Electron. Commun. Probab. **20** (2015), no. 72, 1–13.
DOI: 10.1214/ECP.v20-3829
ISSN: 1083-589X

ELECTRONIC COMMUNICATIONS in PROBABILITY

A note on the Hanson-Wright inequality for random vectors with dependencies*

Radosław Adamczak†

Abstract

We prove that quadratic forms in isotropic random vectors X in $\mathbb{R}^n$, possessing the convex concentration property with constant K, satisfy the Hanson-Wright inequality with constant CK, where C is an absolute constant, thus eliminating the logarithmic (in the dimension) factors in a recent estimate by Vu and Wang. We also show that the concentration inequality for all Lipschitz functions implies a uniform version of the Hanson-Wright inequality for suprema of quadratic forms (in the spirit of the inequalities by Borell, Arcones-Giné and Ledoux-Talagrand). Previous results of this type relied on stronger isoperimetric properties of X and in some cases provided an upper bound on the deviations rather than a concentration inequality.

In the last part of the paper we show that the uniform version of the Hanson-Wright inequality for Gaussian vectors can be used to recover a recent concentration inequality for empirical estimators of the covariance operator of B-valued Gaussian variables, due to Koltchinskii and Lounici.

Keywords: Hanson-Wright inequality; quadratic forms; concentration of measure; empirical covariance operator.
AMS MSC 2010: 60E15; 60B11.
Submitted to ECP on September 30, 2014, final version accepted on September 30, 2015.

1 Introduction

The Hanson-Wright inequality asserts that if $X_1, \ldots, X_n$ are independent mean zero, variance one random variables with sub-Gaussian tail decay, i.e. such that for all $t > 0$,

$$\mathbb{P}(|X_i| \geq t) \leq 2\exp(-t^2/K^2),$$

and $A = [a_{ij}]_{i,j=1}^n$ is an $n \times n$ matrix, then the quadratic form

$$Z = \sum_{i,j=1}^n a_{ij} X_i X_j$$

satisfies the inequality

$$\mathbb{P}(|Z - \mathrm{tr}A| \geq t) \leq 2\exp\Big(- \min\Big(\frac{t^2}{CK^4\|A\|_{HS}^2}, \frac{t}{CK^2\|A\|}\Big)\Big)$$

*Research partially supported by the NCN grant no. 2012/05/B/ST1/00412
†Institute of Mathematics, University of Warsaw, Poland. E-mail: R.Adamczak@mimuw.edu.pl

for all $t > 0$, where C is a universal constant. Here and in what follows $\|A\|_{HS} = (\sum_{i,j\le n} a_{ij}^2)^{1/2}$ is the Hilbert-Schmidt norm of A, whereas $\|A\| = \sup_{|x|\le 1} |Ax|$ is the operator norm of A ($|\cdot|$ denotes the standard Euclidean norm in $\mathbb{R}^n$). Actually, Hanson and Wright [12] proved a somewhat weaker inequality in which $\|A\|$ was replaced by the operator norm of the matrix $\tilde{A} = [|a_{ij}|]_{i,j=1}^n$. The original argument worked also only for symmetric random variables, the general mean zero case was proved by Wright in [33]. The above version with the operator norm of A appeared in many works under different sets of assumptions. For Gaussian variables it follows from estimates for general Banach space valued polynomials by Borell [8] and Arcones-Giné [4]. Independent proofs were also provided by Ledoux-Talagrand [21] and Latała [16, 17]. It is also well known that the general case can be reduced to the Gaussian one by comparison of moments or a decoupling and contraction approach [18, 5, 2, 26]. As observed by Latała [16], in the Gaussian case the Hanson-Wright inequality can be reversed (up to universal constants). Latała provided also two-sided moment and tail inequalities for higher degree homogeneous forms in Gaussian variables [17] (see also [2]).

The interest in Hanson-Wright type estimates has been recently revived in connection with non-asymptotic theory of random matrices and related statistical problems [32, 25]. Since in many applications one considers quadratic forms in random vectors with dependencies among coefficients, some recent work has been devoted to proving counterparts of the Hanson-Wright inequality in a dependent setting. In particular in [14] a corresponding upper tail inequality is proved for positive definite matrices and sub-Gaussian random vectors X (we recall that a random vector X in $\mathbb{R}^n$ is sub-Gaussian with constant K if for all $u \in S^{n-1}$, and all $t > 0$, $\mathbb{P}(|\langle X, u\rangle| \ge t) \le 2\exp(-t^2/K^2)$, where $\langle\cdot,\cdot\rangle$ stands for the standard inner product on $\mathbb{R}^n$). It is easy to see that in this setting one cannot hope for a lower tail estimate as a sub-Gaussian random vector can vanish with probability separated from zero. In [32], Vu and Wang consider vectors satisfying the convex concentration property (see Definition 2.2 below) and prove that if X is a random vector in $\mathbb{R}^n$ in the isotropic position (i.e. with mean zero and covariance matrix equal to identity) which has the convex concentration property with constant K, then for all $t > 0$,

$$\mathbb{P}(|X^TAX - \mathrm{tr}A| \ge t) \le C\log n \exp\Big(-CK^{-2}\min\Big(\frac{t^2}{\|A\|_{HS}^2\log n}, \frac{t}{\|A\|}\Big)\Big). \qquad (1.1)$$

(We remark that Vu and Wang considered complex random vectors with complex conjugate-transpose operation instead of transpose, but since we are interested here primarily in the real case, we do not state their result in this version. In fact it is not difficult to pass from the real version to the complex one).

One of the objectives of this paper is to remove the dependence on dimension in the above estimate (Theorem 2.5 below) as well as to prove corresponding uniform estimates for suprema of quadratic forms under some stronger assumptions on the random vector X (Theorem 2.10). Such uniform versions (corresponding to Banach space valued quadratic forms) for Gaussian random vectors were considered e.g. by Borell [8] and Arcones-Giné [4], whereas the Rademacher case was studied by Talagrand [31] and Bousquet-Boucheron-Lugosi-Massart [9]. In Theorem 2.10 we prove that a uniform estimate is a consequence of the concentration property for Lipschitz functions.

The estimates provided by uniform Hanson-Wright inequalities are expressed in terms of expectations of suprema of certain empirical processes. Since estimating these expectations is in general difficult, direct applications of such inequalities are limited. However, there are examples where it is possible to effectively bound the empirical process involved in the estimate. We present one of them in Section 4, Theorem 4.1, where we recover a recent concentration result for empirical approximations of the covariance operator for Banach space valued Gaussian variables, by Koltchinskii and Lounici [15]. We remark that the original proof in [15] used different methods.

