$

for any collection $r_k \geq 2$ satisfying (2.1). Thus, the function f is integrable so that the random vector X has a bounded continuous density described by the Fourier inversion formula

$$p(x) = \frac{1}{(2\pi)^d} \int_{\mathbb{R}^d} e^{-i\langle t, x \rangle} f(t)\, dt, \quad x = (x_1, \dots, x_d) \in \mathbb{R}^d. \tag{2.6}$$

To proceed, first suppose that both $p(x)$ and $f(t)$ have a Gaussian decay at infinity. In this case the functions V_k are all finite. In particular, the characteristic functions

$$f_k(z) = \int_{\mathbb{R}^d} e^{i\langle t,x\rangle - \langle y,x\rangle} p_k(x)\,dx, \qquad z = t + iy \quad (t, y \in \mathbb{R}^d, k = 1, \dots, n)$$

represent entire functions in the complex space $\mathbb{C}^d$ as well as the characteristic function

$$f(z) = \int_{\mathbb{R}^d} e^{i\langle z,x\rangle} p(x)\,dx, \quad z \in \mathbb{C}^d.$$

One can write down another formula, instead of (2.6), by using a contour integration. Fix $T > 0$ and $y = (y_1, \dots, y_d)$. Assuming for definiteness that $y_1 \geq 0$ and applying Cauchy's formula along the first coordinate, we have

$$\tag{2.7} \begin{aligned} &\int_{-T}^{T} e^{-it_1x_1} f(t)\,dt_1 + \int_0^{y_1} e^{-i(T+ih_1)x_1} f(T + ih_1, t_2, \dots, t_d)\,dh_1 \\ &\quad = \int_{-T}^{T} e^{-i(t_1+iy_1)x_1} f(t_1 + iy_1, t_2, \dots, t_d)\,dt_1 \\ &\qquad + \int_0^{y_1} e^{-i(-T+ih_1)x_1} f(-T + ih_1, t_2, \dots, t_d)\,dh_1, \end{aligned}$$

where $t = (t_1, \dots, t_d)$. For every $h \in \mathbb{R}^d$, the function $p_h(x) = e^{-\langle h,x\rangle} p(x)$ is integrable and has the Fourier transform $\hat{p}_h(t) = f(t + ih)$. Hence, by the Riemann–Lebesgue lemma, $f(t + ih) \to 0$ as $|t| \to \infty$. This convergence is actually uniform over any ball $|h| \leq r$, since the family $\{p_h\}_{|h|\leq r}$ is compact in $L^1(\mathbb{R}^d)$ in view of the continuity of the map $h \to p_h$. Thus, for any $y_1 > 0$,

$$\sup_{|h_1|\leq y_1} \big|f(\pm T + ih_1, t_2, \dots, t_n)\big| \to 0 \quad \text{as } T \to \infty.$$

We may conclude that the two integrals in (2.7) over the interval $[0, y_1]$ are vanishing as $T \to \infty$ so that, in the limit, this identity leads to

$$\int_{-\infty}^{\infty} e^{-it_1x_1} f(t)\,dt_1 = \int_{-\infty}^{\infty} e^{-i(t_1+iy_1)x_1} f(t_1 + iy_1, t_2, \dots, t_d)\,dt_1$$

for every $t_2, \dots, t_d \in \mathbb{R}$. By the decay assumption on $f(t)$, both sides of this equality have a sub-Gaussian behavior with respect to $t_2, \dots, t_n$. Hence, after multiplication of the equality by $e^{-it_2x_2}$, one can perform a similar contour integration with respect to the second coordinate. Continuing the process and recalling (2.6), we arrive at the following variant of the inversion formula:

$$p(x) = \frac{e^{\langle y,x\rangle}}{(2\pi)^d} \int_{\mathbb{R}^d} e^{-i\langle t,x\rangle} f(t + iy)\,dt,$$

which readily implies that

$$\tag{2.8} p(x) \leq \frac{e^{\langle y,x\rangle}}{(2\pi)^d} \int_{\mathbb{R}^d} \big|f(t + iy)\big|\,dt, \quad x, y \in \mathbb{R}^d.$$

On this step, one can remove the assumptions about the decay of p and f. Given $\varepsilon > 0$ and $\delta > 0$, consider the density $p_{\varepsilon,\delta} = p_\varepsilon * q_\delta$, where q_δ is a normal density on $\mathbb{R}^d$ with mean zero and covariance matrix $\delta^2 I_d$ and $p_\varepsilon(x) = c_\varepsilon p(x) e^{-\varepsilon|x|^2/2}$. Here the normalizing constants $c_\varepsilon \to 1$ as $\varepsilon \to 0$. Since $p(x)$ is bounded, $p_\varepsilon(x)$ has a Gaussian decay at infinity, and the same is true for $p_{\varepsilon,\delta}(x)$. Indeed, if $p(x) \leq M$, we get

$$p_{\varepsilon,\delta}(x) \leq c_\varepsilon M \int_{\mathbb{R}^d} e^{-\varepsilon|x-y|^2/2} q_\delta(y)\,dy = Be^{-b|x|^2}$$

