AN EXTENSION TO THE TANGENT SEQUENCE MARTINGALE INEQUALITY

STEPHEN MONTGOMERY-SMITH AND SHIH-CHI SHEN

ABSTRACT. For each $1 < p < \infty$, there exists a positive constant c_p, depending only on p, such that the following holds. Let (d_k), (e_k) be real-valued martingale difference sequences. If for for all bounded nonnegative predictable sequences (s_k) and all positive integers k we have

$$E[s_k \vee |e_k|] \le E[s_k \vee |d_k|]$$

then for all positive integers n we have

$$\left\| \sum_{k=1}^{n} e_k \right\|_p \le c_p \left\| \sum_{k=1}^{n} d_k \right\|_p .$$

1. INTRODUCTION

Let $(\Omega, \mathcal{F}, P)$ be a probability space, and let $(\mathcal{F}_k)$ be a filtration on $(\Omega, \mathcal{F}, P)$. (We will suppose that $\mathcal{F}_0 = \{\emptyset, \Omega\}$.) If an adapted sequence (d_k) is a real-valued martingale difference sequence, Burkholder's inequality [2] shows that for any $1 < p < \infty$ that there exists a positive constant c_p depending only on p such that such that for all $\varepsilon_k \in \{1, -1\}$ and all positive integers n

$$\left\| \sum_{k=1}^{n} \varepsilon_k d_k \right\|_p \le c_p \left\| \sum_{k=1}^{n} d_k \right\|_p .$$

More generally, if (v_k) is a predictable sequence bounded in absolute value by 1, then

$$\left\| \sum_{k=1}^{n} v_k d_k \right\|_p \le c_p \left\| \sum_{k=1}^{n} d_k \right\|_p .$$

1991 *Mathematics Subject Classification.* Primary 60G42; Secondary 15A51, 46B70.

Key words and phrases. Martingale inequalities, tangent sequences, decreasing rearrangement, K-functional, doubly stochastic matrices.

Both authors were supported in part by the NSF and the Research Board of the University of Missouri.

A different approach to this inequality was proposed by Kwapień and Woyczinski [8] (see also [9]). Two adapted sequence (f_k) and (g_k) are said to be *tangent* if for each $k \geq 1$, we have that the law of f_k conditionally on $\mathcal{F}_{k-1}$ is the same as the law of g_k conditionally on $\mathcal{F}_{k-1}$, that is,

$$P(f_k < \lambda | \mathcal{F}_{k-1}) = P(g_k < \lambda | \mathcal{F}_{k-1})$$

for all real numbers λ. Answering a conjecture of Kwapień and Woyczinski [8], it was proved by Hitczenko [4] (see also [14]) that for $1 < p < \infty$ that there exists a positive constant c_p, depending only on p, such that if (d_k) and (e_k) are martingale difference sequences and (d_k), (e_k) are tangent, then for all positive integers n

$$\left\| \sum_{k=1}^{n} e_k \right\|_p \leq c_p \left\| \sum_{k=1}^{n} d_k \right\|_p . \tag{1}$$

The purpose of this paper is to provide a common generalization to these two results.

Theorem 1. *For each $1 < p < \infty$, there exists a positive constant c_p, depending only on p, such that the following holds. Let (d_k), (e_k) be real-valued martingale difference sequences. If for for all bounded nonnegative predictable sequence (s_k) and all positive integers k we have*

$$E[s_k \vee |e_k|] \leq E[s_k \vee |d_k|] \tag{2}$$

then for all positive integers n we have equation (1).

This is essentially equivalent to another result, which concerns martingales in a specific situation. We will consider the probability space $[0,1]^{\mathbf{N}}$ equipped with the product Lebesgue measure $\mathcal{L}$, and consider the filtration $(\mathcal{L}_k)$, where $\mathcal{L}_k$ is the minimal σ-field for which the first k coordinate functions of $[0,1]^{\mathbf{N}}$ are measurable. Then two sequences (d_k) and (e_k) are tangent if

$$e_k(x_1, ..., x_k) = d_k(x_1, ..., x_{k-1}, \phi_k(x_1, ..., x_k))$$

where $(\phi_k : [0,1]^k \to [0,1])$ is a sequence of measurable functions such that $\phi_k(x_1, ..., x_{k-1}, \cdot)$ is a measure preserving map for almost all x_1, ..., x_{k-1}.

