Annals of Functional Analysis

Ann. Funct. Anal. 6 (2015), no. 4, 1–29
http://doi.org/10.15352/afa/06-4-1
ISSN: 2008-8752 (electronic)
http://projecteuclid.org/afa

BEST POSSIBLE BOUNDS OF THE VON BAHR–ESSEEN TYPE

IOSIF PINELIS

Communicated by K. Berenhaut

Abstract. The well-known von Bahr–Esseen bound on the absolute pth moments of martingales with $p \in (1, 2]$ is extended to a large class of moment functions, and now with a best possible constant factor (which depends on the moment function). As an application, measure concentration inequalities for separately Lipschitz functions on product spaces are obtained. Relations with p-uniformly smooth and q-uniformly convex normed spaces are discussed.

1. Summary and discussion

1.1. **Summary.** Given any sequence $(S_j)_{j=1}^n$ of (real-valued) r.v.'s, let $X_j := S_j - S_{j-1}$ denote the corresponding differences, for $j \in \overline{1, n}$, with the convention $S_0 := 0$, so that $X_1 = S_1$; here and in what follows, for any m and n in the set $\{0, 1, \dots, \infty\}$ we let $\overline{m, n}$ stand for the set of all integers i such that $m \leqslant i \leqslant n$.

If $\mathsf{E}\,|X_j| < \infty$ and $\mathsf{E}(X_j|S_{j-1}) = 0$ for all $j \in \overline{2, n}$, let us say that the sequence $(S_j)_{j=1}^n$ is a *v-martingale* (where "v" stands for "virtual"); in such a case, let us also say that $(X_j)_{j=1}^n$ is a *v-martingale difference sequence*, or simply that the X_j's are v-martingale differences. Note that, for a general v-martingale difference sequence $(X_j)_{j=1}^n$, X_1 may be any r.v. whatsoever; in particular, its mean (if it exists) may or may not be 0. It is clear that any martingale $(S_j)_{j=1}^n$ is a v-martingale. Quite similarly one can define v-martingales with values in a normed space.

Date: Received: Nov. 10, 2014; Accepted: Feb. 2, 2015.

2010 *Mathematics Subject Classification.* Primary 60E15; Secondary 46B09, 46B20, 46B10.

Key words and phrases. Probability inequality, concentration of measure, product space, p-uniformly smooth normed space, q-uniformly convex normed space.

Introduce the following class of generalized moment functions:

$$
\begin{aligned}
\mathcal{F}_{1,2} := \big\{f\in C^1(\mathbb{R})\colon\ & f(0)=0,\ f \text{ is even},\\
& f' \text{ is nondecreasing and concave on } [0,\infty)\big\}\\
= \big\{f\in C^1(\mathbb{R})\colon\ & f(0)=0,\ f \text{ is even},\\
& f'' \text{ is nonnegative and nonincreasing on } (0,\infty)\big\};
\end{aligned} \tag{1.1}
$$

here, as usual, $C^1(\mathbb{R})$ is the class of all continuously differentiable real-valued functions on $\mathbb{R}$, and then f'' denotes the right derivative on $(0,\infty)$ of f'; on $(-\infty,0)$, f'' will denote the left derivative of f'. It is clear that each function $f\in\mathcal{F}_{1,2}$ is convex and hence nonnegative. Also, for each function $f\in\mathcal{F}_{1,2}$ one has $f'(0)=0$. It follows that $f'>0$ on $(0,\infty)$ and hence $f>0$ on $\mathbb{R}\setminus\{0\}$ for any function $f\in\mathcal{F}_{1,2}\setminus\{0\}$.

Theorem 1.1.

(I) For any $f\in\mathcal{F}_{1,2}\setminus\{0\}$, $n\in\overline{2,\infty}$, and v-martingale $(S_j)_{j=1}^n$,

$$\operatorname{\mathsf{E}} f(S_n)\leqslant \operatorname{\mathsf{E}} f(X_1)+C\sum_{j=2}^n \operatorname{\mathsf{E}} f(X_j) \tag{1.2}$$

with $C=C_f$, where

$$C_f:=\sup_{0<x<s<\infty}\frac{L_{f;s}(x)}{f(s)}, \tag{1.3}$$

$$L_{f;s}(x):=f(x-s)-f(x)+sf'(x). \tag{1.4}$$

(II) The constant factor C_f is the best possible in the sense that, for each $f\in\mathcal{F}_{1,2}\setminus\{0\}$ and each $n\in\overline{2,\infty}$, the number C_f is the smallest value of C such that inequality (1.2) holds for all v-martingales $(S_j)_{j=1}^n$; in fact, C_f is the best possible even if the differences $X_1,\dots,X_n$ are assumed to be any independent zero-mean r.v.'s.

(III) For each $f\in\mathcal{F}_{1,2}\setminus\{0\}$,

$$1\leqslant C_f\leqslant 2. \tag{1.5}$$

(IV) For each $C\in[1,2]$ there is some $f\in\mathcal{F}_{1,2}\setminus\{0\}$ such that $C_f=C$; in particular, it follows that the bounds 1 and 2 on C_f in (1.5) are the best possible ones.

Since all functions f in $\mathcal{F}_{1,2}$ are nonnegative, the expressions on both sides of inequality (1.2) are well defined. At that, it is possible for the right-hand side, or for both sides, of (1.2) to equal ∞. In the case when the differences $X_1,\dots,X_n$ are independent zero-mean r.v.'s, if the left-hand side of (1.2) is finite then (by Jensen's inequality) $\operatorname{\mathsf{E}} f(X_j)<\infty$ for each $j\in\overline{1,n}$, so that the right-hand side is finite as well; thus, for independent zero-mean $X_1,\dots,X_n$, the two sides of (1.2) are either both finite or both infinite.

1.2. **Discussion.** In this subsection, we shall

1. describe the structure of the class $\mathcal{F}_{1,2}$ as a convex cone, which will be useful in most of the proofs, and provide examples of functions in the class $\mathcal{F}_{1,2}$, including the (absolute) power functions and "extreme" functions (that is, functions belonging to the extreme rays of the convex cone $\mathcal{F}_{1,2}$);
2. present a general approach to effective calculation of the best possible constant C_f, with further information on this constant for the power functions and "extreme" functions;
3. give an application to the concentration of measure for separately Lipschitz functions on product spaces;
4. state other corollaries of the main theorem and relate the results with the relevant ones in the literature, by von Bahr and Esseen (vBE) and other authors.

Each of these items will be presented in a separate subsubsection.

1.2.1. *Structure of the class $\mathcal{F}_{1,2}$ and examples of functions in this class.* The following proposition describes the convex-cone structure of the class $\mathcal{F}_{1,2}$.

Proposition 1.2.

(I) A function $f\colon \mathbb{R}\to\mathbb{R}$ belongs to the class $\mathcal{F}_{1,2}$ if and only if there exists a (nonnegative, possibly infinite) Borel measure $\gamma=\gamma_f$ on $(0,\infty]$ such that $\int_{(0,\infty]}(t\wedge 1)\gamma(\,\mathrm{d}t)<\infty$ and

$$f(x)=\int_{(0,\infty]}\psi_t(x)\gamma(\,\mathrm{d}t) \tag{1.6}$$

for all $x\in\mathbb{R}$, where

$$\psi_t(x):=x^2-(|x|-t)_+^2,$$

assuming the conventions $u_+:=0\vee u$, $u_+^p:=(u_+)^p$, $u-\infty:=-\infty$, and $(-\infty)_+:=0$, for all real u, so that $\psi_\infty(x)=x^2$ for all $x\in\mathbb{R}$. Also,

$$\tfrac{1}{2t}\,\psi_t(x)\underset{t\downarrow 0}{\longrightarrow}|x| \tag{1.7}$$

uniformly in $x\in\mathbb{R}$.

(II) For each $f\in\mathcal{F}_{1,2}$, the corresponding measure $\gamma=\gamma_f$ is unique and determined by the condition that

$$\gamma\big((x,\infty]\big)=\tfrac12\,f''(x) \tag{1.8}$$

for all $x\in(0,\infty)$.

(III) For any $f\in\mathcal{F}_{1,2}$ and $x\in[0,\infty)$,

$$f'(x)=\int_{(0,\infty]}\psi_t'(x)\gamma(\,\mathrm{d}t)=2\int_{(0,\infty]}(x\wedge t)\gamma(\,\mathrm{d}t). \tag{1.9}$$

Proposition 1.2 will be used in the proofs of most of the other results of this paper.

Note that the rays $\mathbb{R}_+\psi_t$ corresponding to the functions ψ_t $\big($for $t\in(0,\infty]\big)$ are precisely the extreme rays of the convex cone $\mathcal{F}_{1,2}$, where $\mathbb{R}_+f:=\{\lambda f\colon \lambda\in(0,\infty)\}$, for any $f\in\mathcal{F}_{1,2}\setminus\{0\}$. This follows because the rays $\mathbb{R}_+\gamma_{\psi_t}=\mathbb{R}_+\delta_t$ $\big($with $t\in(0,\infty]\big)$ are precisely the extreme rays of the corresponding convex cone

$\{\gamma_f : f \in \mathcal{F}_{1,2}\}$ of measures, where δ_t stands for the Dirac measure at the point t. $\big($A ray $\mathbb{R}_+ f$ of a convex cone is called extreme if, for any nonzero f_1 and f_2 in the cone such that $f_1 + f_2 = f$, both f_1 and f_2 must lie on the ray.$\big)$

Also, note that $\psi_t(x) = x^2\,\mathbf{I}\{|x| < t\} + (2t|x| - t^2)\,\mathbf{I}\{|x| \geqslant t\}$, so that $\psi_t(x)$ equals x^2 for small enough $|x|$ and is asymptotic to $2t|x|$ as $|x| \to \infty$. Thus, the "extreme" function ψ_t is in a sense intermediate between the absolute powers $|\cdot|$ and $|\cdot|^2$. So, by (1.6), all functions $f \in \mathcal{F}_{1,2}$ inherit such a property. This should explain the choice of the notation $\mathcal{F}_{1,2}$.

Classes of moment functions similar to $\mathcal{F}_{1,2}$ arise naturally in extremal problems in probability and statistics; see e.g. [14, 39, 28, 16, 4, 31, 30] and further references therein; $\mathcal{F}_{1,2}$ is especially similar to the class $\mathcal{O}_{2,3}$ considered in [16].

Let us now give some examples of functions f in $\mathcal{F}_{1,2}$. The "extreme" functions ψ_t have been already mentioned. Perhaps the most important members of the class $\mathcal{F}_{1,2}$ are the power functions $|\cdot|^p$ with $p \in (1,2]$. The function $|\cdot|$ is not in $\mathcal{F}_{1,2}$, since it is not in $C^1(\mathbb{R})$.

It is easy to construct many other kinds of examples of functions $f \in \mathcal{F}_{1,2}$ by (i) letting f'' be $\big($on $(0,\infty)\big)$ any function, say g, which is nonnegative, nonincreasing, right-continuous, and integrable on any interval of the form $(0,u]$, for any $u \in (0,\infty)$; then (ii) finding f on $[0,\infty)$ as the solution to the following initial value problem: $f(0) = f'(0) = 0$ and $f'' = g$ on $(0,\infty)$; and finally (iii) extending f to the entire real line $\mathbb{R}$ as an even function.

E.g., taking $g(x) = (1+x)^{p-2}$ for $p \in (1,2)$ and $x \in (0,\infty)$, one ends up with $f(x) = \frac{1}{p(p-1)}\,[(1+|x|)^p - 1 - p|x|]$ for all $x \in \mathbb{R}$, which is asymptotic to $\frac12 x^2$ as $x \to 0$ and to $\frac{1}{p(p-1)}\,|x|^p$ as $|x| \to \infty$; if the condition $p \in (1,2)$ is replaced here by $p \in (-\infty,0) \cup (0,1)$, then $f(x)$ is asymptotic to $\frac{|x|}{1-p}$ as $|x| \to \infty$. Similarly one can get $f(x) \equiv e^{-|x|} - 1 + |x|$ $\big($by starting with $g(x) = e^{-x}$ for $x \in (0,\infty)\big)$; $f(x) \equiv |x| - \ln(1+|x|)$ $\big($with $g(x) \equiv \frac{1}{(1+x)^2}\big)$; $f(x) \equiv |x|\ln(1+|x|)$ $\big($with $g(x) \equiv \frac{1}{1+x} + \frac{1}{(1+x)^2}\big)$.

Perhaps a more interesting example is the following family of functions, which are parabolic splines (and will also be used in Remark 1.5):

$$f_{\text{alt}}(x) := \frac{(|x| - x_j)^2}{2(x_j+1)^{2/3}} + \sum_{k=0}^{j-1} \frac{\big[|x| - \frac12(x_k + x_{k+1})\big]\,(x_{k+1} - x_k)}{(x_k+1)^{2/3}} \tag{1.10}$$

if $x_j \leqslant |x| < x_{j+1}$ and $j \in \overline{0,\infty}$, where $x_0 := 0$, x_1 is any positive real number, and $x_j := q^{2^{j-1}} - 1$ for $q := x_1 + 1$ and all $j \in \overline{2,\infty}$, so that $x_{j+1} + 1 = (x_j+1)^2$ for all $j = 1, 2, \dots$ (we use the standard conventions $a^{b^c} := a^{(b^c)}$ and $\sum_{k=0}^{-1} \dots := 0$).

It is easy to check that $f_{\text{alt}} \in \mathcal{F}_{1,2}$ and $f''_{\text{alt}}(x) = (x_j+1)^{-2/3} = (x_{j+1}+1)^{-1/3}$ if $x_j \leqslant |x| < x_{j+1}$ and $j \in \overline{0,\infty}$, so that the function f''_{alt} alternates between the powers $(|\cdot|+1)^{-2/3}$ and $(|\cdot|+1)^{-1/3}$, as shown in the left panel of Figure 1. So, one might expect that the function f_{alt} alternates (far away from 0) between something like the powers $|\cdot|^{-2/3+2} = |\cdot|^{4/3}$ and $|\cdot|^{-1/3+2} = |\cdot|^{5/3}$. This expectation is only partially justified.

Indeed, introduce the (local) "effective" exponent of the function f_{alt} at a point $x \in \mathbb{R} \setminus \{0\}$ by the formula

$$p_{\text{eff}}(x) := \log_{|x|} f_{\text{alt}}(x), \quad \text{so that} \quad f_{\text{alt}}(x) = |x|^{p_{\text{eff}}(x)}.$$

The following proposition shows that the effective exponent p_{eff} eventually, "in the limit", alternates between $\frac{3}{2}$ (rather than the expected $\frac{4}{3}$) and $\frac{5}{3}$. In this sense, one might say that f''_{alt} stays closer to $(|\cdot|+1)^{-1/3}$ than to $(|\cdot|+1)^{-2/3}$, "most of the time".

Proposition 1.3.

(i) $p_{\text{eff}}(x) = \tilde{p}_{\text{eff}}\big(\rho(x)\big) + o(1)$ *as* $x \to \infty$, *where* $\tilde{p}_{\text{eff}}(r) := (2 - \frac{2}{3r}) \vee (1 + \frac{2}{3r})$ *and* $\rho(x) := 2^{1-j} \log_q(x+1)$ *for* $x \in (x_j, x_{j+1}]$.