for some constants $B, b > 0$, which do not depend on x. In addition, the Fourier transform $f_{\varepsilon,\delta}$ of $p_{\varepsilon,\delta}$ admits the bound $|f_{\varepsilon,\delta}(t)| \le e^{-\delta^2|t|^2/2}$. Hence, we are in position to apply the previous step to obtain the pointwise bounds

$$\begin{aligned} p_{\varepsilon,\delta}(x) &\le \frac{e^{\langle y,x\rangle}}{(2\pi)^d} \int_{\mathbb{R}^n} \big|f_{\varepsilon,\delta}(t+iy)\big|\,dt \\ &= \frac{e^{\langle y,x\rangle}}{(2\pi)^d} \int_{\mathbb{R}^d} \big|f_\varepsilon(t+iy)\big| e^{-\delta^2(|t|^2-|y|^2)/2}\,dt, \end{aligned}$$

where f_ε is the Fourier transform of p_ε. Since $p_{\varepsilon,\delta}(x) \to p_\varepsilon(x)$ for all x, in the limit as $\delta \to 0$ we then get a similar bound for the couple $(p_\varepsilon, f_\varepsilon)$. It will remain to send $\varepsilon \to 0$ and apply Fatou's lemma to $f_\varepsilon(t+iy)$, which leads to (2.8) without the decay constraints.

In order to bound the integral in (2.8), one may apply Hölder's inequality, which yields

$$\int_{\mathbb{R}^d} \big|f(t+iy)\big|\,dx \le \prod_{k=1}^n \left(\int_{\mathbb{R}^d} \big|f_k(t+iy)\big|^{r_k}\,dx\right)^{1/r_k} \tag{2.9}$$

for any collection of real numbers $r_k > 1$, as in (2.1). Recall that the functions $t \to f_k(t+iy)$ represent the Fourier transforms of the functions

$$p_{k,y}(x) = e^{-\langle y,x\rangle} p_k(x), \quad x \in \mathbb{R}^d.$$

We now involve the Hausdorff–Young inequality with optimal constants, due to Babenko [1] (for the values $r = 2, 4, 6, \dots$) and Beckner [3] (in the general case). It asserts that if a function q belongs to $L^{r'}(\mathbb{R}^d)$, $r \ge 2$, then its Fourier transform

$$\hat{q}(t) = \int_{\mathbb{R}^d} e^{i\langle t,x\rangle} q(x)\,dx, \quad t \in \mathbb{R}^d,$$

belongs to $L^r(\mathbb{R}^d)$ and has the norm satisfying

$$\|\hat{q}\|_r \le (2\pi)^{d/r} A_r^{d/2} \|q\|_{r'} \tag{2.10}$$

with constants defined in (2.3). Here an equality is attained for the normal densities $q = q_\delta$ as above (see also [2, 8]). Using (2.10) with $q = p_{k,y}$, we conclude that the L^{r_k}-norm of $f_k(t+iy)$ in (2.9) with $r_k \ge 2$ is bounded by

$$\begin{aligned} &(2\pi)^{d/r_k} A_{r_k}^{d/2} \left(\int_{\mathbb{R}^d} e^{-r_k'\langle y,x\rangle} p_k(x)^{r_k'}\,dx\right)^{1/r_k'} \\ &\quad\le (2\pi)^{d/r_k} A_{r_k}^{d/2} \left(\int_{\mathbb{R}^d} e^{-r_k'\langle y,x\rangle} M_k^{r_k'-1} p_k(x)\,dx\right)^{1/r_k'} \\ &\quad= (2\pi)^{d/r_k} A_{r_k}^{d/2} M_k^{1/r_k} \exp\left\{\frac{1}{r_k'} V_k\big(-r_k' y\big)\right\}. \end{aligned}$$

Therefore, according to the definition (2.2),

$$\int_{\mathbb{R}^d} \big|f(t+iy)\big|\,dx \le (2\pi)^d \prod_{k=1}^n A_{r_k}^{d/2} M_k^{1/r_k} \exp\{W(-y)\},$$

and, by (2.8)

$$p(x) \le M \exp\big\{\langle y,x\rangle + W(-y)\big\}$$

with constant M defined in (2.5). It remains to optimize this inequality over all $y \in \mathbb{R}^d$. $\square$

3. Proofs of Theorems 1.1–1.2: Refinement. Since $\mathbb{E}X = 0$, we may assume, without loss of generality, that $\mathbb{E}X_k = 0$ for all $k \le n$. In this case all $V_k(t)$ are nonnegative (by Jensen's inequality), with $V_k(0) = 0$. In view of the convexity of these functions, the functions $\alpha \to \frac{1}{\alpha} V_k(\alpha h)$ are nondecreasing on the half-axis $\alpha > 0$. Since $1 < r_k' \le 2$, we get

$$\frac{1}{r_k'} V_k\big(r_k' y\big) \le \frac{1}{2} V_k(2y),$$

and, therefore, the function W from Lemma 2.1 admits a simple upper bound

$$W(y) \le \frac{1}{2} V(2y), \quad y \in \mathbb{R}^d,$$

where $V = V_1 + \cdots + V_n$. Equivalently, $W^*(x) \ge \frac{1}{2} V^*(x)$ for all $x \in \mathbb{R}^d$. Thus, subject to the condition (2.1) on the collection $r_k \ge 2$, from (2.4), we obtain the desired upper bound

$$p(x) \le M \exp\left\{-\frac{1}{2} V^*(x)\right\} \tag{3.1}$$

with constant M, as in (2.5). Theorem 1.1 is thus proved.

