The Annals of Probability
1994, Vol. 22, No. 4, 1745–1765

CONTRACTION AND DECOUPLING INEQUALITIES FOR MULTILINEAR FORMS AND U-STATISTICS

V. H. DE LA PEÑA[1], S. J. MONTGOMERY-SMITH[2] AND JERZY SZULGA

Columbia University, University of Missouri and Auburn University

We prove decoupling inequalities for random polynomials in independent random variables with coefficients in vector space. We use various means of comparison, including rearrangement invariant norms (e.g., Orlicz and Lorentz norms), tail distributions, tightness, hypercontractivity and so forth.

1. Introduction.

1.1. *Background and scope of the paper.* Decoupling principles stem from the theory of martingale transforms (cf. [2]). For homogeneous random forms of rank $k \geq 2$, decoupling principles were introduced in [11], [18] and [19] (in some special cases, they were known to Pisier cf. [20]), and subsequently became essential tools in multiple integration (cf. [9], [12] and [23]–[25]). One of the most appealing interpretations of such a principle is the reducibility of a study of multiple random series (respectively, of multiple stochastic integrals) to a consecutive treatment of single random series (respectively, of single stochastic integrals that allows one to treat a multiple integral as an Itô-type iterate integral). The concept of a random chaos goes back to Wiener [27] (see also [28]), who elaborated what we call here a real-valued coupled Gaussian chaos. Decoupling inequalities may be viewed as embedding-projection procedures, since a decoupled random chaos is nothing but a lacunary random chaos. In comparison to the classical L^2-theory of multiple summation or integration, decoupling principles make up for the lack of L^2-isometries.

Since the first publication of the aforementioned decoupling principle, the theory has branched into several directions. For example, comparison of tangent processes (cf. [5], [6] and [8]) is akin to the classical decoupling principle. Further contributions can be found, for example, in [3], [4], [8], [11], [21] and [29]. In some of the aforementioned papers (e.g., [3], [12], [18], [19]) the symmetry of random variables is essential for the fulfillment of the decoupling principle,

Received November 1992; revised April 1994.

[1]Research supported in part by NSF Grant DMS-91-08006 and a grant from the CNRS of France.

[2]Research supported in part by NSF Grants DMS-90-01796 and DMS-92-01357.

AMS 1991 *subject classifications.* Primary 60B11, 46M05; secondary 60H07, 46E30, 60E15, 62H05, 62G30.

Key words and phrases. Decoupling principle, symmetric tensor products, random polynomials, multiple random series, multiple stochastic integrals, random multilinear forms, random chaos, tail inequalities, polarization, symmetrization, Banach space, rearrangement invariant, Orlicz space, Lorentz space, Rademacher sequence, U-statistics.

while other papers (e.g., [8], [9], [29]) point out the role of positivity. Norms of L^p-spaces, or more general, of Orlicz spaces (basically, subject to growth restriction) have become typical means of comparison of two classes of vector random variables.

In this paper, we prove decoupling inequalities for random variables that are not necessarily symmetric. Theorems 2.1 and 2.3 in Section 2 and Theorem 3.8 in Section 3 are our main results. The decoupling principle by means of probability tails, Theorem 3.8, immediately ensures the parity of tightness of two types of chaos (that Gaussian decoupled and coupled chaos are simultaneously tight was proved in [11]).

A number of decoupling results are obtained for arbitrary rearrangement invariant norms and Orlicz functionals. In particular, we provide one extended example regarding certain Lorentz norms (important in the approximation theory). Another application is the decoupling principle for U-statistics (a result as in Theorem 2.3 was proven in [4]).

The utilized techniques are based on ideas borrowed from [11], while some are taken from [13]. Proofs are straightforward and point out the algebraic nature of decoupling that is fruitfully merged with a widely understood context of convexity. A rule of thumb is that, in the field of random diagonal-free polynomials, a "definable" is "decouplable." The obtained robust constants are tightly estimated and are sharper than previously known constants.

In the last section, we show tail probability decoupling results for polynomials of symmetric random variables. This section makes use of techniques from [1].x

1.2. *Notation.* Random variables in this paper are defined on a separable probability space $(\Omega, \mathcal{A}, \mathsf{P})$ that is rich enough to carry independent sequences. A sequence of real random variables is denoted by $\boldsymbol{\xi} = (\xi_1, \xi_2, \ldots)$ and a matrix of real random variables is denoted by $\mathbb{X} = [\boldsymbol{\xi}_1, \ldots, \boldsymbol{\xi}_n]$, where $\boldsymbol{\xi}_j = (\xi_{j1}, \xi_{j2}, \ldots)$. We will make use of one particular sequence, the Rademacher sequence $\varepsilon = (\varepsilon_i)$, where ε_i are independent random variables taking values ± 1 with probability $1/2$.

Let $\mathbb{E} = (\mathbb{E}, \|\cdot\|)$ denote a real Banach space. We will be considering $\mathbb{E}$-valued random variables, that is, strongly measurable mappings from Ω into $\mathbb{E}$.

Let k be a positive integer and let $f = (f_{i_1,\ldots,i_k})$ be an array of vectors from $\mathbb{E}$ taking only finitely many nonzero values. Throughout the paper, all such arrays are assumed to vanish on diagonals (we will say *diagonal-free*); that is, $f_{i_1,\ldots,i_k} = 0$ if at least two indices $i_j, i_{j'}$ are equal.

The main object of our interest will be the k-homogeneous random polynomial

$$Q(f; \mathbb{X}) \stackrel{\text{df}}{=} Q(f; \boldsymbol{\xi}_1, \ldots, \boldsymbol{\xi}_k) \stackrel{\text{df}}{=} \sum_{i_1,\ldots,i_k} f_{i_1,\ldots,i_k} \xi_{1i_1} \cdots \xi_{ki_k}.$$

We desire to compare this random polynomial with the "undecoupled" version, that is,

$$Q(f;\xi^k) \overset{\text{df}}{=} Q(f;\xi,\dots,\xi) \overset{\text{df}}{=} \sum_{i_1,\dots,i_k} f_{i_1,\dots,i_k}\xi_{i_1}\cdots\xi_{i_k}. \tag{1.1}$$

The first term in the above definition will be introduced as a notational convenience. We will feel quite free to stretch the use of this notation. So for example, we might write

$$Q(f;\xi^r,\eta^{k-r}) = Q(f;\xi,\dots,\xi,\eta,\dots,\eta) = \sum_{i_1,\dots,i_k} f_{i_1,\dots,i_k}\xi_{i_1}\cdots\xi_{i_r}\eta_{i_{r+1}}\cdots\eta_{i_k}.$$

We hope to convince the reader of the value of this notation, as it enables us to write many of the proofs in a more compact form, and may ultimately lead to clearer thinking on the subject. For the unconvinced reader, we hope that we have explained the notation sufficiently that he/she will be able to rewrite all the subsequent proofs and statements in a more familiar form.

Many of the inequalities that we introduce require the array f to satisfy certain symmetry conditions; for this reason we introduce the symmetrized version of f:

$$\widehat{f}_{i_1,\dots,i_k} \overset{\text{df}}{=} \frac{1}{k!}\sum_{\sigma} f_{i_{\sigma_1},\dots,i_{\sigma_k}},$$

where the sum is taken over all permutations of the set $[1,k] = \{1,\dots,k\}$, and let $\widehat{Q}(f;\cdot) = Q(\widehat{f},\cdot)$. Note that for the undecoupled random polynomial, symmetry makes no change: $\widehat{Q}(f;\xi^k) = Q(f;\xi^k)$.

In the sequel, we will occasionally refer to tetrahedral arrays, that is, f such that $f_{i_1,\dots,i_k} = 0$, if indices fail to satisfy $i_1 < \cdots < i_k$.

We will make frequent use of the following identity, which is known as the Mazur–Orlicz polarization formula [17]:

$$\widehat{Q}(f;\xi_1,\dots,\xi_k) = \frac{1}{k!}\sum_{\delta=(\delta_1,\dots,\delta_k)\in\{0,1\}^k} (-1)^{k-|\delta|} Q\big(f;(\delta_1\xi_1+\cdots+\delta_k\xi_k)^k\big), \tag{1.2}$$

where $|\delta| = \sum_i \delta_i$. Switching to a Rademacher sequence ε, we can write

$$\widehat{Q}(f,\xi_1,\dots,\xi_k) = \frac{1}{k!}\mathsf{E}\varepsilon_1\cdots\varepsilon_k Q\left(f,\left(\sum_{i=1}^k \varepsilon_i\xi_i\right)^k\right), \tag{1.3}$$

where the expectation is only over the Rademacher sequence ε.

