The Annals of Probability
1994, Vol. 22, No. 1, 453–468

ON A DOMINATION OF SUMS OF RANDOM VARIABLES BY SUMS OF CONDITIONALLY INDEPENDENT ONES[1]

BY PAWEŁ HITCZENKO

North Carolina State University

It is known that if (X_n) and (Y_n) are two $(\mathscr{F}_n)$-adapted sequences of random variables such that for each $k \geq 1$ the conditional distributions of X_k and Y_k, given $\mathscr{F}_{k-1}$, coincide a.s., then the following is true:

$$\Big\| \sum X_k \Big\|_p \leq B_p \Big\| \sum Y_k \Big\|_p, \qquad 1 \leq p < \infty,$$

for some constant B_p depending only on p. The aim of this paper is to show that if a sequence (Y_n) is conditionally independent, then the constant B_p may actually be chosen to be independent of p. This significantly improves all hitherto known estimates on B_p and extends an earlier result of Klass on randomly stopped sums of independent random variables as well as our recent result dealing with martingale transforms of Rademacher sequences.

1. Introduction and statement of the result. In this paper we study the L_p-norm inequalities for sums of arbitrarily dependent random variables. In order to describe our results in detail, let us recall some definitions and notation.

Let $(\mathscr{F}_n)$ be an increasing sequence of σ-algebras on some probability space $(\Omega, \mathscr{F}, P)$. We shall assume that $\mathscr{F}_0 = \{\varnothing, \Omega\}$. A sequence (X_n) of random variables is called $(\mathscr{F}_n)$-adapted [resp., $(\mathscr{F}_n)$-predictable] if X_n is $\mathscr{F}_n$-measurable (resp., $\mathscr{F}_{n-1}$-measurable) for each $n = 1, \dots$. In the sequel we shall simply write adapted (resp., predictable) if there is no risk of confusion. All equalities and inequalities between random variables are assumed to hold almost surely. For any sequence (X_n) of random variables we shall write $X^* = \sup_n |X_n|$ and $X_n^* = \max_{1 \leq k \leq n} |X_k|$. The L_p-norm, $1 \leq p \leq \infty$, of a random variable X is denoted by $\|X\|_p$, and $I(A)$ or I_A will denote the indicator function of a set A. Given a σ-algebra $\mathscr{A} \subset \mathscr{F}$ and an integrable random variable X we shall denote the conditional expectation of X, given $\mathscr{A}$, by $E_{\mathscr{A}} X$. If $\mathscr{A} = \mathscr{F}_k$, then we shall simply write $E_k X$ for $E_{\mathscr{F}_k} X$. The conditional distribution of a random variable X, given $\mathscr{A}$, is denoted by $\mathscr{L}(X|\mathscr{A})$. Thus $\mathscr{L}(X|\mathscr{A}) = \mathscr{L}(Y|\mathscr{A})$ means that for each real number t we have

$$P(X < t|\mathscr{A}) = P(Y < t|\mathscr{A}).$$

Received September 1992.
[1]Supported in part by an NSF grant.
AMS 1991 *subject classifications*. 60E15, 60G42.
Key words and phrases. Moment inequalities, martingale, tangent sequences.

The following definition was introduced by Kwapień and Woyczyński (1986):

DEFINITION. Let $(\mathscr{F}_n)$ be an increasing sequence of σ-algebras on $(\Omega, \mathscr{F}, P)$.

(a) Two adapted sequences (X_n) and (Y_n) of random variables are tangent if, for each $n = 1, \dots,$ we have

$$\mathscr{L}(X_n | \mathscr{F}_{n-1}) = \mathscr{L}(Y_n | \mathscr{F}_{n-1}).$$

(b) An adapted sequence (Y_n) of random variables satisfies condition (CI) if there exists a σ-algebra $\mathscr{G} \subset \mathscr{F}$ such that, for each $n = 1, \dots,$

$$\mathscr{L}(Y_n | \mathscr{F}_{n-1}) = \mathscr{L}(Y_n | \mathscr{G})$$

and such that (Y_n) is a sequence of $\mathscr{G}$-conditionally independent random variables.

A sequence (X_n) is conditionally symmetric if (X_n) and $(-X_n)$ are tangent sequences of random variables. With a slight abuse of terminology, a martingale will be called conditionally symmetric if its difference sequence is conditionally symmetric. A class of all conditionally symmetric martingales will be denoted by $\mathscr{M}\mathscr{S}$.

Every sequence of random variables (X_n) admits (possibly on an enlarged probability space) a tangent sequence that satisfies condition (CI) [cf., e.g., Kwapień and Woyczyński (1986) or Hitczenko (1990a), Lemma 2.3]. Throughout this paper such a sequence will be denoted by $(\tilde{X}_n)$ and will be called a decoupled version of (X_n). It is useful to note that the σ-algebra $\mathscr{G}$ can be chosen so that the random variables X_k, $k = 1, \dots,$ are $\mathscr{G}$-measurable. As an example consider the following situation: Let $(\mathscr{A}_n)$ be an increasing sequence of σ-algebras, and let (ξ_n) be a sequence of independent random variables, such that ξ_n is $\mathscr{A}_n$-measurable and independent of $\mathscr{A}_{n-1}$, $n = 1, \dots$. Let (v_n) be an $(\mathscr{A}_n)$-predictable sequence and suppose that (ξ'_n) is a copy of the sequence (ξ_n) and is independent of $\sigma(\bigcup \mathscr{A}_n)$. If we let $\mathscr{F}_n = \sigma(\mathscr{A}_n, \xi'_k;\ 1 \le k \le n)$ and $X_n = v_n \xi_n$, $n = 1, \dots,$ then the decoupled version of (X_n) is given by $(v_n \xi'_n)$, where $\mathscr{G}$ can be taken to be $\sigma(\bigcup \mathscr{A}_n)$. Particular choices of the form of the sequence (v_k) lead to different situations considered by various authors. If we take, for example, $v_k = \sum_{j=1}^{k-1} a_{jk} \xi_j$, then we obtain a bilinear form in (ξ_n) [cf. Kwapień (1987) and references therein], while the choice $v_k = I(\tau \ge k)$ deals with the case of randomly stopped sums of independent random variables [cf. Klass (1988, 1990)].