The organization of the paper is as follows. In the next section we present our main results together with some additional discussion. Next, in Section 3 we provide proofs, deferring some technical parts to the Appendix. Finally, in Section 4 we present the aforementioned application of uniform estimates for quadratic forms.

Acknowledgements The author would like to thank Vladimir Koltchinskii and Karim Lounici for interesting conversations during The Seventh International Conference on High Dimensional Probability. The results of this paper grew directly out of those conversations. Separate thanks go to the organizers of the conference. The author would also like to express his gratitude to the anonymous Referee, whose comments allowed to greatly improve the presentation of results.

2 Main results

To introduce the setting for our estimates let us first recall the standard definitions of concentration properties of random vectors.

Definition 2.1 (Concentration property)**.** *Let X be a random vector in $\mathbb{R}^n$. We will say that X has the concentration property with constant K if for every 1-Lipschitz function $\varphi\colon \mathbb{R}^n \to \mathbb{R}$, we have $\mathbb{E}|\varphi(X)| < \infty$ and for every $t > 0$,*

$$\mathbb{P}(|\varphi(X) - \mathbb{E}\varphi(X)| \geq t) \leq 2\exp(-t^2/K^2).$$

The concentration property of random vectors has been extensively studied in the recent forty years, starting with the celebrated results by Borell [7] and Sudakov-Tsirelson [28] who established it for Gaussian random vectors. Many efficient techniques for proving concentration have been discovered, including e.g. isoperimetric techniques, functional inequalities, transportation of measure, semigroup tools. We refer to the monograph [20] by Ledoux for a thorough discussion of this topic.

Definition 2.2 (Convex concentration property)**.** *Let X be a random vector in $\mathbb{R}^n$. We will say that X has the convex concentration property with constant K if for every 1-Lipschitz convex function $\varphi\colon \mathbb{R}^n \to \mathbb{R}$, we have $\mathbb{E}|\varphi(X)| < \infty$ and for every $t > 0$,*

$$\mathbb{P}(|\varphi(X) - \mathbb{E}\varphi(X)| \geq t) \leq 2\exp(-t^2/K^2).$$

Remark 2.3. Let us now provide a few examples of random vectors satisfying the convex concentration property:

- Any random vector $X = (X_1, \ldots, X_n)$ such that X_i's are independent and for all i, $|X_i| \leq 1$ a.s., satisfies the convex concentration property with constant K independent of the dimension n (as proved in [27], one can take $K = 2\sqrt{2}$). This fact was first obtained by Talagrand [29, 30] by means of his celebrated convex distance inequality.
- An extension of Talagrand's result has been provided by Samson [27] and Paulin [24], who obtained concentration for vectors with bounded coordinates, satisfying some uniform mixing conditions or Dobrushin type criteria. In particular it follows from the results in [27] that if the random variables $X_1, X_2, \ldots,$ with values in a bounded interval, form a uniformly ergodic Markov chain (or more generally are geometrically strongly mixing), then the random vectors $X^{(n)} = (X_1, \ldots, X_n)$ satisfy the convex concentration property with a dimension independent constant. We

refer to the original articles [27, 24] for the precise formulation of the mixing conditions and the parameters controlling the constant K in the convex concentration property.

- From Talagrand's results it also follows that the convex concentration property is satisfied by vectors obtained via sampling without replacement [24, 1]. More precisely, if $x_1, \ldots, x_n \in [0,1]$ and for $m \le n$ the random vector $X = (X_1, \ldots, X_m)$ is obtained by sampling without replacement m numbers from the set $\{x_1, \ldots, x_n\}$, then X satisfies the convex concentration property with an absolute constant K.
- Sub-Gaussian estimates for the upper tails of Lipschitz convex functions of random vectors with independent coordinates were also obtained by Ledoux [19] and later Adamczak in the unbounded case [3] by means of log-Sobolev inequalities for convex functions.

Remark 2.4. It was shown in [29] that even the uniform distribution on the discrete cube $\{0,1\}^n$ does not satisfy the concentration property for all Lipschitz functions with dimension independent constants. In general the convex concentration property is a much weaker condition than the concentration property for all Lipschitz functions, which requires much more regularity of the random vector.

Our first result is

Theorem 2.5. *Let X be a mean zero random vector in $\mathbb{R}^n$. If X has the convex concentration property with constant K, then for any $n \times n$ matrix A and every $t > 0$,*

$$\begin{aligned} \mathbb{P}(|X^T AX - \mathbb{E}(X^T AX)| \ge t) &\le 2\exp\Big(-\frac{1}{CK^2}\min\Big(\frac{t^2}{\|A\|_{HS}^2\|\mathrm{Cov}\,(X)\|}, \frac{t}{\|A\|}\Big)\Big) \qquad (2.1) \\ &\le 2\exp\Big(-\frac{1}{C}\min\Big(\frac{t^2}{2K^4\|A\|_{HS}^2}, \frac{t}{K^2\|A\|}\Big)\Big) \end{aligned}$$

for some universal constant C.

Remark 2.6. The above theorem improves the estimate (1.1) due to Vu-Wang by removing the dimension dependent factors (note that in the isotropic case $\mathbb{E}X^T AX = \mathrm{tr}A$ and $\|\mathrm{Cov}\,(X)\| = \|\mathrm{Id}\| = 1$).

Remark 2.7. The assumption that X is centered is introduced just to simplify the statement of the theorem. Note that if X has the convex concentration property with constant K, then so does $\tilde{X} = X - \mathbb{E}X$. Moreover, a quadratic form in X can be decomposed into a sum of a quadratic form in $\tilde{X}$ and an affine function of X. Since linear functions are convex, Lipschitz, their deviations can be controlled by the convex concentration property. We leave the precise formulation of the corresponding inequality to the Reader.

Remark 2.8. As it will become clear from the proof, similar theorems hold if instead of sub-Gaussian concentration inequality for convex functions one assumes some other rate of decay for the tail probabilities. The whole argument remains then valid, one just has to modify accordingly the right-hand side of (2.1). Convex concentration property with sub-exponential tail decay was studied e.g. in [6].

Remark 2.9. We remark that it is not true that if $X = (X_1, \ldots, X_n)$ where X_i are i.i.d. sub-Gussian random variables, then X has the convex concentration property with a constant independent of dimension (as noted in [3] following [13]). Therefore, Theorem 2.5 does not imply the standard Hanson-Wright inequality.