The choice of equal powers $r_k = n$ leads to (3.1) with $M = (M_1 \cdots M_n)^{1/n}$.

Turning to Theorem 1.2, we need to analyze the expression (2.5). Put $u = (u_1, \dots, u_n)$, $u_k = \frac{1}{r_k}$, $v_k = 1 - u_k$, and rewrite (2.3) with $r = r_k$ as

$$A_{r_k} = \left(\frac{1}{r_k}\right)^{\frac{1}{r_k}} \left(\frac{1}{r_k'}\right)^{-\frac{1}{r_k'}} = u_k^{u_k} v_k^{-v_k}.$$

Thus, (2.5) becomes

$$M^{\frac{2}{d}} = \prod_{k=1}^n u_k^{u_k} v_k^{-v_k} \big(M_k^{\frac{2}{d}}\big)^{u_k}.$$

To simplify this expression, one may take

$$u_k = M_k^{-\frac{2}{d}} \big(M_1^{-\frac{2}{d}} + \cdots + M_n^{-\frac{2}{d}}\big)^{-1}, \tag{3.2}$$

and then

$$M = \big(M_1^{-\frac{2}{d}} + \cdots + M_n^{-\frac{2}{d}}\big)^{-\frac{d}{2}} \psi_n(u)^{-\frac{d}{2}}, \quad \psi_n(u) = \prod_{k=1}^n v_k^{v_k}.$$

As easy to check, the minimum to the function $\psi_n(u)$ on the simplex of all points u with $u_1 + \cdots + u_n = 1$, $u_k \ge 0$ is attained for the point with equal coordinates $u_k = \frac{1}{n}$, at which $\psi_n(u) = (1 - \frac{1}{n})^{n-1} > \frac{1}{e}$.

It remains to note that the condition $r_k \ge 2$, that is, $u_k \le \frac{1}{2}$ is equivalent to the condition (1.5) of Theorem 1.2.

With a similar argument, the inequality (1.3), with constant as in (1.6), may be slightly sharpened. Let us return to the scheme of the weighted sums

$$Z_n = a_1 X_1 + \cdots + a_n X_n, \quad a_1^2 + \cdots + a_n^2 = 1$$

of independent random vectors $X_1, \dots, X_n$ in $\mathbb{R}^d$ with coefficients satisfying

$$\max_{k \le n} a_k^2 \le \frac{1}{2}. \tag{3.3}$$

Assume that, for every $k \le n$, X_k has density p_k such that

$$p_k(x) \le M, \quad x \in \mathbb{R}^d, \tag{3.4}$$

and let

$$\mathbb{E}e^{\langle t,X_k\rangle} \le e^{V(t)}, \quad t \in \mathbb{R}^d \tag{3.5}$$

for a convex function $V : \mathbb{R}^d \to (-\infty, \infty]$. Define

$$W(t) = \sum_{k=1}^{n} (1 - a_k^2) V\left(\frac{a_k t}{1 - a_k^2}\right) \tag{3.6}$$

together with its Legendre transform W^*.

THEOREM 3.1. *Under the conditions* (3.3)–(3.5), *the density q_n of Z_n satisfies*

$$q_n(x) \le e^{d/2} M e^{-W^*(x)}, \quad x \in \mathbb{R}^d. \tag{3.7}$$

The function W in (3.6) corresponds to the definition (2.2) in Lemma 2.1 when it is applied to the random vectors $a_k X_k$ in place of X_k. In that case, in (3.2) one may use $a_k^{-d} M$ with parameter M from (3.4) in place of M_k. This leads to $u_k = a_k^2$, so that $r_k = \frac{1}{a_k^2} \ge 2$ under (3.3) and $r_k' = \frac{r_k}{r_k - 1} = \frac{1}{1 - a_k^2}$.

In the case of equal coefficients $a_k = 1/\sqrt{n}$, the condition (3.3) is satisfied for $n \ge 2$, and (3.6) is simplified to

$$W(t) = (n-1) V\left(\frac{\sqrt{n}\, t}{n-1}\right) \quad \text{with} \quad W^*(x) = (n-1) V^*\left(\frac{x}{\sqrt{n}}\right).$$

Hence, (3.7) leads to the following particular case.

