Remarks on Decoupling Inequalities for Random Multi-linear Forms

Terry R. McConnell
Syracuse University

August 24, 2015

Abstract

We discuss decoupling inequalities for multi-linear forms in non-negative and symmetric random variables

1 Introduction

Let $Q(\mathbf{X}, \mathbf{X})$ be a quadratic form in $\mathbf{X}$, a finite dimensional random vector. It is often quite useful to compare the moments of such forms with those of bilinear forms $Q(\mathbf{X}, \mathbf{Y})$, in which the random vector $\mathbf{Y}$ is independent of $\mathbf{X}$ and both have the same joint distribution:

$$c_p \|Q(\mathbf{X}, \mathbf{Y})\|_p \leq \|Q(\mathbf{X}, \mathbf{X})\|_p \leq C_p \|Q(\mathbf{X}, \mathbf{Y})\|_p \tag{1.1}$$

The right-hand inequalities were termed *decoupling inequalities* by the author and M.S. Taqqu, and they were used as tools in some unpublished work on double stochastic integration with respect to symmetric stable processes. (See [7] for relevant references. The roots of the decoupling idea can be traced to Lemma 1 of [1].) The name seemed apt in that a quantity is estimated by another in which there is less dependence. The result seemed sufficiently useful to warrant a publication [6] of its own.

The left-hand inequalities were not treated in [6], and seem to have first appeared in a paper [5] of S. Kwapień, a paper that contains many other related results. (Kwapień too generously gave the authors of [6] credit for both sides.) The left-hand inequalities might more properly be termed re-coupling inequalities, but the term decoupling inequalities stuck, and is commonly used today to refer to two-sided inequalities like (1.1).

The author believes the left-hand inequalities to be the harder ones to prove. (It is worth mentioning the short proof of the right-hand inequality in more general form, given recently by R. Vershynin[9].) See the monograph [2] for many extensions, related results, and applications of decoupling inequalities. Many (perhaps all) of our results can be obtained by using Theorem 3.4.1 of [2], though our proofs may be of independent interest and provide better, or at least more explicit, values for constants.

To be more precise about terminology and notation, let $Q : \mathbb{R}^N \times \mathbb{R}^N \to \mathbb{R}$ be a bilinear function. Let $a_{ij} = Q(\mathbf{e}_i, \mathbf{e}_j)$, where $\mathbf{e}_i$ are the standard basis vectors of $\mathbb{R}^N$, be the *coefficients* of the bilinear form Q. If $\mathbf{X} = (X_1, X_2, \ldots, X_N)$ is a random vector with component random variables $X_i, i = 1, 2, \ldots, N$, then we may write

$$Q(\mathbf{X}, \mathbf{X}) = \sum_{i,j=1}^{N} a_{ij} X_i X_j,$$

an expression we shall call the *expansion* of $Q(\mathbf{X}, \mathbf{X})$. We shall also consider multi-linear functions, Q, of d variables, and their expansions

$$Q(\mathbf{X}, \ldots, \mathbf{X}) = \sum_{\mathbf{i}} a_{\mathbf{i}} X_{i_1} X_{i_2} \ldots X_{i_d},$$

where the summation is over multi-indices $\mathbf{i} = (i_1, i_2, \ldots, i_d)$ of integers in the set $\{1, 2, \ldots, N\}$. In what follows, we shall usually assume that the forms vanish "on all diagonals", i.e., that the coefficients vanish whenever two or more components of their multi-indices agree. The assumption about diagonals is quite essential, as diagonal terms involve powers of individual random variables and must often be treated separately. Multi-linear forms in infinitely many variables also arise in applications, but we have chosen to stick with the case of a finite N to avoid issues of convergence. Thus, applications of decoupling inequalities sometimes require an additional limiting argument as N tends to infinity.

In this paper we shall give complete proofs of inequalities (1.1) when $0 < p < \infty$ and when component random variables are independent and symmetric. When the random variables are symmetric, Khintchine-type inequalities can be used to reduce to the case of non-negative coefficients and random variables. Accordingly, in the following section we begin with the non-negative case, where we also obtain all inequalities for the full range of p. If there is any novelty in our approach, it is that we use Chebyshev's monotone function inequality[1] to obtain some of the inequalities in the non-negative case.

In section 3 we consider decoupling inequalities for multi-linear forms when the random variables are symmetric.

