

Hanson-Wright inequality and sub-gaussian concentration*

Mark Rudelson[†] Roman Vershynin[‡]

Abstract

In this expository note, we give a modern proof of Hanson-Wright inequality for quadratic forms in sub-gaussian random variables. We deduce a useful concentration inequality for sub-gaussian random vectors. Two examples are given to illustrate these results: a concentration of distances between random vectors and subspaces, and a bound on the norms of products of random and deterministic matrices.

Keywords: subgaussian random variables; concentration inequalities; random matrices.

AMS MSC 2010: 60F10.

Submitted to ECP on June 12, 2013, final version accepted on October 19, 2013.

1 Hanson-Wright inequality

Hanson-Wright inequality is a general concentration result for quadratic forms in sub-gaussian random variables. A version of this theorem was first proved in [9, 19], however with one weak point mentioned in Remark 1.2. In this article we give a modern proof of Hanson-Wright inequality, which automatically fixes the original weak point. We then deduce a useful concentration inequality for sub-gaussian random vectors, and illustrate it with two applications.

Our arguments use standard tools of high-dimensional probability. The reader unfamiliar with them may benefit from consulting the tutorial [18]. Still, we will recall the basic notions where possible. A random variable ξ is called sub-gaussian if its distribution is dominated by that of a normal random variable. This can be expressed by requiring that $\mathbb{E} \exp(\xi^2/K^2) \leq 2$ for some $K > 0$; the infimum of such K is traditionally called the sub-gaussian or ψ_2 norm of ξ . This turns the set of subgaussian random variables into the Orlicz space with the Orlicz function $\psi_2(t) = \exp(t^2) - 1$. A number of other equivalent definitions are used in the literature. In particular, ξ is sub-gaussian if and only if $\mathbb{E} |\xi|^p = O(p)^{p/2}$ as $p \rightarrow \infty$, so we can redefine the sub-gaussian norm of ξ as

$$\|\xi\|_{\psi_2} = \sup_{p \geq 1} p^{-1/2} (\mathbb{E} |X|^p)^{1/p}.$$

One can show that $\|\xi\|_{\psi_2}$ defined this way is within an absolute constant factor from the infimum of $K > 0$ mentioned above, see [18, Section 5.2.3]. One can similarly define sub-exponential random variables, i.e. by requiring that $\|\xi\|_{\psi_1} = \sup_{p \geq 1} p^{-1} (\mathbb{E} |X|^p)^{1/p} < \infty$.

For an $m \times n$ matrix $A = (a_{ij})$, recall that the operator norm of A is $\|A\| = \max_{\|x\|_2 \leq 1} \|Ax\|_2$ and the Hilbert-Schmidt (or Frobenius) norm of A is $\|A\|_{\text{HS}} = (\sum_{i,j} |a_{i,j}|^2)^{1/2}$. Throughout the paper, C, C_1, c, c_1, \dots denote positive absolute constants.

*Partial support: M. R. by NSF grant DMS 1161372. R. V. by NSF grant DMS 1001829 and 1265782.

[†]University of Michigan, U.S.A. E-mail: rudelson@umich.edu.

[‡]University of Michigan, U.S.A. E-mail: romanv@umich.edu.

Theorem 1.1 (Hanson-Wright inequality). *Let $X = (X_1, \dots, X_n) \in \mathbb{R}^n$ be a random vector with independent components X_i which satisfy $\mathbb{E} X_i = 0$ and $\|X_i\|_{\psi_2} \leq K$. Let A be an $n \times n$ matrix. Then, for every $t \geq 0$,*

$$\mathbb{P} \{ |X^\top A X - \mathbb{E} X^\top A X| > t \} \leq 2 \exp \left[-c \min \left(\frac{t^2}{K^4 \|A\|_{\text{HS}}^2}, \frac{t}{K^2 \|A\|} \right) \right].$$

Remark 1.2 (Related results). *One of the aims of this note is to give a simple and self-contained proof of the Hanson–Wright inequality using only the standard toolkit of the large deviation theory. Several partial results and alternative proofs are scattered in the literature.*

Improving upon an earlier result on Hanson-Wright [9], Wright [19] established a slightly weaker version of Theorem 1.1. Instead of $\|A\| = \|(a_{ij})\|$, both papers had $\|(|a_{ij}|)\|$ in the right side. The latter norm can be much larger than the norm of A , and it is often less easy to compute. This weak point went unnoticed in several later applications of Hanson-Wright inequality, however it was clear to experts that it could be fixed.