Our second result concerns a uniform version of the Hanson-Wright inequality for suprema of quadratic forms and is contained in

Theorem 2.10. *Let X be a mean zero random vector in $\mathbb{R}^n$. Assume that X has the concentration property with constant K. Let $\mathcal{A}$ be a bounded set of $n \times n$ matrices and consider the random variable*

$$Z = \sup_{A \in \mathcal{A}} \Big(X^T A X - \mathbb{E} X^T A X \Big).$$

Then, for every $t > 0$,

$$\mathbb{P}(|Z - \mathbb{E}Z| \geq t) \leq 2 \exp\Big(-\frac{1}{CK^2} \min\Big(\frac{t^2}{\|X\|_{\mathcal{A}}^2}, \frac{t}{\sup_{A \in \mathcal{A}} \|A\|} \Big) \Big), \tag{2.2}$$

where

$$\|X\|_{\mathcal{A}} = \mathbb{E} \sup_{A \in \mathcal{A}} |(A + A^T) X|$$

and C is a universal constant.

Remark 2.11. One can easily see that if $\mathcal{A} = \{A\}$, then $\|X\|_{\mathcal{A}} \leq 2\|A\|_{HS}\sqrt{\|\mathrm{Cov}\, X\|}$. If in addition X has the convex concentration property with constant K, then $\|\mathrm{Cov}\, X\| \leq 2K^2$ (see the proof of Theorem 2.5 below). Thus the conclusion of the above theorem is stronger than that of Theorem 2.5. On the other hand the assumption is also stronger. We do not know if (2.2) is implied just by the convex concentration property. This is the case if instead of $\sup_{A \in \mathcal{A}} X^T A X$ one considers $\sup_{A \in \mathcal{A}} X^T A Y$, where Y is an independent copy of X (see [3]).

Remark 2.12. As mentioned in the Introduction, inequalities similar to (2.2) have been proven by many authors under various sets of assumptions. In particular Borell [8] and Arcones-Giné [4] obtained inequalities for Banach space valued polynomials in Gaussian random variables. When specialised to quadratic forms, these inequalities give an upper bound on $\mathbb{P}(\sup_{A \in \mathcal{A}} |X^T A X| \geq M + t)$, where M is a certain quantile of $\sup_{A \in \mathcal{A}} |X^T A X|$. The proofs are based on the Gaussian isoperimetric inequality. We do not see how to adapt their arguments to get concentration around the mean rather then deviation above a multiple of the mean. Talagrand [31] proved a concentration inequality for suprema of quadratic forms in Rademacher variables, which via the Central Limit Theorem implies the concentration inequality in the Gaussian case. The bound on the upper tail in Talagrand's inequality was later generalized to higher order forms by Boucheron, Bousquet, Lugosi and Massart [9].

3 Proofs of the main results

In what follows the letter C will denote an absolute constant, the value of which may change between various occurrences (even in the same line).

Before we proceed with the proofs of Theorems 2.5 and 2.10, let us briefly explain their structure. The main idea is very simple and similar in both cases. Namely, even though the function φ defining the random variable in question is not Lipschitz, $|\nabla\varphi|$ is. Moreover, it is also convex. This allows to use concentration for $|\nabla\varphi(X)|$ and restrict to a small set B where $|\nabla\varphi|$ is bounded by an appropriately chosen constant . Since B is convex, φ is Lipschitz on B and can be extended to a Lipschitz function $\tilde{\varphi}$ on $\mathbb{R}^n$ (in the setting of Theorem 2.5, after some preliminary reductions one can also assume that φ and $\tilde{\varphi}$ are convex). The tail of $\tilde{\varphi}(X)$ can be controlled by the concentration property and the theorem follows by combining the tail inequality for $\tilde{\varphi}(X)$ with estimates on $\mathbb{P}(X \notin B)$. The technical parts of the proofs are related to verifying the quantitative aspects of the above intuition. We remark that a similar strategy has been applied in [23] to provide concentration inequalities for non-commutative polynomials in random matrices.

To keep the main parts of the arguments concise, the proofs of some technical lemmas are postponed to the Appendix.

Proof of Theorem 2.5. Let us first prove the second inequality of (2.1). For any unit vector u, $\langle u, X\rangle$ is a 1-Lipschitz convex function of X. Since we also have $\mathbb{E}\langle u, X\rangle = 0$, by the convex concentration property, we get

$$u^T \mathrm{Cov}\,(X)u = \mathbb{E}\langle u, X\rangle^2 = 2\int_0^\infty t\mathbb{P}(|\langle u, X\rangle| \geq t)dt \leq 4\int_0^\infty te^{-t^2/K^2}dt = 2K^2.$$

This shows that $\|\mathrm{Cov}\,X\| \leq 2K^2$ and proves the second inequality of (2.1).

Let us now proceed with the proof of the first inequality. Since $X^TAX = X^T(\frac{1}{2}(A + A^T))X$, we can assume that A is symmetric. Under this assumption it is easy to see (by diagonalizing A) that $A = A_1 - A_2$, for some nonnegative definite matrices A_1 and A_2 such that $\|A_i\| \leq \|A\|$ and $\|A_i\|_{HS} \leq \|A\|_{HS}$, $i = 1, 2$. Therefore, by the triangle inequality, without loss of generality one can further assume that A is nonnegative definite.

Define $\varphi\colon \mathbb{R}^n \to \mathbb{R}$ by $\varphi(x) = x^TAx$. The function φ is convex (by nonnegative definiteness of A). Moreover, $\nabla\varphi(x) = 2Ax$ (here we use the symmetry of A). In particular, the function $f = |\nabla\varphi|$ is $2\|A\|$-Lipschitz. It is also easy to see that f is convex.