(ii) For each $j \in \overline{1,\infty}$, *the function* ρ *increases from* 1 *to* 2 *on the interval* $(x_j, x_{j+1}]$.

(iii) For each $j \in \overline{1,\infty}$, *the approximate effective exponent* $\tilde{p}_{\text{eff}}\big(\rho(x)\big)$ *decreases from* $\frac{5}{3}$ *to* $\frac{3}{2}$ *and then increases back to* $\frac{5}{3}$ *as* $x+1$ *increases from* x_j+1 *to* $(x_j+1)^{4/3}$ *and then on to* $x_{j+1}+1 = (x_j+1)^2$, *respectively.*

Part of the graph of the (exact) effective exponent p_{eff} (with $x_1 = \frac{1}{10}$) is shown in the right panel of Figure 1. Recall that the x_j's grow very fast in j for large j. Therefore, for better presentation, the horizontal axis in the right panel is nonlinearly rescaled so that the x_j's appear equally spaced. Namely, what is actually shown here is part of the graph $\{\big(\log_2 \log_q(x+1), p_{\text{eff}}(x)\big) : x > x_1\}$; note that $\log_2 \log_q(x_j+1) = j-1$ for all $j \in \overline{1,\infty}$.

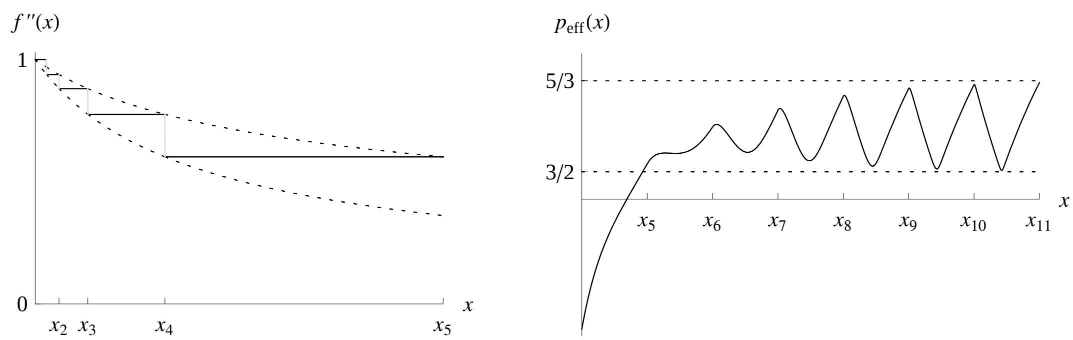


FIGURE 1. Left panel: f'' (solid) for $f = f_{\text{alt}}$ alternates between $(|\cdot|+1)^{-2/3}$ (dotted) and $(|\cdot|+1)^{-1/3}$ (dotted). Right panel: the effective exponent p_{eff} (solid) for $f = f_{\text{alt}}$ eventually alternates between $\frac{3}{2}$ (dotted) and $\frac{5}{3}$ (dotted).

1.2.2. *On the best possible constant* C_f *in general and, in particular, for the power and extreme functions.* The following proposition concerns some general properties of the constant factor C_f for nonzero f in $\mathcal{F}_{1,2}$ except for $f = \psi_\infty$; in the latter, trivial case, one has $C_f = 1$, as also stated in Proposition 1.6; recall that $\psi_\infty(x) = x^2$ for all $x \in \mathbb{R}$.

Proposition 1.4. *Take any $f \in \mathcal{F}_{1,2} \setminus \{0, \psi_\infty\}$. Let $s_f := \inf \operatorname{supp} \gamma$, where $\operatorname{supp} \gamma$ stands for the support of the measure $\gamma = \gamma_f$ defined in Proposition 1.2. Recall the definition of $L_{f;s}(x)$ in (1.4). Then the following statements hold.*

(i) $s_f \in [0, \infty)$.
(ii) *For any $s \in (0, s_f]$, one has $L_{f;s}(x) = f(s)$ for all $x \in (0, s)$.*
(iii) *For any $s \in (s_f, \infty)$, one has $L'_{f;s}(0+) > 0$ and $L'_{f;s}(s-) < 0$.*
(iv) *For any $s \in (0, \infty)$, there is some (not necessarily unique) $x_{f;s} \in (0, s)$ such that $L_{f;s}(x)$ is nondecreasing in $x \in (0, x_{f;s}]$ and nonincreasing in $x \in [x_{f;s}, s)$.*
(v) *One has*

$$\begin{aligned} C_f &= \sup_{s \in (s_f, \infty)} \left[\tfrac{1}{f(s)} \max_{x \in (0,s)} L_{f;s}(x) \right] \\ &= \sup_{s \in (s_f, \infty)} \left[\tfrac{1}{f(s)} L_{f;s}(x_{f;s}) \right] > 1. \end{aligned}$$

Remark 1.5. Proposition 1.4 provides for an effective maximization of $L_{f;s}(x)$ in $x \in (0, s)$, for any given $s \in (0, \infty)$, so that $\mathcal{L}_f(s) := \frac{1}{f(s)} \max_{x \in (0,s)} L_{f;s}(x) = \frac{1}{f(s)} L_{f;s}(x_{f;s})$ can be effectively found. In the important special case when f is a power function $|\cdot|^p$ $\big($with $p \in (1, 2]\big)$, one can also use the homogeneity of f in order to compute the constant C_f quite effectively, as described in Proposition 1.8. However, in general it remains to maximize $\mathcal{L}_f(s)$ in $s \in (s_f, \infty)$. It appears that usually $\mathcal{L}_f(s)$ is monotonically nondecreasing in s, if the function f is not too irregular; one "exceptional" function f for which $\mathcal{L}_f$ lacks such a monotonicity property is a function f_{alt} of the "alternating" family described by formula (1.10). Indeed, take $f = f_{\text{alt}}$ with $x_1 = \frac{1}{5}$. Then $\mathcal{L}(\frac{107}{100}) < \mathcal{L}(\frac{106}{100})$. One may still ask whether it is true for all $f \in \mathcal{F}_{1,2}$ that the limit $\mathcal{L}_f(\infty-)$ exists, and if so, whether it is true that $\mathcal{L}_f(s) \leqslant \mathcal{L}_f(\infty-)$ for all $s \in (s_f, \infty)$, so that C_f be found as $\mathcal{L}_f(\infty-)$. In any case, Theorem 1.1 reduces the problem of finding the optimal constant C in (1.2) to a maximization just in two real variables, s and x, which should not usually be too difficult.

Now let us provide a simple description of the constant C_f in the case when f is an "extreme" function ψ_t, representing the extreme rays of the convex cone $\mathcal{F}_{1,2}$:

Proposition 1.6. *One has $C_{\psi_t} = 2$ for each $t \in (0, \infty)$, whereas $C_{\psi_\infty} = 1$.*

Remark 1.7. Proposition 1.6 might seem quite surprising: whereas, by Theorem 1.1, the range of the values of C_f over all nonzero f in the convex cone $\mathcal{F}_{1,2}$ is the entire interval $[1, 2]$, the only value that C_f takes on all the extreme rays $\mathbb{R}_+ \psi_t$ $\big($which span the cone $\mathcal{F}_{1,2}$ in the sense of (1.6)$\big)$ is 2. This suggests strong nonlinearity of the optimal constant factor C_f in f. However, as seen from the proof of Proposition 1.6, the fact that C_{ψ_t} is the same for all $t \in (0, \infty)$ is due to a simple homogeneity property. Note also the discontinuity of C_{ψ_t} in t at $t = \infty$.

As mentioned earlier, for any $p \in (1, 2]$ the power function $|\cdot|^p$ belongs to the class $\mathcal{F}_{1,2}$; for such p, consider the corresponding constant factor

$$\tilde{C}_p := C_{|\cdot|^p},$$

so that for any v-martingale $(S_j)_{j=1}^n$

$$\mathsf{E}\,|S_n|^p \leqslant \mathsf{E}\,|X_1|^p + \tilde{C}_p \sum_{j=2}^n \mathsf{E}\,|X_i|^p. \tag{1.11}$$

Note that $|\cdot|^2 = \psi_\infty$, so that, by Proposition 1.6,

$$\tilde{C}_2 = 1. \tag{1.12}$$

Proposition 1.8.

(i) *For any* $p \in (1,2)$

$$\tilde{C}_p = \ell(p, x_p) = \max_{x\in(0,1)} \ell(p,x),$$

where

$$\ell(p,x) := L_{|\cdot|^p;1}(x) = (1-x)^p - x^p + px^{p-1} \tag{1.13}$$

for $x \in (0,1)$*, and* x_p *is the only root* $x \in (0,1)$ *of the equation*

$$(1-x)^{p-1} + x^{p-1} = (p-1)x^{p-2}. \tag{1.14}$$

Moreover, $\ell(p,x)$ *is increasing in* $x \in (0,x_p)$ *and decreasing in* $x \in (x_p,1)$*, for each* $p \in (1,2)$*.*

(ii) *In fact,* $x_p \in (\frac{p-1}{5}, \frac{p-1}{2}) \subset (0, \frac{1}{2})$ *for all* $p \in (1,2)$*.*

(iii) *Further,* $\tilde{C}_p$ *is continuously (and strictly) decreasing in* $p \in (1,2]$ *from* $\tilde{C}_{1+} = 2$ *to* $\tilde{C}_2 = 1$*; furthermore,* $\tilde{C}_p$ *is real-analytic in* $p \in (1,2)$*.*

(iv) *The values* $\tilde{C}_p$ *are algebraic for all rational* $p \in (1,2]$*; in particular,* $\tilde{C}_{3/2} = \sqrt{1+\frac{1}{\sqrt{2}}} = 1.306\ldots$ *(with* $x_{3/2} = \frac{1}{4}\big(2-\sqrt{2}\big) = 0.146\ldots$*).*

(v) *Explicit upper and lower bounds on* $\tilde{C}_p$ *are given by the inequalities*

$$\tilde{C}_p^{-,1} \vee \tilde{C}_p^{-,2} < \tilde{C}_p < \tilde{C}_p^{+,1} \wedge \tilde{C}_p^{+,2} \leqslant \tilde{C}_p^{+,2} < W_p \tag{1.15}$$

for all $p \in (1,2)$*, where*

$$\begin{aligned}
\tilde{C}_p^{-,1} :=& 2^{-p}\big((3-p)^p + (p-1)^{p-1}(p+1)\big),\\
\tilde{C}_p^{-,2} :=& 5^{-p}\big((6-p)^p + (p-1)^{p-1}(4p+1)\big),\\
\tilde{C}_p^{+,1} :=& \tfrac{2^{-p}}{50(3-p)}\big((p-1)^{p-1}(150+181p-152p^2+21p^3)\\
&\qquad + (3-p)^{p-1}(450-381p+152p^2-21p^3)\big),\\
\tilde{C}_p^{+,2} :=& \tfrac{5^{-p}}{8(6-p)}\big(4(p-1)^{p-1}(12-35p+94p^2-21p^3)\\
&\qquad + (6-p)^{p-1}(288-15p-94p^2+21p^3)\big),\\
W_p :=& 2^{2-p}.
\end{aligned}$$

The upper bound W_p *on* $\tilde{C}_p$ *is exact at the endpoints of the interval* $(1,2)$ *in the sense that* $\tilde{C}_{1+} = W_{1+}$ *and* $\tilde{C}_2 = \tilde{C}_{2-} = W_{2-} = W_2$*; each of the bounds* $\tilde{C}_p^{-,1}$*,* $\tilde{C}_p^{-,2}$*,* $\tilde{C}_p^{+,1}$*, and* $\tilde{C}_p^{+,2}$ *is also exact in the similar sense.*

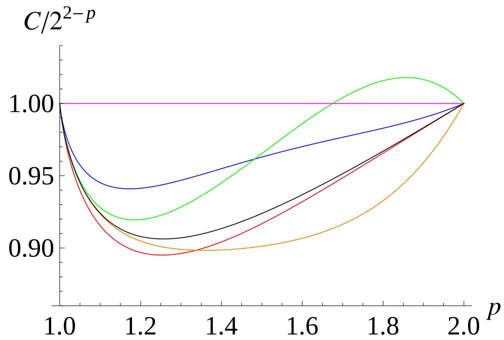


FIGURE 2. The ratios of $\tilde C_p$ (black), $\tilde C_p^{-,1}$ (red), $\tilde C_p^{-,2}$ (orange), $\tilde C_p^{+,1}$ (green), $\tilde C_p^{+,2}$ (blue), and W_p (magenta) to 2^{2-p}.

The graphs of the ratios of $\tilde C_p$, $\tilde C_p^{-,1}$, $\tilde C_p^{-,2}$, $\tilde C_p^{+,1}$, $\tilde C_p^{+,2}$, and W_p to $W_p = 2^{2-p}$ are shown in Figure 2. The graph of $\tilde C_p$, in comparison with W_p and the von Bahr–Esseen constant C_p^{vBE}, is presented in Figure 3.

As mentioned in Subsubsection 1.2.1, the absolute-value function $|\cdot|$ is is not in the class $\mathcal{F}_{1,2}$. However, by (1.7), $|\cdot|$ is in the closure of $\mathcal{F}_{1,2}$ with respect to the uniform convergence on $\mathbb{R}$. It is also clear that inequality (1.2) holds for $f = |\cdot|$ (and any r.v.'s $X_1, \dots, X_n$) with $C = \tilde C_1 := 1$. From this viewpoint, there is a discontinuity of C_p at $p = 1$, namely, $\tilde C_{1+} = 2 \ne 1 = \tilde C_1$.

1.2.3. *Application: concentration inequalities for separately Lipschitz functions on product spaces.* Let $X_1, \dots, X_n$ be independent r.v.'s with values in measurable spaces $\mathfrak{X}_1, \dots, \mathfrak{X}_n$, respectively. Let $g\colon \mathfrak{P} \to \mathbb{R}$ be a measurable function on the product space $\mathfrak{P} := \mathfrak{X}_1 \times \dots \times \mathfrak{X}_n$. Let us say (cf. [5, 30]) that g is *separately Lipschitz* if it satisfies a Lipschitz type condition in each of its arguments:

$$|g(x_1, \dots, x_{i-1}, \tilde x_i, x_{i+1}, \dots, x_n) - g(x_1, \dots, x_n)| \leqslant \rho_i(\tilde x_i, x_i) \tag{1.16}$$

for some measurable functions $\rho_i\colon \mathfrak{X}_i \times \mathfrak{X}_i \to \mathbb{R}$ and all $i \in \overline{1,n}$, $(x_1, \dots, x_n) \in \mathfrak{P}$, and $\tilde x_i \in \mathfrak{X}_i$.

Take now any separately Lipschitz function g and let

$$Y := g(X_1, \dots, X_n).$$

Suppose that the r.v. Y has a finite mean. Then one has the following.