COROLLARY 3.2. *Suppose that the independent random vectors $X_1, \dots, X_n$ have densities satisfying* (3.4)–(3.5). *Then the normalized sums $Z_n = \frac{1}{\sqrt{n}}(X_1 + \cdots + X_n)$, $n \ge 2$ have continuous densities q_n such that*

$$q_n(x) \le e^{d/2} M \exp\left\{-(n-1) V^*\left(\frac{x}{\sqrt{n}}\right)\right\}, \quad x \in \mathbb{R}^d. \tag{3.8}$$

4. Subexponential distributions. Although Theorems 1.1–1.2 are formally applicable without any constraint on the log-Laplace transform V, they make sense when the components of X_k have a finite exponential moment. Let us illustrate the sharpness of the nonuniform bound (1.3) with $n = 2$ on the example of the symmetric exponential distribution with density

$$p(x) = 2^{-d} e^{-\|x\|_1}, \quad x = (x_1, \dots, x_d) \in \mathbb{R}^d, \tag{4.1}$$

where we use the notation for the ℓ^1-norm $\|x\|_1 = |x_1| + \cdots + |x_d|$. The random vector with this distribution may be represented as the difference $X = X_1 - X_2$ of two independent random vectors X_1 and X_2 with density $p_1(x) = p_2(x) = e^{-\|x\|_1}$ supported on the positive octant $x_j > 0$, $1 \le j \le d$. In this case, $M_1 = M_2 = 1$, and for any $t = (t_1, \dots, t_d) \in \mathbb{R}^d$,

$$V(t) = \log \mathbb{E}e^{\langle t,X\rangle} = -\sum_{j=1}^{d} \log(1 - t_j^2), \quad \text{if } \|t\|_\infty = \max_j |t_j| < 1,$$

and $V(t) = \infty$, if $\|t\|_\infty \ge 1$. Hence, for the corresponding Legendre transform we have

$$V^*(x) = \sup_{\|t\|_\infty < 1} \sum_{j=1}^{d} \big(t_j x_j + \log(1 - t_j^2)\big) \ge \frac{1}{5} \sum_{j=1}^{d} \min(|x_j|, x_j^2).$$

Thus, the bound (1.3) with the geometric mean $M = \sqrt{M_1 M_2} = 1$ leads to

$$p(x) \le \exp\left\{-\frac{1}{10}\sum_{j=1}^{d} \min\left(|x_j|, x_j^2\right)\right\}.$$

This is consistent with (4.1) in view of the factor 2^{-d}.

Now, consider a general situation of the weighted sums

$$Z_n = a_1 X_1 + \cdots + a_n X_n, \quad a_1^2 + \cdots + a_n^2 = 1$$

of independent random vectors $X_1, \dots, X_n$ in $\mathbb{R}^d$ with mean zero satisfying

$$\mathbb{E}e^{c|\langle \theta, X_k\rangle|} \le 2, \quad 1 \le k \le n, \tag{4.2}$$

for all $\theta \in \mathbb{R}^d$, $|\theta| = 1$, with some common constant $c > 0$. That is, we assume that all linear functionals $\langle \theta, X_k\rangle$ have the ψ_1-Orlicz norm $\le 1/c$ for the Young function $\psi_1(r) = e^{|r|} - 1$, $r \in \mathbb{R}$. One may specialize Theorem 3.1 to the following assertion.

COROLLARY 4.1. *Under the conditions* (3.3)–(3.4) *and* (4.2), *the density* q_n *of* Z_n ($n \ge 2$) *satisfies*

$$q_n(x) \le e^{d/2} M \exp\left\{-\min\left(\frac{|x|}{12c}, \frac{|x|^2}{8}\right)\right\}, \quad x \in \mathbb{R}^d. \tag{4.3}$$

PROOF. Given a mean zero random variable ξ such that $\mathbb{E}e^{|\xi|} \le 2$, the convex function $R(t) = \log \mathbb{E}e^{t\xi}$ is smooth in $|t| < 1$, with $R(0) = R'(0) = 0$. Using $x^2 e^{t_0 x} \le e^x$ ($x \ge 0$) with $t_0 = 1 - \frac{2}{e}$, we have that, for all $|t| \le t_0$,

$$R''(t) \le \mathbb{E}\xi^2 e^{t\xi} \le \mathbb{E}\xi^2 e^{t_0|\xi|} \le \mathbb{E}e^{|\xi|} \le 2.$$

Hence, by Taylor's formula $R(t) \le t^2$. Applying this with $\xi = c\langle \theta, X_k\rangle$, it follows from (4.2) that

$$\log \mathbb{E}e^{\langle t, X_k\rangle} \le V(t), \qquad V(t) = |t|^2 \quad \text{for } t \in \mathbb{R}^d,\ |t| \le t_0/c.$$

By the assumption (3.3), $\frac{a_k}{1-a_k^2} \le \sqrt{2}$. Hence, in the ball $|t| \le \frac{t_0}{c\sqrt{2}}$, we have $\frac{a_k}{1-a_k^2}|t| \le \frac{t_0}{c}$, and according to (3.6),

$$W(t) = \sum_{k=1}^{n} \frac{a_k^2}{1-a_k^2}|t|^2 \le 2|t|^2, \qquad W^*(x) \ge \min\left(\frac{t_0}{2c\sqrt{2}}|x|, \frac{1}{8}|x|^2\right)$$

for all $x \in \mathbb{R}^d$. It remains to apply (3.7). $\square$

In the case of equal coefficients, the bound (4.3) with its exponential decay at infinity may naturally complement local limit theorems for densities under the moment assumption such as (4.2); compare, for example, [13].