By the comparison of moments and basic linear algebra we get

$$(\mathbb{E}f(X))^2 \leq \mathbb{E}f(X)^2 = \langle 2AX, 2AX\rangle = 4\mathbb{E}X^TA^2X = 4\mathbb{E}\mathrm{tr}(A^2XX^T) = 4\mathrm{tr}(A^2\mathrm{Cov}\,(X)).$$

Now, using the inequality $\mathrm{tr}(QR) \leq (\mathrm{tr}\,Q)\|R\|$, valid for any symmetric, nonnegative definite matrices Q, R, we get

$$\mathbb{E}f(X) \leq \sqrt{4\mathrm{tr}(A^2)\|\mathrm{Cov}\,X\|} = 2\|A\|_{HS}\|\mathrm{Cov}\,(X)\|^{1/2}. \tag{3.1}$$

Let $B = \{x \in \mathbb{R}^n\colon |\nabla\varphi(x)| \leq 2\|A\|_{HS}\|\mathrm{Cov}\,(X)\|^{1/2} + \sqrt{t\|A\|}\}$. By the convex concentration property, applied to the function $f = |\nabla\varphi|$ (recall that f is $2\|A\|$-Lipschitz) and (3.1), we get

$$\mathbb{P}(X \notin B) \leq 2\exp\Big(-\frac{t}{4K^2\|A\|}\Big). \tag{3.2}$$

Define now a new function $\tilde{\varphi}\colon \mathbb{R}^n \to \mathbb{R}$ with the formula

$$\tilde{\varphi}(y) = \max_{x\in B}(\langle\nabla\varphi(x), y - x\rangle + \varphi(x)).$$

Note that $\tilde{\varphi}$ is a convex function, moreover for $y, z \in \mathbb{R}^n$,

$$\begin{aligned}\tilde{\varphi}(y) - \tilde{\varphi}(z) &= \max_{x\in B}(\langle\nabla\varphi(x), y - x\rangle + \varphi(x)) - \max_{x\in B}(\langle\nabla\varphi(x), z - x\rangle + \varphi(x))\\ &\leq \max_{x\in B}\langle\nabla\varphi(x), y - z\rangle \leq \max_{x\in B}|\nabla\varphi(x)||y - z| \leq M|y - z|,\end{aligned}$$

where $M = 2\|A\|_{HS}\|\mathrm{Cov}\,(X)\|^{1/2} + \sqrt{t\|A\|}$ (the last inequality follows by the definition of B). Thus $\tilde{\varphi}$ is convex and M-Lipschitz and so, by the convex concentration property, for all $s > 0$,

$$\mathbb{P}(|\tilde{\varphi}(X) - \mathbb{E}\tilde{\varphi}(X)| \geq s) \leq 2\exp\Big(-\frac{s^2}{K^2M^2}\Big). \tag{3.3}$$

Moreover, by convexity of φ, we have $\tilde{\varphi}(y) \leq \varphi(y)$ and thus for $y \in B$, we have $\tilde{\varphi}(y) = \varphi(y)$.

We will now use the following simple lemma. Its straightforward, but slightly tedious proof is deferred to the Appendix.

Lemma 3.1. *Let S and Z be random variables and $a, b, t > 0$ be such that for all $s > 0$,*

$$\mathbb{P}(|S - \mathbb{E}S| \geq s) \leq 2\exp(-s^2/(a + \sqrt{bt})^2) \tag{3.4}$$

and

$$\mathbb{P}(S \neq Z) \leq 2\exp(-t/b). \tag{3.5}$$

Then

$$\mathbb{P}(|Z - \mathrm{Med}\, Z| \geq t) \leq 2\exp\Big(-\frac{1}{C}\min\Big(\frac{t^2}{a^2}, \frac{t}{b}\Big)\Big), \tag{3.6}$$

where $\mathrm{Med}\, Z$ *is a median of* Z.

Thanks to (3.2) and (3.3) we can now apply Lemma 3.1 with $Z = \varphi(X) = X^T AX$, $S = \tilde{\varphi}(X)$, $a = 2K\|A\|_{HS}\|\mathrm{Cov}\,(X)\|^{1/2}$ and $b = 4K^2\|A\|$, and obtain

$$\mathbb{P}(|X^T AX - \mathrm{Med}\,(X^T AX)| \geq t) \leq 2\exp\Big(-\frac{1}{CK^2}\min\Big(\frac{t^2}{\|A\|_{HS}^2\|\mathrm{Cov}\,(X)\|}, \frac{t}{\|A\|}\Big)\Big). \tag{3.7}$$

We have thus obtained concentration for $X^T AX$, however around a median and not the mean. In the sub-Gaussian case it is classical that at the cost of a universal factor in the constant, one can replace the median with the mean (an vice versa). This fact is true also for two-level Gaussian-exponential concentration of (3.7), as summarized by the following lemma. Since we have not been able to find a proper reference in the literature, we present its proof in the Appendix.

Lemma 3.2. *Assume that a random variable Z satisfies*

$$\mathbb{P}(|Z - \mathrm{Med}\, Z| \geq t) \leq 2\exp\Big(-\min\Big(\frac{t^2}{a^2}, \frac{t}{b}\Big)\Big)$$

for all $t > 0$. Then for some absolute constant C and all $t > 0$,

$$\mathbb{P}(|Z - \mathbb{E}Z| \geq t) \leq 2\exp\Big(-\frac{1}{C}\min\Big(\frac{t^2}{a^2}, \frac{t}{b}\Big)\Big). \tag{3.8}$$

A direct application of the above lemma and (3.7) proves the first inequality of (2.1) and ends the proof of the Theorem. □

Let us now pass to the proof of the uniform version of the Hanson-Wright inequality, stated in Theorem 2.10.

Proof of Theorem 2.10. By the boundedness assumption on the set $\mathcal{A}$ and the integrability assumption on X we can assume that the set $\mathcal{A}$ is finite. Let thus $\mathcal{A} = \{A^{(1)}, \ldots, A^{(m)}\}$, where $A^{(k)} = [a_{ij}^{(k)}]_{i,j\leq n}$. Denote also $a^{(k)} = \mathbb{E}X^T A^{(k)} X$ and define the function $\varphi \colon \mathbb{R}^n \to \mathbb{R}$ with the formula

$$\varphi(x) = \max_{k\leq m}(x^T A^{(k)} x - a^{(k)}). \tag{3.9}$$

Note that φ is locally Lipschitz, moreover as the set of roots of a non-zero multivariate polynomial is of Lebesgue measure zero, for every x outside a set of Lebesgue measure zero, there exists unique $k \leq m$, such that

$$\varphi(x) = x^T A^{(k)} x - a^{(k)}.$$

For $k \leq m$ let B_k be the set of points $x \in R^n$ such that k is the unique maximizer in (3.9). Then $\mathbb{R}^n \setminus (\bigcup_{k\leq m} B_k)$ has Lebesgue measure equal to zero, moreover the sets B_k are open. Since for $x \in B_k$ we have $\nabla\varphi(x) = (A^{(k)} + (A^{(k)})^T)x$, we have Lebesgue-a.e.

$$|\nabla\varphi(x)| \le \max_{A\in\mathcal{A}} |(A+A^T)x|.$$

Let now $B = \{x \in \mathbb{R}^n\colon \max_{A\in\mathcal{A}} |(A+A^T)x| < \|X\|_{\mathcal{A}} + \sqrt{t\max_{A\in\mathcal{A}}\|A\|}\}$ and note that B is an open convex set.

