Hanson–Wright inequality in Banach spaces

Radosław Adamczak[a,b], Rafał Latała[b] and Rafał Meller[b]

[a]*Institute of Mathematics of the Polish Academy of Sciences, Śniadeckich 8, 00-656 Warsaw, Poland. E-mail: r.adamczak@mimuw.edu.pl*
[b]*Institute of Mathematics, University of Warsaw, Banacha 2, 02-097 Warsaw, Poland. E-mail: r.latala@mimuw.edu.pl; r.meller@mimuw.edu.pl*

Abstract. We discuss two-sided bounds for moments and tails of quadratic forms in Gaussian random variables with values in Banach spaces. We state a natural conjecture and show that it holds up to additional logarithmic factors. Moreover in a certain class of Banach spaces (including L_r-spaces) these logarithmic factors may be eliminated. As a corollary we derive upper bounds for tails and moments of quadratic forms in subgaussian random variables, which extend the Hanson–Wright inequality.

Résumé. Nous étudions des bornes bilatères pour les moments et queues de distribution de formes quadratiques de variables aléatoires gaussiennes à valeurs dans des espaces de Banach. Nous formulons une conjecture naturelle, et en proposons une preuve à des facteurs logarithmiques près. De plus nous montrons que ces facteurs logarithmiques sont éliminables pour une certaine classe d'espaces de Banach incluant les espaces L_r. Comme corollaire, nous obtenons une majoration pour les moments et la queue de distribution de formes quadratiques de variables aléatoires sous-gaussiennes, qui étend l'inégalité de Hanson–Wright.

1. Introduction and main results

The Hanson–Wright inequality gives an upper bound for tails of real quadratic forms in independent subgaussian random variables. Recall that a random variable X is called α-*subgaussian* if for every $t > 0$, $\mathbb{P}(|X| \geq t) \leq 2\exp(-t^2/2\alpha^2)$. The Hanson–Wright inequality states that for any sequence of independent mean zero α-subgaussian random variables $X_1, \ldots, X_n$ and any symmetric matrix $A = (a_{ij})_{i,j \leq n}$ one has

$$\mathbb{P}\left(\left|\sum_{i,j=1}^{n} a_{ij}(X_i X_j - \mathbb{E}(X_i X_j))\right| \geq t\right) \leq 2\exp\left(-\frac{1}{C}\min\left\{\frac{t^2}{\alpha^4\|A\|_{\mathrm{HS}}}, \frac{t}{\alpha^2\|A\|_{\mathrm{op}}}\right\}\right), \tag{1}$$

where in the whole article we use the letter C to denote universal constants which may differ at each occurrence. Estimate (1) was essentially established in [12] in the symmetric and in [33] in the mean zero case (in fact in both papers the operator norm of A was replaced by the operator norm of $(|a_{ij}|)_{i,j \leq n}$, which in general could be much bigger, proofs of (1) may be found in [6] and [27]).

The Hanson–Wright inequality has found numerous applications in high-dimensional probability and statistics, as well as in random matrix theory (see, e.g., [32]). However in many problems one faces the need to analyze not a single quadratic form but a supremum of a collection of them or equivalently a norm of a quadratic form with coefficients in a Banach space. Let us mention a few recent developments, relying on the analysis of suprema of quadratic forms in independent random variables. For instance in [16] estimates for such suprema are applied to the theory of compressed sensing with circulant type matrices. In [1] they are used to give an alternate proof of the result from [15] concerning concentration

properties for sample covariance operators corresponding to Banach space-valued Gaussian random variables. In [11] one can find applications to M-estimation in the context of linear random-effects models.

The inequalities concerning tails of Banach space valued quadratic forms available in the literature (see, e.g., ineq. (3) below) are usually expressed in terms of quantities which themselves are troublesome to analyze. The main objective of this article is to provide estimates on vector-valued quadratic forms which can be applied more easily and are of optimal form.

The main step in modern proofs of the Hanson–Wright inequality is to get a bound similar to (1) in the Gaussian case. The extension to general subgaussian variables is then obtained with use of the by now standard tools of probability in Banach spaces, such as decoupling, symmetrization and the contraction principle. Via Chebyshev's inequality to obtain a tail estimate it is enough to bound appropriately the moments of quadratic forms in the case when $X_i = g_i$ are standard Gaussian $\mathcal{N}(0, 1)$ random variables. One may in fact show that (cf. [19,20])

$$\left(\mathbb{E} \left| \sum_{i,j=1}^{n} a_{ij}(g_i g_j - \delta_{ij}) \right|^p \right)^{1/p} \sim p\|A\|_{\mathrm{op}} + \sqrt{p}\|A\|_{\mathrm{HS}}, \tag{2}$$

where δ_{ij} is the Kronecker delta, and $\sim$ stands for a comparison up to universal multiplicative constants.

Following the same line of arguments, in order to extend the Hanson–Wright bound to the Banach space setting we first estimate moments of centered vector-valued Gaussian quadratic forms, i.e., quantities

$$\left\| \sum_{i,j=1}^{n} a_{ij}(g_i g_j - \delta_{ij}) \right\|_p = \left(\mathbb{E} \left\| \sum_{i,j=1}^{n} a_{ij}(g_i g_j - \delta_{ij}) \right\|^p \right)^{1/p}, \quad p \geq 1,$$

where $A = (a_{ij})_{i,j \leq n}$ is a symmetric matrix with values in a normed space $(F, \| \ \|)$. We note that (as mentioned above) there exist two-sided estimates for the moments of Gaussian quadratic forms with vector-valued coefficients. To the best of our knowledge they were obtained first in [7] and then they were reproved in various context by several authors (see, e.g., [5,22,24]). They state that for $p \geq 1$,

$$\left\| \sum_{ij} a_{ij}(g_i g_j - \delta_{ij}) \right\|_p \sim \mathbb{E} \left\| \sum_{ij} a_{ij}(g_i g_j - \delta_{ij}) \right\| + \sqrt{p}\, \mathbb{E} \sup_{\|x\|_2 \leq 1} \left\| \sum_{ij} a_{ij} x_i g_j \right\|$$

$$+ p \sup_{\|x\|_2 \leq 1, \|y\|_2 \leq 1} \left\| \sum_{ij} a_{ij} x_i y_j \right\|. \tag{3}$$

Unfortunately the second term on the right hand side of (3) is usually difficult to estimate. The main effort in this article will be to replace it by quantities which even if still involve expected values of Banach space valued random variables in many situations can be handled more easily. More precisely, we will obtain inequalities in which additional suprema over Euclidean spheres are placed outside the expectations, which reduces the complexity of the involved stochastic processes. As one of the consequences we will derive two-sided bounds in L_r spaces involving only purely deterministic quantities.

Our first observation is a simple lower bound

Proposition 1. *Let $(a_{ij})_{i,j \leq n}$ be a symmetric matrix with values in a normed space $(F, \| \cdot \|)$. Then for any $p \geq 1$ we have*

$$\left\| \sum_{ij} a_{ij}(g_i g_j - \delta_{ij}) \right\|_p \geq \frac{1}{C} \left(\mathbb{E} \left\| \sum_{ij} a_{ij}(g_i g_j - \delta_{ij}) \right\| + \sqrt{p} \sup_{\|x\|_2 \leq 1} \mathbb{E} \left\| \sum_{i \neq j} a_{ij} x_i g_j \right\| \right.$$

$$\left. + \sqrt{p} \sup_{\|(x_{ij})\|_2 \leq 1} \left\| \sum_{ij} a_{ij} x_{ij} \right\| + p \sup_{\|x\|_2 \leq 1, \|y\|_2 \leq 1} \left\| \sum_{ij} a_{ij} x_i y_j \right\| \right).$$

This motivates the following conjecture.

Conjecture 2. *Under the assumptions of Proposition* 1 *we have*

$$\left\| \sum_{ij} a_{ij}(g_i g_j - \delta_{ij}) \right\|_p \leq C \Bigg(\mathbb{E} \left\| \sum_{ij} a_{ij}(g_i g_j - \delta_{ij}) \right\| + \sqrt{p} \sup_{\|x\|_2 \leq 1} \mathbb{E} \left\| \sum_{i \neq j} a_{ij} x_i g_j \right\|$$
$$+ \sqrt{p} \sup_{\|(x_{ij})\|_2 \leq 1} \left\| \sum_{ij} a_{ij} x_{ij} \right\| + p \sup_{\|x\|_2 \leq 1, \|y\|_2 \leq 1} \left\| \sum_{ij} a_{ij} x_i y_j \right\| \Bigg).$$

We are able to show that the conjectured estimate holds up to logarithmic factors.

Theorem 3. *Let* $(a_{ij})_{i,j \leq n}$ *be a symmetric matrix with values in a normed space* $(F, \|\cdot\|)$. *Then for any* $p \geq 1$ *the following two estimates hold*

$$\left\| \sum_{ij} a_{ij}(g_i g_j - \delta_{ij}) \right\|_p \leq C \Bigg(\log(ep) \mathbb{E} \left\| \sum_{ij} a_{ij}(g_i g_j - \delta_{ij}) \right\| + \sqrt{p} \sup_{\|x\|_2 \leq 1} \mathbb{E} \left\| \sum_{i \neq j} a_{ij} x_i g_j \right\|$$
$$+ \sqrt{p} \sup_{\|(x_{ij})\|_2 \leq 1} \left\| \sum_{ij} a_{ij} x_{ij} \right\| + p \log(ep) \sup_{\|x\|_2 \leq 1, \|y\|_2 \leq 1} \left\| \sum_{ij} a_{ij} x_i y_j \right\| \Bigg) \tag{4}$$

and

$$\left\| \sum_{ij} a_{ij}(g_i g_j - \delta_{ij}) \right\|_p \leq C \Bigg(\mathbb{E} \left\| \sum_{ij} a_{ij}(g_i g_j - \delta_{ij}) \right\| + \sqrt{p} \sup_{\|x\|_2 \leq 1} \mathbb{E} \left\| \sum_{i \neq j} a_{ij} x_i g_j \right\|$$
$$+ \sqrt{p} \log(ep) \sup_{\|(x_{ij})\|_2 \leq 1} \left\| \sum_{ij} a_{ij} x_{ij} \right\| + p \sup_{\|x\|_2 \leq 1, \|y\|_2 \leq 1} \left\| \sum_{ij} a_{ij} x_i y_j \right\| \Bigg). \tag{5}$$

One of the main reasons behind the appearance of additional logarithmic factors is lack of good Sudakov-type estimates for Gaussian quadratic forms. Such bounds hold for linear forms and as a result we may show the following $((g_{ij})_{i,j \leq n}$ below denote as usual i.i.d. $\mathcal{N}(0, 1)$ random variables).

Theorem 4. *Under the assumptions of Theorem* 3 *we have*

$$\left\| \sum_{ij} a_{ij}(g_i g_j - \delta_{ij}) \right\|_p \leq C \Bigg(\mathbb{E} \left\| \sum_{ij} a_{ij}(g_i g_j - \delta_{ij}) \right\| + \mathbb{E} \left\| \sum_{i \neq j} a_{ij} g_{ij} \right\|$$
$$+ \sqrt{p} \sup_{\|x\|_2 \leq 1} \mathbb{E} \left\| \sum_{i \neq j} a_{ij} x_i g_j \right\| + \sqrt{p} \sup_{\|(x_{ij})\|_2 \leq 1} \left\| \sum_{ij} a_{ij} x_{ij} \right\|$$
$$+ p \sup_{\|x\|_2 \leq 1, \|y\|_2 \leq 1} \left\| \sum_{ij} a_{ij} x_i y_j \right\| \Bigg). \tag{6}$$

In particular we know that Conjecture 2 holds in Banach spaces, in which Gaussian quadratic forms dominate in mean Gaussian linear forms, i.e., in Banach spaces $(F, \|\,\|)$ for which there exists a constant $\tau < \infty$ such for any finite symmetric matrix $(a_{ij})_{i,j \leq n}$ with values in F one has

$$\mathbb{E} \left\| \sum_{i \neq j} a_{ij} g_{ij} \right\| \leq \tau \mathbb{E} \left\| \sum_{i \neq j} a_{ij} g_i g_j \right\|. \tag{7}$$

This condition has ben introduced in [31] as a one sided version of Pisier's contraction property (see [25]) in the context of multiple stochastic integration in the vector valued setting. It is easy to check (see Proposition 11 below) that it holds for L_r-spaces with $\tau = \tau(r)$, but the class of spaces satisfying it is much larger. For instance (7) holds in all type 2 spaces and for Banach lattices it is equivalent to finite cotype. We refer the reader to [13, Chapter 7] for a more detailed presentation.