Rearrangement invariant spaces. By $(\mathbb{L},\|\cdot\|_{\mathbb{L}})$ we denote a rearrangement invariant Banach space of integrable random variables (so that the norm of a random variable depends only on its probability distribution), $\mathbb{L}\subset L^1(\mathsf{P})$, defined on a separable probability space $(\Omega,\mathcal{A},\mathsf{P})$ that is rich enough to carry independent sequences. For more information on rearrangement invariant spaces, we refer the reader to [15], for example. The basic examples of rearrangement invariant spaces are $\mathbb{L} = L_p$, for $1 \leq p \leq \infty$ (nothing more is needed in many parts of this paper), Orlicz spaces and Lorentz spaces. We will sometimes use the abbreviation r.i. for *rearrangement invariant.*

The important property of rearrangement invariant spaces that we shall use is the following:

(1.4) Conditional expectations are contractions acting on $\mathbb{L}$.

The reader unfamiliar with r.i. spaces should not that this is true of L_p.

We denote by $\mathbb{L}(\mathbb{E})$ the Banach space of $\mathbb{E}$-valued random variables (i.e., strongly measurable mapping from Ω into $\mathbb{E}$) whose norms belong to $\mathbb{L}$, and let $\|\theta\|_{\mathbb{L}(\mathbb{E})} = \|\|\theta\|_{\mathbb{E}}\|_{\mathbb{L}}$. Thus if $\mathbb{L} = L^p$, then $\|\theta\|_{\mathbb{L}(\mathbb{E})} = \|\theta\|_{L_p(\mathbb{E})} = (\mathsf{E}\|\theta\|_{\mathbb{E}}^p)^{1/p}$. In the sequel we sometimes omit the subscript indicating the space if it causes no ambiguity.

2. Decoupling for rearrangement invariant norms.

Interchangeability. In the sequel, we will use several times the following elementary feature of interchangeable random sequences $\boldsymbol{\xi}_1, \ldots, \boldsymbol{\xi}_r$ (i.e., such that each permutation has the same distribution). Suppose that each $\boldsymbol{\xi}_k$ is itself a sequence of independent random variables. Denote by $\mathcal{G}_r$ the σ-field spanned by $\sum_{j=1}^r \boldsymbol{\xi}_j$. Let f be diagonal-free. Then if $j_1, \ldots, j_k \leq r$,

$$\mathsf{E}\Big[Q\big(f, \boldsymbol{\xi}_{j_1}, \ldots, \boldsymbol{\xi}_{j_k}\big) \mid \mathcal{G}_r\Big] = r^{-k} Q\big(f, (\boldsymbol{\xi}_1 + \cdots + \boldsymbol{\xi}_r)^k\big). \tag{2.1}$$

We should point out that the last term represents the random polynomial

$$Q\big(f, (\boldsymbol{\xi}_1 + \cdots + \boldsymbol{\xi}_r)^k\big) = \sum_{i_1, \ldots, i_k} f_{i_1, \ldots, i_k} \big(\xi_{1 i_1} + \cdots + \xi_{r i_1}\big) \cdots \big(\xi_{1 i_k} + \cdots + \xi_{r i_k}\big).$$

Equation (2.1) follows because

$$\mathsf{E}\big(f_{i_1, \ldots, i_k} \xi_{j_1 i_1} \cdots \xi_{j_k i_k} \mid \mathcal{G}_r\big) = f_{i_1, \ldots, i_k} \mathsf{E}\big(\xi_{j_1 i_1} \mid G_r\big) \cdots \mathsf{E}\big(\xi_{j_k i_k} \mid \mathcal{G}_r\big)$$

because f is diagonal-free and hence $\xi_{j_1 i_1} \ldots, \xi_{j_k i_k}$ are independent if $f_{i_1, \ldots, i_k} \neq 0$, and also because for $j \leq r$,

$$\mathsf{E}\big(\xi_{ji} \mid \mathcal{G}_r\big) = r^{-1}(\xi_{1i} + \cdots + \xi_{ri}).$$

We also point out the following easy consequence of the triangle inequality for $\mathbb{L}$:

$$\|\widehat{Q}(f, \mathbb{X})\|_{\mathbb{L}(\mathbb{E})} \leq \|Q(f, \mathbb{X})\|_{\mathbb{L}(\mathbb{E})}. \tag{2.2}$$

Now we are ready to present our first decoupling inequality. This result allows us to decouple random polynomials in the rearrangement invariant norm.

THEOREM 2.1. *Let $f = (f_{i_1, \ldots, i_k})$ be a diagonal-free array of vectors from $\mathbb{E}$. Let $\boldsymbol{\xi}, \boldsymbol{\xi}_1, \ldots, \boldsymbol{\xi}_k$ be sequences of integrable independent real random variables.*

Let $\mathbb{L}$ be a r.i. space of random variables, containing $\xi_1 \cdots \xi_k$ (hence, norms of all finitely supported polynomials spanned by $\xi_1, \dots, \xi_k$).

(A) *Assume that $\xi, \xi_1, \xi_2, \dots$ are independent and identically distributed. Then*

$$\|Q(f, \xi^k)\|_{\mathbb{L}(\mathbb{E})} \le A \|Q(f, \xi_1, \dots, \xi_k)\|_{\mathbb{L}(\mathbb{E})},$$

where $A = A_k \sim (2k)^k$ or, if $\mathsf{E}\,\xi = 0$, $A_k = k^k$.

(B) *Assume that $\xi, \xi_1, \dots, \xi_k$ are interchangeable. Then*

$$\|\widehat{Q}(f, \xi_1, \dots, \xi_k)\|_{\mathbb{L}(\mathbb{E})} \le B \|Q(f, \xi^k)\|_{\mathbb{L}(\mathbb{E})},$$

where $B = B_k \sim k^k/k!$.

PROOF. During this proof, we will suppress the subscript $\mathbb{L}(\mathbb{E})$ on the norms.

(A)

Step 1. *Centering procedure.* Denote $\overline{\xi} = \xi - \mathsf{E}[\xi]$, $\mathbf{m} = (m_1, m_2, \dots)$, where $m_i = \mathsf{E}[\xi_i]$. For $1 \le r \le k$, if $f = (f_{i_1, \dots, i_r})$, then we have

$$\|Q(f, \overline{\xi}_1, \dots, \overline{\xi}_r)\| \le 2^r \|Q(f, \xi_1, \dots, \xi_r)\|. \tag{2.3}$$

Indeed, by interchangeability

$$\begin{aligned}
&\|Q(f, \overline{\xi}_1, \dots, \overline{\xi}_r)\| \\
&\quad = \|Q(f, \xi_1 - \mathbf{m}, \dots, \xi_r - \mathbf{m})\| \\
&\quad = \left\| \sum_{(\delta_1, \dots, \delta_r) \in \{0,1\}^r} \sum_{i_1, \dots, i_r} f_{i_1, \dots, i_r} \xi_{1 i_1}^{\delta_1} \cdots \xi_{r i_r}^{\delta_r} m_{1 i_1}^{1-\delta_1} \cdots m_{r i_r}^{1-\delta_r} \right\| \\
&\quad \le \sum_{j=0}^{r} \binom{r}{j} \left\| \sum_{i_1, \dots, i_r} f_{i_1, \dots, i_r} \xi_{1 i_1} \cdots \xi_{j i_j} m_{j+1, i_{j+1}} \cdots m_{r i_r} \right\| \\
&\quad = \sum_{j=0}^{r} \binom{r}{j} \|Q(f, \xi_1, \dots, \xi_j, m^{(r-j)})\|.
\end{aligned}$$