A proof for the case where X_1, \dots, X_n are independent symmetric Bernoulli random variables appears in the lecture notes of Nelson [14]. The moment inequality which essentially implies the result of [14] can be also found in [6]. A different approach to Hanson-Wright inequality, due to Rauhut and Tropp, can be found in [8, Proposition 8.13]. It is presented for diagonal-free matrices (however this assumption can be removed by treating the diagonal separately as is done below), and for independent symmetric Bernoulli random variables (but the proof can be extended to sub-gaussian random variables).

An upper bound for $\mathbb{P} \{ X^\top A X - \mathbb{E} X^\top A X > t \}$, which is equivalent to what appears in the Hanson–Wright inequality, can be found in [10]. However, the assumptions in [10] are somewhat different. On the one hand, it is assumed that the matrix A is positive-semidefinite, while in our result A can be arbitrary. On the other hand, a weaker assumption is placed on the random vector $X = (X_1, \dots, X_n)$. Instead of assuming that the coordinates of X are independent subgaussian random variables, it is assumed in [10] that the marginals of X are uniformly subgaussian, i.e., that $\sup_{y \in S^{n-1}} \|\langle X, y \rangle\|_{\psi_2} \leq K$.

The paper [3] contains an alternative short proof of Hanson–Wright inequality due to Latala for diagonal-free matrices. Like in the proof below, Latala’s argument uses decoupling of the order 2 chaos. However, unlike the current paper, which uses a simple decoupling argument of Bourgain [2], his proof uses a more general and more difficult decoupling theorem for U -statistics due to de la Peña and Montgomery-Smith [5]. For an extensive discussion of modern decoupling methods see [4].

Large deviation inequalities for polynomials of higher degree, which extend the Hanson-Wright type inequalities, have been obtained by Latala [11] and recently by Adamczak and Wolff [1].

Proof of Theorem 1.1. By replacing X with X/K we can assume without loss of generality that $K = 1$. Let us first estimate

$$p := \mathbb{P} \{ X^\top A X - \mathbb{E} X^\top A X > t \}.$$

Let $A = (a_{ij})_{i,j=1}^n$. By independence and zero mean of X_i , we can represent

$$\begin{aligned} X^\top A X - \mathbb{E} X^\top A X &= \sum_{i,j} a_{ij} X_i X_j - \sum_i a_{ii} \mathbb{E} X_i^2 \\ &= \sum_i a_{ii} (X_i^2 - \mathbb{E} X_i^2) + \sum_{i,j: i \neq j} a_{ij} X_i X_j. \end{aligned}$$

The problem reduces to estimating the diagonal and off-diagonal sums:

$$p \leq \mathbb{P} \left\{ \sum_i a_{ii}(X_i^2 - \mathbb{E} X_i^2) > t/2 \right\} + \mathbb{P} \left\{ \sum_{i,j: i \neq j} a_{ij} X_i X_j > t/2 \right\} =: p_1 + p_2.$$

Step 1: diagonal sum. Note that $X_i^2 - \mathbb{E} X_i^2$ are independent mean-zero sub-exponential random variables, and

$$\|X_i^2 - \mathbb{E} X_i^2\|_{\psi_1} \leq 2\|X_i^2\|_{\psi_1} \leq 4\|X_i\|_{\psi_2}^2 \leq 4.$$

These standard bounds can be found in [18, Remark 5.18 and Lemma 5.14]. Then we can use a Bernstein-type inequality (see [18, Proposition 5.16]) and obtain

$$p_1 \leq \left[-c \min \left(\frac{t^2}{\sum_i a_{ii}^2}, \frac{t}{\max_i |a_{ii}|} \right) \right] \leq \exp \left[-c \min \left(\frac{t^2}{\|A\|_{\text{HS}}^2}, \frac{t}{\|A\|} \right) \right]. \quad (1.1)$$

Step 2: decoupling. It remains to bound the off-diagonal sum

$$S := \sum_{i,j: i \neq j} a_{ij} X_i X_j.$$

The argument will be based on estimating the moment generating function of S by decoupling and reduction to normal random variables.

Let $\lambda > 0$ be a parameter whose value we will determine later. By Chebyshev's inequality, we have

$$p_2 = \mathbb{P} \{ S > t/2 \} = \mathbb{P} \{ \lambda S > \lambda t/2 \} \leq \exp(-\lambda t/2) \mathbb{E} \exp(\lambda S). \quad (1.2)$$

Consider independent Bernoulli random variables $\delta_i \in \{0, 1\}$ with $\mathbb{E} \delta_i = 1/2$. Since $\mathbb{E} \delta_i(1 - \delta_j)$ equals $1/4$ for $i \neq j$ and 0 for $i = j$, we have

$$S = 4 \mathbb{E}_\delta S_\delta, \quad \text{where} \quad S_\delta = \sum_{i,j} \delta_i(1 - \delta_j) a_{ij} X_i X_j.$$