Corollary 1.9. *For each $i \in \overline{1,n}$, take any $x_i \in \mathfrak{X}_i$.*

(I) For any $f \in \mathcal{F}_{1,2} \setminus \{0\}$

$$\mathsf{E}\, f(Y) \leqslant f(\mathsf{E}\, Y) + \kappa_f C_f \sum_{i=1}^n \mathsf{E}\, f\big(\rho_i(X_i, x_i)\big), \tag{1.17}$$

where

$$\kappa_f := \sup\Big\{\frac{U_f(c,s,0)}{U_f(c,s,a)}\colon s \in (0,\infty),\ c \in (0, \tfrac{s}{2}),\ a \in (0,c)\Big\} \in [1,2], \tag{1.18}$$

$$U_f(c,s,a) := cf(s-c+a) + (s-c)f(a-c) \tag{1.19}$$

$\big($the above definition of κ_f is valid, because $f > 0$ on $\mathbb{R} \setminus \{0\}$ and hence $U_f(c,s,a) > 0$ for any $s \in (0,\infty)$, $c \in (0, \frac{s}{2})$, and $a \in (0,c)\big)$.

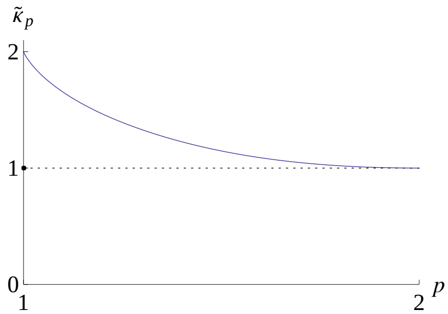
(II) For any $p \in (1,2]$

$$\mathsf{E}\, |Y|^p \leqslant |\,\mathsf{E}\, Y|^p + \tilde\kappa_p \tilde C_p \sum_{i=1}^n \mathsf{E}\, \big|\rho_i(X_i, x_i)\big|^p, \tag{1.20}$$

where

$$\tilde{\kappa}_p := \kappa_{|\cdot|^p} = \max_{c\in[0,1/2]} \Big[\big(c^{p-1} + (1-c)^{p-1}\big)\big(c^{\frac{1}{p-1}} + (1-c)^{\frac{1}{p-1}}\big)^{p-1}\Big]. \tag{1.21}$$

Moreover, $\tilde{\kappa}_p$ *continuously and strictly decreases in* $p \in (1,2]$ *from* 2 *to* 1*. Furthermore, the values of* $\tilde{\kappa}_p$ *are algebraic for all rational* $p \in (1,2]$*; in particular,* $\tilde{\kappa}_{3/2} = \frac{1}{9}\sqrt{51+21\sqrt{7}} = 1.14\dots$*, corresponding to* $c = \frac{1}{6}(3 - \sqrt{1+2\sqrt{7}}) = 0.081\dots$ *in* (1.21)*. The graph of* $\tilde{\kappa}_p$ *is shown below.*



$\tilde{\kappa}_p$, solid; 1, dotted.

One can observe some similarity between $C_f, \tilde{C}_p$ and $\kappa_f, \tilde{\kappa}_p$. Thus, going from the "one-dimensional" inequality (1.2) or (1.11) for v-martingales to the "multi-dimensional" measure concentration inequality (1.17) or (1.20) entails an extra factor, κ_f or $\tilde{\kappa}_p$, whose values are between 1 and 2.

The proof of Corollary 1.9 is partly based on the following proposition, which may be of independent interest.

Proposition 1.10. *For any zero-mean r.v.* X*,* $f \in \mathcal{F}_{1,2} \setminus \{0\}$*, and* $a \in \mathbb{R}$

$$\mathsf{E}\, f(X) \leqslant \kappa\, \mathsf{E}\, f(X+a) \tag{1.22}$$

with $\kappa = \kappa_f$*, and* κ_f *is the best possible constant* κ *in* (1.22)*.*

In turn, the proof of Proposition 1.10 uses

Proposition 1.11. *Take any* $f \in \mathcal{F}_{1,2} \setminus \{0\}$*,* $s \in (0,\infty)$*, and* $c \in (0, \frac{s}{2})$*. Then* $U_f(c,s,a)$ *(defined in* (1.19)*) is convex in* $a \in \mathbb{R}$*. Moreover,* $U_f(c,s,a)$ *attains its minimum over all* $a \in \mathbb{R}$ *at a unique point* $a_{f;c,s} \in [0,c)$*. In particular, for all* $t \in (0,\infty)$*,* $s \in (0,\infty)$*, and* $c \in (0,\frac{s}{2})$

$$a_{\psi_t;c,s} = \tfrac{c}{s-c}\,(s-c-t)_+ \tag{1.23}$$

and $\kappa_{\psi_t} = 2$*.*

On the other hand, Proposition 1.11 obviously complements Corollary 1.9.

A difficulty in proving the uniqueness of the minimizer of $U_f(c,s,a)$ in a in Proposition 1.11 is that, in general, $U_f(c,s,a)$ is not strictly convex in a.

An example of separately Lipschitz functions $g : \mathfrak{X}^n \to \mathbb{R}$ is given by the formula $g(x_1,\dots,x_n) = \|x_1 + \dots + x_n\|$ for all $x_1,\dots,x_n$ in a separable Banach space $(\mathfrak{X}, \|\cdot\|)$. In this case, one may take $\rho_i(\tilde{x}_i, x_i) \equiv \|\tilde{x}_i - x_i\|$. Thus, one obtains

Corollary 1.12. *Let* $X_1,\dots,X_n$ *be independent random vectors in the Banach space* $(\mathfrak{X},\|\cdot\|)$*. Let* $S_n := X_1 + \dots + X_n$*. For each* $i \in \overline{1,n}$*, take any* $x_i \in \mathfrak{X}$*.*

Then for any $f \in \mathcal{F}_{1,2} \setminus \{0\}$

$$\mathsf{E}\, f(\|S_n\|) \leqslant f(\mathsf{E}\,\|S_n\|) + \kappa_f C_f \sum_{i=1}^n \mathsf{E}\, f\big(\|X_i - x_i\|\big).$$

Moreover, for any $p \in (1, 2]$

$$\mathsf{E}\,\|S_n\|^p \leqslant (\mathsf{E}\,\|S_n\|)^p + \kappa_p \tilde{C}_p \sum_{i=1}^n \mathsf{E}\,\|X_i - x_i\|^p. \tag{1.24}$$

For $p = 2$, inequality (1.24) was obtained in [36, Theorem 4], based on an improvement the method of Yurinskiĭ(1974) [19]. The proof of Corollary 1.9 is based in part on the same kind of improvement.

As can be seen from that proof, both Corollaries 1.9 and 1.12 will hold even if the separately-Lipschitz condition (1.16) is relaxed to

$$|\,\mathsf{E}\, g(x_1, \dots, x_{i-1}, \tilde{x}_i, X_{i+1}, \dots, X_n) - \mathsf{E}\, g(x_1, \dots, x_i, X_{i+1}, \dots, X_n)| \leqslant \rho_i(\tilde{x}_i, x_i). \tag{1.25}$$

Note also that in Corollaries 1.9 and 1.12 the r.v.'s X_i do not have to be zero-mean, or even to have any definable mean; at that, the arbitrarily chosen x_i's may act as the centers, in some sense, of the distributions of the corresponding X_i's.

Clearly, the separate-Lipschitz (sep-Lip) condition (1.16) is easier to check than a joint-Lipschitz one. Also, sep-Lip (especially in the relaxed form (1.25)) is more generally applicable. On the other hand, when a joint-Lipschitz condition is satisfied, one can generally obtain better bounds. Literature on the concentration of measure phenomenon, almost all of it for joint-Lipschitz settings, is vast; let us mention here only [24, 22, 21, 6, 23].

1.2.4. *Other corollaries of Theorem 1.1 and comparisons with known results.* Take any $p \in (1, 2]$. A normed space $(\mathfrak{X}, \|\cdot\|)$ (or, briefly, $\mathfrak{X}$) is called p-uniformly smooth [3] if for some constant $D \in (0, \infty)$ (referred to as a p-uniform smoothness constant of $\mathfrak{X}$) and all x and y in $\mathfrak{X}$ one has $\frac{1}{2}\left(\|x+y\|^p + \|x-y\|^p\right) \leqslant \|x\|^p + D^p \|y\|^p$ or, equivalently,

$$\mathsf{E}\,\|x + Xy\|^p \leqslant \|x\|^p + D^p\, \mathsf{E}\,|X|^p \|y\|^p \tag{1.26}$$

for all symmetric(ally distributed) real-valued r.v. X. If $\mathfrak{X}$ is p-uniformly smooth with a p-uniform smoothness constant D, let us say that $\mathfrak{X}$ is (p, D)-uniformly smooth or, simply, (p, D)-smooth. For instance, for any $q \in [2, \infty)$ the space $L^q(\mu)$ is $(2, D)$-smooth with $D = \sqrt{q-1}$, which is the best possible constant of the 2-uniform smoothness as long as the space $L^q(\mu)$ is at least two-dimensional — see e.g. [29, Proposition 2.1], [3, Proposition 3], and [13, Corollary 2.8].

Dual to the notion of (p, D)-uniform smoothness is that of (q, D^{-1})-uniform convexity, whose definition can be obtained by reversing the inequality sign in (1.26) and replacing there p and D by q and D^{-1}, respectively; here, $\frac{1}{p} + \frac{1}{q} = 1$. In particular, a result due to Ball, Carlen, and Lieb [3, Lemma 5] is that $\mathfrak{X}$ is (p, D)-uniformly smooth iff its dual $\mathfrak{X}^*$ is (q, D^{-1})-uniformly convex; cf. e.g. [11, 25]. Note that q-uniform convexity and p-uniform smoothness are refinements of the notions of uniform convexity and uniform smoothness, which go back to Clarkson

[9] and Day [11]; cf. [18, 41]. These notions are important in functional analysis. In particular, Pisier [37] showed that every super-reflexive space is q-uniformly convex and p-uniformly smooth for some q and some p; an earlier result due to Enflo [15] stated that $\mathfrak{X}$ is super-reflexive iff it is isomorphic to a uniformly convex space. Among many other results, Pisier [37] also showed that the super-reflexivity is equivalent to the super-Radon-Nikodym property. Applications of the 2-uniform convexity/2-uniform smoothness to Finsler manifolds were given by Ohta [27].

It is clear that $\mathfrak{X}$ is (p, D)-smooth iff inequality (1.2) with $C = D^p$ and $f = \|\cdot\|^p$ holds for all martingales (or even v-martingales) $(S_j)_{j=1}^n$ with values in $\mathfrak{X}$ and conditionally symmetric differences $X_2, \dots, X_n$; by symmetrization, the same inequality will then hold without the conditional symmetry restriction, but with the worse constant $C = (2D)^p$ instead of $C = D^p$. These considerations suggest the following.

Let us say that the space $\mathfrak{X}$ is *completely* (p, D)*-smooth* if inequality (1.26) holds for all *zero-mean* real-valued r.v.'s X (and all x and y in $\mathfrak{X}$). It is clear that $\mathfrak{X}$ is completely (p, D)-smooth iff inequality (1.2) with $C = D^p$ and $f = \|\cdot\|^p$ holds for all martingales (or even v-martingales) $(S_j)_{j=1}^n$ with values in $\mathfrak{X}$. Also, Proposition 1.8 immediately implies

Corollary 1.13. *Take any $p \in (1, 2]$ and any measure μ on any measurable space. Then the space $L^p(\mu)$ is completely (p, D)-smooth with the best possible constant $D = \tilde{C}_p^{1/p}$. So, for any $n \in \overline{2, \infty}$ and v-martingale $(S_j)_{j=1}^n$ with values in $L^p(\mu)$,*

$$\mathsf{E}\,\|S_n\|_p^p \leqslant \mathsf{E}\,\|X_1\|_p^p + \tilde{C}_p \sum_{j=2}^n \mathsf{E}\,\|X_j\|_p^p$$

(cf. (1.24)).

The above discussion suggests that the form of inequality (1.2) is rather natural in such contexts as concentration of measure, uniform smoothness, and martingales (or v-martingales). Yet, in the case when the differences $X_1, \dots, X_n$ are independent real-valued zero-mean r.v.'s, the form of the following immediate corollary of Theorem 1.1 may be more relevant.

Corollary 1.14. *For any $f \in \mathcal{F}_{1,2}$, $n \in \overline{2, \infty}$, and (real-valued) v-martingale $(S_j)_{j=1}^n$,*

$$\mathsf{E}\, f(S_n) \leqslant K \sum_{j=1}^n \mathsf{E}\, f(X_j) \tag{1.27}$$

with $K = C_f$.

However, in inequality (1.27) the constant factor $K = C_f$ is no longer the best possible one, at least for independent zero-mean X_j's. One way to reduce the constant is as follows. In the conditions of Corollary 1.14, rewrite the right-hand side of (1.2) with $C = C_f$ as $C_f \sum_{j=1}^n \mathsf{E}\, f(X_j) - (C_f - 1)\,\mathsf{E}\, f(X_1)$. Then, assuming that $\mathsf{E}\, f(X_1) \geqslant \frac{\lambda}{n} \sum_{j=1}^n \mathsf{E}\, f(X_j)$ for some $\lambda \in (0, \infty)$, one sees that the constant factor $K = C_f$ in (1.27) can be reduced by spreading the "excess" $C_f - 1 \geqslant 0$ over all the summands $\mathsf{E}\, f(X_1), \dots, \mathsf{E}\, f(X_n)$, to get (1.27) with

$$K = C_f - \tfrac{\lambda}{n}\,(C_f - 1) \leqslant C_f. \tag{1.28}$$

To develop this simple observation a bit further, let us take any $\lambda \in (0, \infty)$ and say that a sequence $(S_j)_{j=1}^n$ is a λ-*good rearranged-v-martingale* if there are (i) some $i \in \overline{1,n}$ such that $\mathsf{E}\, f(X_i) \geqslant \frac{\lambda}{n} \sum_{j=1}^n \mathsf{E}\, f(X_j)$ and (ii) a permutation $(j_1, \dots, j_{n-1})$ of the set $\overline{1,n} \setminus \{i\}$ such that $(X_i, X_{j_1}, \dots, X_{j_{n-1}})$ is the difference sequence of a v-martingale. Note that, if the differences $X_1, \dots, X_n$ of a sequence $(S_j)_{j=1}^n$ are independent zero-mean r.v.'s, then $(S_j)_{j=1}^n$ is a 1-good rearranged-v-martingale. (In general, a λ-good rearranged-v-martingale does not have to be a v-martingale.) Thus, one obtains

Corollary 1.15. *For any $f \in \mathcal{F}_{1,2} \setminus \{0\}$, $n \in \overline{2, \infty}$, and λ-good rearranged-v-martingale $(S_j)_{j=1}^n$, inequality (1.27) holds, again with K as in (1.28).*

In the special case of the power functions $|\cdot|^p$ $\big($with $p \in (1,2)\big)$ in place of general $f \in \mathcal{F}_{1,2} \setminus \{0\}$, an inequality of the form (1.27) was obtained by von Bahr and Esseen (vBE) [2]:

$$\mathsf{E}\,|S_n|^p \leqslant K \sum_{j=1}^n \mathsf{E}\,|X_j|^p, \tag{1.29}$$

with the constant factor $K = 2 - \frac{1}{n} = 2 - \frac{1}{n}(2-1)$, which, by part (iii) of Proposition 1.8, is greater than the K in (1.28), again for $f = |\cdot|^p$ with $p \in (1,2)$. The vBE inequality (1.29) has been used in various kinds of studies.