The latter expression is equal to

$$\begin{aligned}
&\sum_{j=0}^{r} \binom{r}{j} \left\| Q\big(f, \xi_1, \dots, \xi_j, \mathsf{E}[\xi_{j+1} \mid \mathbb{X}_j], \dots, \mathsf{E}[\xi_r \mid \mathbb{X}_j]\big) \right\| \\
&\qquad = \sum_{j=0}^{r} \binom{r}{j} \|\mathsf{E}[Q(f, \xi_1, \dots, \xi_j, \xi_{j+1}, \dots, \xi_r) \mid \mathbb{X}_j]\|,
\end{aligned}$$

where $\mathbb{X}_j$ is the matrix $[\xi_1, \dots, \xi_j]$. Using the contractivity property (1.4), we estimate the preceding term from above by

$$\sum_{j=0}^{r}\binom{r}{j}\|Q(f,\xi_1,\ldots,\xi_r)\| = 2^r\|Q(f,\xi_1,\ldots,\xi_r)\|.$$

Step 2: *Proving* (A). Arguing similarly to Step 1, we note that

$$\|Q(f,\xi^k)\| = \big\|Q\big(f,(\overline{\xi}+\mathbf{m})^k\big)\big\| \le \sum_{r=0}^{k}\binom{k}{r}\|Q(f,\overline{\xi}^r,\mathbf{m}^{k-r})\| := Q_0.$$

Now, using (1.4) and noting that $\mathsf{E}(\overline{\xi}_1+\cdots+\overline{\xi}_r \mid \overline{\xi}_1) = \overline{\xi}_1$, which is distributed as $\overline{\xi}$, it follows that

$$Q_0 \le \sum_{r=0}^{k}\binom{k}{r}\big\|Q\big(f,(\overline{\xi}_1+\cdots+\overline{\xi}_r)^r,\mathbf{m}^{k-r}\big)\big\|.$$

Then, by virtue of (2.1), the latter expression is equal to

$$\sum_{r=0}^{k}\binom{k}{r}\Big\|r^r\mathsf{E}\Big[Q\big(f,\overline{\xi}_1,\ldots,\overline{\xi}_r,\mathbf{m}^{k-r}\big)\mid \mathcal{G}_r\Big]\Big\| =: Q_1.$$

Using (1.4) and applying the centering procedure (2.3), the above term has the upper bounds

$$\begin{aligned} Q_1 &\le \sum_{r=0}^{k}\binom{k}{r}\big\|r^rQ\big(f,\overline{\xi}_1,\ldots,\overline{\xi}_r,\mathbf{m}^{k-r}\big)\big\| \\ &\le \sum_{r=0}^{k}\binom{k}{r}(2r)^r\big\|Q\big(f,\xi_1,\ldots,\xi_r,\mathbf{m}^{k-r}\big)\big\| =: Q_2. \end{aligned}$$

Then, by interchangeability, independence of columns and (1.4) again, we keep estimating:

$$\begin{aligned} Q_2 &= \sum_{r=0}^{k}\binom{k}{r}(2r)^r\big\|\mathsf{E}\big[Q(f,\xi_1,\ldots,\xi_r,\xi_{r+1},\ldots,\xi_k)\mid \mathbb{X}_r\big]\big\| \\ &\le \sum_{i=0}^{k}\binom{k}{r}(2r)^r\|Q(f,\xi_1,\ldots,\xi_r,\xi_{r+1},\ldots,\xi_k)\| \\ &= A_k\|Q(f,\xi_1,\ldots,\xi_k)\|. \end{aligned}$$

Clearly, $A_k \le (2k+1)^k$ [notice that $A_k \ge c(2k)^k$]. If $\mathsf{E}\xi = 0$, the use of the centering procedure and the triangle inequality is superfluous; hence the constant decreases to k^k.

(B) By the Mazur–Orlicz polarization formula (1.2), and (2.1), we obtain the bounds

$$\begin{aligned}
\|\widehat{Q}(f,\xi_1,\dots,\xi_k)\| &= \left\| \frac{1}{k!} Q\left(f, \sum_{\delta} (-1)^{k-|\delta|} (\delta_1, \xi_1 + \cdots + \delta_k \xi_k)^k \right) \right\| \\
&\le \sum_{\delta} \left\| \frac{1}{k!} Q\big(f, (\delta_1 \xi_1 + \cdots + \delta_k \xi_k)^k\big) \right\| \\
&= \sum_{r=0}^{k} \binom{k}{r} \left\| \frac{1}{k!} Q\big(f, (\xi_1 + \cdots + \xi_r)^k\big) \right\| \\
&\le \sum_{r=0}^{k} \binom{k}{r} \left\| \frac{r^k}{k!} Q(f, \xi_1^k) \right\| \\
&= B_k \| Q(f, \xi_1^k) \|.
\end{aligned}$$

That $B_k \sim k^k/k!$ is easy to verify. The proof is complete. □

2.1. *Extended multilinear forms and U-statistics.* In this section we show how to extend Theorem 2.1 to the so called U-statistics. Let k be a positive integer. Let $F = (f_{i_1,\dots,i_k})$ be an array of strongly Borel measurable functions $F_{i_1,\dots,i_k}\colon \mathbb{R}^k \to \mathbb{E}$ such that:

(F0) $F_{i_1,\dots,i_k} = 0$ if some i_j and $i_{j'}$ are identical for $j \neq j'$.
(F1) $F_{i_1,\dots,i_k} = 0$ for all but finitely many $(i_1,\dots,i_k)$.

Then we are going to consider U-statistics, that is, expressions of the form

$$F(\xi_1,\dots,\xi_k) \stackrel{\text{df}}{=} \sum_{i_1,\dots,i_k} F_{i_1,\dots,i_k}(\xi_{1i_1},\dots,\xi_{ki_k}),$$

where $\xi_1,\dots,\xi_k$ are real-valued random variables. (Here, $\mathbb{R}$ could be replaced with any other measure space, but there is no loss of generality to take it as $\mathbb{R}$). We are going to employ the same notational devices as for the random polynomials, so that the undecoupled U-statistic is written

$$F(\xi^k) \stackrel{\text{df}}{=} \sum_{i_1,\dots,i_k} F_{i_1,\dots,i_k}(\xi_{i_1},\dots,\xi_{i_k}).$$

As before, in order to prove the results, we require certain symmetry properties to hold for F. So we define the symmetrized version of F as

$$\widehat{F}(x_1,\dots,x_k) \stackrel{\text{df}}{=} \frac{1}{k!} \sum_{\sigma} F_{i_{\sigma_1},\dots,i_{\sigma_k}}(x_{i_{\sigma_1}},\dots,x i_{i_{\sigma_k}}),$$

where the sum runs over all permutations of $[1,k]$, and we set

$$\widehat{F}(\xi_1,\dots,\xi_k) \stackrel{\text{df}}{=} \sum_{i_1,\dots,i_k} \widehat{F}_{i_1,\dots,i_k}(\xi_{1i_1},\dots,\xi_{ki_k}).$$

Decoupling results were proved by de la Peña [4] for Orlicz modulars (and so by Note 8 in Section 4, one can obtain results for all rearrangement invariant spaces). We will prove similar decoupling results, weakening some of the hypotheses.

More interestingly, we are going to prove the decoupling results for U-statistics as a corollary of Theorem 2.1, which decouples random polynomials. The technique is to approximate the U-statistic as a sum of random polynomials. That is, let D be an integer, and for $1 \le d \le D$, let $f^d = (f^d_{i_1,\dots,i_k})$ be a diagonal-free array of vectors in $\mathbb{E}$ taking only finitely many nonzero values, and let $(\xi_1^d\colon 1 \le d \le D), \dots, (\xi_k^d\colon 1 \le d \le D)$ be sequences of independent random variables. Then we set

$$R(f;\xi_1,\dots,\xi_k) \stackrel{\text{df}}{=} \sum_{d=1}^{d} Q\left(f^d;\xi_1^d,\dots,\xi_k^d\right). \tag{2.4}$$

Then the remarkable thing is that the proof of Theorem 2.1 works for $R(f;\xi_1,\dots,\xi_k)$ exactly as it does for $Q(f;\xi_1,\dots,\xi_k)$; that is, we have the following result.

THEOREM 2.2. *Theorem* 2.1 *is valid for the multilinear form* (2.4).