Remark 5. Let us consider a non-centered Gaussian quadratic form $S = \sum_{i,j} a_{ij} g_i g_j$. By the triangle inequality for $p \geq 1$ one has $\|S\|_p \leq \|\mathbb{E}S\| + \|S - \mathbb{E}S\|_p$, whereas Jensen's inequality implies $\|S\|_p \geq \mathbb{E}\|S\| \geq \|\mathbb{E}S\|$ and as a consequence $\|S - \mathbb{E}S\|_p \leq 2\|S\|_p$. Thus $\|S\|_p \sim \|\mathbb{E}S\| + \|S - \mathbb{E}S\|_p$ and so Proposition 1 yields

$$\left\| \sum_{ij} a_{ij} g_i g_j \right\|_p \geq \frac{1}{C} \left(\mathbb{E} \left\| \sum_{ij} a_{ij} g_i g_j \right\| + \sqrt{p} \sup_{\|x\|_2 \leq 1} \mathbb{E} \left\| \sum_{i \neq j} a_{ij} x_i g_j \right\| \right.$$
$$\left. + \sqrt{p} \sup_{\|(x_{ij})\|_2 \leq 1} \left\| \sum_{ij} a_{ij} x_{ij} \right\| + p \sup_{\|x\|_2 \leq 1, \|y\|_2 \leq 1} \left\| \sum_{ij} a_{ij} x_i y_j \right\| \right)$$

and Theorem 4 implies

$$\left\| \sum_{ij} a_{ij} g_i g_j \right\|_p \leq C \left(\mathbb{E} \left\| \sum_{ij} a_{ij} g_i g_j \right\| + \mathbb{E} \left\| \sum_{i \neq j} a_{ij} g_{ij} \right\| + \sqrt{p} \sup_{\|x\|_2 \leq 1} \mathbb{E} \left\| \sum_{i \neq j} a_{ij} x_i g_j \right\| \right.$$
$$\left. + \sqrt{p} \sup_{\|(x_{ij})\|_2 \leq 1} \left\| \sum_{ij} a_{ij} x_{ij} \right\| + p \sup_{\|x\|_2 \leq 1, \|y\|_2 \leq 1} \left\| \sum_{ij} a_{ij} x_i y_j \right\| \right).$$

Proposition 1 and Theorem 4 may be expressed in terms of tails.

Theorem 6. *Let $(a_{ij})_{i,j \leq n}$ be a symmetric matrix with values in a normed space $(F, \|\cdot\|)$. Then for any $t > 0$,*

$$\mathbb{P}\left(\left\| \sum_{ij} a_{ij}(g_i g_j - \delta_{ij}) \right\| \geq t + \frac{1}{C} \mathbb{E} \left\| \sum_{ij} a_{ij}(g_i g_j - \delta_{ij}) \right\| \right) \geq \frac{1}{C} \exp\left(-C \min\left\{ \frac{t^2}{U^2}, \frac{t}{V} \right\} \right), \tag{8}$$

where

$$U = \sup_{\|x\|_2 \leq 1} \mathbb{E} \left\| \sum_{i \neq j} a_{ij} x_i g_j \right\| + \sup_{\|(x_{ij})\|_2 \leq 1} \left\| \sum_{ij} a_{ij} x_{ij} \right\|, \tag{9}$$

$$V = \sup_{\|x\|_2 \leq 1, \|y\|_2 \leq 1} \left\| \sum_{ij} a_{ij} x_i y_j \right\|. \tag{10}$$

Moreover, for $t > C(\mathbb{E}\| \sum_{ij} a_{ij}(g_i g_j - \delta_{ij})\| + \mathbb{E}\| \sum_{i \neq j} a_{ij} g_{ij}\|)$ we have

$$\mathbb{P}\left(\left\| \sum_{ij} a_{ij}(g_i g_j - \delta_{ij}) \right\| \geq t \right) \leq 2 \exp\left(-\frac{1}{C} \min\left\{ \frac{t^2}{U^2}, \frac{t}{V} \right\} \right). \tag{11}$$

As a corollary to Theorem 4 we get a Hanson–Wright-type inequality for Banach space valued quadratic forms in general independent subgaussian random variables. Its proof as well as proofs of Theorems 3, 4 and 6 is presented in Section 6.

Theorem 7. *Let $X_1, X_2, \ldots, X_n$ be independent mean zero α-subgaussian random variables. Then for any symmetric matrix $(a_{ij})_{i,j \leq n}$ with values in a normed space $(F, \|\cdot\|)$ and $t > C\alpha^2(\mathbb{E}\| \sum_{ij} a_{ij}(g_i g_j - \delta_{ij})\| + \mathbb{E}\| \sum_{i \neq j} a_{ij} g_{ij}\|)$ we have*

$$\mathbb{P}\left(\left\| \sum_{ij} a_{ij}(X_i X_j - \mathbb{E}(X_i X_j)) \right\| \geq t \right) \leq 2 \exp\left(-\frac{1}{C} \min\left\{ \frac{t^2}{\alpha^4 U^2}, \frac{t}{\alpha^2 V} \right\} \right), \tag{12}$$

where U and V are as in Theorem 6.

Remark 8. It is not hard to check that in the case $F = \mathbb{R}$ we have $U \sim \|(a_{ij})\|_{\text{HS}}$ and $V = \|(a_{ij})\|_{\text{op}}$. Moreover,

$$\mathbb{E} \left\| \sum_{ij} a_{ij}(g_i g_j - \delta_{ij}) \right\| + \mathbb{E} \left\| \sum_{i \neq j} a_{ij} g_{ij} \right\| \leq (\sqrt{2} + 1) \|(a_{ij})\|_{\text{HS}},$$

so for $t < C'\alpha^2(\mathbb{E}\|\sum_{ij} a_{ij}(g_i g_j - \delta_{ij})\| + \mathbb{E}\|\sum_{i \neq j} a_{ij} g_{ij}\|)$ and sufficiently large C the right hand side of (12) is at least 1. Hence (12) holds for any $t > 0$ in the real case and is equivalent to the Hanson–Wright bound.

Remark 9. Proposition 22 below shows that we may replace in all estimates above the term $\sup_{\|x\|_2 \leq 1} \mathbb{E}\|\sum_{i \neq j} a_{ij} x_i g_j\|$ by $\sup_{\|x\|_2 \leq 1} \mathbb{E}\|\sum_{ij} a_{ij} x_i g_j\|$.

Remark 10. We are able to derive similar estimates as discussed in this paper for Banach space valued Gaussian chaoses of arbitrary degree. Formulas are however more complicated and the proof is more technical. For these reasons we decided to include details in a separate forthcoming paper [3].

The organization of the paper is as follows. In the next section we discuss a few corollaries of Theorems 4 and 7. In Section 3 we prove Proposition 1 and show that it is enough to bound separately moments of diagonal and off-diagonal parts of chaoses. In Section 4 we reduce Theorems 3 and 4 to the problem of estimating means of suprema of certain Gaussian processes. In Section 5 we show how to bound expectations of such suprema – the main new ingredient are entropy bounds presented in Corollary 27 (derived via volumetric-type arguments). Unfortunately our entropy bounds are too weak to use the Dudley integral bound. Instead, we present a technical chaining argument (of similar type as in [20]). In the last section we conclude the proofs of main Theorems.

2. Consequences and extensions

2.1. L_r-Spaces

We will now restrict our attention to L_r spaces for $1 \leq r < \infty$. It turns out that in this case one may express all parameters of Theorem 4 without any expectations as is shown in the proposition below. As a corollary we will obtain a version of the Hanson–Wright inequality in terms of only deterministic quantities, in particular without the additional term $\mathbb{E}\|\sum_{i \neq j} a_{ij} g_{ij}\|$, which shows that Conjecture 2 is satisfied in these spaces (with constants depending on r). We note that the following proposition implies also that L_r spaces satisfy (7) with $\tau = Cr$. In fact, using the general theory presented in [13], one can improve this to $\tau = C\sqrt{r}$, however we are not going to use directly any of these bounds on τ.

Proposition 11. *For any symmetric matrix $(a_{ij})_{i,j \leq n}$ with values in $L_r = L_r(X, \mu)$, $1 \leq r < \infty$ and $x_1, \ldots, x_n \in \mathbb{R}$ we have*

$$\frac{1}{C}\left\|\sqrt{\sum_{ij} a_{ij}^2}\right\|_{L_r} \leq \mathbb{E}\left\|\sum_{ij} a_{ij} g_{ij}\right\|_{L_r} \leq C\sqrt{r}\left\|\sqrt{\sum_{ij} a_{ij}^2}\right\|_{L_r}, \tag{13}$$

$$\frac{1}{C}\left\|\sqrt{\sum_j \left(\sum_i a_{ij} x_i\right)^2}\right\|_{L_r} \leq \mathbb{E}\left\|\sum_{ij} a_{ij} x_i g_j\right\|_{L_r} \leq C\sqrt{r}\left\|\sqrt{\sum_j \left(\sum_i a_{ij} x_i\right)^2}\right\|_{L_r}, \tag{14}$$

$$\frac{1}{C\sqrt{r}}\left\|\sqrt{\sum_{ij} a_{ij}^2}\right\|_{L_r} \leq \mathbb{E}\left\|\sum_{ij} a_{ij}(g_i g_j - \delta_{ij})\right\|_{L_r} \leq Cr\left\|\sqrt{\sum_{ij} a_{ij}^2}\right\|_{L_r}. \tag{15}$$

Proof. For any a_i's in L_r the Gaussian concentration yields

$$\mathbb{E}\left\|\sum_i a_i g_i\right\|_{L_r} \leq \left(\mathbb{E}\left\|\sum_i a_i g_i\right\|_{L_r}^r\right)^{1/r} \leq C\sqrt{r}\mathbb{E}\left\|\sum_i a_i g_i\right\|_{L_r}.$$

Since

$$\left(\mathbb{E}\left\|\sum_i a_i g_i\right\|_{L_r}^r\right)^{1/r} = \left(\int_X \mathbb{E}\left|\sum_i a_i(x) g_i\right|^r d\mu(x)\right)^{1/r}$$

$$= \left(\int_X \mathbb{E}|g_1|^r \left(\sum_i a_i^2(x)\right)^{r/2} d\mu(x)\right)^{1/r} \sim \sqrt{r}\left\|\sqrt{\sum_i a_i^2}\right\|_{L_r},$$

estimates (13), (14) follow easily. The proof of (15) is analogous. It is enough to observe that (3) yields that

$$\mathbb{E}\left\|\sum_{ij} a_{ij}(g_i g_j - \delta_{ij})\right\|_{L_r} \leq \left(\mathbb{E}\left\|\sum_{ij} a_{ij}(g_i g_j - \delta_{ij})\right\|_{L_r}^r\right)^{1/r} \leq Cr\mathbb{E}\left\|\sum_{ij} a_{ij}(g_i g_j - \delta_{ij})\right\|_{L_r}$$

and (2) implies for any $x \in X$,

$$\frac{\sqrt{r}}{C}\sqrt{\sum_{ij} a_{ij}^2(x)} \leq \left(\mathbb{E}\left|\sum_{ij} a_{ij}(x)(g_i g_j - \delta_{ij})\right|^r\right)^{1/r} \leq Cr\sqrt{\sum_{ij} a_{ij}^2(x)}.$$

$\square$

The above proposition, together with Proposition 1 and Theorem 4 immediately yield the following corollary (in particular they imply that Conjecture 2 holds in L_r spaces with r-dependent constants). Below $\sim^r$ denotes comparison up to constants depending only on r.