We will consider a more general situation. Suppose we have a sequence of linear operators $(T_k(x_1, ..., x_{k-1}))$, depending measurably upon $(x_k) \in [0,1]^{\mathbf{N}}$, that are bounded operators on both $L_1[0,1]$ and $L_\infty[0,1]$ with norm 1. Then consider the condition

$$e_k(x_1, ..., x_{k-1}, \cdot) = [T_k(x_1, ..., x_{k-1})] d_k(x_1, ..., x_{k-1}, \cdot). \tag{3}$$

Theorem 2. *For each $1 < p < \infty$, there exists a positive constant c_p, depending only on p, such that the following holds. If (d_k), (e_k) and (T_k) are as above satisfying (3), then for all positive integers n we have equation (1).*

We will also need the following intermediate result. For any random variable f, let $f^{\#}$ be the decreasing rearrangement of $|f|$, that is,

$$f^{\#}(t) = \sup\{s \in \mathbf{R} : P(|f| < s) < t\}.$$

Theorem 3. *For each $1 < p < \infty$, there exists a positive constant c_p, depending only on p, such that the following holds. Let (d_k), (e_k) be martingale difference sequences on $[0,1]^{\mathbf{N}}$ with respect to $(\mathcal{L}_k)$. Suppose that for each positive integer k*

$$\int_0^t (e_k(x_1, ..., x_{k-1}, \cdot))^{\#}(s)ds \le \int_0^t (d_k(x_1, ..., x_{k-1}, \cdot))^{\#}(s)ds$$

for all $t \in [0,1]$ and almost all x_1,...,x_{k-1}. Then for all positive integers n we have equation (1).

2. The Discrete Type Case

In this section we will prove Theorems 2 and 3 in a special discrete situation, which we now describe. For any positive integer N, let Σ be the σ-field generated by the partition $\{[\frac{i-1}{N}, \frac{i}{N}); i = 1, 2, ..., N\}$. Define a filtration on $[0,1]^{\mathbf{N}}$ by $(\mathcal{F}_k)$ by $\mathcal{F}_k = \mathcal{L}_{k-1} \times \Sigma$. Suppose (d_k), (e_k) are $(\mathcal{F}_k)$-adapted. Then for each k and for each x_1, ..., x_{k-1}, we see that $d_k(x_1, ..., x_{k-1}, \cdot)$ and $e_k(x_1, ..., x_{k-1}, \cdot)$ are Σ-measurable simple functions on $[0, 1)$. Therefore d_k and e_k can be written as N-dimensional vectors and $T_k(x_1, ..., x_{k-1})$ can be represented by a $N \times N$ matrix, that is,

$$\begin{bmatrix} e_k(1) \\ e_k(2) \\ \vdots \\ e_k(N) \end{bmatrix} = \begin{bmatrix} a_k(1,1), & ... & , a_k(1,N) \\ a_k(2,1), & ... & , a_k(2,N) \\ \vdots & & \vdots \\ a_k(N,1), & ... & , a_k(N,N) \end{bmatrix} \begin{bmatrix} d_k(1) \\ d_k(2) \\ \vdots \\ d_k(N) \end{bmatrix}$$

where

$$d_k(i) = d_k(x_1, ..., x_{k-1}, i) = d_k(x_1, ..., x_k) \text{ if } x_k \in [\tfrac{i-1}{N}, \tfrac{i}{N})$$
$$e_k(i) = e_k(x_1, ..., x_{k-1}, i) = e_k(x_1, ..., x_k) \text{ if } x_k \in [\tfrac{i-1}{N}, \tfrac{i}{N})$$
$$T_k = T_k(x_1, ..., x_{k-1}) = [(a_k(x_1, ..., x_{k-1}))(i,j)]_{N \times N} = [a_k(i,j)]_{N \times N}$$

The condition of being martingale difference sequences implies that

$$\sum_{i=1}^{N} d_k(i) = \sum_{i=1}^{N} e_k(i) = 0$$

Theorem 4. *Theorem 2 holds in the case that* (d_k) *and* (e_k) *are adapted to the filtration* $(\mathcal{F}_k)$ *described above.*

In this discrete case, the boundedness of $\|T_k\|_{L_1}$ and $\|T_k\|_{L_\infty}$ by 1 is equivalent to the condition that $\sum_{j=1}^N |a_k(i,j)| \le 1$ for all i and $\sum_{i=1}^N |a_k(i,j)| \le 1$ for all j. We claim that without loss of generality, we can assume that every row sum and column sum of T_k is 0, that is,

$$\sum_{j=1}^N a_k(i,j) = \sum_{i=1}^N a_k(i,j) = 0$$

for all i and j. Suppose the i^{th} row sum $\sum_{j=1}^N a_k(i,j) = R_k(i)$. Let T_k' be the liner operator defined by

$$T_k' = \left[a_k(i,j) - \frac{R_k(i)}{N} \right]_{N\times N}$$

It is clear that every row sum of T_k' is 0 and

$$\begin{aligned} (T_k' d_k)(i) &= \sum_{j=1}^N \left(a_k(i,j) - \frac{R_k(i)}{N} \right) d_k(j) \\ &= \sum_{j=1}^N a_k(i,j) d_k(j) - \frac{R_k(i)}{N} \sum_{j=1}^N d_k(j) \\ &= e_k(i) \end{aligned}$$