A version of the following result for Orlicz modulars $\mathsf{E}\phi(\cdot)$, where ϕ was a moderately increasing function, was proved in [19]. In that paper, terms of the underlying sums were sign-randomized, that is, each $F(\mathbf{i},\cdot)$ was multiplied by Walsh functions $\varepsilon_{i_1}\cdots\varepsilon_{i_k}$. More precisely, the decoupling was proved for

$$(F\circ\varepsilon)(\xi_1,\dots,\xi_k) \stackrel{\text{df}}{=} \sum_{\mathbf{i}} \varepsilon_{i_1}\cdots\varepsilon_{i_k} F_{i_1,\dots,i_k}\left(\xi_{1i_1},\dots,\xi_{ki_k}\right), \tag{2.5}$$

where ε is independent of $\mathbb{X}$. That the presence of Walsh functions is not necessary in the context of Orlicz modulars was shown in [4]. We observe that the following result, generalizing theorems in the mentioned papers, is implicit in the main decoupling principle. Moreover, constants remain the same. For the sake of completeness, we give the full proof.

In the proof we will use the fact that any inequality involving norms of functions of discrete r.v.s, that converge to some limits, is preserved for these limits. Fix $\mathbf{i}$, say $\mathbf{i} = (\mathbf{1},\dots,\mathbf{k})$. Consider $X = F(\xi_1,\dots,\xi_k)$. We may assume that the probability space $(\Omega,\mathcal{F},\mathsf{P})$ and $\mathcal{F}$ is spanned by $\xi_1,\xi_2,\dots$. Also, we may assume that it is separable; that is, $\mathcal{F}_n = \sigma\{\bigcup_n \mathcal{F}_n\}$, where $\mathcal{F}_n$ are finite σ-fields. Put $\xi_i^n = n = \mathsf{E}[\xi \mid \mathcal{F}_n]$. Thus $\mathsf{E}[X\mid\mathcal{F}_n] \to \mathsf{E}[X\mid\mathcal{F}] = X$ a.s. and in $\mathbb{L}$.

THEOREM 2.3. *Let* $F\colon \mathbb{N}^k\times\mathbb{R}^k \to \mathbb{E}$ *satisfy* (F0) *and* (F1), *and also the additional condition*:

(F2) $F\left(\mathbf{i};\xi_{\mathbf{i}_1},\dots,\xi_{\mathbf{i}_\mathbf{k}}\right) \in \mathbb{L}$ *for every* $\mathbf{i}\in\mathbb{N}^\mathbf{k}$.

Let $\xi_1,\xi_2,\dots$ *be sequences of independent random variables.*

(A′) *Let* $\|\cdot\|$ *be a r.i. norm. Let* $\xi_1, \xi_2, \ldots$ *be independent and identically distributed. Then*

$$\|F(\xi^k)\| \le A\|F(\mathbb{X})\|,$$

where A is the constant from Theorem 2.1(A).

(B′) *Let* $\xi_1, \ldots, \xi_k$ *be interchangeable* (*in particular*, *i.i.d*) *and let* $\|\cdot\|$ *be a r.i. norm. Then*

$$\|\widehat{F}(\mathbb{X})\| \le B\|F(\xi^k)\|,$$

where B is the constant from Theorem 2.1(B).

PROOF. By Note 8 in Section 4, we may assume that the rearrangement space $\mathbb{L}$ is separable. In that case, we may assume without loss of generality that $\xi_1, \xi_2, \ldots$ are real discrete random variables.

Thus we may assume that the random variables are defined on a product probability space

$$\left(\prod_{ij} \Omega_{ij}, \left(F^{\otimes N}\right)^{\otimes k}, \left(\mathsf{P}^{\otimes N}\right)^{\otimes k}\right),$$

where the Ω_{ij} are equal, $\Omega_i = \Pi_j \Omega_{ij}$ and the superscript $\otimes$ indicates the product σ-field and the product probability, respectively. So, let

$$\xi_i = \sum_m x_{im} \mathbb{1}_{A_{im}},$$

where $A_{i1}, A_{i2}, \ldots \subset \Omega_i$ are bases of rectangular sets that form a disjoint finite partition of Ω, and let (A_{sim}), $s = 1, \ldots, k$, be independent copies of (A_{im}). Put $I_{sim} = \mathbb{1}_{A_{sim}}$; hence

$$\xi_{si} = \sum_m x_{im} I_{sim}, \qquad s = 1, \ldots, k.$$

Then

$$\sum_{i_1, \ldots, i_k} F_{ik}\left(\xi_{1i_1}, \ldots, \xi_{ki_k}\right) = \sum_{m_1, \ldots, m_k} \sum_{i_1, \ldots, i_k} F_{i_1, \ldots, i_k}\left(x_{i_1 m_1}, \ldots, x_{i_k m_k}\right) I_{1 i_1 m_1} \cdots I_{k i_k m_k}.$$

Now, we can apply Theorem 2.2, and the proof is complete. □

2.2. *An example in a certain Lorentz space.* Motivated by the results in [5], where the problem as to when expectation results imply tail probability results is treated, we obtain the following asymptotic tail probability comparison.

PROPOSITION 2.4. *Let ξ or η be the norm of $\widehat{F}(\boldsymbol{\xi}^k)$ or $\widehat{F}(\mathbb{X})$, and let $W\colon [0,\infty) \to [0,\infty)$ be an increasing function such that there exist constants $p > 1$ and $c > 0$ satisfying $W(st) \le cs^p W(t)$ for all $0 < s < 1$ and all $t > 0$. Then there is a constant C, depending only on p and c, such that*

$$\limsup_{t\to\infty} W(t)\mathsf{P}(\xi \ge t) \le \limsup_{t\to\infty} W(t)\mathsf{P}(\eta \ge Ct).$$

This result is a consequence of Theorem 2.3, and follows by arguments from the theory of Lorentz–Zygmund spaces. If ξ is a random variable, let $F(t) = \mathsf{P}(|\xi| \ge t)$ and define the decreasing rearrangement of ξ to be the function $\xi^*(t) \overset{\text{df}}{=} \sup\{s\colon F(s) > t\}$ (i.e., the right-continuous inverse of F). Obviously, $|\xi|$ and ξ^* are equidistributed. When ξ is integrable, an average operator is often considered:

$$\xi^{**}(t) = \frac{1}{t}\int_0^t \xi^*(u)\,du,$$

which corresponds to a rearrangement invariant norm for every $t > 0$. Therefore, our decoupling inequalities for U-statistics hold for $\Phi(X) = (\|X\|)^{**}$. That is, denoting by ξ or η the norm of $\widehat{F}(\boldsymbol{\xi}^k)$ or $\widehat{F}(\mathbb{X})$, we have

$$\xi^{**}(t) \le C\eta^{**}(t)$$

for some constant $C > 0$.

Now consider the Lorentz–Zygmund space defined by the quasinorm $|||f|||= \sup_x w(x)\,\xi^*(x)$, where $w\colon [0,1] \to [0,\infty)$ is an increasing function. Note that $|||f||| \le 1$ if and only if $\sup_t W(t)\mathsf{P}(|f| \ge t) \le 1$, where $w(t) = 1/(W^{-1}(1/t))$. If W satisfies the relation given in Proposition 2.4, then for some constant c, the function w satisfies the relation $w(x) \le ca^{-1/p}w(xa)$ for $a \le 1$. Then it is possible to show that $|||f||| \le |||f^{**}||| \le C|||f|||$. Indeed, the first inequality is obvious, and for the second,

$$w(x)f^{**}(x) = w(x)\int_0^1 f^*(xa)\,da \le c\int_0^1 a^{-1/p}w(xa)f^*(xa)\,da \le \frac{cp}{p-1}|||f|||.$$

Thus, to show Proposition 2.4, let $F(t) = \mathsf{P}(\xi \ge t)$ and $G(t) = \mathsf{P}(\eta \ge t)$. Then

$$\sup_t W(t)F(t) \le \sup_t W(t)\,G(Ct).$$

If we now set $w(x) = 0$, for $x \ge x_0$, the same argument applies, and letting $x_0 \to \infty$, we obtain

$$\limsup_{t\to\infty} W(t)F(t) \le \limsup_{t\to\infty} W(t)\,G(Ct).$$

3. A discourse on probability tails.

3.1. *L^p-estimates imply tail estimates.*

3.1.1. *Auxiliary results.* The following result can be found in [1].

