Electronic Journal of Probability

Electron. J. Probab. **26** (2021), article no. 11, 1–36.
ISSN: 1083-6489 https://doi.org/10.1214/20-EJP567

Moments of Gaussian chaoses in Banach spaces*

Radosław Adamczak† Rafał Latała† Rafał Meller†

Abstract

We derive moment and tail estimates for Gaussian chaoses of arbitrary order with values in Banach spaces. We formulate a conjecture regarding two-sided estimates and show that it holds in a certain class of Banach spaces including L_q spaces. As a corollary we obtain two-sided bounds for moments of chaoses with values in L_q spaces based on exponential random variables.

Keywords: tail and moment inequalities; polynomials in independent random variables; Gaussian chaoses; Gaussian processes; metric entropy.
MSC2020 subject classifications: Primary 60E15, Secondary 60G15; 60B11.
Submitted to EJP on February 20, 2020, final version accepted on December 6, 2020.
Supersedes arXiv:2002.01790.

1 Introduction

Multivariate polynomials in Gaussian variables have been extensively studied at least since the work of Wiener in the 1930s. They have found numerous applications in the theory of stochastic integration and Malliavin calculus [12, 22, 23], functional analysis [11], limit theory for U-statistics [9] or long-range dependent processes [29], random graph theory [12], and more recently computer science [7, 14, 19, 24]. While early results considered mostly polynomials with real coefficients, their vector-valued counterparts also appear naturally, e.g., in the context of stochastic integration in Banach spaces [20], in the study of weak limits of U-processes [9], as tools in characterization of various geometric properties of Banach spaces [11, 25, 26] or in the analysis of empirical covariance operators [1, 30]. Apart from applications, the theory of Gaussian polynomials has been studied for its rich intrinsic structure, with interesting interplay of analytic, probabilistic, algebraic and combinatorial phenomena, leading to many challenging problems. For a comprehensive presentation of diverse aspects of the theory we refer to the monographs [9, 11, 12, 17].

*The authors were supported by National Science Centre, Poland grants 2015/18/E/ST1/00214 (R.A.), 2015/18/A/ST1/00553 (R.L.) and 2018/29/N/ST1/00454 (R.M.).

†Institute of Mathematics, University of Warsaw.
E-mail: radamcz@mimuw.edu.pl, rlatala@mimuw.edu.pl, rmeller@mimuw.edu.pl

An important aspect of the study of Gaussian polynomials is the order of their tail decay and growth of moments. In the real valued case the first estimates concerning this question, related to the hypercontractivity of the Ornstein-Uhlenbeck semigroup, were obtained by Nelson [21]. For homogeneous tetrahedral (i.e., affine in each variable) forms of arbitrary fixed degree two-sided estimates on the tails and moments were obtained in [15] (in particular generalizing the well-known Hanson-Wright inequality for quadratic forms). In [4] it was shown that the results of [15] in fact allow to obtain such estimates for all polynomials of degree bounded from above. Two-sided estimates for polynomials with values in a Banach space have been obtained independently by Borell [6], Ledoux [16], Arcones-Giné [5]. They are expressed in terms of suprema of certain empirical processes (see formula (1.5) below), which in general may be difficult to estimate (even in the real valued case).

In a recent paper [2] we considered Gaussian quadratic forms with coefficients in a Banach space and obtained upper bounds on their tails and moments, expressed in terms of quantities which are easier to deal with. In the real valued case our estimates reduce to the Hanson-Wright inequality, and for a large class of Banach-spaces (related to Pisier's Gaussian property α and containing all type 2 spaces) they may be reversed. In particular for L_q spaces with $1 \le q < \infty$ they yield two-sided estimates expressed in terms of deterministic quantities. In the present work we generalize these estimates to polynomials of arbitrary degree.

Before presenting our main theorems (which requires an introduction of a rather involved notation) let us describe the setting and discuss in more detail some of the results mentioned above.

To this aim consider a Banach space $(F, \|\cdot\|)$. A (homogeneous, tetrahedral) F-valued Gaussian chaos of order d is a random variable defined as

$$S = \sum_{1\le i_1 < i_2 \ldots < i_d \le n} a_{i_1,\ldots,i_d} g_{i_1} \cdots g_{i_d}, \tag{1.1}$$

where $a_{i_1,\ldots,i_d} \in F$ and $g_1, \ldots, g_n$ are i.i.d. standard Gaussian variables. As explained above the goal of this paper is to derive estimates on moments (defined as $\|S\|_p := (\mathbb{E}\|S\|^p)^{1/p}$) and tails of S, more precisely to establish upper bounds which for some classes of Banach spaces, including L_q spaces, can be reversed (up to constants depending only on d and the Banach space, but not on n or $a_{i_1,\ldots,i_d}$). We restrict to random variables of the form (1.1), however it turns out that estimates on their moments will in fact allow to deduce moment and tail bounds for arbitrary polynomials in Gaussian random variables as well as for homogeneous tetrahedral polynomials in i.i.d. symmetric exponential random variables. In the sequel we will focus on decoupled chaoses

$$S' = \sum_{i_1,\ldots,i_d=1}^{n} a_{i_1,\ldots,i_d} g^{(1)}_{i_1} \cdots g^{(d)}_{i_d}, \tag{1.2}$$

where $(g^{(k)}_i)_{i,k\ge 1}$ are independent $\mathcal{N}(0,1)$ random variables – under natural symmetry assumptions, moments and tails of S, S' are comparable up to constants depending only on d (cf. Theorem A.9 in the Appendix). Moreover, as we will see in Proposition 2.11, estimating moments of general polynomials in i.i.d. standard Gaussian variables can be reduced to estimating moments of variables of the form (1.2).

For $d = 1$ and any $p \ge 1$ one has the following well-known estimate (cf. Lemma A.5)

$$\left\|\sum_{i=1}^n a_i g_i\right\|_p = \left(\mathbb{E}\left\|\sum_{i=1}^n a_i g_i\right\|^p\right)^{1/p} \sim \mathbb{E}\left\|\sum_{i=1}^n a_i g_i\right\| + \sqrt{p}\sup_{x\in B_2^n}\left\|\sum_{i=1}^n a_i x_i\right\|, \tag{1.3}$$

where, B_2^n is the unit standard Euclidean ball in $\mathbb{R}^n$ and $\sim$ stands for a comparison up to universal multiplicative constants.

An iteration of the above inequality yields for chaoses of order 2,

$$\left\|\sum_{i,j=1}^n a_{ij}g_ig'_j\right\|_p \sim \mathbb{E}\left\|\sum_{i,j=1}^n a_{ij}g_ig'_j\right\| + \sqrt{p}\mathbb{E}\sup_{x\in B_2^n}\left(\left\|\sum_{i,j=1}^n a_{ij}g_ix_j\right\| + \left\|\sum_{i,j=1}^n a_{ij}x_ig_j\right\|\right) + p\sup_{x,y\in B_2^n}\left\|\sum_{i,j=1}^n a_{ij}x_iy_j\right\|, \tag{1.4}$$

where in the above formula and in the whole paper, $(g'_j)_{j\geq 1}$ is an independent copy of $(g_i)_{i\geq 1}$.

For chaoses of higher order one gets an estimate

$$\left\|\sum_{i_1,\ldots,i_d=1}^n a_{i_1,\ldots,i_d}g_{i_1}^{(1)}\cdots g_{i_d}^{(d)}\right\|_p \sim^d \sum_{J\subset[d]} p^{d/2}\mathbb{E}\sup\left\|\sum_{i_1,\ldots,i_d=1}^n a_{i_1,\ldots,i_d}\prod_{j\in J}x_{i_j}^{(j)}\prod_{j\in[d]\setminus J}g_{i_j}^{(j)}\right\|, \tag{1.5}$$

where the supremum is taken over $x^{(1)},\ldots,x^{(n)}$ from the Euclidean unit ball and $\sim^a$ stands for comparison up to constants depending only on the parameter a. To the best of our knowledge the above inequality was for the first time established in [6] and subsequently reproved in various context by several authors [5, 16, 17].

The estimate (1.5) gives precise dependence on p, but unfortunately is expressed in terms of expected suprema of certain stochastic processes, which are hard to estimate. In many situations this precludes effective applications. Let us note that even for $d=1$, the estimate (1.3) involves the expectation of a norm of a Gaussian random vector. Estimating such a quantity in general Banach spaces is a difficult task, which requires investigating the geometry of the unit ball of the dual of F (as described by the celebrated majorizing measure theorem due to Fernique and Talagrand). Therefore, in general one cannot hope to get rid of certain expectations in the estimates for moments. Nevertheless, in some classes of Banach spaces (such as, e.g., Hilbert spaces, or more generally type 2 spaces) expectations of Gaussian chaoses can be easily estimated. The difficult part (also for $d=2$ and mentioned class of Banach spaces) is to estimate the terms in (1.4) and (1.5) which involve additional suprema over products of unit balls. Even for $d=2$ and a Hilbert space, the term $\mathbb{E}\sup_{x\in B_2^n}\left\|\sum_{i,j}a_{ij}g_ix_j\right\|$ can be equivalently rewritten as the expected operator norm of a certain random matrix. Such quantities are known to be hard to estimate. Therefore, it is natural to seek inequalities which are expressed in terms of deterministic quantities and expectations of some F-valued polynomial chaoses, but do not involve expectations of additional suprema of such polynomials. This was the motivation behind the article [2], concerning the case $d=2$ and containing the following bound, valid for $p\geq 1$ ([2, Theorem 4]),

$$\left\|\sum_{i,j=1}^n a_{ij}g_ig'_j\right\|_p \leq C\Bigg(\mathbb{E}\left\|\sum_{i,j=1}^n a_{ij}g_ig'_j\right\| + \mathbb{E}\left\|\sum_{i,j=1}^n a_{ij}g_{ij}\right\| + p^{1/2}\sup_{x\in B_2^n}\mathbb{E}\left\|\sum_{i,j=1}^n a_{ij}g_ix_j\right\| + p^{1/2}\sup_{x\in B_2^{n^2}}\left\|\sum_{i,j=1}^n a_{ij}x_{ij}\right\| + p\sup_{x,y\in B_2^n}\left\|\sum_{i,j=1}^n a_{ij}x_iy_j\right\|\Bigg), \tag{1.6}$$

where (with a slight abuse of notation) we denote $B_2^{n^2}=\{(x_{ij})_{i,j=1}^n\colon \sum_{i,j=1}^n x_{ij}^2\leq 1\}$.

Let us point out that even though the inequalities we have presented so far as well as those we are about to discuss in the subsequent part of the article are formulated for general, possibly infinite dimensional, Banach spaces, the random variables involved take values in finite dimensional subspaces spanned by the coefficients of the polynomials in question. Therefore, as long as the constants in the inequalities are universal or do not depend on the particular subspace but just on some numerical characteristic of the space (e.g., the type constant), there is no loss in generality in assuming that the space F is finite dimensional. In particular one can always assume without loss of generality that F as a linear space equals $\mathbb{R}^m$ for some positive integer m. This phenomenon is well known in the local theory of Banach spaces.

It can be shown that in general inequality (1.6) cannot be reversed. However, it turns out to be two-sided in a certain class of Banach spaces containing L_q spaces (see Section 2.1 below). This observation gives rise to the question of obtaining similar results for arbitrary d. Building on ideas and techniques developed in [15] we are able to give an answer to it. In our main result, Theorem 2.1, we provide an upper bound on moments of decoupled chaoses of order d, which generalizes (1.6). We also obtain lower bounds, which we conjecture to be in fact two-sided (see Conjecture 2.2), and in Section 2.1 we identify a large class of Banach spaces for which our upper and lower bounds do match.

Let us briefly comment on the proof of Theorem 2.1. The lower bound for moments relies on a rather straightforward reduction to the real-valued case, treated in [15]. The much more involved proof of the upper bound is based on an inductive approach. The inequality (1.6) serves as the base of induction, while (1.3) allows to reduce the induction step to an estimate of an expectation of a supremum of a certain canonical Gaussian process, which turns out to be the heart of the problem. In order to obtain such an estimate we apply a variant of the chaining method (see the monograph [28]), which requires bounds on the entropy numbers for the indexing set of the process in its intrinsic metric. In our case this metric is given via a norm on a tensor product of F^* and several Euclidean spaces, whereas the indexing set is a Carthesian product of the corresponding unit balls. The bounds on entropy numbers are obtained by a variant of the volumetric argument as well as Sudakov and dual Sudakov minoration leading to expectations of suprema of other Gaussian processes, which can be estimated by using the induction hypotheses. This approach in a sense parallels the one used in [15] for the real-valued case, with (1.6) replacing the classical Hanson-Wright inequality, however it presents some additional difficulties related to the geometry of the unit ball in the space F^*. Let us also remark that this approach cannot be used to pass from $d = 1$, i.e., the inequality (1.3), to $d = 2$, i.e., the inequality (1.6) (see Remark 3.7). While the proof of (1.6) presented in [2] relies on a similar set of tools (chaining arguments, Gaussian concentration) some of the technical estimates of entropy numbers are obtained differently than in the induction step $d \to d + 1$ for $d \geq 2$.

The paper is organized as follows. In the next section we set up the notation and formulate the main results, in particular the pivotal bound for moments of homogeneous tetrahedral Gaussian chaoses in an arbitrary Banach space (Theorem 2.1). We also present its consequences: tail and moment estimates for arbitrary Gaussian polynomials, two-sided bounds in special classes of Banach spaces, inequalities for tetrahedral homogeneous forms in i.i.d. symmetric exponential variables. In Section 3, in Theorem 3.1, we formulate a key inequality for the supremum of a certain Gaussian processes and derive certain entropy bounds to be used in its proof, presented in Section 4. In Section 5 we use Theorem 3.1 to prove Theorem 2.1 from which we deduce all the remaining claims of Section 2. The Appendix contains certain basic facts concerning Gaussian processes and Gaussian polynomials used in the proofs. At the end of the article we provide a glossary explaining the notation.

2 Notation and main results

In this section we introduce the most basic notation used in the article and formulate our main results. Since some additional notation will be introduced as the proofs develop, for the reader's convenience at the end of the article we include a glossary of the most important symbols appearing in the text.

We write $[n]$ for the set $\{1,\ldots,n\}$. Throughout the article C (resp. $C(\alpha)$) will denote an absolute constant (resp. a constant which may depend on α) which may differ at each occurrence. By A we typically denote a finite multi-indexed matrix $(a_{i_1,\ldots,i_d})_{1\le i_1,\ldots,i_d\le n}$ of order d with values in a Banach space $(F,\|\cdot\|)$. If $\mathbf{i}=(i_1,\ldots,i_d),\ i_1,\ldots,i_d\in[n]$ and $I\subset[d]$, then we define $i_I:=(i_j)_{j\in I}$. To simplify the notation we will also often treat i_I as a stand-alone multi-index, with the meaning that each $i_j, j\in I$ runs trough $[n]$. We will also often suppress the range of summation. Unless stated otherwise the sums $\sum_{\mathbf{i}}$ or $\sum_{i_1,\ldots,i_d}$ should be understood as summation over $i_1,\ldots,i_d\in[n]$, whereas $\sum_{i_I}$ should be understood as summation over $i_{j_1},\ldots,i_{j_k}\in[n]$ where $I=\{i_1,\ldots,j_k\}$. The parameter d will not be stated explicitly but will be clear from the context. In particular when we write $x=(x_{i_I})_{i_I}$, it is implicitly assumed that the multi-index i_I ranges over $i_{j_1},\ldots,i_{j_k}\in[n]$ where $I=\{i_1,\ldots,j_k\}$ and $x\in\mathbb{R}^{n^{|I|}}$.

For instance, for $d=3$, $I=\{2,3\}$,

$$\sum_{\mathbf{i}} a_{\mathbf{i}}x_{i_I}=\sum_{i_1,i_2,i_3=1}^{n} a_{i_1i_2i_3}x_{i_2i_3},\quad \sum_{i_I}x_{i_I}^2=\sum_{i_2,i_3=1}^{n}x_{i_2i_3}^2.$$

We note that all the multi-linear forms we consider are given by finite sums. Standard arguments allow to extend our inequalities to the case of infinite multiple series, but we do not pursue this direction.

In what follows we will often identify the space $(\mathbb{R}^n)^{\otimes d}$ of d-indexed matrices with the space $\mathbb{R}^{n^d}$. In particular $B_2^{n^d}$ will stand for the unit Euclidean ball in $(\mathbb{R}^n)^{\otimes d}$, i.e., $B_2^{n^d}=\{(b_{\mathbf{i}})_{\mathbf{i}\in[n]^d}\in\mathbb{R}^{n^d}:\ \sum_{\mathbf{i}}b_{\mathbf{i}}^2\le 1\}$.

If I is a finite set then $|I|$ stands for its cardinality and by $\mathcal{P}(I)$ we denote the family of (unordered) partitions of I into nonempty, pairwise disjoint sets. Note that if $I=\emptyset$ then $\mathcal{P}(I)$ consists only of the empty partition $\emptyset$.

With a slight abuse of notation we write $(\mathcal{P},\mathcal{P}')\in\mathcal{P}(I)$ if $\mathcal{P}\cup\mathcal{P}'\in\mathcal{P}(I)$ and $\mathcal{P}\cap\mathcal{P}'=\emptyset$.

Let $\mathcal{P}=\{I_1,\ldots,I_k\},\ \mathcal{P}'=\{J_1,\ldots,J_m\}$ be such that $(\mathcal{P},\mathcal{P}')\in\mathcal{P}([d])$.Then we define

$$\|A\|_{\mathcal{P}'\,|\,\mathcal{P}}:=\sup\left\{\mathbb{E}\left\|\sum_{i_1,\ldots,i_d}a_{i_1,\ldots,i_d}\prod_{r=1}^{k}x^{(r)}_{i_{I_r}}\prod_{l=1}^{m}g^{(l)}_{i_{J_l}}\right\|\ \Big|\ \forall_{r\le k}\sum_{i_{I_r}}\left(x^{(r)}_{i_{I_r}}\right)^2\le 1\right\},\tag{2.1}$$

$$\|\!|A\|\!|_{\mathcal{P}}:=\sup\left\{\mathbb{E}\left\|\sum_{i_1,\ldots,i_d}a_{i_1,\ldots,i_d}\prod_{r=1}^{k}x^{(r)}_{i_{I_r}}\prod_{l\in[d]\setminus(\bigcup\mathcal{P})}g^{(l)}_{i_l}\right\|\ \Big|\ \forall_{r\le k}\sum_{i_{I_r}}\left(x^{(r)}_{i_{I_r}}\right)^2\le 1\right\},\tag{2.2}$$

where $G^{(l)}=(g^{(l)}_{i_{I_l}})_{i_{I_l}}$, $l=1,\ldots,m$ are independent arrays of i.i.d. standard Gaussian variables .

