The Annals of Probability
1990, Vol. 18, No. 4, 1656–1668

BEST CONSTANTS IN MARTINGALE VERSION OF ROSENTHAL'S INEQUALITY[1,2]

BY PAWEŁ HITCZENKO

Texas A & M University

The following generalization of Rosenthal's inequality was proved by Burkholder:

$$A_p^{-1}\{\|s(f)\|_p + \|d^*\|_p\} \le \|f^*\|_p \le B_p\{\|s(f)\|_p + \|d^*\|_p\},$$

for all martingales (f_n). It is known that A_p grows like $\sqrt{p}$ as $p \to \infty$. In this paper we prove that the growth rate of B_p as $p \to \infty$ is $p/\ln p$.

1. Introduction. Let (d_n) be a martingale difference sequence with respect to an increasing sequence of σ-algebras $(\mathscr{F}_n)$. The following inequality was proved by Burkholder (1973): For $2 \le p < \infty$,

$$A_p^{-1}\left\{\left(E\left(\sum E_{k-1}d_k^2\right)^{p/2}\right)^{1/p} + \left(E\sup_k|d_k|^p\right)^{1/p}\right\}$$
$$\le \left(E\left|\sum d_k\right|^p\right)^{1/p} \le B_p\left\{\left(E\left(\sum E_{k-1}d_k^2\right)^{p/2}\right)^{1/p} + \left(E\sup_k|d_k|^p\right)^{1/p}\right\},$$

where A_p and B_p are constants depending only on p (see the next section for notation). The special case of independent random variables which is a fundamental generalization of Khintchine's inequality was proved by Rosenthal (1970). Rosenthal's proof yielded only exponential of p estimate for the growth rate of B_p as $p \to \infty$. Later on, Johnson, Schechtman and Zinn (1985) showed that the best possible bound on B_p (still, in the independent case) is $p/\ln p$. Using difficult isoperimetric techniques, Talagrand (1989) extended this result to the case of independent Banach-space-valued random variables. Recently, Kwapień and Szulga (1988) obtained a completely elementary proof of Talagrand's result. As to the general case of martingale difference sequences, it seemed that Burkholder's argument, which was based on the good λ inequality, gave exponential-type estimates on B_p. However, as was pointed out by Hall and Marron (1988), page 170, the careful choice of various parameters involved in the good λ inequality shows that $B_p \le K \cdot p(\ln p)^s$, where K is an absolute constant and s is any positive number. We would like to mention that Garsia's (1973) book contains "almost explicitly" a linear of p bound on B_p. More precisely, Garsia (1973), Theorem 3.5.1, proves that for any martingale

Received March 1989; revised July 1989.
[1]Preparation of the final version of this paper was supported in part by NSF Grant DMS-89-02418.
[2]On leave from Academy of Physical Education, Warsaw, Poland.
AMS 1980 *subject classifications*. 60E15, 60G42.
Key words and phrases. Martingale, moment inequalities, good λ inequality.

difference sequence (d_n), the following is true:

$$\left(E\left|\sum d_k\right|^p\right)^{1/p} \le 5p\left(\frac{2p}{p-1}\right)^{1/2}(E|\gamma|^p)^{1/p}, \qquad p \ge 2,$$

where γ is any function in L_p such that

$$E_n\left(\sum_{k=n}^{\infty} d_k\right)^2 \le E_n\gamma^2 \quad \text{a.s.}, n \ge 1.$$

Since

$$E_n\left(\sum_{k=n}^{\infty} d_k\right)^2 = d_n^2 + E_n\left(\sum_{k=n+1}^{\infty} E_{k-1}d_k^2\right) \quad \text{a.s.},$$

one can take

$$\gamma^2 = \sup_k |d_k|^2 + \sum_{k=1}^{\infty} E_{k-1}d_k^2,$$

which shows that $B_p \le 10p$. This bound can be further improved. We will prove below that the actual growth rate of B_p is, as in the independent case, $p/\ln p$. Let us mention in passing that the constant A_p is known to be of order $p^{1/2}$. Indeed, as is well known

$$\left(E\left(\sum E_{k-1}d_k^2\right)^{p/2}\right)^{1/2} \le K\cdot p^{1/2}\cdot\left(E\left|\sum d_k\right|^p\right)^{1/p}$$

holds, for all martingale difference sequences (d_n) [see, e.g., Garsia (1973), Theorem 4.3.1], and, on the other hand, for the "double or nothing" sequence $d_k = (-1)^k \cdot r_k \cdot I_{[0,2^{-k+1})}$, $k \ge 1$, where (r_k) is Rademacher sequence, we have