LEMMA 3.1. *Let $(X, X_1, \dots, X_n)$ be a sequence of positive i.i.d. random variables. Then for all positive integers n, all $\alpha > 0$ and all $0 \leq \theta \leq n$,*

$$\mathsf{P}(X \geq \alpha) \geq \frac{\theta}{n} \quad \Rightarrow \quad \mathsf{P}\Big(\sup_{1 \leq i \leq n} X_i \geq \alpha\Big) \geq \frac{\theta}{1+\theta},$$

$$\mathsf{P}(X \geq \alpha) \leq \frac{\theta}{n} \quad \Rightarrow \quad \mathsf{P}\Big(\sup_{1 \leq i \leq n} X_i \geq \alpha\Big) \leq \theta.$$

PROOF. To show the first inequality, observe first that for $\theta > 0$,

$$\left(1 - \frac{\theta}{n}\right)^n \leq \frac{1}{(1+\theta)}.$$

Hence, by the independence assumption,

$$\begin{aligned}
\mathsf{P}\Big(\sup_j X_j \geq \alpha\Big) &= 1 - \mathsf{P}\Big(\sup_j X_j < \alpha\Big) \\
&= 1 - \prod_{j=1}^{n} \mathsf{P}(X_j < \alpha)r \\
&\geq 1 - \left(1 - \frac{\theta}{n}\right)^n \geq 1 - \frac{1}{(1+\theta)} = \frac{\theta}{(1+\theta)}.
\end{aligned}$$

The second inequality is easy: from the imposed condition, one gets

$$\mathsf{P}\Big(\sup_{1 \leq j \leq n} X_j \geq \alpha\Big) \leq \sum_{i=1}^{n} \mathsf{P}(X_i \geq \alpha) \leq \theta.$$

The proof is complete. □

The following result can be found in [14], Chapter 4.

LEMMA 3.2. *Consider a positive random variable Z such that $\|Z\|_q \leq C\|Z\|_p$ for $q > p > 0$. Then*

$$\mathsf{P}(Z > t) \leq (2C^p)^{q/(p-q)} \quad \Rightarrow \quad \|Z\|_p \leq 2^{1/p} t \quad \text{and} \quad \|Z\|_q \leq 2^{1/p} C t.$$

Putting together Lemmas 3.1 and 3.2 we get the following.

LEMMA 3.3. *Let $(X, X_1, \dots, X_n)$ be a sequence of positive i.i.d. random variables. Assume that there exists a constant c such that for $0 < p < q < \infty$,*

$$\left\| \sup_{1 \le i \le n} \|X_i\| \right\|_q \le c \left\| \sup_{1 \le i \le n} \|X_i\| \right\|_p .$$

Then, letting $\theta = (2c^p)^{q/(p-q)}$,

$$\mathsf{P}(X \ge t) \le \frac{\theta}{n} \quad \Rightarrow \quad \left\| \sup_{1 \le i \le n} X_i \right\|_p \le 2^{1/p} t.$$

For later reference, we also include the next lemma.

LEMMA 3.4. *Let $(X, X_1, \dots, X_n)$ be a sequence of positive i.i.d. random variables. Then*

$$\left\| \sup_{1 \le i \le n} X_i \right\|_p \le t \quad \Rightarrow \quad \mathsf{P}(X \ge 2^{1/p} t) \le \frac{1}{n}.$$

PROOF. Use Chebyshev's inequality and Lemma 3.1 with $\theta = 1$. □

3.1.2. *Main result.* Now we are ready to prove an extension of a result from [1] that deals with strict tail probability comparisons for pairs of random variables.

THEOREM 3.5. *Let $(X, X_1, \dots, X_n)$ and $(Y, Y_1, \dots, Y_n)$ be sequences of positive i.i.d. random variables. For some $0 < p < q$ and all positive integers n, assume that*

$$\left\| \sup_{1 \le i \le n} X_i \right\|_q \le c_1 \left\| \sup_{1 \le i \le n} X_i \right\|_p \tag{3.1}$$

and

$$\left\| \sup_{1 \le i \le n} Y_i \right\|_p \le c_2 \left\| \sup_{1 \le i \le n} X_i \right\|_p .$$

Then there exists c_3, depending only on p, q, c_1 and c_2, such that for all $t \ge 0$,

$$\mathsf{P}(Y \ge c_3 t) \le c_3 \mathsf{P}(X \ge t).$$

PROOF. Given an arbitrary $\alpha = \alpha_1 > 0$ with

$$\mathsf{P}(Y \ge \alpha_1) > 0, \tag{3.2}$$

choose μ to be the smallest positive integer satisfying

$$\frac{1}{2\mu} \le \mathsf{P}(Y \ge \alpha_1) \le \frac{1}{\mu}. \tag{3.3}$$

From Lemma 3.1 it follows that

$$\mathsf{P}\Big(\sup_{1\le j\le \mu} Y_j \ge \alpha_1\Big) \ge \tfrac{1}{3}.$$

Hence, by Chebyshev's inequality,

$$\frac{1}{\alpha_2^p}\Big\|\sup_{1\le j\le \mu} Y_j\Big\|_p^p \ge \frac{1}{3},$$

which, by assumption, yields

$$\frac{c_2^p}{\alpha_1^p}\Big\|\sup_{1\le j\le \mu} X_j\Big\|_p^p \ge \frac{1}{3},$$

and, consequently, for any $\alpha_2 > 0$,

$$\frac{1}{\alpha_2^p}\Big\|\sup_{1\le j\le \mu} X_j\Big\|_p^p \ge \frac{\alpha_1^p}{3\alpha_2^p c_2^p}.$$

In particular, if $\alpha_2^p = \alpha_1^p/(6c_2^p)$, we get from the latter inequality that

$$\frac{1}{\alpha_2^p}\Big\|\sup_{1\le j\le \mu} X_j\Big\|_p^p \ge 2.$$

Now, Lemma 2.3 implies that

$$\mathsf{P}\Big(X \ge \frac{\alpha_1}{(6^{1/p}c_2)}\Big) = \mathsf{P}(X \ge \alpha_2) \ge \big(2c_1^p\big)^{q/(p-q)}\frac{1}{\mu}.$$

Finally, (3.3) gives,

$$\mathsf{P}\Big(X \ge \frac{\alpha_1}{(6^{1/p}c_2)}\Big) \ge \big(2c_1^p\big)^{q/(p-q)}\mathsf{P}(Y \ge \alpha_1). \tag{3.4}$$

Note that (3.4) holds for all α_1 for which (3.2) holds. For any other $\alpha_1 > 0$, (3.4) holds trivially. □

3.1.3. *Contraction for multipliers.* Condition (3.1) yields an example of a class of random variables with the so-called Marcinkiewicz–Paley–Zygmund property (MPZ for short). The concept was studied in [10] and can be traced back to [22] and [16]. A family $\mathcal{Z} \subset L_+^q$ of random variables is said to be in the class MPZ(q) (short for *have MPZ*) if one of the following equivalent conditions is satisfied:

$$\exists\, p < q \text{ (equivalently } \forall\, q \le p), \qquad m_{q,p} \overset{\text{df}}{=} \sup_{Z\in\mathcal{Z}} \frac{\|Z\|_q}{\|Z\|_p} < \infty, \tag{3.5}$$

$$\exists\, \delta > 0, \qquad \inf_{Z\in\mathcal{Z}} \mathsf{P}\big(Z > \delta\|Z\|_q\big) > \delta. \tag{3.6}$$

That is, (3.1) involves $\mathbb{Z} = \{\sup_{1\le i\le n}\|X_i\|: n\in\mathbb{N}\}$. Also, in [10] it was shown that the space of diagonal-free random polynomials of finite degree, spanned by symmetric random variables with so-called semiregular distributions, has MPZ. A random variable ξ is said to have a *semiregular distribution* if its tail $G(t) = \mathsf{P}(|\xi| > t)$ satisfies the relation,

$$V(a) = \limsup_{t\to\infty} G(at)/G(t) < 1$$

for some (or all) $a > 1$ (by convention, $0/0 = 0$). For example, any bounded random variable has semiregular distribution. In particular, (the norm of) any normed space-valued Rademacher polynomial of degree d has MPZ with the constant $m_{qp} = [2(q-1)/(p-1)]^d$ ([10], Corollary 2.7].