5. Sub-Gaussian distributions. A random vector X in $\mathbb{R}^d$ (or its distribution) is called sub-Gaussian, if

$$\mathbb{P}\{|X| \ge r\} \le c_1 e^{-c_2 r^2}, \quad r \ge 0 \tag{5.1}$$

for some constants $c_1, c_2 > 0$, which do not depend on r (one may take $c_1 = 2$ at the expense of a smaller value of c_2 if necessary). If X has mean zero, an equivalent definition is that

$$\mathbb{E}e^{\langle t, X\rangle} \le e^{\sigma^2|t|^2/2}, \quad t \in \mathbb{R}^d \tag{5.2}$$

for some σ^2 (sometimes called the sub-Gaussian constant of X). This form is additive with respect to convolutions: If $X = X_1 + \cdots + X_n$ for independent random vectors satisfying

$$\mathbb{E}e^{\langle t, X_k \rangle} \le e^{\sigma_k^2 |t|^2/2}, \quad t \in \mathbb{R}^d, \tag{5.3}$$

then (5.2) holds true with $\sigma^2 = \sigma_1^2 + \cdots + \sigma_n^2$. In this special situation, Theorems 1.1–1.2 allows us to strengthen (5.1)–(5.2) as follows.

COROLLARY 5.1. *Under the condition* (5.3), *if X_k have bounded densities satisfying* (1.2) *with constants M_k, the density p of X admits the sub-Gaussian upper bound*

$$p(x) \le M \exp\left\{-\frac{|x|^2}{4\sigma^2}\right\}, \quad x \in \mathbb{R}. \tag{5.4}$$

Here the constant M may be defined according to the formula (1.6) from Theorem 1.2 subject to condition (1.5).

Subject to the conditions of Corollary 3.2 with $V(t) = \frac{1}{2}\sigma^2 |t|^2$ in (3.5) and with a common bound M on the densities of X_k, as in (3.4), the bound (5.4) may be sharpened for the densities q_n of the normalized sums

$$Z_n = \frac{X_1 + \cdots + X_n}{\sqrt{n}}. \tag{5.5}$$

Namely, the inequality (3.8) yields

$$q_n(x) \le e^{d/2} M \exp\left\{-\frac{n-1}{2n\sigma^2}|x|^2\right\}, \quad n \ge 2, \tag{5.6}$$

which improves upon the sub-Gaussian bound (5.4) for growing values of n.

REMARK 5.2. If $n = 1$, no moment-type condition, such as (3.5), guarantees the decay of a bounded density $p(x)$ at infinity. Given an unbounded Borel set $A \subset \mathbb{R}$ of a positive finite Lebesgue measure $c = |A|$, one may consider a multiple of the indicator function $p(x) = c^{-1} 1_A(x)$ in which case there is no decay,

$$\limsup_{|x| \to \infty} p(x) = \frac{1}{c}.$$

At the same time, the random variable X with such a density may be sub-Gaussian, for example: For $A = \bigcup_{k \in \mathbb{Z}} (k - h_k, k + h_k)$ with $h_k = \exp\{-k^2\}$, we have $\mathbb{E}\exp\{\varepsilon X^2\} < \infty$ whenever $\varepsilon < 1$.

6. Central limit theorem. The refinement (5.6) for the normalized sums in (5.5) is essential to decide, for example, whether the central limit theorem holds true in terms of the Rényi divergence

$$D_\alpha(Z_n \| Z) = \frac{1}{\alpha - 1} \log \int_{\mathbb{R}^d} \left(\frac{q_n(x)}{\varphi(x)}\right)^\alpha \varphi(x)\, dx, \quad \alpha > 0.$$

Here Z is a standard normal random vector in $\mathbb{R}^d$ with density φ. Let us recall that the quantity D_α is increasing α. For the range $0 < \alpha < 1$, it is (metrically) equivalent to the total variation distance between the distributions of Z_n and Z, and in the case $\alpha = 1$, it becomes the relative entropy (Kullback–Leibler divergence)

$$D(Z_n \| Z) = D_1(Z_n \| Z) = \int_{\mathbb{R}^d} \log\left(\frac{q_n(x)}{\varphi(x)}\right) q_n(x)\, dx.$$

The case $\alpha > 1$ is much stronger. For example, $D(Z_n\|Z)$ is finite under a second moment condition; while for the finiteness of $D_\alpha(Z_n\|Z)$ with $\alpha > 1$, it is necessary that $\mathbb{E}e^{c|Z_n|^2} < \infty$ for all $c < \frac{1}{2\alpha^*}$, where $\alpha^* = \frac{\alpha}{\alpha-1}$ is the conjugate index (in particular, the random vector Z_n and, therefore, all summands X_k should be sub-Gaussian). We refer an interested reader to [7, 17] for basic properties of these information-theoretic distances.