As noted by vBE [2], the special case of inequality (1.29) (with $K = 1$) when the conditional distributions of the differences X_i given S_{i-1} are symmetric for all $i \in \overline{2,n}$ easily follows from Clarkson's inequality [9]

$$|x+y|^p + |x-y|^p \leqslant 2|x|^p + 2|y|^p \tag{1.30}$$

for all real x and y and all $p \in [1,2]$. $\big($As pointed out in [9], inequality (1.30) obviously implies that L^p is uniformly smooth, and in fact p-uniformly smooth.$\big)$ Actually, it is easy to see that Clarkson's inequality (1.30) is equivalent to the symmetric case of (1.29), with $K = 1$.

As mentioned in [2], an inequality of the form (1.29) is not of optimal order in n for independent identically distributed real-valued zero-mean X_i's and may be used together with a Hölder bound such as $\mathsf{E}\,|S_n|^p \leqslant (\mathsf{E}\, S_n^2)^{p/2}$. Using similar considerations together with symmetrization and truncation, Manstavichyus [26] obtained bounds on $\mathsf{E}\,|S_n|^p$ from above and below, which differ from each other by an (unspecified) factor depending only on p. The proof of Theorem 1.1 (and especially that of part (II) of Lemma 2.5) shows that near-extremal r.v.'s $X_1, \dots, X_n$, for which the constant C in (1.2) cannot be non-negligibly less than C_f, are as follows: X_1 and X_2 are independent, zero-mean, and both highly skewed in the same direction (both to the right or both to the left); $|X_2|$ is much smaller than $|X_1|$; and $X_3, \dots, X_n$ are zero or nearly so. This suggests that the inequality (1.29) should be most useful for independent real-valued zero-mean X_i's when the distributions of the X_i's are quite different from one another and/or highly skewed and/or heavy-tailed.

Again in the case when the differences $X_1, \dots, X_n$ are independent zero-mean r.v.'s, von Bahr and Esseen [2] made an effort to improve their constant $K = 2 - \frac{1}{n}$ in (1.29). For such X_i's and the values of p in a left neighborhood of 2 such that $D(p) := \frac{13.52}{\pi(2.6)^p}\Gamma(p)\sin\frac{\pi p}{2} = \frac{2}{\pi}(\frac{13}{5})^{2-p}\Gamma(p)\sin\frac{\pi p}{2} < 1$, they showed that (1.29) holds with $K = C_p^{\mathrm{vBE}} := \frac{1}{\big(1-D(p)\big)_+}$, assuming the convention $\frac{1}{0} := \infty$; in fact, the constant factor C_p^{vBE} may improve on (i.e., may be less than) the factor $2 - \frac{1}{n}$ only for values of p in a left neighborhood of 2 such that $D(p) < \frac{1}{2}$. It is stated (without proof) in [2] that $D(p)$ decreases in $p \in (1, 2)$ and that the mentioned left neighborhood contains the interval $[1.6, 2]$; cf. Figure 3, where the von Bahr–Esseen constant factor $2 \wedge C_p^{\mathrm{vBE}}$ is compared with the optimal (for (1.2)) constant factor $\tilde{C}_p$. (There are a couple of typos in [2]: in [2, (11)], one should have $r(2.6)^r$ instead of $(r2.6)^r$, and also the expression [2, (12)] for $D(p)$ should have $\pi(2.6)^r$ instead of $(\pi 2.6)^r$.)

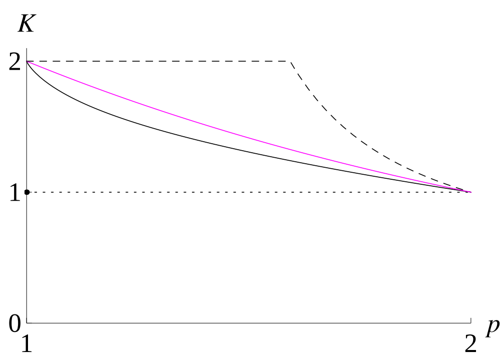


FIGURE 3. $\tilde{C}_p$, solid; W_p, magenta; $2 \wedge C_p^{\mathrm{vBE}}$, dashed; 1, dotted.

The method of [2] is based on a representation of the absolute moment $\mathsf{E}\,|X|^p$ of a r.v. X as a certain integral transform of the Fourier transform of the distribution of X. More general representations, for the positive-part moments $\mathsf{E}\,X_+^p$, were obtained in [7, 33].

Take now again any $p \in (1, 2]$. Woyczyński [40] considered the class $\mathcal{G}_{p-1}$ of Banach spaces $\mathfrak{X}$ defined by the following condition: there exist a map $G\colon \mathfrak{X} \to \mathfrak{X}^*$ and a constant $A = A_{p,\mathfrak{X}} \in (0, \infty)$ such that for all x and y in $\mathfrak{X}$ one has (i) $\|G(x)\| = \|x\|^{p-1}$, (ii) $G(x)x = \|x\|^p$, and (iii) $\|G(x) - G(y)\| \leqslant A\|x - y\|^{p-1}$. The class $\mathcal{G}_1$ was introduced by Fortet and Mourier [17]. Hoffmann-Jørgensen [18] proved that $\mathfrak{X} \in \mathcal{G}_{p-1}$ iff $\mathfrak{X}$ is p-uniformly smooth.

Woyczyński [40] showed that inequality (1.29) holds for any independent zero-mean random vectors $X_1, \dots, X_n$ in any Banach space $\mathfrak{X} \in \mathcal{G}_{p-1}$, with $|\cdot|$ and K replaced by $\|\cdot\|$ and $A_{p,\mathfrak{X}}$. As noted in [40], the space L^p is in $\mathcal{G}_{p-1}$, with the constant $A = 2$; at that, one should take $G(x) = x^{[p-1]} := |x|^{p-1}\operatorname{sign} x \in L^q = (L^p)^*$ for all $x \in L^p$. It is not hard to see that the best possible constant $A = A_{p,\mathfrak{X}}$ for $\mathfrak{X} = L^p$ is

$$W_p := \sup_{u \in (-1,1)} \frac{1 - u^{[p-1]}}{(1-u)^{p-1}} = 2^{2-p},$$

which is in agreement with the definition of W_p in part (v) of Proposition 1.8. Thus, one has (1.29) with $K = W_p = 2^{2-p}$ for independent zero-mean differences $X_1, \dots, X_n$, which may be either real-valued or, equivalently, with values in L^p (in which case $|\cdot|$ is replaced by $\|\cdot\|_p$). The constant $K = W_p$ in (1.29) is not the best possible one, even for independent zero-mean real-valued $X_1, \dots, X_n$, even if n is not fixed; indeed, by part (v) of Proposition 1.8, $W_p > \tilde{C}_p$. On the other hand, the following proposition takes place.

Proposition 1.16. *One has $C_p^{\mathrm{vBE}} > W_p$ for all $p \in [1, 2)$.*

So, $C_p^{\mathrm{vBE}} > W_p > \tilde{C}_p$ for all $p \in (1, 2)$. This comparison is illustrated in Figure 3.

Topchii and Vatutin [38, Theorem 2] obtained inequality (1.27) with $K = 4$. Alsmeyer and Rösler [1] improved Topchii and Vatutin's constant factor $K = 4$ to $K = 2$. In fact, they showed that inequality (1.2) holds with $C = 2$ for all $f \in F_{1,2}$, and they also showed that the constant factor $C = 2$ is optimal over the entire class $F_{1,2}$ of functions. The main difference between Theorem 1.1 in the present paper and the result of [1] is that the factor C_f in Theorem 1.1 is optimal for *each* given moment function f in $F_{1,2}$, and one can see that C_f is strictly less than 2 for all $f \in F_{1,2}$ except f lying on the extreme rays $\mathbb{R}_+\psi_t$ of the convex cone $F_{1,2}$, discussed after the statement of Proposition 1.2. Another advantage of having the individualized optimal factor C_f is that it directly leads to the optimal factor $\tilde{C}_p$ for the pth absolute power moments, and the optimal $\tilde{C}_p$ is different for different values of p between 1 and 2; at that, by part (iii) of Proposition 1.8, $\tilde{C}_p$ is strictly less than 2 for all $p \in (1, 2]$. Recall also that the matter of effective calculation of the constant C_f for any given f in $F_{1,2}$ is addressed in Proposition 1.4. On the other hand, in view of (1.5), the result of [1] immediately follows from our Theorem 1.1.

Note also that a result very similar to the "only if" half of part (I) of our Proposition 1.2 was presented as Lemma 1 in [1]. However, the conclusion in that lemma that the "mixing" measure (denoted by Q_ϕ in [1] and by γ in Proposition 1.2 here) must be finite is mistaken. Indeed, as Proposition 1.2 shows, it is enough that $\int_{(0,\infty]}(t \wedge 1)\gamma(\,\mathrm{d}t) < \infty$. For instance, in the case when f is the absolute power function $|\cdot|^p$ with $p \in (1, 2)$, the corresponding mixing measure $\gamma = \gamma_f$ is given, according to (1.8), by the formula $\gamma\big((x, \infty]\big) = \frac{1}{2}\,p(p-1)x^{p-2}$ for $x \in (0, \infty)$, and so, γ_f is an infinite measure on $(0, \infty]$.

Kemperman and Smit [20] give an expression for the best possible constant factor $\tilde{C}_p$ in a version of the von Bahr–Esseen inequality for the absolute pth power moments. The paper [20] appears to be an abstract of a conference presentation, with only a brief description of the method of proof, which latter seems to consist in deriving, and then working with, a recursion formula. I have been unable to reconstruct the proof based on that description. It appears that the method in the present paper – based on the optimal inequality for moment functions f in the general class $F_{1,2}$ – is quite different from that in [20]. Also, the corresponding inequality in [20] contains the constant factor at the first summand, $\mathsf{E}\,|X_1|^p$, as well, and thus is a bit weaker than inequality (1.11) in the present paper; cf. (1.27) and (1.28).

Cox [10] gives a version of the von Bahr–Esseen inequality for the pth power moment of the norm of a martingale with values in a p-smooth Banach space (with $p \in (1, 2]$). Based on the mentioned result in Kemperman and Smit [20], it is shown in [10] that the constant factor in the inequality in [10] is optimal in Hilbert space case. In contrast, the focus in applications in the present paper is on a different kind of extension of the von Bahr–Esseen inequality, which is valid for all Banach spaces (smooth or not) and also, more generally, for the f-moments for all f in the class $F_{1,2}$, and at that for arbitrary separately Lipschitz functions on product spaces – instead of martingales; indeed, recall our Corollaries 1.91.12. For other, related results on martingales with values in a 2-smooth Banach spaces, one may see [29, 34].

2. Proofs

This section consists of four subsections. In Subsection 2.1, we shall prove 5 propositions, of the 8 ones stated in Section 1; three of these 5 propositions will be used in the proof of Theorem 1.1, in Subsection 2.4. The proof of Proposition 1.8 (which is also used in the proof of Theorem 1.1) is more involved than those of the other propositions, and it will be presented separately, in Subsection 2.2. Corollary 1.9 and the related Propositions 1.10 and 1.11 will be proved in Subsection 2.3.

2.1. Proofs of Propositions 1.2, 1.3, 1.4, 1.6, and 1.16.

Proof of Proposition 1.2. To begin, note that

$$\psi_t'(x) = 2(t \wedge x) \tag{2.1}$$

for all $x \in [0, \infty)$ and $t \in (0, \infty)$. Take any $f \in \mathcal{F}_{1,2}$. Then, by (1.1) and the right continuity of the monotonic right derivative f'' of f', the relation (1.8) defines a nonnegative Borel measure $\gamma = \gamma_f$ on $(0, \infty]$ and, by Fubini's theorem,

$$\begin{aligned} f'(x) &= \int_0^x f''(u)\,\mathrm{d}u = 2\int_0^x \mathrm{d}u \int_{(u,\infty]} \gamma(\,\mathrm{d}t) = 2\int_{(0,\infty]} \gamma(\,\mathrm{d}t) \int_0^{t\wedge x} \mathrm{d}u \\ &= 2\int_{(0,\infty]} (t \wedge x)\gamma(\,\mathrm{d}t) \end{aligned} \tag{2.2}$$

for all $x \in [0, \infty)$. In particular, this proves part (III) of the proposition and (taken with $x = 1$) implies the condition $\int_{(0,\infty]}(t \wedge 1)\gamma(\,\mathrm{d}t) < \infty$ in part (I) of the proposition. Further, for all $x \in [0, \infty)$ (2.2) yields

$$f(x) = \int_0^x f'(u)\,\mathrm{d}u = 2\int_0^x \mathrm{d}u \int_{(0,\infty]} (t \wedge u)\gamma(\,\mathrm{d}t) = 2\int_{(0,\infty]} \gamma(\,\mathrm{d}t) \int_0^x (t \wedge u)\,\mathrm{d}u,$$

which implies (1.6), since $\int_0^x (t \wedge u)\,\mathrm{d}u = \frac{1}{2}\,\psi_t(x)$ for all $x \in [0, \infty)$ and $t \in (0, \infty]$. This proves the "only if" implication in part (I) of the proposition, since the functions f and ψ_t are even.

To prove the "if" implication, assume that (1.6) holds for some nonnegative Borel measure γ on $(0, \infty]$ such that $\int_{(0,\infty]}(t \wedge 1)\gamma(\,\mathrm{d}t) < \infty$ and for all $x \in \mathbb{R}$. In view of (2.1), the condition $\int_{(0,\infty]}(t \wedge 1)\gamma(\,\mathrm{d}t) < \infty$ implies that the integral $\int_{(0,\infty]} \psi_t'(x)\gamma(\,\mathrm{d}t)$ converges uniformly over all x in any given compact subset of the interval $(0, \infty)$. So, one finds that (1.6) implies (1.9), which in turn implies that f' is nondecreasing and concave on $[0, \infty)$ $\big($because the function ψ_t is so, for each $t \in (0,\infty)\big)$. It is also easy to see that $f \in C^1(\mathbb{R})$, $f(0) = 0$, and f is even. Thus, it is checked that $f \in \mathcal{F}_{1,2}$, which completes the proof of the "if" implication in part (I) of the proposition.