Corollary 12. *For any symmetric matrix $(a_{ij})_{ij}$ with values in L_r and $p \geq 1$ we have*

$$\left\|\sum_{ij} a_{ij}(g_i g_j - \delta_{ij})\right\|_p \sim^r \left\|\sqrt{\sum_{ij} a_{ij}^2}\right\|_{L_r} + \sqrt{p}\sup_{\|x\|_2 \leq 1}\left\|\sqrt{\sum_j\left(\sum_{i \neq j} a_{ij}x_i\right)^2}\right\|_{L_r} + \sqrt{p}\sup_{\|(x_{ij})\|_2 \leq 1}\left\|\sum_{ij} a_{ij}x_{ij}\right\|_{L_r}$$

$$+ p\sup_{\|x\|_2 \leq 1, \|y\|_2 \leq 1}\left\|\sum_{ij} a_{ij}x_i y_j\right\|_{L_r}.$$

The implicit constants in the estimates for moments can be taken to be equal to Cr in the upper bound and $r^{-1/2}/C$ in the lower bound.

Theorem 7 and Proposition 11 imply the following Hanson–Wright-type estimate in L_r-spaces.

Corollary 13. *Let $X_1, X_2, \ldots, X_n$ be independent mean zero α-subgaussian random variables. Then for any symmetric finite matrix $(a_{ij})_{i,j \leq n}$ with values in $L_r = L_r(X, \mu)$, $1 \leq r < \infty$ and $t > C\alpha^2 r\|\sqrt{\sum_{ij} a_{ij}^2}\|_{L_r}$ we have*

$$\mathbb{P}\left(\left\|\sum_{ij} a_{ij}(X_i X_j - \mathbb{E}(X_i X_j))\right\|_{L_r} \geq t\right) \leq 2\exp\left(-\frac{1}{C}\min\left\{\frac{t^2}{\alpha^4 r U^2}, \frac{t}{\alpha^2 V}\right\}\right), \tag{16}$$

where

$$U = \sup_{\|x\|_2 \leq 1}\left\|\sqrt{\sum_j\left(\sum_{i \neq j} a_{ij}x_i\right)^2}\right\|_{L_r} + \sup_{\|(x_{ij})\|_2 \leq 1}\left\|\sum_{ij} a_{ij}x_{ij}\right\|_{L_r},$$

$$V = \sup_{\|x\|_2 \leq 1, \|y\|_2 \leq 1}\left\|\sum_{ij} a_{ij}x_i y_j\right\|_{L_r}.$$

2.2. *Spaces of type 2*

Recall that a normed space F is of type 2 with constant λ if for every positive integer n and $v_1, \ldots, v_n \in F$,

$$\mathbb{E}\left\|\sum_{i=1}^n v_i \varepsilon_i\right\| \leq \lambda\sqrt{\sum_{i=1}^n \|v_i\|^2}, \tag{17}$$

where $\varepsilon_1, \varepsilon_2, \ldots$ is a sequence of independent Rademacher variables. The importance of this class of spaces stems from the fact that if F is of type 2 then one can easily bound from above expectations of linear combinations or more generally multilinear forms in independent centered random variables with F-valued coefficients. Let us now introduce the basic (by now classical) tools allowing to prove such estimates and show how they can be combined with Theorem 7 to yield a

tail inequality for quadratic forms with the right hand side expressed in terms of deterministic quantities (i.e., not involving expectations of Banach space valued random variables).

The first tool we are going to use is the following symmetrization inequality (see, e.g., [24, Lemma 6.3] for a more general formulation).

Lemma 14. *Let $Y_1, \ldots, Y_n$ be independent random variables with values in a Banach space F. Let moreover $\varepsilon_1, \ldots, \varepsilon_n$ be i.i.d. Rademacher variables independent of $(Y_i)_{i=1}^n$. Then for any $p \geq 1$,*

$$\left\| \sum_{i=1}^n (Y_i - \mathbb{E}Y_i) \right\|_p \leq 2 \left\| \sum_{i=1}^n \varepsilon_i Y_i \right\|_p.$$

The other fact we need is a decoupling inequality for quadratic forms. For symmetric random variables it has been proved in [17], the general case can be immediately obtained from a decoupling result for U-statistics, proved in [10].

Theorem 15. *Let $X_1, \ldots, X_n$ be a sequence of independent random variables and let $X'_1, \ldots, X'_n$ be its independent copy. Then for any symmetric matrix $(a_{ij})_{i,j \leq n}$ with coefficients in a normed space $(F, \|\cdot\|)$ and any $p \geq 1$,*

$$\frac{1}{C} \left\| \sum_{i \neq j} a_{ij} X_i X'_j \right\|_p \leq \left\| \sum_{i \neq j} a_{ij} X_i X_j \right\|_p \leq C \left\| \sum_{i \neq j} a_{ij} X_i X'_j \right\|_p. \tag{18}$$

The advantage of working with bilinear forms in two independent sequences of random variables instead of quadratic forms in a single sequence is that one can view the former conditionally as a sum of independent random variables, which allows to use a variety of classical tools.

Let us now return to type 2 spaces and show how the above lemmas can be combined with condition (17) in order to bound the expectations appearing in Theorem 7.

First, by Lemma 14 and (17) one easily obtains that if F is of type 2 with constant λ then for any independent random variables X_i,

$$\mathbb{E} \left\| \sum_i a_i \left(X_i^2 - \mathbb{E}X_i^2 \right) \right\| \leq 2\mathbb{E} \left\| \sum_i a_i \varepsilon_i X_i^2 \right\| \leq 2\lambda \mathbb{E} \sqrt{\sum_i \|a_i\|^2 X_i^4} \leq 2\lambda \sqrt{\sum_i \|a_i\|^2 \mathbb{E}X_i^4},$$

where $\varepsilon_1, \ldots, \varepsilon_n$ are independent Rademacher variables, independent of $X_1, \ldots, X_n$.

Moreover, if $\mathbb{E}X_i = 0$, then using the notation from Theorem 15 and denoting now by $\varepsilon_1, \varepsilon'_1, \ldots, \varepsilon_n, \varepsilon'_n$ independent Rademacher variables, independent of $X_1, X'_1, \ldots, X_n, X'_n$, we obtain

$$\left(\mathbb{E} \left\| \sum_{i \neq j} a_{ij} X_i X_j \right\| \right)^2 \leq \mathbb{E} \left\| \sum_{i \neq j} a_{ij} X_i X_j \right\|^2 \leq C \mathbb{E} \left\| \sum_{i \neq j} a_{ij} X_i X'_j \right\|^2$$

$$\leq 16 C \mathbb{E} \left\| \sum_{i \neq j} a_{ij} \varepsilon_i \varepsilon'_j X_i X'_j \right\|^2 \leq 32 C \lambda^2 \sum_i \mathbb{E}(X_i^2) \mathbb{E} \left\| \sum_{j: j \neq i} a_{ij} \varepsilon'_j X'_j \right\|^2$$

$$\leq 64 C \lambda^4 \sum_{i \neq j} \|a_{ij}\|^2 \mathbb{E}X_i^2 \mathbb{E}X_j^2.$$

The second inequality above is an application of Theorem 15, the third one follows from an iterated conditional application of Lemma 14 and the last two by a conditional application of (17) together with the Khintchine-Kahane inequality $\| \sum_i v_i \varepsilon_i \|_2 \leq \sqrt{2} \| \sum_i v_i \varepsilon_i \|_1$ ([14], see [21] for the optimal constant $\sqrt{2}$).

Combining the above estimates with Theorem 7 yields immediately the following

Corollary 16. *Let $X_1, X_2, \ldots, X_n$ be independent mean zero α-subgaussian random variables and let F be a normed space of type two constant λ. Then for any symmetric finite matrix $(a_{ij})_{i,j \leq n}$ with values in F and $t > C\lambda^2 \alpha^2 \sqrt{\sum_{ij} \|a_{ij}\|^2}$ we have*

$$\mathbb{P} \left(\left\| \sum_{ij} a_{ij} (X_i X_j - \mathbb{E}(X_i X_j)) \right\| \geq t \right) \leq 2 \exp \left(-\frac{1}{C} \min \left\{ \frac{t^2}{\alpha^4 U^2}, \frac{t}{\alpha^2 V} \right\} \right), \tag{19}$$

where

$$U = \lambda \sup_{\|x\|_2 \leq 1} \sqrt{\sum_j \left\| \sum_{i \neq j} a_{ij} x_i \right\|^2} + \sup_{\|(x_{ij})\|_2 \leq 1} \left\| \sum_{ij} a_{ij} x_{ij} \right\|,$$

$$V = \sup_{\|x\|_2 \leq 1, \|y\|_2 \leq 1} \left\| \sum_{ij} a_{ij} x_i y_j \right\|.$$

Remark 17. As already mentioned in the introduction, type 2 spaces satisfy (7) with $\tau \leq \lambda$ (cf. [13, Theorem 7.1.20]) and thus also Conjecture 2 with a constant depending on λ.

Remark 18. We note that from Theorem 7 one can also derive similar inequalities for suprema of quadratic forms over VC-type classes of functions appearing, e.g., in the analysis of randomized U-processes (cf., e.g., [9, Chapter 5.4]).

2.3. *Random vectors with dependencies*

Let us now consider a random vector $X = (X_1, \ldots, X_n)$, possibly with dependent coordinates, and assume that there exists an α-Lipschitz map $\psi \colon \mathbb{R}^n \to \mathbb{R}^n$ such that X has the same distribution as $\psi(G)$, where G is a standard Gaussian vector in $\mathbb{R}^n$. An important class of vectors X with this property is provided by the celebrated Caffarelli contraction principle [8], which asserts that the map ψ exists if X has density of the form e^{-V} with $\nabla^2 V \geq \alpha^{-2} \mathrm{Id}$. As observed by Ledoux and Oleszkiewicz [23, Corollary 1], by combining the well known comparison result due to Pisier [26] with a stochastic domination-type argument, one gets that for any smooth function $f \colon \mathbb{R}^n \to F$, and any $p \geq 1$,

$$\left\| f(X) - \mathbb{E} f(X) \right\|_p \leq \frac{\pi \alpha}{2} \left\| \langle \nabla f(X), G \rangle \right\|_p, \tag{20}$$

where here and subsequently G_n is a standard Gaussian vector in $\mathbb{R}^n$ independent of X and for $a \in F^n$, $b \in \mathbb{R}^n$ we denote $\langle a, b \rangle = \sum_{i=1}^n a_i b_i$. This inequality together with Theorem 4 allow us to implement a simple argument from [4] and obtain inequalities for quadratic forms and more general F-valued functions of the random vector X. Below, we will denote the second partial derivatives of f by $\partial_{ij} f$. For the sake of brevity, we will focus on moment estimates, clearly tail bounds follow from them by an application of the Chebyshev inequality.