$$\left(E\left|\sum d_k\right|^p\right)^{1/p} \approx 1, \qquad |d_k| \le 1,$$

and

$$\left(E\left(\sum E_{k-1}d_k^2\right)^{p/2}\right)^{1/p} \approx p^{1/2}.$$

Our proof will exploit some ideas of Burkholder (1973, 1977), Johnson, Schechtman and Zinn (1985), as well as a remark of Hall and Marron mentioned above. Let us recall that in order to conclude the inequality $E|X|^p \le K_p^p E|Y|^p$ for some random variables X and Y, it suffices to establish the following good λ inequality:

$$P(|X| > \beta\lambda, |Y| \le \delta\lambda) \le \varepsilon(\alpha)\cdot P(|X| > \lambda), \tag{1}$$

for some $\delta > 0$, $\beta > 1 + \delta$, all positive λ's and sufficiently small $\varepsilon(\alpha)$, where $\alpha = (\beta - 1 - \delta)/\delta$. If $X = \sum d_k$ and $Y = (\sum E_{k-1}d_k^2)^{1/2} \vee \sup_k|d_k|$, the original proof of Burkholder (1973) gives only polynomial decay of ε as a function of α [actually, he obtained $\varepsilon(\alpha) = \alpha^{-2}$]. It turns out, however, that much more is true. We will prove (1) with $\varepsilon(\alpha) = 2\exp(-(\alpha/2)\ln(1 + \alpha/2))$. Once this is established, minimalization over all possible choices of $0 < \delta < \beta - 1$ yields $K_p \leq K \cdot p/\ln p$, for some absolute constant K.

Let us mention that inequalities of type (1) with exponential dependence of ε as a function of α have already been used in probability [see Burkholder (1977)] as well as in harmonic analysis [see, e.g., Murai and Uchiyama (1986) and references therein].

The main tool used in the proof of our good λ inequality is a refinement of Prokorov's "arc sinh" inequality for martingales obtained by Johnson, Schechtman and Zinn (1985). We also use Davis' decomposition of a martingale and conditional symmetrization argument. The notion of tangent sequences introduced quite recently by Kwapień and Woyczyński (1988, 1989) turned out to be very useful in that context.

The paper is organized as follows: The next section collects some definitions and preliminary material. The good λ inequality, as well as Prokorov's inequality, are proved in Section 3, and the main result is derived in Section 4. In Section 5 we apply our martingale result to obtain optimal constants in similar inequality for nonnegative random variables.

2. Preliminaries. This section contains preliminary facts needed for the proof of our result. We will also fix some notation.

Given an increasing sequence of σ-algebras $(\mathscr{F}_n)$ on some probability space $(\Omega, \mathscr{F}, P)$, we denote by $E_{k-1}(\cdot) = E(\cdot \mid \mathscr{F}_{k-1})$ the conditional expectation operator (with the convention that $E_0 = E$, the expectation operator). The L_p-norm of a random variable X is denoted by $\|X\|_p$, $1 \leq p \leq \infty$. A sequence (f_n) of integrable random variables is a martingale (resp. supermartingale) if f_n is $\mathscr{F}_n$-measurable and $E_{n-1} f_n = f_{n-1}$ (resp. $E_{n-1} f_n \leq f_{n-1}$), $n \geq 1$. A sequence (d_n), where $d_n = f_n - f_{n-1}$, is called a martingale difference sequence [of a martingale (f_n)]. For any martingale (f_n) with difference sequence (d_n), following standard notation, we will write

$$s(f) = \left(\sum_{k=1}^{\infty} E_{k-1} d_k^2 \right)^{1/2}$$

and

$$s_n(f) = \left(\sum_{k=1}^{n} E_{k-1} d_k^2 \right)^{1/2}.$$

For any sequence (X_n) of random variables X^* denotes $\sup_{n \geq 1} |X_n|$ and

$$X_n^* = \max_{1 \leq k \leq n} |X_k|.$$

A sequence (X_n) is $(\mathscr{F}_n)$-adapted [resp. $(\mathscr{F}_n)$-predictable] if X_n is $\mathscr{F}_n$-measurable (resp. $\mathscr{F}_{n-1}$-measurable) random variable, for $n \geq 1$. In the sequel we will simply write adapted (resp. predictable) without any risk of confusion. The indicator function of a set A is denoted by $I(A)$. Recall, that if τ is a stopping time (i.e., a positive integer-valued random variable such that $\{\tau = n\} \in \mathscr{F}_n$, for each $n \geq 1$), then $\mathcal{F}_\tau$ is the σ-algebra of all $\mathcal{F}$-measurable sets A such that $A \cap \{\tau = n\} \in \mathcal{F}_n$, for all $n \geq 1$.