The notion of tangent sequences proved to be a very useful tool in studying sequences of arbitrarily dependent random variables since it allows one to relate the properties of such sequences (X_n) to the corresponding properties $(\tilde{X}_n)$. The latter, being sequences of conditionally independent random variables, are typically much easier to investigate and better understood. From that point of view, a fundamental question to be asked is: How do two tangent sequences of random variables compare? There is by now a rather rich literature on various aspects of such comparison, and a recent book by

Kwapień and Woyczyński (1992) is an excellent source of information on tangent sequences.

In the present paper we shall be interested in one particular aspect of comparison, namely, in the domination of the L_p-norm of a sum of arbitrary random variables by the L_p-norm of its decoupled version. We begin our discussion by recalling the following inequality for martingales: For $1 \le p < \infty$, there exist constants A_p and B_p, depending only on p, such that for every martingale f with difference sequence d one has

$$A_p^{-1}\Big\| \sum \bar{d}_k \Big\|_p \le \| f^* \|_p \le B_p \Big\| \sum \bar{d}_k \Big\|_p.$$

Several different proofs of this inequality are available [cf. Zinn (1985), Kwapień and Woyczyński (1989), McConnell (1990) and Hitczenko (1988, 1993a)]. The proofs given in McConnell (1990) and Hitczenko (1993a) contain additional information on the size of the constants, namely, that both A_p and B_p are bounded above by $O(p)$. It is not hard to see that this bound on A_p is optimal, that is, $A_p = O(p)$ [Hitczenko (1993b)]. From the point of view of applications, however, the right-hand side inequality is more interesting, as one usually wants to dominate a sequence of random variables by its decoupled version. The main purpose of this paper is to address the question about the asymptotic behavior of the constant B_p. First of all, let us note that, using a result of Hitczenko (1990a), one can show that $B_p \le O(p/\ln p)$. Thus, the behavior of B_p is certainly different than that of A_p. Since all hitherto known proofs are "symmetric" [in the sense that they rely only on the fact that (d_n) and $(\bar{d}_n)$ are tangent and do not take advantage of the conditional independence of the latter sequence], they cannot possibly be used to provide any additional information on B_p. On the other hand, in certain particular cases, the question about the size of that constant has been answered satisfactorily. All of those cases deal with martingale difference sequences of the form $d_n = v_n \xi_n$, $n = 1, \dots,$ where (ξ_n) are independent random variables and (v_n) is a predictable sequence. Klass [(1988), Theorem 3.1] studied the case of an arbitrary sequence (ξ_n) and (v_n) of the form $v_n = I(\tau \ge n)$, where τ is a stopping time. The situation of $v_n = \sum_{k=1}^{n-1} a_{kn} \xi_k$ was considered by Kwapień (1987). Finally, Hitczenko (1993b) deals with martingale transforms of either Rademacher or i.i.d. standard Gaussian random variables. In each of those cases the constant B_p turns out to be independent of p. Our aim here is to generalize these results to arbitrary sequences of random variables. We shall prove the following.

THEOREM 1.1. *There exists an absolute constant K such that for every p, $1 \le p < \infty$, and for every adapted sequence (X_n) of random variables one has*

$$\Big\| \sum X_i \Big\|_p \le K \Big\| \sum \bar{X}_i \Big\|_p.$$

Similarly to the special case treated in Hitczenko (1993b), our proof will be based on martingale methods. Let f be a martingale with difference sequence

d. Given two stopping times ν and μ such that $\nu \geq \mu$, for $n = 1, \dots,$ we let

$$^{\mu}f_n^{\nu} = \sum_{k=1}^{n} I(\mu < k \leq \nu) d_k.$$

Then the sequence $^{\mu}f^{\nu} = (^{\mu}f_n^{\nu})$ is also a martingale (referred to as f started at μ and stopped at ν). In particular, $f^n = (f_0, \dots, f_{n-1}, f_n, f_n, \dots)$. Assume that $\mathscr{N}$ is a collection of martingales relative to $(\mathscr{F}_n)$ which is closed under starting and stopping (i.e., $f \in \mathscr{N}$ implies $^{\mu}f^{\nu} \in \mathscr{N}$ for all stopping times ν and μ satisfying $\nu \geq \mu$). We shall consider an operator T defined on a collection $\mathscr{N}$ with values in the class of all nonnegative random variables. We will assume that T satisfies the following conditions [cf. Burkholder and Gundy (1970)]:

(B1) T is quasilinear, that is, $T(f+g) \leq \gamma(T(f) + T(g))$, for some $\gamma > 0$ and all martingales $f, g \in \mathscr{N}$.

(B2) T is local, that is, $T(f) = 0$ on the set $\{s(f) = 0\}$, where

$$s(f) = \left(\sum E\left(d_k^2 | \mathscr{F}_{k-1}\right)\right)^{1/2}$$

is the conditional square function of f.