Corollary 19. *Let X be an α-Lipschitz image of a standard Gaussian vector in $\mathbb{R}^n$ and let $f \colon \mathbb{R}^n \to F$ be a function with bounded derivatives of order two. Assume moreover that $\mathbb{E} \nabla f(X) = 0$. Then for any $p \geq 2$,*

$$\left\| f(X) - \mathbb{E} f(X) \right\|_p \leq C \alpha^2 \sup_{z \in \mathbb{R}^n} \Bigg(\mathbb{E} \left\| \sum_{ij} \partial_{ij} f(z)(g_i g_j - \delta_{ij}) \right\| + \mathbb{E} \left\| \sum_{i \neq j} \partial_{ij} f(z) g_{ij} \right\|$$

$$+ \sqrt{p} \sup_{\|x\|_2 \leq 1} \mathbb{E} \left\| \sum_{i \neq j} \partial_{ij} f(z) x_i g_j \right\| + \sqrt{p} \sup_{\|(x_{ij})\|_2 \leq 1} \left\| \sum_{ij} \partial_{ij} f(z) x_{ij} \right\|$$

$$+ p \sup_{\|x\|_2 \leq 1, \|y\|_2 \leq 1} \left\| \sum_{ij} \partial_{ij} f(z) x_i y_j \right\| \Bigg). \tag{21}$$

In particular if X is of mean zero, then

$$\left\| \sum_{ij} a_{ij} \big(X_i X_j - \mathbb{E}(X_i X_j) \big) \right\|_p \leq C \alpha^2 \Bigg(\mathbb{E} \left\| \sum_{ij} a_{ij} (g_i g_j - \delta_{ij}) \right\| + \mathbb{E} \left\| \sum_{i \neq j} a_{ij} g_{ij} \right\|$$

$$+ \sqrt{p} \sup_{\|x\|_2 \leq 1} \mathbb{E} \left\| \sum_{i \neq j} a_{ij} x_i g_j \right\| + \sqrt{p} \sup_{\|(x_{ij})\|_2 \leq 1} \left\| \sum_{ij} a_{ij} x_{ij} \right\|$$

$$+ p \sup_{\|x\|_2 \leq 1, \|y\|_2 \leq 1} \left\| \sum_{ij} a_{ij} x_i y_j \right\| \Bigg) \tag{22}$$

and the inequality (12) *is satisfied.*

Proof of Corollary 19. Let $G_n = (g_1, \ldots, g_n)$, $G'_n = (g'_1, \ldots, g'_n)$ be independent standard Gaussian vectors in $\mathbb{R}^n$, independent of X. By an iterated application of (20) (the second time conditionally on G_n) we have

$$\mathbb{E}\|f(X) - \mathbb{E}f(X)\|^p \le C^p \alpha^p \mathbb{E}\|\langle \nabla f(X), G_n \rangle\|^p \le C^{2p} \alpha^{2p} \mathbb{E}\left\|\sum_{ij} \partial_{ij} f(X) g_i g'_j\right\|^p$$

$$\le \tilde{C}^{2p} \alpha^{2p} \mathbb{E}\left\|\sum_{ij} \partial_{ij} f(X)(g_i g_j - \delta_{ij})\right\|^p,$$

where the last inequality follows by [5, Theorem 2.2]. To finish the proof of (21) it is now enough to apply Theorem 4 conditionally on X and replace the expectation in X by the supremum over $z \in \mathbb{R}^n$.

The inequality (22) follows by a direct application of (21). □

3. Lower bounds

In this part we show Proposition 1 and the lower bound in Theorem 6. We start with a simple lemma.

Lemma 20. Let $W = \|\sum_{i \ne j} a_{ij} g_i g_j\|_p + \|\sum_i a_{ii}(g_i^2 - 1)\|_p$. Then for any $p \ge 1$,

$$\frac{1}{3}W \le \left\|\sum_{ij} a_{ij}(g_i g_j - \delta_{ij})\right\|_p \le W.$$

Proof. Let $(\varepsilon_i)_i$ be a sequence of i.i.d. Rademacher random variables independent of $(g_i)_i$. We have by symmetry of g_i and Jensen's inequality,

$$\left\|\sum_{ij} a_{ij}(g_i g_j - \delta_{ij})\right\|_p = \left\|\sum_{ij} a_{ij}(\varepsilon_i \varepsilon_j g_i g_j - \delta_{ij})\right\|_p \ge \left\|\mathbb{E}_\varepsilon \sum_{ij} a_{ij}(\varepsilon_i \varepsilon_j g_i g_j - \delta_{ij})\right\|_p$$

$$= \left\|\sum_i a_{ii}(g_i^2 - 1)\right\|_p.$$

To conclude we use the triangle inequality in L_p and get

$$\left\|\sum_{i \ne j} a_{ij} g_i g_j\right\|_p \le \left\|\sum_{ij} a_{ij}(g_i g_j - \delta_{ij})\right\|_p + \left\|\sum_i a_{ii}(g_i^2 - 1)\right\|_p \le 2\left\|\sum_{ij} a_{ij}(g_i g_j - \delta_{ij})\right\|_p.$$

Adding the inequalities above yields the first estimate of the lemma. The second one follows trivially from the triangle inequality. □

Proof of Proposition 1. Obviously

$$\left\|\sum_{ij} a_{ij}(g_i g_j - \delta_{ij})\right\|_p \ge \mathbb{E}\left\|\sum_{ij} a_{ij}(g_i g_j - \delta_{ij})\right\|.$$

Moreover, denoting by $\|\cdot\|_*$ the norm in the dual of F, we have

$$\left\|\sum_{ij} a_{ij}(g_i g_j - \delta_{ij})\right\|_p \ge \sup_{\|\varphi\|_* \le 1}\left\|\sum_{ij} \varphi(a_{ij})(g_i g_j - \delta_{ij})\right\|_p$$

$$\ge \frac{1}{C}\left(\sqrt{p}\sup_{\|\varphi\|_* \le 1}\left\|(\varphi(a_{ij}))_{ij}\right\|_{HS} + p\sup_{\|\varphi\|_* \le 1}\left\|(\varphi(a_{ij}))_{ij}\right\|_{op}\right)$$

$$= \frac{1}{C}\left(\sqrt{p}\sup_{\|(x_{ij})\|_2 \le 1}\left\|\sum_{ij} a_{ij} x_{ij}\right\| + p\sup_{\|x\|_2 \le 1, \|y\|_2 \le 1}\left\|\sum_{ij} a_{ij} x_i y_j\right\|\right),$$

where in the second inequality we used (2).

Lemma 20 and the decoupling inequality (18) yield

$$\left\|\sum_{ij} a_{ij}(g_i g_j - \delta_{ij})\right\|_p \geq \frac{1}{3}\left\|\sum_{i \neq j} a_{ij} g_i g_j\right\|_p \geq \frac{1}{C}\left\|\sum_{i \neq j} a_{ij} g_i g_j'\right\|_p, \tag{23}$$

where $(g_i')_i$ denotes an independent copy of $(g_i)_i$.

For any finite sequence $(b_i)_i$ in $(F, \|\cdot\|)$ we have

$$\left\|\sum_i b_i g_i\right\|_p \geq \sup_{\|\varphi\|_* \leq 1}\left\|\sum_i \varphi(b_i) g_i\right\|_p = \sup_{\|\varphi\|_* \leq 1}\left\|(\varphi(b_i))_i\right\|_2 \cdot \|g_1\|_p \geq \frac{\sqrt{p}}{C} \sup_{\|x\|_2 \leq 1}\left\|\sum_i x_i b_i\right\|. \tag{24}$$

Thus, by (23) and the Fubini Theorem, we get

$$\left\|\sum_{ij} a_{ij}(g_i g_j - \delta_{ij})\right\|_p \geq \frac{\sqrt{p}}{C} \sup_{\|x\|_2 \leq 1}\left\|\sum_{i \neq j} a_{ij} x_i g_j'\right\|_p \geq \frac{\sqrt{p}}{C} \sup_{\|x\|_2 \leq 1} \mathbb{E}\left\|\sum_{i \neq j} a_{ij} x_i g_j\right\|. \qquad \square$$

4. Reduction to a bound on the supremum of a Gaussian process

In this section we will reduce the upper estimates of Theorems 3 and 4 to an estimate on the expected value of a supremum of a certain Gaussian process. The arguments in this part of the article are well-known, we present them for the sake of completeness. In particular we will demonstrate the upper bounds given in (3).

The first lemma shows that we may easily bound the diagonal terms.

Lemma 21. *For $p \geq 1$ we have*

$$\left\|\sum_i a_{ii}(g_i^2 - 1)\right\|_p \leq C\left(\mathbb{E}\left\|\sum_{ij} a_{ij}(g_i g_j - \delta_{ij})\right\| + \sqrt{p} \sup_{\|(x_{ij})\|_2 \leq 1}\left\|\sum_{ij} a_{ij} x_{ij}\right\| + p \sup_{\|x\|_2 \leq 1, \|y\|_2 \leq 1}\left\|\sum_{ij} a_{ij} x_i y_j\right\|\right).$$

Proof. Let X_i be a sequence of i.i.d. standard symmetric exponential r.v.'s. A simple argument (cf. proof of Lemma 9.5 in [2]) shows that

$$\left\|\sum_i a_{ii}(g_i^2 - 1)\right\|_p \sim \left\|\sum_i a_{ii} g_i g_i'\right\|_p \sim \left\|\sum_i a_{ii} X_i\right\|_p. \tag{25}$$

The latter quantity was bounded in [18, Theorem 1], thus

$$\left\|\sum_i a_{ii}(g_i^2 - 1)\right\|_p \sim \left\|\sum_i a_{ii}(g_i^2 - 1)\right\|_1 + \sqrt{p} \sup_{\|x\|_2 \leq 1}\left\|\sum_i a_{ii} x_i\right\| + p \sup_i \|a_{ii}\|$$

$$\leq C\left(\mathbb{E}\left\|\sum_{ij} a_{ij}(g_i g_j - \delta_{ij})\right\| + \sqrt{p} \sup_{\|(x_{ij})\|_2 \leq 1}\left\|\sum_{ij} a_{ij} x_{ij}\right\| + p \sup_{\|x\|_2 \leq 1, \|y\|_2 \leq 1}\left\|\sum_{ij} a_{ij} x_i y_j\right\|\right),$$

where in the last inequality we used Lemma 20. $\qquad \square$

From the next proposition it follows that in all our main results we can replace the term $\sup_{\|x\|_2 \leq 1} \mathbb{E}\|\sum_{i \neq j} a_{ij} x_i g_j\|$ by $\sup_{\|x\|_2 \leq 1} \mathbb{E}\|\sum_{ij} a_{ij} x_i g_j\|$.

Proposition 22. *Under the assumption of Proposition 1 we have for $p \geq 1$,*

$$\sqrt{p} \sup_{\|x\|_2 \leq 1} \mathbb{E}\left\|\sum_i a_{ii} x_i g_i\right\| \leq C\left(\mathbb{E}\left\|\sum_{ij} a_{ij}(g_{ij} - \delta_{ij})\right\| + \sqrt{p} \sup_{\|(x_{ij})\|_2 \leq 1}\left\|\sum_{ij} a_{ij} x_{ij}\right\|\right.$$

$$\left. + p \sup_{\|x\|_2 \leq 1, \|y\|_2 \leq 1}\left\|\sum_{ij} a_{ij} x_i y_j\right\|\right).$$

Proof. Let $(g_i')_i$ be an independent copy of the sequence $(g_i)_i$. Denoting by $\mathbb{E}'$ the expectation with respect to the variables $(g_i')_i$, we may estimate

$$\sqrt{p}\sup_{\|x\|_2\leq 1}\mathbb{E}\left\|\sum_i a_{ii}x_i g_i\right\|\leq \mathbb{E}\sqrt{p}\sup_{\|x\|_2\leq 1}\left\|\sum_i a_{ii}x_i g_i\right\|\leq C\mathbb{E}\left(\mathbb{E}'\left\|\sum_i a_{ii}g_i g_i'\right\|^p\right)^{1/p}$$

$$\leq C\left\|\sum_i a_{ii}g_i g_i'\right\|_p\leq C\left\|\sum_i a_{ii}(g_i^2-1)\right\|_p,$$

where the second inequality follows from (24) applied conditionally on $(g_i)_i$, the third one from Jensen's inequality and the last one from (25). The assertion of the proposition follows now by Lemma 21. $\square$

For the off-diagonal terms we use first the concentration approach.