We will need a notion of tangent sequences which was introduced by Kwapień and Woyczyński (1988, 1989).

DEFINITION 2.1. (a) Two adapted sequences of random variables (X_n) and (Y_n) are said to be tangent if for each real number λ and for all $n \geq 1$, $P(X_n \geq \lambda | \mathcal{F}_{n-1}) = P(Y_n \geq \lambda | \mathcal{F}_{n-1})$ a.s.

(b) An adapted sequence (X_n) is conditionally symmetric if (X_n) and $(-X_n)$ are tangent sequences of random variables.

We will denote the conditional distribution of a random variable X, given a σ-algebra $\mathcal{G}$ by $\mathcal{L}(X|\mathcal{G})$. Thus, the above definition says that (X_n) and (Y_n) are tangent if for each $n \geq 1$, $\mathcal{L}(X_n|\mathcal{F}_{n-1}) = \mathcal{L}(Y_n|\mathcal{F}_{n-1})$ a.s.

The usefulness of that notion stems from the fact that, on one hand, any sequence of random variables admits a tangent sequence which, in a sense, behaves like a sequence of independent random variables and, on the other hand, there is a remarkable similarity in behavior of two tangent sequences. As a consequence, some properties of arbitrarily dependent random variables can be deduced from the corresponding results for independent random variables. For our purpose we will need only the simplest properties of tangent sequences; we refer the reader to the papers of Kwapień and Woyczyński (1988, 1989) for the full exposition of the above ideas. Let us introduce one more notion used by these authors.

DEFINITION 2.2. An adapted sequence (Y_n) of random variables satisfies condition (CI) if there exists a σ-algebra $\mathcal{G} \subset \mathcal{F}$ such hat $\mathcal{L}(Y_n|\mathcal{F}_{n-1}) = \mathcal{L}(Y_n|\mathcal{G})$ a.s., $n \geq 1$ and such that (Y_n) is a sequence of $\mathcal{G}$-conditionally independent random variables.

For a given sequence (X_n) there is a canonical way to construct (perhaps on an enlarged probability space) a sequence (Y_n) which satisfies condition (CI) and is tangent to (X_n). Let us state this fact as a lemma.

LEMMA 2.3. *Let $\mathcal{B}_n$ be the σ-algebra generated by the first n coordinates in $\mathbb{R}^{\mathbb{N}}$, $\mathcal{B} = \sigma(\bigcup \mathcal{B}_n)$ and let (X_n) be any adapted sequence of random variables on $(\Omega, \mathcal{F}, P)$. Define a new probability space $(\overline{\Omega}, \overline{\mathcal{F}}, \overline{P})$ and a sequence $(\overline{\mathcal{F}}_n)$ by the formulas*

$$\overline{\Omega} = \Omega \times \mathbb{R}^{\mathbb{N}}, \qquad \overline{\mathcal{F}} = \mathcal{F} \otimes \mathcal{B}, \qquad \overline{\mathcal{F}}_n = \mathcal{F}_n \otimes \mathcal{B}_n$$

and

$$\overline{P}(A \times B) = \int_A \left(\bigotimes_{n=1}^{\infty} \mathcal{L}(X_n | \mathcal{F}_{n-1}) \right)(B), \qquad A \in \mathcal{F}, B \in \mathcal{B}$$

[*here, $\mathcal{L}(X_n|F_{n-1})$ is a regular version of the conditional distribution of X_n,* *given* $\mathscr{F}_{n-1}$; *cf., e.g., Shiryaev* (1984)]. *If* $\overline{X}_n(\omega,(t_k)) = X_n(\omega)$ *and* $\overline{Y}_n(\omega,(t_k)) = t_n$, *then the sequences* $(\overline{X}_n)$ *and* $(\overline{Y}_n)$ *are tangent, and* $(\overline{Y}_n)$ *satisfies condition* (*CI*) *with* $\mathscr{G} = \mathscr{F} \otimes \{\phi, \mathbb{R}^{\mathbb{N}}\}$.

REMARK. Note that the above construction shows that the random variables $\overline{X}_n$ and $\overline{Y}_n$ are $\overline{\mathscr{F}}_{n-1}$-conditionally independent, $n \ge 1$. In particular, the sequence $(\overline{X}_n - \overline{Y}_n)$ is conditionally symmetric.

Sequences with property (CI) share many properties of sequences of independent random variables. Thus, for example, we have the following lemma.

