TRANSACTIONS OF THE
AMERICAN MATHEMATICAL SOCIETY
Volume 349, Number 3, March 1997, Pages 997–1027
S 0002-9947(97)01789-3

EXTREMAL PROPERTIES OF RADEMACHER FUNCTIONS WITH APPLICATIONS TO THE KHINTCHINE AND ROSENTHAL INEQUALITIES

T. FIGIEL, P. HITCZENKO, W. B. JOHNSON,
G. SCHECHTMAN, AND J. ZINN

ABSTRACT. The best constant and the extreme cases in an inequality of H.P. Rosenthal, relating the p moment of a sum of independent symmetric random variables to that of the p and 2 moments of the individual variables, are computed in the range $2 < p \leq 4$. This complements the work of Utev who has done the same for $p > 4$. The qualitative nature of the extreme cases turns out to be different for $p < 4$ than for $p > 4$. The method developed yields results in some more general and other related moment inequalities.

1. INTRODUCTION

The original motivation for this research was to obtain the best constants in Rosenthal's inequality for sums of independent symmetric random variables. Rosenthal's inequality, [R], which generalizes the classical Khintchine inequality, has the following formulation: Let (X_i) be a sequence of independent symmetric random variables. Then for every p, $2 \leq p < \infty$, there exists a constant D_p such that:

$$M_p(\{X_i\}) \leq \| \sum X_i \|_p \leq D_p M_p(\{X_i\}),$$

where

$$M_p(\{X_i\}) = \max \left\{ \left(\sum \|X_i\|_2^2 \right)^{1/2}, \left(\sum \|X_i\|_p^p \right)^{1/p} \right\}.$$

Here and in the sequel, the symbol $\|X\|_p$ denotes the L_p–norm of a random variable X. Rosenthal's proof gave only exponentials of p estimates on the growth rate of D_p as $p \to \infty$. Later on Johnson, Schechtman and Zinn [JSZ] proved that the actual growth rate of D_p is $p/\log p$. The latter result was later generalized by Talagrand [T] to Banach space valued random variables (with $(\sum \|X_i\|_2^2)^{1/2}$ replaced by $\| \sum X_i \|_2$). In [KS], Kwapień and Szulga obtained a simpler proof of Talagrand's result. After most of our work was completed, we were informed by Kwapień that for $p \geq 4$ the

Received by the editors December 22, 1994.
1991 *Mathematics Subject Classification.* Primary 60E15, 60G50; Secondary 26D07, 46E30.
Key words and phrases. Khintchine inequality, Rosenthal inequality, Orlicz function, extremal problem, Rademacher functions.
The first, second and fourth authors were participants in the NSF Workshop in Linear Analysis & Probability, Texas A&M University.
Professors Hitczenko, Johnson, and Zinn were supported in part by NSF grants.
Johnson, Schechtman and Zinn were supported in part by US–Israel Binational Science Foundation.

problem of finding the exact value of the constant D_p was solved by Utev in [U2]. He showed that D_p can be taken to be equal to $\|\mathcal{P}(1)\|_p$, where $\mathcal{P}(1)$ denotes the symmetrized Poisson random variable with parameter 1. In the present paper we will concentrate on the case $2 \le p \le 4$. We will show that, if $2 < p \le 4$, then the optimal D_p is given by $D_p = (1 + \|g\|_p^p)^{1/p}$, where g is a standard Gaussian random variable. Note that $D_2 = 1$ so that the function $p \to D_p$ is discontinuous at 2.

In trying to solve the problem we found it very helpful to consider Orlicz functions Φ other than just power functions. Also, we found it convenient to look at the problem as an extremal problem related to that inequality. This approach seems to be helpful also in connection with other probabilistic inequalities and, in fact, has been used for quite some time. The general line of argument is as follows: given a moment inequality for sums of random variables, we first try to formulate a related extremal problem and find conditions under which the solution of that problem is given by linear combinations of (at most) two valued random variables. Solving an extremal problem usually requires a convexity type condition on the Orlicz function. Once the extremal problem is solved we try to find an analytic condition on the Orlicz function Φ that would be equivalent to (or at least sufficient for) the convexity condition to be satisfied. Then, using special properties of linear combinations of Rademacher functions, we will identify the best constant. To make our discussion more meaningful, let us consider the classical Khintchine inequality: Let (ε_j) be a Rademacher sequence (i.e. a sequence of i.i.d. symmetric random variables taking on values 1 and -1). Then, for every p, $0 < p < \infty$, there exist constants A_p and B_p such that for every sequence (a_j) of real numbers one has that

$$A_p \left(\sum_{j=1}^n a_j^2 \right)^{1/2} \le \left(E \left| \sum_{j=1}^n a_j \varepsilon_j \right|^p \right)^{1/p} \le B_p \left(\sum_{j=1}^n a_j^2 \right)^{1/2}.$$

Once such an inequality is proved, one may want to find the best possible constant appearing in the above inequality as well as in inequalities which are extensions of Khintchine's inequality to Orlicz functions. This problem has quite a long history. In the case of power function it was solved by Haagerup [H] for the whole range of p, $0 < p < \infty$. Several partial results were obtained earlier by Whittle [W], Eaton [E1], [E2], Stečkin [S], Young [Y], Szarek [Sz].

For $p \ge 2$, Haagerup proved that $B_p = \|g\|_p$, where g is the standard Gaussian random variable. The case $p \ge 3$ was obtained by Whittle [W] [the result is stated for any $p \ge 2$ but the proof is incorrect if $2 < p < 3$]. Whittle's paper apparently was unnoticed. The only reference prior to 1992 we are aware of is in [Pe]. Whittle's argument, based on Schur convexity, was essentially repeated in [E1] and [K].

Given Haagerup's result, the right–hand side of Khintchine's inequality can be rewritten as

$$E \left| \sum_{j=1}^n a_j \varepsilon_j \right|^p \le \|g\|_p^p \left(\sum_{j=1}^n a_j^2 \right)^{p/2} = E \left| \sum_{j=1}^n a_j g_j \right|^p,$$

where (g_j) is a sequence of independent copies of g. One can now ask whether the inequality

$$E \left| \sum_{j=1}^n \sigma_j \varepsilon_j \right|^p \le E \left| \sum_{j=1}^n X_j \right|^p,$$

is true for all sequences of independent, symmetric random variables $\{X_j : j = 1, \dots, n\}$ where $\sigma_j^2 = E\,X_j^2$. More generally, one can ask: for which Orlicz functions Φ do we have

$$E\ \Phi\Big(\sum_{j=1}^{n} \sigma_j\, \varepsilon_j\Big) \le E\ \Phi\Big(\sum_{j=1}^{n} X_j\Big),$$

for all sequences (X_j) as above? In other words we would like to prove that within a certain class of Orlicz functions Φ, given an integer n, and nonnegative numbers $\sigma_1, \sigma_2, \dots, \sigma_n$, the minimum of

$$E\ \Phi\Big(\sum_{i=1}^{n} X_i\Big)$$

over all sequences of independent, symmetric random variables $X_1, \dots, X_n$ such that $E\,X_j^2 = \sigma_j^2$ is achieved when each of the X_j's is (at most) two valued.

Suppose for a moment that the inequality

$$E\ \Phi\Big(\sum_{j=1}^{n} \sigma_j\, \varepsilon_j\Big) \le E\ \Phi\Big(\sum_{i=1}^{n} X_i\Big)$$

holds. Then, since Φ is an even function, for any symmetric random variable X, with $\sigma^2 = E\,X^2$, we have

$$E\,\Phi(b + \sigma\,\varepsilon) = E\,\Phi(b\,\varepsilon' + \sigma\,\varepsilon) \le E\,\Phi(b\,\varepsilon' + X) = E\,\Phi(b + X),$$

where ε' is independent of ε and X. Denoting by E_ε integration with respect to the Rademacher function ε, the above inequality can be equivalently written as

$$E_\varepsilon\,\Phi(b + \sqrt{E\,X^2}\,\varepsilon) \le E\,\Phi(b + \sqrt{X^2}\,\varepsilon).$$

This implies that the necessary condition on Φ is that the function

$$\Psi_b(t) = E_\varepsilon\,\Phi(b + \sqrt{t}\,\varepsilon)$$

is convex on $\mathbf{R}^+$ for every $b \in \mathbf{R}^+$. It is now not hard to notice that it is also a sufficient condition for the minimality of Rademacher functions. Indeed, if $\Psi_b(\cdot)$ is convex on $[0, \infty)$ for a certain b, then for every nonnegative random variable Y with $E\,Y < \infty$, using Jensen's inequality, we get $\Psi_b(E\,Y) \le E\,\Psi_b(Y)$. Let X be a symmetric random variable with $E\,X^2 = \sigma^2 < \infty$ and let $Y = X^2$. Then

$$E\,\Phi(b + X) = E\,\Phi(b + \sqrt{Y}\,\varepsilon) = E\,\Psi_b(Y) \ge \Psi_b(E\,Y) = \Psi_b(\sigma^2) = E\,\Phi(b + \sigma\,\varepsilon).$$

Thus, if $\Psi_b(\cdot)$ is convex on $[0, \infty)$ for each b, then by conditioning, we get for any sequence $X_1, X_2, \dots, X_n$ of independent, symmetric random variables satisfying $E\,X_j^2 = \sigma_j^2 < \infty$,

$$\begin{aligned} E\ \Phi\Big(\sum_{j=1}^{n-1} X_j + X_n\Big) &\ge E\ \Phi\Big(\sum_{j=1}^{n-1} X_j + \sigma_n\,\varepsilon_n\Big) \\ = E\ \Phi\Big(\sum_{j=1}^{n-2} X_j + \sigma_n\,\varepsilon_n + X_{n-1}\Big) &\ge E\ \Phi\Big(\sum_{j=1}^{n-2} X_j + \sigma_n\,\varepsilon_n + \sigma_{n-1}\,\varepsilon_{n-1}\Big) \\ &\ge \cdots \ge E\ \Phi\Big(\sum_{j=1}^{n} \sigma_j\,\varepsilon_j\Big). \end{aligned}$$

Thus, the extremal problem is completely solved and the resulting convexity condition on Φ is that the function

$$t \to \Psi_b(t) = \frac{1}{2}\left(\Phi(b+\sqrt{t}) + \Phi(b-\sqrt{t})\right)$$

should be convex.

The next step is to find an analytic or "b free" characterization of those Orlicz functions Φ for which $\Psi_b(t)$ is convex for each b. As we shall see in Section 3, this condition is equivalent to $\Phi''(t)$ being continuous and convex on $\mathbf{R}$. Thus, we obtain the following

Theorem 1.1. *Let Φ be an Orlicz function and (σ_i) a sequence of nonnegative numbers. Then the inequality*

$$E\ \Phi\left(\sum_{j=1}^{n} \sigma_j\, \varepsilon_j\right) \le E\ \Phi\left(\sum_{j=1}^{n} X_j\right)$$

holds for all sequences (X_j) of independent symmetric random variables with $E\,X_i^2 = \sigma_i^2$ if and only if the function Φ is of class C^2 and the function Φ'' is convex on $\mathbf{R}$.

The above theorem was obtained independently by Pinelis and is contained in [P]. Within the class of functions $\Phi(t) = |t|^p$ the condition is satisfied if and only if $p \ge 3$ or $p = 2$. It should be noted here that for $p \ge 4$ the above result can be deduced from Theorem 5 of [U2]. Clearly, to obtain Whittle's result it suffices to take $X_j = \sigma_j g_j$, where (g_j) is a sequence of i.i.d. copies of a standard Gaussian random variable.

Our approach to Rosenthal's inequality follows the same pattern. We formulate an appropriate extremal problem and find conditions that would guarantee that the solution of the problem with given constraints is given by a linear combination of Rademacher functions. This requires a convexity condition on Φ and we characterize analytically that condition. We then use the minimality property of Rademacher functions (Theorem 1.1 above) to identify the best constant. Unfortunately, there are some problems here. Perhaps the most obvious obstacle is caused by the fact that Theorem 1.1 requires that the function Φ'' be convex, an assumption that is not satisfied for $2 < p < 3$. In the case of power functions one could escape this difficulty by using Haagerup's result, but the only known proof of the latter is quite involved and does not seem to extend to more general functions. For this reason we will take a different path. We will prove some additional inequalities for linear combinations of Rademacher functions that will enable us to find the exact value of D_p. Those inequalities, obtained in Section 4, reveal extreme properties of Rademacher random variables and we believe that they are of independent interest.

Another problem with Rosenthal's inequality is that it is not entirely clear what is the most convenient formulation of the extremal problem. Since Rosenthal's inequality involves two constraints; namely $\sum E\,\Phi(X_j)$ and $\sum E\,X_j^2$, it seems reasonable to subtract one of those two terms from $E\,\Phi(\sum X_j)$ and maximize the difference. This and some computer experiments suggested the following formulation: Given an Orlicz function Φ and positive numbers $\sigma_1, \ldots, \sigma_n$ find

$$\sup\left\{ E\,\Phi(\sum_{j=1}^{n} X_j) - \sum_{j=1}^{n} E\,\Phi(X_j) \right\},$$

where the supremum is taken over all sequences $\{X_j : j = 1, \dots, n\}$ of independent symmetric random variables with $E X_j^2 = \sigma_j^2$, $1 \le j \le n$. More precisely, we will seek a condition on Φ which will ensure that the solution of this extremal problem is given by $X_j = \sigma_j \epsilon_j$. Let us notice that in this formulation we want to solve the extremal problem with $\sigma_1, \dots, \sigma_n$ fixed, rather than having $\sigma = (\sum \sigma_i^2)^{1/2}$ fixed, as one should expect from the very form of Rosenthal inequality. Nonetheless, as will be seen later, in two cases in which we will identify the constant in a generalized form of Rosenthal inequality, the constant will depend only on σ and $\sum E\Phi(X_j)$.

We would like to mention that the point of view of Utev's paper is essentially the same; he deals with the problem of maximizing $E\,\Phi(\sum X_j)$ with $\sum E\,X_j^2$ and $\sum E\Phi(X_j)$ being fixed, and finds a (quite complicated) convexity condition under which the solution is given by three valued symmetric random variables. He then checks that that condition is satisfied for power functions as long as $p \ge 4$, and proves that (X_j) can be taken to be i.i.d. This allows one to identify the optimal value of D_p as the L_p–norm of a constant multiple of a symmetrized Poisson random variable. The constants of both multiplication and parameter depend on the values of $\sum E\,X_j^2$ and $\sum E\Phi(X_j)$. Since Utev's paper is practically unavailable, and since we have obtained a different proof of his result (as well as a similar result for nonnegative random variables), we plan to discuss this in a subsequent publication, now in preparation.