Now the essence of Theorem 3.5 is that the continuity of a certain operator that is fulfilled once by means of L^p-norms, will be also fulfilled by means of probability tails. We will illustrate this concept by the following result.

THEOREM 3.6. *Let $f_{i_1,\dots,i_k}$ be a finitely supported, diagonal-free array, taking values in a Banach space $\mathbb{E}$. Let $\boldsymbol{\xi} = (\xi_1,\xi_2,\dots)$ be a sequence of symmetric independent random variables.*

(i) *Contraction inequality. There is a constant $c > 0$ such that*

$$\mathsf{P}\Big(\big\|(f,(\mathbf{s}\boldsymbol{\xi})^k)\big\| > ct\Big) \le c\mathsf{P}\big(\|Q(f,\boldsymbol{\xi})\| > t\big), \qquad t > 0,$$

where $\mathbf{s}\boldsymbol{\xi} = (s_i\xi_i)$ and $\|\mathbf{s}\|_\infty = \sup_i |s_i| \le 1$.

(ii) *Maximal inequality. There is a constant $C > 0$ such that*

$$\mathsf{P}\Big(\sup_{m_1,\dots,m_k}\|T_{m_1,\dots,m_k}Q(f,\boldsymbol{\xi}^k)\| > Ct\Big) \le C\mathsf{P}\big(\|Q(f,\boldsymbol{\xi}^k)\| > t\big), \qquad t > 0,$$

where

$$T_{m_1,\dots,m_k}Q(f;\mathbb{X}) = \sum_{i_1\le m_1,\dots,i_k\le m_k} f_{i_1,\dots,i_k}\xi_{1i_1}\cdots\xi_{ki_k}.$$

PROOF. Let us first prove (i). By virtue of the symmetry assumption and Fubini's theorem, it suffices to give the proof for the case when $\mathbb{X}$ is a matrix of Rademacher random variables. Let $Q_1,\dots,Q_n$ be independent copies of $Q(f,\boldsymbol{\xi})$. Then the vector $(Q_1,\dots,Q_n)$ is a Rademacher homogeneous polynomial of degree k taking values in $l_n^\infty(\mathbb{E})$. Similarly, let $R_1,\dots,R_n$ be independent copies of $Q(f,\mathbf{s}\boldsymbol{\xi})$. From the contraction principle for L^p-norms, which may be found in [10], Remark 2.9 (essentially, it is due to [11]), it follows that for all $p\ge 1$,

$$\Big\|\|(R_1,\dots,R_n)\|_{l_n^\infty(\mathbb{E})}\Big\|_p \le c\Big\|\|(Q_1,\dots,Q_n)\|_{l_n^\infty(\mathbb{E})}\Big\|_p.$$

From the observation that $\|(Q_1,\ldots,Q_n)\|_{l_n^\infty(\mathbb{E})} = \sup_{1\le i\le n}\|Q_i\|_{\mathbb{L}^r}$ and using the fact that Rademacher polynomials are MPZ, and also citing Theorem 3.6 above, the result follows. □

The proof of part (ii) is the same, using the corresponding result for L^p-norms of polynomials for the Rademacher random variables, which follows easily from [18] and Lévy's inequality.

REMARK 1. In fact, Theorem 3.6 is also valid for the sign randomized U-statistics as in (2.5). The proof is identical.

THEOREM 3.7 (Comparison inequality). *Let $(f_{i_1,\ldots,i_k})$ be a diagonal-free, finitely supported array. Let $\boldsymbol{\xi} = (\xi_i)$ and $\boldsymbol{\eta} = (\eta_i)$ be sequences of symmetric independent random variables such that, for some constant $A > 0$,*

$$\mathsf{P}\big(|\xi_i| > t\big) \le A\mathsf{P}\big(|\eta_i| > t\big), \qquad t > 0,\ i \in \mathbb{N}.$$

Then, for some constant $K = K(c, d, A)$,

$$\mathsf{P}\big(\|Q(f,\boldsymbol{\xi}^k)\| > t\big) \le K\mathsf{P}\big(K\|Q(f,\boldsymbol{\eta}^k)\| > t\big), \qquad t > 0.$$

PROOF. We have

$$Q(f,\boldsymbol{\xi}^k) \overset{\mathcal{D}}{=} Q\Big(f, \big(\varepsilon|\boldsymbol{\xi}|\big)^k\Big)$$

and

$$Q(f,\boldsymbol{\eta}^k) \overset{\mathcal{D}}{=} Q\Big(f, \big(\varepsilon|\boldsymbol{\eta}|\big)^k\Big),$$

where ε is a Rademacher sequence independent of $\boldsymbol{\xi}$ and $\boldsymbol{\eta}$.

If $A = 1$, then we may replace each $|\xi_i|$ and $|\eta_i|$ by their decreasing rearrangements $|\xi_i|^*$ and $|\eta_i|^*$, respectively. The assumption yields $|\xi_i|^* \le |\eta_i|^*$ a.s. Hence, by Theorem 3.6(i), the inequality follows.

Let $A > 1$. Then there exist a sequence $\boldsymbol{\alpha} = (\alpha_i)$ of i.i.d. random variables, independent of $\boldsymbol{\xi}$, such that $\mathsf{P}(\alpha_i = 1) = 1/K$ and $\mathsf{P}(\alpha_i = 0) = 1 - 1/K$, so that $\mathsf{P}(\alpha_i|\xi_i| > t) = \mathsf{P}(|\xi_i| > t)/K$. Therefore, by the first part of the proof,

$$c\mathsf{P}\Big(\Big\|Q\Big(f, \big(\varepsilon\boldsymbol{\alpha}|\boldsymbol{\xi}|\big)^k\Big)\Big\| > t\Big) \le \mathsf{P}\Big(K\Big\|Q\Big(f, \big(\varepsilon|\boldsymbol{\eta}|\big)^k\Big)\Big\| > t\Big), \qquad t > 0.$$

Conditioning on $\boldsymbol{\xi}$, it remains to prove that for every polynomial Q and every diagonal-free array $f_{i_1,\ldots,i_k}$,

$$\mathsf{P}\big(\|Q(f,\varepsilon^k)\| > mt\big) \le m\mathsf{P}\Big(\big\|Q\big(f,(\varepsilon\boldsymbol{\alpha})^k\big)\big\| > t\Big) \tag{3.7}$$

for some constant $m = m_k$. Let $\boldsymbol{\beta} = \varepsilon\boldsymbol{\alpha}$. Then it is clear that $\boldsymbol{\beta}$ is semiregular, as defined earlier, and hence homogeneous random polynomials of degree k over $\boldsymbol{\beta}$ have MPZ. Furthermore, the comparison inequality is true for L^p for $p \ge 1$ (see, e.g., [10], Theorem 2.13). Hence arguing as in the proof of Theorem 3.6, we obtain (3.7). □

3.2. *Decoupling for tails.* In order to prove any tail inequality of the type $\mathsf{P}(\xi > t) \le K\mathsf{P}(\eta > t)$, where ξ and η are real random variables, it is enough to prove it for an arbitrarily chosen conditional probability

$$\mathsf{P}[\xi > t \mid \mathcal{G}] \le K\,\mathsf{P}[\eta > t \mid \mathcal{G}].$$

This observation was used in proving the inequality (6.9.5) in [13]. Denote by $\mathcal{G}$ the σ-field spanned by all random variables of the form $\sum_{j=1}^{i} h(\xi_j)$ (in other words, by the random point measure

$$\sum_{j=1}^{i} \delta_{\xi_j}$$

on $(\mathbb{R}^{\mathbb{N}})^k$; cf. [13], page 182). Then $(\xi_1, \dots, \xi_k)$ is concentrated on a finite permutation invariant subset of $(\mathbb{R}^{\mathbb{N}})^k$. Now (2.1) can be rewritten as [recall the notation preceding (2.1)]

$$\mathsf{E}\Big[Q\big(f, \xi_{j_1}, \dots, \xi_{j_k}\big) \mid \mathcal{G}\Big] = k^{-k} Q\big(f, (\xi_1 + \cdots + \xi_k)^k\big), \tag{3.8}$$

THEOREM 3.8. *Let* $f, \xi, \xi_1, \dots, \xi_k$ *be as in Theorem* 2.1 (*but* we do not assume integrability).