(B3) T is symmetric, that is, $T(f) = T(-f)$, for all martingales $f \in \mathscr{N}$.

If $\gamma = 1$ then T is sublinear. An operator T is called measurable (resp., predictable) if, for $n = 1, \dots,$ $T(f^n)$ is an $\mathscr{F}_n$-measurable (resp., $\mathscr{F}_{n-1}$-measurable) random variable. For example, the square function $S(f)$ defined by $S(f) = (\sum d_k^2)^{1/2}$ is a measurable operator, while the conditional square function is a predictable operator on the collection of all martingales relative to $(\mathscr{F}_n)$.

It was shown in Hitczenko (1990b) that under some mild assumptions on T there is a relationship between the size of the constant C_p appearing in the inequality

$$\|f^*\|_p \leq C_p \|T(f)\|_p$$

and the exponential bounds on the probabilities of the form

$$P(|f_n| \geq c\|T(f)\|_\infty).$$

In this paper, for $1 \leq p < \infty$, we define a sublinear operator T_p, such that

$$\|T_p(f^n)\|_p = \left\| \sum_{k=1}^{n} \bar{d}_k \right\|_p,$$

thus reducing the problem to investigation of the quantity $P(|f_n| \geq c\|T_p(f)\|_\infty)$.

The paper is organized as follows. The next section contains some preliminary materials. In Section 3 we shall define an operator on martingales. A proof of Theorem 1.1 will be given in Section 5. It will be preceded by Section

4, where we shall prove two results for sums of independent random variables. The second of those results will be needed in Section 5. Finally, the last section contains some comments.

2. Preliminaries. As we mentioned in the introduction, we will be interested in the size of the constant K appearing in the following domination inequality for sums of tangent sequences:

$$\left\| \sum X_k \right\|_p \le K \left\| \sum \bar{X}_k \right\|_p.$$

The first lemma allows one to reduce the whole problem to conditionally symmetric sequences.

LEMMA 2.1. *Suppose that the inequality*

$$\left\| \sum Z_k \right\|_p \le K \left\| \sum \bar{Z}_k \right\|_p$$

holds for all conditionally symmetric sequences (Z_k). *Then, for all sequences* (X_k), *we have*

$$\left\| \sum X_k \right\|_p \le (2K+1) \left\| \sum \bar{X}_k \right\|_p.$$

If, in addition, (X_k) *is assumed to be a martingale difference sequence, then the constant* $2K+1$ *can be replaced by* $2K$.

PROOF. We shall use a conditional symmetrization argument which was introduced in Hitczenko (1990a) [cf. also Kwapień and Woyczyński (1992), proof of Proposition 5.7.1]. Let (X_k) be any sequence of random variables and let $(\bar{X}_k)$ be its decoupled version. By definition, there exists a σ-algebra $\mathscr{G}$ such that $(\bar{X}_k)$ is a sequence of $\mathscr{G}$-conditionally independent random variables. Enlarging the probability space [and the filtration $(\mathscr{F}_n)$], if necessary, we can construct another sequence $(\tilde{X}_k)$ of $\mathscr{G}$-conditionally independent random variables such that, for each $j = 1, \dots,$ we have

$$\mathscr{L}\left(\tilde{X}_j | \mathscr{G}\right) = \mathscr{L}\left(\tilde{X}_j | \mathscr{F}_{j-1}\right) = \mathscr{L}\left(\bar{X}_j | \mathscr{F}_{j-1}\right),$$

X_j and $\tilde{X}_j$ are $\mathscr{F}_{j-1}$-conditionally independent and such that the sequences $(\bar{X}_k)$ and $(\tilde{X}_k)$ are $\mathscr{G}$-conditionally independent. Note that the sequence $(X_k - \tilde{X}_k)$ is conditionally symmetric and that $(\bar{X}_k - \tilde{X}_k)$ is its decoupled version. Moreover, the sequences $(\bar{X}_k)$ and $(\tilde{X}_k)$ are equidistributed. Therefore, by our assumption we conclude that

$$\begin{aligned} \left\| \sum X_k \right\|_p &\le \left\| \sum (X_k - \tilde{X}_k) \right\|_p + \left\| \sum \tilde{X}_k \right\|_p \\ &\le K \left\| \sum (\bar{X}_k - \tilde{X}_k) \right\|_p + \left\| \sum \bar{X}_k \right\|_p \\ &\le K \left\{ \left\| \sum \bar{X}_k \right\|_p + \left\| \sum \tilde{X}_k \right\|_p \right\} + \left\| \sum \bar{X}_k \right\|_p \\ &= (2K+1) \left\| \sum \bar{X}_k \right\|_p. \end{aligned}$$

This proves the first statement. For the second statement we observe that, if (X_k) is a martingale difference sequence, then

$$E\left(\sum \tilde{X}_k | \mathscr{G}\right) = \sum E\left(\tilde{X}_k | \mathscr{G}\right) = \sum E\left(X_k | \mathscr{F}_{k-1}\right) = 0.$$

Since we can assume that X_k's are $\mathscr{G}$-measurable, by the conditional version of Jensen's inequality we infer that

$$\begin{aligned}\left\| \sum X_k \right\|_p &= \left\| \sum X_k - E\left(\sum \tilde{X}_k | \mathscr{G}\right) \right\|_p \\ &= \left\| E\left(\sum (X_k - \tilde{X}_k) | \mathscr{G}\right) \right\|_p \\ &\le \left\| \sum (X_k - \tilde{X}_k) \right\|_p,\end{aligned}$$

and the rest follows by exactly the same argument as above. This completes the proof. □

The next lemma was introduced as a tool for proving certain martingale inequalities [cf. Burkholder (1973), Lemma 7.1]. As was realized later [see, e.g., Bañuelos (1988) or Hitczenko (1990a, b)], it could also be used to give quite precise information on the size of the constants involved in some of those inequalities.