The paper is organized as follows. In the next section we discuss briefly some assumptions on Orlicz functions that are needed later and the basic properties they imply. The proof of Theorem 1.1 is finished in Section 3, and there we also derive a corollary to that theorem which generalizes a result of Haagerup ([H], Lemma 5.2). In Section 4 we prove some inequalities involving Rademacher functions that are used in subsequent sections. Section 5 contains a solution of the extremal problem related to Rosenthal's inequality, while in Sections 6 and 7 we derive the value of that constant (under some additional assumptions on Orlicz functions). Section 8 contains an example showing that the constants obtained are the best possible. We have made some effort to find the weakest possible conditions on Orlicz functions. Some negative results showing the impossibility of further extensions of several of our inequalities are gathered in Section 9.

2. Auxiliary facts about Orlicz functions

Let us recall that a convex, even function $\Phi : \mathbf{R} \to \mathbf{R}$ satisfying $\Phi(0) = 0$ is called an Orlicz function. An Orlicz function Φ is 2-convex if the function $t \to \Phi(\sqrt{t})$ is convex. In the case of power functions $\Phi(t) = |t|^p$, this means that $p \ge 2$. Since $\Psi(t)/t$ is an increasing function on $\mathbf{R}^+$ whenever Ψ is convex, our assumption on Φ implies that $\Phi(t)/t^2$ is an increasing function of t for $t > 0$. Hence there exists a finite limit

$$G = \lim_{t \to 0} \frac{\Phi(t)}{t^2},$$

and moreover one has $\Phi(t) \ge Gt^2$. If Φ is twice differentiable then $G = \frac{1}{2}\Phi''(0)$. In most interesting cases, including power functions other than the squaring function, G is equal to zero.

We have

Lemma 2.1. *Suppose that Φ is a 2-convex Orlicz function. Then for every b, x in $\mathbf{R}$ we have that*

$$E\,\Phi(b+x\epsilon) \geq \Phi(b)+\Phi(x).$$

Proof. For every real b, x, using Jensen's inequality and the superadditivity of the convex function $\Phi(\sqrt{x})$ one obtains

$$\begin{aligned} E\,\Phi(b+x\epsilon) =& E\,\Phi(\sqrt{(b+x\epsilon)^2}) \geq \Phi(\sqrt{E\,(b+x\epsilon)^2}) = \Phi(\sqrt{b^2+x^2}) \\ \geq& \Phi(\sqrt{b^2})+\Phi(\sqrt{x^2}) = \Phi(|b|)+\Phi(|x|) = \Phi(b)+\Phi(x). \end{aligned}$$

This completes the proof. □

From the above lemma, by easy induction, using Fubini's theorem, we obtain the following inequality sometimes referred to in this paper as the cotype inequality for Orlicz functions. For every sequence $X_1,\dots,X_n$ of independent symmetric random variables one has

$$E\,\Phi(\sum_{i=1}^{n} X_i) \geq \sum_{i=1}^{n} E\,\Phi(X_i).$$

Observe that if Ψ is an even nonnegative function on $\mathbf{R}$ such that $\Psi(\sqrt{x})$ is concave on $[0,\infty)$ and $\Psi(0)=0$, then the same argument shows that a reverse inequality holds.

Some problems which we consider involve a description of all those Orlicz functions which satisfy a particular family of inequalities. In certain cases the respective sets of Orlicz functions are easily seen to be convex cones; that is, if Φ_1 and Φ_2 are in the set then so is $a\Phi_1+b\Phi_2$ for any $a,b>0$. In some instances, checking that certain conditions are satisfied by all functions in a convex cone is somewhat simpler for specific functions Φ_a than for an arbitrary Orlicz function in the cone. In such cases it suffices to check the conditions for functions Φ_a that are on extreme rays of the closure (in an appropriate topology) of the given cone (cf. eg. [Ph], Chapter 2 for more details).

To illustrate this let us consider the closure in the pointwise topology of the cone of all Orlicz functions Φ such that $\Phi(\sqrt{t})$ is convex. The extreme points of that cone are functions of the form $\{\Phi_a;\, a\geq 0\}$, where $\Phi_a(t)=(t^2-a^2)_+$. Therefore, in order to prove Lemma 2.1 it would be enough to prove it for functions Φ_a; that is, to prove

Lemma 2.1′. *The inequality*

$$E\,\Phi(b+t\epsilon) \geq \Phi(b)+\Phi(t)$$

is valid for every Orlicz function Φ of the form $\Phi(t)=\Phi_a(t)=(t^2-a^2)_+$, where $a\geq 0$.

Proof. It suffices to consider the case where $0\leq b\leq t$. Fix $a\geq 0$ and put

$$G(a,b,t) = 2(E\,\Phi(b+t\epsilon)-\Phi(b)-\Phi(t))$$

where $\Phi=\Phi_a$. We will consider three cases.

If $t\leq a$, then $G(a,b,t)=\Phi(b+t)\geq 0$.

If $a\leq t\leq b+a$, then

$$\begin{aligned} G(a,b,t) =& \Phi(b+t)-2\Phi(t)-2\Phi(b) = a^2+b^2+2bt-t^2-2\Phi(b) \\ \geq& 2(b^2-\Phi(b))\geq 0. \end{aligned}$$

Finally, if $t \geq b + a$, then

$$G(a, b, t) = \Phi(b + t) + \Phi(t - b) - 2\Phi(t) - 2\Phi(b) = 2(b^2 - \Phi(b)) \geq 0,$$

which completes the proof. □

Of course, in this case the direct proof given in Lemma 2.1 is even simpler, but there are situations in which using the extreme rays of certain cones of functions will be useful, or is in fact, the only method of proof we are aware of. Let us briefly describe two cones we will need.

Consider the linear space $\mathcal{C}$ of even functions on $\mathbf{R}$ equipped with the pointwise topology. Let $\mathcal{O} \subset \mathcal{C}$ denote the cone of 2-convex Orlicz functions. Let $\mathcal{O}_{2,3}$ be the subcone of $\mathcal{O}$ consisting of those $\Phi \in \mathcal{O}$ such that the function Φ'' is concave on $(0, \infty)$. Also, consider the cone $\mathcal{O}_{2,4}$ consisting of those Orlicz functions $\Phi \in \mathcal{O}$ for which the function $t \to \Phi''(\sqrt{t})$ is concave on $\mathbf{R}^+$. Our notation is explained by the fact that the Orlicz function $\Phi(t) = |t|^p$ is in $\mathcal{O}_{2,i}$ if and only if $2 \leq p \leq i$, $i = 3, 4$. We will need the following

Lemma 2.2.

(a) $\Phi \in \mathcal{O}_{2,3}$ *if and only if it is in the pointwise closure of the cone of functions generated by the family* $\{\Phi_a : 0 \leq a \leq \infty\}$, *where* $\Phi_0(t) = t^2$ *and for* $a > 0$

$$\Phi_a(t) = |t|^3 - (|t| - a)_+^3 = \begin{cases} |t|^3, & \text{if } |t| \leq a; \\ a(3t^2 - 3a|t| + a^2), & \text{otherwise.} \end{cases}$$

(b) $\Phi \in \mathcal{O}_{2,4}$ *if and only if it is in the pointwise closure of the cone of functions generated by the family* $\{\Phi_a : 0 \leq a \leq \infty\}$, *where* $\Phi_0(t) = t^2$ *and for* $a > 0$

$$\Phi_a(t) = t^4 - (|t| - a)_+^3(3a + |t|) = \begin{cases} t^4, & \text{if } |t| \leq a; \\ a^2(6t^2 - 8a|t| + 3a^2), & \text{otherwise.} \end{cases}$$

Moreover, if $\Phi \in \mathcal{O}_{2,i}$, $i = 3, 4$, *then there exists a sequence* (F_n) *of convex combinations of the respective* Φ_a*'s such that the sequences* (F_n), (F_n'), (F_n'') *converge to* Φ, Φ', Φ'', *respectively, uniformly on compact subsets of* $\mathbf{R}$.

Proof. We will prove part (a) only. The second part follows an analogous pattern, using at some moment the substitution $x \longleftrightarrow x^2$.

For $f \in C^2(\mathbf{R})$, $g \in C(\mathbf{R})$ and $s \in \mathbf{R}$, we let

$$D^2(f) = f'', \qquad \text{and} \qquad Hg(s) = \int_0^s (s - t)g(t)\, dt.$$

Clearly, $D^2(Hg) = g$, if $g \in C(\mathbf{R})$, and $H(D^2(\Phi)) = \Phi$, if $\Phi \in \mathcal{O}_{2,3}$. Observe that $\Phi_a = H(\phi_a)$, where

$$\phi_a(x) = D^2(\Phi_a)(x) = \begin{cases} 6\min(|x|, a), & \text{if } 0 < a \leq \infty; \\ 2, & \text{if } a = 0. \end{cases}$$

Now, let $(a_i)_{i=0}^{\infty}$ be an enumeration of the nonnegative rational numbers (say, with $a_0 = 0$). Given $\Phi \in \mathcal{O}_{2,3}$ and $n \geq 0$, let $F_n = Hf_n$, where $f_n \in C(\mathbf{R})$ is a unique piecewise linear, even function with nodes only at $N_n = \{\pm a_i : 0 \leq i \leq n\}$, such that $f_n(a_i) = D^2(\Phi)(a_i)$ for $0 \leq i \leq n$ and $f_n(t) = D^2(\Phi)(a_j)$ for $t > a_j = \sup N_n$. Since $D^2(\Phi)$ is nonnegative, even and concave on $[0, \infty)$, so are the f_n's.

It follows that one can write $f_n = \sum_{i=0}^{n} s_{i,n}\phi_{a_i}$ where $s_{i,n} \geq 0$ for $0 \leq i \leq n < \infty$ and hence

$$H(f_n) = \sum_{i=0}^{n} s_{i,n} H(\phi_{a_i}) = \sum_{i=0}^{n} s_{i,n}\Phi_{a_i}.$$

It is easy to verify that $f_n \to D^2(\Phi)$ uniformly on compact subsets of $\mathbf{R}$ and this implies the almost uniform convergence of (F_n) and (F_n') to Φ and Φ', respectively. Thus the sequence (F_n) has all the required properties.

It remains to prove that $\mathcal{O}_{2,3}$ is a closed subset of $\mathbf{R}^{\mathbf{R}}$ in the topology of pointwise convergence. To this end, let $(\Phi_\alpha)_{\alpha\in A}$ be a net of elements of $\mathcal{O}_{2,3}$ which converges pointwise to a function Φ. We have to prove that $\Phi \in \mathcal{O}_{2,3}$.

One can show that then there is $\phi \in C(\mathbf{R})$ such that the net $(D^2(\Phi_\alpha))$ converges to ϕ almost uniformly on $\mathbf{R}\setminus\{0\}$, and $H\phi = \Phi$. Since it is clear that $\phi \in D^2(\mathcal{O}_{2,3})$, we can conclude that $\Phi = H\phi \in \mathcal{O}_{2,3}$. However, we shall verify only a weaker fact.

Clearly, we may assume that $\sup_\alpha \Phi_\alpha(1) = M < \infty$. Then, for each α, one has

$$\frac{1}{6}\Phi_\alpha''(1) = \int_0^1 x(1-x)\Phi_\alpha''(1)\,dx \leq \int_0^1 (1-x)\Phi_\alpha''(x)\,dx = \Phi_\alpha(1) \leq M,$$

and hence one has $0 \leq \Phi_\alpha''(x) \leq 6M\max(1,|x|) < \infty$, for $x \in \mathbf{R}$. Therefore the sets $\{\Phi_\alpha' : \alpha \in A\}$ and $\{\Phi_\alpha : \alpha \in A\}$ are uniformly bounded and equicontinuous on each compact subset of $\mathbf{R}$. In particular, using Ascoli's theorem and the diagonal procedure, we can select a sequence (Φ_{α_n}) such that $\lim_n \Phi_{\alpha_n}(x) = \Phi(x)$ for each $x \in \mathbf{R}$. Now, using Helly's theorem, we can find a further subsequence (f_n) of (Φ_{α_n}) such that $\phi(x) = \lim_n D^2 f_n(x)$ exists for each $x \in (-\infty,0)\cup(0,\infty)$. Putting $\phi(0) = \inf_{x\neq 0}\phi(x)$, we have $\phi \in C(\mathbf{R})$, because ϕ is an even, nonnegative function that is concave in $(0,\infty)$ and continuous at 0. Obviously, for each x, one has $(H\phi)(x) = H(\lim_n D^2 f_n)(x) = \lim_n (HD^2 f_n)(x) = \lim_n f_n(x) = \Phi(x)$. This shows that $\Phi \in \mathcal{O}_{2,3}$ and completes the proof. □

As we will see later on (cf. Theorem 5.1 and the discussion following it) we will link the cone $\mathcal{O}_{2,4}$ to a "convexity axiom" on Φ that will appear naturally in the context of an extremal problem for Rosenthal's inequality.

3. More on Khintchine's Inequality

In this section we will complete the proof of Theorem 1.1. In view of the discussion in Section 1 it suffices to prove the following:

Proposition 3.1. *Let Φ be an Orlicz function. The following two conditions are equivalent:*

(a) *For every b in $\mathbf{R}^+$ the function $t \to E\,\Phi(b+\sqrt{t}\epsilon)$ is convex.*
(b) *the function Φ is of class C^2 and the function Φ'' is convex on $\mathbf{R}$.*

Remark. A third equivalent condition could be added; namely,

(c) *Φ is in the pointwise closure of the cone generated by the family of functions $\{\Phi_a : a \geq 0\}$, where*

$$\Phi_a(t) = (t^2 - a^2)_+.$$

This fact, however, will not be used in the paper.

Proposition 3.1 is an immediate consequence of the following two lemmas.

Lemma 3.2. *Let Φ be an Orlicz function. Given $b > 0$, consider the following conditions:*

(i) the function $x \mapsto \Phi(b + \sqrt{|x|}) + \Phi(b - \sqrt{|x|})$ is convex on $\mathbf{R}$.

(ii) Φ is differentiable on $\mathbf{R}$ and the function $x \mapsto \frac{1}{x}(\Phi'(b + x) - \Phi'(b - x))$ is increasing in $(0, \infty)$.