Assuming that the summands X_k are independent and identically distributed in $\mathbb{R}^d$, with mean zero and a unit covariance matrix, it was shown in [7] that, for a given $\alpha > 1$,

$$D_\alpha(Z_n\|Z) \to 0 \quad \text{as } n \to \infty, \tag{6.1}$$

if and only if for some $n = n_0$ (and then for any $n \geq n_0$), Z_n has a density q_n with finite $D_\alpha(Z_n\|Z)$, and

$$\mathbb{E}e^{\langle t,X_1\rangle} < e^{(\alpha^*)^2|t|^2/2}, \quad t \in \mathbb{R}^d, t \neq 0. \tag{6.2}$$

Necessarily, Z_n should have bounded densities, say q_n for some and then for all large n. As for the requirement that $D_\alpha(Z_n\|Z)$ is finite, it can be explored by using the upper bound (5.6). Applying it, we obtain the following sufficient condition.

COROLLARY 6.1. *Suppose that the i.i.d. random vectors X_k in $\mathbb{R}^d$ with mean zero and a unit covariance matrix satisfy*

$$\mathbb{E}e^{\langle t,X_1\rangle} \leq e^{\sigma^2|t|^2/2}, \quad t \in \mathbb{R}^d \tag{6.3}$$

with some $\sigma \geq 1$. If Z_n has a bounded density for some n, then the convergence (6.1) *holds true for any $\alpha < \frac{\sigma^2}{\sigma^2-1}$ (i.e., if $\alpha^* > \sigma^2$).*

Since X_1 is assumed to have a unit covariance matrix, it is necessary that $\sigma \geq 1$ in (6.3). The case $\sigma = 1$ is possible; it describes a rich family of probability distributions, including, for example, arbitrary convolutions of uniform distributions on bounded intervals (in dimension one, subject to the variance assumption). In that case we obtain the convergence (6.1) in the relative α-entropy for all $\alpha > 1$.

Let us also note that the assumption that Z_n has a bounded density for some n can be expressed in terms of the characteristic function $f(t) = \mathbb{E}e^{i\langle t,X_1\rangle}$. Namely (cf., e.g., [4]), this is equivalent to

$$\int_{\mathbb{R}^d} |f(t)|^\nu \, dt < \infty \quad \text{for some } \nu \geq 1.$$

PROOF. We may assume that X_1 has a density bounded by a constant M. By (6.3) the condition (6.2) is fulfilled as long as $\alpha^* > \sigma^2$. Moreover, Z_n has a bounded density q_n satisfying (5.5) for any $n \geq 2$. This bound implies that, for all $n \geq 2$,

$$\int_{\mathbb{R}^d} \left(\frac{q_n(x)}{\varphi(x)}\right)^\alpha \varphi(x)\, dx \leq c \int_{\mathbb{R}^d} \exp\left\{-\frac{1}{2}\left(\alpha\frac{n-1}{n\sigma^2} - (\alpha-1)\right)|x|^2\right\} dx,$$

where the constant c does not depend on x. Hence, $D_\alpha(Z_n\|Z)$ is finite for large n under the same assumption $\alpha^* > \sigma^2$. $\square$

7. Bernoulli convolutions. One possible application of sub-Gaussian density bounds may concern the distributions F_λ of random power series

$$Z_\lambda = \sqrt{1-\lambda^2} \sum_{k=0}^\infty \varepsilon_k \lambda^k, \quad 0 < \lambda < 1,$$

where $(\varepsilon_k)_{k\geq 0}$ are independent Bernoulli random variables taking the values ± 1 with probability $1/2$. Here the normalization $\sqrt{1-\lambda^2}$ is chosen so that $\mathbb{E}Z_\lambda^2 = 1$. A long-standing problem about these Bernoulli sums is to describe the values λ for which F_λ is absolutely continuous (Erdös' problem). In that case one also asks about general properties of the density of Z_λ.

In the range $0 < \lambda < \frac{1}{2}$, F_λ is known to be continuous singular; while for $\lambda = \frac{1}{2}$, we obtain the uniform distribution on the interval $(-\frac{1}{2}\sqrt{3}, \frac{1}{2}\sqrt{3})$. Erdös [9] constructed an infinite sequence of λ in $(\frac{1}{2}, 1)$ such that F_λ is singular. Although these values are bounded away from 1, he conjectured that the collection $\mathfrak{E}$ of all exceptional values, that is, when F_λ is singular, is clustering at $1-$ (which would imply that this collection is dense in this subinterval). On the other hand, for each integer $m \geq 0$, there is a number $\lambda_m \in (\frac{1}{2}, 1)$ such that F_λ has a density of class C^m for almost all $\lambda \in [\lambda_m, 1)$; compare [10]. Important results in this direction were later obtained by Solomyak [16] and Shmerkin [15] who, respectively, showed that the set $\mathfrak{E}$ has Lebesgue measure zero and actually has Hausdorff dimension zero (cf. also the review [12]).