We do not exclude the situation that $\mathcal{P}'$ or $\mathcal{P}$ is an empty partition. If $\mathcal{P}'=\emptyset$, then $\|A\|_{\mathcal{P}'\,|\,\mathcal{P}}=\|\!|A\|\!|_{\mathcal{P}}$ is defined in non-probabilistic terms. Another case when $\|A\|_{\mathcal{P}'\,|\,\mathcal{P}}=\|\!|A\|\!|_{\mathcal{P}}$ is when $\mathcal{P}'$ consists of singletons only.

In particular for $d=3$ we have (note that to shorten the notation we suppress some brackets and write e.g. $\|\!|A\|\!|_{\{2\},\{3\}}$ and $\|A\|_{\{1\}\,|\,\{2\},\{3\}}$, instead of $\|\!|A\|\!|_{\{\{2\},\{3\}\}}$ and

$\|A\|_{\{\{1\}\}\mid\{\{2\},\{3\}\}})$

$$\|A\|_{\emptyset\mid\{1,2,3\}}=\|\!|A|\!\|_{\{1,2,3\}}=\sup_{\sum_{i,j,k}x_{ijk}^2\le 1}\left\|\sum_{i,j,k}a_{ijk}x_{ijk}\right\|,$$

$$\|A\|_{\emptyset\mid\{1,3\},\{2\}}=\|\!|A|\!\|_{\{1,3\},\{2\}}=\sup_{\sum_{i,k}x_{ik}^2\le 1,\sum_j y_j^2\le 1}\left\|\sum_{i,j,k}a_{ijk}x_{ik}y_j\right\|.$$

$$\|A\|_{\emptyset\mid\{1\},\{2\},\{3\}}=\|\!|A|\!\|_{\{1\},\{2\},\{3\}}=\sup_{\sum_i x_i^2\le 1,\sum_j y_j^2\le 1,\sum_k z_k^2\le 1}\left\|\sum_{i,j,k}a_{ijk}x_iy_jz_k\right\|.$$

$$\|A\|_{\{1,2\},\{3\}\mid\emptyset}=\mathbb{E}\left\|\sum_{i,j,k}a_{ijk}g_{ij}^{(1)}g_k^{(2)}\right\|,$$

$$\|A\|_{\{1\}\mid\{2\},\{3\}}=\|\!|A|\!\|_{\{2\},\{3\}}=\sup_{\sum_j x_j^2\le 1,\sum y_k^2\le 1}\mathbb{E}\left\|\sum_{i,j,k}a_{ijk}g_ix_jy_k\right\|,$$

$$\|A\|_{\{1\},\{2\},\{3\}\mid\emptyset}=\|\!|A|\!\|_{\emptyset}=\mathbb{E}\left\|\sum_{i,j,k}a_{ijk}g_i^{(1)}g_j^{(2)}g_k^{(3)}\right\|,$$

$$\|A\|_{\{1\},\{3\}\mid\{2\}}=\|\!|A|\!\|_{\{2\}}=\sup_{\sum_j x_j^2\le 1}\mathbb{E}\left\|\sum_{i,j,k}a_{ijk}g_i^{(1)}x_jg_k^{(2)}\right\|.$$

The main result is the following moment estimate of the variable S'.

Theorem 2.1. *Assume that $A=(a_{i_1,\ldots,i_d})_{i_1,\ldots,i_d}$ is a finite matrix with values in a Banach space $(F,\|\cdot\|)$. Then for any $p\ge 1$,*

$$\frac{1}{C(d)}\sum_{J\subset[d]}\sum_{\mathcal{P}\in\mathcal{P}(J)}p^{|\mathcal{P}|/2}\|\!|A|\!\|_{\mathcal{P}}\le\left\|\sum_{i_1,\ldots,i_d}a_{i_1,\ldots,i_d}g_{i_1}^{(1)}\cdots g_{i_d}^{(d)}\right\|_p$$
$$\le C(d)\sum_{(\mathcal{P},\mathcal{P}')\in\mathcal{P}([d])}p^{|\mathcal{P}|/2}\,\|A\|_{\mathcal{P}'\mid\mathcal{P}}.\tag{2.3}$$

The lower bound in (2.3) motivates the following conjecture (we leave it to the reader to verify that in general Banach spaces it is impossible to reverse the upper bound even for $d=2$).

Conjecture 2.2. *Under the assumption of Theorem 2.1 we have*

$$\left\|\sum_{i_1,\ldots,i_d}a_{i_1,\ldots,i_d}g_{i_1}^{(1)}\cdots g_{i_d}^{(d)}\right\|_p\le C(d)\sum_{J\subset[d]}\sum_{\mathcal{P}\in\mathcal{P}(J)}p^{|\mathcal{P}|/2}\|\!|A|\!\|_{\mathcal{P}}.\tag{2.4}$$

Example 2.3. In particular for $d=3$, Theorem 2.1 yields for symmetric matrices

$$\frac{1}{C}S_1\le\left\|\sum_{ijk}a_{ijk}g_i^{(1)}g_j^{(2)}g_k^{(3)}\right\|_p\le C(S_1+S_2),$$

where (we recall) C is a numerical constant and

$$\begin{aligned}
S_1 :=\ & \mathbb{E}\left\|\sum_{i,j,k} a_{ijk} g_i^{(1)} g_j^{(2)} g_k^{(3)}\right\| + p^{1/2}\left(\sup_{\|x\|_2\le 1}\mathbb{E}\left\|\sum_{i,j,k} a_{ijk} g_i^{(1)} g_j^{(2)} x_k\right\|\right. \\
& \left. + \sup_{\|x\|_2\le 1}\mathbb{E}\left\|\sum_{i,j,k} a_{ijk} g_i^{(1)} x_{jk}\right\| + \sup_{\|x\|_2\le 1}\left\|\sum_{i,j,k} a_{ijk} x_{ijk}\right\|\right) \\
& + p\left(\sup_{\|x\|_2,\|y\|_2\le 1}\mathbb{E}\left\|\sum_{i,j,k} a_{ijk} g_i^{(1)} x_j y_k\right\| + \sup_{\|x\|_2,\|y\|_2\le 1}\left\|\sum_{i,j,k} a_{ijk} x_{ij} y_k\right\|\right) \\
& + p^{3/2}\sup_{\|x\|_2,\|y\|_2,\|z\|_2\le 1}\left\|\sum_{i,j,k} a_{ijk} x_i y_j z_k\right\|, \\
S_2 :=\ & \mathbb{E}\left\|\sum_{ijk} a_{ijk} g_{ijk}^{(1)}\right\| + \mathbb{E}\left\|\sum_{i,j,k} a_{ijk} g_{ij}^{(1)} g_k^{(2)}\right\| + p^{1/2}\sup_{\|x\|_2\le 1}\mathbb{E}\left\|\sum_{i,j,k} a_{ijk} g_{ij}^{(1)} x_k\right\|.
\end{aligned}$$

Remark 2.4. Unfortunately we are able to show (2.4) only for $d = 2$ and with an additional factor $\ln p$ (cf. [2]). It is likely that by a modification of our proof one can show (2.4) for arbitrary d with an additional factor $(\ln p)^{C(d)}$.

By a standard application of Chebyshev's and Paley-Zygmund inequalities, Theorem 2.1 can be expressed in terms of tails.

Theorem 2.5. *Under the assumptions of Theorem 2.1 the following two inequalities hold. For any* $t > C(d)\sum_{\mathcal{P}'\in\mathcal{P}([d])}\|A\|_{\mathcal{P}'|\emptyset}$*,*

$$\mathbb{P}\left(\left\|\sum_{i_1,\ldots,i_d} a_{i_1,\ldots,i_d} g_{i_1}^{(1)}\cdots g_{i_d}^{(d)}\right\| \ge t\right) \le 2\exp\left(-\frac{1}{C(d)}\min_{\substack{(\mathcal{P},\mathcal{P}')\in\mathcal{P}([d])\\ |\mathcal{P}|>0}}\left(\frac{t}{\|A\|_{\mathcal{P}'\mid\mathcal{P}}}\right)^{2/|\mathcal{P}|}\right),$$

and for any $t\ge 0$*,*

$$\begin{aligned}
\mathbb{P}&\left(\left\|\sum_{i_1,\ldots,i_d} a_{i_1,\ldots,i_d} g_{i_1}^{(1)}\cdots g_{i_d}^{(d)}\right\| \ge \frac{1}{C(d)}\mathbb{E}\left\|\sum_{i_1,\ldots,i_d} a_{i_1,\ldots,i_d} g_{i_1}^{(1)}\cdots g_{i_d}^{(d)}\right\| + t\right) \\
&\qquad \ge \frac{1}{C(d)}\exp\left(-C(d)\min_{\emptyset\ne J\subset[d]}\min_{\mathcal{P}\in\mathcal{P}(J)}\left(\frac{t}{\|A\|_{\mathcal{P}}}\right)^{2/|\mathcal{P}|}\right).
\end{aligned}$$

In view of (1.5) and [15] it is clear that to prove Theorem 2.1 one needs to estimate suprema of some Gaussian processes. The next statement is the key element of the proof of the upper bound in (2.3).

Theorem 2.6. *Under the assumptions of Theorem 2.1 we have for any* $p\ge 1$*,*

$$\mathbb{E}\sup_{(x^{(2)},\ldots,x^{(d)})\in(B_2^n)^{d-1}}\left\|\sum_{i_1,\ldots,i_d} a_{i_1,\ldots,i_d} g_{i_1}\prod_{k=2}^{d} x_{i_k}^{(k)}\right\| \le C(d)\sum_{(\mathcal{P},\mathcal{P}')\in\mathcal{P}([d])} p^{\frac{|\mathcal{P}|+1-d}{2}}\|A\|_{\mathcal{P}'\mid\mathcal{P}}. \tag{2.5}$$

We postpone proofs of the above results until Section 5 and discuss now some of their consequences.

2.1 Two-sided estimates in special classes of Banach spaces

We start by introducing a class of Banach spaces for which the estimate (2.3) is two-sided. To this end we restrict our attention to Banach spaces $(F, \|\cdot\|)$ which satisfy the following condition: there exists a constant $K = K(F)$ such that for any $n \in \mathbb{N}$ and any matrix $(b_{ij})_{i,j\le n}$ with values in F,

$$\mathbb{E}\left\|\sum_{ij} b_{ij} g_{i,j}\right\| \le K\mathbb{E}\left\|\sum_{i,j} b_{ij} g_i g'_j\right\|. \tag{2.6}$$

This property appears in the literature under the name *Gaussian property* $(\alpha+)$ (see [20]) and is closely related to Pisier's contraction property [25]. It has found applications, e.g., in the theory of stochastic integration in Banach spaces. We refer to [11, Chapter 7] for a thorough discussion and examples, mentioning only that (2.6) holds for Banach spaces of type 2, and for Banach lattices (2.6) is equivalent to finite cotype.

Remark 2.7. By considering $n = 1$ it is easy to see that $K \ge \sqrt{\pi/2} > 1$.

A simple inductive argument and (2.6) yield that for any $d, n \in \mathbb{N}$ and any F-valued matrix $(b_{i_1,\ldots,i_d})_{i_1,\ldots,i_d\le n}$,

$$\mathbb{E}\left\|\sum_{\mathbf{i}} b_{\mathbf{i}} g_{\mathbf{i}}\right\| \le K^{d-1}\mathbb{E}\left\|\sum_{\mathbf{i}} b_{\mathbf{i}} g^{(1)}_{i_1}\cdots g^{(d)}_{i_d}\right\|, \tag{2.7}$$

where we recall that $\mathbf{i} = (i_1, \ldots, i_d)$ and each $i_1, \ldots, i_d$ runs trough $[n]$. It turns out that under the condition (2.6) our bound (2.3) is actually two-sided.

Proposition 2.8. *Assume that* $(F, \|\cdot\|)$ *satisfies* (2.6) *and* $(\mathcal{P}, \mathcal{P}') \in \mathcal{P}([d])$. *Then*

$$\|A\|_{\mathcal{P}' \mid \mathcal{P}} \le K^{|\bigcup \mathcal{P}'|-|\mathcal{P}'|}\|\!|A|\!\|_{\mathcal{P}}.$$

Proof. Let $\mathcal{P}' = (J_1, \ldots, J_k),\ \mathcal{P} = (I_1, \ldots, I_m)$. Then $|\bigcup \mathcal{P}'| - |\mathcal{P}'| = \sum_{l=1}^k (|J_l| - 1)$. The proof is by induction on $s := |\{l\ :\ |J_l| \ge 2\}|$. If $s = 0$ the assertion follows by the definition of $\|\!|A|\!\|_{\mathcal{P}}$. Assume that the statement holds for s and $|\{l\ :\ |J_l| \ge 2\}| = s + 1$. Without loss of generality $|J_1| \ge 2$. Combining Fubini's Theorem with (2.7) we obtain

$$\begin{aligned}
\|A\|_{\mathcal{P}' \mid \mathcal{P}} &= \sup\left\{\mathbb{E}^{(G^{(2)},\ldots,G^{(m)})}\mathbb{E}^{G^{(1)}}\left\|\sum_{\mathbf{i}} a_{\mathbf{i}} \prod_{r=1}^m x^{(r)}_{i_{I_r}} g^{(1)}_{i_{J_1}} \prod_{r=2}^k g^{(r)}_{i_{J_r}}\right\| \,\middle|\, \forall_{r\le m} \sum_{i_{I_r}} \left(x^{(r)}_{i_{I_r}}\right)^2 \le 1\right\}\\
&\le K^{|J_1|-1}\sup\left\{\mathbb{E}^{(G^{(2)},\ldots,G^{(m)})}\mathbb{E}^{\tilde{G}}\left\|\sum_{\mathbf{i}} a_{\mathbf{i}} \prod_{r=1}^m x^{(r)}_{i_{I_r}} \prod_{j\in J_1} \tilde{g}^{(j)}_{i_j} \prod_{r=2}^k g^{(r)}_{i_{J_r}}\right\| \,\middle|\, \forall_{r\le m} \sum_{i_{I_r}} \left(x^{(r)}_{i_{I_r}}\right)^2 \le 1\right\}\\
&\le K^{|\bigcup \mathcal{P}'|-|\mathcal{P}'|}\|\!|A|\!\|_{\mathcal{P}},
\end{aligned}$$

where $G^{(l)} = (g^{(l)}_{i_{I_l}})_{i_{I_l}}$, $\tilde{G} = (\tilde{g}^{(j)}_{i_j})_{j\in J_1, i_{J_1}}$ are independent families of i.i.d. $\mathcal{N}(0,1)$ random variables and in the last inequality we used (conditionally) the induction assumption. □

The following corollary is an obvious consequence of Proposition 2.8 and Theorems 2.1, 2.5.

Corollary 2.9. *For any Banach space* $(F, \|\cdot\|)$ *satisfying* (2.6) *we have for* $p \ge 1$,

$$\left\|\sum_{i_1,\ldots,i_d} a_{i_1,\ldots,i_d} g^{(1)}_{i_1}\cdots g^{(d)}_{i_d}\right\|_p \le C(d)K^{d-1}\sum_{I\subset[d]}\sum_{\mathcal{P}\in\mathcal{P}(I)} p^{|\mathcal{P}|/2}\|\!|A|\!\|_{\mathcal{P}},$$

and for $t > C(d)K^{d-1}\mathbb{E}\left\|\sum_{\mathbf{i}} a_{\mathbf{i}} g^{(1)}_{i_1}\cdots g^{(d)}_{i_d}\right\|$*,*

$$\mathbb{P}\left(\left\|\sum_{i_1,\ldots,i_d} a_{i_1,\ldots,i_d} g^{(1)}_{i_1}\cdots g^{(d)}_{i_d}\right\| \geq t\right) \leq 2\exp\left(-\frac{1}{C(d)}K^{2-2d}\eta(t)\right),$$

where

$$\eta(t) := \min_{\emptyset\neq I\subset[d]}\min_{\mathcal{P}\in\mathcal{P}(I)}\left(\frac{t}{\|A\|_{\mathcal{P}}}\right)^{2/|\mathcal{P}|}.$$

Thanks to infinite divisibility of Gaussian variables, the above corollary can be in fact generalized to arbitrary polynomials in Gaussian variables, as stated in the following theorem.

Theorem 2.10. *Let* F *be a Banach space. If* G *is a standard Gaussian vector in* $\mathbb{R}^n$ *and* $f\colon \mathbb{R}^n \to F$ *is a polynomial of degree* D*, then for all* $p \geq 2$*,*

$$\begin{aligned}&\|f(G) - \mathbb{E}f(G)\|_p \\ &\geq \frac{1}{C(D)}\Big(\mathbb{E}\|f(G)-\mathbb{E}f(G)\| + \sum_{1\leq d\leq D}\sum_{\emptyset\neq I\subset[d]}\sum_{\mathcal{P}\in\mathcal{P}(I)} p^{\frac{|\mathcal{P}|}{2}}\|\mathbb{E}\nabla^d f(G)\|_{\mathcal{P}}\Big)\end{aligned} \tag{2.8}$$

and for all $t > 0$*,*

$$\mathbb{P}\Big(\|f(G)-\mathbb{E}f(G)\| \geq \frac{1}{C(D)}(\mathbb{E}\|f(G)-\mathbb{E}f(G)\| + t)\Big) \geq \frac{1}{C(D)}\exp\Big(-C(D)\eta_f(t)\Big), \tag{2.9}$$

where

$$\eta_f(t) = \min_{1\leq d\leq D}\min_{\emptyset\neq I\subset[d]}\min_{\mathcal{P}\in\mathcal{P}(I)}\Big(\frac{t}{\|\mathbb{E}\nabla^d f(G)\|_{\mathcal{P}}}\Big)^{2/|\mathcal{P}|}.$$

Moreover, if F *satisfies* (2.6)*, then for all* $p \geq 1$*,*

$$\begin{aligned}&\|f(G) - \mathbb{E}f(G)\|_p \\ &\leq C(D)K^{D-1}\Big(\mathbb{E}\|f(G)-\mathbb{E}f(G)\| + \sum_{1\leq d\leq D}\sum_{\emptyset\neq I\subset[d]}\sum_{\mathcal{P}\in\mathcal{P}(I)} p^{\frac{|\mathcal{P}|}{2}}\|\mathbb{E}\nabla^d f(G)\|_{\mathcal{P}}\Big)\end{aligned} \tag{2.10}$$

and for all $t \geq C(D)K^{D-1}\mathbb{E}\|f(G)-\mathbb{E}f(G)\|$*,*

$$\mathbb{P}\Big(\|f(G)-\mathbb{E}f(G)\| \geq t\Big) \leq 2\exp\Big(-C(D)^{-1}K^{2-2D}\eta_f(t)\Big). \tag{2.11}$$

The above theorem is an easy consequence of results for homogeneous decoupled chaoses and the following proposition, the proof of which (as well as the proof of the theorem) will be presented in Section 5.