(iii) Φ is twice differentiable on $\mathbf{R}$ and for each $x \in \mathbf{R}$ one has

$$x(\Phi''(b + x) + \Phi''(b - x)) - (\Phi'(b + x) - \Phi'(b - x)) \geq 0.$$

(iv) Φ is of class C^2 on $\mathbf{R}$ and the function Φ'' is convex on $\mathbf{R}$.

Then

$$(iv) \Rightarrow (iii) \Rightarrow (ii) \Rightarrow (i).$$

Proof. The implications $(iii) \Rightarrow (ii) \Rightarrow (i)$ are routine, while $(iv) \Rightarrow (iii)$ follows from the estimate

$$\Phi'(b + x) - \Phi'(b - x) = \int_{b-x}^{b+x} \Phi''(u)\, du \leq 2x \frac{\Phi''(b + x) + \Phi''(b - x)}{2}.$$

□

We conclude the proof of Proposition 3.1 by showing that if the condition (i) holds for every $b > 0$ then (iv) is satisfied.

Lemma 3.3. *Let Φ be an Orlicz function that satisfies condition (i) for every $b > 0$. Then Φ'' is continuous and convex on $\mathbf{R}$; that is, Φ satisfies (iv).*

Proof. If one assumes in addition that Φ is of class C^2, then the proof is very simple. Namely, let $(b_n)_{n=1}^{\infty}$ be a sequence of positive numbers tending to 0. Put $f_n(x) = b_n^{-2}(\Phi(\sqrt{|x|} + b_n) + \Phi(\sqrt{|x|} - b_n) - 2\Phi(b_n))$. Then each f_n is a convex function on $\mathbf{R}$ and hence so is the pointwise limit $\lim_{n\to\infty} f_n(x) = \Phi''(\sqrt{|x|})$.

The argument which avoids unnecessary assumptions is somewhat lengthier. Let ϕ be a nonnegative C^4 function on $\mathbf{R}$ such that $\phi(x) = 0$ if $|x| > \frac{1}{2}$ and $\int_{\mathbf{R}} \phi(x)\, dx = 1$. Put for $x \in \mathbf{R}$ and $n = 1, 2, \ldots$

$$\Phi_n(x) = \int_{\mathbf{R}_+} n\phi(n(u - 1))\Phi(\frac{x}{u})\frac{du}{u}.$$

Clearly, the Φ_n's are even functions that have four continuous derivatives in $\mathbf{R}\setminus\{0\}$. It is also clear that $\lim_{n\to\infty} \Phi_n(x) = \Phi(x)$ for $x \in \mathbf{R}$.

Notice that, if $\Phi(x) = |x|^3$, then the Φ_n's are multiples of Φ, hence they do not have a third derivative at 0. We will see, however, that the Φ_n's are of class C^2 on $\mathbf{R}$.

Observe first that the Φ_n's satisfy condition (i) for every $b > 0$. Indeed, one has

$$\Phi_n(b + \sqrt{x}) + \Phi_n(b - \sqrt{x}) = \int_{\mathbf{R}_+} n\phi(n(u - 1))\Psi(x, u)\frac{du}{u},$$

where

$$\Psi(x, u) = \Phi(\frac{b}{u} + \sqrt{\frac{x}{u^2}}) + \Phi(\frac{b}{u} - \sqrt{\frac{x}{u^2}}).$$

By our assumption, for every $u > 0$, the function $\Psi(\cdot, u)$ is convex in $[0, \infty)$. It follows that Φ_n satisfies condition (i) for every $b > 0$.

It is now obvious that $\Phi_n'(0) = 0$. Indeed, if we had $c = \lim_{x\to+0} \Phi_n(x)/x \neq 0$, then the function $x \mapsto \lim_{s\to+0} \Phi_n(sx)/s = c|x|$ would also satisfy condition (i) for

$b > 0$, which is false. Using the monotonicity and the Darboux property of Φ_n', we conclude that Φ_n is of class C^1 on $\mathbf{R}$.

Now observe that Φ_n'' is convex in $(0,\infty)$. Indeed, suppose that for some $x_0 > 0$ we had $(\Phi_n'')''(x_0) < 0$ (here we use the C^4 condition). Then the function Φ_n'' would be strictly concave in some neighborhood U of x_0. Pick numbers $0 < x < b$ so that $[b-x, b+x] \subset U$. Then we obtain

$$\Phi_n'(b+x) - \Phi_n'(b-x) = \int_{b-x}^{b+x} \Phi_n''(u)\,du \geq \frac{(b+x)-(b-x)}{2}(\Phi_n''(b-x) + \Phi_n''(b+x)).$$

On the other hand, differentiating at $t = x$ the function

$$t \mapsto g(t) = (\Phi_n'(b+t) - \Phi_n'(b-t))/t,$$

which by (ii) is nondecreasing, we obtain that

$$0 \leq g'(x) = x(\Phi_n''(b-x) + \Phi_n''(b+x)) - (\Phi_n'(b+x) - \Phi_n'(b-x))/x^2.$$

This contradiction shows that Φ_n'' is convex in $(0,\infty)$. It follows that Φ_n''' is nondecreasing in $(0,\infty)$.

Suppose that $\Phi_n'''(x_0) < 0$ for some $x_0 > 0$. Then Φ_n'' is decreasing in the interval $(0, x_0]$. It follows that $\Phi_n'(x_0) = \Phi_n'(x_0) - \Phi_n'(0) > x_0\Phi_n''(x_0)$. This, however, contradicts the fact that the function $x \mapsto \Phi_n'(x)/x$ has a nonnegative derivative in $(0,\infty)$.

Thus we have proved that Φ_n'' is nondecreasing in $(0,\infty)$. It follows that Φ_n' is convex and nonnegative in $[0,\infty)$. Since $\Phi_n'(0) = 0$, there is a limit $C = \lim_{x\to+0} \Phi_n'(x)/x$. Using L'Hospital's rule we find that $C = \lim_{x\to+0} \Phi_n''(x)$. Since Φ_n' is odd and Φ_n'' is even, this proves that there exists $\Phi_n''(0) = C$ and the function Φ_n'' is continuous at 0.

Hence, using Helly's theorem, we can find a subsequence $(n_i)_{i=1}^{\infty}$ of positive integers so that each of the sequences (Φ_{n_i}'), (Φ_{n_i}''), (Φ_{n_i}''') converges pointwise on $\mathbf{R}$. Using the equicontinuity, one obtains easily that the sequences (Φ_{n_i}'') and (Φ_{n_i}') converge almost uniformly on $\mathbf{R}$. (The pointwise convergence of the sequence of monotone functions (Φ_{n_i}''') provides an easy argument to show that the functions (Φ_{n_i}'') are equicontinuous on every bounded interval.) It follows that the function Φ has two continuous derivatives on $\mathbf{R}$. Moreover, the even function Φ'' is convex on $\mathbf{R}$ and nondecreasing in $[0,\infty)$, being the limit of the sequence (Φ_{n_i}''). This completes the proof of Lemma 3.3 and thus also of Proposition 3.1. □

We would like to close this section with a result that is a direct consequence of Theorem 1.1 and which extends Lemma 5.2 of [H].

Corollary 3.4. *Let $X = \sum_{i=1}^n a_i\epsilon_i$ and suppose that Φ is an Orlicz function of class C^2 and that Φ'' is convex. Then*

$$E\,\Phi''(X)E\,X^2 \geq E\,\Phi'(X)X.$$

Proof. Let Y be a random variable such that $E\,Y^2 = \sigma^2 = E\,X^2$ and $E\,XY = 0$. Observe that for each $\phi \in \mathbf{R}$ one has $E\,(\cos\phi X + \sin\phi Y)^2 = \sigma^2$. Put

$$f(\phi) = E\,\Phi(\cos\phi X + \sin\phi Y).$$

Since Φ is of class C^2, as long as Y satisfies an appropriate integrability condition we have

$$f'(\phi) = E\,\Phi'(\cos\phi X + \sin\phi Y)(-\sin\phi X + \cos\phi Y),$$

and

$$\begin{aligned} f''(\phi) =& E\,\Phi''(\cos\phi X + \sin\phi Y)(-\sin\phi X + \cos\phi Y)^2 \\ &+ E\,\Phi'(\cos\phi X + \sin\phi Y)(-\cos\phi X - \sin\phi Y), \end{aligned}$$

hence

$$f'(0) = E\,\Phi'(X)Y, \qquad f''(0) = E\,\Phi''(X)Y^2 - E\,\Phi'(X)X.$$

Thus if f has a local minimum at 0, then

$$E\,\Phi'(X)Y = 0, \qquad E\,\Phi''(X)Y^2 \geq E\,\Phi'(X)X.$$

Now suppose that $Y = \sum_{i=1}^n a_i\epsilon_i'$ is independent of X and identically distributed as X and that Φ'' is a convex function on $\mathbf{R}$. By Theorem 1.1, the function f has a local minimum at 0, because for each fixed ϕ the random variables $a_1(\cos\phi\epsilon_1 + \sin\phi\epsilon_1'), \ldots, a_n(\cos\phi\epsilon_n + \sin\phi\epsilon_n')$ have the same variances as the sequence $a_1\epsilon_1, \ldots, a_n\epsilon_n$. It follows that

$$E\,\Phi''(X)E\,X^2 \geq E\,\Phi'(X)X,$$

as required.

□

4. Rademacher Inequalities

In this section we establish some inequalities involving the behavior of linear combinations of Rademacher random variables in certain Orlicz spaces. Those inequalities will be used later and we believe that they are interesting in their own right. We begin with a technical lemma:

Lemma 4.1. *If Φ is an Orlicz function such that Φ'' is a concave function in $[0,\infty)$, then for each $b \geq 0$ the following inequality holds for all $0 \leq v \leq u$*

$$E\,\Phi'(b + u\epsilon + v\epsilon')(u\epsilon' - v\epsilon) + v\Phi'(u) - u\Phi'(v) \geq 0.$$

Proof. By Lemma 2.2 (a), it suffices to prove the inequality for the functions $\{\Phi_a : 0 < a < \infty\}$, where for $x \in \mathbf{R}$

$$\Phi_a(x) = |x|^3 - (|x| - a)_+^3.$$

Put for $0 \leq v \leq u$ and $b \geq 0$

$$g_{u,v}(b) = E\,\Phi_a'(b + u\epsilon + v\epsilon')(u\epsilon' - v\epsilon).$$

Hence we should prove that $g_{u,v}(b) + v\Phi_a'(u) - u\Phi_a'(v) \geq 0$. Observe first that Φ_a' is convex in $[0,\infty)$, because Φ_a'' is nondecreasing in $[0,\infty)$. It follows that $\Phi_a'(v) \leq (v/u)\Phi_a'(u)$.

Observe that if $b \geq a + u + v$, then one has

$$g_{u,v}(b) = E\,(6a(b + u\epsilon + v\epsilon') - 3a^2)(u\epsilon' - v\epsilon) = 0.$$

Since $v\Phi_a'(u) - u\Phi_a'(v) \geq 0$, our inequality is true for every $b \geq a + u + v$. Since $g_{u,v}$ is a C^1 function, then the proof of the inequality can be reduced to the case where either $b = 0$ or else $b \in (0, a + u + v)$ is a critical point of the function $g_{u,v}$.

We will find all such critical points b and verify the inequality in each case. Since for every b one has $g_{u,u}(b) = g_{u,0}(b) = 0$, we will assume below that $0 < v < u$. We obtain easily that

$$\begin{aligned}\frac{d}{db}g_{u,v}(b) =& E\,\Phi_a''(b+u\epsilon+v\epsilon')(u\epsilon'-v\epsilon)\\ =&\frac{1}{2}(u-v)E\,\Phi_a''(b+(u+v)\epsilon)\epsilon-\frac{1}{2}(u+v)E\,\Phi_a''(b+(u-v)\epsilon)\epsilon.\end{aligned}$$

Let $l_{b,s}$ denote the unique straight line passing through the points $(b-s, \Phi_a''(b-s))$ and $(b+s, \Phi_a''(b+s))$. Thus, if b is a critical point of $g_{u,v}$, then the slopes of the lines $l_{b,u-v}$ and $l_{b,u+v}$ are equal. This observation leads to an easy classification of all such cases, because Φ_a'' is a rather simple function; namely, $\Phi_a''(x) = 6\min\{|x|, a\}$.

First consider the case where both slopes are zero. Since $b > 0$, the line $l_{b,s}$ cannot be horizontal unless the set $S = \{b-s, b+s\}$ is disjoint from the interval $(-a, a)$. It follows that for $x \in S$ one has $\Phi_a'(x) = 6ax - 3a^2\mathrm{sign}(x)$. Thus we find that

$$g_{u,v}(b) = E\,\Phi_a'(b+u\,\varepsilon\,+v\,\varepsilon\,')(u\,\varepsilon\,'-v\,\varepsilon\,) = -3a^2E\ \mathrm{sign}(b+u\epsilon+v\epsilon')(u\epsilon'-v\epsilon).$$

Since $b < a+u+v$, we have $b-(u+v) \le -a$, hence only two subcases need to be distinguished: $b-(u-v) \le -a$, or else $b-(u-v) \ge a$. The first subcase is easy, because it implies that $g_{u,v}(b) = \frac{3}{2}a^2((v-u)+(u+v)) = 3a^2v > 0$.

In the second subcase we have $g_{u,v}(b) = \frac{3}{2}a^2(v-u) < 0$. However, in this subcase we have $u > v \ge a$, because $2v = b-(u-v)-(b-(u+v)) \ge a-(-a)$. Therefore $v\Phi_a'(u) - u\Phi_a'(v) = v(6au-3a^2) - u(6av-3a^2) = 3a^2(u-v) > -g_{u,v}(b)$.