In what follows we shall omit explicit ranges of summation, it being understood that sums are to be taken over all possible values of the index or multi-index indicated. If $\mathcal{F}$ denotes a sigma field, then $P^{\mathcal{F}}$ and $E^{\mathcal{F}}$ denote, respectively, the conditional probability measure and expectation operator given that sigma field.

2 Decoupling for Non-negative Random Variables

We consider here results in which all random variables are non-negative and all coefficients of the multi-linear form are also non-negative.

A smooth function F of two variables is termed *2-monotone* non-decreasing (non-increasing) if the partial derivative F_{xy} is non-negative (non-positive). Such functions are monotone in the same 2-dimensional sense as joint distribution functions of pairs of random variables. (This type of monotonicity can be generalized to functions of n variables, but we shall only need the 2 variable case here.)

If F is smooth and 2-monotone non-decreasing on the closure of the first quadrant, $\mathbb{R}^2_+$, then it enjoys the representation

$$F(x,y) = f(x) + g(y) + \iint_{\mathbb{R}^2_+} 1_{[a,\infty)}(x) 1_{[b,\infty)}(y)\, \mu(dadb), \tag{2.1}$$

where f and g are continuous functions of one variable and μ is a positive measure. This can easily be shown using the fundamental theorem of calculus, and in this simple setting the measure μ is absolutely continuous with density F_{xy}, and $f(x) = F(x,0), g(y) = F(0,y) - F(0,0)$. (In the case of a 2-monotone non-increasing function there is a minus sign in front of the integral above.)

The elementary inequalities in the following two lemmas are key to our approach. They will be applied iteratively, an unbounded number of times, and so placement of constants is important, as are the values of constants.

Lemma 2.1. *Let A, B, C be non-negative constants. Let X be a non-negative random variable and Y an independent copy of X. Then*

$$E(A + BX + CY)^p \le E(A + 2^{\frac{1}{p}-1}BX + 2^{\frac{1}{p}-1}CX)^p, 0 < p \le 1, \tag{2.2}$$

and

$$E(A + 2^{\frac{1}{p}-1}BX + 2^{\frac{1}{p}-1}CX)^p \le E(A + BX + CY)^p, 1 \le p < \infty. \tag{2.3}$$

[1] Sometimes called "Chebyshev's other inequality"

In order to prove (2.2), it suffices to prove

$$E(A+BX+BY)^p \le E(A+2^{1/p}BX)^p, 0<p\le 1. \tag{2.4}$$

Indeed, by equal distribution and concavity we have

$$E(A+BX+CY)^p = \frac{1}{2}E(A+BX+CY)^p + \frac{1}{2}E(A+CX+BY)^p \le E\left(A+\left(\frac{B+C}{2}\right)X+\left(\frac{B+C}{2}\right)Y\right)^p,$$

and then we obtain (2.2) by using (2.4) with B replaced by $\frac{(B+C)}{2}$.

To prove (2.4), first note that one has

$$(A+Bx+By)^p-(A+Bx)^p-(A+By)^p+A^p\le 0,$$

since the function $F(x,y)=(A+Bx+By)^p$ is 2-monotone non-increasing. (Alternatively, since the left-hand side vanishes when $B=0$ and its derivative with respect to B is everywhere non-positive.) Thus, the left hand side of (2.4) is bounded above by

$$E\left[(A+BX)^p+(A+BY)^p\right]-A^p = 2E(A+BX)^p-A^p.$$

But the last written quantity is in turn bounded above by $E(A+2^{1/p}BX)^p$, since the two are equal when $B=0$, and the derivative of the first with respect to B is bounded above by the derivative of the second with respect to B.

For the proof of (2.3), we first note

$$2E(A+BX)^p-A^p\ge E(A+2^{1/p}BX)^p. \tag{2.5}$$

To see this, compare the derivatives of the two sides with respect to B. Next, since the function $F(x,y)=(A+Bx+Cy)^p$ is 2-monotone non-decreasing, we have

$$E(A+BX+CY)^p\ge E(A+BX)^p+E(A+CY)^p-A^p.$$

But then, by equal distribution and convexity, the latter expression is bounded below by

$$2E\left(A+\left(\frac{B+C}{2}\right)X\right)^p-A^p,$$

and we finish by using (2.5) with B replaced by $\frac{(B+C)}{2}$.