Here \mathbb{E}_δ denotes the expectation with respect to $\delta = (\delta_1, \dots, \delta_n)$. Jensen's inequality yields

$$\mathbb{E} \exp(\lambda S) \leq \mathbb{E}_{X,\delta} \exp(4\lambda S_\delta) \quad (1.3)$$

where $E_{X,\delta}$ denotes expectation with respect to both X and δ . Consider the set of indices $\Lambda_\delta = \{i \in [n] : \delta_i = 1\}$ and express

$$S_\delta = \sum_{i \in \Lambda_\delta, j \in \Lambda_\delta^c} a_{ij} X_i X_j = \sum_{j \in \Lambda_\delta^c} X_j \left(\sum_{i \in \Lambda_\delta} a_{ij} X_i \right).$$

Now we condition on δ and $(X_i)_{i \in \Lambda_\delta}$. Then S_δ is a linear combination of mean-zero sub-gaussian random variables X_j , $j \in \Lambda_\delta^c$, with fixed coefficients. It follows that the conditional distribution of S_δ is sub-gaussian, and its sub-gaussian norm is bounded by the ℓ_2 -norm of the coefficient vector (see e.g. in [18, Lemma 5.9]). Specifically,

$$\|S_\delta\|_{\psi_2} \leq C \sigma_\delta \quad \text{where} \quad \sigma_\delta^2 := \sum_{j \in \Lambda_\delta^c} \left(\sum_{i \in \Lambda_\delta} a_{ij} X_i \right)^2.$$

Next, we use a standard estimate of the moment generating function of centered sub-gaussian random variables, see [18, Lemma 5.5]. It yields

$$\mathbb{E}_{(X_j)_{j \in \Lambda_\delta^c}} \exp(4\lambda S_\delta) \leq \exp(C\lambda^2 \|S_\delta\|_{\psi_2}^2) \leq \exp(C'\lambda^2 \sigma_\delta^2).$$

Taking expectations of both sides with respect to $(X_i)_{i \in \Lambda_\delta}$, we obtain

$$\mathbb{E}_X \exp(4\lambda S_\delta) \leq \mathbb{E}_X \exp(C' \lambda^2 \sigma_\delta^2) =: E_\delta. \tag{1.4}$$

Recall that this estimate holds for every fixed δ . It remains to estimate E_δ .

Step 3: reduction to normal random variables. Consider $g = (g_1, \dots, g_n)$ where g_i are independent $N(0, 1)$ random variables. The rotation invariance of normal distribution implies that for each fixed δ and X , we have

$$Z := \sum_{j \in \Lambda_\delta^c} g_j \left(\sum_{i \in \Lambda_\delta} a_{ij} X_i \right) \sim N(0, \sigma_\delta^2).$$

By the formula for the moment generating function of normal distribution, we have $\mathbb{E}_g \exp(sZ) = \exp(s^2 \sigma_\delta^2 / 2)$. Comparing this with the formula defining E_δ in (1.4), we find that the two expressions are somewhat similar. Choosing $s^2 = 2C' \lambda^2$, we can match the two expressions as follows:

$$E_\delta = \mathbb{E}_{X,g} \exp(C_1 \lambda Z)$$

where $C_1 = \sqrt{2C'}$.

Rearranging the terms, we can write $Z = \sum_{i \in \Lambda_\delta} X_i \left(\sum_{j \in \Lambda_\delta^c} a_{ij} g_j \right)$. Then we can bound the moment generating function of Z in the same way we bounded the moment generating function of S_δ in Step 2, only now relying on the sub-gaussian properties of $X_i, i \in \Lambda_\delta$. We obtain

$$E_\delta \leq \mathbb{E}_g \exp \left[C_2 \lambda^2 \sum_{i \in \Lambda_\delta} \left(\sum_{j \in \Lambda_\delta^c} a_{ij} g_j \right)^2 \right].$$

To express this more compactly, let P_δ denotes the coordinate projection (restriction) of \mathbb{R}^n onto $\mathbb{R}^{\Lambda_\delta}$, and define the matrix $A_\delta = P_\delta A (I - P_\delta)$. Then what we obtained

$$E_\delta \leq \mathbb{E}_g \exp \left(C_2 \lambda^2 \|A_\delta g\|_2^2 \right).$$

Recall that this bound holds for each fixed δ . We have removed the original random variables X_i from the problem, so it now becomes a problem about normal random variables g_i .