Now consider the case where both slopes are equal to 1. This is possible if and only if $0 \le b-(u+v) < b+u+v \le a$. It follows that in this case

$$g_{u,v}(b) = E\,\Phi_a'(b+u\epsilon+v\epsilon')(u\epsilon'-v\epsilon) = 3E\,(b+u\epsilon+v\epsilon')^2(u\epsilon'-v\epsilon) = 0.$$

A special case results if $b = a$. Then the slope of the line $l_{b,s}$ equals $\frac{1}{2}$, for $0 < s \le a$, and hence it does not depend on s. In terms of u and v this corresponds to the case where $0 < 2v \le u+v \le a = b$. Observe that, since $\Phi_a'(a+s)-\Phi_a'(a-s) = 6a(a+s) - 3a^2 - 3(a-s)^2 = 12as - 3s^2$, we have

$$\begin{aligned}g_{u,v}(b) =& E\,\Phi_a'(b+u\epsilon+v\epsilon')(u\epsilon'-v\epsilon)\\ =&\frac{1}{4}((u-v)(\Phi_a'(a+u+v)-\Phi_a'(a-u-v))\\ &\qquad -(u+v)(\Phi_a'(a+u-v)-\Phi_a'(a-u+v)))\\ =&\frac{1}{4}(u-v)(u+v)(-3(u+v)+3(u-v)) = -\frac{3}{2}v(u-v)(u+v).\end{aligned}$$

Hence in this case we have

$$g_{u,v}(b) + v\Phi'(u) - u\Phi'(v) = -\frac{3}{2}v(u-v)(u+v) + 3(vu^2-uv^2) = \frac{3}{2}v(u-v)^2 > 0.$$

Now an elementary geometric argument (or a simple computation) shows that no other case is possible (besides those we have listed) in which $0 < b < a+u+v$, $0 < v < u$ and the lines $l_{b,u+v}$ and $l_{b,u-v}$ are parallel.

Hence the proof of our inequality will be complete once we verify it in the case where $b = 0$ and $0 < v < u$. Observe that, since Φ_a' is an odd function, we have

$$g_{u,v}(0) = E\,\Phi_a'(u\epsilon+v\epsilon')(u\epsilon'-v\epsilon) = \frac{1}{2}(\Phi_a'(u+v)(u-v) - \Phi_a'(u-v)(u+v)).$$

Put

$$h(a,u,v)=\frac{1}{2}(\Phi'_a(u+v)(u-v)-\Phi'_a(u-v)(u+v))+v\Phi'_a(u)-u\Phi'_a(v).$$

Since $h(0,u,v)=0$ and h is of class C^1, it will suffice if we show that $\frac{\partial}{\partial a}h(a,u,v)\geq 0$ for $a>0$, $0<v<u$.

Note that $\frac{\partial}{\partial a}\Phi'_a(x)=6(x-a)_+$ for $x\geq 0$. Therefore,

$$\begin{aligned}\frac{1}{3}\frac{\partial}{\partial a}h(a,u,v)&=(u-v)(u+v-a)_+-(u+v)(u-v-a)_+\\&\quad+2v(u-a)_+-2u(v-a)_+.\end{aligned}$$

If $0<a\leq\min\{u-v,v\}$, then the above expression is positive, being equal to $2au$. If $a\geq\max\{u-v,v\}$, then the expression is nonnegative. Finally, if $v<a<u-v$, then the expression equals $2au+2u(v-a)=2uv>0$ and if $u-v<a<v$, then it equals $2au+(u+v)(u-v-a)=(u+v+a)(u-v)>0$. This completes the proof of the lemma. □

Now we are ready to prove the following:

Proposition 4.2. *If Φ is an Orlicz function such that Φ'' is concave in $[0,\infty)$, then for each $n\geq 1$ and b, $x_1,\dots,x_n\in\mathbf{R}$ one has*

$$\begin{aligned}E\,\Phi(b+x\epsilon)-\Phi(x)&\leq E\,\Phi(b+\sum_{i=1}^n x_i\epsilon_i)-\sum_{i=1}^n\Phi(x_i)\\&\leq E\,\Phi(b+\frac{x}{\sqrt{n}}\sum_{i=1}^n\epsilon_i)-n\Phi(\frac{x}{\sqrt{n}}),\end{aligned}$$

where $x=(\sum_{i=1}^n x_i^2)^{1/2}$.

Proof. Both estimates follow easily from the following fact, which is a consequence of Lemma 4.1. If u, v are subject to the constraint $u^2+v^2=r^2$, then the value of the expression $E\,\Phi(b+u\epsilon+v\epsilon')-(\Phi(u)+\Phi(v))$ is an increasing function of v along every arc of the form $u^2+v^2=r^2$ contained in the region $u\geq v\geq 0$.

Indeed, the derivative of that expression along the vector field $-v\frac{\partial}{\partial u}+u\frac{\partial}{\partial v}$ is simply

$$\begin{aligned}&(-v\frac{\partial}{\partial u}+u\frac{\partial}{\partial v})(E\,\Phi(b+u\epsilon+v\epsilon')-(\Phi(u)+\Phi(v)))\\&\quad=E\,\Phi'(b+u\epsilon+v\epsilon')(u\epsilon'-v\epsilon)+v\Phi'(u)-u\Phi'(v),\end{aligned}$$

which is nonnegative by Lemma 4.1.

Now the lower estimate in Proposition 4.2 can be obtained by induction with respect to n, provided that the case where $n=2$ has been established. So let $n=2$. Since Φ is an even function, we may assume without loss of generality that $b\geq 0$. We put $u=\max\{|x_1|,|x_2|\}$, $v=\min\{|x_1|,|x_2|\}$ and observe that the expression $E\,\Phi(b+u\epsilon+v\epsilon')-(\Phi(u)+\Phi(v))$ is minimized when $v=0$.

To prove the upper estimate we may assume that all the x_i's are nonnegative. If one of them, say x_k, is greater than $x/\sqrt{n}$, then another, say x_j, must be less than $x/\sqrt{n}$. Let $\sum'$ be the sum extended over $\{1,2,\dots,n\}\setminus\{j,k\}$. Using monotonicity of the expression $E\,\Phi(b+u\epsilon+v\epsilon')-(\Phi(u)+\Phi(v))$ with $u=x_k$, $v=x_j$ and b replaced by $\sum' x_i\epsilon_i$, we can replace our combination $\sum x_i\epsilon_i$, by another one $\sum x'_i\epsilon_i$, so that the L_2-norm is preserved, but $E\,\Phi(\sum x'_i\epsilon_i)\geq E\,\Phi(\sum x_i\epsilon_i)$ and the number of those

coefficients x_i which are equal to $x/\sqrt{n}$ is increased either by 1 or by 2. If not all of the x_i''s are equal to $x/\sqrt{n}$, then we repeat this procedure. After at most $n-1$ applications of the procedure, we must come to the linear combination with all the coefficients equal to $x/\sqrt{n}$. This completes the proof. □

In fact, the above proof shows even more, namely that the function

$$(x_1, \dots, x_n) \mapsto E\,\Phi(b + \sum_{i=1}^n \sqrt{x_i}\epsilon_i) - \sum_{i=1}^n \Phi(\sqrt{x_i})$$

is Schur-concave. Indeed, we have verified Schur's condition (cf. Theorem 3.A.4 in [MO]).

Proposition 4.2 yields the following:

Corollary 4.3. *If Φ is an Orlicz function such that Φ'' is concave in $[0,\infty)$, then for each $n \geq 1$, b, $x_1, \dots, x_n, y_1, \dots, y_n \in \mathbf{R}$ and $t_1, \dots, t_n \geq 0$ such that $\sum_{i=1}^n t_i = 1$ and $\sum_{i=1}^n t_i y_i^2 = \sum_{i=1}^n x_i^2$, one has*

$$\sum_{i=1}^n t_i E\,(\Phi(b + y_i\epsilon_i) - \Phi(y_i)) \leq E\,\Phi(b + \sum_{i=1}^n x_i\epsilon_i) - \sum_{i=1}^n \Phi(x_i).$$

Proof. By the preceding proposition, the right-hand side is greater than or equal to

$$E\,\Phi(b + (\sum_{i=1}^n t_i y_i^2)^{1/2}\epsilon) - \Phi((\sum_{i=1}^n t_i y_i^2)^{1/2}).$$

The latter number is greater than or equal to the left-hand side, because our assumption on Φ implies that the function $\phi(s) = E\,\Phi(b + \sqrt{s}\epsilon) - \Phi(\sqrt{s})$ is concave (see e.g. Proposition 5.3 below). Hence Jensen's inequality can be applied. □

Remark 4.4. Here is one possible interpretation of Corollary 4.3 that reveals the extremal character of Rademacher random variables: Suppose that $Y_1, \dots, Y_n$ are disjointly supported simple symmetric random variables, and for $i = 1, \dots, n$, Y_i takes on the values $\pm y_i$, each with probability $t_i/2$, and 0 with probability $1 - t_i$. Let $X_i = x_i\epsilon_i$, $i = 1, \dots, n$, satisfy $E\sum_{i=1}^n X_i^2 = E\sum_{i=1}^n Y_i^2$; that is, $\{x_i\}$ satisfy the relation:

$$\sum_{i=1}^n t_i y_i^2 = \sum_{i=1}^n x_i^2.$$

Then we have

$$E\left\{\Phi(b + \sum_{i=1}^n Y_i) - \sum_{i=1}^n \Phi(Y_i)\right\} \leq E\left\{\Phi(b + \sum_{i=1}^n X_i) - \sum_{i=1}^n \Phi(X_i)\right\}.$$

We will also need the following "doubling" estimate:

Proposition 4.5. *Let Φ be a 2-convex Orlicz function such that Φ'' is concave. Then for nonnegative numbers b and σ we have*

$$E\,\Phi(b + \sigma\epsilon) - \Phi(\sigma) \leq E\,\Phi(b + \frac{\sigma}{\sqrt{2}}(\epsilon + \epsilon')) - 2\Phi(\frac{\sigma}{\sqrt{2}}).$$

Proof. This is immediate from Proposition 4.2. Alternatively, and more simply, compute the expectation on both sides and move all terms to the right–hand side to see that the above inequality is equivalent to

$$\Phi(b+\sqrt{2}\sigma)+\Phi(b-\sqrt{2}\sigma)+2(\Phi(b)-\Phi(b+\sigma)-\Phi(b-\sigma))$$
$$+4(\Phi(\sigma)-2\Phi(\frac{1}{\sqrt{2}}\sigma))\geq 0.$$

The next lemma shows that even stronger inequality holds. Note that, since $\Phi(u)/u^2$ is increasing, $\Phi(\sigma)-2\Phi(\sigma/\sqrt{2})\geq 0$. □

Lemma 4.6. *Let Φ be as in Proposition 4.5. Then, for all $b\geq 0$ and $\sigma\geq 0$,*

$$\Phi(b+\sqrt{2}\sigma)+\Phi(b-\sqrt{2}\sigma)+2(\Phi(b)-\Phi(b+\sigma)-\Phi(b-\sigma))$$
$$+2\sqrt{2}(\Phi(\sigma)-2\Phi(\frac{1}{\sqrt{2}}\sigma))\geq 0.$$

Proof. We will prove this lemma by verifying the inequality for the extreme functions Φ_a, where $a>0$ and

$$\Phi_a(t)=|t|^3-(|t|-a)_+^3,$$

of the cone $\mathcal{O}_{2,3}$ consisting of all 2-convex Orlicz functions Φ such that Φ'' is concave and nondecreasing.

Observe first that, thanks to the homogeneity property, $\Phi_{ca}(ct)=c^3\,\Phi_a(t)$ which holds for all $t\in\mathbf{R}$ and each choice of $a>0$ and $c>0$, we can reduce the number of the parameters in the inequality. Without loss of generality we will assume that $\sigma=1$.

The inequality to be proved is therefore $\Psi(a,b)\geq 0$, for every $a>0$ and $b\geq 0$, where by $\Psi(a,b)$ we have denoted the value of the expression

$$\Phi_a(b+\sqrt{2})+\Phi_a(b-\sqrt{2})+2(\Phi_a(b)-\Phi_a(b+1)-\Phi_a(b-1))$$
$$+2\sqrt{2}(\Phi_a(1)-2\Phi_a(\frac{1}{\sqrt{2}})).$$

Since $2\sqrt{2}\Phi_a(t)=\Phi_{\sqrt{2}a}(\sqrt{2}t)$, letting for $a,b\geq 0$ and $\lambda>0$

$$H(a,b,\lambda)=\lambda^{-2}(\Phi_a(b+\lambda)+\Phi_a(b-\lambda)-2\Phi_a(b)+\Phi_{\sqrt{2}a}(\lambda)),$$

we obtain that the estimate $\Psi(a,b)\geq 0$ will follow, as soon as we verify that $H(a,b,\sqrt{2})\geq H(a,b,1)$.

It follows from the homogeneity of the function H that there is a function Ξ such that, for $a,b\geq 0$ and $\lambda>0$, one has

$$\frac{\partial}{\partial\lambda}H(a,b,\lambda)=\Xi(\frac{a}{\lambda},\frac{b}{\lambda}).$$

Thus in order to prove that $H(a,b,\lambda)$ is a nondecreasing function of λ it suffices to verify that $\Xi(a,b)\geq 0$ for $a,b\geq 0$.

Here is an explicit formula for the function Ξ.

$$\Xi(a,b)=G(0)+2G(b)-G(a+b)-G(a-b)-G(\sqrt{2}a),$$

where G denotes the function on $\mathbf{R}$ defined by

$$G(x)=3(1-|x|)_+^2-2(1-|x|)_+^3=\begin{cases}1-3x^2+2|x|^3, & \text{if } 0\leq|x|\leq 1;\\ 0, & \text{otherwise.}\end{cases}$$

Before we proceed, let us observe that for any real A and all positive $\lambda \geq |A|$ one has

$$\frac{\partial}{\partial\lambda}(\lambda^{-2}((\lambda+A)^3 - 2A_+^3)) = G(\frac{A}{\lambda}).$$

The above formula for Ξ can be verified rather easily, if H is decomposed into parts that allow simple treatment. The following two formulas hold for $a, b \geq 0$ and $\lambda > 0$.

$$\begin{aligned}\frac{\partial}{\partial\lambda}(\lambda^{-2}\Phi_{\sqrt{2}a}(\lambda)) =& \frac{\partial}{\partial\lambda}\begin{cases} 3\sqrt{2}a - 6\lambda^{-1}a^2 + 2\sqrt{2}a^3\lambda^{-2}, & \text{if } \lambda > \sqrt{2}a \geq 0, \\ \lambda, & \text{if } 0 < \lambda \leq \sqrt{2}a; \end{cases} \\ =& G(0) - G(\frac{\sqrt{2}a}{\lambda}),\end{aligned}$$

$$\begin{aligned}\frac{\partial}{\partial\lambda}(\lambda^{-2}(|b+\lambda|^3 + |b-\lambda|^3 - 2b^3)) =& \frac{\partial}{\partial\lambda}\begin{cases} 2\lambda + 6b^2\lambda^{-1} - 2b^3\lambda^{-2}, & \text{if } \lambda > b \geq 0, \\ 6b, & \text{if } 0 < \lambda \leq b, \end{cases} \\ =& 2G(\frac{b}{\lambda}).\end{aligned}$$

The next two formulas hold for $a, b \geq 0$ and $\lambda > (b-a)_+$

$$\begin{aligned}\frac{\partial}{\partial\lambda}(\lambda^{-2}((|b-\lambda|-a)_+^3)) =& G(\frac{a+b}{\lambda}), \\ \frac{\partial}{\partial\lambda}(\lambda^{-2}((|b+\lambda|-a)_+^3 - 2(b-a)_+^3)) =& G(\frac{a-b}{\lambda}).\end{aligned}$$

Finally, if $b > a$ and $0 < \lambda \leq b - a$, then one has

$$\frac{\partial}{\partial\lambda}(\lambda^{-2}((|b-\lambda|-a)_+^3 + (|b+\lambda|-a)_+^3 - 2(b-a)_+^3)) = 0 = G(\frac{a+b}{\lambda}) + G(\frac{a-b}{\lambda}).$$

Let us show why the function Ξ is nonnegative for $a, b \geq 0$. Observe first that, since G is of class C^1 on its domain (actually it is piecewise polynomial), so is Ξ. Hence, if we show that Ξ is nonnegative on the "boundary" (its domain is unbounded), then we will reduce the problem to verifying that $\Xi(a,b) \geq 0$ at every critical point (a,b) of Ξ in the remaining set.