It remains to prove part (II). Take indeed any $f \in \mathcal{F}_{1,2}$. Take also any nonnegative Borel measure γ on $(0,\infty]$ such that $\int_{(0,\infty]}(t\wedge 1)\gamma(\,dt) < \infty$ and (1.6) holds for all $x \in \mathbb{R}$. We have to show that (1.8) takes place for all $x \in (0,\infty)$. Take indeed any such x. Then, as has been shown, one has identities (1.9). Therefore, for any $h \in (0,\infty)$

$$\frac{1}{2}\,\frac{f'(x+h)-f'(x)}{h} = \int_{(0,\infty]} r_t(x,h)\gamma(\,dt), \tag{2.3}$$

where $r_t(x,h) := \frac{1}{h}\big[\big((x+h)\wedge t\big) - (x\wedge t)\big]$, which is bounded (between 0 and 1) and converges to $\mathbf{I}\{t > x\}$ as $h \downarrow 0$. So, (1.8) follows from (2.3) by dominated convergence. This completes the proof of part (II) of the proposition as well. $\square$

Proof of Proposition 1.3. Part (ii) of the proposition is obvious on recalling that $x_j = q^{2^{j-1}} - 1$ for $j \in \overline{1,\infty}$. Note also that $\rho\big((x_j+1)^{4/3} - 1\big) = \frac{4}{3}$ for $j \in \overline{1,\infty}$. So, to prove then part (iii), it is enough to show that $\tilde p_{\text{eff}}(r)$ decreases from $\frac{5}{3}$ to $\frac{3}{2}$ and then increases back to $\frac{5}{3}$ as r increases from 1 to $\frac{4}{3}$ and then to 2, which follows because the expressions $2 - \frac{2}{3r}$ and $1 + \frac{2}{3r}$ are, respectively, increasing and decreasing in $r \in [1,2]$, and they are equal to each other at $r = \frac{4}{3}$.

It remains to prove part (i) of the proposition, which is equivalent to

$$f_{\text{alt}}(x) = x^{\tilde p_{\text{eff}}(r)+o(1)} \tag{2.4}$$

as $x \to \infty$, where $r := \rho(x) \in (1,2]$, so that $x = q^{r2^{j-1}} - 1$. In other words, it suffices to prove that the convergence (2.4) with $x = q^{r2^{j-1}} - 1$ takes place uniformly in $r \in (1,2]$ as $j \to \infty$. Assume indeed that $j \to \infty$ and $x = q^{r2^{j-1}} - 1$. Introduce $y_j := x_j + 1$, so that $y_j = q^{2^{j-1}}$ for $j = 1,2,\dots$. Then $x = y_j^{r+o(1)}$, and uniformly over all $k \in \{0,\dots,j-1\}$ one has $x - \frac{1}{2}(x_k + x_{k+1}) = x^{1+o(1)}$; moreover, if at that $k \to \infty$ then $x_{k+1} - x_k = x_{k+1}^{1+o(1)} = y_k^{2+o(1)}$, which shows that the kth summand in the sum $\sum_{k=0}^{j-1}\dots$ in (1.10) is $(xy_k^{2-2/3})^{1+o(1)} = (y_j^r\, y_k^{4/3})^{1+o(1)}$ as $k \to \infty$. So, the sum $\sum_{k=0}^{j-1}\dots$ in (1.10) is $(y_j^r\, y_{j-1}^{4/3})^{1+o(1)} = (y_j^r\, y_j^{2/3})^{1+o(1)} = y_j^{r+\frac{2}{3}+o(1)}$.

To estimate the difference $x - x_j$, which appears on the right-hand side of (1.10), we need to distinguish two possible cases: $r \in [1,\frac{4}{3})$ and $r \in [\frac{4}{3},2]$. Uniformly over all $r \in [\frac{4}{3},2]$ one has $x - x_j = x^{1+o(1)} = y_j^{r+o(1)}$, so that the term on the right-hand side of (1.10) before the sum $\sum_{k=0}^{j-1}\dots$ is $y_j^{2r-\frac{2}{3}+o(1)}$, which yields $f_{\text{alt}}(x) = y_j^{2r-\frac{2}{3}+o(1)} + y_j^{r+\frac{2}{3}+o(1)} = y_j^{(2r-\frac{2}{3})\vee(r+\frac{2}{3})+o(1)} = y_j^{r\tilde p_{\text{eff}}(r)+o(1)} = x^{\tilde p_{\text{eff}}(r)+o(1)}$, as in (2.4).

It remains to consider the values $r \in [1,\frac{4}{3})$. For such values of r, the relation $x - x_j = x^{1+o(1)}$ no longer holds; for instance, $x - x_j = 0$ if $r = 1$. However, in this case one can obviously write $0 \leqslant x - x_j \leqslant x$ and also $\tilde p_{\text{eff}}(r) = 1 + \frac{2}{3r} > 2 - \frac{2}{3r}$. So, the term on the right-hand side of (1.10) before the sum $\sum_{k=0}^{j-1}\dots$ is $\leqslant y_j^{2r-\frac{2}{3}+o(1)} \leqslant$

$y_j^{r+\frac{2}{3}+o(1)}$, whereas still $\sum_{k=0}^{j-1}\dots = y_j^{r+\frac{2}{3}+o(1)}$; so, $y_j^{r+\frac{2}{3}+o(1)} \leqslant f_{\text{alt}}(x) \leqslant y_j^{r+\frac{2}{3}+o(1)} + y_j^{r+\frac{2}{3}+o(1)}$, whence $f_{\text{alt}}(x) = y_j^{r+\frac{2}{3}+o(1)} = y_j^{r\tilde{p}_{\text{eff}}(r)+o(1)} = x^{\tilde{p}_{\text{eff}}(r)+o(1)}$, thus proving (2.4) uniformly over all $r \in [1, \frac{4}{3})$ as well. □

Proof of Proposition 1.4.

(i) Since the function f is nonzero, the set $\operatorname{supp}\gamma$ is a nonempty subset of $(0,\infty]$. So, $s_f = \inf\operatorname{supp}\gamma \in [0,\infty]$. If $s_f = \infty$ then $\operatorname{supp}\gamma = \{\infty\}$, which implies, in view of (1.6), that $f = \psi_\infty$, which contradicts the assumption on f in Proposition 1.4. This proves part (i) of the proposition.

(ii) Take any $s \in (0, s_f]$ and $t \in \operatorname{supp}\gamma$, so that $t \in [s_f, \infty]$. Then $s_f > 0$ and it is straightforward to check that $L_{\psi_t;s}(x) = \psi_t(s)$ for any $x \in (0,s)$. Hence, by (1.6) and (1.9),

$$L_{f;s}(x) = \int_{(0,\infty]} L_{\psi_t;s}(x)\gamma(\,\mathrm{d}t) = \int_{(0,\infty]} \psi_t(s)\gamma(\,\mathrm{d}t) = f(s),$$

which proves part (ii) of Proposition 1.4.

(iii) Take any $s \in (s_f, \infty)$. Then $L'_{\psi_t;s}(0+) = 2(s-t)_+$ for any $t \in (0,\infty]$. So, by (1.9) and (1.8),

$$L'_{f;s}(0+) = \int_{(0,\infty]} L'_{\psi_t;s}(0+)\gamma(\,\mathrm{d}t) = 2\int_{(0,\infty]} (s-t)_+\gamma(\,\mathrm{d}t) > 0,$$

since for any $s \in (s_f,\infty)$ one has $\gamma\big((0,s)\big) > 0$. Similarly,

$$L'_{f;s}(s-) = \int_{(0,\infty]} L'_{\psi_t;s}(s-)\gamma(\,\mathrm{d}t) = -2\int_{(0,\infty]} t\,\mathbf{I}\{t<s\}\gamma(\,\mathrm{d}t) < 0.$$

This proves part (iii) of Proposition 1.4.

(iv) In view of the rescaling identity $L_{f;s}(x) = L_{f_s;1}(\frac{x}{s})$ with $f_s(u) := f(su)$, without loss of generality (w.l.o.g.) $s = 1$. Then part (iv) of the proposition follows by parts (ii) and (iii) and the observation that $\ell_f(z) := L_{f;1}(1-\sqrt{z})$ is concave in $z \in (0,1)$. In view of (1.6), it is enough to prove this observation for $f = \psi_t$ with $t \in (0,\infty]$; at that, by part (ii) of Proposition 1.4 and because $s_{\psi_t} = t$, w.l.o.g. let us assume that $0 < t < s = 1$. Observe that the second derivative $\ell''_{\psi_t}(z)$ in z admits of a piecewise-algebraic expression, which may be quickly obtained by using the Mathematica command `PiecewiseExpand`. Applying then a `Reduce` command, one finds that $\ell''_{\psi_t}(z) \leqslant 0$ for all $t \in (0,1)$ and $z \in (0,1)$. Now part (iv) of Proposition 1.4 follows.

(v) Part (v) of the proposition follows by parts (i)–(iv), on recalling (1.3) and taking into account that $L_{f;s}(0+) = f(s)$, for all $s \in (0,\infty)$.

Proposition 1.4 is now completely proved. □

Proof of Proposition 1.6. Take any $t \in (0,\infty]$. That $C_{\psi_\infty} = 1$ follows immediately by (1.3). So, w.l.o.g. $t \in (0,\infty)$, and then, by (1.3) and homogeneity, w.l.o.g. $t = 1$. Thus, it remains to show that $C_{\psi_1} = 2$. Take any $s \in (s_{\psi_1},\infty) = (1,\infty)$ and observe that $L'_{\psi_1;s}(1) = -2(s\wedge 2) < 0$, whereas $L'_{\psi_1;s}(1-) = -2(s\wedge 2) + 2s \geqslant 0$. Therefore, by part (iv) of Proposition 1.4,

$\max_{x\in(0,s)} L_{\psi_1;s}(x) = L_{\psi_1;s}(1) = s^2-(s-2)_+^2$. Now, using part (v) of Proposition 1.4, it is easy to see that $C_{\psi_1} = \sup_{s\in(1,\infty)} \frac{s^2-(s-2)_+^2}{s^2-(s-1)_+^2} = \lim_{s\to\infty} \frac{s^2-(s-2)_+^2}{s^2-(s-1)_+^2} = 2$. □

Proof of Proposition 1.16. Take any $p\in[1,2)$. It suffices to show that

$$\beta(p) := \big(1-D(p)\big)2^{2-p} \overset{(?)}{<} 1. \tag{2.5}$$

Observe that

$$\begin{aligned}\beta'(p) &= -2^{2-p}\ln 2 + (\tfrac{26}{5})^{2-p}\,\tfrac{\Gamma(p)}{\pi}\big[2(\sin\tfrac{\pi p}{2})\big(\ln\tfrac{26}{5}-(\ln\Gamma)'(p)\big)-\pi\cos\tfrac{\pi p}{2}\big]\\ &> -2^{2-p}\ln 2 > -2\ln 2 > -1.4;\end{aligned}$$

the first inequality here follows because $\cos\frac{\pi p}{2}\leqslant 0$, $\sin\frac{\pi p}{2}>0$, and $\ln\frac{26}{5}-(\ln\Gamma)'(p)\geqslant \ln\frac{26}{5}-(\ln\Gamma)'(2)>0$, taking into account that $\ln\Gamma$ is convex and hence $(\ln\Gamma)'$ is increasing. It is easy to see that $\max\{\beta(1+\frac{i}{4})\colon i\in\overline{1,2}\}<1-0.49$. So, $\beta(p)<\beta(1+\frac{i}{4})+(1.4)\frac14<1-0.49+(1.4)\frac14<1$ for $p\in[1+\frac{i-1}{4},1+\frac{i}{4}]$ and $i\in\overline{1,2}$; thus, (2.5) holds for all $p\in[1,\frac32]$.

Next,

$$\beta_2(p) := 25\pi\,\beta''(p)\,2^{p-1} = A + B(E_1+E_2+E_3+E_4),$$

where

$$\begin{gathered} A := 50\pi\,\ln^2 2,\quad B := 169\,\Gamma(p)\,(\tfrac{5}{13})^p,\\ E_1 := 4\pi\,(\cos\tfrac{\pi p}{2})\,\ln\tfrac{26}{5},\quad E_2 := \kappa\,\sin\tfrac{\pi p}{2},\\ E_3 := -4\big((\ln\Gamma)'(p)^2+(\ln\Gamma)''(p)\big)\sin\tfrac{\pi p}{2},\\ E_4 := (\ln\Gamma)'(p)\big(-4\pi\cos\tfrac{\pi p}{2}+8\ln\tfrac{26}{5}\,\sin\tfrac{\pi p}{2}\big),\end{gathered}$$

and $\kappa := \pi^2-4\ln^2 2-4\ln^2\frac{13}{5}-8\ln 2\,\ln\frac{13}{5}<0$, whence $E_2<0$. Also, $E_3<0$, because $(\ln\Gamma)''>0$. Let us next bound E_1 and E_4 from above, assuming that $p\in[\frac32,2]$. Then $E_1\leqslant 4\pi\,(\cos(\pi\frac34)\,\ln\frac{26}{5}<-14.6$; also, $(\ln\Gamma)'(p)\geqslant(\ln\Gamma)'(\frac32)>0$ and $(\ln\Gamma)'(p)\leqslant(\ln\Gamma)'(2)$, so that $E_4\leqslant(\ln\Gamma)'(2)\big(4\pi+8\ln\frac{26}{5}\big)<10.9$. Thus, for all $p\in[\frac32,2]$

$$\beta_2(p)\leqslant 50\pi\,\ln^2 2+169\,\Gamma(\tfrac32)\,(\tfrac{5}{13})^2(-14.6+10.9)<-6<0$$

and hence $\beta''(p)<0$, so that β is strictly concave on $[\frac32,2]$. At that, $\beta(2)=1$ and $\beta'(2)=1-\ln 2>0$; so, (2.5) holds for all $p\in[\frac32,2)$ as well. □

2.2. **Proof of Proposition 1.8.** Of the 5 parts of the proposition, the most difficult to prove are parts (iii) and (v), which are based to a certain extent on several lemmas. To state these lemmas, we need more notation. Recall the definition of $\ell(p,x)$ in (1.13) and introduce

$$\ell_p(p,x) := \frac{\partial}{\partial p}\ell(p,x),\quad \ell_x(p,x) := \frac{\partial}{\partial x}\ell(p,x),$$

$$\ell_{x,x}(p,x) := \frac{\partial}{\partial x}\ell_x(p,x) = \frac{\partial^2}{\partial x^2}\ell(p,x)$$

and also

$$p_x^* := \tfrac{1}{4}(25x+2) \quad \text{and} \quad x_p^* := \tfrac{2}{25}\,(2p-1),$$

so that $x = x_p^* \iff p = p_x^*$. Now we are ready to state the lemmas:

Lemma 2.1. *For all $p \in (1,2)$ and $x \in (0, \frac{1}{2})$, one has $\ell_{x,x}(p,x) < 0$ and hence $\ell_{x,x}(p,x) \neq 0$.*

Lemma 2.2. *For all $p \in (1,2)$,*

$$B(p) := 4(p-1)^{p-1} - (6-p)^{p-1} > 0. \tag{2.6}$$

Lemma 2.3. *For all $p \in (1,2)$ and $x \in (0, \frac{1}{2})$ such that $x \geqslant x_p^*$, one has $\ell_x(p,x) < 0$.*

Lemma 2.4. *For all $p \in (1,2)$ and $x \in (0, \frac{1}{2})$ such that $x < x_p^*$, one has $\ell_p(p,x) < 0$.*

The proofs of these lemmas are deferred to the end of this subsection. Let us now consider the four parts of Proposition 1.8.

