AN EXTENSION TO THE STRONG DOMINATION 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 [3] shows that for any $1 < p < \infty$, if (v_k) is a predictable sequence bounded in absolute value by 1, then there exists a positive constant c_p, depending only on p, such that such that for all positive integers n

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

Later Burkholder [4] extended this result to *subordination* martingales: if (d_k), (e_k) are two martingale difference sequences such that (e_k) is subordinate to (d_k), that is, for all $k \ge 1$,

$$|e_k| \le |d_k| \tag{1}$$

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.

then there exists a positive constant c_p, depending only on p, such that for all positive integers n

$$\left\| \sum_{k=1}^{n} e_k \right\|_p \leq c_p \left\| \sum_{k=1}^{n} d_k \right\|_p . \tag{2}$$

A different approach to this inequality was proposed by Kwapień and Woycziński [9] (see also [10]). Two adapted sequence (d_k) and (e_k) are said to be *tangent* if for each $k \geq 1$, we have that the law of d_k conditionally on $\mathcal{F}_{k-1}$ is the same as the law of e_k conditionally on $\mathcal{F}_{k-1}$, that is,

$$P(d_k > \lambda | \mathcal{F}_{k-1}) = P(e_k > \lambda | \mathcal{F}_{k-1}) \tag{3}$$

for all real numbers λ. Answering a conjecture of Kwapień and Woycziński [9], it was proved by Hitczenko [6] and Zinn [15] 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 we have equation (2).

Given two adapted sequences, (e_k) is said to be *strongly dominated by* (d_k) if for each $k \geq 1$,

$$P(|e_k| > \lambda | \mathcal{F}_{k-1}) \leq P(|d_k| > \lambda | \mathcal{F}_{k-1}) \tag{4}$$

for all $\lambda \geq 0$. It is obvious that the case of (1) and the case of (3) are contained in the cases of (4). Thus the following result of Kwapień and Woycziński [9] is a common generalization of these two results: if (d_k), (e_k) are two martingale difference sequences such that (e_k) is strongly dominated by (d_k), then there exists a positive constant c_p, depending only on p, such that for all positive integers n equation (2) holds.

The purpose of this paper is to use a different approach to provide another common generalization of those two results, an even a further extension to Kwapień and Woycziński's result.

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*

$$E[s_k \vee |e_k|] \leq E[s_k \vee |d_k|] \tag{5}$$

then for all positive integers n we have equation (2).

Remark (a). We have that (5) is equivalent to

$$E[(\lambda \vee |e_k|) | \mathcal{F}_{k-1}] \leq E[(\lambda \vee |d_k|) | \mathcal{F}_{k-1}] \tag{6}$$

for all $\lambda \geq 0$. This is because for any $A_k \in \mathcal{F}_{k-1}$ and $a \geq 0$ we have that $(a\chi_{A_k^c} \vee \lambda)$ is predictable, and hence

$$E[(a\chi_{A_k^c} \vee \lambda) \vee |e_k| - a\chi_{A_k^c}] \leq E[(a\chi_{A_k^c} \vee \lambda) \vee |d_k| - a\chi_{A_k^c}]$$

When a intends to infinity, we obtain

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

which is equivalent to (6).

Remark (b). To see that Theorem 1 is really an extension to Kwapień and Woycziński's result, we just simply observe that (4) is equivalent to

$$P(\{|e_k| > \lambda\} \cap A_k) \leq P(\{|d_k| > \lambda\} \cap A_k),$$

and (6) is equivalent to

$$\int_\lambda^\infty P(\{|e_k| > t\} \cap A_k)dt \leq \int_\lambda^\infty P(\{|d_k| > t\} \cap A_k)dt$$

for all $A_k \in \mathcal{F}_{k-1}$.

Remark (c). Once we have Theorem 1, we can obtain that for $\kappa \geq 1$, if

$$P(|e_k| > \lambda|\mathcal{F}_{k-1}) \leq \kappa P(|d_k| > \lambda|\mathcal{F}_{k-1}),$$

we have

$$\left\| \sum_{k=1}^n e_k \right\|_p \leq \kappa c_p \left\| \sum_{k=1}^n d_k \right\|_p . \tag{7}$$

This is because

$$\begin{aligned} \int_\lambda^\infty P(\{|e_k| > t\} \cap A_k)dt &\leq \kappa \int_\lambda^\infty P(\{|d_k| > t\} \cap A_k)dt \\ &= \kappa \int_{\frac{\lambda}{\kappa}}^\infty P\left(\left\{|d_k| > \frac{t}{\kappa}\right\} \cap A_k\right) d\left(\frac{t}{\kappa}\right) \\ &\leq \int_\lambda^\infty P(\{\kappa|d_k| > t\} \cap A_k)dt \end{aligned}$$

Hence

$$E[(\lambda \vee |e_k|)|\mathcal{F}_{k-1}] \leq E[(\lambda \vee \kappa|d_k|)|\mathcal{F}_{k-1}]$$

and equation (7) follows.