We will need the following, standard lemma, whose proof we present for completeness in the Appendix.

Lemma 3.3. *Let $B \subseteq \mathbb{R}^n$ be an open, convex set and $\varphi\colon B \to \mathbb{R}$ be a locally Lipschitz function, such that Lebesgue-a.e. on B, $|\nabla\varphi(x)| \le M$. Then, φ is M-Lipschitz on B.*

Applying this lemma, we get that φ is M-Lipschitz on B with

$$M = \|X\|_{\mathcal{A}} + \sqrt{t\max_{A\in\mathcal{A}}\|A\|}.$$

Let now $\tilde{\varphi}\colon \mathbb{R}^n \to \mathbb{R}$ be any M-Lipschitz function, which coincides with φ on B (it exists by McShane's lemma, see e.g. Lemma 7.3. in [22]). By the concentration property of X we have for all $s > 0$,

$$\mathbb{P}(|\tilde{\varphi}(X) - \mathbb{E}\tilde{\varphi}(X)| \ge s) \le 2\exp\Big(-\frac{s^2}{K^2M^2}\Big)$$

and

$$\mathbb{P}(X \notin B) = \mathbb{P}\Big(\max_{A\in\mathcal{A}}|(A+A^T)X| \ge \|X\|_{\mathcal{A}} + \sqrt{t\max_{A\in\mathcal{A}}\|A\|}\Big) \le 2\exp\Big(-\frac{t}{4K^2\max_{A\in\mathcal{A}}\|A\|}\Big),$$

where we used that the function $x \mapsto \max_{A\in\mathcal{A}}|(A+A^T)x|$ has the Lipschitz constant bounded by $\max_{A\in\mathcal{A}}\|A+A^T\| \le 2\max_{A\in\mathcal{A}}\|A\|$. Thus, Lemma 3.1 with $S = \tilde{\varphi}(X)$, $Z = \varphi(X)$, $a = K\|X\|_{\mathcal{A}}$ and $b = 4K^2\max_{A\in\mathcal{A}}\|A\|$ gives

$$\mathbb{P}(|\varphi(X) - \mathrm{Med}\,\varphi(X)| \ge t) \le 2\exp\Big(-\frac{1}{CK^2}\min\Big(\frac{t^2}{\|X\|_{\mathcal{A}}^2}, \frac{t}{\max_{A\in\mathcal{A}}\|A\|}\Big)\Big).$$

Since the above inequality holds for arbitrary $t > 0$, we can use Lemma 3.2 to complete the proof. $\square$

4 Application. Concentration inequalities for the empirical covariance operator

Let us conclude with an application of Theorem 2.10 in the Gaussian setting, by providing a new proof of the concentration inequality for empirical approximations of the covariance operator of a Banach space valued random variable, proved recently in [15] by other methods. Since this part serves mostly as an illustration of applicability of Theorem 2.10, we do not present the general setting and motivation for this type of results, referring the Reader to the original paper [15].

In the formulation of the following theorem we use $\|\cdot\|$ to denote both the norm of a vector in a Banach space E and the operator norm. By E^* we will denote the dual of E. For $x \in E$ and $u \in E^*$, $\langle x, u\rangle$ will stand for the value of u on x, i.e. $\langle x, u\rangle = u(x)$.

Theorem 4.1. *Let G be a Gaussian vector with values in a separable Banach space E and let $\Sigma\colon E^* \to E$ be its covariance operator, i.e.*

$$\Sigma u = \mathbb{E}\langle G, u\rangle G,\ u \in E^*.$$

Let $G_1, \ldots, G_n$ be i.i.d. copies of G and define the empirical covariance operator $\hat{\Sigma}\colon E^* \to E$ with the formula

$$\hat{\Sigma}u = \frac{1}{n}\sum_{k=1}^{n}\langle G_k, u\rangle G_k,\ u \in E^*.$$

Then, for any $t \geq 1$,

$$\mathbb{P}\Big(\Big|\|\hat{\Sigma}-\Sigma\| - \mathbb{E}\|\hat{\Sigma}-\Sigma\|\Big| \geq C\Big(\|\Sigma\|\Big(1+\sqrt{\frac{r(\Sigma)}{n}}\Big)\sqrt{\frac{t}{n}} + \|\Sigma\|\frac{t}{n}\Big)\Big) \leq e^{-t},$$

where $r(\Sigma) = \frac{(\mathbb{E}\|G\|)^2}{\|\Sigma\|}$ is the effective rank of Σ.

Proof. By the Karhunen-Loève theorem, there exists a sequence $x_k \in E$, such that almost surely

$$G = \sum_{j=1}^{\infty} x_j g_j,$$

where g_j are i.i.d. standard Gaussian variables. Let $\{g_{ij}\}_{1\leq i\leq n, j\in\mathbb{N}}$ be an array of i.i.d. standard Gaussian variables. We can assume that

$$G_i = \sum_{j=1}^{\infty} x_j g_{ij}.$$

Then

$$\Sigma u = \sum_{j=1}^{\infty}\langle x_j, u\rangle x_j \quad \text{and} \quad \hat{\Sigma}u = \frac{1}{n}\sum_{k=1}^{n}\sum_{i=1}^{\infty}\sum_{j=1}^{\infty}\langle x_i, u\rangle x_j g_{ki} g_{kj}.$$

Therefore, denoting by B^* the unit ball of E^*, we get

$$\|\hat{\Sigma}-\Sigma\| = \sup_{u,v\in B^*}\Big(\frac{1}{n}\sum_{k=1}^{n}\sum_{i=1}^{\infty}\sum_{j=1}^{\infty}\langle x_i, u\rangle\langle x_j, v\rangle g_{ki}g_{kj} - \mathbb{E}\frac{1}{n}\sum_{k=1}^{n}\sum_{i=1}^{\infty}\sum_{j=1}^{\infty}\langle x_i, u\rangle\langle x_j, v\rangle g_{ki}g_{kj}\Big),$$

which puts us in position to use Theorem 2.10 with

$$\mathcal{A} = \{[n^{-1}\langle x_i, u\rangle\langle x_j, v\rangle \mathbf{1}_{\{k=l\}}]_{(k,i),(l,j)} \colon u, v \in B^*\}$$

and $X = (g_{ki})_{k\leq n, i\leq\infty}$ (we skip the standard details of approximation by finite dimensional vectors).