Step 4: calculation for normal random variables. By the rotation invariance of the distribution of g , the random variable $\|A_\delta g\|_2^2$ is distributed identically with $\sum_i s_i^2 g_i^2$ where s_i denote the singular values of A_δ . Hence by independence,

$$E_\delta = \mathbb{E}_g \exp \left(C_2 \lambda^2 \sum_i s_i^2 g_i^2 \right) = \prod_i \mathbb{E}_g \exp \left(C_2 \lambda^2 s_i^2 g_i^2 \right).$$

Note that each g_i^2 has the chi-squared distribution with one degree of freedom, whose moment generating function is $\mathbb{E} \exp(tg_i^2) = (1 - 2t)^{-1/2}$ for $t < 1/2$. Therefore

$$E_\delta \leq \prod_i (1 - 2C_2 \lambda^2 s_i^2)^{-1/2} \quad \text{provided} \quad \max_i C_2 \lambda^2 s_i^2 < 1/2.$$

Using the numeric inequality $(1 - z)^{-1/2} \leq e^z$ which is valid for all $0 \leq z \leq 1/2$, we can simplify this as follows:

$$E_\delta \leq \prod_i \exp(C_3 \lambda^2 s_i^2) = \exp \left(C_3 \lambda^2 \sum_i s_i^2 \right) \quad \text{provided} \quad \max_i C_3 \lambda^2 s_i^2 < 1/2.$$

Since $\max_i s_i = \|A_\delta\| \leq \|A\|$ and $\sum_i s_i^2 = \|A_\delta\|_{\text{HS}}^2 \leq \|A\|_{\text{HS}}^2$, we have proved the following:

$$E_\delta \leq \exp(C_3 \lambda^2 \|A\|_{\text{HS}}^2) \quad \text{for } \lambda \leq c_0 / \|A\|.$$

This is a uniform bound for all δ . Now we take expectation with respect to δ . Recalling (1.3) and (1.4), we obtain the following estimate on the moment generating function of S :

$$\mathbb{E} \exp(\lambda S) \leq \mathbb{E}_\delta E_\delta \leq \exp(C_3 \lambda^2 \|A\|_{\text{HS}}^2) \quad \text{for } \lambda \leq c_0 / \|A\|.$$

Step 5: conclusion. Putting this estimate into the exponential Chebyshev's inequality (1.2), we obtain

$$p_2 \leq \exp(-\lambda t / 2 + C_3 \lambda^2 \|A\|_{\text{HS}}^2) \quad \text{for } \lambda \leq c_0 / \|A\|.$$

Optimizing over λ , we conclude that

$$p_2 \leq \exp\left[-c \min\left(\frac{t^2}{\|A\|_{\text{HS}}^2}, \frac{t}{\|A\|}\right)\right] =: p(A, t).$$

Now we combine with a similar estimate (1.1) for p_1 and obtain

$$p = p_1 + p_2 \leq 2p(A, t).$$

Repeating the argument for $-A$ instead of A , we get $\mathbb{P}\{X^\top A X - \mathbb{E} X^\top A X < -t\} \leq 2p(A, t)$. Combining the two events, we obtain $\mathbb{P}\{|X^\top A X - \mathbb{E} X^\top A X| > t\} \leq 4p(A, t)$. Finally, one can reduce the factor 4 to 2 by adjusting the constant c in $p(A, t)$. The proof is complete. \square

2 Sub-gaussian concentration

Hanson-Wright inequality has a useful consequence, a concentration inequality for random vectors with independent sub-gaussian coordinates.

Theorem 2.1 (Sub-gaussian concentration). *Let A be a fixed $m \times n$ matrix. Consider a random vector $X = (X_1, \dots, X_n)$ where X_i are independent random variables satisfying $\mathbb{E} X_i = 0$, $\mathbb{E} X_i^2 = 1$ and $\|X_i\|_{\psi_2} \leq K$. Then for any $t \geq 0$, we have*

$$\mathbb{P}\{|\|AX\|_2 - \|A\|_{\text{HS}}| > t\} \leq 2 \exp\left(-\frac{ct^2}{K^4 \|A\|^2}\right).$$

Remark 2.2. *The consequence of Theorem 2.1 can be alternatively formulated as follows: the random variable $Z = \|AX\|_2 - \|A\|_{\text{HS}}$ is sub-gaussian, and $\|Z\|_{\psi_2} \leq CK^2 \|A\|$.*

Remark 2.3. *A few special cases of Theorem 2.1 can be easily deduced from classical concentration inequalities. For Gaussian random variables X_i , this result is a standard consequence of Gaussian concentration, see e.g. [13]. For bounded random variables X_i , it can be deduced in a similar way from Talagrand's concentration for convex Lipschitz functions [15], see [16, Theorem 2.1.13]. For more general random variables, one can find versions of Theorem 2.1 with varying degrees of generality scattered in the literature (e.g. the appendix of [7]). However, we were unable to find Theorem 2.1 in the existing literature.*