(i,ii) Take any $p \in (1,2)$. Observe that $\ell_x(p, \frac{p-1}{2}) = 2^{1-p}\big((p-1)^{p-1} - (3-p)^{p-1}\big)p < 0$, since $p-1 < 3-p$. On the other hand, $\ell_x(p, \frac{p-1}{5}) = 5^{1-p} pB(p) > 0$, by Lemma 2.2. So, any value of $x_{f;s}$ as in part (iv) of Proposition 1.4 (for $f = |\cdot|^p$) must be in the interval $(\frac{p-1}{5}, \frac{p-1}{2}) \subset (0, \frac{1}{2})$. By Lemma 2.1 and part (iii) of Proposition 1.4 (with $s_f = 0$), $\ell_x(p,x)$ is strictly decreasing in $x \in (0, \frac{1}{2})$ from a positive value to a negative one. Now, in view of part (v) of Proposition 1.4, parts (i) and (ii) of Proposition 1.8 follow, taking also into account that the equation (1.14) is equivalent to $\ell_x(p,x) = 0$.

(iii) By part (i) of Proposition 1.8, x_p is the only root $x \in (0, \frac{1}{2})$ of the equation $\ell_x(p,x) = 0$, for each $p \in (1,2)$. So, by Lemma 2.1 and the implicit function theorem, $\tilde{C}_p$ is differentiable, and even real-analytic, and hence continuous in $p \in (1,2)$.

Next, by Lemma 2.3, for any $p \in (1,2)$ and $x \in (0, \frac{1}{2})$ the equality $\ell_x(p,x) = 0$ implies $x < x_p^*$, which in turn implies $\ell_p(p,x) < 0$, by Lemma 2.4. So, for any $p \in (1,2)$ one has $\ell_p(p, x_p) < 0$, whence $\frac{\mathrm{d}}{\mathrm{d}p}\tilde{C}_p = \frac{\mathrm{d}}{\mathrm{d}p}\ell(p, x_p) = \ell_p(p, x_p) + \ell_x(p, x_p)\frac{\partial}{\partial p}x_p = \ell_p(p, x_p) < 0$, which verifies that $\tilde{C}_p$ is decreasing in $p \in (1,2)$.

Thus, to complete the proof of part (iii) of the proposition, it remains to show that $\tilde{C}_{1+} = 2$ and $\tilde{C}_{2-} = 1$ (recall that $\tilde{C}_2 = 1$, by (1.12)). Here, consider first the case $p \downarrow 1$. Observe that then $\ell(p-1, p) = (2-p)^p - (p-1)^p + p(p-1)^{p-1} \to 2$; on the other hand, by (1.5), $\tilde{C}_p \leqslant 2$ for all $p \in (1,2]$. It indeed follows that $\tilde{C}_{1+} = 2$. Next, for all $x \in (0,1)$ and $p \in (\frac{3}{2}, 2)$, one has $\ell(2,x) = 1$ and $|x^p \ln x| < |x^{p-1} \ln x| < |x^{1/2} \ln x| < \frac{2}{e} < 1$, whence $|\ell_p(p,x)| = |x^{p-1} + px^{p-1}\ln x - x^p \ln x + (1-x)^p \ln(1-x)| \leqslant |x^{p-1}| + |px^{p-1}\ln x| + |x^p \ln x| + |(1-x)^p \ln(1-x)| \leqslant 1 + 2 + 1 + 1 = 5$; so, letting $p \uparrow 2$, one has $\ell(p,x) = \ell(2,x) - \int_p^2 \ell_p(r,x)\,\mathrm{d}r \leqslant 1 + 5(2-p) \to 1$, whence $\limsup_{p\uparrow 2} \tilde{C}_p = \limsup_{p\uparrow 2} \ell(p, x_p) \leqslant 1$. It remains to refer, again, to (1.5).

(iv) The proof of part (iv) of the proposition is straightforward.

(v) The equalities $\tilde C_{1+}=W_{1+}$ and $\tilde C_2=\tilde C_{2-}=W_{2-}=W_2$, and the similar equalities for the upper and lower bounds $\tilde C_p^{-,1}$, $\tilde C_p^{-,2}$, $\tilde C_p^{+,1}$, and $\tilde C_p^{+,2}$ on $\tilde C_p$ follow immediately by part (iii) of the proposition. Take now any $p\in(1,2)$. Consider $\tilde\ell(p,z):=\ell(p,1-\sqrt{z}\,)$, where $z\in(0,1)$. By parts (i) and (ii) of Proposition 1.8,

$$\tilde C_p=\max_{z\in(0,1)}\tilde\ell(p,z)=\max_{z\in(z_1,z_2)}\tilde\ell(p,z),$$

where $z_1:=z_1(p):=(\frac{3-p}{2})^2$ and $z_2:=z_2(p):=(\frac{6-p}{5})^2$ $\big($since the values $\frac{p-1}{2}$ and $\frac{p-1}{5}$ of x correspond, respectively, to the values z_1 and z_2 of z under the correspondence given by the formula $x=1-\sqrt{z}.\big)$ Hence, $\tilde C_p>\tilde\ell(p,z_1)\vee\tilde\ell(p,z_2)=\tilde C_p^{-,1}\vee\tilde C_p^{-,2}$, which proves the first inequality in (1.15). It follows from the proof of part (iv) of Proposition 1.4 that $\tilde\ell(p,z)$ is concave in $z\in(0,1)$. Also, in the proof of parts (i) and (ii) of the proposition it was observed that $\ell_x(p,\frac{p-1}{5})>0>\ell_x(p,\frac{p-1}{2})$, which is equivalent to $\tilde\ell_z(p,z_2)<0<\tilde\ell_z(p,z_1)$, where $\tilde\ell_z:=\frac{\partial\tilde\ell}{\partial z}$. Therefore, $\tilde\ell(p,z)\leqslant\tilde\ell(p,z_1)+\tilde\ell_z(p,z_1)(z-z_1)<\tilde\ell(p,z_1)+\tilde\ell_z(p,z_1)(z_2-z_1)=\tilde C_p^{+,1}$ and $\tilde\ell(p,z)\leqslant\tilde\ell(p,z_2)+\tilde\ell_z(p,z_2)(z-z_2)<\tilde\ell(p,z_2)+\tilde\ell_z(p,z_2)(z_1-z_2)=\tilde C_p^{+,2}$ for all $z\in(z_1,z_2)$, which yields the second inequality in (1.15). The third inequality in (1.15) is trivial.

So, it remains to prove the last inequality in (1.15). It is enough to show that $\rho(p)<0$, where

$$\begin{aligned}\rho(p)&:=2\times5^p\big(\tilde C_p^{+,2}-2^{2-p}\big)\\&=A(p)+\tfrac34\,\tfrac{27-7p}{6-p}\,p(p-1)B(p),\\A(p)&:=10p(p-1)^{p-1}-2(p-1)^p-2^{3-p}5^p+2(6-p)^p,\end{aligned}$$

and $B(p)$ is as in (2.6). Observe next that $27-7p\leqslant\frac{49}{60}(6-p)^2$. Hence and in view of Lemma 2.2,

$$4\rho(p)\leqslant\tilde\rho(p):=4A(p)+\tfrac{49}{20}(6-p)p(p-1)B(p);$$

thus, it suffices to show that $\tilde\rho(p)<0$, which can be rewritten as $\hat\rho(r)<0$ for $r\in(0,\frac25)$, where

$$\hat\rho(r):=16(\tfrac25)^{1+\frac52 r}\tilde\rho(1+\tfrac52 r).$$

One has

$$\rho_1(s):=\hat\rho'(r)\frac{(1+s)^3}{r^{5r/2}}=A_1(s)+4B_1(s)s^{\frac{5}{s+1}},$$

where

$$A_1(s):=16(-62+2202s+1160s^2+121s^3)+80(40+382s+105s^2+8s^3)\ln\tfrac{2}{1+s},$$
$$B_1(s):=1572-367s-795s^2-81s^3+(-1310s+75s^2+160s^3)\ln\tfrac{2s}{1+s},$$

and $s:=\frac2r-1$, so that $r=\frac{2}{1+s}$, and $r\in(0,\frac25)$ iff $s>4$. Using a `Reduce` command, one finds that $B_1(s)$ switches in sign from $-$ to $+$ as s increases from 4 to ∞, and the switch occurs at a certain point $s_*=31.4\dots$. With

$$\tilde\rho_1(s):=\frac{\rho_1(s)}{s^{5/(1+s)}B_1(s)}=\frac{A_1(s)}{s^{5/(1+s)}B_1(s)}+4,$$

another `Reduce` command shows (in about 12 sec) that

$$\rho_2(s) := \tilde{\rho}_1'(s) B_1(s)^2 s^{(6+s)/(1+s)} \tfrac{(1+s)^2}{80}$$

switches in sign from $+$ to $-$ to $+$ to $-$ as s increases from 4 to ∞, and the switches occur at certain points $s_1 = 5.2\dots$, $s_2 = 21.5\dots$, and $s_3 = 42.7\dots$. So, $\tilde{\rho}_1(s)$ switches from increase to decrease to increase as s increases from 4 to $s_1 = 5.2\dots$ to $s_2 = 21.5\dots$ to $s_* = 31.4\dots$, and then $\tilde{\rho}_1(s)$ switches from increase to decrease as s increases from $s_* = 31.4\dots$ to $s_3 = 42.7\dots$ to ∞. Next, $\tilde{\rho}_1(s) < 0$ for $s \in \{4, s_1, s_2, s_3\}$; also, $\rho_1(s_*) < 0$, whence $\tilde{\rho}_1(s_*-) = \infty > 0$ and $\tilde{\rho}_1(s_*+) = -\infty < 0$ (on recalling the definitions of $\tilde{\rho}_1(s)$ and s_*). It follows that $\tilde{\rho}_1(s)$ switches in sign from $-$ to $+$ as s increases from 4 to s_*, and $\tilde{\rho}_1 < 0$ on (s_*, ∞). Therefore, $\rho_1(s)$ switches in sign from $+$ to $-$ as s increases from 4 to ∞. Equivalently, $\hat{\rho}'(r)$ switches in sign from $-$ to $+$ as r increases from 0 to $\frac{2}{5}$. This implies that $\hat{\rho}(r)$ switches from decrease to increase as r increases from 0 to $\frac{2}{5}$. Equivalently, $(\frac{2}{5})^p \tilde{\rho}(p)$ switches from decrease to increase as p increases from 1 to 2. Note also that $\tilde{\rho}(1+) = \tilde{\rho}(2-) = \tilde{\rho}(2) = 0$. So, indeed $\tilde{\rho}(p) < 0$, for all $p \in (1,2)$. This proves part (v) and thus the entire proposition, modulo Lemmas 2.1–2.4.

Proof of Lemma 2.1. Introduce the new variable $y := \frac{1-x}{x}$, so that $y > 1$ for $x \in (0, \frac{1}{2})$. Then, for any $p \in (1,2)$ and $x \in (0, \frac{1}{2})$,

$$\begin{aligned}\ell_{x,x}(p,x)\,\frac{(1-x)^{2-p}}{p(p-1)} &= 1 - (2-p)y^{3-p} - (3-p)y^{2-p} \\ &< 1 - (2-p) - (3-p) = 2(p-2) < 0,\end{aligned}$$

which proves the lemma. □

Proof of Lemma 2.2. Take indeed any $p \in (1,2)$. Note that (2.6) is equivalent to $\tilde{B}(p) := \ln\left(4(p-1)^{p-1}\right) - \ln\left((6-p)^{p-1}\right) > 0$. Next, $\tilde{B}'(p) = 1 + r + \ln r$, where $r := \frac{p-1}{6-p}$, so that $\tilde{B}'(p)$ is increasing in p, and $\tilde{B}'(2) < 0$, which implies that $\tilde{B}'(p) < 0$ and hence $\tilde{B}(p)$ is decreasing in p, with $\tilde{B}(2) = 0$. Thus, indeed $\tilde{B}(p) > 0$. □

Proof of Lemma 2.3. Throughout the proof, it is assumed that indeed $p \in (1,2)$ and $x \in (0, \frac{1}{2})$. Let

$$(D_x\ell)(p,x) := \frac{\ell_x(p,x)}{p(1-x)^{p-1}},$$

so that $D_x\ell$ equals ℓ_x in sign. Then $\frac{\partial}{\partial x}(D_x\ell)(p,x) = (p-2)(p-1)(1-x)^{-p}x^{p-3} < 0$, so that $(D_x\ell)(p,x)$ decreases in x. Consider now

$$H(p) := (D_x\ell)(p, x_p^*) = (27-4p)^{1-p}(4p-2)^{p-2}(21p-23) - 1.$$

Obviously, $H(p)<0$ for $p\leqslant\frac{23}{21}$. Let us show that $H(p)<0$ for $p\in(\frac{23}{21},2)$ as well. Observe that

$$\begin{aligned} H'(p)&\frac{4(27-4p)^{p-1}(2p-1)^2(4p-2)^{-p}}{21p-23} \\ &= H_1(p):=\frac{25\,(42p^2-92p+73)}{(27-4p)(2p-1)(21p-23)}+\ln\frac{4p-2}{27-4p}. \end{aligned}$$

Using the Mathematica command `Minimize`, one finds that $H_1(p)>0$ and hence $H'(p)>0$ for $p\in(\frac{23}{21},2]$. Since $H(2)=0$, it indeed follows that $H(p)<0$ for $p\in(\frac{23}{21},2)$ and thus for all $p\in(1,2)$. So, one has $(D_x\ell)(p,x^*_p)<0$. Recalling that $(D_x\ell)(p,x)$ decreases in x, one has $(D_x\ell)(p,x)<0$ or, equivalently, $\ell_x(p,x)<0$ — provided that $x\geqslant x^*_p$. $\square$

Proof of Lemma 2.4. Throughout the proof, it is assumed that indeed $p\in(1,2)$ and $x\in(0,\frac12)$. Let

$$\begin{aligned} (D_p\ell)(p,x)&:=\frac{\ell_p(p,x)}{-(1-x)^p\ln(1-x)}=\frac{x^{p-1}(1+(p-x)\ln x)}{-(1-x)^p\,\ln(1-x)}-1, \\ (D_pD_p\ell)(p,x)&:=\frac{\partial(D_p\ell)(p,x)}{\partial p}\,\frac{(1-x)^p}{x^{p-1}}\,\frac{\ln(1-x)}{\ln x}, \end{aligned}$$

so that $D_p\ell$ and $D_pD_p\ell$ equal ℓ_p and $\frac{\partial(D_p\ell)}{\partial p}$ in sign, respectively. Then $\frac{\partial}{\partial p}(D_pD_p\ell)(p,x)=\ln(1-x)-\ln x>0$ $\big($since $x\in(0,\frac12)\big)$, so that $(D_pD_p\ell)(p,x)$ increases in p. Consider now

$$(D_pD_p\ell)(p^*_x,x)=\frac{[4+(21x+2)\ln x]\ln(1-x)-[8+(21x+2)\ln x]\ln x}{4\ln x}.$$

Observe that $1<p^*_x<2\iff\frac{2}{25}<x<\frac{6}{25}$, and then use the Mathematica command `Reduce` to find that $(D_pD_p\ell)(p^*_x,x)>0$ provided that $\frac{2}{25}<x<\frac{6}{25}$. Similarly, $(D_pD_p\ell)(1,x)>0$ provided that $0<x\leqslant\frac{2}{25}$. Thus, $(D_pD_p\ell)(1\vee p^*_x,x)>0$ for all $x\in(0,\frac{6}{25})$. Recalling that $(D_pD_p\ell)(p,x)$ increases in p, one has $(D_pD_p\ell)(p,x)>0$ for all $p\in[1\vee p^*_x,2)$. It follows that $(D_p\ell)(p,x)$ increases in $p\in[1\vee p^*_x,2)$. Now use `Reduce` to check that $(D_p\ell)(2,x)<0$, which yields $(D_p\ell)(p,x)<0$ or, equivalently, $\ell_p(p,x)<0$ for $p\in[1\vee p^*_x,2)$ or, equivalently, for $x\leqslant x^*_p$. $\square$

2.3. **Proofs of Corollary 1.9 and Propositions 1.10 and 1.11.** First in this subsection we shall prove Proposition 1.11, then Proposition 1.10, and finally Corollary 1.9.