Proposition 23. *For $p\geq 1$ we have*

$$\left\|\sum_{i\neq j}a_{ij}g_i g_j\right\|_p\leq C\left(\mathbb{E}\left\|\sum_{i\neq j}a_{ij}g_i g_j\right\|+\sqrt{p}\mathbb{E}\sup_{\|x\|_2\leq 1}\left\|\sum_{i\neq j}a_{ij}x_i g_j\right\|+p\sup_{\|x\|_2\leq 1,\|y\|_2\leq 1}\left\|\sum_{ij}a_{ij}x_i y_j\right\|\right).$$

Proof. Let

$$A:=\left\{z\in\mathbb{R}^n:\left\|\sum_{i\neq j}a_{ij}z_i z_j\right\|\leq 4\mathbb{E}\left\|\sum_{i\neq j}a_{ij}g_i g_j\right\|,\ \sup_{\|x\|_2\leq 1}\left\|\sum_{i\neq j}a_{ij}x_i z_j\right\|\leq 4\mathbb{E}\sup_{\|x\|_2\leq 1}\left\|\sum_{i\neq j}a_{ij}x_i g_j\right\|\right\}.$$

Then $\gamma_n(A)\geq\frac{1}{2}$ by the Chebyshev inequality. Gaussian concentration gives $\gamma_n(A+tB_2^n)\geq 1-e^{-t^2/2}$ for $t\geq 0$. It is easy to check that for $z\in A+tB_2^n$ we have

$$\left\|\sum_{i\neq j}a_{ij}z_i z_j\right\|\leq 4S(t),$$

where

$$S(t)=\mathbb{E}\left\|\sum_{i\neq j}a_{ij}g_i g_j\right\|+2t\mathbb{E}\sup_{\|x\|_2\leq 1}\left\|\sum_{i\neq j}a_{ij}x_i g_j\right\|+t^2\sup_{\|x\|_2\leq 1,\|y\|_2\leq 1}\left\|\sum_{i\neq j}a_{ij}x_i y_j\right\|.$$

So

$$\mathbb{P}\left(\left\|\sum_{i\neq j}a_{ij}g_i g_j\right\|>4S(t)\right)\leq e^{-t^2/2}\quad\text{for }t\geq 0.$$

Integrating by parts we get $\|\sum_{i\neq j}a_{ij}g_i g_j\|_p\leq CS(\sqrt{p})$ for $p\geq 1$, which ends the proof, since

$$\sup_{\|x\|_2\leq 1,\|y\|_2\leq 1}\left\|\sum_{i\neq j}a_{ij}x_i y_j\right\|\leq\sup_{\|x\|_2\leq 1,\|y\|_2\leq 1}\left\|\sum_{i,j}a_{ij}x_i y_j\right\|+\max_i\|a_{ii}\|\leq 2\sup_{\|x\|_2\leq 1,\|y\|_2\leq 1}\left\|\sum_{i,j}a_{ij}x_i y_j\right\|.\qquad(26)$$

$\square$

For any symmetric matrix by (18) we have $\mathbb{E}\|\sum_{i\neq j}a_{ij}g_i g_j\|\sim\mathbb{E}\|\sum_{i\neq j}a_{ij}g_i g_j'\|$. Moreover introducing decoupled chaos enables us to release the assumptions of the symmetry of the matrix and zero diagonal.

Taking into account the above observations, Conjecture 2 reduces to the statement that for any $p\geq 1$ and any finite matrix (a_{ij}) in $(F,\|\cdot\|)$ we have

$$\mathbb{E}\sup_{\|x\|_2\leq 1}\left\|\sum_{ij}a_{ij}g_i x_j\right\|\leq C\left(\frac{1}{\sqrt{p}}\mathbb{E}\left\|\sum_{ij}a_{ij}g_i g_j'\right\|+\sup_{\|x\|_2\leq 1}\mathbb{E}\left\|\sum_{ij}a_{ij}g_i x_j\right\|\right.$$

$$\left.+\sup_{\|(x_{ij})\|_2\leq 1}\left\|\sum_{ij}a_{ij}x_{ij}\right\|+\sqrt{p}\sup_{\|x\|_2\leq 1,\|y\|_2\leq 1}\left\|\sum_{ij}a_{ij}x_i y_j\right\|\right).\qquad(27)$$

Let us rewrite (27) in another language. We may assume that $F = \mathbb{R}^m$ for some finite m and $a_{ij} = (a_{ijk})_{k \leq m}$. Let $T = B_{F^*}$ be the unit ball in the dual space F^*. Then (27) takes the following form.

Conjecture 24. *Let $p \geq 1$. Then for any triple indexed matrix $(a_{ijk})_{i,j \leq n, k \leq m}$ and bounded nonempty set $T \subset \mathbb{R}^m$ we have*

$$
\mathbb{E} \sup_{\|x\|_2 \leq 1, t \in T} \left| \sum_{ijk} a_{ijk} g_i x_j t_k \right| \leq C \left(\frac{1}{\sqrt{p}} \mathbb{E} \sup_{t \in T} \left| \sum_{ijk} a_{ijk} g_i g'_j t_k \right| \right.
$$

$$
+ \sup_{\|x\|_2 \leq 1} \mathbb{E} \sup_{t \in T} \left| \sum_{ijk} a_{ijk} g_i x_j t_k \right| + \sup_{t \in T} \left(\sum_{ij} \left(\sum_k a_{ijk} t_k \right)^2 \right)^{1/2}
$$

$$
\left. + \sqrt{p} \sup_{\|x\|_2 \leq 1, t \in T} \left(\sum_i \left(\sum_{jk} a_{ij} x_j t_k \right)^2 \right)^{1/2} \right). \tag{28}
$$

Obviously it is enough to show this for finite sets T.

5. Estimating suprema of Gaussian processes

To estimate the supremum of a centered Gaussian process $(G_v)_{v \in V}$ one needs to study the distance on V given by $d(v, v') := (\mathbb{E}|G_v - G_{v'}|^2)^{1/2}$ (cf. [30]). In the case of the Gaussian process from Conjecture 24 this distance is defined on $B_2^n \times T \subset \mathbb{R}^n \times \mathbb{R}^m$ by the formula

$$
d_A\big((x, t), (x', t')\big) := \left(\sum_i \left(\sum_{jk} a_{ijk} (x_j t_k - x'_j t'_k) \right)^2 \right)^{1/2} = \alpha_A(x \otimes t - x' \otimes t'),
$$

where $x \otimes t = (x_j t_k)_{j,k} \in \mathbb{R}^{nm}$ and α_A is a norm on $\mathbb{R}^{nm}$ given by

$$
\alpha_A(y) := \left(\sum_i \left(\sum_{jk} a_{ijk} y_{jk} \right)^2 \right)^{1/2}
$$

(as in Conjecture 24 in this section we do not assume that the matrix $(a_{ijk})_{ijk}$ is symmetric or that it has 0 on the generalized diagonal).

Let

$$
B\big((x, t), d_A, r\big) = \left\{ (x', t') \in \mathbb{R}^n \times T : \alpha_A(x \otimes t - x' \otimes t') \leq r \right\}
$$

be the closed ball in d_A with center at (x, t) and radius r.

Observe that

$$
\operatorname{diam}\big(B_2^n \times T, d_A\big) \sim \sup_{\|x\|_2 \leq 1, t \in T} \left(\sum_i \left(\sum_{jk} a_{ijk} x_j t_k \right)^2 \right)^{1/2}.
$$

Now we try to estimate entropy numbers $N(B_2^n \times T, d_A, \varepsilon)$ for $\varepsilon > 0$ (recall that $N(S, \rho, \varepsilon)$ is the smallest number of closed balls with diameter ε in metric ρ that cover the set S). To this end we first introduce some notation. For a nonempty bounded set S in $\mathbb{R}^m$ let

$$
\beta_{A,S}(x) := \mathbb{E} \sup_{t \in S} \left| \sum_{ijk} a_{ijk} g_i x_j t_k \right|, \quad x \in \mathbb{R}^n.
$$

Observe that $\beta_{A,S}$ is a norm on $\mathbb{R}^n$. Moreover, by the classical Sudakov minoration ([28] or [24, Theorem 3.18]) for any $x \in \mathbb{R}^n$ and $\varepsilon > 0$ there exists a set $S_{x,\varepsilon} \subset S$ of cardinality at most $\exp(C\varepsilon^{-2})$ such that

$$
\forall_{t \in S} \exists_{t' \in S_{x,\varepsilon}} \quad \alpha_A\big(x \otimes (t - t')\big) \leq \varepsilon \beta_{A,S}(x).
$$

For a finite set $S \subset \mathbb{R}^m$ and $\varepsilon > 0$ define a measure $\mu_{\varepsilon,S}$ on $\mathbb{R}^n \times S$ in the following way

$$\mu_{\varepsilon,S}(C) := \int_{\mathbb{R}^n} \sum_{t \in S_{x,\varepsilon}} \delta_{(x,t)}(C) \, d\gamma_{n,\varepsilon}(x),$$

where $\gamma_{n,\varepsilon}$ is the distribution of the vector εG_n (recall that G_n is the standard Gaussian vector in $\mathbb{R}^n$). Since S is finite, we can choose sets $S_{x,\varepsilon}$ in such a way that there are no problems with measurability.

To bound $N(B_2^n \times T, d_A, \varepsilon)$ we need two lemmas.

Lemma 25 ([20, Lemma 1]). *For any norms α_1, α_2 on $\mathbb{R}^n$, $y \in B_2^n$ and $\varepsilon > 0$,*

$$\gamma_{n,\varepsilon}\big(x \colon \alpha_1(x - y) \le 4\varepsilon \mathbb{E}\alpha_1(G_n), \alpha_2(x) \le 4\varepsilon \mathbb{E}\alpha_2(G_n) + \alpha_2(y)\big) \ge \frac{1}{2}\exp\big(-\varepsilon^{-2}/2\big).$$

Lemma 26. *For any finite set S in $\mathbb{R}^m$, any $(x, t) \in B_2^n \times S$ and $\varepsilon > 0$ we have*

$$\mu_{\varepsilon,S}\big(B\big((x,t), d_A, r(\varepsilon)\big)\big) \ge \frac{1}{2}\exp\big(-\varepsilon^{-2}/2\big),$$

where

$$r(\varepsilon) = r(A, S, x, t, \varepsilon) = 4\varepsilon^2 \mathbb{E}\beta_{A,S}(G_n) + \varepsilon\beta_{A,S}(x) + 4\varepsilon \mathbb{E}\alpha_A(G_n \otimes t).$$

Proof. Let

$$U = \big\{x' \in \mathbb{R}^n \colon \beta_{A,S}(x') \le 4\varepsilon \mathbb{E}\beta_{A,S}(G_n) + \beta_{A,S}(x), \alpha_A\big((x - x') \otimes t\big) \le 4\varepsilon \mathbb{E}\alpha_A(G_n \otimes t)\big\}.$$

For any $x' \in U$ there exists $t' \in S_{x',\varepsilon}$ such that $\alpha_A(x' \otimes (t - t')) \le \varepsilon\beta_{A,S}(x')$. By the triangle inequality

$$\alpha_A\big(x \otimes t - x' \otimes t'\big) \le \alpha_A\big((x - x') \otimes t\big) + \alpha_A\big(x' \otimes (t - t')\big) \le r(\varepsilon).$$

Thus, by Lemma 25, $\mu_{\varepsilon,S}(B((x,t), d_A, r(\varepsilon))) \ge \gamma_{n,\varepsilon}(U) \ge \frac{1}{2}\exp(-\varepsilon^{-2}/2)$. $\qquad\square$

Having Lemma 26 we can estimate the entropy numbers by a version of the usual volumetric argument.

Corollary 27. *For any $\varepsilon > 0$, $U \subset B_2^n$ and $S \subset \mathbb{R}^m$,*

$$N\left(U \times S, d_A, 8\varepsilon^2 \mathbb{E}\beta_{A,S}(G_n) + 2\varepsilon \sup_{x \in U} \beta_{A,S}(x) + 8\varepsilon \sup_{t \in S} \mathbb{E}\alpha_A(G_n \otimes t)\right) \le \exp\big(C\varepsilon^{-2}\big) \tag{29}$$

and for any $\delta > 0$,

$$\sqrt{\log N(U \times S, d_A, \delta)} \le C\left(\delta^{-1}\left(\sup_{x \in U} \beta_{A,S}(x) + \sup_{t \in S} \mathbb{E}\alpha_A(G_n \otimes t)\right) + \delta^{-1/2}\big(\mathbb{E}\beta_{A,S}(G_n)\big)^{1/2}\right).$$

Proof. Let $r = 4\varepsilon^2 \mathbb{E}\beta_{A,S}(G_n) + \varepsilon \sup_{x \in U} \beta_{A,S}(x) + 4\varepsilon \sup_{t \in S} \mathbb{E}\alpha_A(G_n \otimes t)$ and $N = N(U \times S, d_A, 2r)$. Then there exist points $(x_i, t_i)_{i=1}^N$ in $U \times S$ such that $d_A((x_i, t_i), (x_j, t_j)) > 2r$. To show (29) we consider two cases.