Now we can assume that every row sum of T_k is 0. Similarly suppose the j^{th} column sum $\sum_{i=1}^N a_k(i,j) = C_k(j)$. Let T_k'' be the linear operator defined by

$$T_k'' = \left[a_k(i,j) - \frac{C_k(j)}{N} \right]_{N\times N}$$

Again it is clear that every row sum and column sum of T_k'' is 0 and

$$\begin{aligned} (T_k'' d_k)(i) &= \sum_{j=1}^N \left(a_k(i,j) - \frac{C_k(j)}{N} \right) d_k(j) \\ &= \sum_{j=1}^N a_k(i,j) d_k(j) - \frac{1}{N} \sum_{j=1}^N C_k(j) d_k(j) \\ &= e_k(i) \end{aligned}$$

since

$$\sum_{i=1}^N e_k(i) = \sum_{j=1}^N C_k(j) d_k(j) = 0$$

After adjusting T_k, it is easy to check that the norms of T_k may be enlarged up to 4. Of course, we can pick up $T_k/4$ instead and absorb the 4 into the constant c_p.

A nonnegative real matrix is said to be *doubly stochastic* if each of its row and column sum is 1. A sub-doubly stochastic matrix means that each of its row and column sum is less than or equal to 1. Therefore we can change the assumption in Theorem 4 to be that: "for almost all x_1, ..., x_{k-1}, every row sum and column sum of the matrix from T_k is 0, and the matrix from $|T_k|$ is sub-doubly stochastic for each positive integer k"

One of the fundamental results in the theory of doubly stochastic matrices was introduced by Birkhoff (see for example [11, p. 117]).

Theorem 5. *If M is a doubly stochastic matrix, then*

$$M = \sum_{i=1}^{S} \theta_i P_i$$

where P_i are permutation matrices, and the θ_i are nonnegative numbers satisfying $\sum_{i=1}^{S} \theta_i = 1$.

Lemma 1. *If M is a $n \times n$ sub-doubly stochastic matrix, then there exists a $2n \times 2n$ doubly stochastic matrix such that its upper left $n \times n$ sub-matrix is M.*

Proof. Suppose that $R(i)$ is the i^{th} row sum of M, $C(j)$ is the j^{th} column sum and S is the sum of all entries. Let

$$A = \begin{bmatrix} \frac{1-R(1)}{n}, & \cdots & , \frac{1-R(1)}{n} \\ \vdots & & \vdots \\ \frac{1-R(n)}{n}, & \cdots & , \frac{1-R(n)}{n} \end{bmatrix}_{n \times n}$$

$$B = \begin{bmatrix} \frac{1-C(1)}{n}, & \cdots & , \frac{1-C(n)}{n} \\ \vdots & & \vdots \\ \frac{1-C(1)}{n}, & \cdots & , \frac{1-C(n)}{n} \end{bmatrix}_{n \times n}$$

$$C = \mathrm{Diag} \left[\begin{matrix} \frac{S}{n}, & \cdots & , \frac{S}{n} \end{matrix} \right]_{n \times n}$$

Then define

$$M' = \begin{bmatrix} M & A \\ B & C \end{bmatrix}_{2n \times 2n}$$

It is easy to check that M' is a doubly stochastic matrix. $\square$

Lemma 2. *If M is a sub-doubly stochastic matrix, then there exists a sub-doubly stochastic matrix N such that $M + N$ is doubly stochastic.*

Proof. Let M' be the $2n \times 2n$ doubly stochastic matrix such that its upper left $n \times n$ sub-matrix is M. By Theorem 5,

$$M' = \sum_{i=1}^{S} \theta_i P_i'$$

where P_i' are $2n \times 2n$ permutation matrices and $\sum_{i=1}^{S} \theta_i = 1$. Suppose that P_i is the upper left $n \times n$ sub-permutation matrix of P_i', then

$$M = \sum_{i=1}^{S} \theta_i P_i$$

Let Q_i be a $n \times n$ sub-permutation matrix such that $P_i + Q_i$ is a permutation matrix, say R_i. Define

$$N = \sum_{i=1}^{S} \theta_i Q_i$$

thus

$$M + N = \sum_{i=1}^{S} \theta_i R_i$$

which is a doubly stochastic matrix. □

Theorem 6. *Let M be an $n \times n$ matrix. If every row sum and column sum of M is 0 and $|M|$ is sub-doubly stochastic, then*

$$M = \sum_{i=1}^{S} \theta_i P_i$$

where P_i are permutation matrices, $\sum_{i=1}^{S} \theta_i = 0$ and $\sum_{i=1}^{S} |\theta_i| = 1$

Proof. Let

$$A = \frac{|M| + M}{2}$$
$$B = \frac{|M| - M}{2}$$

so A and B are nonnegative, and $2A$ and $2B$ are sub-doubly stochastic. By Lemma 2, there exists a sub-doubly stochastic matrix C such that $2(A+C)$ is a doubly stochastic. But A and B have the same row sums and column sums, and hence $2(B+C)$ is also a doubly stochastic. By applying Theorem 5, we have

$$2(A + C) = \sum_{i=1}^{m} \lambda_i Q_i$$

$$2(B+C)=\sum_{i=1}^{m'}\lambda_i' Q_i'$$

where Q_i, Q_i' are permutation matrices, and the λ_i, λ_i' are nonnegative numbers satisfying $\sum_{i=1}^m \lambda_i = \sum_{i=1}^{m'} \lambda_i' = 1$. Then the result follows because

$$M=(A+C)-(B+C)=\sum_{i=1}^{m}\frac{\lambda_i}{2}Q_i-\sum_{i=1}^{m'}\frac{\lambda_i'}{2}Q_i'$$