Now observe that G is an even function which is nonincreasing on $[0,\infty)$ and vanishes on $[1,\infty)$. Moreover, one has $G(1-x) = G(0) - G(x)$ for $0 \leq x \leq 1$.

If $b \geq 1 \geq a \geq 0$, then $G(b) = 0 = G(a+b)$ and $G(b-a) \leq G(1-a) = G(0) - G(a)$, hence

$$\Xi(a,b) \geq \Xi(a,1) \geq G(a) - G(\sqrt{2}a) \geq 0.$$

If $a \geq \frac{1}{\sqrt{2}}$ and $a + b \geq 1$, then $G(a+b) = 0 = G(\sqrt{2}a)$, and hence

$$\Xi(a,b) \geq G(0) - G(b-a) \geq 0.$$

If $\frac{1}{2} \leq a \leq a+b \leq 1$, then $0 \leq b \leq \frac{1}{2} \leq a$. So

$$\Xi(a,b) = (G(0) - G(a-b)) + (G(b) - G(a+b)) + (G(b) - G(\sqrt{2}a)) \geq 0,$$

since G is nonincreasing on $[0,\infty)$. If $0 \leq a \leq b \leq a+b \leq 1$, then

$$\Xi(a,b) = 12a^2 - 12a^2b - 4\sqrt{2}a^3 \geq (12 - 4\sqrt{2})a^3 \geq 0.$$

For each (a,b) such that $0 \leq b \leq a \leq \frac{1}{2}$ we have $\frac{\partial}{\partial b}\Xi(a,b) = 12b(b-2a) \leq 0$, and hence we infer that $\Xi(a,b) \geq \Xi(a,a) \geq 0$ in this region.

Therefore the proof that Ξ is nonnegative will be complete if we show that this function has no local minimum in the open set

$$\Omega = \{(a,b) : 0 < a < \frac{1}{\sqrt{2}},\ 0 < b < 1 < a+b\}.$$

In fact, the function Ξ has only one critical point in the set Ω; namely, $(a,b) = (\frac{2}{21}(5\sqrt{2}-1), \frac{2}{21}(3\sqrt{2}+5))$. That point is easily seen to be a saddle point of Ξ (the second order partial derivatives at (a,b) have different signs). This will complete the proof.

So suppose that $(a,b) \in \Omega$ is a critical point of Ξ. Recall that $G'(x) = 0$ if $|x| \geq 1$, and $G'(x) = 6x(|x|-1)$ if $|x| \leq 1$. The partial derivatives of Ξ at (a,b) are

$$\begin{aligned}\frac{\partial}{\partial a}\Xi(a,b) =& - G'(a+b) - G'(a-b) - \sqrt{2}G'(\sqrt{2}a) = 0,\\ \frac{\partial}{\partial b}\Xi(a,b) =& 2G'(b) - G'(a+b) + G'(a-b) = 0.\end{aligned}$$

Since $(a,b) \in \Omega$, we have $G'(a+b) = 0$. Thus the equations for the critical point yield

$$-G'(a-b) - \sqrt{2}G'(\sqrt{2}a) = 0, \quad 2G'(b) + G'(a-b) = 0.$$

It follows that $G'(a-b) = -2G'(b) > 0$, and hence $-1 < a-b < 0$, so that

$$G'(a-b) = 6(a-b)(b-a-1).$$

Since $G'(b) = 6(b-b^2)$ and $G'(\sqrt{2}a) = 6(\sqrt{2}a - 2a^2)$, the relation $2G'(b) = \sqrt{2}G'(\sqrt{2}a)$ implies that

$$b - b^2 = a - \sqrt{2}a^2.$$

Analogously, the condition $\frac{\partial}{\partial b}\Xi(a,b) = 0$ yields $2(b^2-b) + (-(a-b)^2 - (a-b)) = 0$, i.e.,

$$b^2 + 2ab - a^2 - a - b = 0.$$

Combining this with the equation $b - b^2 = a - \sqrt{2}a^2$, we obtain $(\sqrt{2}-1)a^2 + 2ab - 2a = 0$. Since $a > 0$, this yields a linear equation

$$\frac{1}{2}(\sqrt{2}-1)a + b = 1.$$

Using this we obtain that $a - \sqrt{2}a^2 = (1-b)b = \frac{1}{2}(\sqrt{2}-1)ab$ which, after dividing both sides by the positive number a, yields another linear equation, namely

$$\sqrt{2}a + \frac{1}{2}(\sqrt{2}-1)b = 1.$$

Solving the two linear equations we obtain $a = \frac{2}{21}(5\sqrt{2}-1)$ and $b = \frac{2}{21}(3\sqrt{2}+5)$. This completes the proof of Lemma 4.6. □

5. Extremal problem for Rosenthal's inequality

In this section we discuss an extremal problem that allows us to find the exact values of the constant D_p in Rosenthal's inequality for sums of independent, symmetric random variables:

$$M_p(\{X_i\}) \leq \|\sum X_i\|_p \leq D_p M_p(\{X_i\}),$$

where

$$M_p(\{X_i\}) = \max\left\{\left(\sum \|X_i\|_2^2\right)^{1/2}, \left(\sum \|X_i\|_p^p\right)^{1/p}\right\}.$$

We are also interested in a more general formulation: Given a 2-convex Orlicz function Φ and positive constants c_1 and c_2, find

$$\sup\left\{E\,\Phi(\sum X_j) : \sum E\,\Phi(X_j) \le c_1, \sum E\,X_j^2 \le c_2\right\}.$$

As mentioned in the introduction, we consider a related extremal problem (and we will show the relationship between these problems in Section 6): For an Orlicz function Φ and positive numbers $\sigma_1, \dots, \sigma_n$, find

$$\sup\left\{E\,\Phi(\sum_{j=1}^n X_j) - \sum_{j=1}^n E\,\Phi(X_j)\right\},$$

where the supremum is taken over all sequences $\{X_j : j = 1, \dots, n\}$ of independent symmetric random variables with $E\,\Phi(X_j) < \infty$ and $E\,X_j^2 = \sigma_j^2$, $1 \le j \le n$. More precisely, we shall characterize (under natural, rather weak, assumptions on Φ) those Φ such that, for each choice of n and positive numbers $\sigma_1, \dots, \sigma_n$, the solution of the above problem is given by appropriate multiples of Rademacher random variables; that is, that we have

$$(*) \qquad E\,\Phi(\sum_{j=1}^n X_j) - \sum_{j=1}^n E\,\Phi(X_j) \le E\,\Phi(\sum_{j=1}^n \sigma_j\,\varepsilon_j) - \sum_{j=1}^n \Phi(\sigma_j).$$

Note that we *do not* fix $\sum_{j=1}^n E\,\Phi(X_j)$ in advance. This is in contrast with Utev [U2], who found the best value of D_p for $p \ge 4$ by solving the following extremal problem: Given positive numbers A and D and a sequence of independent symmetric random variables (X_i) such that

$$\sum \|X_i\|_2^2 \le D^2 \quad \text{and} \quad \sum \|X_i\|_p^p \le A,$$

maximize $\|\sum X_i\|_p$.

It turns out that the solution of the latter problem is given by the L_p–norm of a multiple of a symmetrized Poisson random variable with both constant of multiplication and parameter depending on A and D. For an announcement of the results contained in [U2]; see [U1]. For p an even integer the question was settled in [PU].

In order to formulate a necessary condition on Φ, we put for $b, t \ge 0$

$$\Theta_b(t) = E_{\,\varepsilon}\,\Phi(b + \sqrt{t}\,\varepsilon) - \Phi(\sqrt{t}).$$

Lemma 5.1. *Assume that Φ is an even function on $\mathbf{R}$ such that $\Phi(0) = 0$ and (*) holds for $n = 2$, whenever X_1 is 2-valued and X_2 is k-valued with $k \le 4$. Then, for every $b \in \mathbf{R}^+$, the function Θ_b is concave on $\mathbf{R}^+$.*

Proof. Fix nonnegative numbers b, a_1, a_2, x_1, x_2 such that $a_1 + a_2 = 1$ and $x_1 \ne x_2$. We shall verify that $a_1\Theta_b(x_1) + a_2\Theta_b(x_2) \le \Theta_b(a_1x_1 + a_2x_2)$.

Put $\sigma_i = \sqrt{x_i}$ for $i = 1, 2$ and let $\sigma^2 = a_1\sigma_1^2 + a_2\sigma_2^2$, $\sigma \ge 0$. Let $X_1 = b\,\varepsilon$ and let X_2 be a symmetric random variable independent of X_1 such that $P(|X_2| = \sigma_i) = a_i$, $i = 1, 2$. Observe that $E\,X_2^2 = a_1\sigma_1^2 + a_2\sigma_2^2 = \sigma^2$, $E\,\Phi(X_1) = \Phi(b)$ and

$$a_1\Theta_b(x_1) + a_2\Theta_b(x_2) = E\,\Phi(X_1 + X_2) - E\,\Phi(X_2),$$

$$\Theta_b(a_1x_1 + a_2x_2) = \Theta_b(\sigma^2) = E\,\Phi(b\,\varepsilon + \sigma\,\varepsilon') - E\,\Phi(\sigma\,\varepsilon').$$

Thus the inequality $a_1\Theta_b(x_1)+a_2\Theta_b(x_2) \le \Theta_b(a_1x_1+a_2x_2)$ follows easily from (*). This proves the lemma. □

Conversely, assuming that the continuous even function Φ is bounded from below, it is easy to see that, if $\Theta_b(t)$ is concave for every $b \in \mathbf{R}^+$, then the estimate (*) holds. Indeed, if the function $\Theta_b(t)$ is concave for every $b \in \mathbf{R}^+$, then, for all independent, symmetric random variables Y and X such that $E\,\Phi(X) < \infty$, $E\,\Phi(Y) < \infty$ and $E\,X^2 = \sigma^2 < \infty$, we have

$$E\,\Phi(Y + X) - E\,\Phi(X) \le E\,\Phi(Y + \sigma\,\varepsilon\,) - \Phi(\sigma).$$

Therefore, if $\{X_i : i = 1, \dots, n\}$ are independent symmetric random variables satisfying $E\,\Phi(X_i) < \infty$ and $E\,X_i^2 = \sigma_i^2$, $i = 1, \dots, n$, then we have

$$\begin{aligned}
&E\ \Phi\Big(\sum_{j=1}^{n} X_j\Big) - \sum_{j=1}^{n} E\,\Phi(X_j)\\
=&E\ \Phi\Big(\sum_{j=1}^{n-1} X_j + X_n\Big) - E\,\Phi(X_n) - \sum_{j=1}^{n-1} E\,\Phi(X_j)\\
\le&E\ \Phi\Big(\sum_{j=1}^{n-1} X_j + \sigma_n\,\varepsilon_{\,n}\Big) - \Phi(\sigma_n) - \sum_{j=1}^{n-1} E\,\Phi(X_j)\\
=&E\ \Phi\Big(\sum_{j=1}^{n-2} X_j + \sigma_n\,\varepsilon_{\,n} + X_{n-1}\Big) - E\,\Phi(X_{n-1}) - \sum_{j=1}^{n-2} E\,\Phi(X_j) - \Phi(\sigma_n)\\
\le&\cdots\\
\le&E\ \Phi\Big(\sum_{j=1}^{n} \sigma_j\,\varepsilon_{\,j}\Big) - \sum_{j=1}^{n} \Phi(\sigma_j).
\end{aligned}$$

Our next goal is to find a verifiable characterization of even functions Φ for which the function

$$\Theta_b(t) = E_{\,\varepsilon}\,\Phi(b + \sqrt{t}\,\varepsilon\,) - \Phi(\sqrt{t}),$$

is concave on $\mathbf{R}^+$ for every $b \in \mathbf{R}^+$. We will show that, if Φ is continuous and bounded from below, then the latter property is equivalent to the condition that $\Phi''(\sqrt{t})$ is concave and nonnegative on $\mathbf{R}^+$. In particular, it will follow that Φ must then be a 2-convex Orlicz function (cf. Lemma 5.8 below). This will complete the proof of the following:

Theorem 5.2. *Let Φ be a continuous, even, bounded from below function on $\mathbf{R}$. Then the inequality*

$$E\,\Phi(\sum_{j=1}^{n} X_j) - \sum_{j=1}^{n} E\,\Phi(X_j) \le E\,\Phi(\sum_{j=1}^{n} \sigma_j\,\varepsilon_{\,j}) - \sum_{j=1}^{n} \Phi(\sigma_j),$$

holds for all sequences (X_j) of independent symmetric random variables such that $E\,\Phi(X_j) < \infty$ and $E\,X_j^2 = \sigma_j^2$ if and only if Φ is of class C^2 and the function $t \to \Phi''(\sqrt{t})$ is concave and nonnegative on $\mathbf{R}^+$.