LEMMA 2.4. *Let* (g_n) *be a martingale, such that the difference sequence* (e_n) *satisfies condition* (*CI*). *Then, for* $2 \le p < \infty$ *we have*

$$\|g^*\|_p \le \frac{K \cdot p}{\operatorname{Log} p}\{\|s(g)\|_p + \|e^*\|_p\},$$

where $\operatorname{Log} x = \max\{1, \ln x\}$ *and* K *is an absolute constant.*

PROOF. By the results of Johnson, Schechtman and Zinn (1985) mentioned above, for every sequence (ξ_n) of independent mean-zero random variables we have

$$\left\|\sum \xi_n\right\|_p \le \frac{K \cdot p}{\operatorname{Log} p}\left\{\left\|\sum \xi_n\right\|_2 + \|\xi^*\|_p\right\}.$$

Applying this inequality to the $\mathscr{G}$-conditionally independent sequence (e_n), we get

$$\left(E\left|\sum e_n\right|^p|\mathscr{G}\right)^{1/p} \le \frac{K \cdot p}{\operatorname{Log} p}\left\{\left(E\left(\sum e_n\right)^2|\mathscr{G}\right)^{1/2} + (Ee^{*p}|\mathscr{G})^{1/p}\right\}.$$

Since

$$\begin{aligned}\left(E\left(\sum e_n\right)^2|\mathscr{G}\right)^{1/2} &= \left(E\sum e_n^2|\mathscr{G}\right)^{1/2}\\ &= \left(\sum Ee_n^2|\mathscr{G}\right)^{1/2} = \left(\sum E_{n-1}e_n^2\right)^{1/2} = s(g),\end{aligned}$$

we obtain

$$\begin{aligned}\left\|\sum e_k\right\|_p = \left\|\left(E\left|\sum e_k\right|^p|\mathscr{G}\right)^{1/p}\right\|_p &\le \frac{K \cdot p}{\operatorname{Log} p}\{\|s(g)\|_p + \|(Ee^{*p}|\mathscr{G})^{1/p}\|_p\}\\ &\le \frac{K \cdot p}{\operatorname{Log} p}\{\|s(g)\|_p + \|e^*\|_p\},\end{aligned}$$

as desired. □

The last lemma in this section is a particular case of the Lemma 1 in Hitczenko (1988).

LEMMA 2.5. *If* (X_n) *and* (Y_n) *are any tangent sequences of random variables, then for each positive number* λ *the following inequality holds*:

$$P(X^* > \lambda) \leq 2P(Y^* > \lambda).$$

Throughout, the letter K will always denote an absolute constant, not necessarily the same from one use to the next.

3. The good λ inequality. We begin this section with the following version of Prokorov's "arc sinh" inequality for martingales. Our result refines Proposition 3.1 of Johnson, Schechtman and Zinn (1985).

PROPOSITION 3.1. *Let* (f_n) *be a mean-zero martingale such that* $|d_k| \leq M$ *a.s.*, $k \geq 1$ *and* $\|s^2(f)\|_\infty = K^2 < \infty$. *Then, for each* $\lambda > 0$,

$$P\left(\left|\sum d_k\right| \geq \lambda\right) \leq 2\exp\left(\frac{-\lambda}{2M} \cdot \operatorname{arc\,sinh}\left(\frac{M\lambda}{2K^2}\right)\right).$$

PROOF. Let, for $c > 0$,

$$Y_n = \exp\left(c \cdot \sum_{k=1}^{n} d_k - \frac{c}{M}\sinh cM \cdot s_n^2(f)\right).$$

Then (Y_n) is a supermartingale. Indeed,

$$E_n Y_{n+1} = Y_n \exp\left(-\frac{c}{M}\sinh cM \cdot E_n d_{n+1}^2\right) \cdot E_n e^{cd_{n+1}},$$

and it suffices to check that

$$E_n \exp cd_{n+1} \leq \exp\left(\frac{c}{M}\sinh cM \cdot E_n d_{n+1}^2\right).$$

But, since $x \leq e^{x-1}$ and $e^x - x - 1 \leq e^x + e^{-x} - 2 \leq x \sinh x$, for all real x, we can write

$$E_n \exp cd_{n+1} \leq \exp\left(E_n\left(e^{cd_{n+1}} - 1\right)\right),$$

and then

$$\begin{aligned} E_n\left(c^{cd_{n+1}} - 1\right) &= E_n\left(e^{cd_{n+1}} - cd_{n+1} - 1\right) \leq E_n cd_{n+1} \cdot \sinh cd_{n+1} \\ &\leq E_n c|d_{n+1}|\sinh c|d_{n+1}| \\ &= E_n c^2 d_{n+1}^2 \frac{\sinh c|d_{n+1}|}{c|d_{n+1}|} \leq E_n d_{n+1}^2 \cdot \frac{c}{M}\sinh cM, \end{aligned}$$