We will first express the operator norm of Σ in terms of the sequence x_k. We have

$$\|\Sigma\| = \sup_{u\in B^*}\|\Sigma u\| = \sup_{u,v\in B^*}\langle \Sigma u, v\rangle = \sup_{u,v\in B^*}\sum_{j=1}^{\infty}\langle x_j, u\rangle\langle x_j, v\rangle = \sup_{u\in B^*}\sum_{i=1}^{\infty}\langle x_i, u\rangle^2, \tag{4.1}$$

where the last equality follows from the Schwarz inequality.

Let us now estimate the parameters of Theorem 2.10. Using the fact that each $A \in \mathcal{A}$ is a block matrix with blocks of the form $\frac{1}{n}(\langle x_i, u\rangle)_{i=1}^{\infty} \otimes (\langle x_j, v\rangle)_{j=1}^{\infty}$, one easily gets that

$$\sup_{A\in\mathcal{A}}\|A\| = \frac{1}{n}\sup_{u,v\in B^*}\Big(\sum_{i=1}^{\infty}\langle x_i, u\rangle^2\Big)^{1/2}\Big(\sum_{j=1}^{\infty}\langle x_j, v\rangle^2\Big)^{1/2} = \frac{1}{n}\sup_{u\in B^*}\sum_{i=1}^{\infty}\langle x_i, u\rangle^2 = \frac{1}{n}\|\Sigma\|, \tag{4.2}$$

where in the last equality we used (4.1). Passing to $\|X\|_{\mathcal{A}}$, we have

$$\|X\|_{\mathcal{A}} \leq \mathbb{E}\sup_{A\in\mathcal{A}}|AX| + \mathbb{E}\sup_{A\in\mathcal{A}}|A^T X|. \tag{4.3}$$

Now,

$$\mathbb{E}\sup_{A\in\mathcal{A}}|A^TX| = \frac{1}{n}\mathbb{E}\sup_{u,v\in B^*}\Big(\sum_{k=1}^n\sum_{j=1}^\infty\langle x_j,v\rangle^2\Big(\sum_{i=1}^\infty\langle x_i,u\rangle g_{ki}\Big)^2\Big)^{1/2} \tag{4.4}$$
$$= \frac{1}{n}\sup_{v\in B^*}\Big(\sum_{j=1}^\infty\langle x_j,v\rangle^2\Big)^{1/2}\mathbb{E}\sup_{u\in B^*}\Big(\sum_{k=1}^n\Big(\sum_{i=1}^\infty\langle x_i,u\rangle g_{ki}\Big)^2\Big)^{1/2}$$
$$= \frac{1}{n}\|\Sigma\|^{1/2}\mathbb{E}\sup_{u\in B^*}\Big(\sum_{k=1}^n\Big(\sum_{i=1}^\infty\langle x_i,u\rangle g_{ki}\Big)^2\Big)^{1/2},$$

where in the last inequality we again used (4.1). To bound the last expectation above, we can use the Gordon-Chevet inequality [10, 11], which asserts that for any Banach spaces E, F and points $x_i\in E$, $y_k\in F$, the random operator

$$\Gamma = \sum_{i,k} g_{ki}x_i\otimes y_k\colon E^*\to F,$$

satisfies

$$\mathbb{E}\|\Gamma\|_{E^*\to F} \le \sup\{\|\sum_i t_ix_i\|_E\colon \sum_i t_i^2=1\}\mathbb{E}\|\sum_k g_ky_k\|_F$$
$$+\sup\{\|\sum_k t_ky_k\|_F\colon \sum_k t_k^2=1\}\mathbb{E}\|\sum_i g_ix_i\|_E,$$

where g_i's are i.i.d. standard Gaussian variables.

Applying this inequality with $\Gamma=\sum_{k,i}g_{ki}x_i\otimes y_k\colon E^*\to\ell_2^n$, where $y_1,\ldots,y_n$ is the standard basis of ℓ_2^n, we get

$$\mathbb{E}\sup_{u\in B^*}\Big(\sum_{k=1}^n\Big(\sum_{i=1}^\infty\langle x_i,u\rangle g_{ki}\Big)^2\Big)^{1/2} = \mathbb{E}\|\Gamma\|_{E^*\to\ell_2^n}$$
$$\le \sup\{\|\sum_{i=1}^\infty t_ix_i\|\colon \sum_{i=1}^\infty t_i^2=1\}\mathbb{E}|\sum_{k=1}^n g_ky_k| + \sup\{|\sum_{k=1}^n t_ky_k|\colon \sum_{k=1}^n t_k^2=1\}\mathbb{E}\|\sum_{i=1}^\infty g_ix_i\|$$
$$\le \sup_{u\in B^*}\Big(\sum_{i=1}^\infty\langle x_i,u\rangle^2\Big)^{1/2}\sqrt{n}+1\cdot\mathbb{E}\|G\| = \|\Sigma\|^{1/2}\sqrt{n}+\mathbb{E}\|G\|.$$

Going back to (4.4), we get

$$\mathbb{E}\sup_{A\in\mathcal{A}}|A^TX| \le \frac{\|\Sigma\|}{\sqrt{n}}+\frac{\|\Sigma\|^{1/2}\mathbb{E}\|G\|}{n}.$$

By symmetry, an analogous bound holds for the other expectation on the right-hand side of (4.3), hence (recall that $r(\Sigma)=\frac{(\mathbb{E}\|G\|)^2}{\|\Sigma\|}$)

$$\|X\|_{\mathcal{A}} \le 2\frac{\|\Sigma\|}{\sqrt{n}}+2\frac{\|\Sigma\|^{1/2}\mathbb{E}\|G\|}{n} = 2\frac{\|\Sigma\|}{\sqrt{n}}+2\frac{\|\Sigma\|}{\sqrt{n}}\sqrt{\frac{r(\Sigma)}{n}}. \tag{4.5}$$

Note that by a change of variable, up to universal constants, Theorem 2.10 can be equivalently stated as

$$\mathbb{P}\Big(|Z-\mathbb{E}Z|\ge C\Big(K\|X\|_{\mathcal{A}}\sqrt{t}+K^2(\sup_{A\in\mathcal{A}}\|A\|)t\Big)\Big)\le e^{-t}$$

for all $t\ge 1$. Moreover in the Gaussian case $K=\sqrt{2}$.