Before passing to the proof, let us remark that the above theorem solves the extremal problem with $\sigma_1, \dots, \sigma_n$ fixed, rather than having $\sigma = (\sum \sigma_i^2)^{1/2}$ fixed, as in Utev's formulation of the extremal problem concerning Rosenthal's inequality. Nonetheless, as will be seen in two subsequent sections, in the two cases in which we will identify the constant in the generalized form of Rosenthal inequality, the constant will depend only on σ and $\sum E\Phi(X_j)$. Let us also remark that within the class $\Phi(t) = |t|^p$, $p > 0$, the assumption that $\Phi''(\sqrt{t})$ is concave is equivalent to $2 \le p \le 4$. On the other hand, the examples $\Phi(t) = -t^{2k}$ where $k = 1, 2, \dots$, show that there is a reason to make the assumption that $\inf_x \Phi(x) > -\infty$.

In view of the discussion preceding the statement of Theorem 5.2, in order to complete the proof it suffices to show that the conditions:

(a) $\Theta_b(t) = E_{\,\varepsilon}\, \Phi(b + \sqrt{t}\,\varepsilon\,) - \Phi(\sqrt{t})$, is concave on $\mathbf{R}^+$ for every $b \in \mathbf{R}^+$, and
(b) the function $t \to \Phi''(\sqrt{t})$ is concave on $\mathbf{R}^+$

are equivalent. We first prove

Proposition 5.3. *Let Φ be an even function on $\mathbf{R}$. Given $b > 0$, consider the following conditions:*

(i) the function $x \mapsto \Phi(b + \sqrt{x}) + \Phi(b - \sqrt{x}) - 2\Phi(\sqrt{x})$ is concave in $[0, \infty)$.
(ii) the function $x \mapsto \frac{1}{x}(\Phi'(b+x) - \Phi'(b-x) - 2\Phi'(x))$ is nonincreasing in $(0, \infty)$.
(iii) for each $x > 0$ one has

$$x(\Phi''(b + x) + \Phi''(b - x) - 2\Phi''(x)) - (\Phi'(b + x) - \Phi'(b - x) - 2\Phi'(x)) \le 0.$$

(iv) the function $x \mapsto \Phi''(\sqrt{x})$ is concave and nonnegative in $[0, \infty)$.

Then

$$(iv) \Rightarrow (iii) \Rightarrow (ii) \Rightarrow (i).$$

Proof. Only the implication $(iv) \Rightarrow (iii)$ needs a proof. We will make use of the extreme rays for the cone defined by (iv) [cf. Lemma 2.2(b)].

We should check that the following function $\Psi(x, b)$ is nonpositive for x, $b \ge 0$.

$$\Psi(x, b) = x(\Phi''(b + x) + \Phi''(x - b) - 2\Phi''(x)) - (\Phi'(b + x) + \Phi'(x - b) - 2\Phi'(x)).$$

By Lemma 2.2 (b), it will suffice if we do this in the case where $\Phi = \Phi_a$ for some $0 < a < \infty$ is an extreme element of the cone of those Orlicz functions Φ such that the $\Phi''(\sqrt{x})$ is concave on $[0, \infty)$. (Once this is known, the two missing cases, where $\Phi(x) = x^2$, and $\Phi(x) = x^4$ can either be checked directly or obtained in the limit as a tends to 0 or to ∞).

Thus $0 < a < \infty$ is fixed and we have for $x \ge 0$

$$\Phi(x) = x^4 - (x - a)_+^3(3a + x),$$

and hence

$$\Phi''(x) = 12 \min\{x^2, a^2\}, \qquad \Phi'(x) = \min\{4x^3, a^2(12x - 8a)\}.$$

Then for each $b \ge 0$ and every $x \ge 0$ one has

$$\Psi(x, b) \le \max\{\Psi(0, b), \Psi(a + b, b)\} = \max\{0, 0\} = 0.$$

The last inequality requires some explanation. Observe that, using the formula

$$\frac{\partial}{\partial x}\Psi(x, b) = x\,(\Phi'''(b + x) - \Phi'''(b - x) - 2\Phi'''(x))$$

which is valid for $x \in [0, \infty) \setminus \{|a - b|, a + b\}$, we obtain

$$\frac{1}{24x}\frac{\partial}{\partial x}\Psi(x,b) = \begin{cases} b+x, & b+x<a, \\ 0, & b+x>a, \end{cases} + \begin{cases} x-b, & |x-b|<a, \\ 0, & |x-b|>a, \end{cases} - \begin{cases} 2x, & x<a, \\ 0, & x>a, \end{cases}$$
$$= \begin{cases} x-b>0, & \max\{a,b\}<x<a+b, \\ \le 0, & \text{otherwise.} \end{cases}$$

If $b \ge 0$ is fixed, then the continuous function $x \mapsto \Psi(x,b)$ is increasing in the interval $(\max\{a,b\}, a+b)$ and nonincreasing in the complement of that interval. This proves our inequality for the case where Φ is an extreme element. □

To complete the proof of Theorem 5.2 we will prove the following:

Proposition 5.4. *Let Φ be a continuous, even, bounded from below function on $\mathbf{R}$ that satisfies condition (i) of Proposition 5.3 for every $b>0$. Then Φ'' is continuous and nonnegative on $\mathbf{R}$ and the function $x \mapsto \Phi''(\sqrt{x})$ is concave in $[0,\infty)$.*

We will need the following four lemmas.

Lemma 5.5. *Let h be a positive continuous function defined in an open interval $(a,b) \subseteq \mathbf{R}$. Suppose that f is a nonincreasing function in (a,b), and the function hf is nondecreasing in (a,b). Then the function f is continuous in (a,b).*

Proof. Let $x \in (a,b)$. Then, since $f(x) \ge \lim_{t\to x+0} f(t)$ and

$$h(x)f(x) \le \lim_{t\to x+0} h(t)f(t) = h(x) \lim_{t\to x+0} f(t),$$

we have $f(x) = \lim_{t\to x+0} f(t)$. The proof that $f(x) = \lim_{t\to x-0} f(t)$ is similar. □

Lemma 5.6. *Let h_1, h_2 be positive continuous functions defined in an open interval $(a,b) \subseteq \mathbf{R}$. Let $f_1, f_2, \dots$ be a sequence of real-valued functions that converges pointwise in (a,b) to a real-valued function f_0. Suppose the f_i's are twice differentiable in (a,b), and for every $i = 1,2,\dots$ the function $h_1 f_i''$ is nonincreasing and the function $h_2 f_i''$ is nondecreasing in (a,b). Then the function f_0 is twice differentiable in (a,b) and f_0'' is continuous. Moreover, the sequence (f_i'') converges to f_0'' almost uniformly in (a,b).*

Proof. Observe first that f_i'' is continuous in (a,b) for every $i = 1,2,\dots$. This follows by applying the previous lemma with $f = h_1 f_i''$ and $h = h_2/h_1$.

Note that the sequence $f_i''(x)$ is bounded for each $x \in (a,b)$. Indeed, let $x_0 \in (a,b)$. We will show that $\sup_i f_i''(x) < \infty$ (the proof that $\inf_i f_i''(x) > -\infty$ is analogous). Fix $x_1 \in (a, x_0)$. Observe that for each i there is $\xi_i \in (x_1, x_0)$ such that

$$f_i(x_1) + f_i(x_0) - 2f_i(\frac{x_1+x_2}{2}) = \frac{1}{4}(x_1-x_0)^2 f_i''(\xi_i).$$

Then the left-hand side is bounded by a constant, say M, because the sequence (f_i) is pointwise convergent. Hence we can estimate

$$f_i''(x_0)h_1(x_0) \le f_i''(\xi_i)h_1(\xi_i) \le 4M(x_1-x_0)^{-2} \sup\{h_1(x) : x \in [x_1, x_0]\}.$$

Hence, using Helly's theorem, we can extract a subsequence (n_i) of the indices such that the limit

$$j(x) = \lim_{i\to\infty} f_{n_i}''(x)h_1(x)$$

exists for each $x \in (a,b)$. Again it follows from the previous lemma that the function j is continuous in (a,b). Using the continuity and monotonicity of the functions j and $f''_{n_i}(x)h_1(x)$, $i = 1,2,\dots$, we obtain easily that the convergence $f''_{n_i}(x)h_1(x) \to j$ is almost uniform on (a,b).

It is now clear that the sequence $f''_{n_i}(x)$ converges almost uniformly to the function $k(x) = j(x)/h_1(x)$. It follows that the limit f_0 of the sequence f_i is twice differentiable, its second derivative being the continuous function k.

Since every subsequence of the original sequence (f_i) has a subsequence, say (f_{k_i}), such that $\lim_{i\to\infty} f''_{k_i}(x) = f''_0(x)$, for each $x \in (a,b)$, it follows that $\lim_{i\to\infty} f''_i(x) = f''_0(x)$, for each $x \in (a,b)$, and that the convergence is almost uniform. □

Lemma 5.7. *Let Φ be an even function on $\mathbf{R}$ that satisfies condition (i) of Proposition 5.3 for $b = b_i$, $i = 1,2,\dots$, where (b_i) decreases to 0. If Φ is twice differentiable in $(0,\infty)$, then the function $x \mapsto \Phi''(\sqrt{x})$ is concave in $(0,\infty)$.*

Proof. Since a pointwise limit of concave functions is concave, it suffices to notice that for $x > 0$ one has

$$\Phi''(\sqrt{x}) = \lim_{i\to\infty} \frac{1}{b_i^2}\big(\Phi(\sqrt{x}+b_i) + \Phi(\sqrt{x}-b_i) - 2\Phi(\sqrt{x})\big).$$

□

Lemma 5.8. *Assume that Φ is a twice differentiable, even function on $\mathbf{R}$ such that $\Phi(0) = 0$ and Φ'' is nondecreasing on $\mathbf{R}^+$. Then Φ is a 2-convex Orlicz function.*

Proof. We should verify that the function $t \mapsto \Phi(\sqrt{t})$ is convex. Observe that $\Phi'(0) = 0$, because Φ is even. Thus, if $x > 0$, then $\Phi'(x) = \int_0^x \Phi''(u)\,du \le x\Phi''(x)$. Using this, it is routine to verify that the second derivative of the function $t \mapsto \Phi(\sqrt{t})$ is nonnegative on $\mathbf{R}^+$. □

Proof of Proposition 5.4. We may and do assume that $\Phi(0) = 0$. Let ϕ be a nonnegative C^2 function on $\mathbf{R}$ such that $\phi(x) = 0$ if $|x| > \frac{1}{2}$ and $\int_{\mathbf{R}} \phi(x)\,dx = 1$. Put for $x \in \mathbf{R}$ and $n = 1,2,\dots$

$$\Phi_n(x) = \int_{\mathbf{R}_+} n\phi(n(u-1))\Phi(\frac{x}{u})\frac{du}{u}.$$

Clearly, the Φ_n's are even functions that have two continuous derivatives in $(0,\infty)$ and are bounded from below. Since Φ is continuous, it is clear that $\lim_{n\to\infty} \Phi_n(x) = \Phi(x)$ for $x \in \mathbf{R}$.

Note that, when we prove that the function $x \mapsto \Phi''_n(\sqrt{x})$ is concave in $(0,\infty)$, it will follow that $\Phi''_n(x) \ge 0$ for $x > 0$. Indeed, otherwise we would have for some $A, B, x_0 > 0$ that $\Phi''_n(\sqrt{x}) \le A - Bx$ for $x \ge x_0$.The latter inequality means that $\Phi''_n(u) \le A - Bu^2$ for $u \ge \sqrt{x_0}$. By integrating twice, we would infer that $\lim_{x\to\infty} \Phi_n(x) = -\infty$, while we know that Φ_n is bounded from below.

Observe that the Φ_n's satisfy condition (i) of Proposition 5.3 for every $b > 0$. Indeed, one has

$$\Phi_n(b+\sqrt{x}) + \Phi_n(b-\sqrt{x}) - 2\Phi_n(\sqrt{x}) = \int_{\mathbf{R}_+} n\phi(n(u-1))\Psi(x,u)\frac{du}{u},$$

where

$$\Psi(x,u) = \Phi(\frac{b}{u} + \sqrt{\frac{x}{u^2}}) + \Phi(\frac{b}{u} - \sqrt{\frac{x}{u^2}}) - 2\Phi(\sqrt{\frac{x}{u^2}}).$$

By our assumption, for every $u > 0$ the function $\Psi(\cdot, u)$ is concave in $[0, \infty)$. It follows that Φ_n satisfies condition (i) for every $b > 0$.

Using Lemma 5.7, we obtain that $\Phi_n''(\sqrt{x})$ is concave in $(0, \infty)$. Hence Φ_n'' is nondecreasing in $(0, \infty)$, while the function $x \mapsto x^{-2}\Phi_n''(x)$ is nonincreasing, because so is $x \mapsto \frac{1}{x}\Phi_n''(\sqrt{x})$. By Lemma 5.6, this proves that Φ is of class C^2 in $(0, \infty)$. In order to complete the proof of Proposition 5.4 and thus also of Theorem 5.2 it remains to verify that Φ is of class C^2 on $\mathbf{R}$.

Observe that, by Lemma 5.8, each Φ_n is a 2-convex Orlicz function. It follows that $\Phi = \lim_n \Phi_n$ is a 2-convex Orlicz function too.

Now we can show that $\Phi'(0)$ exists and equals 0. Let us write $c = \lim_{t \to +0} \Phi(t)/t$. Clearly, $c \geq 0$. Since Φ is even, it suffices to show that $c = 0$. To this end observe that for $x \in \mathbf{R}$ one has $\lim_{s \to +0} s^{-1}\Phi(sx) = c|x|$. Since for every $s > 0$ the function $x \mapsto s^{-1}\Phi(sx)$ satisfies property (i) for every $b > 0$, so does the limit of those functions as s tends to 0, i.e., the function $x \mapsto c|x|$. Now, if one had $c > 0$, then it would follow that the function $x \mapsto |x|$ has property (i) for each $b > 0$, which is obviously false. Consequently, $\Phi'(0) = 0$, and hence Φ is differentiable everywhere on $\mathbf{R}$. Since the derivative Φ' has the Darboux property and is nondecreasing, it follows that Φ' is continuous on $\mathbf{R}$.