which gives the desired inequality. Therefore, for all positive numbers c and λ

we have

$$P\left(\sum_{k=1}^{n} d_n \geq \lambda\right) \leq P\left\{\exp\left(c \cdot \sum_{k=1}^{n} d_n - \frac{c}{M}\sinh cM \cdot s_n^2(f)\right)\right.$$
$$\left.\geq \exp\left(c\lambda - \frac{c}{M}\sinh cM \cdot K^2\right)\right\}$$
$$\leq \exp\left(-c\lambda + \frac{c}{M}\sinh cM \cdot K^2\right) \cdot EY_1$$
$$\leq \exp\left(-c\lambda + \frac{c}{M}\sinh cM \cdot K^2\right)$$

and if

$$c_0 = \frac{1}{M}\operatorname{arc\,sinh}\frac{M\lambda}{2K^2}$$

we have

$$\frac{\lambda}{2} = \frac{\sinh c_0 \cdot M}{M} \cdot K^2,$$

so that

$$P\left(\sum_{k=1}^{n} d_k \geq \lambda\right) \leq \exp\left(-\frac{c_0\lambda}{2}\right) = \exp\left(\frac{-\lambda}{2M}\operatorname{arc\,sinh}\frac{\lambda M}{2K^2}\right)$$

and finally

$$P\left(\left|\sum_{k=1}^{n} d_k\right| \geq \lambda\right) \leq 2\exp\left(\frac{-\lambda}{2M}\operatorname{arc\,sinh}\frac{\lambda M}{2K^2}\right),$$

which completes the proof. □

The above proposition and a stopping-time argument yield the following good λ inequality.

PROPOSITION 3.2. *Let (d_n) be a martingale difference sequence, such that $|d_n| \leq w_n$, where (w_n) is a predictable sequence of random variables. Then, for all $\delta > 0$, $\beta > 1 + \delta$ and $\lambda > 0$ the following inequality holds*:

$$P(f^* > \beta\lambda,\ s(f) \vee w^* \leq \delta\lambda) \leq 2\varepsilon(\alpha) \cdot P(f^* > \lambda),$$

where $\alpha = (\beta - 1 - \delta)/\delta$ and $\varepsilon(\alpha) = \exp(-(\alpha/2)\ln(1 + \alpha/2))$.

PROOF. Following Burkholder (1973), we write

$$\mu = \inf\{n: |f_n| > \lambda\},$$
$$\nu = \inf\{n: |f_n| > \beta\lambda\},$$
$$\tau = \inf\{n: s_{n+1}(f) > \delta\lambda \text{ or } w_{n+1} > \delta\lambda\}.$$

Since both sequences $(s_n(f))$ and (w_n) are predictable, all of μ, ν and τ are stopping times. Thus $v_k = I(\mu < k \le \nu \wedge \tau)$ is a $\mathscr{F}_{k-1}$-measurable random variable and $(\sum_{k=1}^n v_k d_k)$ is a martingale. Therefore,

$$\begin{aligned} P(f^* > \beta\lambda, s(f) \vee w^* \le \delta\lambda) &= P(\nu < \infty, \tau = \infty) \\ &\le P\Big(\Big|\sum v_k d_k\Big| \ge (\beta - 1 - \delta)\lambda\Big) \\ &= EP\left(\left|\sum_{k=\mu+1}^{\nu\wedge\tau} d_k\right| \ge (\beta - 1 - \delta)\lambda | \mathscr{F}_\mu\right). \end{aligned}$$

But, conditionally on $\mathscr{F}_\mu$, $(\sum_{k=\mu+1}^{\nu\wedge\tau} d_k)$ is a mean-zero martingale, and since

$$\sum_{k=1}^{\nu\wedge\tau} E_{k-1} d_k^2 \le \delta^2\lambda^2$$

and $w_{\nu\wedge\tau}^* \le \delta\lambda$ by the preceding proposition, we get

$$\begin{aligned} &P\left(\left|\sum_{k=\mu+1}^{\nu\wedge\tau} d_k\right| \ge (\beta - 1 - \delta)\lambda | \mathscr{F}_\mu\right) \\ &\quad\le \begin{cases} 2\exp\left\{-\dfrac{(\beta-1-\delta)\lambda}{2\delta\lambda} \operatorname{arc\,sinh} \dfrac{\delta\lambda^2(\beta-1-\delta)}{2\delta^2\lambda^2}\right\} & \text{if } \mu < \infty, \\ 0 & \text{if } \mu = \infty \end{cases} \\ &\quad= 2\exp\left(-\frac{\alpha}{2}\operatorname{arc\,sinh}\frac{\alpha}{2}\right) \cdot I(\mu < \infty). \end{aligned}$$

Integrating both sides and using the inequality $\operatorname{arc\,sinh} x \ge \ln(1 + x)$, $x \ge 0$, we conclude that

$$P\left(\left|\sum_{k=\mu+1}^{\nu\wedge\tau} d_n\right| \ge (\beta - 1 - \delta)\lambda\right) \le 2E\varepsilon\left(\frac{\alpha}{2}\right) I(\mu < \infty) = 2\varepsilon\left(\frac{\alpha}{2}\right) P(f^* > \lambda),$$

which completes the proof. □

The next lemma was suggested by the observation of Hall and Marron (1988).