$\square$

Proof of Theorem 4. From Theorem 6, we know that for each $k \geq 1$ and almost all x_1, ..., x_{k-1}

$$T_k(x_1,...,x_{k-1})=\sum_{i_k=1}^{S_k}\theta_{k,i_k}(x_1,...,x_{k-1})\cdot P_{k,i_k}(x_1,...,x_{k-1})$$

where P_{k,i_k} are permutation matrices, $\sum_{i=1}^{S_k}\theta_{k,i_k}=0$, and $\sum_{i=1}^{S_k}|\theta_{k,i_k}|=1$. Let

$$h_{k,i_k}(x_1,...,x_{k-1},\cdot)=[P_{k,i_k}(x_1,...,x_{k-1})]d_k(x_1,...,x_{k-1},\cdot). \tag{4}$$

Then

$$\begin{aligned} e_k &= \left[\sum_{i_k=1}^{S_k}\theta_{k,i_k}P_{k,i_k}\right]d_k \\ &= \sum_{i_k=1}^{S_k}|\theta_{k,i_k}|\varepsilon_{k,i_k}h_{k,i_k} \end{aligned}$$

where $\varepsilon_{k,i_k}=\operatorname{sgn}(\theta_{k,i_k})$.

Now we need to consider the probability space $\Omega_1\times\Omega_2$, where $\Omega_1=\Omega_2=[0,1]^{\mathbf{N}}$. We consider all of the previous random variables considered as random variables on this new probability space, depending only upon the first coordinate ω_1. We define a filtration $(\mathcal{G}_k)$ where $\mathcal{G}_k=\mathcal{F}_k\otimes\mathcal{L}_{k+1}$.

We define a predictable sequence of random variables (I_k) so that for each $\omega_1\in\Omega_1$, the random variable $I_k(\omega_1,\cdot)$ takes the value i with probability $|\theta_{k,i}(\omega_1)|$. Then we see that

$$e_k=E(\epsilon_{k,I_k}h_{k,I_k}|\mathcal{L}\otimes\{\emptyset,\Omega_2\}).$$

Hence, since conditional expectation is a contraction on L_p

$$\left\|\sum_{k=1}^{n}e_k\right\|_p\leq\left\|\sum_{k=1}^{n}\epsilon_{k,I_k}h_{k,I_k}\right\|_p.$$

Now we see that (ϵ_{k,I_k}) is a predictable sequence bounded by 1. Hence by Burkholder's inequality, we see that

$$\left\| \sum_{k=1}^{n} \epsilon_{k,I_k} h_{k,I_k} \right\|_p \leq c_p \left\| \sum_{k=1}^{n} h_{k,I_k} \right\|_p .$$

Next, observing (4), since P_{k,i_k} are permutation matrices, for each $k \geq 1$, $i_k = 1, 2, ..., S_k$, h_{k,i_k} is just an x_k-rearrangement of d_k. that is

$$h_{k,i_k}(x_1, ..., x_{k-1}, j) = d_k(x_1, ..., x_{k-1}, \pi_{k,i_k}(j))$$

for some permutation π_{k,i_k}. Thus for any sequence (i_k) we have that (h_{k,i_k}) and (d_k) are tangent sequences. But then we see that (h_{k,I_k}) and (d_k) are tangent sequences. Hence there exists a positive constant c_p such that

$$\left\| \sum_{k=1}^{n} h_{k,I_k} \right\|_p \leq c_p \left\| \sum_{k=1}^{n} d_k \right\|_p .$$

The result follows. □

Theorem 7. *Theorem 3 holds in the case that (d_k) and (e_k) are adapted to the filtration $(\mathcal{F}_k)$ described above.*

This will follow immediately from the following well-known result [10, p. 124].

Theorem 8. *$f = (f_1, f_2, ..., f_N)$, $g = (g_1, g_2, ..., g_N)$ are N-dimensional real-valued vectors. $f^\# = (f_1^\#, f_2^\#, ..., f_N^\#)$ is the decreasing rearrangement of $|f| = (|f_1|, |f_2|, ..., |f_N|)$. Then*

$$\sum_{k=1}^{n} g_k^\# \leq \sum_{k=1}^{n} f_k^\#$$

for all $n = 1, 2, ..., N$ if and only if there exists a matrix $T = [a_{ij}]_{N \times N}$ such that $Tf = g$, $\sum_{i=1}^{N} |a_{ij}| \leq 1$ and $\sum_{j=1}^{N} |a_{ij}| \leq 1$.