It follows that the function Φ' is convex in $[0, \infty)$, because Φ'' is a nondecreasing function in $(0, \infty)$ (being concave and nonnegative). Hence the function $x \mapsto \frac{1}{x}\Phi'(x)$ is nondecreasing and nonnegative in $(0, \infty)$, so it has a limit at 0. Since Φ' is odd, this implies that $\Phi''(0)$ exists. Again, using the Darboux property and the monotonicity of Φ'' in $[0, \infty)$, we obtain that the function Φ'' is continuous at 0, which completes the proof that Φ'' is continuous on $\mathbf{R}$ and that $\Phi''(\sqrt{x})$ is concave in $[0, \infty)$. □

Now, given the solution to the extremal problem associated with Rosenthal's inequality we would like to identify the best constants in this inequality, and we would like to work in the context of Orlicz functions Φ. In order to do that, we will need additional assumptions on Φ. As we will show in the next two sections, the sufficient condition is that Φ'' is either convex or concave. Let us remark that in the case of power functions and $2 \leq p \leq 4$ the above conditions on Φ are equivalent to $3 \leq p \leq 4$ and $2 \leq p \leq 3$, respectively. This, combined with Utev's result for $p > 4$, provides a complete solution to the question of the best possible constants in Rosenthal's inequality for all p, $2 \leq p < \infty$. We will treat those two cases in separate sections, and we begin with the easier case where Φ'' is convex.

6. Upper bound for the constant in Rosenthal's inequality: Φ'' convex

Through this section we will assume that our Orlicz function $\Phi \in \mathcal{O}_{2,4}$ has the property that Φ'' is convex. Therefore, using Lemma 5.8, we infer that $\Phi(\sqrt{t})$ is convex, so that there exists a finite limit

$$G = \lim_{t \to 0} \frac{\Phi(t)}{t^2} = \frac{1}{2}\Phi''(0),$$

and the function $t \to \Phi(t) - Gt^2$ is increasing. Using the solution to the extremal problem we will obtain the best constant in the generalized version of Rosenthal's inequality. We will prove the following:

Theorem 6.1. *Let Φ be an Orlicz function such that Φ'' is convex and $\Phi''(\sqrt{t})$ is concave. Let (σ_j) be a sequence of nonnegative numbers such that $\sum \sigma_j^2 < \infty$. Then for all sequences (X_n) of independent symmetric random variables with $E\,X_j^2 = \sigma_j^2$ the following inequality is true:*

$$E\,\Phi(\sum_{j=1}^{n} X_j) \le \sum_{j=1}^{n} E\,\Phi(X_j) + E\,\Phi(\sigma g) - G\sigma^2,$$

where g is a standard Gaussian random variable, and $\sigma^2 = \sum \sigma_j^2$.

Proof. Since $\Phi''(\sqrt{t})$ is concave, by Theorem 5.2 we have that

$$E\,\Phi(\sum_{j=1}^{n} X_j) - \sum_{j=1}^{n} E\,\Phi(X_j) \le E\,\Phi(\sum_{j=1}^{n} \sigma_j\,\varepsilon_j) - \sum_{j=1}^{n} \Phi(\sigma_j).$$

Since Φ'' is convex, Theorem 1.1 implies that

$$E\,\Phi(\sum_{j=1}^{n} \sigma_j\,\varepsilon_j) \le E\,\Phi(\sum_{j=1}^{n} \sigma_j g_j) = E\,\Phi(\sigma g),$$

where (g_j) is a sequence of independent, standard Gaussian random variables. Moreover, we have $\Phi(\sigma_i)/\sigma_i^2 \ge G$, if $\sigma_i > 0$. Combining these estimates, we obtain

$$E\,\Phi(\sum_{j=1}^{n} X_j) - \sum_{j=1}^{n} E\,\Phi(X_j) \le E\,\Phi(\sigma g) - \sum_{j=1}^{n} \Phi(\sigma_j) \le E\,\Phi(\sigma g) - G\sigma^2.$$

This completes the proof. □

When specified to power functions, the assumptions about Φ force p to satisfy $3 \le p \le 4$. Let us note, however, that in the course of the proof of Theorem 6.1 we have not really used the full strength of Theorem 1.1 but just the fact that

$$E\,\Phi(\sum \sigma_j \epsilon_j) \le E\,\Phi(\sum \sigma_j g_j).$$

Thus, the above inequality and Theorem 1.1 suffice to obtain a sharp version of Rosenthal's inequality. In particular, Haagerup [H] proved that

$$E\,|\sum \sigma_j \epsilon_j|^p \le E\,|\sum \sigma_j g_j|^p$$

holds for all $2 \le p < \infty$. Hence, we obtain

Corollary 6.2. *Let $2 \le p \le 4$. Let c_1, c_2 be two positive constants. Then*

$$\sup\left\{ E\,|\sum_{j=1}^{n} X_j|^p : \sum_{j=1}^{n} E\,|X_j|^p \le c_1, \sum_{j=1}^{n} E\,X_j^2 \le c_2 \right\} \le c_1 + \|g\|_p^p c_2^{p/2}.$$

It should be stressed, however, that the proof of Haagerup's result for $2 < p < 3$ is difficult. Therefore, in the next section we will provide a different, not relying on Haagerup's result, proof of a sharp version of Rosenthal's inequality.

Remark 6.3. Suppose that $\sum_{j=1}^{n} E\,\Phi(X_j) \le c_1$ and $\sum_{j=1}^{n} E\,X_j^2 \le c_2$. Then, Theorem 6.1 implies that

$$\begin{aligned} E\,\Phi(\sum_{j=1}^{n} X_j) &\le \sum_{j=1}^{n} E\,\Phi(X_j) + E\,\Phi(\sigma g) - G\sigma^2 \\ &\le c_1 + E\,\Phi(\sqrt{c_2}\,g) - Gc_2, \end{aligned}$$

since the function $E\,\Phi(\sigma g)-\sigma^2 G$ is increasing in σ. If $c_1-c_2G<0$ then the right–hand side can be replaced by $E\,\Phi(\sqrt{c_1/G}g)$. Indeed, the inequality $Gt^2\le\Phi(t)$ implies that

$$\sigma^2 G = G\sum E\,X_j^2 \le \sum E\,\Phi(X_j)\le c_1.$$

Whence, $\sigma\le\sqrt{c_1/G}$, so that

$$\begin{aligned}E\,\Phi(\sigma g)-\sigma^2 G+\sum E\,\Phi(X_j) \le& E\,\Phi(\sqrt{c_1/G}g)-\frac{c_1}{G}G+\sum E\,\Phi(X_j)\\ \le& E\,\Phi(\sqrt{c_1/G}g)-c_1+c_1 = E\,\Phi(\sqrt{c_1/G}g).\end{aligned}$$

Therefore, under the above constraints on $\sum E\,X_j^2$ and $\sum E\,\Phi(X_j)$ we have that

$$E\,\Phi\left(\sum_{j=1}^n X_j\right)\le\begin{cases}c_1+E\,\Phi(\sqrt{c_2}g)-Gc_2, & \text{if } c_1-c_2G\ge 0;\\ E\,\Phi(\sqrt{c_1/G}g), & \text{if } c_1-c_2G<0.\end{cases}$$

As we will see in Section 8 below the above estimate is sharp.

7. Upper bound for the constant in Rosenthal's inequality: Φ'' concave

In this section we will obtain an analog of Theorem 6.1 under the assumption that our Orlicz function Φ has the property that Φ'' is concave. Here we do not have the minimality property of Rademacher functions (Theorem 1.1), so that the method used in the last section does not carry over. Instead, we will rely on estimates for Rademacher functions proved in Section 4. Of course, the solution of the extremal problem of Section 5 is again crucial. Note that the assumption Φ'' – concave automatically implies that $\Phi''(\sqrt{\cdot})$ is concave, so that Theorem 5.2 applies. Before stating the theorem, let us note, that under our assumption on Φ there exists a finite limit

$$G=\lim_{t\to 0}\frac{\Phi(t)}{t^2}=\frac{1}{2}\Phi''(0).$$

Also, since the function Φ'' is nondecreasing, so is $t\to\Phi(t)-t^2G$. Here is the main result of this section:

Theorem 7.1. *Let Φ be an Orlicz function such that Φ'' is concave. Let (σ_j) be a sequence of nonnegative numbers such that $\sum\sigma_j^2<\infty$. Then, for all sequences (X_n) of independent symmetric random variables with $E\,X_j^2=\sigma_j^2$ the following inequality is true:*

$$E\,\Phi(\sum_{j=1}^n X_j)\le\sum_{j=1}^n E\,\Phi(X_j)+E\,\Phi(\sigma g)-G\sigma^2,$$

where g is a standard Gaussian random variable, and $\sigma^2=\sum\sigma_j^2$.

Proof. Since $\Phi''(\sqrt{t})$ is concave, by Theorem 5.2 we have that

$$E\,\Phi(\sum_{j=1}^n X_j)-\sum_{j=1}^n E\,\Phi(X_j)\le E\,\Phi(\sum_{j=1}^n\sigma_j\,\varepsilon_j)-\sum_{j=1}^n\Phi(\sigma_j).$$

In order to estimate the latter quantity we will use the following estimate

$$E\,\Phi(b+\sigma\epsilon)-\Phi(\sigma)\le E\,\Phi(b+\frac{\sigma}{\sqrt{2}}(\epsilon+\epsilon'))-2\Phi(\frac{\sigma}{\sqrt{2}}),$$

which holds for each $b\ge 0$ and each $\sigma\ge 0$ (see Proposition 4.5). □

Using repeatedly the above estimate and the Central Limit Theorem, we obtain

$$E\,\Phi(b+\sigma\epsilon)-\Phi(\sigma)\le E\,\Phi(b+\sigma g)-\lim_{n\to\infty}2^n\Phi(\frac{\sigma}{2^{n/2}})=E\,\Phi(b+\sigma g)-\sigma^2G.$$

Now, by Fubini's theorem one can easily conclude as follows:

$$\begin{aligned}
E\,\Phi(\sum_{j=1}^n\sigma_j\,\varepsilon_j)-&\sum_{j=1}^n\Phi(\sigma_j\,\varepsilon_j)\\
&=E\,\Phi(\sum_{j=1}^{n-1}\sigma_j\,\varepsilon_j+\sigma_n\,\varepsilon_n)-\Phi(\sigma_n\,\varepsilon_n)-\sum_{j=1}^{n-1}\Phi(\sigma_j\,\varepsilon_j)\\
&\le E\,\Phi(\sum_{j=1}^{n-1}\sigma_j\,\varepsilon_j+\sigma_n g_n)-\sigma_n^2G-\sum_{j=1}^{n-1}\Phi(\sigma_j\,\varepsilon_j)\\
&\le\dots\\
&\le E\,\Phi(\sum_{j=1}^n\sigma_j g_j)-\sum_{j=1}^n\sigma_j^2G\\
&=E\,\Phi(\sigma g)-\sigma^2G.
\end{aligned}$$

Combining the above estimates we obtain the desired inequality. This completes the proof.

Remark 7.2. 1. When specified to power functions, the assumption that Φ'' is concave means that $2\le p\le 3$. Assuming that $p\ne 2$, we have then $G=0$ and thus we get

$$E\,|\sum_{j=1}^n X_j|^p\le\sum_{j=1}^n E\,|X_j|^p+\|g\|_p^p\left(\sum_{j=1}^n\sigma_j^2\right)^{p/2}.$$

This gives an alternative proof of the part of Corollary 6.2 dealing with the case $2\le p\le 3$. One advantage of this argument is that it the bypasses difficult Haagerup result for $2<p<3$.

2. An alternative proof of Theorem 7.1 based on Proposition 4.2 rather than Proposition 4.5 can be given, too. Indeed, after applying Theorem 5.2 at the beginning of the proof one can use Proposition 4.2 to write

$$\begin{aligned}
E\,\Phi(\sum_{j=1}^n\sigma_j\,\varepsilon_j)-&\sum_{j=1}^n\Phi(\sigma_j)\\
&\le E\,\Phi\left(\frac{\sigma}{\sqrt{n}}\sum_{j=1}^n\varepsilon_j\right)-n\Phi\left(\frac{\sigma}{\sqrt{n}}\right)\to E\,\Phi(\sigma g)-\sigma^2G,
\end{aligned}$$

as $n\to\infty$, by the Central Limit Theorem.

We will show in the next section that the estimate of Theorem 7.1 is sharp.

8. A lower bound

In this section we will exhibit an example showing the sharpness of the estimates obtained in Theorems 6.1 and 7.1, and Corollary 6.2. Recall that the conditions imposed on the Orlicz function Φ in either of these theorems imply that there exists a finite limit $G = \lim_{t\to 0} t^{-2}\Phi(t)$. Here we let more generally

$$G = \liminf_{t\to 0} t^{-2}\Phi(t).$$

Proposition 8.1. *Let Φ be a 2-convex Orlicz function and suppose that*

$$\limsup_{t\to\infty} t^{-2}\Phi(t) = \infty.$$

Then, for $c_1, c_2 > 0$,

$$\sup\Big\{E\,\Phi\Big(\sum_{i=1}^n X_i\Big) : \sum_{i=1}^n E\,\Phi(X_i) \le c_1 \text{ and } \sum_{i=1}^n E\,X_i^2 \le c_2,\ n \in \mathbf{N}\Big\}$$
$$= \begin{cases} E\,\Phi(\sqrt{c_2}g) - c_2G + c_1, & \text{if } c_1 - c_2G \ge 0; \\ E\,\Phi(\sqrt{\frac{c_1}{G}}), & \text{if } c_1 - c_2G < 0. \end{cases}$$

Therefore, if Φ satisfies our sufficient conditions of Sections 6 or 7, then the above number is the exact value of the Rosenthal's constant for Φ.

Proof. Consider first the case where $\Phi(t) = |t|^p$ for some $p > 2$. Of course, in this case $G = 0$.

Let c be a fixed number, $0 < c < \sqrt{c_2}$. For a positive integer n and positive constants $\mu = \mu_n \le n$, $\gamma = \gamma_n$ to be specified later, we define independent symmetric random variables $X_1, \dots, X_{2n}$ by the formulas

$$X_i = \frac{c}{\sqrt{n}}\epsilon_i, \qquad X_{n+i} = \gamma\,\epsilon_{n+i}\,1_{A_i^{(n)}},$$

for $i = 1, \dots, n$, where $A_1^{(n)}, \dots, A_n^{(n)}$ are independent sets of measure μ/n which are independent of the Rademacher functions $\epsilon_1, \dots, \epsilon_{2n}$.