LEMMA 3.3. *Assume that X and Y are nonnegative random variables such that for all $\delta > 0$, $\beta > 1 + \delta$, $\lambda > 0$ the following is true*:

$$P(X > \beta\lambda, Y \le \delta\lambda) \le 2\varepsilon(\alpha) P(X > \lambda),$$

where ε and α are as in Proposition 3.2. Then, for all $p \ge e^2$,

$$\|X\|_p \le K \cdot \frac{p}{\ln p} \|Y\|_p.$$

PROOF. Since $P(X > \beta\lambda) \le P(Y > \delta\lambda) + P(X > \beta\lambda,\ Y \le \delta\lambda)$, by our assumption we get

$$P(X > \beta\lambda) \le P(Y > \delta\lambda) + 2\varepsilon P(X > \lambda).$$

Multiplying both sides by $p\lambda^{p-1}$ and integrating over $\mathbb{R}_+$ with respect to λ, we obtain

$$E(X/\beta)^p \le E(Y/\delta)^p + 2\varepsilon EX^p$$

or

$$EX^p \le (\beta/\delta)^p \cdot (1 - 2\beta^p\varepsilon)^{-1} \cdot EY^p,$$

whenever $\beta^p\varepsilon < \frac{1}{2}$. Choosing now $\delta \approx \ln p/p$ and $\beta \approx$ const and using the inequality $(1 + x)^p \le \exp(px)$, we obtain the lemma for $p \ge p_0$, where p_0 is large enough [the choice

$$\delta = \frac{1}{3p}\ln\frac{p}{\ln p}, \qquad \beta = 1 + \delta\left(\frac{2p}{\ln p} - 1\right)$$

gives $p_0 = e^2$]. □

4. Rosenthal's inequality. This section contains a proof of the following theorem.

THEOREM 4.1. *There exists an absolute constant K, such that for all martingales (f_n) and all p, $2 \le p < \infty$, the following inequality is true*:

$$\|f^*\|_p \le K \cdot p/\mathrm{Log}\, p\{\|s(f)\|_p + \|d^*\|_p\}.$$

REMARK. Of course the order of growth rate $p/\mathrm{Log}\, p$ is best possible, since it is already best possible in the case of independent random variables [cf. Johnson, Schechtman and Zinn (1985)].

PROOF OF THEOREM 4.1. First of all let us observe that for $2 \le p < e^2$, by the result of Burkholder (1973) mentioned in Section 1, $\|f^*\|_p \le B_p\{\|s(f)\|_p + \|d^*\|_p\}$ for some constant B_p. In that range of p we, of course, have $B_p \le Kp/\mathrm{Log}\, p$, so we can assume without loss of generality, that $p \ge e^2$. It is convenient to split the proof in two steps. In the first step we will prove our theorem for conditionally symmetric sequences, while in the second we explain how to reduce the general case to the case of conditionally symmetric random variables.

STEP 1. Assume that (d_n) is conditionally symmetric. We will use Davis' decomposition of a martingale [cf., e.g., Burkholder (1973) or the original paper of Davis (1970)]: Write $d_n = d'_n + d''_n$, where $d'_n = d_n I(|d_n| \le 2d^*_{n-1})$ and $d''_n = d_n I(|d_n| > 2d^*_{n-1})$. Note that by the conditional symmetry of (d_n), $E_{n-1}d'_n = E_{n-1}d''_n = 0$, so that both sequences (d'_n) and (d''_n) are martingale differences. To estimate $\|(\sum d_k'')^*\|_p$, observe that on the set $\{|d_n| > 2d_{n-1}^*\}$,

$$|d_n| + 2d_{n-1}^* < 2|d_n| \le 2d_n^*.$$

Hence, $|d_n''| \le 2(d_n^* - d_{n-1}^*)$ and $\sum |d_n''| \le 2d^*$. Consequently,

$$\Big\| \sum d_k'' \Big\|_p \le \Big\| \sum |d_k''| \Big\|_p \le 2\|d^*\|_p.$$

For (d_n') we first use Proposition 3.2 with $w_n = 2d_{n-1}^*$, and then Lemma 3.3 to conclude that

$$\begin{aligned}\Big\|\Big(\sum d_k'\Big)^*\Big\|_p &\le Kp/\mathrm{Log}\, p\{\|s(f')\|_p + 2\|d^*\|_p\}\\ &\le Kp/\mathrm{Log}\, p\{\|s(f)\|_p + \|d^*\|_p\}.\end{aligned}$$

Collecting the above estimates, we obtain

$$\|f^*\|_p \le Kp/\mathrm{Log}\, p\{\|s(f)\|_p + \|d^*\|_p\},$$

which completes the proof of this part.