3. The General Case

The following theorem was proved by Crowe, Zweibel and Rosenbloom [3].

Theorem 9. *Suppose f, g are random variables on $[0, 1]$, then for $1 \leq p \leq \infty$,*

$$\left\| f^\# - g^\# \right\|_p \leq \| f - g \|_p$$

Proof of Theorem 3. We will prove this theorem by using the discrete case. For each $k \geq 1$, we approximate d_k and e_k by functions $d'_k(x_k)$ and $e'_k(x_k)$ such that (d'_k) and (e'_k) are adapted to $(\mathcal{L}_{k-1} \times \Sigma)$, keep the martingale property, and

$$\int_0^t (e'_k(x_1, ..., x_{k-1}, \cdot))^{\#}(s)ds \leq \int_0^t (d'_k(x_1, ..., x_{k-1}, \cdot))^{\#}(s)ds$$

for all $t \in [0, 1]$. Then we apply Theorem 7.

We set $\epsilon \in (0, 1/3)$ to be an arbitrarily small number. First we know that for the quantities to make sense that d_k and e_k are in L_p. Hence there exist simple functions

$$d''_k = \sum_{i=1}^{S} \alpha_{k,i} \chi_{A_{k,i}}$$

$$e''_k = \sum_{i=1}^{S} \beta_{k,i} \chi_{B_{k,i}}$$

such that $\|d_k - d''_k\|_p < \epsilon \|d_k\|_1$ and $\|e_k - e''_k\|_p < \epsilon \|e_k\|_1$. We may suppose without loss of generality that the sets $A_{k,i}$ and $B_{k,i}$ are cylinder sets in $\mathcal{L}_k$ defined by rational numbers, that is, sets of the form $\{(x_i) : x_1 \in (r_1, s_1), \ldots, x_k \in (r_k, s_k)\}$, where the r_i and s_i are rational numbers. Furthermore, we will suppose that $A_{k,i_1} \cap A_{k,i_2} = \emptyset$ and $B_{k,i_1} \cap B_{k,i_2} = \emptyset$ for $i_1 \neq i_2$.

Let N' be the least common denominator of all these rational numbers. For each x_1,...,x_{k-1}, since $d_k^{\#}$ is Reimann integrable as a function of x_k, there is a number $N_k = N_k(x_1, \ldots, x_{k-1})$ that is a multiple of N' and such that if $N \geq N_k$, and

$$\alpha''_{k,i} = N \cdot \int_{\frac{i-1}{N}}^{\frac{i}{N}} d_k^{\#} \tag{5}$$

then

$$\left\| \sum_{i=1}^{N} \alpha''_{k,i} \chi_i - d_k^{\#} \right\|_p < \epsilon \|d_k\|_1 \tag{6}$$

and also the analogous statement holds for e_k. Note that in this case that

$$\int_0^t \sum_{i=1}^{N} \alpha''_{k,i} \chi_i = \int_0^t d_k^{\#}$$

if $t = \frac{i}{N}$ for some $i = 0, 1, 2, ..., N$. Since N_k is measurable with respect to a finitely generated measure space, it follows that there exists a number N that is a multiple of N' that is larger than each N_k.

Now let us fix $x_1,...,x_{k-1}$, and regard the functions as functions as functions of only one variable x_k on $[0,1]$. Similarly, define $A'_{k,i}$ and $B'_{k,i}$ as the set of $x_k \in [0,1]$ for which $(x_1,\ldots,x_k) \in A_{k,i}$ or $B_{k,i}$ respectively. Hence $[\frac{i-1}{N},\frac{i}{N})$ is either contained in some $A'_{k,j}$ or disjoint to all $A'_{k,j}$. Let $\alpha'_{k,i} = \alpha_{k,j}$ if $[\frac{i-1}{N},\frac{i}{N}) \subset A'_{k,j}$ for some j, and $\alpha'_{k,i} = 0$ otherwise. Let $\chi_i = \chi_{[\frac{i-1}{N},\frac{i}{N})}$. Thus

$$\left\| \sum_{i=1}^{N} \alpha'_{k,i}\chi_i - d_k \right\|_p \le \epsilon \|d_k\|_1 \tag{7}$$

Now

$$\left(\sum_{i=1}^{N} \alpha'_{k,i}\chi_i \right)^{\#} = \sum_{i=1}^{N} \varepsilon_{\sigma(i)}\alpha'_{k,\sigma(i)}\chi_i$$

for some permutation σ, where $\varepsilon_j = \operatorname{sgn}(\alpha'_{k,j})$. By Theorem 9,

$$\left\| \sum_{i=1}^{N} \varepsilon_{\sigma(i)}\alpha'_{k,\sigma(i)}\chi_i - d_k^{\#} \right\|_p < \epsilon \|d_k\|_1 \tag{8}$$