In [16] it was actually shown that the characteristic function of Z_λ,

$$f_\lambda(t) = \mathbb{E}e^{itZ_\lambda} = \prod_{k=0}^{\infty} \cos\bigl(\sqrt{1-\lambda^2}\lambda^k t\bigr), \quad t \in \mathbb{R},$$

is square integrable for almost all $\lambda \in (\frac{1}{2}, 1)$, that is,

$$I(\lambda) = \|f_\lambda\|_2^2 = \int_{-\infty}^{\infty} f_\lambda(t)^2\, dt < \infty.$$

By Plancherel's theorem this ensures that Z_λ has a square integrable density p_λ. However, not much is known about the shape or decay of such densities. Nevertheless, using the self-similarity of F_λ and Corollary 5.1, one may speak about the Gaussian decay of p_λ for almost all λ that are sufficiently close to 1. More precisely, Solomyak's theorem may be complemented with the following observation.

COROLLARY 7.1. *For all $\lambda \in (2^{-1/4}, 1)$ such that $I(\lambda^4)$ is finite, the density p_λ of Z_λ is continuous and admits the sub-Gaussian upper bound*

$$p_\lambda(x) < \frac{1}{4} I\bigl(\lambda^4\bigr) e^{-x^2/4}, \quad x \in \mathbb{R}. \tag{7.1}$$

This bound is consistent with the property that $F_\lambda(x) = \mathbb{P}\{Z_\lambda \leq x\}$ approaches the standard normal distribution function $\Phi(x)$ in the weak sense as $\lambda \to 1$. This normal approximation may be quantified by applying the Berry–Esseen inequality, which readily yields

$$\sup_x \bigl|F_\lambda(x) - \Phi(x)\bigr| \leq c\sqrt{1-\lambda^2}$$

for all $\lambda \in (0, 1)$ up to some absolute constant c. It seems that a nonuniform bound similar to (7.1) also holds for the difference of densities $p_\lambda(x) - \varphi(x)$. One should, however, stress that (7.1) may only be useful once we are able to control the factor $I(\lambda^4)$ (which is finite for almost all $\lambda \in (2^{-1/4}, 1)$ by Solomyak's theorem).

PROOF. First note that, since $\mathbb{E}e^{t\varepsilon_k} = \cosh(t) \leq e^{t^2/2}$, $t \in \mathbb{R}$, a similar sub-Gaussian bound on the Laplace transform holds for any convergent Bernoulli sum $X = \sum_{k=0}^{\infty} \varepsilon_k a_k$ ($a_k \in \mathbb{R}$), that is,

$$\mathbb{E}e^{tX} \leq e^{\sigma^2 t^2/2}, \quad \sigma^2 = \mathbb{E}X^2 = \sum_{k=0}^{\infty} a_k^2.$$

The series $S_\lambda = \sum_{k=0}^{\infty} \varepsilon_k \lambda^k = X_1 + X_2$ may be represented as the sum of two independent random variables with

$$X_1 = S_{\lambda^4} + \lambda S'_{\lambda^4}, \qquad X_2 = \lambda^2 S''_{\lambda^4} + \lambda^3 \lambda S'''_{\lambda^4},$$

where S'_{λ^4}, S''_{λ^4}, S'''_{λ^4} are independent copies of S_{λ^4}. Being particular cases of Bernoulli sums, these random variables are sub-Gaussian, and moreover, for all $t \in \mathbb{R}$,

$$\mathbb{E}e^{tX_j} \le e^{\sigma_j^2 t^2/2}, \quad \sigma_j^2 = \mathbb{E}X_j^2 \ (j = 1, 2).$$

Denote by $g_\lambda(t) = f_\lambda(\frac{1}{\sqrt{1-\lambda^2}} t)$ the characteristic function of S_λ, and assume that $I(\lambda^4)$ is finite. The characteristic function h_1 of X_1 is given by $h_1(t) = g_{\lambda^4}(t) g_{\lambda^4}(\lambda t)$ so that, by Cauchy's inequality,

$$\begin{aligned}\int_{-\infty}^{\infty} |h_1(t)|\, dt &\le \|g_{\lambda^4}(t)\|_2 \|g_{\lambda^4}(\lambda t)\|_2 \\ &= \frac{\sqrt{1-\lambda^2}}{\sqrt{\lambda}} \|f_{\lambda^4}(t)\|_2^2 = \frac{\sqrt{1-\lambda^2}}{\sqrt{\lambda}} I(\lambda^4).\end{aligned}$$

Hence, X_1 has a continuous bounded density q_1. Moreover, by the Fourier inversion formula,

$$q_1(x) \le M_1 = \frac{\sqrt{1-\lambda^2}}{2\pi\sqrt{\lambda}} I(\lambda^4).$$

By a similar argument, the random variable X_2 has a continuous bounded density q_2 satisfying

$$q_2(x) \le M_2 = \frac{\sqrt{1-\lambda^2}}{2\pi\sqrt{\lambda^5}} I(\lambda^4).$$