Combining this with the estimate (4.2) on $\sup_{A\in\mathcal{A}}\|A\|$ and (4.5), we get for $t\ge 1$,

$$\mathbb{P}\Big(\Big|\|\hat\Sigma-\Sigma\|-\mathbb{E}\|\hat\Sigma-\Sigma\|\Big|\ge C\Big(\|\Sigma\|\Big(1+\sqrt{\frac{r(\Sigma)}{n}}\Big)\sqrt{\frac{t}{n}}+\|\Sigma\|\frac{t}{n}\Big)\Big)\le e^{-t},$$

which ends the proof. □

A Proofs of technical lemmas

We will now provide the proofs of Lemmas 3.1, 3.2 and 3.3. Let us start with a simple observation given in

Lemma A.1. *Assume that a random variable Z satisfies*

$$\mathbb{P}(|Z - \mathbb{E}Z| \ge t) \le 2\exp(-t^2/K^2)$$

for all $t > 0$. Consider $p \in (0,1)$ and let $q_p Z = \inf\{t \in \mathbb{R} \colon \mathbb{P}(Z \le t) \ge p\}$ be the smallest p-th quantile of Z. Then

$$q_p Z \ge \mathbb{E}Z - K\sqrt{\log(2/p)}.$$

Proof. Assume that $q_p Z < \mathbb{E}Z - K\sqrt{\log(2/p)}$. Then

$$\mathbb{P}(Z \le q_p Z) < 2\exp(-K^2 \log(2/p)/K^2) = p,$$

which contradicts the standard inequality $\mathbb{P}(Z \le q_p Z) \ge p$. □

Proof of Lemma 3.1. Set

$$M = a + \sqrt{bt}.$$

Assume first that $t > \max(3b, 2M\sqrt{\log 8})$. We then have $\mathbb{P}(S \ne Z) \le 1/4$ and so $\mathbb{P}(S \le \operatorname{Med} Z) \ge 1/4$, which means that $\operatorname{Med} Z \ge q_{1/4} S$, where $q_p S = \inf\{t \colon \mathbb{P}(S \le t) \ge p\}$. By Lemma A.1, $\operatorname{Med} Z \ge q_{1/4} S \ge \mathbb{E}S - M\sqrt{\log 8}$ and thus

$$\mathbb{P}(Z - \operatorname{Med} Z \ge t) \le \mathbb{P}(S \ne Z) + \mathbb{P}(S - \mathbb{E}S \ge t - M\sqrt{\log 8}) \le \mathbb{P}(S \ne Z) + \mathbb{P}(S - \mathbb{E}S \ge t/2).$$

Using (3.4) with $s = t/2$ and (3.5), we obtain

$$\mathbb{P}(Z - \operatorname{Med} Z \ge t) \le 2\exp(-\frac{t}{b}) + 2\exp(-\frac{t^2}{4M^2}).$$

Similarly, by replacing S, Z, with $-S, -Z$ and using the fact that $-\operatorname{Med} Z$ is a median for $-Z$, we obtain

$$\mathbb{P}(Z - \operatorname{Med} Z \le -t) \le 2\exp(-\frac{t}{b}) + 2\exp(-\frac{t^2}{4M^2}).$$

Thus we have obtained that if $t > \max(3b, 2M\sqrt{2\log 8})$, then

$$\mathbb{P}(|Z - \operatorname{Med} Z| \ge t) \le 4\exp(-\frac{t}{b}) + 4\exp(-\frac{t^2}{4M^2}).$$

Using the definition of M and simple calculations, one can easily see that this implies (3.6) for some universal constant C. This ends the proof in the case $t > \max(3b, 2M\sqrt{\log 8})$.

Note that for $t \le \max(3b, 2M\sqrt{\log 8})$, we have $\exp(-t^2/(4M^2)) \ge 1/8$ or $\exp(-t/b) \ge 1/27$, so trivially

$$\mathbb{P}(|Z - \operatorname{Med} Z| \ge t) \le 27\exp(-\min\Big(\frac{t^2}{4M^2}, \frac{t}{b}\Big)),$$

which again implies (3.6) for C large enough. □

Proof of Lemma 3.2. We have

$$|\mathbb{E}Z - \operatorname{Med} Z| \le \mathbb{E}|Z - \operatorname{Med} Z| \le 2\int_0^\infty \exp\Big(-\min\Big(\frac{t^2}{a^2}, \frac{t}{b}\Big)\Big)dt \le \sqrt{\pi}a + 2b.$$

Thus for $t > 2\sqrt{\pi}a + 4b$, we have

$$\mathbb{P}(|Z - \mathbb{E}Z| \ge t) \le \mathbb{P}(|Z - \operatorname{Med} Z| \ge t/2) \le 2\exp\Big(-\min\Big(\frac{t^2}{4a^2}, \frac{t}{2b}\Big)\Big).$$

On the other hand, there exists an absolute constant C, such that if $t \leq 2\sqrt{\pi}a + 4b \leq 8\max(a,b)$, then

$$\frac{1}{C}\min\left(\frac{t^2}{a^2},\frac{t}{b}\right) \leq \log 2,$$

which implies that (3.8) is trivially satisfied in this case. This ends the proof of the lemma. □

Proof of Lemma 3.3. Let λ_k denote the Lebesgue measure on $\mathbb{R}^k$. By the Fubini theorem and the convexity of B, for λ_{2n} almost all pairs $(x,y) \in B \times B$ we have

$$\lambda_1(\{u \in [0,1]\colon \nabla\varphi(ux+(1-u)y) \text{ exists and } |\nabla\varphi(ux+(1-u)y)| \leq M\}) = 1.$$

Since $u \mapsto \varphi(ux+(1-u)y)$ is locally Lipschitz and thus absolutely continuous, we have for such x,y,

$$\begin{aligned}\varphi(x)-\varphi(y) &= \int_0^1 \frac{d}{du}\varphi(ux+(1-u)y)du = \int_0^1 \langle\nabla\varphi(ux+(1-u)y), x-y\rangle du\\ &\leq M|x-y|.\end{aligned}$$

By continuity and density arguments, the above inequality clearly extends to all $x,y \in B$, allowing us to conclude the proof. □

References

[1] R. Adamczak, D. Chafaï, and P. Wolff, *Circular law for random matrices with exchangeable entries*, Random Structures & Algorithms (2015), Published online. DOI: 10.1002/rsa.20599.

[2] R. Adamczak and P. Wolff, *Concentration inequalities for non-Lipschitz functions with bounded derivatives of higher order*, Probab. Theory Related Fields **162** (2015), no. 3, 531–586.