LEMMA 2.2. *Fix $0 < p < \infty$. Suppose that X and Y are nonnegative random variables and that there exist positive numbers δ, $\beta > 1 + \delta$ and $\varepsilon < 1/\beta^p$ such that for all positive λ's one has*

$$P(X \ge \beta\lambda, Y < \delta\lambda) \le \varepsilon P(X \ge \lambda).$$

Then

$$\|X\|_p^p \le \frac{(\beta/\delta)^p}{1 - \beta^p \varepsilon} \|Y\|_p^p.$$

The following lemma gives a sufficient condition for a constant appearing in the inequality $\|f^*\|_p \le K\|T(f)\|_p$ to be independent of p.

LEMMA 2.3. *Fix p, $1 \le p < \infty$. Let $\mathscr{N}$ be a class of martingales, closed under starting and stopping, and suppose that T is a predictable sublinear operator on $\mathscr{N}$ satisfying*

$$\|d^*\|_p \le C\|T(f)\|_p,$$

for some absolute constant C and all $f \in \mathscr{N}$. Assume that there exist constants $A > 0$ and $\varepsilon < 1/2$ such that, for each $p \ge 1$ and all conditionally symmetric martingales f, we have

$$P(|f_n| \ge A\|T(f^n)\|_\infty) \le \varepsilon^p, \qquad n = 1, \dots .$$

Then the following inequality is true:

$$\|f_n^*\|_p \le K\|T(f^n)\|_p = K\|\bar{f}_n\|_p,$$

where

$$K \le 2^{2/p}\frac{A+1}{1-2\varepsilon}(C^p+1)^{1/p} \le \frac{4(A+1)}{1-2\varepsilon}(C+1).$$

PROOF (Sketch). Let A and ε be as above. Choose β and δ so that

$$\frac{\beta-1-\delta}{\delta} = A \quad\text{and}\quad \delta = \frac{1-2\varepsilon}{2\varepsilon(A+1)} > 0.$$

Then

$$\beta = \frac{1}{2\varepsilon} = 1 + \frac{1-2\varepsilon}{2\varepsilon}\frac{A+1}{A+1} = 1 + (A+1)\delta > 1+\delta.$$

Repeating now the usual stopping time argument given in Hitczenko [(1993b), proof of Lemma 2.2], which goes back to Burkholder (1973), we get the following good-λ inequality

$$P(|f_n| > \beta\lambda, T(f^n) \vee d_n^* \le \delta\lambda) \le \varepsilon^p P(|f_n| > \lambda), \qquad \lambda > 0.$$

This, in view of Lemma 2.2, implies that

$$\|f_n\|_p^p \le \frac{(\beta/\delta)^p}{1-\beta^p\varepsilon^p}\{\|T(f^n)\|_p^p + \|d_n^*\|_p^p\},$$

as long as $\beta^p\varepsilon^p < 1$. In our case

$$\beta^p\varepsilon^p = \left(\frac{1}{2\varepsilon}\right)^p\varepsilon^p = \left(\frac{1}{2}\right)^p \le \frac{1}{2}.$$

This completes the proof of Lemma 2.3. □

The final lemma in this section is a particular case of Lemma 1 in Hitczenko (1988) and is stated here for easy reference.

LEMMA 2.4. *Let (X_n) and (Y_n) be two tangent sequences of random variables. Then, for each $t > 0$, the following is true*:

$$P(X^* \ge t) \le 2P(Y^* \ge t).$$

3. An operator on martingales. In this section we introduce a new operator on martingales which is related to our domination inequality. For a martingale $f = (f_n)$ with difference sequence (d_n), we define the operator $T_p(f)$ by the formula

$$T_p(f^n) = \left(E_{\mathscr{G}}\left|\sum_{k=1}^n \bar{d}_k\right|^p\right)^{1/p}.$$

Then T_p satisfies conditions (B1)–(B3), with $\gamma = 1$. Note that for any martingale f the random variable $T_p(f^n)$ is $\mathscr{F}_{n-1}$-measurable; thus, T_p is a predictable operator. Let us also remark that if $p = 2$, then

$$\begin{aligned} T_2(f) &= \left(E_{\mathscr{G}}\left(\sum \bar{d}_k^2\right)\right)^{1/2} \\ &= \left(\sum E_{\mathscr{G}}(\bar{d}_k^2)\right)^{1/2} \\ &= \left(\sum E_{k-1}(\bar{d}_k^2)\right)^{1/2} = \left(\sum E_{k-1}(d_k^2)\right)^{1/2} \end{aligned}$$

so that T_2 is nothing but the conditional square function. Moreover, for $n = 1, \dots,$ we have that

$$\begin{aligned} \|T_p(f^n)\|_p &= \left(E\left(E_{\mathscr{G}}\left|\sum_{k=1}^n \bar{d}_k\right|^p\right)\right)^{1/p} \\ &= \left\|\sum_{k=1}^n \bar{d}_k\right\|_p. \end{aligned}$$

Therefore, the inequality

$$\|f_n^*\|_p \le K\left\|\sum_{k=1}^n \bar{d}_k\right\|_p$$

is equivalent to

$$\|f_n^*\|_p \le K\|T_p(f^n)\|_p.$$

When restricted to a class $\mathscr{M}\mathscr{S}$ of all conditionally symmetric martingales, the operator T_p has the following property:

LEMMA 3.1. *For every p, $1 \le p < \infty$, and all $f \in \mathscr{M}\mathscr{S}$ we have that*

$$\|d^*\|_p \le 2 \cdot 2^{2/p}\|T_p(f)\|_p \le 8\|T_p(f)\|_p.$$

This follows from Lemma 2.4 and Lévy's inequality.