Proof of Proposition 1.11. The convexity of $U_f(c,s,a)$ in $a\in\mathbb{R}$ follows immediately from that of f. Since f' is strictly positive and nondecreasing on $(0,\infty)$, it follows that $f(\infty-)=\infty$; similarly (or because f is even), $f(-\infty+)=\infty$. So, $U_f(c,s,a)\to\infty$ as $|a|\to\infty$. Therefore and by continuity, there is a minimizer of $U_f(c,s,a)$ in $a\in\mathbb{R}$. Take any such minimizer, say a_*. Since

$f \in \mathcal{C}^1(\mathbb{R})$, the partial derivative of $U_f(c,s,a)$ in a at $a = a_*$ is 0; that is, $cf'(s-c+a_*) + (s-c)f'(a_*-c) = 0$, which can be rewritten as

$$cf'(s-c+a_*) = (s-c)f'(c-a_*), \tag{2.7}$$

since f is even and hence f' is odd. Recall also that f' is strictly positive and hence nowhere zero on $(0,\infty)$. It follows that the arguments $s-c+a_*$ and $c-a_*$ of f' in (2.7) must be of the same sign; noting that the sum of these arguments is $s>0$, one concludes that they must be both positive; equivalently, $a_* \in (c-s,c)$. Moreover, f' is positive and nondecreasing on $(0,\infty)$ and $0<c<s-c$, so that (2.7) yields $f'(s-c+a_*) > f'(c-a_*)$ and hence

$$s-c+a_* > c-a_*. \tag{2.8}$$

If a minimizer of $U_f(c,s,a)$ in a is not unique, then the first two partial derivatives of $U_f(c,s,a)$ in a are identically zero for all a in some nonempty open interval $(a_1,a_2) \subset (c-s,c)$. That is, $cf'(s-c+a) = (s-c)f'(c-a)$ and $cf''(s-c+a) + (s-c)f''(a-c) = 0$ for all $a \in (a_1,a_2)$. Since f'' is nonnegative and even, it follows that $f''(c-a) = f''(a-c) = 0$ for all $a \in (a_1,a_2)$, so that $f'' = 0$ on the interval $(c-a_2, c-a_1)$. Because $a_2 \leqslant c$ and f'' is nonnegative and nonincreasing on $(0,\infty)$, one has $f''=0$ on the interval $(c-a_2,\infty)$, so that f' is constant on the same interval. On recalling (2.8), one has $s-c+a > c-a > c-a_2$ for any $a \in (a_1,a_2)$, which shows that $f'(s-c+a) = f'(c-a)$; however, this contradicts the previously obtained inequality $f'(s-c+a_*) > f'(c-a_*)$ for any minimizer a_*.

Next, the formula (1.23) for the unique minimizer of $U_{\psi_t}(c,s,a)$ in a is easy to verify by noting that the partial derivative of $U_{\psi_t}(c,s,a)$ in a at $a = \frac{c}{s-c}\,(s-c-t)_+$ is 0. Moreover, for any real c an t such that $c>t>0$ one has $\dfrac{U_{\psi_1}(c,s,0)}{U_{\psi_1}(c,s,a_{\psi_1;c,s})} \underset{s\to\infty}{\longrightarrow} 2-\frac{t}{2c}$, and then $2-\frac{t}{2c} \underset{c\to\infty}{\longrightarrow} 2$, which shows that $\kappa_{\psi_t} = 2$.

It remains to prove that the unique minimizer $a = a_{f;c,s}$ is nonnegative. Equivalently, it remains to show that the partial derivative of $U_f(c,s,a)$ in a is no greater than 0 at $a=0$, that is,

$$cf'(s-c) \geqslant (s-c)f'(c). \tag{2.9}$$

By the linearity relation (1.9) and homogeneity, w.l.o.g. $f = \psi_t$ for some $t \in (0,\infty)$, in which case (2.9) is equivalent to $a_{\psi_t;c,s} \geqslant 0$, and that is obvious from (1.23). □

Proof of Proposition 1.10. Take indeed any $f \in \mathcal{F}_{1,2} \setminus \{0\}$. By e.g. [32, Proposition 3.18], any zero-mean probability distribution on $\mathbb{R}\setminus\{0\}$ is a mixture of zero-mean probability distributions on 2-point sets. Therefore, w.l.o.g. the zero-mean r.v. X takes on only two values, so that $X = X_{c,d}$, where c and d are positive real numbers, and $X_{c,d}$ is a r.v. such that $\mathsf{P}(X_{c,d} = -c) = \frac{d}{c+d}$ and $\mathsf{P}(X_{c,d} = d) = \frac{c}{c+d}$. Take now any c and s such that $0<c<s<\infty$, and introduce

$$R_f(c,s,a) := \frac{U_f(c,s,0)}{U_f(c,s,a)} = \frac{\mathsf{E}\, f(X_{c,s-c})}{\mathsf{E}\, f(X_{c,s-c}+a)}.$$

So, the best constant κ in (1.22) is given by a formula similar to (1.18), but with the restrictions $c \in (0, s)$ and $a \in \mathbb{R}$ instead of $c \in (0, \frac{s}{2})$ and $a \in (0, c)$. That $c \in (0, s)$ can be reduced to $c \in (0, \frac{s}{2})$ follows by the symmetry relation $R_f(c, s, a) \equiv R_f(s-c, s, -a)$ and the continuity of $R_f(c, s, a)$ in c. Finally, the condition $a \in \mathbb{R}$ can be reduced to $a \in (0, c)$ by Proposition 1.11 and the continuity of $R_f(c, s, a)$ in a. □

Proof of Corollary 1.9.

(I) Take indeed any $f \in \mathcal{F}_{1,2} \setminus \{0\}$. Consider the martingale expansion

$$Y = \mathsf{E}\, Y + \xi_1 + \dots + \xi_n$$

of Y with the martingale-differences

$$\xi_i := \mathsf{E}_i\, Y - \mathsf{E}_{i-1}\, Y$$

for $i \in \overline{1, n}$, where E_i stands for the conditional expectation given the σ-algebra generated by $(X_1, \dots, X_i)$, with $\mathsf{E}_0 := \mathsf{E}$. For each $i \in \overline{1, n}$ introduce the r.v. $\eta_i := \mathsf{E}_i(Y - \tilde{Y}_i)$, where $\tilde{Y}_i := g(X_1, \dots, X_{i-1}, x_i, X_{i+1}, \dots, X_n)$; then, in view of (1.16) or (1.25), $|\eta_i| \leqslant \rho_i(X_i, x_i)$; because $f(u)$ is increasing in $|u|$, it follows that $f(\eta_i) \leqslant f\big(\rho_i(X_i, x_i)\big)$ and hence $\mathsf{E}\, f(\eta_i) \leqslant \mathsf{E}\, f\big(\rho_i(X_i, x_i)\big)$; also, $\xi_i = \eta_i - \mathsf{E}_{i-1}\, \eta_i$, since the r.v.'s $X_1, \dots, X_n$ are independent. Now (1.17) follows from Theorem 1.1 and Proposition 1.10, which latter yields $\mathsf{E}_{i-1}\, f(\xi_i) \leqslant \kappa_f\, \mathsf{E}_{i-1}\, f(\eta_i)$ and hence $\mathsf{E}\, f(\xi_i) \leqslant \kappa_f\, \mathsf{E}\, f(\eta_i)$.

To check the inclusion $\kappa_f \in [1, 2]$ in (1.18), note first that the inequality $\kappa_f \geqslant 1$ follows by the continuity of $U_f(c, s, a)$ in a, at $a = 0$. As for the inequality $\kappa_f \leqslant 2$, it can be rewritten as

$$U_f(c, s, 0) \leqslant 2U_f(c, s, a) \tag{2.10}$$

for all $s \in (0, \infty)$, $c \in (0, \frac{s}{2})$, and $a \in (0, c)$, where w.l.o.g. $f = \psi_t$ $\big($for some $t \in (0, \infty)$, by (1.19) and (1.6)$\big)$ and $s = 1$ (by homogeneity). Take then indeed any $c \in (0, \frac{1}{2})$ and $a \in (0, c)$. By Proposition 1.11, w.l.o.g. $a = a_{\psi_t;c,1}$. Using a `Simplify` Mathematica command for $U_{\psi_t}(c, 1, a_{\psi_t;c,1})$ and then following with a `Reduce`, one quickly verifies that (2.10) indeed holds for $f = \psi_t$. This completes the proof of part (I) of Corollary 1.9.

(II) To obtain the expression in (1.21) for $\tilde{\kappa}_p = \kappa_{|\cdot|^p}$, note first that, by homogeneity of the power function $f = |\cdot|^p$, w.l.o.g. $s = 1$. Then solve the equation (2.7) to find the unique minimizer

$$a_* = \tilde{a}_{p;c} := a_{|\cdot|^p;c} = c - \frac{c^{1/(p-1)}}{c^{1/(p-1)} + (1-c)^{1/(p-1)}}$$

of $\tilde{U}_p(c, a) := U_{|\cdot|^p}(c, 1, a)$ in a. Finally, substitute this minimizer for a in $\tilde{R}_p(c, a) := \frac{\tilde{U}_p(c,0)}{\tilde{U}_p(c,a)}$ and simplify, to show that $\hat{r}_c(p) := \tilde{R}_p(c, \tilde{a}_{p;c})$ equals the expression under the max sign in (1.21).

The continuity of $\tilde{\kappa}_p$ in p follows because $\hat{r}_c(p)$ is continuous in $p \in (1, 2]$ uniformly in $c \in [0, \frac{1}{2}]$ $\big($indeed, the derivative, $\hat{r}_c'(p)$, of $\hat{r}_c(p)$ in p is bounded over all $c \in [0, \frac{1}{2}]$ and all p in any compact subinterval of $(1, 2]\big)$. That $\tilde{\kappa}_2 = 1$ is trivial. To check that $\tilde{\kappa}_{1+} = 2$, observe that $\tilde{R}_p(p-1, p) \to 2$ as $p \downarrow 1$ and recall that

$\kappa_f \leqslant 2$ for all $f \in \mathcal{F}_{1,2} \setminus \{0\}$. The statements that the values of $\tilde{\kappa}_p$ are algebraic for all rational $p \in (1,2]$ and $\tilde{\kappa}_{3/2} = \frac{1}{9}\sqrt{51+21\sqrt{7}} = 1.14\dots$, corresponding to $c = \frac{1}{6}(3 - \sqrt{1+2\sqrt{7}}) = 0.081\dots$, are straightforward to check.

It remains to prove that $\tilde{\kappa}_p$ strictly decreases in $p \in (1,2]$. To accomplish this, it is enough to show that $\hat{r}_c(p)$ does so for each $c \in (0, \frac{1}{2})$, since $\hat{r}_0(p) = \hat{r}_{1/2}(p) = 1$ for all $p \in (1,2]$ and $\hat{r}_c(2) = 1$ for all $c \in [0, \frac{1}{2}]$. Take indeed any $p \in (1,2)$ and $c \in (0, \frac{1}{2})$ and observe that $(\ln \hat{r}_c)'(p) = r_1 + r_2 - \frac{1}{p-1} r_3$, where

$$\begin{aligned}
r_1 &:= \frac{c^{p-1}\ln c + (1-c)^{p-1}\ln(1-c)}{c^{p-1} + (1-c)^{p-1}}, \\
r_2 &:= \ln\big(c^{1/(p-1)} + (1-c)^{1/(p-1)}\big), \\
r_3 &:= \frac{c^{1/(p-1)}\ln c + (1-c)^{1/(p-1)}\ln(1-c)}{c^{1/(p-1)} + (1-c)^{1/(p-1)}}.
\end{aligned}$$

Note that $r_1 + r_2 - \frac{1}{p-1} r_3 = R_1 + R_2$, where $R_1 := r_1 - r_3$ and $R_2 := r_2 + (1 - \frac{1}{p-1}) r_3$. Observe that

$$R_1 = \frac{\Big(\big(\frac{1-c}{c}\big)^{p-1} - \big(\frac{1-c}{c}\big)^{\frac{1}{p-1}}\Big)\, c^{p-1+\frac{1}{p-1}} \ln \frac{1-c}{c}}{\Big(c^{\frac{1}{p-1}} + (1-c)^{\frac{1}{p-1}}\Big)\,(c^{p-1} + (1-c)^{p-1})} < 0,$$

since $\frac{1-c}{c} > 1$ and $p - 1 < 1 < \frac{1}{p-1}$.

It remains to show that $R_2 < 0$. Consider the new variable

$$b := \frac{c^{1/(p-1)}}{c^{1/(p-1)} + (1-c)^{1/(p-1)}},$$

so that $b \in (0, \frac{1}{2})$ and $c = \frac{b^{p-1}}{b^{p-1} + (1-b)^{p-1}}$. Then one can check that

$$R_2 = h(b) := (p-2)\big(b\ln b + (1-b)\ln(1-b)\big) - \ln\big(b^{p-1} + (1-b)^{p-1}\big)$$

and

$$h''(b) b^{2-p} (1-b)^{2-p} (b^{p-1} + (1-b)^{p-1})^2 = h_{21}(b) h_{22}(b),$$

where

$$h_{21}(b) := \big(\tfrac{2-p}{b} - 1\big)\big(\tfrac{b}{1-b}\big)^{2-p} + 1, \quad h_{22}(b) := \big(\tfrac{b}{1-b}\big)^{p-1}\big(\tfrac{p-1}{b} - 1\big) - 1,$$

with $h_{21}'(b) = (p-2)(p-1)\big(\frac{b}{1-b}\big)^{-p}(1-b)^{-3} < 0$ and $h_{22}'(b) = (p-2) \times (p-1)\big(\frac{b}{1-b}\big)^{p} b^{-3} < 0$, so that both $h_{21}(b)$ and $h_{22}(b)$ are decreasing in b. Since $h_{21}(\frac{1}{2}) = 2(2-p) > 0$, it follows that $h_{21} > 0$ on $(0, \frac{1}{2})$. So, $h''(b)$ equals $h_{22}(b)$ in sign. Since $h_{22}(0+) = \infty > 0$ and $h_{22}(\frac{1}{2}) = 2(p-2) < 0$, both $h_{22}(b)$ and $h''(b)$ switch from $+$ to $-$ as b increases from 0 to $\frac{1}{2}$. Therefore, $h(b)$ switches from convexity to concavity in $b \in (0, \frac{1}{2})$. At that, $h(0+) = h(\frac{1}{2}) = h'(\frac{1}{2}) = 0$. It follows that $h < 0$ and hence $R_2 < 0$. This completes the proof of part (II) and thus that of the entire Corollary 1.9. □

2.4. **Proof of Theorem 1.1.**

(I, II) By induction and conditioning, parts (I) and (II) of Theorem 1.1 follow immediately from

Lemma 2.5. *Take any $f \in \mathcal{F}_{1,2} \setminus \{0\}$.*

(I) For any $x \in \mathbb{R}$ and zero-mean r.v. Y

$$\mathsf{E}\, f(x+Y) \leqslant f(x) + C_f\, \mathsf{E}\, f(Y).$$

(II) If a constant factor C is such that

$$\mathsf{E}\, f(X+Y) \leqslant \mathsf{E}\, f(X) + C\, \mathsf{E}\, f(Y) \tag{2.11}$$

for all independent zero-mean r.v.'s X and Y, then $C \geqslant C_f$.