Proposition 2.11. *Let* F *be a Banach space,* G *a standard Gaussian vector in* $\mathbb{R}^n$ *and* $f\colon \mathbb{R}^n \to F$ *be a polynomial of degree* D*. Then for* $p \geq 1$*,*

$$\|f(G)-\mathbb{E}f(G)\|_p \sim^D \sum_{d=1}^{D}\Big\|\sum_{i_1,\ldots,i_d=1}^{n} a^{(d)}_{i_1,\ldots,i_d} g^{(1)}_{i_1}\cdots g^{(d)}_{i_d}\Big\|_p,$$

where the d*-indexed* F*-valued matrices* $A_d = (a^{(d)}_{i_1,\ldots,i_d})_{i_1,\ldots,i_d\leq n}$ *are defined as* $A_d = \mathbb{E}\nabla^d f(G)$*.*

Remark 2.12. Let us stress that Proposition 2.11 as well as inequalities (2.8) and (2.9) of Theorem 2.10 hold in arbitrary Banach spaces. The assumption (2.6) is needed for inequalities (2.10) and (2.11). A positive answer to Conjecture 2.2 would allow to eliminate this assumption and remove the constant K from the inequalities.

2.2 L_q spaces

It turns out that L_q spaces satisfy (2.6) and as a result upper and lower bounds in (2.3) are comparable. Moreover, as is shown in Lemma 2.14 below, in this case one may express all the parameters without any expectations. For the sake of brevity, we will focus on moment estimates, clearly tail bounds follow from them by standard arguments (cf. the proof of Theorem 2.5).

Proposition 2.13. *For $q \geq 1$ the space $L_q(V,\mu)$ satisfies* (2.6) *with $K = C\sqrt{q}$.*

Proof. From [11, Theorem 7.1.20] it follows that if F is of type 2 with constant T_2, then it satisfies (2.6) with $K = T_2$, while it is well known that the type 2 constant of $L_q(V,\mu)$ is of order $\sqrt{q}$. □

For a multi-indexed matrix A of order d with values in $L_q(V,\mu)$ and $J \subset [d]$, $\mathcal{P} = (I_1,\ldots,I_k) \in \mathcal{P}([J])$ we define

$$\|A\|_{L_q,\mathcal{P}} = \sup\left\{ \left\| \sqrt{\sum_{i_{[d]\setminus J}} \left(\sum_{i_J} a_{\mathbf{i}} \prod_{r=1}^{k} x^{(r)}_{i_{I_r}} \right)^2} \right\|_{L_q} \;\Big|\; \forall_{r\leq k} \sum_{i_{I_r}} \left(x^{(r)}_{i_{I_r}}\right)^2 \leq 1 \right\}.$$

For $J = [d]$ and $\mathcal{P} \in \mathcal{P}([d])$ we obviously have $\|A\|_{L_q,\mathcal{P}} = \|A\|_{\mathcal{P}}$. The following lemma asserts that for general J the corresponding two norms are comparable.

Lemma 2.14. *For any $J \subsetneq [d]$, $\mathcal{P} = (I_1,\ldots,I_k) \in \mathcal{P}(J)$ and any multi-indexed matrix A of order d with values in $L_q(V,\mu)$ we have*

$$C(d)^{-1} q^{\frac{1-d+|J|}{2}} \|A\|_{L_q,\mathcal{P}} \leq \|A\|_{\mathcal{P}} \leq C(d) q^{\frac{d-|J|}{2}} \|A\|_{L_q,\mathcal{P}}.$$

Proof. By Jensen's inequality and Corollary A.7 we get

$$\begin{aligned} \|A\|_{\mathcal{P}} &\leq \sup\left\{ \left(\int_V \mathbb{E}\left| \sum_{\mathbf{i}} a_{\mathbf{i}}(v) \prod_{j\in[d]\setminus J} g^{(j)}_{i_j} \prod_{r=1}^{k} x^{(r)}_{i_{I_r}} \right|^q d\mu(v) \right)^{1/q} \;\Big|\; \forall_{r\leq k} \sum_{i_{I_r}} \left(x^{(r)}_{i_{I_r}}\right)^2 \leq 1 \right\} \\ &\leq C(d) q^{\frac{d-|J|}{2}} \sup\left\{ \left\| \sqrt{\sum_{i_{[d]\setminus J}} \left(\sum_{i_J} a_{\mathbf{i}} \prod_{r=1}^{k} x^{(r)}_{i_{I_r}} \right)^2} \right\|_{L_q} \;\Big|\; \forall_{r\leq k} \sum_{i_{I_r}} \left(x^{(r)}_{i_{I_r}}\right)^2 \leq 1 \right\}. \end{aligned}$$

On the other hand Theorem A.1 (applied with $p = 1$) and Corollary A.7 yield

$$\begin{aligned} \|A\|_{\mathcal{P}} &\geq \frac{q^{\frac{|J|-d}{2}}}{C(d)} \sup\left\{ \left(\mathbb{E}\left\| \sum_{\mathbf{i}} a_{\mathbf{i}} \prod_{j\in[d]\setminus J} g^{(j)}_{i_j} \prod_{r=1}^{k} x^{(r)}_{i_{I_r}} \right\|_{L_q}^q \right)^{1/q} \;\Big|\; \forall_{r\leq k} \sum_{i_{I_r}} \left(x^{(r)}_{i_{I_r}}\right)^2 \leq 1 \right\} \\ &\geq \frac{q^{\frac{1-d+|J|}{2}}}{C(d)} \sup\left\{ \left\| \sqrt{\sum_{i_{[d]\setminus J}} \left(\sum_{i_J} a_{\mathbf{i}} \prod_{r=1}^{k} x^{(r)}_{i_{I_r}} \right)^2} \right\|_{L_q} \;\Big|\; \forall_{r\leq k} \sum_{i_{I_r}} \left(x^{(r)}_{i_{I_r}}\right)^2 \leq 1 \right\}. \end{aligned}$$

□

Theorem 2.15. *Let $q \geq 1$ and let $A = (a_{i_1,\ldots,i_d})_{i_1,\ldots,i_d}$ be a multi-indexed matrix with values in $L_q(V,\mu)$. Then for any $p \geq 1$ we have*

$$\frac{1}{C(d)} q^{\frac{1-d}{2}} \sum_{J\subset[d]} \sum_{\mathcal{P}\in\mathcal{P}([J])} p^{\frac{|\mathcal{P}|}{2}} \|A\|_{L_q,\mathcal{P}} \leq \Big\| \sum_{i_1,\ldots,i_d} a_{i_1,\ldots,i_d} g^{(1)}_{i_1}\cdots g^{(d)}_{i_d} \Big\|_p \leq C(d) q^{d-\frac{1}{2}} \sum_{J\subset[d]} \sum_{\mathcal{P}\in\mathcal{P}([J])} p^{\frac{|\mathcal{P}|}{2}} \|A\|_{L_q,\mathcal{P}}.$$

Proof. This is an obvious consequence of Theorem 2.1, Corollary 2.9, Proposition 2.13 and Lemma 2.14. □

Using Proposition 2.11 we can extend the above result to general polynomials.

Theorem 2.16. *Let G be a standard Gaussian vector in $\mathbb{R}^n$ and let $f\colon \mathbb{R}^n \to L_q(V,\mu)$ ($q \geq 1$) be a polynomial of degree D. Then for $p \geq 1$, we have*

$$\frac{1}{C(D)} \sum_{d=1}^{D} q^{\frac{1-d}{2}} \sum_{J\subset[d]} \sum_{\mathcal{P}\in\mathcal{P}([J])} p^{\frac{|\mathcal{P}|}{2}} \|\mathbb{E}\nabla^d f(G)\|_{L_q,\mathcal{P}} \leq \|f(G) - \mathbb{E}f(G)\|_p \leq C(D) \sum_{d=1}^{D} q^{d-\frac{1}{2}} \sum_{J\subset[d]} \sum_{\mathcal{P}\in\mathcal{P}([J])} p^{\frac{|\mathcal{P}|}{2}} \|\mathbb{E}\nabla^d f(G)\|_{L_q,\mathcal{P}}.$$

Example 2.17. Consider a general polynomial of degree 3, i.e.,

$$f(G) = \sum_{i,j,k=1}^{n} a_{ijk} g_i g_j g_k + \sum_{i,j=1}^{n} b_{ij} g_i g_j + \sum_{i=1}^{n} c_i g_i + d,$$

where the coefficients a_{ijk}, b_{ij}, c_i, d take values in a Banach space and the matrices $(a_{ijk})_{ijk}$, $(b_{ij})_{ij}$ are symmetric. Then one checks that

$$\begin{aligned}
\mathbb{E}\nabla f(G) &= \Big(c_i + 3\sum_{j=1}^{n} a_{ijj}\Big)_{i=1}^{n},\\
\mathbb{E}\nabla^2 f(G) &= 2(b_{ij})_{i,j=1}^{n},\\
\mathbb{E}\nabla^3 f(G) &= \nabla^3 f(G) = 6(a_{ijk})_{i,j,k=1}^{n}.
\end{aligned}$$

2.3 Exponential variables

Theorem 2.15 together with Lemma A.8 allows us to obtain inequalities for chaoses based on i.i.d standard symmetric exponential random variables (i.e., variables with density $2^{-1}\exp(-|t|)$) which are denoted by $(E^{(i)}_j)_{i,j\in\mathbb{N}}$ below. Similarly as in the previous section we concentrate only on the moment estimates.

Proposition 2.18. *Let $A = (a_{i_1,\ldots,i_d})_{i_1,\ldots,i_d}$ be a matrix with values in $L_q(V,\mu)$. Then for any $p \geq 1,\ q \geq 2$ we have*

$$\Big\| \sum_{\mathbf{i}} a_{\mathbf{i}} \prod_{k=1}^{d} E^{(k)}_{i_k} \Big\|_p \sim^{d,q} \sum_{I\subset[d]} \sum_{J\subset[d]\setminus I} \sum_{\mathcal{P}\in\mathcal{P}([d]\setminus(I\cup J))} p^{|I|+|\mathcal{P}|/2} \max_{i_I} \|(a_{i_1,\ldots,i_d})_{i_{I^c}}\|_{L_q,\mathcal{P}}.$$

One can take $C^{-1}(d)q^{1/2-d}$ in the lower bound and $C(d)q^{2d-1/2}$ in the upper bound.

Example 2.19. If $d = 2$ then Proposition 2.18 reads for a symmetric matrix $A = (a_{ij})_{ij}$ as

$$\begin{aligned}\Big\|\sum_{ij} a_{ij} E_i^{(1)} E_j^{(2)}\Big\|_p &\sim^q p^2 \max_{i,j} \|a_{ij}\|_{L_q} + p^{3/2} \max_i \sup_{x\in B_2^n} \Big\|\sum_j a_{ij} x_j\Big\|_{L_q} \\ &+ p\left(\max_{x,y\in B_2^n} \Big\|\sum_{ij} a_{ij} x_i y_j\Big\|_{L_q} + \max_i \Big\|\sqrt{\sum_j a_{ij}^2}\Big\|_{L_q}\right) \\ &+ p^{1/2}\left(\sup_{x\in B_2^n} \Big\|\sqrt{\sum_i \Big(\sum_j a_{ij} x_j\Big)^2}\Big\|_{L_q} + \sup_{x\in B_2^{n^2}} \Big\|\sum_{i,j} a_{ij} x_{ij}\Big\|_{L_q}\right) \\ &+ \Big\|\sqrt{\sum_{ij} a_{ij}^2}\Big\|_{L_q}.\end{aligned}$$

The proof of Proposition 2.18 is postponed until Section 5.

3 Reformulation of Theorem 2.6 and entropy estimates

Let us rewrite Theorem 2.6 in a different language. As explained in the introduction, since the variables we consider take values in the finite dimensional subspace spanned by the coefficients $a_{\mathbf{i}}$, we may assume without loss of generality that $F = \mathbb{R}^m$ for some finite m and $a_{i_1,\ldots,i_d} = (a_{i_1,\ldots,i_d,i_{d+1}})_{i_{d+1}\le m}$. For this reason from now on the multi-index $\mathbf{i}$ will take values in $[n]^d \times [m]$ and all summations over $\mathbf{i}$ should be understood as summations over this set. Accordingly, the matrix A will be treated as a $(d+1)$-indexed matrix with real coefficients. Let $T = B_{F^*}$ be the unit ball in the dual space F^* (where duality is realized on $\mathbb{R}^m$ through the standard inner product). In the sequel we will therefore assume that T is a fixed nonempty symmetric bounded subset of $\mathbb{R}^m$.

In this setup we have

$$\begin{aligned}\mathbb{E} \sup_{(x^{(2)},\ldots,x^{(d)})\in(B_2^n)^{d-1}} \Big\|\sum_{\mathbf{i}} a_{\mathbf{i}} g_{i_1} \prod_{k=2}^d x_{i_k}^{(k)}\Big\| &= \mathbb{E} \sup_{(x^{(2)},\ldots,x^{(d)})\in(B_2^n)^{d-1}} \sup_{t\in T} \sum_{\mathbf{i}} a_{\mathbf{i}} g_{i_1} \prod_{k=2}^d x_{i_k}^{(k)} t_{i_{d+1}}, \\ \|A\|_{\mathcal{P}' \mid \mathcal{P}} = \sup\Bigg\{ \mathbb{E}\sup_{t\in T} \sum_{\mathbf{i}} a_{\mathbf{i}} \prod_{k=1}^r x_{i_{I_k}}^{(k)} \prod_{l=1}^s g_{i_{J_l}}^{(l)} t_{i_{d+1}} \;\Big|\; &\forall_{k=1,\ldots,r} \sum_{i_{I_j}} \Big(x_{i_{I_k}}^{(k)}\Big)^2 = 1 \Bigg\},\end{aligned} \tag{3.1}$$

where $\mathcal{P} = (I_1,\ldots,I_r), \mathcal{P}' = (J_1,\ldots,J_s)$, $(\mathcal{P}',\mathcal{P}) \in \mathcal{P}([d])$.

To make the notation more compact we define

$$s_k(A) = \sum_{\substack{(\mathcal{P},\mathcal{P}')\in\mathcal{P}([d]) \\ |\mathcal{P}|=k}} \|A\|_{\mathcal{P}' \mid \mathcal{P}}.$$

The next statement is a reformulation of Theorem 2.6 in the introduced setup. The proof of it will be presented in Section 4.

Theorem 3.1. *For any $p \ge 1$ we have*

$$\mathbb{E} \sup_{(x^{(2)},\ldots,x^{(d)},t)\in(B_2^n)^{d-1}\times T} \sum_{\mathbf{i}} a_{\mathbf{i}} g_{i_1} \prod_{k=2}^d x_{i_k}^{(k)} t_{i_{d+1}} \le C(d) \sum_{k=0}^d p^{\frac{k+1-d}{2}} s_k(A). \tag{3.2}$$

To estimate the supremum of a centered Gaussian process $(G_v)_{v\in V}$ one needs to study the distance on V given by $d(v,v') := (\mathbb{E}|G_v - G_{v'}|^2)^{1/2}$ (we refer to the monograph [28] for an extensive presentation of chaining techniques related to estimates of suprema of stochastic processes). In the case of the Gaussian process from (3.2) this distance is defined on $(B_2^n)^{d-1} \times T \subset \mathbb{R}^{n(d-1)} \times \mathbb{R}^m$ by the formula

$$
\begin{aligned}
&\rho_A((x^{(2)},\ldots,x^{(d)},t),(y^{(2)},\ldots,y^{(d)},t')) \\
&:= \left(\sum_{i_1}\left(\sum_{i_2,\ldots,i_{d+1}} a_{i_1,\ldots,i_{d+1}}\left(\prod_{k=2}^{d} x^{(k)}_{i_k} t_{i_{d+1}} - \prod_{k=2}^{d} y^{(k)}_{i_k} t'_{i_{d+1}}\right)\right)^2\right)^{1/2} \\
&= \alpha_A\left(\left(\bigotimes_{k=2}^{d} x^{(k)}\right)\otimes t - \left(\bigotimes_{k=2}^{d} y^{(k)}\right)\otimes t'\right),
\end{aligned} \tag{3.3}
$$

where $\left(\bigotimes_{k=2}^{d} x^{(k)}\right)\otimes t = (x^{(2)}_{i_2}\cdots x^{(d)}_{i_d} t_{i_{d+1}})_{i_2,\ldots,i_{d+1}} \in \mathbb{R}^{n^{d-1}m}$ and α_A is a norm defined on $(\mathbb{R}^n)^{\otimes(d-1)} \otimes \mathbb{R}^m \simeq \mathbb{R}^{n^{d-1}m}$ given by

$$
\alpha_A(\mathbf{x}) := \sqrt{\sum_{i_1}\left(\sum_{i_{[d+1]\setminus\{1\}}} a_{\mathbf{i}} \mathbf{x}_{i_{[d+1]\setminus\{1\}}}\right)^2}. \tag{3.4}
$$

We will now provide estimates for the entropy numbers $N(U,\rho_A,\varepsilon)$ for $\varepsilon > 0$ and $U \subset (B_2^n)^{d-1} \times T$ (recall that $N(S,\rho,\varepsilon)$ is the minimal number of closed balls with diameter ε in metric ρ that cover the set S). To this end let us introduce some new notation. From now on $G_n = (g_1,\ldots,g_n)$ and $G_n^{(i)} = (g_1^{(i)},\ldots,g_n^{(i)})$ stand for independent standard Gaussian vectors in $\mathbb{R}^n$. For $s > 0$, $U = \{(x^{(2)},\ldots,x^{(d)},t) \in U\} \subset (\mathbb{R}^n)^{d-1} \times T$ we set

$$
W_d^U(\alpha_A, s) := \sum_{k=1}^{d-1} s^k \sum_{I\subset\{2,\ldots,d\}:|I|=k} W_I^U(\alpha_A), \tag{3.5}
$$

where

$$
W_I^U(\alpha_A) := \sup_{(x^{(2)},\ldots,x^{(d)},t)\in U} \mathbb{E}\alpha_A\left(\left(\bigotimes_{k=2}^{d}\left(x^{(k)}(1-\mathbf{1}_I(k)) + G^{(k)}\mathbf{1}_I(k)\right)\right)\otimes t\right).
$$

We define a norm β_A on $(\mathbb{R}^n)^{\otimes(d-1)} \simeq \mathbb{R}^{n^{d-1}}$ by (recall that we assume symmetry of the set T)

$$
\beta_A(\mathbf{y}) := \mathbb{E}\sup_{t\in T}\sum_{\mathbf{i}} a_{\mathbf{i}} g_{i_1} \mathbf{y}_{i_{[d]\setminus\{1\}}} t_{i_{d+1}} = \mathbb{E}\sup_{t\in T}\left|\sum_{\mathbf{i}} a_{\mathbf{i}} g_{i_1} \mathbf{y}_{i_{[d]\setminus\{1\}}} t_{i_{d+1}}\right|. \tag{3.6}
$$

Following (3.5) we denote

$$
V_d^U(\beta_A, s) := \sum_{k=0}^{d-1} s^{k+1} \sum_{I\subset\{2,\ldots,d\}:|I|=k} V_I^U(\beta_A), \tag{3.7}
$$

where

$$
V_I^U(\beta_A) := \sup_{(x^{(2)},\ldots,x^{(d)},t)\in U} \mathbb{E}\beta_A\left(\bigotimes_{k=2}^{d}\left(x^{(k)}(1-\mathbf{1}_I(k)) + G^{(k)}\mathbf{1}_I(k)\right)\right).
$$

Let us note that $V_I^U(\beta_A)$ depends on the set U only through its projection on the first $d-1$ coordinates.