4. Inequalities for sums of independent random variables. Through this section (ξ_k) will be a sequence of independent random variables which, for simplicity, will be assumed to be symmetric. We let

$$S_k = \sum_{j=1}^k \xi_j, \qquad k = 1, \dots.$$

The first inequality follows easily from a result established by Kwapień and Woyczyński [(1992), proof of Proposition 1.4.2].

PROPOSITION 4.1. *Let $p \ge q \ge 1$ be real numbers. Then there exists an absolute constant K such that, for every sequence $(\xi_k) \subset L_p$, the following* *inequality is true*:

$$\|S_n^*\|_p \leq K\frac{p}{q}\{\|S_n\|_q + \|\xi_n^*\|_p\}, \qquad n = 1, \dots .$$

PROOF. Our proof is just a repetition of an argument given by Kwapień and Woyczyński. They showed that for every positive t we have that

$$\|S_n^*\|_p^p \leq \left\{ \frac{t^{p/(p+1)} + \|\xi_n^*\|_p^{p/(p+1)}}{1 - \left(2P(|S_n| \geq t)\right)^{1/(p+1)}} \right\}^{p+1}.$$

Selecting $t = 4\|S_n\|_q$ and using Chebyshev's inequality, we infer that

$$\|S_n^*\|_p^p \leq \left\{ \frac{(4\|S_n\|_q)^{p/(p+1)} + \|\xi_n^*\|_p^{p/(p+1)}}{1 - (1/2)^{q/(p+1)}} \right\}^{p+1}.$$

Since

$$\frac{1}{1 - x^\alpha} \leq \frac{1}{\alpha(1 - x)},$$

for $0 \leq x \leq 1$ and $0 < \alpha < 1$, and

$$\left(u^{p/(p+1)} + v^{p/(p+1)}\right)^{(p+1)/p} \leq 2^{1/p}(u + v), \qquad u, v \geq 0,$$

we obtain that

$$\begin{aligned}
\|S_n^*\|_p &\leq 2^{(p+1)/p}\left(\frac{p+1}{q}\right)^{(p+1)/p} \left\{(4\|S_n\|_q)^{p/(p+1)} + \|\xi_n^*\|_p^{p/(p+1)}\right\}^{(p+1)/p} \\
&\leq 2^{(p+2)/p}\left(\frac{p+1}{q}\right)^{(p+1)/p} \{4\|S_n\|_q + \|\xi_n^*\|_p\} \\
&\leq K\frac{p}{q}\{\|S_n\|_q + \|\xi_n^*\|_p\}.
\end{aligned}$$

This completes the proof. □

Proposition 4.1 easily implies the following inequality, which will be needed in the next section.

PROPOSITION 4.2. *Suppose* $\|\xi^*\|_\infty < \infty$. *Then there exist absolute constants* $B > 0$ *and* $\delta > 0$ *such that, for every* $p \geq 1$ *and every*

$$\lambda \leq \frac{\delta p}{\max\{\|S_n\|_p, \|\xi^*\|_\infty\}},$$

we have that

$$E\exp\{\lambda|S_n|\} \leq 5\exp\{B\lambda(\|S_n\|_p + \|\xi^*\|_\infty)\}.$$

PROOF. Let r be the largest integer less than or equal to p. Expanding the exponential function into power series and integrating term by term, we obtain that

$$\begin{aligned} E\exp\{\lambda|S_n|\} &= 1 + \sum_{j=1}^{\infty}\frac{\lambda^j}{j!}\|S_n\|_j^j \\ &= 1 + \sum_{j=1}^{r}\frac{\lambda^j}{j!}\|S_n\|_j^j + \sum_{j=r+1}^{\infty}\frac{\lambda^j}{j!}\|S_n\|_j^j \\ &\le 1 + \sum_{j=1}^{r}\frac{\lambda^j}{j!}\|S_n\|_p^j + \sum_{j=r+1}^{\infty}\frac{\lambda^j}{j!}\|S_n\|_j^j \\ &\le \exp\{\lambda\|S_n\|_p\} + \sum_{j=r+1}^{\infty}\frac{\lambda^j}{j!}\|S_n\|_j^j, \end{aligned}$$

and we need to estimate the second sum. By Proposition 4.1 we infer that, for $j \ge p$,

$$\|S_n\|_j^j \le \left(\frac{Kj}{p}\right)^j \{\|S_n\|_p^j + \|\xi^*\|_j^j\} \le \left(\frac{Kj}{p}\right)^j \{\|S_n\|_p^j + \|\xi^*\|_\infty^j\}.$$

Since $j^j/j! \le e^j$, we can write

$$\begin{aligned} \sum_{j=r+1}^{\infty}\frac{\lambda^j}{j!}\|S_n\|_j^j &\le \sum_{j=r+1}^{\infty}\frac{j^j}{j!}\left(\frac{K\lambda}{p}\right)^j (\|S_n\|_p^j + \|\xi^*\|_\infty^j) \\ &\le \sum_{j=r+1}^{\infty} e^jK^j\left(\frac{\lambda}{p}\right)^j (\|S_n\|_p^j + \|\xi^*\|_\infty^j). \end{aligned}$$