Let us give an application of Theorem 1. In fact this application is essentially equivalent to Theorem 1, and indeed will play a large role in its proof. We will consider the probability space $[0,1]^{\mathbb{N}}$ 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

$[0,1]^{\mathbb{N}}$ are measurable. Then two sequences (d_k) and (e_k) are tangent if

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

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

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

$$e_k(x_1, \dots, x_{k-1}, \cdot) = [T_k(x_1, \dots, x_{k-1})] d_k(x_1, \dots, x_{k-1}, \cdot). \tag{8}$$

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 (8), then for all positive integers n we have equation (2).*

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

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

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 $[0,1]^{\mathbb{N}}$ with respect to $(\mathcal{L}_k)$. Suppose that for each positive integer k*

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

for all $t \in [0,1]$ and almost all $x_1, \dots, x_{k-1}$. Then for all positive integers n we have equation (2).

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_N be the σ-field generated by the partition $\{[\frac{i-1}{N}, \frac{i}{N}) : i = 1, 2, \dots, N\}$. Define a filtration $(\mathcal{F}_k)$ on $[0,1]^{\mathbb{N}}$ by $\mathcal{F}_k = \mathcal{L}_{k-1} \otimes \Sigma_N$. Suppose (d_k), (e_k) are $(\mathcal{F}_k)$-adapted. Then for each k and for each $x_1, \dots, x_{k-1}$, we see that $d_k(x_1, \dots, x_{k-1}, \cdot)$ and $e_k(x_1, \dots, x_{k-1}, \cdot)$ are Σ_N-measurable simple functions on $[0,1)$. Therefore d_k and e_k can be written as N-dimensional vectors and $T_k(x_1, \dots, x_{k-1})$ can be represented by a $N \times N$

matrix, that is,

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

where

$$d_k(i) = d_k(x_1, \dots, x_{k-1}, i) = d_k(x_1, \dots, x_k) \text{ if } x_k \in [\tfrac{i-1}{N}, \tfrac{i}{N})$$

$$e_k(i) = e_k(x_1, \dots, x_{k-1}, i) = e_k(x_1, \dots, x_k) \text{ if } x_k \in [\tfrac{i-1}{N}, \tfrac{i}{N})$$

$$T_k = T_k(x_1, \dots, x_{k-1}) = [(a_k(x_1, \dots, x_{k-1}))(i,j)]_{N\times N} = [a_k(i,j)]_{N\times N}$$

The condition of being martingale difference sequences implies that

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

Proposition 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 $\|T_k\|_{L_1([0,1])}$ and $\|T_k\|_{L_\infty([0,1])}$ 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 Proposition 4 to be that: "for almost all $x_1, \ldots, x_{k-1}$, 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 [1] (or see for example [12]).

Theorem A. *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 5. *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

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

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

$$C = \text{Diag}\left[\ \frac{S}{n}, \quad \dots \quad , \frac{S}{n}\ \right]_{n\times n}$$

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. □

Lemma 6. *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 A,

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

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. □

Lemma 7. *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 6, 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 A, 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

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

□

Proof of Proposition 4. From Lemma 7, we know that for each $k \geq 1$ and almost all $x_1, \ldots, x_{k-1}$

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

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

$$h_{k,i_k}(x_1, \ldots, x_{k-1}, \cdot) = [P_{k,i_k}(x_1, \ldots, x_{k-1})] d_k(x_1, \ldots, x_{k-1}, \cdot). \tag{9}$$

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 $\Omega_1 = \Omega_2 = [0,1]^{\mathbb{N}}$. We consider all of the previous random variables 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

$$e_k = E[\varepsilon_{k,I_k} h_{k,I_k} | \mathcal{L} \otimes \{\emptyset, \Omega_2\}].$$

Hence, since conditional expectation is a contraction on L_p

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

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

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

Next, observing (9), since P_{k,i_k} are permutation matrices, for each $k \ge 1$, $i_k = 1, 2, \ldots, S_k$, h_{k,i_k} is just an x_k-rearrangement of d_k. that is

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

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

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

The result follows. □

Proposition 8. *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 [11].

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

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

for all $n = 1, 2, \ldots, N$ 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 [5].