We are in position to apply Corollary 5.1 to the couple (X_1, X_2) and conclude that S_λ has a density q satisfying the bound (5.4). That is, for all $x \in \mathbb{R}$, we have

$$q(x) \le M e^{-x^2/4\sigma^2}, \quad \sigma^2 = \sigma_1^2 + \sigma_2^2 = \mathbb{E}S_\lambda^2 = \frac{1}{1-\lambda^2}. \tag{7.2}$$

Here, according to (1.4), one may take

$$M = \sqrt{M_1 M_2} = \frac{\sqrt{1-\lambda^2}}{2\pi\lambda^{3/2}} I(\lambda^4). \tag{7.3}$$

Rescaling the variable, we thus obtain from (7.2)–(7.3) that

$$p_\lambda(x) = \frac{1}{\sqrt{1-\lambda^2}} q\left(\frac{x}{\sqrt{1-\lambda^2}}\right) \le \frac{1}{2\pi\lambda^{3/2}} I(\lambda^4) e^{-x^2/4}.$$

It remains to simplify the constant by using $\frac{1}{2\pi\lambda^{3/2}} \le \frac{2^{3/8}}{2\pi} < \frac{1}{4}$. $\square$

Funding. This work was supported by BSF Grant 2016050 and NSF Grant DMS-2154001.

REFERENCES

[1] BABENKO, K. I. (1961). An inequality in the theory of Fourier integrals. *Izv. Akad. Nauk SSSR Ser. Mat.* **25** 531–542. MR0138939

[2] BARTHE, F. (1998). Optimal Young's inequality and its converse: A simple proof. *Geom. Funct. Anal.* **8** 234–242. MR1616143 https://doi.org/10.1007/s000390050054

[3] BECKNER, W. (1975). Inequalities in Fourier analysis. *Ann. of Math.* (2) **102** 159–182. MR0385456 https://doi.org/10.2307/1970980

[4] BHATTACHARYA, R. N. and RANGA RAO, R. (1976). *Normal Approximation and Asymptotic Expansions*. *Wiley Series in Probability and Mathematical Statistics*. Wiley, New York. MR0436272

[5] BOBKOV, S. G. and CHISTYAKOV, G. P. (2012). 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** 62–73, 324. MR3032208 https://doi.org/10.1007/s10958-014-1836-9

[6] BOBKOV, S. G. and CHISTYAKOV, G. P. (2015). Entropy power inequality for the Rényi entropy. *IEEE Trans. Inf. Theory* **61** 708–714. MR3332742 https://doi.org/10.1109/TIT.2014.2383379

[7] BOBKOV, S. G., CHISTYAKOV, G. P. and GÖTZE, F. (2019). Rényi divergence and the central limit theorem. *Ann. Probab.* **47** 270–323. MR3909970 https://doi.org/10.1214/18-AOP1261

[8] BRASCAMP, H. J. and LIEB, E. H. (1976). Best constants in Young's inequality, its converse, and its generalization to more than three functions. *Adv. Math.* **20** 151–173. MR0412366 https://doi.org/10.1016/0001-8708(76)90184-5

[9] ERDÖS, P. (1939). On a family of symmetric Bernoulli convolutions. *Amer. J. Math.* **61** 974–976. MR0000311 https://doi.org/10.2307/2371641

[10] ERDÖS, P. (1940). On the smoothness properties of a family of Bernoulli convolutions. *Amer. J. Math.* **62** 180–186. MR0000858 https://doi.org/10.2307/2371446

[11] LI, J. (2018). Rényi entropy power inequality and a reverse. *Studia Math.* **242** 303–319. MR3794336 https://doi.org/10.4064/sm170521-5-8

[12] PERES, Y., SCHLAG, W. and SOLOMYAK, B. (2000). Sixty years of Bernoulli convolutions. In *Fractal Geometry and Stochastics*, *II* (*Greifswald/Koserow*, 1998). *Progress in Probability* **46** 39–65. Birkhäuser, Basel. MR1785620

[13] PETROV, V. V. (1975). *Sums of Independent Random Variables*. *Ergebnisse der Mathematik und Ihrer Grenzgebiete*, *Band* 82. Springer, New York. Translated from the Russian by A. A. Brown. MR0388499

[14] RAM, E. and SASON, I. (2016). On Rényi entropy power inequalities. *IEEE Trans. Inf. Theory* **62** 6800–6815. MR3599071 https://doi.org/10.1109/TIT.2016.2616135

[15] SHMERKIN, P. (2014). On the exceptional set for absolute continuity of Bernoulli convolutions. *Geom. Funct. Anal.* **24** 946–958. MR3213835 https://doi.org/10.1007/s00039-014-0285-4

[16] SOLOMYAK, B. (1995). On the random series $\sum \pm\lambda^n$ (an Erdős problem). *Ann. of Math.* (2) **142** 611–625. MR1356783 https://doi.org/10.2307/2118556

[17] VAN ERVEN, T. and HARREMOËS, P. (2014). Rényi divergence and Kullback–Leibler divergence. *IEEE Trans. Inf. Theory* **60** 3797–3820. MR3225930 https://doi.org/10.1109/TIT.2014.2320500