If λ is chosen so that

$$\frac{eK\lambda}{p}\max\{\|S_n\|_p, \|X^*\|_\infty\} \le \frac{1}{2},$$

then the above sum does not exceed

$$\begin{aligned} &\frac{e^{r+1}K^{r+1}(\lambda/p)^{r+1}\|S_n\|_p^{r+1}}{1 - eK(\lambda/p)\|S_n\|_p} + \frac{e^{r+1}K^{r+1}(\lambda/p)^{r+1}\|\xi^*\|_\infty^{r+1}}{1 - eK(\lambda/p)\|\xi^*\|_\infty} \\ &\le 4e^pK^p\left(\frac{\lambda}{p}\right)^p (\|S_n\|_p^p + \|\xi^*\|_\infty^p). \end{aligned}$$

Since $x^p \le \exp\{px\}$ for $x > 0$, the last quantity can be estimated from above by

$$4\exp\{eK\lambda(\|S_n\|_p + \|\xi^*\|_\infty).$$

This completes the proof. □

5. Proof of Theorem 1.1. It follows from the previous discussion that, in order to complete the proof of Theorem 1.1, it suffices to check the assumptions of Lemma 2.3. In view of Lemma 3.1 this will be done once we establish the following result.

THEOREM 5.1. *There exist absolute constants* $A > 0$ *and* $0 < \varepsilon < 1/2$ *such that, for every* $p \geq 1$ *and for all conditionally symmetric martingales* f, *the following inequality is true*:

$$P\big(|f_n| \geq A\|T_p(f^n)\|_\infty\big) \leq \varepsilon^p, \qquad n = 1, \dots .$$

PROOF. We shall show first that it is enough to prove the inequality of the theorem under the additional assumption that

$$\|d^*\|_\infty \leq C\|T_p(f)\|_\infty,$$

for some $C > 4$. This follows easily from Davis' decomposition of a martingale and Lemma 2.4. Indeed, given a conditionally symmetric martingale difference sequence (d_n), let us write

$$d_n = d'_n + d''_n = d_n I(|d_n| \leq 2d^*_{n-1}) + d_n I(|d_n| > 2d^*_{n-1}).$$

Since (d_n) is conditionally symmetric, both (d'_n) and (d''_n) are martingale difference sequences. We denote the corresponding martingales by (f'_n) and (f''_n), respectively. It was shown by Davis (1970) [cf. also Burkholder (1973)] that

$$|f''_n| \leq \sum |d''_k| \leq 4d^*.$$

Therefore, for every $A > 0$,

$$\begin{aligned} &P\big(|f_n| \geq A\|T_p(f)\|_\infty\big) \\ &\quad \leq P\big(|f'_n| \geq (A/2)\|T_p(f)\|_\infty\big) + P\big(|f''_n| \geq (A/2)\|T_p(f)\|_\infty\big) \\ &\quad \leq P\big(|f'_n| \geq (A/2)\|T_p(f)\|_\infty\big) + P\big(d^* \geq (A/8)\|T_p(f)\|_\infty\big). \end{aligned}$$

To estimate the first term let us define a stopping time τ by

$$\tau = \inf\{n \geq 1 : |d_n| > (C/2)\|T_p(f)\|_\infty\}.$$

Then we have that

$$\begin{aligned} &P\big(|f'_n| \geq (A/2)\|T_p(f)\|_\infty\big) \\ &\quad \leq P\big(|f'_n| \geq (A/2)\|T_p(f)\|_\infty, \tau = \infty\big) + P(\tau < \infty) \\ &\quad \leq P\Big(\Big|\sum I(\tau \geq k) d'_k\Big| \geq (A/2)\|T_p(f)\|_\infty\Big) + P\big(d^* \geq (C/2)\|T_p(f)\|_\infty\big). \end{aligned}$$

Combining these estimates, we obtain that

$$\begin{aligned} P\big(|f_n| \ge A\|T_p(f)\|_\infty\big) &\le P\big(\big|\sum I(\tau \ge k)d'_k\big| \ge (A/2)\|T_p(f)\|_\infty\big) \\ &\quad + P\big(d^* \ge (C/2)\|T_p(f)\|_\infty\big) \\ &\quad + P\big(d^* \ge (A/8)\|T_p(f)\|_\infty\big). \end{aligned}$$

Now, observe that the martingale difference sequence $(I(\tau \ge k)d'_k)$ has the required property since

$$|d'_k| I(\tau \ge k) \le 2d^*_{k-1} I(\tau \ge k) \le C\|T_p(f)\|_\infty.$$

Thus, in order to complete this part of the proof it suffices to show that, for $\gamma > 1$,

$$P\big(d^* \ge \gamma\|T_p(f)\|_\infty\big) \le B\gamma^{-p},$$

for some absolute constant B. This follows immediately from the estimate

$$\|T_p(f)\|_\infty \ge \|T_p(f)\|_p,$$

and Lemmas 2.4 and 3.1. This completes the first part of the proof.