STEP 2. Let (d_n) be an arbitrary martingale difference sequence and denote by (e_n) a sequence which is tangent to (d_n) and satisfies condition (CI). Assume also, that for each $n \ge 1$ the random variables d_n and e_n are $\mathscr{F}_{n-1}$-conditionally independent (see Lemma 2.3 above and the remark following it). Let us write $\bar{d}_n = d_n - e_n$ and denote by $(\bar{f}_n)$ and (g_n) sequences of partial sums of $(\bar{d}_n)$ and (e_n), respectively. Obviously, (g_n) is a martingale and since d_n and e_n are $\mathscr{F}_{n-1}$-conditionally independent, $(\bar{d}_n)$ is conditionally symmetric sequence of martingale differences. Therefore, by the first part of the proof

$$\begin{aligned}\|\bar{f}^*\|_p &\le Kp/\mathrm{Log}\, p\{\|s(\bar{f})\|_p + \|\bar{d}^*\|_p\}\\ &\le Kp/\mathrm{Log}\, p\{\|s(g)\|_p + \|s(f)\|_p + \|d^*\|_p + \|e^*\|_p\}.\end{aligned}$$

Since (d_n) and (e_n) are tangent sequences, we have: $s(f) = s(g)$, and by Lemma 2.5 $\|e^*\|_p \le 2^{1/p}\|d^*\|_p$. Therefore, the right-hand side of the above inequality is dominated by $Kp/\mathrm{Log}\, p\{\|s(f)\|_p + \|d^*\|_p\}$.

Applying Lemma 2.4 to the martingale (g_n), we see that

$$\|g^*\|_p \le Kp/\mathrm{Log}\, p\{\|s(g)\|_p + \|e^*\|_p\},$$

and the same argument as above yields

$$\|g^*\|_p \le Kp/\mathrm{Log}\, p\{\|s(f)\|_p + \|d^*\|_p\}.$$

Finally,

$$\|f^*\|_p \le \|\bar{f}^*\|_p + \|g^*\|_p \le K \cdot p/\mathrm{Log}\, p\{\|s(f)\|_p + \|d^*\|_p\},$$

which completes the proof. □

REMARK. It turns out that the above method can be applied to other quasilinear operators on martingales rather than $s(f) \vee d^*$. The details will appear elsewhere.

5. The case of nonnegative random variables. In this section we will use our result to obtain the optimal constant in similar inequality for nonnegative random variables.

THEOREM 5.1. *For any adapted sequence* (X_k) *of nonnegative random variables the following inequality is true*:

$$\Big\| \sum X_k \Big\|_p \le K \cdot p/\mathrm{Log}\, p \Big\{ \Big\| \sum E_{k-1} X_k \Big\|_p + \|X^*\|_p \Big\}, \qquad p \ge 1,$$

for some absolute constant K. *The growth rate* $p/\mathrm{Log}\, p$ *is best possible.*

Apart from the best constant the above inequality was first proved by Burkholder (1971) [or Burkholder (1973), page 40] and, under the assumption that (X_k) are independent (in which case the first term on the right-hand side is equal to $\|\sum X_k\|_1$) by Rosenthal (1970). Later on, Johnson, Schechtman and Zinn (1985), using an "unbalanced" version of Rosenthal's argument, proved that the optimal choice of a constant in the independent case is $p/\mathrm{Log}\, p$. Unfortunately, this approach does not seem to be useful in the general case, but as we will show below, the result can be easily deduced from Theorem 4.1 and Davis' decomposition.