[3] Radosław Adamczak, *Logarithmic Sobolev inequalities and concentration of measure for convex functions and polynomial chaoses*, Bull. Pol. Acad. Sci. Math. **53** (2005), no. 2, 221–238. MR-2163396

[4] Miguel A. Arcones and Evarist Giné, *On decoupling, series expansions, and tail behavior of chaos processes*, J. Theoret. Probab. **6** (1993), no. 1, 101–122. MR-1201060

[5] Franck Barthe and Emanuel Milman, *Transference principles for log-Sobolev and spectral-gap with applications to conservative spin systems*, Comm. Math. Phys. **323** (2013), no. 2, 575–625. MR-3096532

[6] S. G. Bobkov and F. Götze, *Discrete isoperimetric and Poincaré-type inequalities*, Probab. Theory Related Fields **114** (1999), no. 2, 245–277. MR-1701522

[7] Christer Borell, *The Brunn-Minkowski inequality in Gauss space*, Invent. Math. **30** (1975), no. 2, 207–216. MR-0399402

[8] ______, *On the Taylor series of a Wiener polynomial*, Seminar Notes on multiple stochastic integration, polynomial chaos and their integration. Case Western Reserve Univ., Cleveland (1984).

[9] Stéphane Boucheron, Olivier Bousquet, Gábor Lugosi, and Pascal Massart, *Moment inequalities for functions of independent random variables*, Ann. Probab. **33** (2005), no. 2, 514–560. MR-2123200

[10] S. Chevet, *Séries de variables aléatoires gaussiennes à valeurs dans $E\hat{\otimes}_\varepsilon F$. Application aux produits d'espaces de Wiener abstraits*, Séminaire sur la Géométrie des Espaces de Banach (1977–1978), École Polytech., Palaiseau, 1978, pp. Exp. No. 19, 15. MR-520217

[11] Yehoram Gordon, *Some inequalities for Gaussian processes and applications*, Israel J. Math. **50** (1985), no. 4, 265–289. MR-800188

[12] D. L. Hanson and F. T. Wright, *A bound on tail probabilities for quadratic forms in independent random variables*, Ann. Math. Statist. **42** (1971), 1079–1083. MR-0279864

[13] Paweł Hitczenko, Stanisław Kwapień, Wenbo V. Li, Gideon Schechtman, Thomas Schlumprecht, and Joel Zinn, *Hypercontractivity and comparison of moments of iterated maxima and minima of independent random variables*, Electron. J. Probab. **3** (1998), No. 2, 26 pp. (electronic). MR-1491527

[14] Daniel Hsu, Sham M. Kakade, and Tong Zhang, *A tail inequality for quadratic forms of subgaussian random vectors*, Electron. Commun. Probab. **17** (2012), no. 52, 6. MR-2994877

[15] V. Koltchinskii and K. Lounici, *Concentration Inequalities and Moment Bounds for Sample Covariance Operators*, To appear in Bernoulli, http://arxiv.org/abs/1405.2468.

[16] Rafał Latała, *Tail and moment estimates for some types of chaos*, Studia Math. **135** (1999), no. 1, 39–53. MR-1686370

[17] ______, *Estimates of moments and tails of Gaussian chaoses*, Ann. Probab. **34** (2006), no. 6, 2315–2331. MR-2294983

[18] Rafal Latala, Piotr Mankiewicz, Krzysztof Oleszkiewicz, and Nicole Tomczak-Jaegermann, *Banach-Mazur distances and projections on random subgaussian polytopes*, Discrete Comput. Geom. **38** (2007), no. 1, 29–50. MR-2322114

[19] Michel Ledoux, *On Talagrand's deviation inequalities for product measures*, ESAIM Probab. Statist. **1** (1995/97), 63–87 (electronic). MR-1399224

[20] ______, *The concentration of measure phenomenon*, Mathematical Surveys and Monographs, vol. 89, American Mathematical Society, Providence, RI, 2001. MR-1849347

[21] Michel Ledoux and Michel Talagrand, *Probability in Banach spaces*, Ergebnisse der Mathematik und ihrer Grenzgebiete (3) [Results in Mathematics and Related Areas (3)], vol. 23, Springer-Verlag, Berlin, 1991, Isoperimetry and processes. MR-1102015

[22] Francesco Maggi, *Sets of finite perimeter and geometric variational problems*, Cambridge Studies in Advanced Mathematics, vol. 135, Cambridge University Press, Cambridge, 2012, An introduction to geometric measure theory. MR-2976521

[23] Mark W. Meckes and Stanisław J. Szarek, *Concentration for noncommutative polynomials in random matrices*, Proc. Amer. Math. Soc. **140** (2012), no. 5, 1803–1813. MR-2869165

[24] Daniel Paulin, *The convex distance inequality for dependent random variables, with applications to the stochastic travelling salesman and other problems*, Electron. J. Probab. **19** (2014). MR-3248197

[25] Holger Rauhut, Justin Romberg, and Joel A. Tropp, *Restricted isometries for partial random circulant matrices*, Appl. Comput. Harmon. Anal. **32** (2012), no. 2, 242–254. MR-2880281

[26] Mark Rudelson and Roman Vershynin, *Hanson-Wright inequality and sub-Gaussian concentration*, Electron. Commun. Probab. **18** (2013), no. 82, 9. MR-3125258

[27] Paul-Marie Samson, *Concentration of measure inequalities for Markov chains and Φ-mixing processes*, Ann. Probab. **28** (2000), no. 1, 416–461. MR-1756011

[28] V. N. Sudakov and B. S. Cirel′son, *Extremal properties of half-spaces for spherically invariant measures*, Zap. Naučn. Sem. Leningrad. Otdel. Mat. Inst. Steklov. (LOMI) **41** (1974), 14–24, 165, Problems in the theory of probability distributions, II. MR-0365680

[29] Michel Talagrand, *An isoperimetric theorem on the cube and the Kintchine-Kahane inequalities*, Proc. Amer. Math. Soc. **104** (1988), no. 3, 905–909. MR-964871

[30] ______, *Concentration of measure and isoperimetric inequalities in product spaces*, Inst. Hautes Études Sci. Publ. Math. (1995), no. 81, 73–205. MR-1361756

[31] ______, *New concentration inequalities in product spaces*, Invent. Math. **126** (1996), no. 3, 505–563. MR-1419006

[32] Van Vu and Ke Wang, *Random weighted projections, random quadratic forms and random eigenvectors*, Random Structures & Algorithms (2014), Published online. DOI: 10.1002/rsa.20561.

[33] F. T. Wright, *A bound on tail probabilities for quadratic forms in independent random variables whose distributions are not necessarily symmetric*, Ann. Probability **1** (1973), no. 6, 1068–1070. MR-0353419