If $\varepsilon > 2$ then

$$2r \ge 4\sup_{x \in U} \beta_{A,S}(x) \ge 4 \sup_{(x,t) \in U \times S} \mathbb{E}\left|\sum_{ijk} a_{ijk} g_i t_j x_k\right| = 4\sqrt{\frac{2}{\pi}} \sup_{(x,t) \in U \times S} \left(\sum_i \left(\sum_{jk} a_{ijk} t_j x_k\right)^2\right)^{1/2}$$

$$\ge \operatorname{diam}(U \times S, d_A)$$

so $N = 1 \le \exp(C\varepsilon^{-2})$.

If $\varepsilon < 2$, note that the balls $B((x_i, t_i), d_A, r)$ are disjoint and, by Lemma 26, each of these balls has $\mu_{\varepsilon,S}$ measure at least $\frac{1}{2}\exp(-\varepsilon^{-2}/2) \ge \exp(-5\varepsilon^{-2})$. On the other hand we obviously have $\mu_{\varepsilon,S}(\mathbb{R}^n \times S) \le \exp(C\varepsilon^{-2})$. Comparing the upper and lower bounds on $\mu_{\varepsilon,S}(\mathbb{R}^n \times S)$ gives (29) in this case.

The second estimate from the assertion is an obvious consequence of the first one. $\qquad\square$

Remark 28. The classical Dudley's bound on suprema of Gaussian processes (see, e.g., [9, Corollary 5.1.6]) gives

$$\mathbb{E} \sup_{\|x\|_2 \leq 1, t \in T} \left| \sum_{ijk} a_{ijk} g_i x_j t_k \right| \leq C \int_0^{\operatorname{diam}(B_2^n \times T, d_A)} \sqrt{\log N(B_2^n \times T, d_A, \delta)} \, d\delta.$$

Observe that

$$\int_0^{\operatorname{diam}(B_2^n \times T, d_A)} \delta^{-1/2} \left(\mathbb{E} \beta_{A,T}(G_n) \right)^{1/2} d\delta = 2\sqrt{\operatorname{diam}(B_2^n \times T, d_A) \mathbb{E} \beta_{A,T}(G_n)}$$

$$\leq \frac{1}{\sqrt{p}} \mathbb{E} \beta_{A,T}(G_n) + \sqrt{p} \operatorname{diam}(B_2^n \times T, d_A)$$

appears on the right hand side of (28). Unfortunately the other term in the estimate of $\log^{1/2} N(B_2^n \times T, d_A, \delta)$ is not integrable. The remaining part of the proof is devoted to improve on Dudley's bound.

We will now continue along the lines of [20]. We will need in particular to partition the set T into smaller pieces T_i such that $\sup_{t,s \in T_i} \mathbb{E} \alpha_A(G_n \otimes (t - t'))$ is small on each piece. To this end we apply the following Sudakov-type estimate for chaoses, derived by Talagrand ([29] or [30, Section 8.2]).

Theorem 29. *Let $\mathcal{A}$ be a subset of n by n real valued matrices and d_2, d_∞ be distances associated to the Hilbert-Schmidt and operator norms respectively. Then*

$$\varepsilon \log^{1/4} N(\mathcal{A}, d_2, \varepsilon) \leq C \mathbb{E} \sup_{a \in \mathcal{A}} \sum_{ij} a_{ij} g_i g_j' \quad \text{for } \varepsilon > 0$$

and

$$\varepsilon \log^{1/2} N(\mathcal{A}, d_2, \varepsilon) \leq C \mathbb{E} \sup_{a \in \mathcal{A}} \sum_{ij} a_{ij} g_i g_j' \quad \text{for } \varepsilon > C \sqrt{\operatorname{diam}(\mathcal{A}, d_\infty) \mathbb{E} \sup_{a \in \mathcal{A}} \sum_{ij} a_{ij} g_i g_j'}.$$

To make the notation more compact let for $T \subset \mathbb{R}^m$ and $V \subset \mathbb{R}^n \times \mathbb{R}^m$,

$$s_A(T) := \mathbb{E} \beta_{A,T}(G_n) = \mathbb{E} \sup_{t \in T} \left| \sum_{ijk} a_{ijk} g_i g_j' t_k \right|,$$

$$F_A(V) := \mathbb{E} \sup_{(x,t) \in V} \sum_{ijk} a_{ijk} g_i x_j t_k,$$

$$\Delta_{A,\infty}(T) := \sup_{\|x\|_2 \leq 1, \|y\|_2 \leq 1, t, t' \in T} \left| \sum_{ijk} a_{ijk} x_i y_j (t_k - t_k') \right|,$$

$$\Delta_A(V) := \operatorname{diam}(V, d_A) = \sup_{(x,t),(x',t') \in V} \alpha_A(x \otimes t - x' \otimes t').$$

Corollary 30. *Let T be a subset of $\mathbb{R}^m$. Then for any $r > 0$ there exists a decomposition $T - T = \bigcup_{i=1}^N T_i$ such that, $N \leq e^{Cr}$ and*

$$\sup_{t,t' \in T_i} \mathbb{E} \alpha_A(G_n \otimes (t - t')) \leq \min\{r^{-1/4} s_A(T), r^{-1/2} s_A(T) + C\sqrt{s_A(T) \Delta_{A,\infty}(T)}\}.$$

Proof. We use Theorem 29 with $\mathcal{A} = \{(\sum_k a_{ijk} t_k)_{ij} : t \in T - T\}$. It is enough to observe that

$$\mathbb{E} \sup_{b \in \mathcal{A}} \left| \sum_{ij} b_{ij} g_i g_j' \right| = s_A(T - T) \leq 2s_A(T), \qquad \operatorname{diam}(\mathcal{A}, d_\infty) = 2\Delta_{A,\infty}(T)$$

and

$$\mathbb{E} \alpha_A(G_n \otimes (t - t')) \leq \left\| \left(\sum_k a_{ijk}(t_k - t_k') \right)_{ij} \right\|_{HS}.$$

$$\square$$

On the other hand the dual Sudakov minoration (cf. formula (3.15) in [24]) yields the following

Corollary 31. *Let U be a subset of B_2^n. Then for any $r > 0$ there exists a decomposition $U = \bigcup_{i=1}^N U_i$ such that $N \leq e^{Cr}$ and*

$$\sup_{x,x' \in U_i} \beta_{A,T}(x - x') \leq r^{-1/2} s_A(T).$$

Putting the above two corollaries together with Corollary 27 we get the following decomposition of subsets of $B_2^n \times T$.

Corollary 32. *Let $V \subset \mathbb{R}^n \times \mathbb{R}^m$ be such that $V - V \subset B_2^n \times (T - T)$. Then for $r \geq 1$ we may find a decomposition $V = \bigcup_{i=1}^N ((x_i, t_i) + V_i)$ such that $N \leq e^{Cr}$ and for each $1 \leq i \leq N$,*

(i) $(x_i, t_i) \in V$, $V_i - V_i \subset V - V$, $V_i \subset B_2^n \times (T - T)$,
(ii) $\sup_{(x,t) \in V_i} \beta_{A,T}(x) \leq r^{-1/2} s_A(T)$,
(iii) $\sup_{(x,t) \in V_i} \mathbb{E} \alpha_A(G_n \otimes t) \leq \min\{r^{-1/4} s_A(T), r^{-1/2} s_A(T) + C\sqrt{s_A(T)\Delta_{A,\infty}(T)}\}$,
(iv) $\Delta_A(V_i) \leq \min\{r^{-3/4} s_A(T), r^{-1} s_A(T) + r^{-1/2}\sqrt{s_A(T)\Delta_{A,\infty}(T)}\}$.

Proof. The assertion is invariant under translations of the set V thus we may assume that $(0, 0) \in V$ and so $V \subset V - V \subset B_2^n \times (T - T)$. By Corollaries 30 and 31 we may decompose $B_2^n = \bigcup_{i=1}^{N_1} U_i$, $T - T = \bigcup_{i=1}^{N_2} T_i$ in such a way that $N_1, N_2 \leq e^{Cr}$ and

$$\sup_{x,x' \in U_i} \beta_{A,T}(x - x') \leq r^{-1/2} s_A(T),$$

$$\sup_{t,t' \in T_i} \mathbb{E} \alpha_A(G_n \otimes (t - t')) \leq \min\{r^{-1/4} s_A(T), r^{-1/2} s_A(T) + C\sqrt{s_A(T)\Delta_{A,\infty}(T)}\}.$$

Let $V_{ij} := V \cap (U_i \times T_j)$. If $V_{ij} \neq \varnothing$ we take any point $(x_{ij}, y_{ij}) \in V_{ij}$ and using Corollary 27 with $\varepsilon = r^{-1/2}/C$ we decompose

$$V_{ij} - (x_{ij}, y_{ij}) = \bigcup_{k=1}^{N_3} V_{ijk}$$

in such a way that $N_3 \leq e^{Cr}$ and

$$\Delta_A(V_{ijk}) \leq \frac{1}{C}\left(r^{-1} s_A(T) + r^{-1/2} \sup_{x' \in U_i} \beta_{A,T}(x' - x_{ij}) + r^{-1/2} \sup_{y' \in T_j} \mathbb{E}\alpha_A(G_n \otimes (y' - y_{ij}))\right)$$

$$\leq \min\{r^{-3/4} s_A(T), r^{-1} s_A(T) + r^{-1/2}\sqrt{s_A(T)\Delta_{A,\infty}(T)}\}.$$

The final decomposition is obtained by relabeling of the decomposition $V = \bigcup_{ijk}((x_{ij}, y_{ij}) + V_{ijk})$. $\square$

Remark 33. We may also use a trivial bound in (iii):

$$\sup_{(x,t) \in V_i} \mathbb{E}\alpha_A(G_n \otimes t) \leq \sup_{t,t' \in T} \mathbb{E}\alpha_A(G_n \otimes (t - t')) \leq 2\sup_{t \in T} \mathbb{E}\alpha_A(G_n \otimes t),$$

this will lead to the following bound in (iv):

$$\Delta_A(V_i) \leq r^{-1} s_A(T) + r^{-1/2} \sup_{t \in T} \mathbb{E}\alpha_A(G_n \otimes t).$$

Remark 34. By using Sudakov minoration instead of Theorem 29 we may decompose the set $T = \bigcup_{i=1}^N T_i$, $N \leq \exp(Cr)$ in such a way that

$$\forall_{i \leq N} \sup_{t,t' \in T_i} \mathbb{E}\alpha_A(G_n \otimes (t - t')) \leq r^{-1/2} \mathbb{E}\sup_{t \in T} \sum_{ijk} a_{ijk} g_{ij} t_k.$$

This will lead to the following bounds in (iii) and (iv):

$$\sup_{(x,t)\in V_i} \mathbb{E}\alpha_A(G_n \otimes t) \leq r^{-1/2}\mathbb{E}\sup_{t\in T}\sum_{ijk} a_{ijk}g_{ij}t_k,$$

$$\Delta_A(V_i) \leq r^{-1}\left(\mathbb{E}\sup_{t\in T}\sum_{ijk} a_{ijk}g_{ij}t_k + s_A(T)\right).$$

Lemma 35. *Let V be a subset of $B_2^n \times (T - T)$. Then for any $(y, s) \in \mathbb{R}^n \times \mathbb{R}^m$ we have*

$$F_A\big(V + (y, s)\big) \leq F_A(V) + 2\beta_{A,T}(y) + C\mathbb{E}\alpha_A(G_n \otimes s).$$

Proof. We have

$$F_A\big(V + (y, s)\big) \leq F_A(V) + \mathbb{E}\sup_{(x,t)\in V}\sum_{ijk} a_{ijk}g_i y_j t_k + \mathbb{E}\sup_{(x,t)\in V}\sum_{ijk} a_{ijk}g_i x_j s_k.$$

Obviously,

$$\mathbb{E}\sup_{(x,t)\in V}\sum_{ijk} a_{ijk}g_i y_j t_k \leq \mathbb{E}\sup_{t,t'\in T}\left|\sum_{ijk} a_{ijk}g_i y_j\big(t_k - t_k'\big)\right| \leq 2\beta_{A,T}(y).$$