We remark that Lemma 2.1 holds if we only assume X and Y are exchangeable.

Lemma 2.2. *With the same notation as in Lemma 2.1, we have*

$$E(A+BX+CY)^p\le E(A+BX+CX)^p, 1\le p<\infty, \tag{2.6}$$

and

$$E(A+BX+CX)^p\le E(A+BX+CY)^p, 0<p\le 1. \tag{2.7}$$

For the proof of (2.6), note that the function $F(x,y)=(A+Bx+Cy)^p$ is 2-monotone non-decreasing. Thus, we have by (2.1) and Fubini's Theorem,

$$E(A+BX+CX)^p = Ef(X)+Eg(X)+\iint_{\mathbb{R}_+^2} E(1_{[a,\infty)}(X)1_{[b,\infty)}(X))\,\mu(dadb).$$

By Chebyshev's monotone function inequality (see, e.g., [4], section 2.17,)

$$E(1_{[a,\infty)}(X)1_{[b,\infty)}(X))\ge E(1_{[a,\infty)}(X))E(1_{[b,\infty)}(X)) = E(1_{[a,\infty)}(X))E(1_{[b,\infty)}(Y)). \tag{2.8}$$

Substituting this inequality and $Eg(X) = Eg(Y)$ above leaves us with

$$E(A + BX + CX)^p \geq E(A + BX + CY)^p.$$

The proof of (2.7) is similar, only in this range of p the function F is 2-monotone non-increasing.

Like Lemma 2.1, Lemma 2.2 does not really require independence of X and Y. For (2.6) it is sufficient that X and Y be identically distributed and negatively correlated in the sense that

$$P(X \geq a, Y \geq b) \leq P(X \geq a)P(Y \geq b),$$

and for (2.7) it is sufficient that X and Y be identically distributed and positively correlated in the sense that

$$P(X \geq a, Y \geq b) \geq P(X \geq a)P(Y \geq b).$$

Theorem 2.1. *Let $\mathbf{X} = (X_1, X_2, \ldots, X_N)$ be a random vector whose components are independent non-negative random variables. Let $\mathbf{Y}^j, j = 1, 2, \ldots, d$, be independent copies of $\mathbf{X}$. Assume all coefficients of the multi-linear form Q are also non-negative and that Q vanishes on all diagonals. Then we have*

$$EQ(\mathbf{Y}^1, \ldots, \mathbf{Y}^d)^p \leq EQ(\mathbf{X}, \ldots, \mathbf{X})^p,\ 1 \leq p < \infty, \tag{2.9}$$

$$2^{-d(d-1)(1-p)} EQ(\mathbf{Y}^1, \ldots, \mathbf{Y}^d)^p \leq EQ(\mathbf{X}, \ldots, \mathbf{X})^p,\ 0 < p \leq 1, \tag{2.10}$$

$$EQ(\mathbf{X}, \ldots, \mathbf{X})^p \leq EQ(\mathbf{Y}^1, \ldots, \mathbf{Y}^d)^p,\ 0 < p \leq 1, \tag{2.11}$$

and,

$$EQ(\mathbf{X}, \ldots, \mathbf{X})^p \leq 2^{d(d-1)(p-1)} EQ(\mathbf{Y}^1, \ldots, \mathbf{Y}^d)^p,\ 1 \leq p < \infty. \tag{2.12}$$

The proofs of these inequalities are similar, so we shall only give the proof of (2.10) in detail. It is based on (2.2), or more precisely, on the following consequence of (2.2):

$$E(A + 2^{1-1/p}BX + 2^{1-1/p}CY)^p \leq E(A + BX + CX)^p, 0 < p \leq 1. \tag{2.13}$$

The idea is to replace each component random variable in each of the slots of Q by the corresponding random variable from an independent copy. Let us begin with the last component X_N. Group the terms in the expansion of $Q(\mathbf{X}, \ldots, \mathbf{X})$ as

$$A + \sum_{j=1}^{d} B_j X_N = A + BX_N + CX_N,$$

where B_j includes all terms with the jth component of the multi-index equal to N, A includes all terms with no component of the multi-index equal to N, $B = \sum_{j=1}^{d-1} B_j$, and $C = B_d$.