PROOF OF THEOREM 5.1. Given (X_k) let us write

$$X_k' = X_k I(X_k \le 2X_{k-1}^*)$$

and

$$X_k'' = X_k I(X_k > 2X_{k-1}^*).$$

Then

$$\Big\| \sum X_k \Big\|_p \le \Big\| \sum X_k' \Big\|_p + \Big\| \sum X_k'' \Big\|_p$$

and, by the same computation as in the proof of Theorem 4.1 we have

$$\Big\| \sum X_k'' \Big\|_p \le 2\|X^*\|_p.$$

To estimate $\|\sum X_k'\|_p$, we write

$$\begin{aligned} \Big\| \sum X_k' \Big\|_p &\le \Big\| \sum X_k' - E_{k-1} X_k' \Big\|_p + \Big\| \sum E_{k-1} X_k' \Big\|_p \\ &\le \Big\| \sum X_k' - E_{k-1} X_k' \Big\|_p + \Big\| \sum E_{k-1} X_k \Big\|_p \end{aligned}$$

and apply Theorem 4.1 to the martingale difference sequence $(X_k' - E_{k-1} X_k')$ to get

$$\begin{aligned} \Big\| \sum X_k' - E_{k-1} X_k' \Big\|_p \le Kp/\mathrm{Log}\, p \Big\{ \Big\| \Big(\sum E_{k-1} (X_k' - E_{k-1} X_k')^2 \Big)^{1/2} \Big\|_p \\ + \|\sup_k |X_k' - E_{k-1} X_k'| \|_p \Big\}, \qquad p \ge 2. \end{aligned}$$

Since

$$|X_k' - E_{k-1}X_k'| \le 4X_{k-1}^* \le 4X^*,$$

the second term on the right-hand side above does not exceed $4 \cdot \|X^*\|_p$ and the first is dominated by

$$\begin{aligned}
&\Big\|\Big(\sum 4X_{k-1}^* E_{k-1}|X_k' - E_{k-1}X_k'|\Big)^{1/2}\Big\|_p \\
&\quad \le \Big\|(4X^*)^{1/2}\Big(\sum E_{k-1}(X_k' + E_{k-1}X_k')\Big)^{1/2}\Big\|_p \\
&\quad \le \Big\|(8X^*)^{1/2}\Big(\sum E_{k-1}X_k'\Big)^{1/2}\Big\|_p \\
&\quad \le \Big\|(8X^*)^{1/2}\Big(\sum E_{k-1}X_k\Big)^{1/2}\Big\|_p.
\end{aligned}$$

This last quantity, in view of Schwarz inequality applied to the random variables $(8X^*)^{p/2}$ and $(\sum E_{k-1}X_k)^{p/2}$ is no greater than

$$\Big(\|8X^*\|_p \cdot \Big\|\sum E_{k-1}X_k\Big\|_p\Big)^{1/2} \le 8\|X^*\|_p + \Big\|\sum E_{k-1}X_k\Big\|_p.$$

This, combined with our earlier estimates, completes the proof for $p \ge 2$. For $1 \le p < 2$ the result follows by the same argument as at the very beginning of the proof of Theorem 4.1 above. □

Acknowledgments. I would like to thank Gideon Schechtman and Joel Zinn for several enlightening conversations on the subject of this paper. In particular, it was Joel Zinn who brought the paper by Hall and Marron (1988) to my attention. Also, thanks are due to Don Burkholder and the referee for pointing out the reference Burkholder (1977) (in which "exponential" good λ inequalities appeared for the first time).

REFERENCES

BURKHOLDER, D. L. (1971). Inequalities for operators on martingales. *Proc. Internat. Congr. Math.* (Nice, 1970) **2** 551–557. Gauthier-Villars, Paris.

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

BURKHOLDER, D. L. (1977). Exit times of Brownian motion, harmonic majorization and Hardy spaces. *Adv. in Math.* **26** 182–205.

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

GARSIA, A. M. (1973). *Martingale Inequalities. Seminar Notes on Recent Progress.* Benjamin, Reading, Mass.

HALL, P. and MARRON, J. S. (1988). Choice of kernel order in density estimation. *Ann. Statist.* **16** 161–172.

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

JOHNSON, W. B., SCHECHTMAN, G. and ZINN, J. (1985). Best constants in moment inequalities for linear combination of independent and exchangeable random variables. *Ann. Probab.* **13** 234–253.

REFERENCES

BURKHOLDER, D. L. (1971). Inequalities for operators on martingales. *Proc. Internat. Congr. Math.* (Nice, 1970) **2** 551–557. Gauthier-Villars, Paris.

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

MURAI, T. and UCHIYAMA, A. (1986). Good λ inequalities for the area integral and the nontangential maximal function. *Studia Math.* **83** 251–262.

ROSENTHAL, H. P. (1970). On the subspaces of L^p $(p > 2)$ spanned by sequences of independent random variables. *Israel J. Math.* **8** 273–303.

SHIRYAYEV, A. N. (1984). *Probability*. Springer, Berlin.

TALAGRAND, M. (1989). Isoperimetry and integrability of the sum of independent Banach-space valued random variables. *Ann. Probab.* **17** 1546–1570.

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