Moreover,

$$\mathbb{E}\sup_{(x,t)\in V}\sum_{ijk} a_{ijk}g_i x_j s_k \leq \left(\mathbb{E}\sup_{x\in B_2^n}\left|\sum_{ijk} a_{ijk}g_i x_j s_k\right|^2\right)^{1/2} = \left(\sum_{ij}\left(\sum_k a_{ijk}s_k\right)^2\right)^{1/2}$$

$$= \big(\mathbb{E}\alpha_A(G_n \otimes s)^2\big)^{1/2} \leq C\mathbb{E}\alpha_A(G_n \otimes s),$$

where in the second inequality we used the comparison of moments of Gaussian variables [24, Corollary 3.2]. $\qquad\square$

Proposition 36. *For any nonempty finite set T in $\mathbb{R}^m$ and $p \geq 1$ we have*

$$F_A\big(B_2^n \times T\big) \leq C\left(\frac{\log(ep)}{\sqrt{p}}s_A(T) + \sup_{\|x\|_2 \leq 1}\beta_{A,T}(x) + \sup_{t\in T}\mathbb{E}\alpha_A(G_n \otimes t) + \log(ep)\sqrt{p}\Delta_A\big(B_2^n \times T\big)\right), \tag{30}$$

$$F_A\big(B_2^n \times T\big) \leq C\left(\frac{1}{\sqrt{p}}s_A(T) + \sup_{\|x\|_2 \leq 1}\beta_{A,T}(x) + \log(ep)\sup_{t\in T}\mathbb{E}\alpha_A(G_n \otimes t) + \sqrt{p}\Delta_A\big(B_2^n \times T\big)\right), \tag{31}$$

$$F_A\big(B_2^n \times T\big) \leq C\left(\frac{1}{\sqrt{p}}s_A(T) + \frac{1}{\sqrt{p}}\mathbb{E}\sup_{t\in T}\sum a_{ijk}g_{ij}t_k + \sup_{\|x\|_2 \leq 1}\beta_{A,T}(x)\right.$$

$$\left. + \sup_{t\in T}\mathbb{E}\alpha_A(G_n \otimes t) + \sqrt{p}\Delta_A\big(B_2^n \times T\big)\right). \tag{32}$$

Proof. First we prove (30). Let $l_0 \in \mathbb{N}$ be such that $2^{l_0-1} \leq p < 2^{l_0}$. Define

$$\Delta_0 := \Delta_A\big(B_2^n \times T\big), \qquad \tilde{\Delta}_0 := \sup_{x\in B_2^n}\beta_{A,T}(x) + \sup_{t\in T}\mathbb{E}\alpha_A(G_n \otimes t),$$

$$\Delta_l = 2^{-3l/4}p^{-3/4}s_A(T), \qquad \tilde{\Delta}_l = 2^{-l/4}p^{-1/4}s_A(T), \quad l > l_0.$$

and for $1 \leq l \leq l_0$,

$$\Delta_l := 2^{-l}p^{-1}s_A(T) + 2^{-l/2}p^{-1/2}\sqrt{s_A(T)\Delta_{A,\infty}(T)},$$

$$\tilde{\Delta}_l := 2^{-l/2}p^{-1/2}s_A(T) + C\sqrt{s_A(T)\Delta_{A,\infty}(T)}.$$

Let for $l = 0, 1, \dots$ and $m = 1, 2, \dots$

$$c(l, m) := \sup\Big\{ F_A(V) \colon V - V \subset B_2^n \times (T - T), \#V \leq m,$$

$$\Delta_A(V) \leq \Delta_l, \sup_{(x,t) \in V} \big(\beta_{A,T}(x) + \mathbb{E}\alpha_A(G_n \otimes t)\big) \leq 2\tilde{\Delta}_l\Big\}.$$

Obviously $c(l, 1) = 0$. We will show that for $m > 1$ and $l \geq 0$ we have

$$c(l, m) \leq c(l + 1, m - 1) + C\big(2^{l/2}\sqrt{p}\Delta_l + \tilde{\Delta}_l\big). \tag{33}$$

To this end take any set V as in the definition of $c(l, m)$ and apply to it Corollary 32 with $r = 2^{l+1}p$ to obtain decomposition $V = \bigcup_{i=1}^N((x_i, t_i) + V_i)$. We may obviously assume that all V_i have smaller cardinality than V. Conditions (i)–(iv) from Corollary 32 easily imply that $F_A(V_i) \leq c(l + 1, m - 1)$.

Gaussian concentration (cf. [20, Lemma 3]) yields

$$F_A(V) = F_A\Big(\bigcup_i\big((x_i, t_i) + V_i\big)\Big) \leq C\sqrt{\log N}\Delta_A(V) + \max_i F_A\big((x_i, t_i) + V_i\big).$$

Estimate (33) follows since

$$\sqrt{\log N}\Delta_A(V) \leq C2^{l/2}\sqrt{p}\Delta_l$$

and for each i by Lemma 35 we have (recall that $(x_i, t_i) \in V$)

$$F_A\big((x_i, t_i) + V_i\big) \leq F_A(V) + 2\beta_{A,T}(x_i) + C\mathbb{E}\alpha_A(G_n \otimes t_i) \leq c(l + 1, m - 1) + C\tilde{\Delta}_l.$$

Hence

$$c(0, m) \leq C\Bigg(\sum_{l=0}^{\infty} 2^{l/2}\sqrt{p}\Delta_l + \sum_{l=0}^{\infty}\tilde{\Delta}_l\Bigg)$$

$$\leq C\Bigg(\sqrt{p}\Delta_0 + \tilde{\Delta}_0 + \frac{1}{\sqrt{p}}s_A(T) + l_0\sqrt{s_A(T)\Delta_{A,\infty}(T)} + 2^{-l_0/4}p^{-1/4}s_A(T)\Bigg).$$

Since $\log_2 p < l_0 \leq \log_2 p + 1$ and $\sqrt{s_A(T)\Delta_{A,\infty}(T)} \leq \frac{1}{\sqrt{p}}s_A(T) + \sqrt{p}\Delta_{A,\infty}(T)$ and clearly $\Delta_{A,\infty}(T) \leq \Delta_A(B_2^n \times T)$ we get for all $m \geq 1$,

$$c(0, m) \leq C\Bigg(\frac{\log(ep)}{\sqrt{p}}s_A(T) + \sup_{\|x\|_2 \leq 1}\beta_{A,T}(x) + \sup_{t \in T}\mathbb{E}\alpha_A(G_n \otimes t) + \log(ep)\sqrt{p}\Delta_A\big(B_2^n \times T\big)\Bigg).$$

To conclude the proof of (30) it is enough to observe that

$$F_A\big(B_2^n \times T\big) = 2F_A\Big(\frac{1}{2}B_2^n \times T\Big) \leq 2\sup_{m \geq 1} c(0, m).$$

The proofs of (31) and (32) are the same as the proof of (30). The only difference is that for $1 \leq l \leq l_0$ we change the definitions of Δ_l, $\tilde{\Delta}_l$ and we use Remarks 33 and 34 respectively. In the first case we take

$$\Delta_l := 2^{-l}p^{-1}s_A(T) + 2^{-l/2}p^{-1/2}\sup_{t \in T}\mathbb{E}\alpha_A(G_n \otimes t),$$

$$\tilde{\Delta}_l := 2^{-l/2}p^{-1/2}s_A(T) + \sup_{t \in T}\mathbb{E}\alpha_A(G_n \otimes t),$$

while in the second

$$\Delta_l := 2^{-l}p^{-1}\Big(s_A(T) + \mathbb{E}\sup_{t \in T}\sum_{ijk} a_{ijk}g_{ij}t_k\Big),$$

$$\tilde{\Delta}_l = 2^{-l/2}p^{-1/2}\Big(s_A(T) + \mathbb{E}\sup_{t \in T}\sum_{ijk} a_{ijk}g_{ij}t_k\Big). \qquad \square$$

Remark 37. Note that the proof of Proposition 36 gives in fact the following estimate for any set $U \subset B_2^n \times T$. For any $p \geq 1$,

$$F_A(U) \leq C\left(\frac{1}{\sqrt{p}} s_A(T) + \frac{1}{\sqrt{p}} \mathbb{E} \sup_{t \in T} \sum a_{ijk} g_{ij} t_k + \sup_{\|x\|_2 \leq 1} \beta_{A,T}(x) + \sup_{t \in T} \mathbb{E} \alpha_A(G_n \otimes t) + \sqrt{p} \Delta_A(U)\right). \quad (34)$$

Indeed, as in the proof above one can reduce the problem to the case of $U \subset \frac{1}{2} B_2^n \times T$ and then it is enough to set $\Delta_0 = \Delta_A(U)$ and add the condition $V - V \subset U - U$ to the definition of $c(l, m)$.

6. Proofs of main results

Proof of Theorems 3 and 4. By Lemmas 20, 21 and Proposition 23 we need only to establish (4)-(6) with $\|\sum_{ij} a_{ij}(g_i g_j - \delta_{ij})\|_p$ replaced by $\sqrt{p} \mathbb{E} \sup_{\|x\|_2 \leq 1} \|\sum_{i \neq j} a_{ij} g_i x_j\|$. We may assume that $F = \mathbb{R}^m$ and by (26) also that $a_{ii} = 0$, so taking for T the unit ball in the dual space F^* we have

$$\left\|\sum_{i \neq j} a_{ij} g_i x_j\right\| = \sup_{t \in T} \sum_{ijk} a_{ijk} g_i x_j t_k.$$

Then, using the notation introduced in Section 5,

$$\mathbb{E} \sup_{\|x\|_2 \leq 1} \left\|\sum_{i \neq j} a_{ij} g_i x_j\right\| = F_A(B_2^n \times T), \qquad \mathbb{E} \left\|\sum_{i \neq j} a_{ij} g_i x_j\right\| = \beta_{A,T}(x),$$

$$\sup_{\|(x_{ij})\|_2 \leq 1} \left\|\sum_{ij} a_{ij} x_{ij}\right\| = \sup_{t \in T} (\mathbb{E} \alpha_A^2(G_n \otimes t))^{1/2} \sim \sup_{t \in T} \mathbb{E} \alpha_A(G_n \otimes t),$$

$$\sup_{\|x\|_2 \leq 1, \|y\|_2 \leq 1} \left\|\sum_{ij} a_{ij} x_i y_j\right\| \sim \Delta_A(B_2^n \times T),$$

$$\mathbb{E} \left\|\sum_{i \neq j} a_{ij} g_{ij}\right\| = \mathbb{E} \sup_{t \in T} \sum_{ijk} a_{ijk} g_{ij} t_k \quad \text{and} \quad \mathbb{E} \left\|\sum_{i \neq j} a_{ij} g_i g_j\right\| \sim s_A(T),$$

where the last estimate follows by decoupling. We conclude the proof invoking Proposition 36. $\qquad \square$

Proof of Theorem 6. The proof of the lower bound (8) will rely on two well known results, which we will state here in special cases, sufficient for our goals. The first one is the Paley–Zygmund inequality (see, e.g., [9, Corollary 3.3.2]), which asserts that for any non-negative random variable ξ, such that $0 < \mathbb{E}\xi^2 < \infty$ and for any $\lambda \in [0, 1]$,

$$\mathbb{P}(\xi > \lambda \mathbb{E}\xi) \geq (1 - \lambda)^2 \frac{(\mathbb{E}\xi)^2}{\mathbb{E}\xi^2}.$$

The other result we will need is a comparison of moments for Gaussian quadratic forms (see, e.g., [9, Theorem 3.2.10]), where the case of general polynomials is considered), which states that if $S = \|\sum_{ij} a_{ij}(g_i g_j - \delta_{ij})\|$, then for any $q > p > 1$,

$$\|S\|_q \leq \frac{q-1}{p-1} \|S\|_p.$$

Passing to the proof of (8), set $p = C \min\{t^2/U^2, t/V\}$. Note that it is enough to prove (8) under the additional assumption that $p \geq 2$, since the left-hand side of this inequality is non-increasing with t, while the right-hand side is bounded from above, so the case $p < 2$ can be easily reduced to $p = 2$ at the cost of increasing the constant C.