We have

$$V_d^U(\beta_A, s) \geq s \cdot V_\emptyset^U(\beta_A) = s \cdot \sup_{(x^{(2)},\dots,x^{(d)},t)\in U} \beta_A\left(\bigotimes_{k=2}^d x^{(k)}\right). \tag{3.8}$$

Observe that by the classical Sudakov minoration (see Theorem A.2), for any $x^{(k)} \in \mathbb{R}^n$, $k = 2,\dots,d$ there exists $T_{\bigotimes x^{(k)},\varepsilon} \subset T$ such that $|T_{\bigotimes x^{(k)},\varepsilon}| \leq \exp(C\varepsilon^{-2})$ and

$$\forall_{t\in T}\exists_{t'\in T_{\bigotimes x^{(k)},\varepsilon}} \alpha_A\left(\bigotimes_{k=2}^d x^{(k)} \otimes (t-t')\right) \leq \varepsilon\beta_A\left(\bigotimes_{k=2}^d x^{(k)}\right).$$

We define a measure $\mu^d_{\varepsilon,T}$ on $\mathbb{R}^{(d-1)n}\times T$ by the formula

$$\mu^d_{\varepsilon,T}(C) := \int_{\mathbb{R}^{(d-1)n}} \sum_{t\in T_{\bigotimes x^{(k)},\varepsilon}} \mathbf{1}_C\big((x^{(2)},\dots,x^{(d)},t)\big)d\gamma_{(d-1)n,\varepsilon}\big((x^{(k)})_{k=2,\dots,d}\big),$$

where $\gamma_{n,t}$ is the distribution of $tG_n = t(g_1,\dots,g_n)$. Clearly,

$$\mu^d_{\varepsilon,T}((\mathbb{R}^{d-1})^n\times T) \leq \exp(C\varepsilon^{-2}). \tag{3.9}$$

To bound $N(U,\rho_A,\varepsilon)$ for $\varepsilon>0$ and $U\subset (B_2^n)^{d-1}\times T$ we need two lemmas.

Lemma 3.2 ([15, Lemma 2])**.** *For any* $\mathbf{x} = (x^{(1)},\dots,x^{(d)}) \in (B_2^n)^d$*, norm* α' *on* $\mathbb{R}^{n^d}$ *and* $\varepsilon>0$ *we have*

$$\gamma_{dn,\varepsilon}\left(B_{\alpha'}(\mathbf{x}, r(4\varepsilon,\alpha'))\right) \geq 2^{-d}\exp(-d\varepsilon^{-2}/2),$$

where

$$B_{\alpha'}(\mathbf{x}, r(\varepsilon,\alpha')) = \left\{\mathbf{y} = (y^{(1)},\dots,y^{(d)}) \in (\mathbb{R}^n)^d \;\middle|\; \alpha'\left(\bigotimes_{k=1}^d x^{(k)} - \bigotimes_{k=1}^d y^{(k)}\right) \leq r(\varepsilon,\alpha')\right\}$$

and

$$r(\varepsilon,\alpha') = \sum_{k=1}^d \varepsilon^k \sum_{I\subset[d]\colon |I|=k} \mathbb{E}\alpha'\left(\bigotimes_{k=1}^d \left(x^{(k)}(1-\mathbf{1}_{k\in I}) + G^{(k)}\mathbf{1}_{k\in I}\right)\right).$$

Lemma 3.3. *For any* $(\mathbf{x},t) = (x^{(2)},\dots,x^{(d)},t) \in (B_2^n)^{d-1}\times T$ *and* $\varepsilon>0$ *we have*

$$\mu^d_{\varepsilon,T}\left(B\left((\mathbf{x},t),\rho_A, W_d^{\{(\mathbf{x},t)\}}(\alpha_A,8\varepsilon) + V_d^{\{(\mathbf{x},t)\}}(\beta_A,8\varepsilon)\right)\right) \geq c^d\exp\left(-C(d)\varepsilon^{-2}\right).$$

Proof. Fix $(\mathbf{x},t)\in (B_2^n)^{d-1}\times T$, $\varepsilon>0$ and consider

$$\begin{aligned} U = \Bigg\{(y^{(2)},\dots,y^{(d)}) \in \mathbb{R}^{(d-1)n}\colon \alpha_A\left(\left(\bigotimes_{k=2}^d x^{(k)} - \bigotimes_{k=2}^d y^{(k)}\right)\otimes t\right) \\ + \varepsilon\beta_A\left(\bigotimes_{k=2}^d x^{(k)} - \bigotimes_{k=2}^d y^{(k)}\right) \leq W_d^{\{(\mathbf{x},t)\}}(\alpha_A,4\varepsilon) + V_d^{\{(\mathbf{x},t)\}}(\beta_A,4\varepsilon)\Bigg\}. \end{aligned}$$

For any $(y^{(2)},\dots,y^{(d)})\in U$ there exists $t'\in T_{\bigotimes y^{(k)},\varepsilon}$ such that

$$\alpha_A\left(\bigotimes_{k=2}^d y^{(k)}\otimes(t-t')\right) \leq \varepsilon\beta_A\left(\bigotimes_{k=2}^d y^{(k)}\right).$$

By the triangle inequality,

$$\begin{aligned}
&\alpha_A\left(\bigotimes_{k=2}^d x^{(k)}\otimes t-\bigotimes_{k=2}^d y^{(k)}\otimes t'\right)\\
&\quad\le \alpha_A\left(\left(\bigotimes_{k=2}^d x^{(k)}-\bigotimes_{k=2}^d y^{(k)}\right)\otimes t\right)+\alpha_A\left(\bigotimes_{k=2}^d y^{(k)}\otimes(t-t')\right)\\
&\quad\le \alpha_A\left(\left(\bigotimes_{k=2}^d x^{(k)}-\bigotimes_{k=2}^d y^{(k)}\right)\otimes t\right)+\varepsilon\beta_A\left(\bigotimes_{k=2}^d x^{(k)}-\bigotimes_{k=2}^d y^{(k)}\right)+\varepsilon\beta_A\left(\bigotimes_{k=2}^d x^{(k)}\right)\\
&\quad\le W_d^{\{(\mathbf{x},t)\}}(\alpha_A,4\varepsilon)+2V_d^{\{(\mathbf{x},t)\}}(\beta_A,4\varepsilon)\le W_d^{\{(\mathbf{x},t)\}}(\alpha_A,8\varepsilon)+V_d^{\{(\mathbf{x},t)\}}(\beta_A,8\varepsilon),
\end{aligned}$$

where in the third inequality we used (3.8). Thus,

$$\begin{aligned}
&\mu^d_{\varepsilon,T}\left(B\left((\mathbf{x},t),\rho_A,W_d^{\{(\mathbf{x},t)\}}(\alpha_A,8\varepsilon)+V_d^{\{(\mathbf{x},t)\}}(\beta_A,8\varepsilon)\right)\right)\\
&\qquad\qquad\qquad\qquad\ge\gamma_{(d-1)n,\varepsilon}(U)\ge c^d\exp(-C(d)\varepsilon^{-2}),
\end{aligned}$$

where the last inequality follows by Lemma 3.2 applied to the norm $\alpha_A(\cdot\otimes t)+\varepsilon\beta_A(\cdot)$. □

Corollary 3.4. *Let $U\subset(B_2^n)^{d-1}\times T$. Then for any $\varepsilon>0$,*

$$N\left(U,\rho_A,W_d^U(\alpha_A,\varepsilon)+V_d^U(\beta_A,\varepsilon)\right)\le\exp(C(d)\varepsilon^{-2}) \tag{3.10}$$

and for any $\delta>0$,

$$\begin{aligned}
&\sqrt{\log N(U,\rho_A,\delta)}\\
&\le C(d)\left(\sum_{k=1}^{d-1}\left(\sum_{\substack{I\subset\{2,\ldots,d\}\\|I|=k}}W_I^U(\alpha_A)\right)^{\frac1k}\delta^{-\frac1k}+\sum_{k=0}^{d-1}\left(\sum_{\substack{I\subset\{2,\ldots,d\}\\|I|=k}}V_I^U(\beta_A)\right)^{\frac{1}{k+1}}\delta^{-\frac{1}{k+1}}\right). \quad (3.11)
\end{aligned}$$

Proof. It suffices to show (3.10), since it easily implies (3.11). Consider first $\varepsilon\le 8$. Obviously, $W_d^U(\alpha_A,\varepsilon)+V_d^U(\beta_A,\varepsilon)\ge\sup_{(\mathbf{x},t)\in U}(W_d^{\{(\mathbf{x},t)\}}(\alpha_A,\varepsilon)+V_d^{\{(\mathbf{x},t)\}}(\beta_A,\varepsilon))$. Therefore, by Lemma 3.3 (applied with $\varepsilon/16$) we have for any $(\mathbf{x},t)\in U$,

$$\mu^d_{\varepsilon,T}\left(B\left((\mathbf{x},t),\rho_A,W_d^U(\alpha_A,\varepsilon/2)+V_d^U(\beta_A,\varepsilon/2)\right)\right)\ge C(d)^{-1}\exp\left(-C(d)\varepsilon^{-2}\right). \tag{3.12}$$

Suppose that there exist $(\mathbf{x}_1,t_1),\ldots,(\mathbf{x}_N,t_N)\in U$ such that

$$\rho_A((\mathbf{x}_i,t_i),(\mathbf{x}_j,t_j))>W_d^U(\alpha_A,\varepsilon)+V_d^U(\beta_A,\varepsilon)\ge 2W_d^U(\alpha_A,\varepsilon/2)+2V_d^U(\beta_A,\varepsilon/2)\text{ for } i\ne j.$$

Then the sets $B\left((\mathbf{x}_i,t_i),\rho_A,W_d^U(\alpha_A,\varepsilon/2)+V_d^U(\beta_A,\varepsilon/2)\right)$ are disjoint, so by (3.9) and (3.12), we obtain $N\le C(d)\exp(C(d)\varepsilon^{-2})\le\exp(C(d)\varepsilon^{-2})$.

If $\varepsilon\ge 8$ then (3.8) gives

$$\begin{aligned}
W_d^U(\alpha_A,\varepsilon)+V_d^U(\beta_A,\varepsilon)&\ge 8\sup_{(x^{(2)},\ldots,x^{(d)},t)\in U}\mathbb{E}\left|\sum_{\mathbf{i}}a_{\mathbf{i}}g_{i_1}\prod_{k=2}^d x^{(k)}_{i_k}t_{i_{d+1}}\right|\\
&=\sqrt{\frac{128}{\pi}}\sup_{(x^{(2)},\ldots,x^{(d)},t)\in U}\left(\sum_{i_1}\left(\sum_{i_2,\ldots,i_{d+1}}a_{\mathbf{i}}\prod_{k=2}^d x^{(k)}_{i_k}t_{i_{d+1}}\right)^2\right)^{1/2}\\
&\ge\operatorname{diam}(U,\rho_A).
\end{aligned}$$

So, $N(U,\rho_A,W_d^U(\alpha_A,\varepsilon)+V_d^U(\beta_A,\varepsilon))=1\le\exp(\varepsilon^{-2})$. □

Remark 3.5. The classical Dudley's bound on suprema of Gaussian processes (see e.g., [9, Corollary 5.1.6]) gives

$$\mathbb{E} \sup_{(x^{(2)},\ldots,x^{(d)},t)\in(B_2^n)^{d-1}\times T} \sum_{\mathbf{i}} a_{\mathbf{i}} g_{i_1} \prod_{k=2}^{d} x_{i_k}^{(k)} t_{i_{d+1}} \leq C \int_0^{\Delta} \sqrt{\log N((B_2^n)^{d-1}\times T, \rho_A, \delta)} d\delta,$$

where Δ is equal to the diameter of the set $(B_2^n)^{d-1}\times T$ in the metric ρ_A. Unfortunately the entropy bound derived in Corollary 3.4 involves a nonintegrable term δ^{-1}. The remaining part of the proof of Theorem 3.1 is devoted to improving on Dudley's bound.

For $\mathbf{x},\mathbf{y}\in(\mathbb{R}^n)^{d-1}$ we define a norm $\hat{\alpha}_A$ on $(\mathbb{R}^n)^{d-1}=\mathbb{R}^{(d-1)n}$ by the formula

$$\begin{aligned}\hat{\alpha}_A(x^{(2)},\ldots,x^{(d)}) &:= \sum_{j=2}^{d} \sum_{\substack{(\mathcal{P},\mathcal{P}')\in\mathcal{P}([d]\setminus\{j\})\\ |\mathcal{P}|=d-2}} \Big\| \sum_{i_j} a_{\mathbf{i}} x_{i_j}^{(j)} \Big\|_{\mathcal{P}' \mid \mathcal{P}} \\ &= \sum_{j=2}^{d} \sum_{\substack{\mathcal{P}\in\mathcal{P}([d]\setminus\{j\})\\ |\mathcal{P}|=d-2}} \Big\| \sum_{i_j} a_{\mathbf{i}} x_{i_j}^{(j)} \Big\|_{\emptyset \mid \mathcal{P}} + \sum_{j=2}^{d} \sum_{j\neq k=1}^{d} \sum_{\substack{\mathcal{P}\in\mathcal{P}([d]\setminus\{j,k\})\\ |\mathcal{P}|=d-2}} \Big\| \sum_{i_j} a_{\mathbf{i}} x_{i_j}^{(j)} \Big\|_{\{k\} \mid \mathcal{P}}.\end{aligned}$$

Proposition 3.6. *For any $d\geq 3$, $\varepsilon>0$ and $U\subset(B_2^n)^{d-1}\times T$,*

$$N\left(U, \rho_A, \sum_{k=0}^{d-2} \varepsilon^{d-k} s_k(A) + \varepsilon \sup_{(x^{(2)},\ldots,x^{(d)},t)\in U} \hat{\alpha}_A(x^{(2)},\ldots,x^{(d)})\right) \leq \exp(C(d)\varepsilon^{-2}).$$

Proof. We will estimate the quantities $W_d^U(\alpha_A,\varepsilon)$ and $V_d^U(\beta_A,\varepsilon)$ appearing in Corollary 3.4.

Since $U\subset(B_2^n)^{d-1}\times T$, Jensen's inequality yields for $I\subset\{2,\ldots,d\}$,

$$\begin{aligned} W_I^U(\alpha_A) &= \sup_{(x^{(2)},\ldots,x^{(d)},t)\in U} \mathbb{E}\alpha_A\left(\left(\bigotimes_{k=2}^{d}\left(x^{(k)}(1-\mathbf{1}_I(k)) + G^{(k)}\mathbf{1}_I(k)\right)\right)\otimes t\right) \\ &\leq \sup_{(x^{(2)},\ldots,x^{(d)},t)\in U} \sqrt{\mathbb{E}\sum_{i_1}\left(\sum_{i_2,\ldots,i_{d+1}} a_{\mathbf{i}} \prod_{k\in I} g_{i_k}^{(k)} \prod_{k\in[d]\setminus(I\cup\{1\})} x_{i_k}^{(k)} t_{i_{d+1}}\right)^2} \\ &= \sup_{(x^{(2)},\ldots,x^{(d)},t)\in U} \sqrt{\sum_{i_{I\cup\{1\}}}\left(\sum_{i_{[d+1]\setminus(I\cup\{1\})}} a_{\mathbf{i}} \prod_{k\in[d]\setminus(I\cup\{1\})} x_{i_k}^{(k)} t_{i_{d+1}}\right)^2} \\ &\leq \|A\|_{\emptyset \mid I\cup\{1\},\{k\}:k\in[d]\setminus(I\cup\{1\})} \leq s_{d-|I|}(A). \end{aligned} \tag{3.13}$$

By estimating a little more accurately in the second inequality in (3.13) we obtain for $2\leq j\leq d$,

$$\begin{aligned} &W_{\{j\}}^U(\alpha_A) \\ &\leq \sup_{(x^{(2)},\ldots,x^{(d)},t)\in U} \sum_{\substack{2\leq l\leq d\\ l\neq j}} \sup_{(y^{(2)},\ldots,y^{(d)})\in(B_2^n)^{d-1}} \sqrt{\sum_{i_1,i_j}\left(\sum_{i_{[d+1]\setminus\{1,j\}}} a_{\mathbf{i}} x_{i_l}^{(l)} \prod_{\substack{2\leq k\leq d\\ k\neq j,l}} y_{i_k}^{(k)} t_{i_{d+1}}\right)^2} \\ &\leq \sup_{(x^{(2)},\ldots,x^{(d)},t)\in U} \sum_{l=2}^{d} \sum_{\substack{\mathcal{P}\in\mathcal{P}([d]\setminus\{l\})\\ |\mathcal{P}|=d-2}} \Big\| \sum_{i_l} a_{\mathbf{i}} x_{i_l}^{(l)} \Big\|_{\emptyset \mid \mathcal{P}}. \end{aligned} \tag{3.14}$$

Observe that (3.14) is not true for $d=2$ (cf. Remark 3.7).

Let us now pass to the quantity $V_d^U(\beta_A,\varepsilon)$. The definition of V_I^U and the inclusion $U\subset (B_2^n)^{d-1}\times T$ yield

$$V_I^U(\beta_A)\le \|A\|_{\{1\}\{i\}:\ i\in I\ |\ \{k\}:k\in[d]\setminus(I\cup\{1\})}\le s_{d-|I|-1}(A)\quad \text{for } I\neq\emptyset \tag{3.15}$$

and

$$\begin{aligned} V_\emptyset^U(\beta_A)&\le \sup_{(x^{(2)},\ldots,x^{(d)},t)\in U}\sum_{l=2}^d \sup_{(y^{(2)},\ldots,y^{(d)})\in(B_2^n)^{d-1}}\mathbb{E}\sup_{t'\in T}\sum_{\mathbf{i}} a_{\mathbf{i}}g_{i_1}x_{i_l}^{(l)}\prod_{\substack{2\le k\le d\\ k\neq l}} y_{i_k}^{(k)}t'_{i_{d+1}}\\ &\le \sup_{(x^{(2)},\ldots,x^{(d)},t)\in U}\sum_{l=2}^d \left\|\sum_{i_j} a_{\mathbf{i}}x_{i_l}^{(l)}\right\|_{\{1\}\ |\ \{k\}:\ k\in[d]\setminus\{1,l\}}. \end{aligned}\tag{3.16}$$

Inequalities (3.13)–(3.16) imply that

$$\begin{aligned} &W_d^U(\alpha_A,\varepsilon)+V_d^U(\beta_A,\varepsilon)\\ &=\sum_{k=2}^{d-1}\varepsilon^k\sum_{I\subset\{2,\ldots,d\}:|I|=k} W_I^U(\alpha_A)+\sum_{k=1}^{d-1}\varepsilon^{k+1}\sum_{I\subset\{2,\ldots,d\}:|I|=k} V_I^U(\beta_A)\\ &+\varepsilon\left(\sum_{j=2}^d W_{\{j\}}^U(\alpha_A)+V_\emptyset^U(\beta_A)\right)\\ &\le C(d)\sum_{k=0}^{d-2}\varepsilon^{d-k}s_k(A)+d\varepsilon\sup_{(x^{(2)},\ldots,x^{(d)},t)\in U}\left(\sum_{l=2}^d\sum_{\substack{(\mathcal{P},\mathcal{P}')\in\mathcal{P}([d]\setminus\{l\})\\ |\mathcal{P}|=d-2}}\left\|\sum_{i_l} a_{\mathbf{i}}x_{i_l}^{(l)}\right\|_{\mathcal{P}'\ |\ \mathcal{P}}\right). \end{aligned}$$

Hence the assertion is a simple consequence of Corollary 3.4. □

Remark 3.7. Proposition 3.6 is not true for $d=2$. The problem arises in (3.14) – for $d=2$ there does not exist $\mathcal{P}\in\mathcal{P}([d]\setminus\{l\})$ such that $|\mathcal{P}|=d-2$. This is the main reason why proofs for chaoses of order $d=2$ (cf. [2]) have a different nature than for higher order chaoses.