(A″) *Let* $\xi, \xi_1, \xi_2, \dots$ *be independent and symmetric. Then there exists a constant* A'', *depending only on* k, *such that, for all* $t \ge 0$,

$$\mathsf{P}\big(\|Q(f, \xi^k)\| \ge A''t\big) \le A''\mathsf{P}\big(\|Q(f, \xi_1, \dots, \xi_k)\| \ge t\big).$$

(B″) *Let* $\xi, \xi_1, \dots, \xi_k$ *be interchangeable. Then there exists some constant* B'', *depending only on* k, *such that for all* $t \ge 0$,

$$\mathsf{P}\big(\|\widehat{Q}(f, \xi_1, \dots, \xi_k)\| \ge B''t\big) \le B''\mathsf{P}\big(\|Q(f, \xi_k)\| \ge t\big).$$

PROOF. (A″) By symmetry, using Theorem 3.6.(ii), with $\eta = \xi_1 + \cdots + \xi_k$ and $A = k$, we obtain that

$$\mathsf{P}\big(\|Q(f, \xi^k)\| \ge tK\big) \le K\,\mathsf{P}\Big[k^k \big\|Q\big(f, (\xi_1 + \cdots + \xi_k)^k\big)\big\| \ge tK\Big].$$

By (3.8) and inequality (6.9.5) in [13], the latter quantity can be estimated from below by

$$c_k K \mathsf{P}\big(k^{2k} k^k \|Q(f, \xi_1, \dots, \xi_k)\| \ge t\big),$$

which completes the proof of (A″).

(B″) By the Mazur–Orlicz polarization formula (1.2) and by (2.1), we obtain the estimates

$$\begin{aligned}
&\mathsf{P}(\|\widehat{Q}(f,\xi_1,\dots,\xi_k)\| \ge t)\\
&= \mathsf{P}\left(\left\|\frac{1}{k!}Q\left(f,\sum_{\delta}(-1)^{k-|\delta|}(\delta_1\xi_1+\cdots+\delta_k\xi_k)^k\right)\right\| \ge t\right)\\
&\le \sum_{\delta}\mathsf{P}\left(2^k\left\|\frac{1}{k!}Q(f,(\delta_1\xi_1+\cdots+\delta_k\xi_k)^k)\right\| \ge t\right)\\
&= \sum_{i=0}^{k}\binom{k}{i}\mathsf{P}\left(2^k\left\|\frac{1}{k!}Q(f,(\xi_1+\cdots+\xi_i)^k)\right\| \ge t\right).
\end{aligned}$$

By (3.8), with $j_1 = \cdots = j_k$ and the inequality (6.9.5) in [13], we estimate the preceding expression from above by

$$c_k^{-1}\sum_{i=0}^{k}\binom{k}{i}\mathsf{P}\left(\frac{(2i)^k}{k!}\|Q(f,\xi_1^k)\| \ge t\right) \le c_k^{-1}2^k\mathsf{P}\left(\frac{(2k)^k}{k!}\|Q(f,\xi_1^k)\| \ge t\right),$$

which completes the proof. □

REMARK 2. While the symmetry assumption is irrelevant in condition (B″) [or in (B), before], the symmetrization procedure used in the proof of (A) fails. The reason is that we use the conditioning on $\mathcal{G}$, which destroys the independence which is essential in applications of (2.3).

4. Notes.

1. The inverse estimate in (2.3) is not true, in general, even if $k = 1$. For example, let $\xi_1,\dots,\xi_n$ be Bernoulli random variables with $p = \mathsf{P}(\xi_1 = 1) = 1/2$ and $f(i) = 1$. Then $\mathsf{E}|f\bar{\xi}|^2 = n/4$ and $\mathsf{E}|f\xi|^2 = (n+n^2)/4$.
2. The symmetry of functions f is essential in Theorem 2.1(B) and its analogs, as was pointed out in [19]. Bourgain's counterexample, given there, involves $\mathbb{E} = l^2 \otimes l^2$ endowed with the projective norm $\|\boldsymbol{a}\| = \inf\{\sum_{i,j}\|a_i^1\| \times\|a_j^2\|\colon \boldsymbol{a} = \sum_{i,j}a_i^1 \otimes a_j^2\}$, Rademacher chaos and tetrahedral functions f. However, the inequalities (B) of both Theorems 2.1 and 2.3 hold for tetrahedral Rademacher chaos induced by ξ and $\mathbb{X}$ (with independent columns), whenever $\mathbb{E}$ is (a) a Banach lattices with no subspace isomorphic to c_0 or (b) a UMD space.
3. The full analog of Theorem 2.1 is valid in locally convex spaces.
4. The decoupling results from Section 2 can be carried over to linear spaces over the field of complex numbers. To obtain similar results for Section 3 is more difficult. One approach is to show that if ε denotes a sequence of independent Rademacher random variables and if σ denotes a sequence of independent Steinhaus random variables (that is, σ_i is uniformly distributed over the complex unit circle), then $\|Q(f;\varepsilon^k\| \approx \|Q(f;\sigma^k)\|$. We omit the details of the development.

5. In the case when the tail decoupling holds, that is, in Theorem 3.8, 2.3(A‴) and 2.3(B‴), we obtain the comparison of tightness. That is, for a family of functions $\{f: f \in F\}$, we have that if one type of chaos $\{Q_d(f): f \in F\}$ is tight, so is the other, $\{\tilde{Q}_d(f): f \in F\}$, subject to restrictions listed in the above theorems. That remark also applies to functions f taking values in a locally convex space.
6. In the context discussed above, we immediately obtain the comparison of generalized Orlicz modulars, that is, functionals of the form $\Phi(\cdot) = \mathsf{E}\,\phi(\|\cdot\|)$, where ϕ is a nondecreasing function on the positive half-line, $\phi(0) = 0$.
7. Multiple stochastic integrals of deterministic multivariate functions (cf. e.g., [9]) can be seen as limits of multilinear random forms. Therefore, if $\boldsymbol{\xi}, \boldsymbol{\xi}_1, \boldsymbol{\xi}_k$ are stochastic processes with independent increments, and the symbols $\langle f\boldsymbol{\xi}_1 \otimes \cdots \otimes \boldsymbol{\xi}_d\rangle$ and $\langle f\boldsymbol{\xi}^{\otimes k}\rangle$ are understood as such integrals, then all decoupling inequalities carry over word-for-word.
8. Our decoupling inequalities involve a certain means of domination. Essentially, we show that the domination by means of L^p-norms yields the same for probability tails. The passing from one to another type of domination may be of an intrinsic interest. Recall the definition of f^{**} mentioned in Section 2.2. Let us note the following result, which can be applied in a wider context than ours. Suppose that ξ and η are two given nonnegative random variables, and define quantities $c_1, \dots, c_5$:

   Let c_1 be the smallest constant such that for every Orlicz function, $\|\xi\|_\phi \le c_1\|\eta\|_\phi$.
   Let c_2 be the smallest constant such that for all $t > 0$, if $\phi_t(x) = (x-1)_+/t$ then $\|\xi\|_{\phi t} \le c_2\|\eta\|_{\phi t}$.
   Let c_3 be the smallest constant such that $\xi^{**} \le c_3\eta^{**}$.
   Let c_4 be the smallest constant such that for every r.i. norm, $\|\xi\| \le c_4\|\eta\|$.
   Let c_5 be the smallest constant such that for every separable r.i. norm, $\xi\| \le c_5\|\eta\|$.