Let $\mathcal{F}$ be the sigma field generated by all components of $\mathbf{X}$ except X_N. Then by (2.13) we have

$$E^{\mathcal{F}}(A + BX_N + CX_N)^p \geq E^{\mathcal{F}}(A + 2^{1-1/p}BX_N + 2^{1-1/p}CY_N^d)^p,$$

hence, taking expectation of both sides,

$$E(A + BX_N + CX_N)^p \geq E(A + 2^{1-1/p}BX_N + 2^{1-1/p}CY_N^d)^p.$$

Next, we incorporate the term $2^{1-1/p}CY_N^d$ into A, split BX_N as $BX_N = (\sum_{j=1}^{d-2} B_j)X_N + B_{d-1}X_N$, and repeat the whole argument so as to trade the last X_N for Y_N^{d-1}. (This time, the random variable Y_N^d is included among the generators of $\mathcal{F}$.) Continuing in this way, we reach

$$E(A + BX_N + CX_N)^p \geq E(A + 2^{(d-1)(p-1)/p} \sum_{j=1}^{d} B_j Y_N^j)^p,$$

after $d-1$ steps in total.

The random variable X_N has now been completely eliminated, and we continue the argument with X_{N-1} playing the role of X_N, continuing until all X_j have been replaced by components of the vectors $\mathbf{Y}^i$. In the course of the argument, each term in the expansion of Q is multiplied by the factor $2^{(d-1)(p-1)/p}$ exactly d times.

The other inequalities are derived in a similar way, using the other parts of Lemmas 2.1 and 2.2.

The requirement of independence among the components, and between random vectors, in Theorem 2.1 can be relaxed as follows: For a given index i, let $\mathcal{F}$ be the sigma field generated by all components of $\mathbf{X}$ and $\mathbf{Y}^j, j = 1, 2, \ldots, d$ except for the i-th components. The proof of each part of Theorem 2.1 may proceed under a suitable pair of the following assumptions, each to hold for any given i:

$$\text{Conditional on } \mathcal{F}, X_i, Y_i^j, j = 1, 2, \ldots, d \text{ are identically distributed} \tag{2.14}$$

$$\text{Conditional on } \mathcal{F}, (X_i, Y_i^j) \text{ are each exchangeable, } j = 1, 2, \ldots, d \tag{2.15}$$

$$P^{\mathcal{F}}(X_i > a, Y_i^j > b) \le P^{\mathcal{F}}(X_i > a) P^{\mathcal{F}}(Y_i^j > b), j = 1, 2, \ldots, d, a \ge 0, b \ge 0 \tag{2.16}$$

$$P^{\mathcal{F}}(X_i > a, Y_i^j > b) \ge P^{\mathcal{F}}(X_i > a) P^{\mathcal{F}}(Y_i^j > b), j = 1, 2, \ldots, d, a \ge 0, b \ge 0 \tag{2.17}$$

The ones with unit constants, (2.9) and (2.11), require (2.14) and either (2.16) (for (2.9)), or (2.17) (for (2.11).) The ones with non-unit constants, (2.10) and (2.12), require (2.15) and either (2.16) (for (2.10),) or (2.17) (for (2.12).) In the proofs, one applies the appropriate parts of Lemmas 2.1 or 2.2, with expectation E replaced by conditional expectation $E^{\mathcal{F}}$.

Moreover, in the unit constant inequalities, we may replace the p-th power function by a non-negative non-decreasing function ϕ having the property that $\phi(A+Bx+Cy)$ is 2-monotone non-decreasing (for (2.11)) or 2-monotone non-increasing (for (2.9)), for each possible choice of non-negative constants A, B, C.

For example, we have that

$$Ee^{\lambda Q(\mathbf{Y}^1, \ldots, \mathbf{Y}^d)} \le Ee^{\lambda Q(\mathbf{X}, \ldots, \mathbf{X})}, \ \lambda > 0, \tag{2.18}$$

provided (2.14) and (2.17) hold. If components are assumed independent, then (2.18) holds assuming each pair X_i, Y_i^j is identically distributed and positively tail correlated, i.e., (2.17) holds without the $\mathcal{F}$.