Proof. Let us apply Hanson-Wright inequality, Theorem 1.1, for the matrix $Q = A^\top A$. Since $X^\top Q X = \|AX\|_2^2$, we have $\mathbb{E} X^\top Q X = \|A\|_{\text{HS}}^2$. Also, note that since all X_i have unit variance, we have $K \geq 2^{-1/2}$. Thus we obtain for any $u \geq 0$ that

$$\mathbb{P}\{|\|AX\|_2^2 - \|A\|_{\text{HS}}^2| > u\} \leq 2 \exp\left[-\frac{C}{K^4} \min\left(\frac{u}{\|A\|^2}, \frac{u^2}{\|A^\top A\|_{\text{HS}}^2}\right)\right].$$

Let $\varepsilon \geq 0$ be arbitrary, and let us use this estimate for $u = \varepsilon \|A\|_{\text{HS}}^2$. Since $\|A^\top A\|_{\text{HS}}^2 \leq \|A^\top\|^2 \|A\|_{\text{HS}}^2 = \|A\|^2 \|A\|_{\text{HS}}^2$, it follows that

$$\mathbb{P} \left\{ \left| \|AX\|_2^2 - \|A\|_{\text{HS}}^2 \right| > \varepsilon \|A\|_{\text{HS}}^2 \right\} \leq 2 \exp \left[-c \min(\varepsilon, \varepsilon^2) \frac{\|A\|_{\text{HS}}^2}{K^4 \|A\|^2} \right]. \quad (2.1)$$

Now let $\delta \geq 0$ be arbitrary; we shall use this inequality for $\varepsilon = \max(\delta, \delta^2)$. Observe that the (likely) event $\left| \|AX\|_2^2 - \|A\|_{\text{HS}}^2 \right| \leq \varepsilon \|A\|_{\text{HS}}^2$ implies the event $\left| \|AX\|_2 - \|A\|_{\text{HS}} \right| \leq \delta \|A\|_{\text{HS}}$. This can be seen by dividing both sides of the inequalities by $\|A\|_{\text{HS}}^2$ and $\|A\|_{\text{HS}}$ respectively, and using the numeric bound $\max(|z-1|, |z-1|^2) \leq |z^2-1|$, which is valid for all $z \geq 0$. Using this observation along with the identity $\min(\varepsilon, \varepsilon^2) = \delta^2$, we deduce from (2.1) that

$$\mathbb{P} \left\{ \left| \|AX\|_2 - \|A\|_{\text{HS}} \right| > \delta \|A\|_{\text{HS}} \right\} \leq 2 \exp \left(-c \delta^2 \frac{\|A\|_{\text{HS}}^2}{K^4 \|A\|^2} \right).$$

Setting $\delta = t/\|A\|_{\text{HS}}$, we obtain the desired inequality. □

2.1 Small ball probabilities

Using a standard symmetrization argument, we can deduce from Theorem 2.1 some bounds on small ball probabilities. The following result is due to Latala et al. [12, Theorem 2.5].

Corollary 2.4 (Small ball probabilities). *Let A be a fixed $m \times n$ matrix. Consider a random vector $X = (X_1, \dots, X_n)$ where X_i are independent random variables satisfying $\mathbb{E} X_i = 0$, $\mathbb{E} X_i^2 = 1$ and $\|X_i\|_{\psi_2} \leq K$. Then for every $y \in \mathbb{R}^m$ we have*

$$\mathbb{P} \left\{ \|AX - y\|_2 < \frac{1}{2} \|A\|_{\text{HS}} \right\} \leq 2 \exp \left(-\frac{c \|A\|_{\text{HS}}^2}{K^4 \|A\|^2} \right).$$

Remark 2.5. *Informally, Corollary 2.4 states that the small ball probability decays exponentially in the stable rank $r(A) = \|A\|_{\text{HS}}^2/\|A\|^2$.*

Proof. Let X' denote an independent copy of the random vector X . Denote $p = \mathbb{P} \left\{ \|AX - y\|_2 < \frac{1}{2} \|A\|_{\text{HS}} \right\}$. Using independence and triangle inequality, we have

$$\begin{aligned} p^2 &= \mathbb{P} \left\{ \|AX - y\|_2 < \frac{1}{2} \|A\|_{\text{HS}}, \|AX' - y\|_2 < \frac{1}{2} \|A\|_{\text{HS}} \right\} \\ &\leq \mathbb{P} \left\{ \|A(X - X')\|_2 < \|A\|_{\text{HS}} \right\}. \end{aligned} \quad (2.2)$$

The components of the random vector $X - X'$ have mean zero, variances bounded below by 2 and sub-gaussian norms bounded above by $2K$. Thus we can apply Theorem 2.1 for $\frac{1}{\sqrt{2}}(X - X')$ and conclude that

$$\mathbb{P} \left\{ \|A(X - X')\|_2 < \sqrt{2}(\|A\|_{\text{HS}} - t) \right\} \leq 2 \exp \left(-\frac{ct^2}{K^4 \|A\|^2} \right), \quad t \geq 0.$$

Using this with $t = (1 - 1/\sqrt{2})\|A\|_{\text{HS}}$, we obtain the desired bound for (2.2). □

The following consequence of Corollary 2.4 is even more informative. It states that $\|AX - y\|_2 \gtrsim \|A\|_{\text{HS}} + \|y\|_2$ with high probability.