   Then $c_1 = c_2 \le c_3 = c_4 = c_5 \le 2c_1$. Indeed, inequalities $c_2 \le c_1 \le c_4$ and $c_3 \le c_5 \le c_4$ are obvious. That $c_1 \le c_2$ follows immediately from the formula

$$\phi(x) = \int_0^\infty \phi_t(x)\,\bigl(\phi'(t)\bigr).$$

   That $c_4 \le c_3$ was proved in [15], Proposition 2.a.8. That $c_3 \le 2c_2$ follows from the formula $\|\xi\|_{\phi_t} \le \xi^{**}(t) \le 2\|\xi\|_{\phi_t}$. To show the left-hand side, suppose that $\xi^{**}(t) \le 1$. Then

$$\int_0^t \xi^*(s)\,ds \le t.$$

   Thus we have that $\xi^*(t) \le 1$ and hence

$$\mathsf{E}_{\phi_t}(\xi) = \frac{1}{t}\int_0^1 \bigl(\xi^*(s) - 1\bigr)_+ ds = \frac{1}{t}\int_0^t \bigl(\xi^*(s) - 1\bigr)_+ ds \le \frac{1}{t}\int_0^t \xi^*(s)\,ds \le 1.$$

To show the right-hand side, suppose that $\|\xi\|_{\phi_t} \le 1$. Thus

$$\int_0^{t_0} \left(\xi^*(s) - 1\right) ds \le t,$$

where $t_0 = \mathsf{P}(\xi > 1)$. If $t_0 \ge t$, then it follows that

$$\int_0^{t} \left(\xi^*(s) - 1\right) ds \le t,$$

whence it follows that

$$\int_0^{t} \xi^*(s)\, ds \le 2t.$$

If $t_0 < t$, then

$$\int_0^{t} \xi^*(s)\, ds = \int_0^{t_0} \left(\xi^*(s) - 1\right) ds + \int_0^{t_0} ds + \int_{t_0}^{t} \xi^*(s)\, ds \le 2t,$$

because $\xi^*(s) \le 1$ if $s > t_0$.

9. A decoupling principle for multivalued functions (proved in [4]) also follows from our basic decoupling inequalities. Suppose that $F(\cdot, \boldsymbol{\xi})$ is a countably multivalued function, that is, a countable family of functions $\mathcal{F}_{\mathbf{i}}$ is associated with each $\mathbf{i}$. In equivalent terms, one may think of a decision function $\tau\colon D \times \mathbb{N}^k \to \prod_{\mathbf{i} \in \mathbb{N}^k} \mathcal{F}_{\mathbf{i}}$ (D is countable). Then the statements of Theorem 2.3 hold uniformly with respect to τ, that is, the norm $\|F(\cdot)\|$ is replaced by $\sup_\tau \sup_d \|\tau(d, \cdot)(\cdot)\|$. The theorem follows for a finite collection of decision functions $\{\tau_1, \dots, \tau_n\}$, since this means the replacement of the underlying Banach space $\mathbb{E}$ by another Banach space $l_n^\infty(\mathbb{E})$. In the full statement we need the Banach lattice $\mathbb{L}(l^\infty)$ to satisfy the property $\sup_n \|x_n\| = \|\sup_n x_n\|$, *for an increasing sequence of nonnegative vectors.* In view of the preceding note, we may choose a family of Orlicz spaces, and the required property holds.

Other sequential functionals on $\mathbb{R}^D$, for example, l^p, Orlicz l^ψ and so forth, yield numerous variations of Theorem 2.3.

REMARK 3. This paper represents the combination of the papers [7] and [26].

REFERENCES

[1] ASMAR, N. H. and MONTGOMERY-SMITH, S. J. (1993). On the distribution of Sidon series. *Ark. Mat.* **31** 13–26.

[2] BURKHOLDER, D. L. (1986). *Martingales and Fourier Analysis in Banach Space. Lecture Notes in Math.* **1206** 61–108. Springer, Berlin.

[3] DE ACOSTA, A. (1987). A decoupling inequality for multilinear forms of stable vectors. *Probab. Math. Statist.* **8** 71–76.

[4] DE LA PEÑA, V. H. (1992). Decoupling and Khintchine's inequalities for U-statistics. *Ann. Probab.* **20** 1877–1892.

[5] DE LA PEÑA, V. H. (1993). Inequalities for tails of adapted process with an application to Wald's lemma. *J. Theoret. Probab.* **6**.

[6] DE LA PEÑA, V. H. (1994). A bound on the moment generating function of a sum of dependent variables with an application to simple random sampling without replacement. *Ann. Inst. H. Poincaré.* To appear.

[7] DE LA PEÑA, V. H. and MONTGOMERY-SMITH, S. J. (1992). Decoupling inequalities for tail probabilities of multilinear forms in symmetric and hypercontractive variables. Preprint.

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

[9] KALLENBERG, O. and SZULGA, J. (1989). Multiple integration with respect to Poisson and Lévy processes. *Probab. Theory Related Fields* **83** 101–134.

[10] KRAKOWIAK, W. and SZULGA, J. (1988). Hypercontraction principle and random multilinear forms in Banach spaces. *Probab. Theory Related Fields* **77** 325–342.

[11] KWAPIEŃ, S. (1987). Decoupling inequalities and polynomial chaos. *Ann. Probab.* **15** 1062–1071.

[12] KWAPIEŃ, S. and WOYCZYŃSKI, W. A. (1987). Double stochastic integrals, random quadratic forms and random series in Orlicz spaces. *Ann. Probab.* **15** 1072–1096.

[13] KWAPIEŃ, S. and WOYCZYŃSKI, W. A. (1992). *Random Series and Stochastic Integrals.* Birkhäuser, Boston.

[14] LEDOUX, M. and TALAGRAND, M. (1991). *Probability in Banach Spaces.* Springer, Berlin.

[15] LINDENSTRAUSS, J. and TZAFRIRI, L. (1979). *Classical Banach spaces II. Function Spaces.* Springer, Berlin.

[16] MARCINKIEWICZ, J. and ZYGMUND, A. (1937). Sur les fonctions independantes. *Fund. Math.* **29** 60–90.

[17] MAZUR, S. and ORLICZ, W. (1935). Grundlegende Eigenschaften der polynomischen Operationen. *Studia Math.* **5** 50–68, 179–189.

[18] MCCONNELL, T. R. and TAQQU, M. S. (1986). Decoupling inequalities for multilinear forms in independent symmetric random variables. *Ann. Probab.* **14** 943–954.

[19] MCCONNELL, T. and TAQQU, M. S. (1987). Decoupling of Banach-valued multilinear forms in independent symmetric Banach-valued random variables. *Probab. Theory Related Fields* **75** 499–507.

[20] MCCONNELL, T. and TAQQU, M. (1986). Dyadic approximation of double integrals with respect to symmetric stable processes. *Stochastic Processes. Appl.* **22** 323–331.

[21] NOLAN, D. and POLLARD, D. (1987). U-processes: Rates and convergence. *Ann. Statist.* **15** 780–799.

[22] PALEY, R. E. A. C. and ZYGMUND, A. (1932). A note on analytic functions on the circle. *Proc. Cambridge Philos. Soc.* **28** 266–272.

[23] ROSIŃSKI, J. and WOYCZYŃSKI, W. A. (1986). On Itô stochastic integration with respect to p-stable motion: Inner clock, integrability of sample paths, double and multiple integrals. *Ann. Probab.* **14** 271–286.

[24] ROSIŃSKI, J. SAMORODNITSKY, G. and TAQQU, M. S. (1991). Sample path properties of stochastic processes respresented as multiple stable integrals. *J. Multivariate Anal.* **37** 115–134.

[25] SZULGA, J. (1991). Limit theorems for some randomized nonlinear functionals of empirical measures. Preprint, Auburn Univ.

[26] SZULGA, J. (1992). Robust decoupling of homogeneous random chaoses. Preprint, Auburn Univ.

[27] WIENER, N. (1930). The homogeneous chaos. *Amer. J. Math.* **60** 897–936.

[28] WIENER, W. and WINTNER, A. (1943). The discrete chaos. *Amer. J. Math.* **65** 279–298.

[29] ZINN, J. (1986). Comparison of martingale difference sequences. *Probability on Banach Spaces. Lecture Notes in Math.* **1153** 453–451. Springer, Berlin.

Victor H. de la Peña
Department of Statistics
Columbia University
New York, New York 10027

Stephen J. Montgomery-Smith
Department of Mathematics
University of Missouri
Columbia, Missouri 65211

Jerzy Szulga
Department of Mathematics
Auburn University
Auburn, Alabama 36849-3501