3 Decoupling for Symmetric Random Variables

Let $\epsilon_1, \epsilon_2, \ldots$ be i.i.d. symmetric Bernoulli random variables. Denote by ϵ the entire sequence, and by $\epsilon^j, j = 1, 2, \ldots$ independent copies of ϵ. Let Q be a multi-linear form as in the previous section, except we no longer assume coefficients to be non-negative. To avoid issues of convergence we shall assume that all but finitely many of the coefficients of Q vanish. The following collection of inequalities is well-known:

$$\alpha_{p,d} \|Q(\epsilon, \ldots, \epsilon)\|_2 \le \ \|Q(\epsilon, \ldots, \epsilon)\|_p \ \le \ \beta_{p,d} \|Q(\epsilon, \ldots, \epsilon)\|_2, \ 0 < p < \infty, \tag{3.1}$$

and

$$\alpha_{p,d} \|Q(\epsilon^1, \ldots, \epsilon^d)\|_2 \ \le \ \|Q(\epsilon^1, \ldots, \epsilon^d)\|_p \ \le \ \beta_{p,d} \|Q(\epsilon^1, \ldots, \epsilon^d)\|_2, \ 0 < p < \infty. \tag{3.2}$$

The form Q is said to be *symmetric* if $a_{\mathbf{i}} = a_{\mathbf{j}}$ whenever some permutation of $\mathbf{j}$ is equal to $\mathbf{i}$. If Q vanishes on diagonals and is symmetric, then we have

$$\|Q(\epsilon, \dots, \epsilon)\|_2 = \left(\sum_{\mathbf{i}} a_{\mathbf{i}}^2\right)^{1/2}, \tag{3.3}$$

since all terms in the expansion of Q over multi-indices with strictly increasing components are orthogonal. The identity

$$\|Q(\epsilon^1, \dots, \epsilon^d)\|_2 = \left(\sum_{\mathbf{i}} a_{\mathbf{i}}^2\right)^{1/2}$$

holds in general. Since (3.1) and (3.2) will be used to derive decoupling inequalities, and since the constants in the latter depend on the constants in the former, it pays to be somewhat fussy about the values of these constants. In case $d = 1$, inequalities (3.1) are the famous *Khintchine Inequalities*, and in this case the best possible constants were found by Uffe Haagerup[3]. There are two trivial cases: We have $\alpha_{p,d} = 1$ for $p \geq 2$ and $\beta_{p,d} = 1$ for $p \leq 2$. Haagerup proved that

$$\beta_{p,1}^p = 2^{p/2} \frac{\Gamma\left(\frac{p+1}{2}\right)}{\sqrt{\pi}}, \; p \geq 2, \tag{3.4}$$

the p-th moment of the standard normal distribution, is the best constant. He also proved that $\alpha_{p,1}^p$ is given by the same formula in a certain range of p, $p_0 < p < 2$, and by $2^{(p-2)/2}$ in the complementary range $0 < p \leq p_0$. (The latter expression is the p-th moment of a linear combination of ϵ_1 and ϵ_2 that has unit variance.) The changeover occurs at the value p_0, approximately 1.84742, where the two expressions become equal.

For $d > 1$, there is an argument based on Minkowski's inequality showing that $\beta_{p,d} = \beta_{p,1}^d$. See, e.g., pp 277-8 in the appendix of Stein's monograph [8]. This is the best constant, as can be seen by considering Q such that the center member of (3.1) factors into a product of d random variables. Explicit (though not optimal) values for the $\alpha_{p,d}$ can be found by a simple duality argument:

$$\alpha_{p,d} = \beta_{4-p,d}^{1-4/p} = \beta_{4-p,1}^{d(1-4/p)}, \; 0 < p < 2.$$

(The duality argument given by Stein on p.278 of [8] appears to work only for $1 < p < 2$, but can be adapted to the full range $0 < p < 2$ by using the Cauchy-Schwartz inequality in place of Hölder's inequality.)

Inequalities (3.2) are a special case of inequalities (3.1).