4 Proof of Theorem 3.1

We will prove Theorem 3.1 by induction on d (recall that the matrix A has order $d+1$). To this end we need to amplify the induction hypothesis. For $U\subset(\mathbb{R}^n)^{d-1}\times\mathbb{R}^m$ we define

$$F_A(U)=\mathbb{E}\sup_{(x^{(2)},\ldots,x^{(d)},t)\in U}\sum_{i_1,\ldots,i_{d+1}} a_{i_1,\ldots,i_{d+1}}g_{i_1}\prod_{k=2}^d x_{i_k}^{(k)}t_{i_{d+1}}.$$

Theorem 4.1. *For any $U\subset(B_2^n)^{d-1}\times T$ and any $p\ge 1$*

$$F_A(U)\le C(d)\left(\sqrt{p}\Delta_A(U)+\sum_{k=0}^{d-1}p^{\frac{k+1-d}{2}}s_k(A)\right), \tag{4.1}$$

where

$$\begin{aligned}\Delta_A(U)&=\sup_{(x^{(2)},\ldots,x^{(d)},t),(y^{(2)},\ldots,y^{(d)},t')\in U}\rho_A((x^{(2)},\ldots,x^{(d)},t),(y^{(2)},\ldots,y^{(d)},t'))\\ &=\operatorname{diam}(A,\rho_A).\end{aligned}$$

Clearly it is enough to prove Theorem 4.1 for finite sets U. Observe that

$$\Delta_A((B_2^n)^{d-1} \times T) \leq 2 \|A\|_{\emptyset \mid \{j\}: j \in [d]} = 2 s_d(A),$$

thus Theorem 4.1 implies Theorem 3.1. We will prove (4.1) by induction on d, but first we will show several consequences of the theorem. In the next three lemmas, we shall assume that Theorem 4.1 (and thus also Theorem 3.1) holds for all matrices of order smaller than $d+1$.

Lemma 4.2. *Let $p \geq 1$, $l \geq 0$ and $d \geq 3$. Then*

$$N\left((B_2^n)^{d-1}, \hat{\rho}_A, 2^{-l} \sum_{k=0}^{d-1} p^{\frac{k+1-d}{2}} s_k(A)\right) \leq \exp(C(d) 2^{2l} p),$$

where $\hat{\rho}_A$ is the distance on $(\mathbb{R}^n)^{d-1}$ corresponding to the norm $\hat{\alpha}_A$.

Proof. Note that

$$\mathbb{E} \hat{\alpha}_A \left(G^{(2)}, \ldots, G^{(d)}\right) = \sum_{j=2}^{d} \sum_{\substack{(\mathcal{P}, \mathcal{P}') \in \mathcal{P}([d] \setminus \{j\}) \\ |\mathcal{P}| = d-2}} \mathbb{E} \left\| \sum_{i_j} a_{\mathbf{i}} g_{i_j} \right\|_{\mathcal{P}' \mid \mathcal{P}}. \tag{4.2}$$

Up to a permutation of the indexes we have two possibilities

$$\left\| \sum_{i_j} a_{\mathbf{i}} g_{i_j} \right\|_{\mathcal{P}' \mid \mathcal{P}} = \begin{cases} \left\| \sum_{i_j} a_{\mathbf{i}} g_{i_j} \right\|_{\emptyset \mid \{1,2\}, \{\{l\}:\ 3 \leq l \leq d,\ l \neq j\}} & \text{or} \\ \left\| \sum_{i_j} a_{\mathbf{i}} g_{i_j} \right\|_{\{1\} \mid \{l\}: 2 \leq l \leq d,\ l \neq j}. & \end{cases} \tag{4.3}$$

First assume that $\left\| \sum_{i_j} a_{\mathbf{i}} g_{i_j} \right\|_{\mathcal{P}' \mid \mathcal{P}} = \left\| \sum_{i_j} a_{\mathbf{i}} g_{i_j} \right\|_{\emptyset \mid \{1,2\}, \{\{l\}:\ 3 \leq l \leq d,\ l \neq j\}}$. In this case

$$\left\| \sum_{i_j} a_{\mathbf{i}} g_{i_j} \right\|_{\emptyset \mid \{1,2\}, \{\{l\}:\ 3 \leq l \leq d,\ l \neq j\}} = \left\| \sum_{i_1} b_{i_1, \ldots, i_d} g_{i_1} \right\|_{\emptyset \mid \{2\}, \ldots, \{d-1\}}$$

for an appropriately chosen matrix $B = (b_{i_1, \ldots, i_d})$ (we treat a pair of indices $\{1,2\}$ as a single index and renumber the indices in such a way that j, $\{1,2\}$ and $d+1$ would become 1, 2 and d respectively).

Clearly,

$$\sum_{\substack{(\mathcal{P}', \mathcal{P}) \in \mathcal{P}([d-1]) \\ |\mathcal{P}| = k}} \|B\|_{\mathcal{P}' \mid \mathcal{P}} = \sum_{\substack{(\mathcal{P}', \mathcal{P}) \in \mathcal{C} \\ |\mathcal{P}| = k}} \|A\|_{\mathcal{P}' \mid \mathcal{P}} \leq \sum_{\substack{(\mathcal{P}', \mathcal{P}) \in \mathcal{P}([d]) \\ |\mathcal{P}| = k}} \|A\|_{\mathcal{P}' \mid \mathcal{P}} = s_k(A), \tag{4.4}$$

where $\mathcal{C} \subset \mathcal{P}([d])$ is the set of partitions which do not separate 1 and 2.

Thus, Theorem 3.1 applied to the matrix B of order d yields

$$\begin{aligned} \mathbb{E} \left\| \sum_{i_j} a_{\mathbf{i}} g_{i_j} \right\|_{\emptyset \mid \{1,2\}, \{\{l\}:\ 3 \leq l \leq d,\ l \neq j\}} &= \mathbb{E} \left\| \sum_{i_1} b_{i_1, \ldots, i_d} g_{i_1} \right\|_{\emptyset \mid \{2\}, \ldots, \{d-1\}} \\ &\leq C(d) \sum_{(\mathcal{P}', \mathcal{P}) \in \mathcal{P}([d-1])} p^{\frac{|\mathcal{P}|+2-d}{2}} \|B\|_{\mathcal{P}' \mid \mathcal{P}} \leq C(d) \sum_{k=0}^{d-1} p^{\frac{k+2-d}{2}} s_k(A). \end{aligned} \tag{4.5}$$

Now assume that $\left\|\sum_{i_j} a_{\mathbf{i}} g_{i_j}\right\|_{\mathcal{P}' \mid \mathcal{P}} = \left\|\sum_{i_j} a_{\mathbf{i}} g_{i_j}\right\|_{\{1\} \mid \{l\}: 2\le l\le d,\ l\neq j}$ and observe that

$$\begin{aligned}\mathbb{E}\left\|\sum_{i_j} a_{\mathbf{i}} g_{i_j}\right\|_{\{1\} \mid \{l\}: 2\le l\le d,\ l\neq j} &= \mathbb{E}^{g} \sup_{x^{(l)}\in B_2^n,\ 2\le l\le d,\ l\neq j} \mathbb{E}^{g'} \sup_{t\in T} \sum_{\mathbf{i}} a_{\mathbf{i}} g'_{i_1} g_{i_j} \prod_{2\le l\le d,\ l\neq j} x^{(l)}_{i_l} t_{i_{d+1}} \\ &= \mathbb{E} \sup_{x^{(l)}\in B_2^n,\ 2\le l\le d,\ l\neq j} \sup_{m\in\mathcal{M}} \sum_{\mathbf{i}} a_{\mathbf{i}} g_{i_j} \prod_{2\le l\le d,\ l\neq j} x^{(l)}_{i_l} m_{i_1,i_{d+1}} \\ &= \mathbb{E} \sup_{x^{(l)}\in B_2^n,\ 2\le l\le d-1} \sup_{m\in\widetilde{\mathcal{M}}} \sum_{i_1,\dots,i_d} d_{i_1,\dots,i_d} g_{i_1} \prod_{l=2}^{d-1} x^{(l)}_{i_l} m_{i_d},\end{aligned}$$

where $D = (d_{i_1,\dots,i_d})_{i_1,\dots,i_d}$ is an appropriately chosen matrix of order d, the symmetric set $\mathcal{M} \subset \mathbb{R}^n \otimes \mathbb{R}^m$ satisfies

$$\mathbb{E}\sup_{t\in T} \sum_{i,j} b_{i,j} g_i t_j = \sup_{m\in\mathcal{M}} \sum_{i,j} b_{i,j} m_{i,j} \text{ for any matrix } (b_{i,j})_{i\le n, j\le m},$$

and $\widetilde{\mathcal{M}}$ corresponds to $\mathcal{M}$ under a natural identification of $\mathbb{R}^n \otimes \mathbb{R}^m$ with $\mathbb{R}^{nm}$.

Applying Theorem 3.1 to the matrix D of order d gives

$$\begin{aligned}\mathbb{E}\left\|\sum_{i_j} a_{\mathbf{i}} g_{i_j}\right\|_{\{1\} \mid \{l\}: 2\le l\le d,\ l\neq j} &= \mathbb{E} \sup_{x^{(l)}\in B_2^n,\ 2\le l\le d-1} \sup_{m\in\widetilde{\mathcal{M}}} \sum_{i_1,\dots,i_d} d_{i_1,\dots,i_d} g_{i_1} \prod_{l=2}^{d-1} x^{(l)}_{i_l} m_{i_d} \\ &\le C(d) \sum_{(\mathcal{P}',\mathcal{P})\in\mathcal{P}([d-1])} p^{\frac{|\mathcal{P}|+2-d}{2}} \|D\|^{\widetilde{\mathcal{M}}}_{\mathcal{P}' \mid \mathcal{P}} \\ &\le C(d) \sum_{(\mathcal{P}',\mathcal{P})\in\mathcal{P}([d])} p^{\frac{|\mathcal{P}|+2-d}{2}} \|A\|_{\mathcal{P}' \mid \mathcal{P}} \\ &= C(d) \sum_{k=0}^{d-1} p^{\frac{k+2-d}{2}} s_k(A), \end{aligned} \tag{4.6}$$

where $\|D\|^{\widetilde{\mathcal{M}}}_{\mathcal{P}' \mid \mathcal{P}}$ is defined in the same manner as $\|A\|_{\mathcal{P}' \mid \mathcal{P}}$ (see (3.1)) but the supremum is taken over the set $\widetilde{\mathcal{M}}$ instead of T. The second inequality in (4.6) can be justified analogously as (4.4).

Combining (4.2), (4.3), (4.5), (4.6) and the dual Sudakov inequality (Theorem A.3, note that $(B_2^n)^{d-1} \subseteq \sqrt{d-1} B_2^{n(d-1)}$) we obtain

$$\begin{aligned} &N\left((B_2^n)^{d-1}, \hat{\rho}_A, t\sum_{k=0}^{d-1} p^{\frac{k+2-d}{2}} s_k(A)\right) \\ &\qquad\le N\left((B_2^n)^{d-1}, \hat{\rho}_A, C(d)^{-1} t\mathbb{E}\hat{\alpha}_A(G^{(2)},\dots,G^{(n)})\right) \le \exp(C(d)t^{-2}).\end{aligned}$$

It is now enough to choose $t = (\sqrt{p}2^l)^{-1}$. □

From now on for $U \subseteq (\mathbb{R}^n)^{d-1} \times \mathbb{R}^m$ we denote

$$\hat{\alpha}_A(U) = \sup_{(x^{(2)},\dots,x^{(d)},t)\in U} \hat{\alpha}_A\left(x^{(2)},\dots,x^{(d)}\right).$$

Lemma 4.3. *Suppose that $d \ge 3$, $\mathbf{y} = (y^{(2)},\dots,y^{(d)}) \in (B_2^n)^{d-1}$ and $U \subset (B_2^n)^{d-1} \times T$. Then for any $p \ge 1$ and $l \ge 0$, we can find a decomposition*

$$U = \bigcup_{j=1}^{N} U_j,\ N \le \exp(C(d)2^{2l}p)$$

such that for each $j \leq N$*,*

$$F_A((\mathbf{y},0)+U_j) \leq F_A(U_j) + C(d)\left(\hat{\alpha}_A(\mathbf{y}) + \hat{\alpha}_A(U) + 2^{-l}\sum_{k=0}^{d-2} p^{\frac{k+1-d}{2}} s_k(A)\right) \tag{4.7}$$

and

$$\Delta_A(U_j) \leq 2^{-l} p^{-1/2}\hat{\alpha}_A(U) + 2^{-2l}\sum_{k=0}^{d-2} p^{\frac{k-d}{2}} s_k(A). \tag{4.8}$$

Proof. Fix $\mathbf{y} \in (B_2^n)^{d-1}$ and $U \subset (B_2^n)^{d-1}\times T$. For $I \subset \{2,\ldots,d\}$, $\mathbf{x} = (x^{(2)},\ldots,x^{(d)},t)$, $\tilde{\mathbf{x}} = (\tilde{x}^{(2)},\ldots,\tilde{x}^{(d)},t') \in (\mathbb{R}^n)^{d-1}\times\mathbb{R}^m$ and $S \subset (\mathbb{R}^n)^{d-1}\times\mathbb{R}^m$, we define

$$\rho_A^{\mathbf{y},I}(\mathbf{x},\tilde{\mathbf{x}}) := \sqrt{\sum_{i_1}\left(\sum_{i_2,\ldots,i_{d+1}} a_{\mathbf{i}}\prod_{k\in I} y_{i_k}^{(k)}\left(t_{i_{d+1}}\prod_{\substack{2\leq j\leq d\\ j\notin I}} x_{i_j}^{(j)} - t'_{i_{d+1}}\prod_{\substack{2\leq j\leq d\\ j\notin I}}\tilde{x}_{i_j}^{(j)}\right)\right)^2},$$

$$\Delta_A^{\mathbf{y},I}(S) := \sup\left\{\rho_A^{\mathbf{y},I}(\mathbf{x},\tilde{\mathbf{x}}) : \mathbf{x},\tilde{\mathbf{x}}\in S\right\}$$

and

$$F_A^{\mathbf{y},I}(S) := \mathbb{E}\sup_{(x^{(2)},\ldots,x^{(d)},t)\in S}\sum_{\mathbf{i}} a_{\mathbf{i}} g_{i_1}\prod_{k\in I} y_{i_k}^{(k)}\left(\prod_{\substack{2\leq j\leq d\\ j\notin I}} x_{i_j}^{(j)}\right) t_{i_{d+1}}.$$

If $I = \{2,\ldots,d\}$ then for $S \subset (B_2^n)^{d-1}\times T$ we have

$$\begin{aligned}
F_A^{\mathbf{y},\{2,\ldots,d\}}(S) &\leq \mathbb{E}\sup_{t\in T}\sum_{\mathbf{i}} a_{\mathbf{i}} g_{i_1}\prod_{k=2}^{d} y_{i_k}^{(k)} t_{i_{d+1}} \\
&\leq \sup_{(x^{(2)},\ldots,x^{(d-1)})\in(B_2^n)^{d-2}} \mathbb{E}\sup_{t\in T}\sum_{\mathbf{i}} a_{\mathbf{i}} g_{i_1}\left(\prod_{j=2}^{d-1} x_{i_j}^{(j)}\right) y_{i_d}^{(d)} t_{i_{d+1}} \\
&= \left\|\sum_{i_d} a_{\mathbf{i}} y_{i_d}^{(d)}\right\|_{\{1\}\mid\{k\}\,:\,k=2,\ldots,d-1} \leq \hat{\alpha}_A(\mathbf{y}).
\end{aligned} \tag{4.9}$$

If $\emptyset \neq I \subsetneq \{2,\ldots,d\}$ then Theorem 4.1 applied to the matrix

$$A(\mathbf{y},I) := \left(\sum_{i_I} a_{\mathbf{i}}\prod_{k\in I} y_{i_k}^{(k)}\right)_{\mathbf{i}_{[d+1]\setminus I}}$$

of order $d-|I|+1 < d+1$ gives for any $S \subset (B_2^n)^{d-1}\times T$ and $q \geq 1$,

$$F_A^{\mathbf{y},I}(S) \leq C(d-|I|)\left(q^{1/2}\Delta_A^{\mathbf{y},I}(S) + \sum_{k=0}^{d-|I|-1} q^{\frac{k+1-d+|I|}{2}} s_k(A(\mathbf{y},I))\right).$$

For any $2 \leq k \leq d$, $y^{(k)} \in B_2^n$, thus $s_k(A(\mathbf{y},I)) \leq s_{k+|I|}(A)$ for $k < d-|I|-1$ and $s_{d-|I|-1}(A(\mathbf{y},I)) \leq \hat{\alpha}_A(\mathbf{y})$.