By the Paley–Zygmund inequality (applied to $\xi = S^p$) and comparison of moments of Gaussian quadratic forms we have for $p \geq 2$,

$$\mathbb{P}\left(S \geq \frac{1}{2}(\mathbb{E}|S|^p)^{1/p}\right) = \mathbb{P}\left(|S|^p \geq \frac{1}{2^p} \mathbb{E} S^p\right) \geq \left(1 - \frac{1}{2^p}\right)^2 \frac{(\mathbb{E} S^p)^2}{\mathbb{E} S^{2p}} \geq C_1^{-2p}.$$

Thus (8) follows by Proposition 1.

To derive the upper bound we use Theorem 4, estimate $\mathbb{P}(S \geq e\|S\|_p) \leq e^{-p}$ for $p \geq 1$ and make an analogous substitution. □

Proof of Theorem 7. Recall that for $r > 0$ the ψ_r-norm of a random variable Y is defined as

$$\|Y\|_{\psi_r} = \inf\left\{a > 0 : \mathbb{E}\exp\left(\left(\frac{|Y|}{a}\right)^r\right) \leq 2\right\}$$

(formally for $r < 1$ this is a quasi-norm, but it is customary to use the name ψ_r-norm for all r). By [4, Lemma 5.4] if k is a positive integer and $Y_1, \ldots, Y_n$ are symmetric random variables such that $\|Y\|_{\psi_{2/k}} \leq M$, then

$$\left\|\sum_{i=1}^n a_i Y_i\right\|_p \leq C_k M \left\|\sum_{i=1}^n a_i g_{i1} \cdots g_{ik}\right\|_p, \tag{35}$$

where g_{ik} are i.i.d. standard Gaussian variables (we remark that the lemma in [4] is stated only for $F = \mathbb{R}$ but its proof, based on the contraction principle, works in any normed space).

To prove the theorem we will again establish a moment bound and then combine it with Chebyshev's inequality. Similarly as in the Gaussian setting we will treat the diagonal and off-diagonal part separately. Let $\varepsilon_1, \ldots, \varepsilon_n$ be a sequence of i.i.d. Rademacher variables independent of X_i's. For $p \geq 1$ we have

$$\left\|\sum_i a_{ii}\left(X_i^2 - \mathbb{E}X_i^2\right)\right\|_p \leq 2\left\|\sum_i a_{ii}\varepsilon_i X_i^2\right\|_p \leq C\alpha^2\left\|\sum_i a_{ii}g_{i1}g_{i2}\right\|_p,$$

where in the first inequality we used symmetrization and in the second one (35) together with the observation $\|\varepsilon_i X_i^2\|_{\psi_1} \leq C\alpha^2$ (which can be easily proved by integration by parts).

Now by (25),

$$\left\|\sum_i a_{ii}g_{i1}g_{i2}\right\|_p \leq C\left\|\sum_i a_{ii}\left(g_i^2 - 1\right)\right\|_p,$$

and thus

$$\left\|\sum_i a_{ii}\left(X_i^2 - \mathbb{E}X_i^2\right)\right\|_p \leq C\alpha^2\left\|\sum_i a_{ii}\left(g_i^2 - 1\right)\right\|_p. \tag{36}$$

The estimate of the off-diagonal part is analogous, the only additional ingredient is decoupling. Denoting by $(X_i')_{i=1}^n$ an independent copy of the sequence $(X_i)_{i=1}^n$ and by $(\varepsilon_i)_{i=1}^n$, $(\varepsilon_i')_{i=1}^n$ (resp. $(g_i)_{i=1}^n$, $(g_i')_{i=1}^n$) independent sequences of Rademacher (resp. standard Gaussian) random variables, we have

$$\left\|\sum_{i \neq j} a_{ij} X_i X_j\right\|_p \sim \left\|\sum_{i \neq j} a_{ij} X_i X_j'\right\|_p \sim \left\|\sum_{i \neq j} a_{ij}\varepsilon_i X_i \varepsilon_i' X_j'\right\|_p$$

$$\leq C\alpha^2\left\|\sum_{i \neq j} a_{ij} g_i g_j'\right\|_p \sim \alpha^2\left\|\sum_{i \neq j} a_{ij} g_i g_j\right\|_p, \tag{37}$$

where in the first and last inequality we used decoupling, the second one follows from iterated conditional application of symmetrization inequalities and the third one from iterated conditional application of (35) (note that by integration by parts we have $\|\varepsilon_i X_i\|_{\psi_2} \leq C\alpha$).

Combining inequalities (36) and (37) with Lemma 20 we obtain

$$\left\|\sum_{ij} a_{ij}(X_i X_j - \mathbb{E}X_i X_j)\right\|_p \leq C\alpha^2\left\|\sum_{ij} a_{ij}(g_i g_j - \delta_{ij})\right\|_p.$$

To finish the proof of the theorem it is now enough to invoke moment estimates of Theorem 4 and use Chebyshev's inequality in L_p. □

Acknowledgement

References

[1] R. Adamczak. A note on the Hanson–Wright inequality for random vectors with dependencies. *Electron. Commun. Probab.* **20** (2015), no. 72, 13 pp. MR3407216 https://doi.org/10.1214/ECP.v20-3829

[2] R. Adamczak and R. Latała. Tail and moment estimates for chaoses generated by symmetric random variables with logarithmically concave tails. *Ann. Inst. Henri Poincaré Probab. Stat.* **48** (2012) 1103–1136. MR3052405 https://doi.org/10.1214/11-AIHP441

[3] R. Adamczak, R. Latała and R. Meller. Tail and moment estimates for Gaussian chaoses in Banach spaces. Preprint. Available at arXiv:2002.01790.

[4] R. Adamczak and P. Wolff. Concentration inequalities for non-Lipschitz functions with bounded derivatives of higher order. *Probab. Theory Related Fields* **162** (2015) 531–586. MR3383337 https://doi.org/10.1007/s00440-014-0579-3

[5] M. A. Arcones and E. Giné. On decoupling, series expansions, and tail behavior of chaos processes. *J. Theoret. Probab.* **6** (1993) 101–122. MR1201060 https://doi.org/10.1007/BF01046771

[6] F. Barthe and E. Milman. Transference principles for log-Sobolev and spectral-gap with applications to conservative spin systems. *Comm. Math. Phys.* **323** (2013) 575–625. MR3096532 https://doi.org/10.1007/s00220-013-1782-2

[7] C. Borell. On the Taylor series of a Wiener polynomial. In *Seminar Notes on Multiple Stochastic Integration, Polynomial Chaos and Their Integration*. Case Western Reserve Univ., Cleveland, 1984.

[8] L. A. Caffarelli. Monotonicity properties of optimal transportation and the FKG and related inequalities. *Comm. Math. Phys.* **214** (2000) 547–563. MR1800860 https://doi.org/10.1007/s002200000257

[9] V. H. de la Peña and 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. MR1666908 https://doi.org/10.1007/978-1-4612-0537-1

[10] V. H. de la Peña and S. J. Montgomery-Smith. Decoupling inequalities for the tail probabilities of multivariate U-statistics. *Ann. Probab.* **23** (1995) 806–816. MR1334173

[11] L. H. Dicker and M. A. Erdogdu. Flexible results for quadratic forms with applications to variance components estimation. *Ann. Statist.* **45** (2017) 386–414. MR3611496 https://doi.org/10.1214/16-AOS1456

[12] D. L. Hanson and F. T. Wright. A bound on tail probabilities for quadratic forms in independent random variables. *Ann. Math. Stat.* **42** (1971) 1079–1083. MR0279864 https://doi.org/10.1214/aoms/1177693335

[13] T. Hytönen, J. van Neerven, M. Veraar and L. Weis. *Analysis in Banach Spaces. Vol. II. Probabilistic Methods and Operator Theory. Ergebnisse der Mathematik und ihrer Grenzgebiete. 3. Folge. A Series of Modern Surveys in Mathematics* **67**. Springer, Cham, 2017. MR3752640 https://doi.org/10.1007/978-3-319-69808-3

[14] J.-P. Kahane. *Some Random Series of Functions*. Cambridge University Press, Cambridge, 1985. MR0833073

[15] V. Koltchinskii and K. Lounici. Concentration inequalities and moment bounds for sample covariance operators. *Bernoulli* **23** (2017) 110–133. MR3556768 https://doi.org/10.3150/15-BEJ730

[16] F. Krahmer, S. Mendelson and H. Rauhut. Suprema of chaos processes and the restricted isometry property. *Comm. Pure Appl. Math.* **67** (2014) 1877–1904. MR3263672 https://doi.org/10.1002/cpa.21504

[17] S. Kwapień. Decoupling inequalities for polynomial chaos. *Ann. Probab.* **15** (1987) 1062–1071. MR0893914

[18] R. Latała. Tail and moment estimates for sums of independent random vectors with logarithmically concave tails. *Studia Math.* **118** (1996) 301–304. MR1388035 https://doi.org/10.4064/sm-118-3-301-304

[19] R. Latała. Tail and moment estimates for some types of chaos. *Studia Math.* **135** (1999) 39–53. MR1686370 https://doi.org/10.4064/sm-135-1-39-53

[20] R. Latała. Estimates of moments and tails of Gaussian chaoses. *Ann. Probab.* **34** (2006) 2315–2331. MR2294983 https://doi.org/10.1214/009117906000000421

[21] R. Latała and K. Oleszkiewicz. On the best constant in the Khinchin–Kahane inequality. *Studia Math.* **109** (1) (1994) 101–104. MR1267715

[22] M. Ledoux. A note on large deviations for Wiener chaos. In *Séminaire de Probabilités, XXIV, 1988/89* 1–14. *Lecture Notes in Math.* **1426**. Springer, Berlin, 1990. MR1071528 https://doi.org/10.1007/BFb0083753

[23] M. Ledoux and K. Oleszkiewicz. On measure concentration of vector-valued maps. *Bull. Pol. Acad. Sci. Math.* **55** (2007) 261–278. MR2346103 https://doi.org/10.4064/ba55-3-7

[24] M. Ledoux and M. Talagrand. *Probability in Banach Spaces. Isoperimetry and Processes*. Springer-Verlag, Berlin, 1991. MR1102015 https://doi.org/10.1007/978-3-642-20212-4

[25] G. Pisier. Some results on Banach spaces without local unconditional structure. *Compos. Math.* **37** (1978) 3–19. MR0501916

[26] G. Pisier *Probabilistic Methods in the Geometry of Banach Spaces. Probability and Analysis. Lecture Notes in Math.* **1206**, 167–241. Springer, Berlin, 1986. MR0864714 https://doi.org/10.1007/BFb0076302

[27] M. Rudelson and R. Vershynin. Hanson–Wright inequality and sub-Gaussian concentration. *Electron. Commun. Probab.* **18** (2013), no. 82, 9 pp. MR3125258 https://doi.org/10.1214/ECP.v18-2865

[28] V. N. Sudakov. Gaussian measures, Cauchy measures and ε-entropy. *Sov. Math., Dokl.* **10** (1969) 310–313. MR0247034

[29] M. Talagrand. Sudakov-type minoration for Gaussian chaos processes. *Israel J. Math.* **79** (1992) 207–224. MR1248914 https://doi.org/10.1007/BF02808216

[30] M. Talagrand. *Upper and Lower Bounds for Stochastic Processes Modern Methods and Classical Problems. Ergebnisse der Mathematik und ihrer Grenzgebiete. 3. Folge. A Series of Modern Surveys in Mathematics* **60**. Springer, Heidelberg, 2014. MR3184689 https://doi.org/10.1007/978-3-642-54075-2

[31] J. M. A. M. van Neerven and L. Weis. Stochastic integration of operator-valued functions with respect to Banach space-valued Brownian motion. *Potential Anal.* **29** (2008) 65–88. MR2421495 https://doi.org/10.1007/s11118-008-9088-2

[32] R. Vershynin. *High-Dimensional Probability: An Introduction with Applications in Data Science*. *Cambridge Series in Statistical and Probabilistic Mathematics*. Cambridge University Press, Cambridge, 2018. MR3837109 https://doi.org/10.1017/9781108231596

[33] F. T. Wright. A bound on tail probabilities for quadratic forms in independent random variables whose distributions are not necessarily symmetric. *Ann. Probab.* **1** (1973) 1068–1070. MR0353419 https://doi.org/10.1214/aop/1176996815