Define $\alpha''_{k,i}$ using equation (5), then by (8) and (6),

$$\left\| \sum_{i=1}^{N} \alpha''_{k,i}\chi_i - \sum_{i=1}^{N} \varepsilon_{\sigma(i)}\alpha'_{k,\sigma(i)}\chi_i \right\|_p < 2\epsilon \|d_k\|_1$$

By doing the reverse process of taking decreasing rearrangement of $|\sum_{i=1}^{N} \alpha'_{k,i}\chi_i|$, and setting

$$\hat{\alpha}_{k,i} = \varepsilon_{\sigma^{-1}(i)}\alpha''_{k,\sigma^{-1}(i)}$$

we have

$$\left\| \sum_{i=1}^{N} \hat{\alpha}_{k,i}\chi_i - \sum_{i=1}^{N} \alpha'_{k,i}\chi_i \right\|_p < 2\epsilon \|d_k\|_1 \tag{9}$$

From (7) and (9),

$$\left\| \sum_{i=1}^{N} \hat{\alpha}_{k,i}\chi_i - d_k \right\|_p < 3\epsilon \|d_k\|_1$$

For $t = \frac{j}{N}$, $j = 0,1,2,...,N$, it is clear that

$$\int_0^t \left(\sum_{j=1}^{N} \hat{\alpha}_{k,j}\chi_j \right)^{\#} = \int_0^t \sum_{j=1}^{N} \alpha''_{k,j}\chi_j = \int_0^t d_k^{\#}.$$

Furthermore, if we set

$$\zeta = E\left[\sum_{j=1}^{N} \hat{\alpha}_{k,j}\chi_j\right]$$

then $|\zeta| \leq 3\epsilon\|d_k\|_1$. Thus we see that for $t = \frac{j}{N}$, $j = 0, 1, 2, ..., N$ that

$$\begin{aligned}
(10) \qquad & \int_0^t \left(\sum_{j=1}^{N} \left(\hat{\alpha}_{k,j} - \zeta\right)\chi_j\right)^{\#} \\
\leq \quad & \int_0^t \left(\sum_{j=1}^{N} \left(|\hat{\alpha}_{k,j}| + |\zeta|\right)\chi_j\right)^{\#} \\
\leq \quad & \int_0^t \left(\sum_{j=1}^{N} \hat{\alpha}_{k,j}\chi_j\right)^{\#} + 3\epsilon\|d_k\|_1 \cdot t \\
\leq \quad & \int_0^t d_k^{\#} + 3\epsilon\|d_k\|_1 \cdot t \\
\leq \quad & (1 + 3\epsilon)\int_0^t d_k^{\#}
\end{aligned}$$

and similarly

$$\begin{aligned}
(11) \qquad \int_0^t \left(\sum_{j=1}^{N} \left(\hat{\alpha}_{k,j} - \zeta\right)\chi_j\right)^{\#} \quad & \geq \quad \int_0^t d_k^{\#} - 3\delta\|d_k\|_1 \cdot t \\
& \geq \quad (1 - 3\epsilon)\int_0^t d_k^{\#}
\end{aligned}$$

We can also perform this same construction for e_k, the analogues of $\hat{\alpha}_{k,i}$ and ζ being $\hat{\beta}_{k,i}$ and η. Thus, we are ready to define d'_k and e'_k. Let

$$d'_k = (1 + 3\epsilon)\sum_{j=1}^{N}(\hat{\alpha}_{k,i} - \zeta)\chi_i$$

$$e'_k = (1 - 3\epsilon)\sum_{j=1}^{N}(\hat{\beta}_{k,i} - \eta)\chi_i$$

It is clear that $E[d'_k] = E[e'_k] = 0$. Combining (10) and (11), we have for $t = \frac{j}{N}$, $j = 0, 1, 2, ..., N$

$$\int_0^t (e'_k)^{\#} \leq \int_0^t (d'_k)^{\#}.$$

But then by linear interpolation, this follows for all $t \in [0,1]$. By Theorem 7, there exist a positive constant c_p such that

$$\left\| \sum_{k=1}^{n} e'_k \right\|_p \le c_p \left\| \sum_{k=1}^{n} d'_k \right\|_p$$

Now an easy argument shows that $\|d_k - d'_k\|_p \to 0$ and $\|e_k - e'_k\|_p \to 0$ as $\epsilon \to 0$. The result follows. □

Proof of Theorem 2. If f is a random variable on $(\Omega, \mathcal{F}, P)$, $1 \le p < \infty$, $0 \le t \le 1$, we define the K-functional by

$$K(t, f; L_p, L_\infty) = \inf_{f_0 + f_1 = f} \{\|f_0\|_p + t\|f_1\|_\infty\}.$$

J. Peetre [13] has shown that

$$K(t, f; L_1, L_\infty) = \int_0^t f^{\#}(s) ds.$$

Hence it follows that if T is an operator on both $L_1([0,1])$ and $L_\infty([0,1])$ with norm bounded by 1, then for $t \ge 0$

$$\int_0^t (Tf)^{\#}(s) ds \le \int_0^t f^{\#}(s) ds.$$

Thus the result follows from Theorem 3. □

Lemma 3. *Let f and g are real-valued random variables on $(\Omega, \mathcal{F}, P)$. Then*

$$E\left[M \vee |g|\right] \le E\left[M \vee |f|\right] \tag{12}$$

for all nonnegative number M if and only if

$$\int_0^t g^{\#}(s) ds \le \int_0^t f^{\#}(s) ds$$

for all $t \in [0,1]$.