Corollary 2.6 (Small ball probabilities, improved). *Let A be a fixed $m \times n$ matrix. Consider a random vector $X = (X_1, \dots, X_n)$ where X_i are independent random variables satisfying $\mathbb{E} X_i = 0$, $\mathbb{E} X_i^2 = 1$ and $\|X_i\|_{\psi_2} \leq K$. Then for every $y \in \mathbb{R}^m$ we have*

$$\mathbb{P} \left\{ \|AX - y\|_2 < \frac{1}{6} (\|A\|_{\text{HS}} + \|y\|_2) \right\} \leq 2 \exp \left(-\frac{c \|A\|_{\text{HS}}^2}{K^4 \|A\|^2} \right).$$

Proof. Denote $h := \|A\|_{\text{HS}}$. Combining the conclusions of Theorem 2.1 and Corollary 2.4, we obtain that with probability at least $1 - 4\exp(-ch^2/K^4\|A\|^2)$, the following two estimates hold simultaneously:

$$\|AX\|_2 \leq \frac{3}{2}h \quad \text{and} \quad \|AX - y\|_2 \geq \frac{1}{2}h. \tag{2.3}$$

Suppose this event occurs. Then by triangle inequality, $\|AX - y\|_2 \geq \|y\|_2 - \|AX\|_2 \geq \|y\|_2 - \frac{3}{2}h$. Combining this with the second inequality in (2.3), we obtain that

$$\|AX - y\|_2 \geq \max\left(\frac{1}{2}h, \|y\|_2 - \frac{3}{2}h\right) \geq \frac{1}{6}(h + \|y\|_2).$$

The proof is complete. □

3 Two applications

Concentration results like Theorem 2.1 have many useful consequences. We include two applications in this article; the reader will certainly find more.

The first application is a concentration of distance from a random vector to a fixed subspace. For random vectors with bounded components, one can find a similar result in [16, Corollary 2.1.19], where it was deduced from Talagrand’s concentration inequality.

Corollary 3.1 (Distance between a random vector and a subspace). *Let E be a subspace of \mathbb{R}^n of dimension d . Consider a random vector $X = (X_1, \dots, X_n)$ where X_i are independent random variables satisfying $\mathbb{E} X_i = 0$, $\mathbb{E} X_i^2 = 1$ and $\|X_i\|_{\psi_2} \leq K$. Then for any $t \geq 0$, we have*

$$\mathbb{P} \left\{ |d(X, E) - \sqrt{n-d}| > t \right\} \leq 2\exp(-ct^2/K^4).$$

Proof. The conclusion follows from Theorem 2.1 for $A = P_{E^\perp}$, the orthogonal projection onto E . Indeed, $d(X, E) = \|P_{E^\perp} X\|_2$, $\|P_{E^\perp}\|_{\text{HS}} = \dim(E^\perp) = \sqrt{n-d}$ and $\|P_{E^\perp}\| = 1$. □

Our second application of Theorem 2.1 is for operator norms of random matrices. The result essentially states that an $m \times n$ matrix BG obtained as a product of a deterministic matrix B and a random matrix G with independent sub-gaussian entries satisfies

$$\|BG\| \lesssim \|B\|_{\text{HS}} + \sqrt{n}\|B\|$$

with high probability. For random matrices with heavy-tailed rather than sub-gaussian components, this problem was studied in [17].

Theorem 3.2 (Norms of random matrices). *Let B be a fixed $m \times N$ matrix, and let G be an $N \times n$ random matrix with independent entries that satisfy $\mathbb{E} G_{ij} = 0$, $\mathbb{E} G_{ij}^2 = 1$ and $\|G_{ij}\|_{\psi_2} \leq K$. Then for any $s, t \geq 1$ we have*

$$\mathbb{P} \left\{ \|BG\| > CK^2(s\|B\|_{\text{HS}} + t\sqrt{n}\|B\|) \right\} \leq 2\exp(-s^2r - t^2n).$$

Here $r = \|B\|_{\text{HS}}^2/\|B\|^2$ is the stable rank of B .