Hence,

$$F_A^{\mathbf{y},I}(S) \leq C(d-|I|)\left(q^{1/2}\Delta_A^{\mathbf{y},I}(S) + \hat{\alpha}_A(\mathbf{y}) + \sum_{k=0}^{d-2} q^{\frac{k+1-d}{2}} s_k(A)\right). \tag{4.10}$$

By the triangle inequality,

$$F_A((y^{(2)},\ldots,y^{(d)},0)+S)-F_A(S)\le\sum_{\emptyset\neq I\subset\{2,\ldots,d\}}F_A^{\mathbf{y},I}(S).$$

Combining (4.9) and (4.10) we obtain for $S\subset(B_2^n)^{d-1}\times T$ and $q\ge1$,

$$\begin{aligned}&F_A((y^{(2)},\ldots,y^{(d)},0)+S)\\&\quad\le F_A(S)+C(d)\left(\hat{\alpha}_A(\mathbf{y})+\sum_{\emptyset\neq I\subsetneq\{2,\ldots,d\}}q^{1/2}\Delta_A^{\mathbf{y},I}(S)+\sum_{k=0}^{d-2}q^{\frac{k+1-d}{2}}s_k(A)\right).\end{aligned}\tag{4.11}$$

Fix $I\subsetneq\{2,\ldots,d\}$, $|I|<d-2$ (we do not exclude $I=\emptyset$). Taking supremum over $\mathbf{y}\in(B_2^n)^{d-1}$ we conclude that

$$\sup_{(x^{(2)},\ldots,x^{(d)},t)\in U}\hat{\alpha}_{A(\mathbf{y},I)}((x^{(k)})_{k\in\{2,\ldots,d\}\setminus I})\le\sup_{(x^{(2)},\ldots,x^{(d)},t)\in U}\hat{\alpha}_A(x^{(2)},\ldots,x^{(d)}).$$

Recall also that $s_k(A(\mathbf{y},I))\le s_{k+|I|}(A)$, thus we may apply $2^{d-1}-d$ times Proposition 3.6 with $\varepsilon=2^{-l}p^{-1/2}$ and find a decomposition $U=\bigcup_{j=1}^{N_1}U_j'$, $N_1\le\exp(C(d)2^{2l}p)$ such that for each j and $I\subset\{2,\ldots,d\}$ with $|I|<d-2$,

$$\Delta_A^{\mathbf{y},I}(U_j')\le2^{-l}p^{-1/2}\hat{\alpha}_A(U)+2^{-2l}\sum_{k=0}^{d-2}p^{\frac{k-d}{2}}s_k(A).\tag{4.12}$$

If $|I|=d-2$ then the distance $\rho_A^{\mathbf{y},I}$ corresponds to a norm $\alpha_{A(\mathbf{y},I)}$ on $\mathbb{R}^{nm}$ given by

$$\alpha_{A(\mathbf{y},I)}(\mathbf{x})=\sqrt{\sum_{i_1}\left(\sum_{i_2,\ldots,i_{d+1}}a_{\mathbf{i}}x_{i_{\{j,d+1\}}}\prod_{k\in I}y_{i_k}^{(k)}\right)^2},$$

where j is defined by the condition $\{1,j\}=[d]\setminus I$ (cf. (3.3), (3.4) and observe that $A(\mathbf{y},I)$ is an $n\times m$ matrix). We define also (as in (3.6))

$$\beta_{A(\mathbf{y},I)}(\mathbf{x})=\mathbb{E}\sup_{t\in T}\sum_{\mathbf{i}}a_{\mathbf{i}}g_{i_1}x_{i_j}\prod_{l\in I}y_{i_l}^{(l)}t_{i_{d+1}}.$$

Recall the definitions (3.5) and (3.7) and note that (denoting by $\widetilde{U}$ the projection of U onto the j-th and $(d+1)$-th coordinate)

$$\begin{aligned}W_2^{\widetilde{U}}(\alpha_{A(\mathbf{y},I)},\varepsilon)&=\varepsilon\sup_{(x^{(2)},\ldots,x^{(d)},t)\in U}\mathbb{E}\sqrt{\sum_{i_1}\left(\sum_{i_2,\ldots,i_{d+1}}a_{\mathbf{i}}g_{i_j}t_{i_{d+1}}\prod_{k\in I}y_{i_k}^{(k)}\right)^2}\\&\le\varepsilon\sup_{(x^{(2)},\ldots,x^{(d)},t)\in U}\sqrt{\sum_{i_1,i_j}\left(\sum_{\mathbf{i}_I}a_{\mathbf{i}}\prod_{k\in I}y_{i_k}^{(k)}t_{i_{d+1}}\right)^2}\le\varepsilon\hat{\alpha}(\mathbf{y}),\end{aligned}\tag{4.13}$$

where we again used that $y_k\in B_2^n$, $U\subset(B_2^n)^{d-1}\times T$.

We also have

$$V_{\{j\}}^{\widetilde{U}}(\beta_{A(\mathbf{y},I)})=\mathbb{E}\sup_{t\in T}\sum_{\mathbf{i}}a_{\mathbf{i}}g_{i_1}^{(1)}g_{i_j}^{(2)}\prod_{k\in I}y_{i_k}^{(k)}t_{i_{d+1}}\le s_{d-2}(A)$$

and

$$V_{\emptyset}^{\widetilde{U}}(\beta_{A(\mathbf{y},I)}) = \sup_{(x^{(2)},\ldots,x^{(d)},t)\in U} \mathbb{E}\sup_{t'\in T}\sum_{\mathbf{i}} a_{\mathbf{i}} g_{i_1} x^{(j)}_{i_j}\prod_{k\in I} y^{(k)}_{i_k} t'_{i_{d+1}} \le \hat{\alpha}(\mathbf{y}).$$

Thus

$$V_2^{\widetilde{U}}(\beta_{A(\mathbf{y},I)},\varepsilon) \le \varepsilon\hat{\alpha}(\mathbf{y}) + \varepsilon^2 s_{d-2}(A).$$

Taking $\varepsilon = 2^{-l-1}p^{-1/2}$ and combining the above estimate with (4.13) and Corollary 3.4 (applied $d-1$ times) we obtain a partition $U = \bigcup_{j=1}^{N_2} U_j''$ with $N_2 \le \exp(C(d)2^{2l}p)$ and

$$\Delta_A^{\mathbf{y},I}(U_j'') \le 2^{-l}p^{-1/2}\hat{\alpha}_A(\mathbf{y}) + 2^{-2l}p^{-1}s_{d-2}(A) \tag{4.14}$$

for any $I\subset\{2,\ldots,d\}$ with $|I| = d-2$ and $j\le N_2$.

Intersecting the partition $(U_i')_{i\le N_1}$ (which fulfills (4.12)) with $(U_j'')_{j\le N_2}$ we obtain a partition $U = \bigcup_{i=1}^{N} U_i$ with $N \le N_1N_2 \le \exp(C(d)2^{2l}p)$ and such that for every $i\le N$ there exist $j\le N_1$ and $l\le N_2$ such that $U_i\subset U_j'\cap U_l''$.

Inequality (4.7) follows by (4.11) with $q = 2^{2l}p$, (4.12) and (4.14). Observe that (4.8) follows by (4.12) for $I=\emptyset$. □

Lemma 4.4. *Suppose that U is a finite subset of $(B_2^n)^{d-1}\times T$, with $|U|\ge 2$ and $U-U\subset (B_2^n)^{d-1}\times(T-T)$. Then for any $p\ge 1, l\ge 0$ there exist finite sets $U_i\subset (B_2^n)^{d-1}\times T$ and $(\mathbf{y}_i,t_i)\in U$, $i=1,\ldots,N$ such that*

(i) $2\le N\le \exp(C(d)2^{2l}p)$,

(ii) $U = \bigcup_{i=1}^N((\mathbf{y}_i,0)+U_i), (U_i-U_i)\subset U-U, |U_i|\le |U|-1$,

(iii) $\Delta_A(U_i)\le 2^{-2l}\sum_{k=0}^{d-1}p^{\frac{k-d}{2}}s_k(A)$,

(iv) $\hat{\alpha}_A(U_i)\le 2^{-l}\sum_{k=0}^{d-1}p^{\frac{k+1-d}{2}}s_k(A)$,

(v) $F_A((\mathbf{y}_i,0)+U_i)\le F_A(U_i)+C(d)\left(\hat{\alpha}_A(U)+2^{-l}\sum_{k=0}^{d-1}p^{\frac{k+1-d}{2}}s_k(A)\right)$.

Proof. By Lemma 4.2 we get

$$(B_2^n)^{d-1} = \bigcup_{i=1}^{N_1} B_i,\ N_1\le\exp(C(d)2^{2l}p),$$

where the diameter of the sets B_i in the norm $\hat{\alpha}$ satisfies

$$\operatorname{diam}(B_i,\hat{\alpha}_A)\le 2^{-l}\sum_{k=0}^{d-1}p^{\frac{k+1-d}{2}}s_k(A).$$

Let $U_i = U\cap(B_i\times T)$. Selecting arbitrary $(\mathbf{y}_i,t_i)\in U_i$ (we can assume that these sets are nonempty) and using Lemma 4.3 (with $l+1$ instead of l) we decompose $U_i-(\mathbf{y}_i,0)$ into $\bigcup_{j=1}^{N_2}U_{ij}$ in such a way that $N_2\le\exp(C(d)2^{2l}p)$,

$$\begin{aligned}F_A((\mathbf{y}_i,0)+U_{ij}) &\le F_A(U_{ij})+C(d)\left(\hat{\alpha}_A(\mathbf{y}_i)+\hat{\alpha}_A(U_i-(\mathbf{y}_i,0))+2^{-l}\sum_{k=0}^{d-2}p^{\frac{k+1-d}{2}}s_k(A)\right)\\ &\le F_A(U_{ij})+C(d)\left(\hat{\alpha}_A(\mathbf{y}_i)+\operatorname{diam}(B_i,\hat{\alpha}_A)+2^{-l}\sum_{k=0}^{d-2}p^{\frac{k+1-d}{2}}s_k(A)\right)\\ &\le F_A(U_{ij})+C(d)\left(\hat{\alpha}_A(U)+2^{-l}\sum_{k=0}^{d-1}p^{\frac{k+1-d}{2}}s_k(A)\right)\end{aligned}$$

and

$$\Delta_A(U_{ij}) \le 2^{-l-1}p^{-1/2}\hat{\alpha}_A\left(U_i-(\mathbf{y}_i,0)\right)+2^{-2l-2}\sum_{k=0}^{d-2}p^{\frac{k-d}{2}}s_k(A) \le 2^{-2l}\sum_{k=0}^{d-1}p^{\frac{k-d}{2}}s_k(A).$$

We take the decomposition $U=\bigcup_{i,j}((\mathbf{y}_i,0)+U_{ij})$. We have $N=N_1N_2\le\exp(C(d)2^{2l}p)$. Without loss of generality we can assume $N\ge 2$ and $|U_{i,j}|\le|U|-1$. Obviously, $U_{ij}-U_{ij}\subset U_i-U_i\subset U-U$ and $\hat{\alpha}_A\left(U_{ij}\right)\le\hat{\alpha}_A\left(U_i-(\mathbf{y}_i,0)\right)\le 2^{-l}\sum_{k=0}^{d-1}p^{\frac{k+1-d}{2}}s_k(A)$. A relabeling of the obtained decomposition concludes the proof. □

Proof of Theorem 4.1. In the case of $d=2$ Theorem 4.1 is proved in [2] (see Remark 37 therein).

Assuming (4.1) to hold for $\{2,\dots,d-1\}$, we will prove it for $d\ge 3$. Let $U\subset(\mathbb{R}^n)^{d-1}\times T$ and let us put $\Delta_0=\Delta_A(U)$, $\hat{\Delta}_0=\hat{\alpha}_A\left((B_2^n)^{d-1}\times T\right)\le C(d)s_{d-1}(A)$,

$$\Delta_l:=2^{2-2l}\sum_{k=0}^{d-1}p^{\frac{k-d}{2}}s_k(A),\ \hat{\Delta}_l:=2^{1-l}\sum_{k=0}^{d-1}p^{\frac{k+1-d}{2}}s_k(A)\text{ for }l\ge 1.$$

Suppose first that $U\subset(\frac{1}{2}(B_2^n)^{d-1})\times T$ and define

$$c_U(r,l):=\sup\Big\{F_A(S):\ S\subset(B_2^n)^{d-1}\times T, S-S\subset U-U,\\ |S|\le r,\Delta_A(S)\le\Delta_l,\hat{\alpha}_A\left(S\right)\le\hat{\Delta}_l\Big\}.$$

Note that any subset $S\subset U$ satisfies $\Delta_A(S)\le\Delta_0$ and $\hat{\alpha}_A\left(S\right)\le\hat{\Delta}_0$, therefore,

$$c_U(r,0)\ge\sup\{F_A(S):S\subset U,|S|\le r\}. \tag{4.15}$$

We will now show that for $r\ge 2$,

$$c_U(r,l)\le c_U(r-1,l+1)+C(d)\left(\hat{\Delta}_l+2^l\sqrt{p}\Delta_l+2^{-l}\sum_{k=0}^{d-1}p^{\frac{k+1-d}{2}}s_k(A)\right). \tag{4.16}$$

Indeed, let us take $S\subset(B_2^n)^{d-1}\times T$ as in the definition of $c_U(r,l)$. Then by Lemma 4.4 we may find a decomposition $S=\bigcup_{i=1}^N((\mathbf{y}_i,0)+S_i)$ satisfying (i)–(v) with U, U_i replaced by S, S_i. Hence, by Lemma A.4, we have

$$\begin{aligned}F_A(S)&\le C\sqrt{\log N}\Delta_A(S)+\max_i F_A((\mathbf{y}_i,0)+S_i)\\ &\le C(d)\left(2^l\sqrt{p}\Delta_l+\hat{\alpha}_A\left(S\right)+2^{-l}\sum_{k=0}^{d-1}p^{\frac{k+1-d}{2}}s_k(A)\right)+\max_i F_A(S_i).\end{aligned} \tag{4.17}$$

We have $\Delta_A(S_i)\le\Delta_{l+1}$, $\hat{\alpha}_A\left(S_i\right)\le\hat{\Delta}_{l+1},\ S_i-S_i\subset S-S\subset U-U$ and $|S_i|\le|S|-1\le r-1$, thus $\max_i F_A(S_i)\le c_U(r-1,l+1)$ and (4.17) yields (4.16). Since $c_U(1,l)=0$, (4.16) yields

$$c_U(r,0)\le C(d)\sum_{l=0}^{\infty}\left(\hat{\Delta}_l+2^l\sqrt{p}\Delta_l+2^{-l}\sum_{k=0}^{d-1}p^{\frac{k+1-d}{2}}s_k(A)\right).$$

For $U\subset(\frac{1}{2}(B_2^n)^{d-1})\times T$, we have by (4.15)

$$\begin{aligned}F_A(U)=\sup\{F_A(S):S\subset U,|S|<\infty\}&\le\sup_r c_U(r,0)\\ &\le C(d)\left(\sqrt{p}\Delta_A(U)+\sum_{k=0}^{d-1}p^{\frac{k+1-d}{2}}s_k(A)\right).\end{aligned}$$

Finally, if $U \subset (B_2^n)^{d-1} \times T$, then $U' := \{(\mathbf{y}/2, t) \colon (\mathbf{y}, t) \in U\} \subset (\frac{1}{2}(B_2^n)^{d-1}) \times T$ and $\Delta_A(U') = 2^{1-d}\Delta_A(U)$, hence,

$$F_A(U) = 2^{d-1}F_A(U') \le C(d)\left(\sqrt{p}\Delta_A(U) + \sum_{k=0}^{d-1} p^{\frac{k+1-d}{2}} s_k(A)\right). \qquad \square$$

5 Proofs of main results

We return to the notation used Section 2. In particular in this section the multi-index $\mathbf{i}$ takes values in $[n]^d$ (instead of $[n]^d \times [m]$ as we had in the two previous sections) and all summations over $\mathbf{i}$ should be understood as summations over $[n]^d$.

5.1 Proofs of Theorems 2.1 and 2.5

Proof of Theorem 2.1. We start with the lower bound. Fix $J \subset [d]$, $\mathcal{P} \in \mathcal{P}([d] \setminus J)$ and observe that

$$\begin{aligned}
\left\|\sum_{\mathbf{i}} a_{\mathbf{i}} \prod_{k=1}^d g_{i_k}^{(k)}\right\|_p &\ge \left(\mathbb{E}^{(G^{(j)}):j\in J} \sup_{\substack{\varphi \in F^* \\ \|\varphi\| \le 1}} \mathbb{E}^{(G^{(j)}):j\in[d]\setminus J}\left|\varphi\left(\sum_{\mathbf{i}} a_{\mathbf{i}} \prod_{k=1}^d g_{i_k}^{(k)}\right)\right|^p\right)^{1/p} \\
&\ge c(d)\left(\mathbb{E}^{(G^{(j)}):j\in J} p^{\frac{p|\mathcal{P}|}{2}} \left\|\left(\sum_{i_J} a_{\mathbf{i}} \prod_{j\in J} g_{i_j}^{(j)}\right)_{i_{[d]\setminus J}}\right\|_{\mathcal{P}}^p\right)^{1/p} \\
&\ge c(d) p^{\frac{|\mathcal{P}|}{2}} \mathbb{E}\left\|\left(\sum_{i_J} a_{\mathbf{i}} \prod_{j\in J} g_{i_j}^{(j)}\right)_{i_{[d]\setminus J}}\right\|_{\mathcal{P}} = c(d) p^{\frac{|\mathcal{P}|}{2}} \| A \|_{\mathcal{P}},
\end{aligned}$$

where F^* is the dual space and in the second inequality we used Theorem A.6.