Proof. Equation (12) is equivalent to $E\left[M \vee g^{\#}\right] \le E\left[M \vee f^{\#}\right]$. For the "if" part, let

$$\alpha = \sup\left\{t : f^{\#}(t) \ge M\right\}$$

$$\beta = \sup\left\{t : g^{\#}(t) \ge M\right\}.$$

Then

$$\begin{aligned} E\left[M \vee f^{\#}\right] &= \int_0^\alpha f^{\#} + (1-\alpha)M \\ &= \int_0^\beta f^{\#} + (1-\beta)M + \int_\beta^\alpha (f^{\#} - M) \\ &\geq \int_0^\beta g^{\#} + (1-\beta)M + \int_\beta^\alpha (f^{\#} - M) \\ &= E\left[M \vee g^{\#}\right] + \int_\beta^\alpha (f^{\#} - M). \end{aligned}$$

If $\alpha \leq \beta$, then for all $x \in (\alpha, \beta)$ we have $f^{\#}(x) \leq M$, and if $\beta \leq \alpha$, then for all $x \in (\beta, \alpha)$ we have $f^{\#}(x) \geq M$. Either way, we see that $\int_\beta^\alpha (f^{\#} - M) \geq 0$, and the result follows.

To show the "only if", for any $\alpha \in [0,1]$, let

$$M = f^{\#}(\alpha)$$
$$\beta = \inf\left\{t : g^{\#}(t) \geq M\right\}.$$

Then

$$\begin{aligned} \int_0^\alpha g^{\#} &= \int_0^\beta g^{\#} + \int_\beta^\alpha (g^{\#} - M) + M(1-\beta) + M(\alpha - 1) \\ &= E\left[M \vee g^{\#}\right] + M(\alpha - 1) + \int_\beta^\alpha (g^{\#} - M) \\ &\leq E\left[M \vee f^{\#}\right] + M(\alpha - 1) + \int_\beta^\alpha (g^{\#} - M) \\ &= \int_0^\alpha f^{\#} + \int_\beta^\alpha (g^{\#} - M). \end{aligned}$$

Arguing as above, we see that $\int_\beta^\alpha (g^{\#} - M) \leq 0$, and again the result follows. $\square$

Given a random variable f and a sigma field $\mathcal{G}$, we will say that f is nowhere constant with respect to $\mathcal{G}$ if $P(f = g) = 0$ for every $\mathcal{G}$ measurable function g. The following theorem [12] shows a concrete representation of a sequence of random variables.

Theorem 10. *Let (f_n) be a sequence of random variables takeing values in a separable sigma filed $(S, \mathcal{S})$. Then there exists a sequence of measurable functions $(g_n : [0,1]^n \to S)$ that has the same law as (f_n). If further we have that f_{n+1} is nowhere constant with respect to $\sigma(f_1, ..., f_n)$ for all $n \geq 0$, then we may suppose that $\sigma(g_1, ..., g_n) = \mathcal{L}_n$ for all $n \geq 0$.*

Proof of Theorem 1. First we claim that (2) is equivalent to

$$E[(M \vee |e_k|)|\mathcal{F}_{k-1}] \le E[(M \vee |d_k|)|\mathcal{F}_{k-1}] \tag{13}$$

This is because for any $A_k \in \mathcal{F}_{k-1}$, $\lambda \ge 0$, $(\lambda\chi_{A_k} \vee M)$ is predictable, and hence

$$E[(\lambda\chi_{A_k} \vee M) \vee |e_k| - \lambda\chi_{A_k}] \le E[(\lambda\chi_{A_k} \vee M) \vee |d_k| - \lambda\chi_{A_k}]$$

When λ intends to infinity, we obtain

$$E[(M \vee |e_k|)\chi_{A_k^c}] \le E[(M \vee |d_k|)\chi_{A_k^c}]$$

which is equivalent to (13).

Consider the map $D_k = (d_k, e_k, f_k) : \Omega \times [0,1]^{\mathbf{N}} \to \mathbf{R}^3$ by $(\omega, (x_k)) \mapsto (d_k(\omega), e_k(\omega), x_k)$. It is clear that D_k is nowhere constant with respect to $\sigma(D_1, D_2, ..., D_{k-1})$. Apply the previous theorem to get $\widetilde{D}_k = (\widetilde{d}_k, \widetilde{e}_k, \widetilde{f}_k) : [0,1]^k \to \mathbf{R}^3$ such that $(\widetilde{D}_k)$ has the same law as (D_k) and $\sigma(\widetilde{D}_1, \widetilde{D}_2, ..., \widetilde{D}_k) = \mathcal{L}_k$.