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

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

Lemma 9. $1 \leq p < \infty$. *$(\Omega, \mathcal{F}, P)$ is a probability space. Let $d(\omega, x)$,$e(\omega, x) \in L_p(\Omega \times [0,1])$ be two random variables such that for almost every $\omega \in \Omega$ that*

$$\int_0^t (e(\omega, \cdot))^\# \leq \int_0^t (d(\omega, \cdot))^\#$$

and

$$\int_0^1 d(\omega, \cdot) = \int_0^1 d(\omega, \cdot) = 0.$$

Then given $\epsilon > 0$, there exists a positive integer N and $d', e' \in L_p(\Omega \times [0,1])$ that are measurable with respect to $\mathcal{F} \otimes \Sigma_N$ such that $\|d - d'\|_p$, $\|e - e'\| \leq \epsilon$,

$$\int_0^t (e'(\omega, \cdot))^\# \leq \int_0^t (d'(\omega, \cdot))^\#$$

and

$$\int_0^1 d'(\omega, \cdot) = \int_0^1 d'(\omega, \cdot) = 0.$$

Proof. For every $\epsilon > 0$, pick $0 < \gamma < \min\{\epsilon/[7(\|d\|_p \vee \|e\|_p)], 1/3\}$. Fix $\omega \in \Omega$, and regard the functions as functions of only one variable x on $[0,1]$. Hence there exist simple functions

$$\bar{d} = \sum_{i=1}^S \bar{\alpha}_i \chi_{A_i}$$

$$\bar{e} = \sum_{i=1}^{S} \bar{\beta}_i \chi_{B_i}$$

such that

$$\|\bar{d} - d\|_{L_p([0,1])} \le \gamma \|d\|_{L_1([0,1])}.$$
$$\|\bar{e} - e\|_{L_p([0,1])} \le \gamma \|e\|_{L_1([0,1])}.$$

We may suppose without loss of generality that the sets A_i and B_i are the sets of the form $[r_1, s_1)$, where the r_i and s_i are rational numbers. Furthermore, we will suppose that $A_{i_1} \cap A_{i_2} = \emptyset$ and $B_{i_1} \cap B_{i_2} = \emptyset$ for $i_1 \neq i_2$.

Let $N_0 = N_0(\omega)$ be the least common denominator of all these rational numbers. For each ω, since $d(\omega, \cdot)^\#$, $e(\omega, \cdot)^\#$ are Reimann integrable as a function of x, there is a number $N_1 = N_1(\omega)$ that is a multiple of N_0 and such that for all $n \ge N_1$ that

$$\|E[d(\omega, \cdot)^\# | \Sigma_n] - d(\omega, \cdot)^\#\|_{L_p([0,1])} \le \gamma \|d\|_{L_1([0,1])}.$$
$$\|E[e(\omega, \cdot)^\# | \Sigma_n] - e(\omega, \cdot)^\#\|_{L_p([0,1])} \le \gamma \|e\|_{L_1([0,1])}.$$

Now let $d_n = d\chi_{\{N_1(\omega) \le n\}}$ and $e_n = e\chi_{\{N_1(\omega) \le n\}}$. Then $d_n \to d$ and $e_n \to e$ in $L_p(\Omega \times [0,1])$. So pick N such that

$$\|d_N - d\|_{L_p(\Omega \times [0,1])} < \epsilon/7,$$
$$\|e_N - e\|_{L_p(\Omega \times [0,1])} < \epsilon/7.$$

For each fixed $\omega \in \{N_1(\omega) \le N\}$, $[\frac{i-1}{N}, \frac{i}{N})$ is either contained in some A_j or disjoint to all A_j. Let $\alpha_i = \bar{\alpha}_j$ if $[\frac{i-1}{N}, \frac{i}{N}) \subset A_j$ for some j, and $\alpha_i = 0$ otherwise. Let $\chi_i = \chi_{[\frac{i-1}{N}, \frac{i}{N})}$. Thus

$$\left\| \sum_{i=1}^{N} \alpha_i \chi_i - d_N \right\|_{L_p([0,1])} \le \gamma \|d\|_{L_1([0,1])} \tag{10}$$

and

$$\left(\sum_{i=1}^{N} \alpha_i \chi_i \right)^\# = \sum_{i=1}^{N} \varepsilon_{\sigma(i)} \alpha_{\sigma(i)} \chi_i$$

for some permutation σ, where $\varepsilon_j = \operatorname{sgn}(\alpha_j)$. By Theorem C,

$$\left\| \sum_{i=1}^{N} \varepsilon_{\sigma(i)} \alpha_{\sigma(i)} \chi_i - d_N^\# \right\|_{L_p([0,1])} \le \gamma \|d\|_{L_1([0,1])} \tag{11}$$

and also the analogous statement holds for e.