Clearly, one has for any $r > 0$

$$\sum_{i=1}^{2n} E\,|X_i|^r = c^r n^{1-\frac{r}{2}} + \gamma^r\mu.$$

We choose μ and γ so that $\sum_{i=1}^{2n} E\,|X_i|^p = c_1$, and $\sum_{i=1}^{2n} E\,|X_i|^2 = c_2$. This leads to equations

$$\gamma^2\mu = c_2 - c^2, \qquad \gamma^p\mu = c_1 - c^p n^{1-\frac{p}{2}}.$$

Thus we obtain

$$\gamma_n^{p-2} = \frac{c_1 - c^p n^{1-\frac{p}{2}}}{c_2 - c^2}, \qquad \mu_n = \frac{c_2 - c^2}{\gamma_n^2}.$$

By the cotype property of the function $\Phi(t) = |t|^p$, where $p \geq 2$, we have

$$\begin{aligned} E\,|\sum_{i=1}^{2n} X_i|^p \geq & E\,|\sum_{i=1}^{n} X_i|^p + \sum_{i=n+1}^{2n} E\,|X_i|^p \\ = & c^p E\,|\frac{1}{\sqrt{n}} \sum_{i=1}^{n} \epsilon_i|^p + \gamma_n^p n \frac{\mu_n}{n} \\ = & c^p E\,|\frac{1}{\sqrt{n}} \sum_{i=1}^{n} \epsilon_i|^p + \gamma_n^p \mu_n. \end{aligned}$$

Writing now $X_i = X_i^{(n,c)}$ for $i = 1, \ldots, 2n$ and letting n tend to infinity, we obtain

$$\liminf_{n\to\infty} E\,|\sum_{i=1}^{2n} X_i^{(n,c)}|^p \geq c^p E\,|g|^p + \lim_{n\to\infty} \gamma_n^p \mu_n = c^p E\,|g|^p + c_1.$$

Since c can be chosen arbitrarily close to $\sqrt{c_2}$, this proves our lower estimate in the case where $\Phi(t) = |t|^p$ for some $p > 2$.

Now let Φ be an Orlicz function. Suppose first that $c_1 - c_2 G \geq 0$. Consider the following system of inequalities

$$c^2 + \mu\gamma^2 \leq c_2, \qquad n\Phi(\frac{c}{\sqrt{n}}) + \mu\Phi(\gamma) \leq c_1, \qquad 0 < \mu \leq n.$$

Given a positive solution (c, γ, μ, n) of this system, define a sequence $X_1, \ldots, X_{2n}$ of independent symmetric random variables by the same formulas as before in the case $\Phi(t) = |t|^p$. Clearly, this sequence will satisfy the constraints

$$\sum_{i=1}^{2n} E\,X_i^2 \leq c_2, \qquad \sum_{i=1}^{2n} E\,\Phi(X_i) \leq c_1.$$

If n is large enough, then by the Central Limit Theorem we will have

$$\begin{aligned} E\,\Phi(\sum_{i=1}^{2n} X_i) \geq & E\,\Phi(\sum_{i=1}^{n} X_i) + \sum_{i=n+1}^{2n} E\,\Phi(X_i) = E\,\Phi(\frac{c}{\sqrt{n}} \sum_{i=1}^{n} \varepsilon_i) + n\Phi(\gamma)\frac{\mu}{n} \\ = & E\,\Phi(\frac{c}{\sqrt{n}} \sum_{i=1}^{n} \varepsilon_i) + \mu\Phi(\gamma) \approx E\,\Phi(cg) + \mu\Phi(\gamma). \end{aligned}$$

Now, with $c < \sqrt{c_2}$ being as close to $\sqrt{c_2}$ as one wants, a suitable choice of n will assure that $n\Phi(\frac{c}{\sqrt{n}}) \approx c^2 G$. If $\limsup_{t\to\infty} \Phi(t)/t^2 = \infty$, then a large γ can be chosen so that the ratio $\Phi(\gamma)/\gamma^2$ is greater than $(c_1 - c^2 G)/(c_2 - c^2)$. Finally, μ is assigned a small value such that both inequalities in the system are satisfied and the second one is close to being equality. In this situation, one obtains that $\mu\Phi(\gamma) \approx c_1 - c_2 G$.

The case where $c_1 - c_2 G < 0$ is much easier, because we may simply take $\gamma = \mu = 0$ and let $c = \sqrt{n}\Phi^{-1}(c_1/n) \approx \sqrt{c_1/G}$. This proves our statement. □

9. Remarks

We conclude this paper by exhibiting some negative results showing limitations of some of our arguments. Probably the most intriguing question concerns Orlicz functions for which the minimality of Rademacher functions (or at least domination

of Rademachers by Gaussian variables) holds. Our first example shows that the condition that Φ'' is concave is not enough.

Remark 9.1. Let $X = \sum_{i=1}^n \sigma_i \epsilon_i$ be a linear combination of Rademacher functions. If

$$E\,X^2 = \sigma^2, \qquad E\,|X| < \sigma E\,|g| = \sqrt{\frac{2}{\pi}}\sigma,$$

and $\Phi_a(x) = |x|^3 - (|x| - a)_+^3$, then $E\,\Phi_a(X) > E\,\Phi_a(\sigma g)$ for sufficiently small $a > 0$.

This means that the exact analog of Khintchine's inequality fails for some Orlicz functions Φ such that $\Phi''(x)$ is concave in $[0, \infty)$. We prove it using the following Taylor expansion.

If Y is a random variable such that $E\,Y^2 < \infty$, then for $a > 0$ one has

$$E\,\Phi_a(Y) = a^3 - 3a^2 E\,|Y| + 3aE\,Y^2 + \rho(a), \qquad 0 \geq \rho(a) \geq -a^3,$$

which implies that for small $a > 0$ one has

$$E\,\Phi_a(X) - E\,\Phi_a(\sigma g) = 3a^2(E\,|\sigma g| - E\,|X|) + O(a^3).$$

The latter expression is clearly positive for sufficiently small $a > 0$. Taylor's formula for $E\,\Phi_a(Y)$ is a consequence of the following easily verifiable identity valid for $a > 0$ and every real x

$$\Phi_a(x) = 3ax^2 - 3a^2|x| + a^3 - (a - |x|)_+^3.$$

If it were true that for $0 < b, \sigma < \infty$ one had the estimate

$$E\,\Phi(b + \sigma\epsilon) \leq E\,\Phi(b + \sigma g) + \Phi(\sigma),$$

for all Orlicz functions Φ such that $\Phi''(\sqrt{x})$ is concave in $[0, \infty)$, then Rosenthal's inequality would follow easily. However, the above estimate fails, even in a weaker form. Namely, we have

Remark 9.2. Let $a > 0$ and let $\Phi_a(x) = x^4$ for $x \in [-a, a]$, $\Phi_a(x) = 6a^2x^2 - 8a^3|x| + 3a^4$ for $|x| > a$. If C is any positive constant, then for every $a > 0$ one has

$$E\,\Phi_a(a + \sigma\epsilon) - E\,\Phi_a(a + \sigma g) > C\Phi_a(\sigma),$$

if $\sigma > 0$ is sufficiently small.

Again this is proved using a suitable Taylor expansion. A homogeneity argument shows that it suffices to consider the case where $a = 1$. Put

$$f(\sigma) = E\,\Phi_1(1 + \sigma g) - E\,\Phi_1(1 + \sigma\epsilon).$$

Write $\rho(x) = x^4 - \Phi_1(x)$, so that $\rho(x) = (|x| - 1)_+^3(3 + |x|)$. If Y is a symmetric random variable such that $E\,Y^4 < \infty$, one has for $\sigma > 0$

$$E\,\Phi_1(1 + Y) = E\,(1 + Y)^4 - E\,\rho(1 + Y) = 1 + 6E\,Y^2 + E\,Y^4 - E\,\rho(1 + Y),$$

which implies that

$$f(\sigma) = \sigma^4(E\,g^4 - 1) - E\,\rho(1 + \sigma g) + E\,\rho(1 + \sigma\epsilon).$$

Now if $0 \leq \sigma \leq 1$, then $E\,\rho(1 + \sigma\epsilon) = \frac{1}{2}\rho(1 + \sigma) = \frac{1}{2}\sigma^3(4 + \sigma)$, and

$$E\,\rho(1 + \sigma g)1_{g>0} = E\,(4\sigma^3 g^3 + \sigma^4 g^4)1_{g>0} = 2\sigma^3 E\,|g|^3 + \frac{1}{2}\sigma^4 E\,|g|^4,$$

while $E\,\rho(1+\sigma g)1_{g\le 0} = o(\sigma^4)$ as σ tends to 0. Therefore

$$\begin{aligned} f(\sigma) =& 2\sigma^4 + \frac{1}{2}\sigma^3(4+\sigma) - (2\sigma^3 E\,|g|^3 + \frac{1}{2}\sigma^4 E\,g^4 + o(\sigma^4)) \\ =& -2(E\,|g|^3 - 1)\sigma^3 + \sigma^4 + o(\sigma^4). \end{aligned}$$

This proves our claim, because $E\,|g|^3 > 1 = E\,g^2$ and $\Phi_1(\sigma) = \sigma^4$ for $\sigma \in [-1,1]$.

We close by observing that the assumption about concavity of Φ'' in Lemma 4.6 cannot be weakened to concavity of the function $t \to \Phi''(\sqrt{t})$.

Remark 9.3. There exist Orlicz functions Φ of class C^2 such that $\Phi''(\sqrt{x})$ is concave in $[0,\infty)$ and yet there is no $C < \infty$ so that for all b,$x \ge 0$

$$\begin{aligned} \Phi(b+\sqrt{2}x) + \Phi(b-\sqrt{2}x) + 2(\Phi(b) - \Phi(b+x) - \Phi(b-x)) \\ + C(\Phi(x) - 2\Phi(\frac{1}{\sqrt{2}}x)) \ge 0, \end{aligned}$$

i.e., for those Φ no choice of a positive constant C will make the inequality of Lemma 4.6 true for Φ. Indeed, if $b > 0$ is given, we let $\Phi = \Phi_b$ be the extreme function given by

$$\Phi_b(t) = \begin{cases} t^4, & \text{if } |t| \le b; \\ b^2(6t^2 - 8b|t| + 3b^2), & \text{otherwise.} \end{cases}$$

Then for $0 < x < b/\sqrt{2}$, the left-hand side equals $(2+C/2)x^4 - 8b(\sqrt{2}-1)x^3$, which is negative if $0 < x < 16b(\sqrt{2}-1)/(C+4)$.

Acknowledgments. We would like to thank S. Kwapień for bringing Utev's paper [U2] to our attention.

References

[E1] M. R. Eaton, *A note on symmetric Bernoulli random variables*, Ann. Math. Statist. **41** (1970), 1223–1226. MR **42:**3827

[E2] M. R. Eaton, *A probability inequality for linear combinations of bounded random variables*, Ann. Statist. **2** (1974), 609–613.

[H] U. Haagerup, *Best constants in the Khintchine's inequality*, Studia Math. **70** (1981), 231–283. MR **83m:**60031

[JSZ] W. B. Johnson, G. Schechtman, and J. Zinn, *Best constants in moment inequalities for linear combinations of independent and exchangeable random variables*, Ann. Probab. **13** (1985), 234–253. MR **86i:**60054

[K] R. Komorowski, *On the best possible constants in the Khintchine inequality for $p \ge 3$*, Bull. London Math. Soc. **20** (1988), 73–75. MR **89e:**60037

[KS] S. Kwapień and J. Szulga, *Hypercontraction methods in moment inequalities for series of independent random variables in normed spaces*, Ann. Probab. **19** (1991), 1–8. MR **92a:**60051

[MO] A. W. Marshall and I. Olkin, *Inequalities: Theory of Majorization and its Application*, Academic Press, New York, 1979. MR **81b:**00002

[Pe] V. V. Petrov, *Sums of Independent Random Variables*, Springer, Berlin, Heidelberg, 1975. MR **52:**9335

[Ph] R. R. Phelps, *Lectures on Choquet's Theorem*, Van Nostrand, Princeton, 1966. MR **33:**1690

[P] I. F. Pinelis, *Extremal probabilistic problems and Hotelling's T^2 test under a symmetry condition*, Ann. Statist. **22** (1994), 357–368. MR **95m:**62115

[PU] I. F. Pinelis and S. A. Utev, *Estimates of the moments of sums of independent random variables*, Theory Probab. Appl. **29** (1984), 574–577. MR **85m:**60034

[R] H. P. Rosenthal, *On the subspaces of L_p ($p > 2$) spanned by sequences of independent random variables*, Israel J. Math. **8** (1970), 273–303. MR **42:**6602

[S] S. B. Stečkin, *On the best lacunary system of functions*, Izv. Akad. Nauk. SSSR, Ser. Mat. **25** (1961), 357–366 (in Russian). MR **24:**A951

[Sz] S. Szarek, *On the best constant in the Khintchine inequality*, Studia Math. **58** (1976), 197–208. MR **55:**3672

[T] M. Talagrand, *Isoperimetry and integrability of the sum of independent Banach – space valued random variables*, Ann. Probab. **17** (1989), 1546–1570. MR **91e:**60054

[U1] S. A. Utev, *Extremal problems in moment inequalities*, Theory Probab. Appl. **28** (1984), 421–422. MR **87d:**60021

[U2] S. A. Utev, *Extremal problems in moment inequalities*, Limit Theorems in Probability Theory, Trudy Inst. Math., Novosibirsk, 1985, pp. 56–75 (in Russian).

[W] P. Whittle, *Bounds for the moments of linear and quadratic forms in independent random variables*, Theory Probab. Appl. **5** (1960), 302–305. MR **24:**A3673

[Y] R. M. G. Young, *On the best possible constants in the Khintchine inequality*, J. London Math. Soc. **14** (1976), 496–504. MR **55:**11008

Institute of Mathematics, Polish Academy of Sciences, ul. Abrahama 18, 81–825 Sopot, Poland

E-mail address: T.Figiel@IMPAN.Gda.pl

Department of Mathematics, North Carolina State University, Raleigh, North Carolina 27695–8205

E-mail address: pawel@math.ncsu.edu

Department of Mathematics, Texas A&M University, College Station, Texas 77843

E-mail address: johnson@math.tamu.edu

Department of Theoretical Mathematics, The Weizmann Institute of Science, Rehovot, Israel

E-mail address: mtschech@weizmann.weizmann.ac.il

Department of Mathematics, Texas A&M University, College Station, Texas 77843

E-mail address: jzinn@plevy.math.tamu.edu