We shall turn to the proof of this lemma in a moment, after the proof of parts (III) and (IV) of Theorem 1.1 is completed.

(III) Take any $f \in \mathcal{F}_{1,2} \setminus \{0\}$. The inequality $C_f \geqslant 1$ follows by (1.3), since $L_{f;s}(x) \to f(s)$ as $x \downarrow 0$. On the other hand, in view of Proposition 1.6 and (1.3), one has $L_{\psi_t;s}(x) \leqslant 2\psi_t(s)$ for any $t \in (0,\infty]$ and x, s such that $0 < x < s < \infty$; so, (1.6) implies $L_{f;s}(x) \leqslant 2f(s)$, whence, by (1.3), $C_f \leqslant 2$.

(IV) Part (IV) of Theorem 1.1 follows immediately from Propositions 1.6 and 1.8.

Thus, Theorem 1.1 is proved, modulo Lemma 2.5.

Proof of Lemma 2.5. The main idea of this proof is to use appropriate Taylor expansions. A similar approach was used e.g. in [12, 8, 35, 29, 28, 34].

(I) Clearly, for all real z and y,

$$f(z+y) \leqslant f(z) + yf'(z) + \hat{C}_f f(y), \tag{2.12}$$

where

$$\hat{C}_f := \sup_{\substack{z\in\mathbb{R},\\ y\in\mathbb{R}\setminus\{0\}}} R_f(z,y) \quad \text{and} \quad R_f(z,y) := \frac{f(z+y) - f(z) - yf'(z)}{f(y)}. \tag{2.13}$$

Here one may recall that, as was noted at the end of the paragraph containing (1.1), $f > 0$ on $\mathbb{R} \setminus \{0\}$. Concerning the validity of (2.12) when $y = 0$, recall that $f(0) = 0$ and assume that $\hat{C}_f f(y) = 0$ if $y = 0$ and $\hat{C}_f = \infty$ (in fact, later it will be seen that $\hat{C}_f$ is always between 1 and 2.

It is not hard to see that

$$\hat{C}_f = C_f. \tag{2.14}$$

Indeed, because f is an even function and hence f' is an odd function, it follows that $R_f(-z,-y) = R_f(z,y)$ for any $z \in \mathbb{R}$ and $y \in \mathbb{R} \setminus \{0\}$. So, one may replace the condition $y \in \mathbb{R} \setminus \{0\}$ in (2.13) by $y \in (-\infty, 0)$. Take indeed any such y and consider the Taylor expansion

$$R_f(z,y)f(y) = f(z+y) - f(z) - yf'(z) = y^2 \int_0^1 (1-t) f''(z+ty)\,\mathrm{d}t. \tag{2.15}$$

By (1.1), f'' is nondecreasing on the interval $(-\infty, 0)$. Next, note that $z + ty < 0$ whenever $z \in (-\infty, 0]$, $y \in (-\infty, 0)$, and $t \in (0,1)$. Therefore, in view of

(2.15) and the continuity of f and f', $R_f(z,y)$ is nondecreasing in $z \in (-\infty, 0]$. Similarly, $R_f(z,y)$ is nonincreasing in $z \in [-y, \infty)$, because f'' is nonincreasing on the interval $(0,\infty)$ and $z + ty > 0$ whenever $z \in [-y,\infty)$, $y \in (-\infty, 0)$, and $t \in (0,1)$. Hence, the condition $z \in \mathbb{R}, y \in \mathbb{R}\setminus\{0\}$ in (2.13) can be replaced by $y \in (-\infty,0), z \in (0,-y)$. Thus, (2.14) follows by replacing s and x in (1.3) by $-y$ and z, respectively.

Now part (I) of Lemma 2.5 follows immediately from (2.12) and (2.14).

(II) For any positive real numbers c and d, let $X_{c,d}$ stand for any r.v. such that $\mathsf{P}(X_{c,d} = -c) = \frac{d}{c+d}$ and $\mathsf{P}(X_{c,d} = d) = \frac{c}{c+d}$. Take now any c and s such that $0 < c < s < \infty$ and introduce

$$g_{f;c,s}(x) := \mathsf{E}\, f(x + X_{c,s-c}) - f(x) \quad \text{and} \quad J_{f;c,s}(x) := \frac{g_{f;c,s}(x)}{g_{f;c,s}(0)};$$

the latter definition is correct, because $f > 0$ on $\mathbb{R}\setminus\{0\}$ and hence $g_{f;c,s}(0) = \mathsf{E}\, f(X_{c,s-c}) > 0$.

In view of the Taylor expansion in (2.15), for any $x \in \mathbb{R}$

$$\begin{aligned} s g_{f;c,s}(x) &= cf(x+s-c) + (s-c)f(x-c) - sf(x) \\ &= (s-c)c\int_0^1 (1-t)\big[(s-c)f''\big(x+(s-c)t\big) + cf''(x-ct)\big]\,\mathrm{d}t. \end{aligned} \tag{2.16}$$

Since f'' is even on $\mathbb{R}$ and nonnegative and nonincreasing on $(0,\infty)$, the identity (2.16) implies that $g_{f;c,s}(x)$ converges to a finite limit as $x \to -\infty$, and then so does $J_{f;c,s}(x)$. Let now a and b be any positive real numbers. Then

$$\frac{\mathsf{E}\, f(X_{a,b} + X_{c,s-c}) - \mathsf{E}\, f(X_{a,b})}{\mathsf{E}\, f(X_{c,s-c})} = \frac{b}{a+b} J_{f;c,s}(-a) + \frac{a}{a+b} J_{f;c,s}(b) \underset{a\to\infty}{\longrightarrow} J_{f;c,s}(b),$$

assuming that the r.v.'s $X_{a,b}$ and $X_{c,s-c}$ are independent. So, the constant C in (2.11) cannot be less than $J_{f;c,s}(b)$, for any c, s, b such that $0 < c < s < \infty$ and $0 < b < \infty$.

On the other hand, by l'Hospital's rule, for any $x \in \mathbb{R}$,

$$J_{f;c,s}(x) \underset{c\uparrow s}{\longrightarrow} \frac{L_{f;s}(x)}{f(s)},$$

with $L_{f;s}(x)$ as in (1.4) So, in view of (1.3), $C \geqslant C_f$. So, part (II) of Lemma 2.5 is proved as well. □

Now Theorem 1.1 is completely proved.

References

1. G. Alsmeyer and U. Rösler, *The best constant in the Topchii-Vatutin inequality for martingales*, Statist. Probab. Lett. **65** (2003), no. 3, 199–206.
2. B. von Bahr and C.-G. Esseen, *Inequalities for the rth absolute moment of a sum of random variables,* $1 \leqslant r \leqslant 2$, Ann. Math. Statist **36** (1965) 299–303.
3. K. Ball, E.A. Carlen and E.H. Lieb, *Sharp uniform convexity and smoothness inequalities for trace norms*, Invent. Math. **115** (1994), no. 3, 463–482.
4. V. Bentkus, *On Hoeffding's inequalities*, Ann. Probab. **32** (2004), no. 2, 1650–1673.
5. V. Bentkus, *On measure concentration for separately Lipschitz functions in product spaces*, Israel J. Math. **158** (2007) 1–17.

6. S. Boucheron, O. Bousquet, G. Lugosi and P. Massart, *Moment inequalities for functions of independent random variables*, Ann. Probab. **33** (2005), no. 2, 514–560.
7. B.M. Brown, *Characteristic functions, moments, and the central limit theorem*, Ann. Math. Statist. **41** (1970) 658–664.
8. S.D. Chatterji, *An L^p-convergence theorem*, Ann. Math. Statist. **40** (1969) 1068–1070.
9. J.A. Clarkson, *Uniformly convex spaces*, Trans. Amer. Math. Soc. **40** (1936), no. 3, 396–414.
10. D.C. Cox, *Note on a martingale inequality of Pisier*, Math. Proc. Cambridge Philos. Soc. **92** (1982), no. 1, 163–165.
11. M.M. Day, *Uniform convexity in factor and conjugate spaces*, Ann. of Math. (2) **45** (1944) 375–385.
12. S.W. Dharmadhikari, V. Fabian and K. Jogdeo, *Bounds on the moments of martingales*, Ann. Math. Statist. **39** (1968) 1719–1723.
13. L. Dümbgen, S.A. van de Geer, M.C. Veraar and J.A. Wellner, *Nemirovski's inequalities revisited*, Amer. Math. Monthly **117** (2010), no. 2, 138–160.
14. M.L. Eaton, *A note on symmetric Bernoulli random variables*, Ann. Math. Statist. **41** (1970) 1223–1226.
15. P. Enflo, *Banach spaces which can be given an equivalent uniformly convex norm*, in *Proceedings of the International Symposium on Partial Differential Equations and the Geometry of Normed Linear Spaces (Jerusalem, 1972)*, vol. 13, 281–288 (1973) (1972).
16. T. Figiel, P. Hitczenko, W.B. Johnson, G. Schechtman and J. Zinn, *Extremal properties of Rademacher functions with applications to the Khintchine and Rosenthal inequalities*, Trans. Amer. Math. Soc. **349** (1997), no. 3, 997–1027.
17. R. Fortet and E. Mourier, *Les fonctions aléatoires comme éléments aléatoires dans les espaces de Banach*, Studia Math. **15** (1955) 62–79.
18. J. Hoffmann-Jørgensen, *On the modulus of smoothness and the g_α-conditions in B-spaces (Aarhus Universitet, Matematisk Institut, Preprint Series, September 1974)*.
19. V.V. Jurinskiĭ, *Exponential estimates for large deviations*, Teor. Verojatnost. i Primenen. **19** (1974) 152–154.
20. J.H.B. Kemperman and J.C. Smit, *Sharp upper and lower bounds for the moments of a martingale*, Advances in Applied Probability **6** (1974), no. 2, 244–245.
21. R. Latała and K. Oleszkiewicz, *Between Sobolev and Poincaré*, in *Geometric aspects of functional analysis*, vol. 1745 of *Lecture Notes in Math.*, Springer, Berlin, 147–168 (2000).
22. M. Ledoux, *The concentration of measure phenomenon*, vol. 89 of *Mathematical Surveys and Monographs*, American Mathematical Society, Providence, RI (2001).
23. M. Ledoux and K. Oleszkiewicz, *On measure concentration of vector-valued maps*, Bull. Pol. Acad. Sci. Math. **55** (2007), no. 3, 261–278.
24. M. Ledoux and M. Talagrand, *Probability in Banach spaces*, vol. 23 of *Ergebnisse der Mathematik und ihrer Grenzgebiete (3) [Results in Mathematics and Related Areas (3)]*, Springer-Verlag, Berlin, isoperimetry and processes (1991).
25. J. Lindenstrauss, *On the modulus of smoothness and divergent series in Banach spaces*, Michigan Math. J. **10** (1963) 241–252.
26. È. Manstavichyus, *Inequalities for the pth moment, $0 < p < 2$, of a sum of independent random variables*, Lithuanian Math. J. **22** (1982), no. 1, 64–67.
27. S.-i. Ohta, *Uniform convexity and smoothness, and their applications in Finsler geometry*, Math. Ann. **343** (2009), no. 3, 669–699.
28. I. Pinelis, *Extremal probabilistic problems and Hotelling's T^2 test under a symmetry condition*, Ann. Statist. **22** (1994), no. 1, 357–368.
29. I. Pinelis, *Optimum bounds for the distributions of martingales in Banach spaces*, Ann. Probab. **22** (1994), no. 4, 1679–1706.
30. I. Pinelis, *On normal domination of (super)martingales*, Electron. J. Probab. **11** (2006), no. 39, 1049–1070.
31. I. Pinelis, *Exact inequalities for sums of asymmetric random variables, with applications*, Probab. Theory Related Fields **139** (2007), no. 3-4, 605–635.

32. I. Pinelis, *Optimal two-value zero-mean disintegration of zero-mean random variables*, Electron. J. Probab. **14** (2009), no. 26, 663–727.
33. I. Pinelis, *Positive-part moments via the Fourier–Laplace transform*, J. Theoret. Probab. **24** (2011), no. 2, 409–421.
34. I. Pinelis, *Rosenthal-type inequalities for martingales in* 2*-smooth Banach spaces*, Teor. Veroyatnost. i Primenen **59** (2014), no. 4, 814-821.
35. I.F. Pinelis, *Estimates for moments of infinite-dimensional martingales*, Math. Notes **27** (1980), no. 5–6, 459–462.
36. I.F. Pinelis and A.I. Sakhanenko, *Remarks on inequalities for probabilities of large deviations*, Theory Probab. Appl. **30** (1985), no. 1, 143–148.
37. G. Pisier, *Martingales with values in uniformly convex spaces*, Israel J. Math. **20** (1975), no. 3-4, 326–350.
38. V.A. Topchiĭ and V.A. Vatutin, *The maximum of critical Galton-Watson processes, and left-continuous random walks*, Theory Probab. Appl. **42** (1997), no. 1, 17–27.
39. S.A. Utev, *Extremal problems in moment inequalities*, in *Limit theorems of probability theory*, vol. 5 of *Trudy Inst. Mat.*, "Nauka" Sibirsk. Otdel., Novosibirsk, 56–75, 175 (1985).
40. W.A. Woyczyński, *Random series and laws of large numbers in some Banach spaces*, Theory Probab. Appl. **18** (1973) 350–355.
41. W.A. Woyczyński, *Geometry and martingales in Banach spaces*, in *Probability—Winter School (Proc. Fourth Winter School, Karpacz, 1975)*, Springer, Berlin, 229–275. Lecture Notes in Math., Vol. 472 (1975).

Department of Mathematical Sciences, Michigan Technological University, Houghton, Michigan 49931, USA.
E-mail address: ipinelis@mtu.edu