Theorem 3.1. *Let $\mathbf{X} = (X_1, X_2, \dots, X_N)$ be a random vector whose components are independent symmetric random variables. Let $\mathbf{Y}^j, j = 1, 2, \dots, d$, be independent copies of $\mathbf{X}$. Assume that the multi-linear form Q vanishes on all diagonals. Then we have*

$$\alpha_{p,d}^p EQ(\mathbf{X}, \dots, \mathbf{X})^p \leq EQ(\mathbf{Y}^1, \dots, \mathbf{Y}^d)^p, \; 0 < p \leq 2, \tag{3.5}$$

and,

$$EQ(\mathbf{X}, \dots, \mathbf{X})^p \leq 2^{d(d-1)(p/2-1)} \beta_{p,d}^p EQ(\mathbf{Y}^1, \dots, \mathbf{Y}^d)^p, \; 2 \leq p < \infty. \tag{3.6}$$

If Q is also symmetric, then we have

$$EQ(\mathbf{Y}^1, \dots, \mathbf{Y}^d)^p \leq \beta_{p,d}^p EQ(\mathbf{X}, \dots, \mathbf{X})^p, \; 2 \leq p < \infty, \tag{3.7}$$

and,

$$2^{-d(d-1)(1-p/2)} \alpha_{p,d}^p EQ(\mathbf{Y}^1, \dots, \mathbf{Y}^d)^p \leq EQ(\mathbf{X}, \dots, \mathbf{X})^p, \; 0 < p \leq 2. \tag{3.8}$$

We prove only (3.8) since the proofs of all statements are similar. Let $\mathcal{F}$ be the sigma field generated by $\{X_i, Y_i^j, i = 1, 2, \ldots, j = 1, 2, \ldots, d\}$ and assume ϵ is independent of $\mathcal{F}$. Denote by $\epsilon\mathbf{X}$ the random vector whose components are $\epsilon_i X_i, i = 1, 2, \ldots$. Similarly define $\epsilon^j \mathbf{Y}^j$. By the symmetry and independence assumptions, the joint distribution of $(\mathbf{X}, \mathbf{Y}^1, \ldots, \mathbf{Y}^d)$ is the same as that of $(\epsilon\mathbf{X}, \epsilon^1\mathbf{Y}^1, \ldots, \epsilon^d\mathbf{Y}^d)$. Then by the right-hand inequality in (3.2),

$$E^{\mathcal{F}} \left| \sum_{\mathbf{i}} a_{\mathbf{i}} \epsilon^1_{i_1} \ldots \epsilon^d_{i_d} Y^1_{i_1} \ldots Y^d_{i_d} \right|^p \leq \beta^p_{p,d} E^{\mathcal{F}} \left(\sum_{\mathbf{i}} a^2_{\mathbf{i}} (Y^1_{i_1} \ldots Y^d_{i_d})^2 \right)^{p/2} \tag{3.9}$$

The expected value of the left-hand side of (3.9) is $\|Q(\epsilon^1\mathbf{Y}^1, \ldots, \epsilon^d\mathbf{Y}^d)\|_p^p = \|Q(\mathbf{Y}^1, \ldots, \mathbf{Y}^d)\|_p^p$. In turn, using the case $p/2$ of (2.9), the expected value of the right-hand side of (3.9) is bounded above by $\beta^p_{p,d}\|Q(\mathbf{X}, \ldots, \mathbf{X})\|_p^p$, and the proof is complete.

References

[1] D.L. Burkholder, A geometric condition that implies the existence of certain singular integrals of Banach-space-valued functions, *Conference on Harmonic Analysis in Honor of Antoni Zygmund,* University of Chicago, 1981, Wadsworth International Group, Belmont, California, **1**(1983), 270-286.

[2] V. de la Peña and E. Giné, *Decoupling: From Dependence to Independence*, Springer, New York, 1999.

[3] U. Haagerup, The Best Constants in the Khintchine Inequality, *Studia Math.* **70**(1982), 231-283.

[4] G.H. Hardy, J.E. Littlewood, and Pólya, *Inequalities*, Cambridge University Press, Cambridge, 1934.

[5] S. Kwapień, Decoupling Inequalities for Polynomial Chaos, *Ann. Probab.***15** (1987), 1062-1071.

[6] T.R. McConnell and M.S. Taqqu, Decoupling inequalities for multilinear forms in independent symmetric random variables, *Ann. Probab.***14** (1986), 943-954.

[7] T.R. McConnell and M.S. Taqqu, Decoupling of Banach valued multilinear forms in independent symmetric Banach valued random variables, *Probab. Theory Rel. Fields***75** (1987), 499-507.

[8] E.M. Stein, *Singular Integrals and Differentiability Properties of Functions*, Princeton University Press, Princeton, N.J., 1970.

[9] R. Vershynin, A Simple Decoupling Inequality in Probability Theory, preprint, 2011.