Proof. We need to bound $\|BGx\|_2$ uniformly for all $x \in S^{n-1}$. Let us first fix $x \in S^{n-1}$. By concatenating the rows of G , we can view G as a long vector in \mathbb{R}^{Nn} . Consider the linear operator $T : \ell_2^{Nn} \rightarrow \ell_2^m$ defined as $T(G) = BGx$, and let us apply Theorem 2.1 for $T(G)$. To this end, it is not difficult to see that the Hilbert-Schmidt norm of T equals $\|B\|_{\text{HS}}$ and the operator norm of T is at most $\|B\|$. (The latter follows from

$\|BGx\| \leq \|B\| \|G\| \|x\|_2 \leq \|B\| \|G\|_{\text{HS}}$, and from the fact the $\|G\|_{\text{HS}}$ is the Euclidean norm of G as a vector in ℓ_2^{Nn} . Then for any $u \geq 0$, we have

$$\mathbb{P} \left\{ \|BGx\|_2 > \|B\|_{\text{HS}} + u \right\} \leq 2 \exp \left(- \frac{cu^2}{K^4 \|B\|^2} \right).$$

The last part of the proof is a standard covering argument. Let \mathcal{N} be an $1/2$ -net of S^{n-1} in the Euclidean metric. We can choose this net so that $|\mathcal{N}| \leq 5^n$, see [18, Lemma 5.2]. By a union bound, with probability at least

$$5^n \cdot 2 \exp \left(- \frac{cu^2}{K^4 \|B\|^2} \right), \tag{3.1}$$

every $x \in \mathcal{N}$ satisfies $\|BGx\|_2 \leq \|B\|_{\text{HS}} + u$. On this event, the approximation lemma (see [18, Lemma 5.2]) implies that every $x \in S^{n-1}$ satisfies $\|BGx\|_2 \leq 2(\|B\|_{\text{HS}} + u)$. It remains to choose $u = CK^2(s\|B\|_{\text{HS}} + t\sqrt{n}\|B\|)$ with sufficiently large absolutely constant C in order to make the probability bound (3.1) smaller than $2 \exp(-s^2r - t^2n)$. This completes the proof. \square

Remark 3.3. A couple of special cases in Theorem 3.2 are worth mentioning. If $B = P$ is a projection in \mathbb{R}^N of rank r then

$$\mathbb{P} \left\{ \|PG\| > CK^2(s\sqrt{r} + t\sqrt{n}) \right\} \leq 2 \exp(-s^2r - t^2n).$$

The same holds if $B = P$ is an $r \times N$ matrix such that $PP^T = I_r$.

In particular, if $B = I_N$ we obtain

$$\mathbb{P} \left\{ \|G\| > CK^2(s\sqrt{N} + t\sqrt{n}) \right\} \leq 2 \exp(-s^2N - t^2n).$$

3.1 Complexification

We formulated the results in Sections 2 and 3 for real matrices and real valued random variables. Using a standard complexification trick, one can easily obtain complex versions of these results. Let us show how to complexify Theorem 2.1; the other applications follow from it.

Suppose A is a complex matrix while X is a real-valued random vector as before. Then we can apply Theorem 2.1 for the real $2m \times n$ matrix $\tilde{A} := \begin{bmatrix} \text{Re } A \\ \text{Im } A \end{bmatrix}$. Note that $\|\tilde{A}X\|_2 = \|AX\|_2$, $\|\tilde{A}\| = \|A\|$ and $\|\tilde{A}\|_{\text{HS}} = \|A\|_{\text{HS}}$. Then the conclusion of Theorem 2.1 follows for A .

Suppose now that both A and X are complex. Let us assume that the components X_i have independent real and imaginary parts, such that

$$\text{Re } X_i = 0, \quad \mathbb{E}(\text{Re } X_i)^2 = \frac{1}{2}, \quad \|\text{Re } X_i\|_{\psi_2} \leq K,$$

and similarly for $\text{Im } X_i$. Then we can apply Theorem 2.1 for the real $2m \times 2n$ matrix $A' := \begin{bmatrix} \text{Re } A & -\text{Im } A \\ \text{Im } A & \text{Re } A \end{bmatrix}$ and vector $X' = \sqrt{2}(\text{Re } X \ \text{Im } X) \in \mathbb{R}^{2n}$. Note that $\|A'X'\|_2 = \sqrt{2}\|AX\|_2$, $\|A'\| = \|A\|$ and $\|A'\|_{\text{HS}} = \sqrt{2}\|A\|_{\text{HS}}$. Then the conclusion of Theorem 2.1 follows for A .