The upper bound will be proved by an induction on d. For $d = 2$ it is shown in [2]. Suppose that $d \ge 3$ and the estimate holds for F-valued matrices of order $2, \ldots, d-1$. By the induction assumption, we have

$$\left\|\sum_{\mathbf{i}} a_{\mathbf{i}} \prod_{k=1}^d g_{i_k}^{(k)}\right\|_p \le C(d) \sum_{(\mathcal{P},\mathcal{P}')\in\mathcal{P}([d-1])} p^{\frac{|\mathcal{P}|}{2}} \left\|\left\|\sum_{i_d} a_{\mathbf{i}} g_{i_d}\right\|_{\mathcal{P}' \mid \mathcal{P}}\right\|_p. \tag{5.1}$$

Since $\|\cdot\|_{\mathcal{P}' \mid \mathcal{P}}$ is a norm Lemma A.5 yields

$$\left\|\left\|\sum_{i_d} a_{\mathbf{i}} g_{i_d}\right\|_{\mathcal{P}' \mid \mathcal{P}}\right\|_p \le C\mathbb{E}\left\|\sum_{i_d} a_{\mathbf{i}} g_{i_d}\right\|_{\mathcal{P}' \mid \mathcal{P}} + C\sqrt{p}\,\|A\|_{\mathcal{P}' \mid \mathcal{P}\cup\{d\}}. \tag{5.2}$$

Choose $\mathcal{P} = (I_1 \ldots, I_k), \mathcal{P}' = (J_1, \ldots, J_m)$ and denote $J = \bigcup \mathcal{P}'$. By the definition of $\|A\|_{\mathcal{P}' \mid \mathcal{P}}$ we have

$$\begin{aligned}
&\left\|\sum_{i_d} a_{\mathbf{i}} g_{i_d}\right\|_{\mathcal{P}' \mid \mathcal{P}} \\
&= \sup\left\{\mathbb{E}^{(G^{(1)},\ldots,G^{(m)})}\left\|\sum_{\mathbf{i}} a_{\mathbf{i}} x_{i_{I_1}}^{(1)} \cdots x_{i_{I_k}}^{(k)} \prod_{l=1}^m g_{i_{J_l}}^{(l)} g_{i_d}^{(d)}\right\| \;\Big|\; \forall_{j=1\ldots,k} \sum_{i_{I_j}} \left(x_{i_{I_j}}^{(j)}\right)^2 = 1\right\} \\
&= \sup\left\{\left\|\left(\sum_{i_{[d]\setminus J}} a_{\mathbf{i}} x_{i_{I_1}}^{(1)} \cdots x_{i_{I_k}}^{(k)} g_{i_d}^{(d)}\right)_{i_J}\right\| \;\Big|\; \forall_{j=1\ldots,k} \sum_{i_{I_j}} \left(x_{i_{I_j}}^{(j)}\right)^2 = 1\right\},
\end{aligned} \tag{5.3}$$

where $G^{(l)} = (g^{(l)}_{i_{J_l}})_{i_{J_l}}$ and $\|\!|\cdot|\!\|$ is a norm on $F^{n^{|J|}}$ given by

$$\|\!|(b_{i_J})_{i_J}|\!\| = \mathbb{E}\left\| \sum_{i_J} b_{i_J} \prod_{l=1}^{m} g^{(l)}_{i_{J_l}} \right\|.$$

Theorem 2.6 implies that

$$\begin{aligned}
&\mathbb{E}\sup\left\{ \left|\!\left|\!\left| \left(\sum_{i_{[d]\setminus J}} a_{\mathbf{i}} x^{(1)}_{i_{I_1}} \cdots x^{(k)}_{i_{I_k}} g^{(d)}_{i_d} \right)_{i_J} \right|\!\right|\!\right| \;\Big|\; \forall_{j=1\ldots,k} \sum_{i_{I_j}} \left(x^{(j)}_{i_{I_j}}\right)^2 = 1 \right\} \\
&\quad \le C(k) \sum_{(\mathcal{R}',\mathcal{R})\in\mathcal{P}([d]\setminus J)} p^{\frac{|\mathcal{R}|-k}{2}} \|\!|\hat{A}|\!\|_{\mathcal{R}' \mid \mathcal{R}} = C(k) \sum_{(\mathcal{R}',\mathcal{R})\in\mathcal{P}([d]\setminus J)} p^{\frac{|\mathcal{R}|-k}{2}} \|A\|_{\mathcal{R}'\cup\mathcal{P}' \mid \mathcal{R}} \\
&\quad \le C(k) \sum_{(\mathcal{R}',\mathcal{R})\in\mathcal{P}([d])} p^{\frac{|\mathcal{R}|-k}{2}} \|A\|_{\mathcal{R}' \mid \mathcal{R}},
\end{aligned}$$

where $\hat{A} := (\hat{a}_{i_{[d]\setminus J}})_{i_{[d]\setminus J}}$ is $F^{n^{|J|}}$-valued matrix of order $d-|J|$ given by $\hat{a}_{i_{[d]\setminus J}} = (a_{\mathbf{i}})_{i_J}$ and $\|\!|A|\!\|_{\mathcal{R}' \mid \mathcal{R}}$ is defined in a similar way as $\|A\|_{\mathcal{R}' \mid \mathcal{R}}$ but under the expectation occurs the norm $\|\!|\cdot|\!\|$.

The above and (5.3) yield

$$\mathbb{E}\left\| \sum_{i_d} a_{\mathbf{i}} g_{i_d} \right\|_{\mathcal{P}' \mid \mathcal{P}} \le C(k) \sum_{(\mathcal{R}',\mathcal{R})\in\mathcal{P}([d])} p^{\frac{|\mathcal{R}|-k}{2}} \|A\|_{\mathcal{R}' \mid \mathcal{R}}. \tag{5.4}$$

Since $|\mathcal{P}| = k$ the theorem follows from (5.1), (5.2) and (5.4). □

Proof of Theorem 2.5. Let $S = \left\| \sum a_{\mathbf{i}} g^{(1)}_{i_1} \cdots g^{(d)}_{i_d} \right\|$. Chebyshev's inequality and Theorem 2.1 yield for $p > 0$,

$$\mathbb{P}\left(S \ge C(d) \sum_{(\mathcal{P},\mathcal{P}')\in\mathcal{P}([d])} p^{|\mathcal{P}|/2} \|A\|_{\mathcal{P}' \mid \mathcal{P}} \right) \le e^{1-p}. \tag{5.5}$$

Now we substitute

$$t = C(d) \sum_{\mathcal{P}'\in\mathcal{P}([d])} \|A\|_{\mathcal{P}' \mid \emptyset} + C(d) \sum_{\substack{(\mathcal{P},\mathcal{P}')\in\mathcal{P}([d]) \\ |\mathcal{P}|\ge 1}} p^{|\mathcal{P}|/2} \|A\|_{\mathcal{P}' \mid \mathcal{P}} := t_1 + t_2$$

and observe that if $t_1 < t_2$ then

$$p \ge \frac{1}{C(d)} \min_{\substack{(\mathcal{P},\mathcal{P}')\in\mathcal{P}([d]) \\ |\mathcal{P}|>0}} \left(\frac{t}{\|A\|_{\mathcal{P}' \mid \mathcal{P}}} \right)^{2/|\mathcal{P}|}.$$

The first inequality of the theorem follows then by adjusting the constants.

On the other hand by the Paley-Zygmund inequality we get for $p \ge 2$,

$$\begin{aligned}
\mathbb{P}\left(S \ge C^{-1}(d) \sum_{J\in[d]} \sum_{\mathcal{P}\in\mathcal{P}(J)} p^{|\mathcal{P}|/2} \|\!|A|\!\|_{\mathcal{P}} \right) &\ge \mathbb{P}\left(S^p \ge \frac{1}{2^p} \mathbb{E}S^p \right) \\
&\ge \left(1 - \frac{1}{2^p}\right)^2 \frac{(\mathbb{E}S^p)^2}{\mathbb{E}S^{2p}} \ge e^{-C(d)p},
\end{aligned}$$

where in the last inequality we used Theorem A.1. The inequality follows by a similar substitution as for the upper bound. □

5.2 Proof of Proposition 2.11 and Theorem 2.10

Let us first note that Proposition 2.11 reduces (2.8) of Theorem 2.10 to the lower estimate given in Theorem 2.1, while (2.10) is reduced to Corollary 2.9. The tail bounds (2.9) and (2.11) can be then obtained by Chebyshev's and Paley-Zygmund inequalities as in the proof of Theorem 2.5. The rest of this section will be therefore devoted to the proof of Proposition 2.11.

The overall strategy of the proof is similar to the one used in [4] to obtain the real valued case of Theorem 2.10. It relies on a reduction of inequalities for general polynomials of degree D to estimates for decoupled chaoses of degree $d = 1, \dots, D$. To this end we will approximate general polynomials by tetrahedral ones and split the latter into homogeneous parts of different degrees, which can be decoupled. The splitting may at first appear crude but it turns out that up to constants depending on D one can in fact invert the triangle inequality, which is formalized in the following result due to Kwapień (see [13, Lemma 2]). Recall that a multivariate polynomial is called tetrahedral, if it is affine in each variable.

Theorem 5.1. *If $X = (X_1, \dots, X_n)$ where X_i are independent symmetric random variables, Q is a multivariate tetrahedral polynomial of degree D with coefficients in a Banach space E and Q_d is its homogeneous part of degree d, then for any symmetric convex function $\Phi\colon E \to \mathbb{R}_+$ and any $d \in \{0, 1, \dots, D\}$,*

$$\mathbb{E}\Phi(Q_d(X)) \le \mathbb{E}\Phi(C_D Q(X)).$$

It will be convenient to have the polynomial f represented as a combination of multivariate Hermite polynomials:

$$f(x_1, \dots, x_n) = \sum_{d=0}^{D} \sum_{\mathbf{d} \in \Delta_d^n} a_{\mathbf{d}} h_{d_1}(x_1) \cdots h_{d_n}(x_n), \tag{5.6}$$

where

$$\Delta_d^n = \{\mathbf{d} = (d_1, \dots, d_n) \colon \forall_{k \in [n]}\ d_k \ge 0 \text{ and } d_1 + \cdots + d_n = d\}$$

and $h_m(x) = (-1)^m e^{x^2/2} \frac{d^m}{dx^m} e^{-x^2/2}$ is the m-th Hermite polynomial. Recall that Hermite polynomials are orthogonal with respect to the standard Gaussian measure, in particular if g is a standard Gaussian variable, then for $m \ge 1$, $\mathbb{E}h_m(g) = 0$ (we will use this property several times without explicitly referring to it).

In what follows, we will use the following notation. For a set I, by $I^{\underline{k}}$ we will denote the set of all one-to-one sequences of length k with values in I. For an F-valued d-indexed matrix $A = (a_{i_1,\dots,i_d})_{i_1,\dots,i_d \le n}$ and $x \in \mathbb{R}^{n^d} \simeq (\mathbb{R}^n)^{\otimes d}$ we will denote

$$\langle A, x \rangle = \sum_{i_1,\dots,i_d} a_{i_1,\dots,i_d} x_{i_1,\dots,i_d}.$$

Let $(W_t)_{t \in [0,1]}$ be a standard Brownian motion. Consider standard Gaussian random variables $g = W_1$ and, for any positive integer N,

$$g_{j,N} = \sqrt{N}(W_{\frac{j}{N}} - W_{\frac{j-1}{N}}), \quad j = 1, \dots, N.$$

For any $d \ge 0$, we have the following representation of $h_d(g) = h_d(W_1)$ as a multiple stochastic integral (see [12, Example 7.12 and Theorem 3.21]),

$$h_d(g) = d! \int_0^1 \int_0^{t_d} \cdots \int_0^{t_2} dW_{t_1} \cdots dW_{t_{d-1}} dW_{t_d}.$$

Approximating the multiple stochastic integral leads to

$$\begin{aligned} h_d(g) &= d! \lim_{N\to\infty} N^{-d/2} \sum_{1\le j_1<\cdots<j_d\le N} g_{j_1,N}\cdots g_{j_d,N} \\ &= \lim_{N\to\infty} N^{-d/2} \sum_{j\in[N]^{\underline{d}}} g_{j_1,N}\cdots g_{j_d,N}, \end{aligned} \tag{5.7}$$

where the limit is in $L^2(\Omega)$ (see [12, Theorem 7.3. and formula (7.9)]) and actually the convergence holds in any L^p (see [12, Theorem 3.50]).

Now, consider n independent copies $(W_t^{(i)})_{t\in[0,1]}$ of the Brownian motion ($1\le i\le n$) together with the corresponding Gaussian random variables: $g^{(i)} = W_1^{(i)}$ and, for $N\ge 1$,

$$g_{j,N}^{(i)} = \sqrt{N}(W_{\frac{j}{N}}^{(i)} - W_{\frac{j-1}{N}}^{(i)}), \quad j = 1,\ldots,N.$$

Let also

$$G^{(n,N)} = (g_{1,N}^{(1)},\ldots,g_{N,N}^{(1)},\ g_{1,N}^{(2)},\ldots,g_{N,N}^{(2)},\ \ldots,\ g_{1,N}^{(n)},\ldots,g_{N,N}^{(n)}) = (g_{j,N}^{(i)})_{(i,j)\in[n]\times[N]}$$

be a Gaussian vector with nN coordinates. We identify here the set $[nN]$ with $[n]\times[N]$ via the bijection $(i,j)\leftrightarrow(i-1)N+j$. We will also identify the sets $([n]\times[N])^d$ and $[n]^d\times[N]^d$ in a natural way. For $d\ge 0$ and $\mathbf{d}\in\Delta_d^n$, let

$$I_{\mathbf{d}} = \big\{ i\in[n]^d \colon \forall_{l\in[n]}\ \#i^{-1}(\{l\}) = d_l \big\},$$

and define a d-indexed matrix $B_{\mathbf{d}}^{(N)}$ of n^d blocks each of size N^d as follows: for $i\in[n]^d$ and $j\in[N]^d$,

$$\big(B_{\mathbf{d}}^{(N)}\big)_{(i,j)} = \begin{cases} \frac{d_1!\cdots d_n!}{d!} N^{-d/2} & \text{if } i\in I_{\mathbf{d}} \text{ and } (i,j) := \big((i_1,j_1),\ldots,(i_d,j_d)\big)\in([n]\times[N])^{\underline{d}}, \\ 0 & \text{otherwise.} \end{cases}$$

Proof of Proposition 2.11. Assume that f is of the form (5.6), By [4, Lemma 4.3], for any $p>0$,

$$\big\langle B_{\mathbf{d}}^{(N)}, (G^{(n,N)})^{\otimes d}\big\rangle \xrightarrow{N\to\infty} h_{d_1}(g^{(1)})\cdots h_{d_n}(g^{(n)}) \quad \text{in } L^p(\Omega),$$

which together with the triangle inequality implies that

$$\lim_{N\to\infty} \Big\| \sum_{d=1}^{D} \Big\langle \sum_{\mathbf{d}\in\Delta_d^n} a_{\mathbf{d}} B_{\mathbf{d}}^{(N)}, \big(G^{(n,N)}\big)^{\otimes d} \Big\rangle \Big\|_p = \big\| f(G) - \mathbb{E}f(G) \big\|_p$$

for any $p>0$, where $G=(g^{(1)},\ldots,g^{(n)})$ and we interpret multiplication of an element of F and a real valued d indexed matrix in a natural way. Thus, by Theorem 5.1 and the triangle inequality we obtain

$$\begin{aligned} C_D^{-1} \lim\nolimits_{N\to\infty} \sum_{d=1}^{D} \Big\| \Big\langle \sum_{\mathbf{d}\in\Delta_d^n} a_{\mathbf{d}} B_{\mathbf{d}}^{(N)}, \big(G^{(n,N)}\big)^{\otimes d} \Big\rangle \Big\|_p & \\ \le \|f(G) - \mathbb{E}f(G)\|_p & \\ \le \lim\nolimits_{N\to\infty} \sum_{d=1}^{D} \Big\| \Big\langle \sum_{\mathbf{d}\in\Delta_d^n} a_{\mathbf{d}} B_{\mathbf{d}}^{(N)}, \big(G^{(n,N)}\big)^{\otimes d} \Big\rangle \Big\|_p & \end{aligned}$$

(recall that the matrices $B_{\mathbf{d}}^{(N)}$ have zeros on generalized diagonals and so do their linear combinations).

Denote by $G^{(n,N,1)},\dots,G^{(n,N,d)}$ independent copies of $G^{(n,N)}$.

By decoupling inequalities of Theorem A.9 we have

$$\Big\|\Big\langle \sum_{\mathbf{d}\in\Delta_d^n} a_{\mathbf{d}} B_{\mathbf{d}}^{(N)}, \big(G^{(n,N)}\big)^{\otimes d}\Big\rangle\Big\|_p \sim^d \Big\|\Big\langle \sum_{\mathbf{d}\in\Delta_d^n} a_{\mathbf{d}} B_{\mathbf{d}}^{(N)}, G^{(n,N,1)}\otimes\cdots\otimes G^{(n,N,d)}\Big\rangle\Big\|_p. \quad (5.8)$$

To finish the proof it is therefore enough to show that for any $d\le D$,

$$\lim_{N\to\infty}\Big\|\Big\langle \sum_{\mathbf{d}\in\Delta_d^n} a_{\mathbf{d}} B_{\mathbf{d}}^{(N)}, G^{(n,N,1)}\otimes\cdots\otimes G^{(n,N,d)}\Big\rangle\Big\|_p = \frac{1}{d!}\|\langle A_d, G_1\otimes\cdots\otimes G_d\rangle\|_p, \quad (5.9)$$

where $G_1,\dots,G_D$ are independent copies of G (recall that $A_d=\mathbb{E}\nabla^d f(G)$).

Fix $d\ge 1$. For any $\mathbf{d}\in\Delta_d^n$ define a symmetric d-indexed matrix $(b_{\mathbf{d}})_{i\in[n]^d}$ as

$$(b_{\mathbf{d}})_i = \begin{cases} \frac{d_1!\cdots d_n!}{d!} & \text{if } i\in I_{\mathbf{d}},\\ 0 & \text{otherwise,}\end{cases}$$

and a symmetric d-indexed matrix $(\tilde B_{\mathbf{d}}^{(N)})_{(i,j)\in([n]\times[N])^d}$ as

$$(\tilde B_{\mathbf{d}}^{(N)})_{(i,j)} = N^{-d/2}(b_{\mathbf{d}})_i \quad \text{for all } i\in[n]^d \text{ and } j\in[N]^d.$$

Using the convolution properties of Gaussian distributions one easily obtains

$$\Big\|\Big\langle \sum_{\mathbf{d}\in\Delta_d^n} a_{\mathbf{d}} \tilde B_{\mathbf{d}}^{(N)}, G^{(n,N,1)}\otimes\cdots\otimes G^{(n,N,d)}\rangle\Big\|_p = \Big\|\Big\langle \sum_{\mathbf{d}\in\Delta_d^n} a_{\mathbf{d}} (b_{\mathbf{d}})_{i\in[n]^d}, G_1\otimes\cdots\otimes G_d\Big\rangle\Big\|_p \quad (5.10)$$

On the other hand, for any $\mathbf{d}\in\Delta_d^n$, the matrices $\tilde B_{\mathbf{d}}^{(N)}$ and $B_{\mathbf{d}}^{(N)}$ differ at no more than $|I_{\mathbf{d}}|\cdot|([N]^d\setminus[N]^{\underline{d}})|$ entries. Thus

$$\begin{aligned}
&\Big\|a_{\mathbf{d}}\Big\langle \tilde B_{\mathbf{d}}^{(N)} - B_{\mathbf{d}}^{(N)}, G^{(n,N,1)}\otimes\cdots\otimes G^{(n,N,d)}\Big\rangle\Big\|_p\\
&\le C(d)p^{\frac d2}\|a_{\mathbf{d}}\|\cdot\Big\|\Big\langle \tilde B_{\mathbf{d}}^{(N)} - B_{\mathbf{d}}^{(N)}, G^{(n,N,1)}\otimes\cdots\otimes G^{(n,N,d)}\Big\rangle\Big\|_2\\
&\le C(d)p^{\frac d2}\|a_{\mathbf{d}}\|\cdot\sqrt{|I_{\mathbf{d}}|\Big(\frac{d_1!\cdots d_n!}{d!}\Big)^2 N^{-d}\Big(N^d-\frac{N!}{(N-d)!}\Big)} \longrightarrow 0
\end{aligned}$$

as $N\to\infty$, where in the first inequality we used Theorem A.1.