Now if we set

$$E[d_N^\# | \Sigma_N] = \sum_{i=1}^{N} \alpha_i'' \chi_i$$

then

$$\left\| \sum_{i=1}^{N} \alpha_i'' \chi_i - d_N^{\#} \right\|_{L_p([0,1])} \leq \gamma \|d\|_{L_1([0,1])} \tag{12}$$

Note that in this case that

$$\int_0^t \sum_{i=1}^{N} \alpha_i'' \chi_i = \int_0^t d_N^{\#}$$

if $t = \frac{j}{N}$, $j = 0, 1, 2, \ldots, N$. Then by (11) and (12),

$$\left\| \sum_{i=1}^{N} \alpha_i'' \chi_i - \sum_{i=1}^{N} \varepsilon_{\sigma(i)} \alpha_{\sigma(i)} \chi_i \right\|_{L_p([0,1])} \leq 2\gamma \|d\|_{L_1([0,1])}$$

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

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

we have

$$\left\| \sum_{i=1}^{N} \hat{\alpha}_i \chi_i - \sum_{i=1}^{N} \alpha_i \chi_i \right\|_{L_p([0,1])} \leq 2\gamma \|d\|_{L_1([0,1])} \tag{13}$$

From (10) and (13),

$$\left\| \sum_{i=1}^{N} \hat{\alpha}_i \chi_i - d_N \right\|_{L_p([0,1])} \leq 3\gamma \|d\|_{L_1([0,1])}$$

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

$$\int_0^t \left(\sum_{i=1}^{N} \hat{\alpha}_i \chi_i \right)^{\#} = \int_0^t \sum_{i=1}^{N} \alpha_i'' \chi_i = \int_0^t d_N^{\#}.$$

Furthermore, if we set

$$\zeta = E\left[\sum_{i=1}^{N} \hat{\alpha}_i \chi_i \right]$$

then

$$|\zeta| \leq 3\gamma \|d\|_{L_1([0,1])}.$$

We can also perform this same construction for e, the analogues of $\hat{\alpha}_i$ and ζ being $\hat{\beta}_i$ and η. Thus we see that for $t = \frac{j}{N}$, $j = 0, 1, 2, \ldots, N$

that

$$\begin{aligned}
(14) \qquad & \int_0^t \left(\sum_{i=1}^N \left(\hat{\alpha}_i - \zeta \right) \chi_i \right)^\# \\
\leq \quad & \int_0^t \left(\sum_{i=1}^N \left(|\hat{\alpha}_i| + |\zeta| \right) \chi_i \right)^\# \\
\leq \quad & \int_0^t \left(\sum_{i=1}^N \hat{\alpha}_i \chi_i \right)^\# + 3\gamma \|d\|_{L_1([0,1])} \cdot t \\
= \quad & \int_0^t d_N^\# + 3\gamma \|d\|_{L_1([0,1])} \cdot t \\
\leq \quad & (1+3\gamma) \int_0^t d_N^\#
\end{aligned}$$

and similarly

$$\begin{aligned}
(15) \qquad \int_0^t \left(\sum_{i=1}^N \left(\hat{\beta}_i - \eta \right) \chi_i \right)^\# \quad & \geq \quad \int_0^t e_N^\# - 3\gamma \|e\|_{L_1([0,1])} \cdot t \\
& \geq \quad (1-3\gamma) \int_0^t e_N^\#
\end{aligned}$$

Thus, we are ready to define d' and e'. Let

$$d' = (1+3\gamma) \sum_{i=1}^N (\hat{\alpha}_i - \zeta) \chi_i$$

$$e' = (1-3\gamma) \sum_{i=1}^N (\hat{\beta}_i - \eta) \chi_i$$

It is clear that $E[d'] = E[e'] = 0$. Combining (14) and (15), we have for $t = \frac{j}{N}$, $j = 0, 1, 2, \ldots, N$

$$\int_0^t (e')^\# \leq \int_0^t e_N^\# = \int_0^t d_N^\# \leq \int_0^t (d')^\#.$$