Assume now that (d_n) is uniformly bounded. Then, for each positive λ and each $k = 1, \dots,$

$$E\exp\{\lambda d_k\} < \infty,$$

so that a sequence (Y_n) defined by

$$Y_n = \frac{\exp\{\lambda \sum_{k=1}^n d_k\}}{\prod_{k=1}^n E_{k-1}\exp\{\lambda d_k\}}$$

is a martingale with $EY_n = 1$. Consequently,

$$\begin{aligned} P\left(\sum_{k=1}^n d_k \ge A\|T_p(f)\|_\infty\right) &= P\left(\exp\left\{\lambda \sum_{k=1}^n d_k\right\} \ge \exp\big\{A\lambda\|T_p(f)\|_\infty\big\}\right) \\ &= P\left(Y_n \ge \frac{\exp\big\{A\lambda\|T_p(f)\|_\infty\big\}}{\prod_{k=1}^n E_{k-1}\exp\{\lambda d_k\}}\right) \\ &\le P\left(Y_n \ge \frac{\exp\big\{A\lambda\|T_p(f)\|_\infty\big\}}{\|\prod_{k=1}^n E_{k-1}\exp\{\lambda d_k\}\|_\infty}\right) \\ &\le \exp\big\{-A\lambda\|T_p(f)\|_\infty\big\}\left\|\prod_{k=1}^n E_{k-1}\exp\{\lambda d_k\}\right\|_\infty EY_n \\ &= \exp\big\{-A\lambda\|T_p(f)\|_\infty\big\}\left\|\sum_{k=1}^n E_{k-1}\exp\{\lambda d_k\}\right\|_\infty. \end{aligned}$$

Since d_k and $\bar{d}_k$ have identical conditional distributions, given $\mathscr{F}_{k-1}$, we find that

$$E_{k-1}\exp\{\lambda d_k\} = E_{k-1}\exp\{\lambda \bar{d}_k\} = E_{\mathscr{G}}\exp\{\lambda \bar{d}_k\},$$

and it follows from the $\mathscr{G}$-conditional independence of the sequence $(\bar{d}_k)$ that

$$\begin{aligned}\prod_{k=1}^{n} E_{k-1}\exp\{\lambda d_k\} &= \prod_{k=1}^{n} E_{k-1}\exp\{\lambda \bar{d}_k\}\\ &= \prod_{k=1}^{n} E_{\mathscr{G}}\exp\{\lambda \bar{d}_k\}\\ &= E_{\mathscr{G}}\exp\left\{\lambda \sum_{k=1}^{n} \bar{d}_k\right\}.\end{aligned}$$

Since

$$\|d^*\|_\infty \le C\|T_p(f)\|_\infty,$$

Lemma 2.4 implies that

$$\|\bar{d}^*\|_\infty \le C\|T_p(f)\|_\infty,$$

as well. Choose

$$\lambda = \frac{\kappa p}{\|T_p(f)\|_\infty},$$

with $\kappa > 0$ to be specified later. Applying Proposition 4.2 to the sequence $(\bar{d}_k)$ and the conditional measure $P(\cdot|\mathscr{G})$, we get

$$\begin{aligned}\left\|E_{\mathscr{G}}\exp\left\{\lambda \sum_{k=1}^{n} \bar{d}_k\right\}\right\|_\infty &\le 5\exp\left\{B\lambda\left(\|T_p(f)\|_\infty + C\|T_p(f)\|_\infty\right)\right\}\\ &\le 5\exp\left\{2BC\lambda\|T_p(f)\|_\infty\right\}.\end{aligned}$$

Substituting the last quantity into our previous estimate and using our choice of λ, we obtain

$$\begin{aligned}P\left(f_n \ge A\|T_p(f)\|_\infty\right) &\le 5\exp\{-A\lambda\|T_p(f)\|_\infty\}\exp\{5BC\lambda\|T_p(f)\|_\infty\}\\ &\le 5\exp\{-(A-2BC)\lambda\|T_p(f)\|_\infty\}\\ &\le 5\exp\{-(A-2BC)\kappa p\},\end{aligned}$$

as long as $\kappa < 1/(2BC)$. Repeating the same argument for the sequence $(-d_k)$ and putting together all of the above estimates, we get

$$P\left(|f_n| \ge A\|T_p(f^n)\|_p\right) \le 4(2/C)^p + 4(8/A)^p + 10\exp\{-(A-2BC)\kappa p\}.$$

Now it is clear that, for $C > 16$, it is enough to let $\kappa = 1/(2BC)$ and $A = \eta BC$, and η sufficiently large. This completes our proof. □

6. Concluding remarks.

(i) Since all constants appearing in the course of our proof are explicit, a numerical upper bound on a constant K can be constructed. No attempts were made to optimize that bound, as methods based on the good-λ inequality are not believed to produce sharp estimates. As for a lower bound, by considering a trivial example $(X_1 = X_2 = I_A)$ and letting $P(A) \to 0$, we see that $K \geq 2^{(p-1)/p}$. A question about the exact value of that constant is open even in the special case of randomly stopped sums of independent random variables.

(ii) Consider the following result of Klass: There exists a constant K such that for every sequence (X_k) of independent, Banach space valued random variables and every stopping time τ, one has

$$E \max_{1 \leq n \leq \tau} \Phi\left(\left\| \sum_{k=1}^{n} X_k \right\|\right) \leq K^\alpha E \max_{1 \leq n \leq \tau} \Phi\left(\left\| \sum_{k=1}^{n} X_k' \right\|\right).$$

Here, (X_k') is an independent copy of (X_k), and Φ is any increasing continuous function of moderate growth [i.e., $\Phi(0) = 0$ and $\Phi(cx) \leq c^\alpha \Phi(x)$, for all $x \geq 0$ and $c \geq 2$]. Our method can be used to give an alternative proof in the special case, namely, $\Phi(x) = |x|^p$, $1 \leq p < \infty$. For convenience, we shall consider real-valued random variables; it should be clear, however, that the proof carries over to Banach spaces with no essential changes. Let (ξ_k) be a sequence of independent mean-zero random variables, and let τ be a stopping time with respect to the natural filtration generated by a sequence (ξ_k) such that $\|\tau\|_\infty = N$. Then,

$$\|T_p(f)\|_\infty = \left\| \sum_{k=1}^{N} \xi_k \right\|_p.$$

Therefore, by a submartingale property, we have that

$$P\left(\max_{k \leq N} \left| \sum_{k=1}^{\tau \wedge k} \xi_k \right| \geq t \right) \leq t^{-p} E \left| \sum_{k=1}^{N} \xi_k \right|^p;$$

selecting $t = c\|\sum_{k=1}^{N} \xi_k\|_p$, we obtain

$$P\left(\max_{k \leq N} \left| \sum_{j=1}^{k \wedge \tau} \xi_k \right| \geq c \left\| T_p\left(\sum_{k=1}^{\tau} \xi_k \right) \right\|_\infty \right) \leq (1/c)^p.$$