References

[1] R. Adamczak, P. Wolff, *Concentration inequalities for non-Lipschitz functions with bounded derivatives of higher order*, arXiv:1304.1826

- [2] J. Bourgain, *Random points in isotropic convex sets*. In: *Convex geometric analysis*, Berkeley, CA, 1996, Math. Sci. Res. Inst. Publ., Vol. 34, 53–58, Cambridge Univ. Press, Cambridge (1999). MR-1665576
- [3] F. Barthe, E. Milman, *Transference Principles for Log-Sobolev and Spectral-Gap with Applications to Conservative Spin Systems*, arXiv:1202.5318.
- [4] V. H. de la Peña, E. Giné, *Decoupling*. From dependence to independence. Randomly stopped processes. U-statistics and processes. Martingales and beyond. Probability and its Applications (New York). Springer-Verlag, New York, 1999. MR-1666908
- [5] V. H. de la Peña, S. J. Montgomery-Smith, *Decoupling inequalities for the tail probabilities of multivariate U-statistics*, Ann. Probab. 23 (1995), no. 2, 806–816. MR-1334173
- [6] I. Diakonikolas, D. M. Kane, J. Nelson, *Bounded independence fools degree-2 threshold functions*, Proceedings of the 51st Annual IEEE Symposium on Foundations of Computer Science (FOCS 2010), Las Vegas, NV, October 23-26, 2010. MR-3024771
- [7] L. Erdős, H.-T. Yau, J. Yin, *Bulk universality for generalized Wigner matrices*, Probability Theory and Related Fields 154, 341–407. MR-2981427
- [8] A. Foucart, H. Rauhut, *A Mathematical Introduction to Compressive Sensing*. Applied and Numerical Harmonic Analysis. Birkhäuser, 2013.
- [9] D. L. Hanson, E. T. Wright, *A bound on tail probabilities for quadratic forms in independent random variables*, Ann. Math. Statist. 42 (1971), 1079–1083. MR-0279864
- [10] D. Hsu, S. Kakade, T. Zhang, *A tail inequality for quadratic forms of subgaussian random vectors*, Electron. Commun. Probab. 17 (2012), no. 52, 1–6. MR-2994877
- [11] R. Latała, *Estimates of moments and tails of Gaussian chaoses*, Ann. Probab. 34 (2006), no. 6, 2315–2331. MR-2294983
- [12] R. Latała, P. Mankiewicz, K. Oleszkiewicz, N. Tomczak-Jaegermann, *Banach-Mazur distances and projections on random subgaussian polytopes*, Discrete Comput. Geom. 38 (2007), 29–50. MR-2322114
- [13] M. Ledoux, *The concentration of measure phenomenon*. Mathematical Surveys and Monographs, 89. Providence: American Mathematical Society, 2005. MR-1849347
- [14] J. Nelson, *Johnson–Lindenstrauss notes*, http://web.mit.edu/minilek/www/jl_notes.pdf.
- [15] M. Talagrand, *Concentration of measure and isoperimetric inequalities in product spaces*, IHES Publ. Math. No. 81 (1995), 73–205. MR-1361756
- [16] T. Tao, *Topics in random matrix theory*. Graduate Studies in Mathematics, 132. American Mathematical Society, Providence, RI, 2012. MR-2906465
- [17] R. Vershynin, *Spectral norm of products of random and deterministic matrices*, Probability Theory and Related Fields 150 (2011), 471–509. MR-2824864
- [18] R. Vershynin, *Introduction to the non-asymptotic analysis of random matrices*. Compressed sensing, 210–268, Cambridge Univ. Press, Cambridge, 2012. MR-2963170
- [19] E. T. Wright, *A bound on tail probabilities for quadratic forms in independent random variables whose distributions are not necessarily symmetric*, Ann. Probability 1 (1973), 1068–1070. MR-0353419

Electronic Journal of Probability

Electronic Communications in Probability

Advantages of publishing in EJP-ECP

- Very high standards
- Free for authors, free for readers
- Quick publication (no backlog)

Economical model of EJP-ECP

- Low cost, based on free software (OJS¹)
- Non profit, sponsored by IMS², BS³, PKP⁴
- Purely electronic and secure (LOCKSS⁵)

Help keep the journal free and vigorous

- Donate to the IMS open access fund⁶ (click here to donate!)
- Submit your best articles to EJP-ECP
- Choose EJP-ECP over for-profit journals

¹OJS: Open Journal Systems <http://pkp.sfu.ca/ojs/>

²IMS: Institute of Mathematical Statistics <http://www.imstat.org/>

³BS: Bernoulli Society <http://www.bernoulli-society.org/>

⁴PK: Public Knowledge Project <http://pkp.sfu.ca/>

⁵LOCKSS: Lots of Copies Keep Stuff Safe <http://www.lockss.org/>

⁶IMS Open Access Fund: <http://www.imstat.org/publications/open.htm>