But then by linear interpolation, this follows for all $t \in [0,1]$. Now an easy argument shows that

$$\|d' - d\|_{L_p(\Omega \times [0,1])} \leq 6\gamma \|d\|_{L_1(\Omega \times [0,1])} + \epsilon/7$$

$$\|e' - e\|_{L_p(\Omega \times [0,1])} \leq 6\gamma \|e\|_{L_1(\Omega \times [0,1])} + \epsilon/7$$

and we are done. □

Proof of Theorem 3. For each $1 \le k \le n$, apply Lemma 9, there exists an integer N_k and functions d'_k, e'_k satisfying $\|d'_k - d_k\|_p$, $\|e'_k - e_k\|_p \le \epsilon$ such that (d'_k) and (e'_k) are adapted to $(\mathcal{L}_{k-1} \otimes \Sigma_N)$, where N is the least common multiple of N_k, keep the martingale property, and

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

for all $t \in [0,1]$. By Proposition 8, there exist a positive constant c_p such that

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

$\|d_k - d'_k\|_p \to 0$ 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 [14] 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 10. *Let f and g be real-valued random variables on $(\Omega, \mathcal{F}, P)$. Then*

$$E\left[\lambda \vee |g|\right] \le E\left[\lambda \vee |f|\right] \tag{16}$$

for all nonnegative number λ 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 (16) is equivalent to $E\left[\lambda \vee g^{\#}\right] \le E\left[\lambda \vee f^{\#}\right]$. For the "if" part, let

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

Then

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

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

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

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

Then

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

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

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 [13] shows a concrete representation of a sequence of random variables.

Theorem D. *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 $\sigma(f_1, \ldots, f_n)$ for all $n \geq 0$, then we may suppose that $\sigma(g_1, \ldots, g_n) = \mathcal{L}_n$ for all $n \geq 0$.*

Proof of Theorem 1. We will prove this theorem under the assumption (6). Consider the map $D_k = (d_k, e_k, f_k) : \Omega \times [0,1]^{\mathbb{N}} \to \mathbb{R}^3$ 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 $\sigma(D_1, \ldots, D_{k-1})$. Apply the previous theorem to get $\widetilde{D}_k = (\widetilde{d}_k, \widetilde{e}_k, \widetilde{f}_k) : [0,1]^k \to \mathbb{R}^3$ such that $(\widetilde{D}_k)$ has the same law as (D_k) and $\sigma(\widetilde{D}_1, \ldots, \widetilde{D}_k) = \mathcal{L}_k$.

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

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

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

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

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

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

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

The result then follows from Lemma 10 and Theorem 3. □

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. Also, we would like to express our thanks to Mitch Taibleson for bringing this problem to our attention.

References

1. G. Birkhoff, *Tres observaciones sobre el algebra lineal.* Univ. Nac. Tucumán Rev. Ser **A5** (1946), 147–150.

2. D. L. Burkholder, *Distribution function inequalities for martingales.* Ann. Probability **1** (1973), 19–42.
3. D. L. Burkholder, *A geomtrical characterization of banach spaces in which marti difference sequences are unconditional.* Ann. Probability **9** (1981), 997–1011.
4. D. L. Burkholder, *Sharp inequalities for martingales and stochastic integrals.* Colloque Paul Lévy sur les Processus Stochastiques (Palaiseau, 1987). Astérisque (1988), 75–94.
5. L. A. Crowe, J. A. Zweibel and P. C. Rosenbloom, *Rearrangements of functions*, Journal of functional analysis **66** (1986), 43291–438.
6. P. Hitczenko, *Comparison of moments for tangent sequences of random variables.* Probab. Theory Related Fields **78** (1988), 223–230.
7. P. Hitczenko, *On a domination of sums of random variables by sums of conditionally independent ones.* Ann. Probability **22** (1994), 453–468.
8. P. Hitczenko and S. J. Montgomery-Smith, *Tangent sequences in Orlicz and rearrangement invariant spaces.* Math. Proc. Camb. Phil. Soc. **119** (1996), 91–101.
9. S. Kwapień and W. A. Woyczyńsi, *Semimartingale integrals via decoupling inequalities and tangent processes.* Probab. Math. Statist. **12** (1991).
10. S. Kwapień and W. A. Woyczyńsi, *Random series and stochastic integrals. Single and multiple.* Birkhauser, Boston. (1996).
11. J. Lindenstrauss ans L. Tzafriri, *Classical Banach spaces II.* Springer-Verlag. (1979).
12. H. Minc, *Nonnegative matrices.* Wiley Interscience. (1988).
13. S. J. Montgomery-Smith, *Concrete representation of martingales.* Electronic J. Probab. **3**, (1998), paper 15
14. J. Peetre, *Espaces d'interpolation, généralisations, applications.* Rend. Sem. Mat. Fis. Milano **34**, (1964), 83-92.
15. 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