The result follows by Lemma 2.3. One advantage of our approach is that it gives a better constant then the original proof of Klass. This may be of some interest in the context of identifying the best value of the constant in domination inequality for martingales. It may be (as happens with other martingale inequalities) that the special situation of randomly stopped sums of independent random variables, in fact, gives rise to the worst case.

(iii) Theorem 1.1 yields an inequality

$$E\Phi\left(\left| \sum X_k \right|\right) \leq E\Phi\left(K \left| \sum \bar{X}_k \right|\right)$$

for more general functions Φ: $\mathbb{R} \to \mathbb{R}$ than powers. For example, taking $\Phi(t) = \exp\{t\}$ and using Taylor expansion, we get a domination inequality for the moment generating function

$$E \exp\left\{\left|\sum X_k\right|\right\} \le E \exp\left\{K\left|\sum \bar{X}_k\right|\right\}.$$

This is closely related to a recent result of de la Peña (1992), who showed that

$$E \exp\left\{\left|\sum X_k\right|\right\} \le 2\sqrt{E \exp\left\{2\left|\sum \bar{X}_k\right|\right\}}.$$

We would like to mention at this point that an upper bound for the probability

$$P\left(\sum_{k=1}^{n} d_k \ge A\|T_p(f)\|_\infty\right)$$

obtained in the course of the proof of Theorem 5.1 could also be deduced from de la Peña's result. Both arguments are based on essentially the same ideas and were given independently. For the sake of self-containment of the present paper we have chosen to present our version of the argument rather than to refer simply to de la Peña's inequality.

Acknowledgments. It is a pleasure to thank S. Kwapień and M. Klass for their useful remarks.

REFERENCES

Bañuelos, R. (1988). A sharp good-λ inequality with an application to Riesz transforms. *Michigan Math. J.* **35** 117–125.

Burkholder, D. L. (1973). Distribution function inequalities for martingales. *Ann. Probab.* **1** 19–42.

Burkholder, D. L., Davis, B. and Gundy, R. F. (1972). Integral inequalities for convex functions of operators on martingales. *Proc. Sixth Berkeley Symp. Math. Statist. Probab.* **2** 223–240. Univ. California. Press, Berkeley.

Burkholder, D. L. and Gundy, R. F. (1970). Extrapolation and interpolation of quasilinear operators on martingales. *Acta Math.* **124** 250–304.

Davis, B. (1970). On the integrability of the martingale square function. *Israel J. Math.* **8** 187–190.

de la Peña, V. H. (1992). A bound on the moment generating function of a sum of dependent random variables with an application to simple random sampling without replacement. Unpublished manuscript.

Hitczenko, P. (1988). Comparison of moments for tangent sequences of random variables. *Probab. Theory Related Fields* **78** 223–230.

Hitczenko, P. (1990a). Best constants in martingale version of Rosenthal's inequality. *Ann. Probab.* **18** 1656–1668.

Hitczenko, P. (1990b). Upper bounds for the L_p-norms of martingales. *Probab. Theory Related Fields* **86** 225–238.

Hitczenko, P. (1993a). On the behavior of the constant in a decoupling inequality for martingales. *Proc. Amer. Math. Soc.* To appear.

Hitczenko, P. (1993b). Domination inequality for martingale transforms of Rademacher sequence. *Israel J. Math.* to appear.

Klass, M. J. (1988). A best possible improvement of Wald's equation. *Ann. Probab.* **16** 840–853.

Klass, M. J. (1990). Uniform lower bounds for randomly stopped sums of Banach space valued random sums. *Ann. Probab.* **18** 790–809.

KWAPIEŃ, S. (1987). Decoupling inequalities for polynomial chaos. *Ann. Probab.* **15** 1062–1071.

KWAPIEŃ, S. and WOYCZYŃSKI, W. A. (1986). Semimartingale integrals via decoupling inequalities and tangent processes. Case Western Reserve Univ. Preprint 86–56.

KWAPIEŃ, S and WOYCZYŃSKI, W. A. (1989). Tangent sequences of random variables: Basic inequalities and their applications. In *Proceedings of the Conference on Almost Everywhere Convergence in Probability and Ergodic Theory* (G. A. Edgar and L. Sucheston, eds.) 237–265. Academic, New York.

KWAPIEŃ, S. and WOYCZYŃSI, W. A. (1992). *Random Series and Stochastic Integrals. Single and Multiple*. Birkhäuser, Boston.

MCCONNELL, T. R. (1990). Decoupling and integration in UMD Banach space. *Probab. Math. Statist.* **10** 283–295.

ZINN, J. (1985). Comparison of martingale differences. *Probability in Banach Spaces 5. Lecture Notes in Math.* **1153** 453–457. Springer, Berlin.

DEPARTMENT OF MATHEMATICS
NORTH CAROLINA STATE UNIVERSITY
RALEIGH, NORTH CAROLINA 27695-8205