Next, we show that for almost every $x_1, \ldots, x_{k-1}$ and $M \ge 0$ that

$$\int_0^1 M \vee |\widetilde{e}_k(x_1, \ldots, x_k)| \, dx_k \le \int_0^1 M \vee |\widetilde{d}_k(x_1, \ldots, x_k)| \, dx_k$$

which will follow from showing that for any bounded non-negative measurable function $\phi_k : [0,1]^{k-1} \to [0, \infty)$ that

$$E[\phi_k \vee |\widetilde{e}_k|] \le E[\phi_k \vee |\widetilde{d}_k|].$$

But then there exists a bounded Borel measurable function $\theta_k : \mathbf{R}^{3(k-1)} \to [0, \infty)$ such that $\phi = \theta(\widetilde{D}_1, \widetilde{D}_2, ..., \widetilde{D}_{k-1})$ almost everywhere in $[0,1]^{k-1}$. Thus

$$\begin{aligned}
\int_{[0,1]^k} \phi_k \vee |\widetilde{e}_k| &= \int_{[0,1]^k} \theta(\widetilde{D}_1, ..., \widetilde{D}_{k-1}) \vee |\widetilde{e}_k| \\
&= E[\theta(D_1, ..., D_{k-1}) \vee |e_k|] \\
&\le E[\theta(D_1, ..., D_{k-1}) \vee |d_k|] \\
&= \int_{[0,1]^k} \theta(\widetilde{D}_1, ..., \widetilde{D}_{k-1}) \vee |\widetilde{d}_k| \\
&= \int_{[0,1]^k} \phi_k \vee |\widetilde{d}_k|
\end{aligned}$$

Also to show that $E[\widetilde{d}_k|\mathcal{L}_{k-1}] = E[\widetilde{d}_k|\mathcal{L}_{k-1}] = 0$, it is sufficient to show that for any bounded measurable function $\phi_k : [0,1]^{k-1} \to \mathbf{R}$ that $E[\phi_k\widetilde{d}_k] = E[\phi_k\widetilde{d}_k] = 0$. Thus follows by a very similar argument to that above.

The result then follows from Lemma 3 and Theorem 3. $\square$

Acknowledgments. We would like to mention the help of Jim Reeds and David Boyd in obtaining the argument about doubly stochastic matrices and other useful remarks.

References

1. D. L. Burkholder, *Distribution function inequalities for martingales.* Ann. Probability **1** (1973), 19–42.
2. D. L. Burkholder, *A geomtrical characterization of banach spaces in which marti difference sequences are unconditional.* Ann. Probability **9** (1981), 997–1011.
3. L. A. Crowe, J. A. Zweibel and P. C. Rosenbloom, *Rearrangements of functions*, Journal of functional analysis **66** (1986), 43291–438.
4. P. Hitczenko, *Comparison of moments for tangent sequences of random variables.* Probab. Theory Related Fields **78** (1988), 223–230.
5. P. Hitczenko, *On a domination of sums of random variables by sums of conditionally independent ones.* Ann. Probability **22** (1994), 453–468.
6. P. Hitczenko and S. J. Montgomery-Smith, *Tangent sequences in Orlicz and rearrangement invariant spaces.* Math. Proc. Camb. Phil. Soc. **119** (1996), 91–101.
7. P. Kree, *Interpolation d'espaces qui ne sont mi normes, ni complets.* Applications, Seminaire Lions-Schwartz, semestre 1964-1965, Secretariat Mathematatique, 11 rue Pierre Curie, Paris 5e.
8. S. Kwapień and W. A. Woyczynsi, *Semimartingale integrals via decoupling inequalities and tangent processes.* Case Western Reserve University, preprint, (1986).
9. S. Kwapień and W. A. Woyczynsi, *Random series and stochastic integrals. Single and multiple.* Birkhauser, Boston. (1996).
10. J. Lindenstrauss ans L. Tzafriri, *Classical Banach spaces II.* Springer-Verlag. (1979).
11. H. Minc, *Nonnegative matrices.* Wiley Interscience. (1988).
12. S. J. Montgomery-Smith, *Concrete representation of martingales.* Electronic J. Probab. **3**, (1998), paper 15
13. J. Peetre, *Espaces d'interpolation, generalisations, applications.* Rend. Sem. Mat. Fis. Milano **34** (1964), 83-92.
14. J. Zinn, *Comparison of martingale difference sequences.* Probability in Banach spaces, V (Medford, Mass., 1984), 453–457, Lecture Notes in Math., 1153, Springer, Berlin, 1985.

Department of Mathematics, University of Missouri, Columbia, Missouri 65211, USA.
E-mail address: stephen@math.missouri.edu
URL: http://www.math.missouri.edu/~stephen

Department of Mathematics, University of Missouri, Columbia, Missouri 65211, USA.
E-mail address: mathgr75@math.missouri.edu