Together with the triangle inequality and (5.10) this gives

$$\begin{aligned}
&\lim_{N\to\infty}\Big\|\Big\langle \sum_{\mathbf{d}\in\Delta_d^n} a_{\mathbf{d}} B_{\mathbf{d}}^{(N)}, G^{(n,N,1)}\otimes\cdots\otimes G^{(n,N,d)}\Big\rangle\Big\|_p\\
&\qquad = \Big\|\Big\langle \sum_{\mathbf{d}\in\Delta_d^n} a_{\mathbf{d}} (b_{\mathbf{d}})_{i\in[n]^d}, G_1\otimes\cdots\otimes G_d\Big\rangle\Big\|_p. \quad (5.11)
\end{aligned}$$

Finally, we have

$$A_d = \mathbb{E}\nabla^d f(G) = d!\sum_{\mathbf{d}\in\Delta_d^n} a_{\mathbf{d}}(b_{\mathbf{d}})_{i\in[n]^d}. \quad (5.12)$$

Indeed, using the identity on Hermite polynomials, $\frac{d}{dx}h_k(x) = kh_{k-1}(x)$ ($k\ge 1$), we obtain $\mathbb{E}\frac{d^l}{dx^l}h_k(g) = k!\mathbf{1}_{k=l}$ for $k,l\ge 0$, and thus, for any $d,l\le D$ and $\mathbf{d}\in\Delta_l^n$,

$$\big(\mathbb{E}\nabla^d h_{d_1}(g^{(1)})\cdots h_{d_n}(g^{(n)})\big)_i = d!(b_{\mathbf{d}})_i\mathbf{1}_{d=l} \quad \text{for each } i\in[n]^d.$$

Now (5.12) follows by linearity. Combining it with (5.11) yields (5.9) and ends the proof. $\square$

5.3 Proof of the bound for chaoses in exponential variables

Proof of Proposition 2.18. Lemma A.8 implies

$$\left\| \sum_{\mathbf{i}} a_{\mathbf{i}} \prod_{k=1}^{d} E_{i_k}^{(k)} \right\|_p \sim^d \left\| \sum_{i_1,\ldots,i_{2d}} \hat{a}_{i_1,\ldots,i_{2d}} \prod_{k=1}^{2d} g_{i_k}^{(k)} \right\|_p, \tag{5.13}$$

where

$$\hat{a}_{i_1,\ldots,i_{2d}} := a_{i_1,\ldots,i_d} \mathbf{1}_{\{i_1=i_{d+1},\ldots,i_d=i_{2d}\}}.$$

Let $\hat{A} = (\hat{a}_{i_1,\ldots,i_{2d}})_{i_1,\ldots,i_{2d}}$.

Theorem 2.15 and (5.13) yield

$$\begin{aligned} \frac{1}{C(d)} q^{1/2-d} \sum_{J\subset[2d]} \sum_{\mathcal{P}\in\mathcal{P}([J])} p^{\frac{|\mathcal{P}|}{2}} \|\hat{A}\|_{L_q,\mathcal{P}} &\le \left\| \sum_{\mathbf{i}} a_{\mathbf{i}} \prod_{k=1}^{d} E_{i_k}^{(k)} \right\|_p \\ &\le C(d) q^{2d-\frac{1}{2}} \sum_{J\subset[2d]} \sum_{\mathcal{P}\in\mathcal{P}([J])} p^{\frac{|\mathcal{P}|}{2}} \|\hat{A}\|_{L_q,\mathcal{P}}. \end{aligned} \tag{5.14}$$

We will now express $\sum_{J\subset[2d]} \sum_{\mathcal{P}\in\mathcal{P}([J])} p^{\frac{|\mathcal{P}|}{2}} \|\hat{A}\|_{L_q,\mathcal{P}}$ in terms of the matrix A. To this end we need to introduce new notation. Consider a finite sequence $\mathcal{M} = (J, I_1, \ldots, I_k)$ of subsets of $[d]$, such that $J \cup I_1 \cup \ldots \cup I_k = [d]$, $I_1, \ldots, I_k \neq \emptyset$ and each number $m \in [d]$ belongs to at most two of the sets $J, I_1, \ldots, I_k$. Denote the family of all such sequences by $\mathcal{M}([d])$. For $\mathcal{M} = (J, I_1, \ldots, I_k)$ set $|\mathcal{M}| = k+1$ and

$$\langle A \rangle_{L_q,\mathcal{M}} := \sup \left\{ \left\| \sqrt{\sum_{i_J} \left(\sum_{i_{[d]\setminus J}} a_{\mathbf{i}} \prod_{r=1}^{k} x_{i_{I_r}}^{(r)} \right)^2} \right\|_{L_q} \,\middle|\, \forall_{r\le k} \sum_{i_{I_r}} \left(x_{i_{I_r}}^{(r)}\right)^2 \le 1 \right\},$$

where we do not exclude that $J = \emptyset$. By a straightforward verification

$$\sum_{J\subset[2d]} \sum_{\mathcal{P}\in\mathcal{P}([J])} p^{\frac{|\mathcal{P}|}{2}} \|\hat{A}\|_{L_q,\mathcal{P}} \sim^d \sum_{\mathcal{M}\in\mathcal{M}([d])} p^{\frac{|\mathcal{M}|-1}{2}} \langle A \rangle_{L_q,\mathcal{M}}. \tag{5.15}$$

To finish the proof it is enough to show that

$$\sum_{\mathcal{M}\in\mathcal{M}([d])} p^{\frac{|\mathcal{M}|-1}{2}} \langle A \rangle_{L_q,\mathcal{M}} \sim^d \sum_{\mathcal{M}\in\mathcal{C}} p^{\frac{|\mathcal{M}|-1}{2}} \langle A \rangle_{L_q,\mathcal{M}}, \tag{5.16}$$

where

$$\begin{aligned} \mathcal{C} = \Bigg\{ \mathcal{M} = (J, I_1, \ldots, I_k) \in \mathcal{M}([d]) \,\Bigg|\, & J \cap \left(\bigcup_{l=1}^{k} I_l \right) = \emptyset, \\ & \forall_{l,m\le k}\ I_m \cap I_l \neq \emptyset \Rightarrow (|I_l| = |I_m| = 1,\ I_l = I_m) \Bigg\}. \end{aligned}$$

Indeed assume that (5.16) holds and choose $\mathcal{M} = (J, I_1, \ldots, I_k) \in \mathcal{C}$. Consider

$I = \{i \mid \exists_{l<m\le k}\ \{i\} = I_l = I_m\}$. Then $J \cap I = \emptyset$ and we have

$$\langle A\rangle^q_{L_q,\mathcal{M}} = \sup\Bigg\{\int_V \Bigg(\sum_{i_J}\Bigg(\sum_{i_{J^c}} a_{i_1,\ldots,i_d}(v)\prod_{l\in I} y^{(l)}_{i_l}x^{(l)}_{i_l}\prod_{\substack{l\le k\\ I_l\cap I=\emptyset}} x^{(l)}_{i_{I_l}}\Bigg)^2\Bigg)^{q/2} d\mu(v)) \ \Bigg|$$

$$\forall_{1\le l\le k}\sum_{i_{I_l}}(x^{(l)}_{i_{I_l}})^2\le 1,\ \forall_{l\in I}\ \sum_{i_l}(y^{(l)}_{i_l})^2\le 1\Bigg\}$$

$$= \max_{i_I}\sup\Bigg\{\int_V\Bigg(\sum_{i_J}\Bigg(\sum_{i_{J^c\setminus I}} a_{i_1,\ldots,i_d}(v)\prod_{\substack{l\le k\\ I_l\cap I=\emptyset}} x^{(l)}_{i_{I_l}}\Bigg)^2\Bigg)^{q/2} d\mu(v)\ \Bigg|\ \forall_{\substack{1\le l\le k\\ I_l\cap I=\emptyset}}\ \sum_{i_{I_l}}(x^{(l)}_{i_{I_l}})^2\le 1\Bigg\}$$

$$= \max_{i_I}\left(\|\!|(a_{i_1,\ldots,i_d})_{i_{I^c}}\|\!|_{L_q,\{I_l\,:\,I_l\cap I=\emptyset\}}\right)^q =: \max_{i_I}\left(\|\!|(a_{i_1,\ldots,i_d})_{i_{I^c}}\|\!|_{L_q,\mathcal{P}}\right)^q, \tag{5.17}$$

where in the second equality we used the fact that

$$(y^{(l)}_{i_l}x^{(l)}_{i_l})_{i_l}\in B^n_1 = \{x\in\mathbb{R}^N:\ \sum_{i=1}^N |x_i|\le 1\}$$

together with convexity and homogeneity of the norm

$$\|(f_{i_J})_{i_J}\|_{L_q(\ell_2)} = \left(\int_X\left(\sum_{i_J} f^2_{i_J}\right)^{q/2}\right)^{1/q}.$$

By combining the above with (5.14)–(5.16) and comparing the exponents of p we conclude the assertion of the proposition.

The proof is completed by showing that

$$\sum_{\mathcal{M}\in\mathcal{M}([d])} p^{\frac{|\mathcal{M}|-1}{2}}\langle A\rangle_{L_q,\mathcal{M}} \le C(d)\sum_{\mathcal{M}\in\mathcal{C}} p^{\frac{|\mathcal{M}|-1}{2}}\langle A\rangle_{L_q,\mathcal{M}}$$

(the other estimate in (5.16) is trivial), which will be done in two steps. Let us fix $\mathcal{M} = (J, I_1,\ldots,I_k)\in\mathcal{M}([d])$.

Step 1. Assume first that $J\cap(\bigcup_{i=1}^k I_i)\ne\emptyset$. Without loss of generality we can assume that $1\in J\cap I_1$. Denote $\hat{I}_1 = I_1\setminus\{1\}$ and for any matrix $(x^{(1)}_{i_I})_{i_I}$ such that $\sum_{i_{I_1}}(x^{(1)}_{i_{I_1}})^2\le 1$, set $(b_{i_1})_{i_1} := (\sqrt{\sum_{i_{I_1\setminus\{1\}}}(x^{(1)}_{i_{I_1}})^2})_{i_1}$. Clearly,

$$\left(b^2_{i_1}\right)_{i_1}\in B^n_1 \quad\text{and}\quad \sum_{i_{I_1\setminus\{1\}}}\left(\frac{x^{(1)}_{i_{I_1}}}{b_{i_1}}\right)^2\le 1.$$

Observe that for any $f_1,\ldots,f_n\in L^q(X,d\mu)$ the function

$$[0,+\infty)^n\ni t\to\int_V\left(\sum_i f^2_i(v)t_i\right)^{q/2} d\mu(v)$$

is convex (recall that $q \geq 2$). Since $B_1^n \cap [0,\infty)^n = \mathrm{conv}(0, e_1, e_2, \ldots, e_n)$, we have

$$\langle A\rangle_{L_q,\mathcal{M}}^q = \sup\left\{ \int_V \left(\sum_{i_J} b_{i_1}^2 \left(\sum_{i_{J^c}} a_{i_1,\ldots,i_d}(v) \frac{x^{(1)}_{i_{I_1}}}{b_{i_1}} \prod_{l=2}^k x^{(l)}_{i_{I_l}} \right)^2 \right)^{q/2} d\mu(v) \;\Big|\; \forall_{1\leq l\leq k} \sum_{i_{I_l}} (x^{(l)}_{i_{I_l}})^2 \leq 1 \right\}$$

$$\leq \max_{i_1} \sup\left\{ \int_V \left(\sum_{i_{J\setminus\{1\}}} \left(\sum_{i_{J^c}} a_{i_1,\ldots,i_d}(v) \frac{x^{(1)}_{i_{I_1}}}{b_{i_1}} \prod_{l=2}^k x^{(l)}_{i_{I_l}} \right)^2 \right)^{q/2} d\mu(v) \;\Big|\; \forall_{1\leq l\leq k} \sum_{i_{I_l}} (x^{(l)}_{i_{I_l}})^2 \leq 1 \right\}$$

$$\leq \max_{i_1} \sup\left\{ \int_V \left(\sum_{i_{J\setminus\{1\}}} \left(\sum_{i_{J^c}} a_{i_1,\ldots,i_d}(v) y_{i_{\hat{I}_1}} \prod_{l=2}^k x^{(l)}_{i_{I_l}} \right)^2 \right)^{q/2} d\mu(v) \;\Big|\; \sum_{i_{\hat{I}_1}} (y_{i_{\hat{I}_1}})^2 \leq 1,\ \forall_{1\leq l\leq k} \sum_{i_{I_l}} (x^{(l)}_{i_{I_l}})^2 \leq 1 \right\}.$$

Let

$$\mathcal{M}' = \begin{cases} (J\setminus\{1\}, \{1\}, \{1\}, \hat{I}_1, I_2, \ldots, I_k) & \text{if } \hat{I}_1 \neq \emptyset \\ (J\setminus\{1\}, \{1\}, \{1\}, I_2, \ldots, I_k) & \text{if } \hat{I}_1 = \emptyset \end{cases}$$

By the same argument as was used for the second equality in (5.17) we obtain that the right-hand side above equals $\langle A\rangle_{L_q,\mathcal{M}'}$, which gives

$$\langle A\rangle_{L_q,\mathcal{M}} \leq \langle A\rangle_{L_q,\mathcal{M}'}.$$

Observe that

$$p^{(|\mathcal{M}|-1)/2} \langle A\rangle_{L_q,\mathcal{M}} \leq p^{(|\mathcal{M}|-1)/2} \langle A\rangle_{L_q,\mathcal{M}'} \leq p^{(|\mathcal{M}'|-1)/2} \langle A\rangle_{L_q,\mathcal{M}'}.$$

By iterating this argument we obtain that $p^{(|\mathcal{M}|-1)/2} \langle A\rangle_{L_q,\mathcal{M}} \leq p^{(|\mathcal{M}''|-1)/2} \langle A\rangle_{L_q,\mathcal{M}''}$ for some $\mathcal{M}'' = (J'', I_1'', \ldots, I_m'')$ such that $J'' \cap (\bigcup_{l=1}^m I_l'') = \emptyset$.

Step 2. Assume that for some $l, m \leq k$ $I_l \cap I_m \neq \emptyset$ and $|I_l| \geq 2$ or $|I_m| \geq 2$. Without loss of generality assume that $1 \in I_1 \cap I_2$ and $|I_1| \geq 2$. Clearly,

$$\langle A\rangle_{L_q,\mathcal{M}}^q = \sup\left\{ \int_V \left(\sum_{i_J} \left(\sum_{i_{J^c}} a_{i_1,\ldots,i_d}(v) b_{i_1} c_{i_1} \frac{x^{(1)}_{i_{I_1}}}{b_{i_1}} \frac{x^{(2)}_{i_{I_2}}}{c_{i_1}} \prod_{l=3}^k x^{(l)}_{i_{I_l}} \right)^2 \right)^{q/2} d\mu(v) \;\Big|\; \forall_{1\leq l\leq k} \sum_{i_{I_l}} (x^{(l)}_{i_{I_l}})^2 \leq 1 \right\},$$

where $(b_{i_1})_{i_1} := (\sqrt{\sum_{i_{I_1\setminus\{1\}}} (x^{(1)}_{i_{I_1}})^2})_{i_1}$, $(c_{i_1})_{i_1} := (\sqrt{\sum_{i_{I_2\setminus\{1\}}} (x^{(2)}_{i_{I_2}})^2})_{i_1} \in B_2^n$. Because $(b_{i_1} c_{i_1})_{i_1} \in B_1^n$,

$$\forall_{i_1} \sum_{i_{I_1\setminus\{1\}}} \left(\frac{x^{(1)}_{i_{I_1}}}{b_{i_1}} \right)^2 \leq 1, \quad \sum_{i_{I_2\setminus\{1\}}} \left(\frac{x^{(2)}_{i_{I_2}}}{c_{i_1}} \right)^2 \leq 1,$$

and for any $(f_{ij})_{ij}$ in $L^q(X, d\mu)$, the function

$$\mathbb{R}^n \ni t \to \int_V \left(\sum_i \left(\sum_j t_j f_{ij}(v) \right)^2 \right)^{q/2} d\mu(v)$$

is convex, we obtain similarly as in Step 1,

$$p^{(|\mathcal{M}|-1)/2} \left\langle A \right\rangle_{L_q, \mathcal{M}} \leq p^{(|\mathcal{M}|-1)/2} \ \left\langle A \right\rangle_{L_q, \mathcal{M}'} \leq p^{(|\mathcal{M}'|-1)/2} \ \left\langle A \right\rangle_{L_q, \mathcal{M}'}$$

where

$$\mathcal{M}' := \begin{cases} (J, \{1\}, \{1\}, I_1 \setminus \{1\}, I_2 \setminus \{1\}, I_3, \ldots, I_k) & \text{if } I_2 \setminus \{1\} \neq \emptyset \\ (J, \{1\}, \{1\}, I_1 \setminus \{1\}, I_3, \ldots, I_k) & \text{otherwise} \end{cases}.$$

An iteration of this argument shows that indeed one can assume that $\mathcal{M}$ satisfies the implication $I_m \cap I_l \neq \emptyset \Rightarrow (|I_l| = |I_m| = 1, \ I_l = I_m)$.

Combining Steps 1 and 2 we obtain that for any $\mathcal{M} \in \mathcal{M}([d])$ there exists $\mathcal{M}' \in \mathcal{C}$ such that $p^{(|\mathcal{M}|-1)/2} \left\langle A \right\rangle_{L_q, \mathcal{M}} \leq p^{(|\mathcal{M}'|-1)/2} \ \left\langle A \right\rangle_{L_q, \mathcal{M}'}$ which yields (5.16). □

A Appendix

In this section we gather technical facts that are used in the proof.

Theorem A.1 (Hypercontractivity of Gaussian chaoses)**.** *Let*

$$S = a + \sum_{i_1} a_{i_1} g_{i_1} + \sum_{i_1, i_2} a_{i_1, i_2} g_{i_1} g_{i_2} + \ldots + \sum_{i_1, \ldots, i_d} a_{i_1, \ldots, i_d} g_{i_1} \cdots g_{i_d},$$

be a non-homogeneous Gaussian chaos of order d with values in a Banach space $(F, \|\cdot\|)$. Then for any $1 \leq p < q < \infty$, we have

$$\left(\mathbb{E} \left\| S \right\|^q \right)^{1/q} \leq C(d) \left(\frac{q}{p} \right)^{d/2} \left(\mathbb{E} \left\| S \right\|^p \right)^{1/p}.$$

Proof. It is an immediate consequence of [9, Theorem 3.2.10] and Hölder's inequality. □

Theorem A.2 (Sudakov minoration [27])**.** *For any set $T \subset \mathbb{R}^n$ and $\varepsilon > 0$ we have*

$$\varepsilon \sqrt{\ln N(T, d_2, \varepsilon)} \leq C \mathbb{E} \sup_{t \in T} \sum_i t_i g_i,$$

where d_2 is the Euclidean distance.

Theorem A.3 (Dual Sudakov minoration [17, formula (3.15)])**.** *Let α be a norm on $\mathbb{R}^n$ and $\rho_\alpha(x, y) = \alpha(x - y)$ for $x, y \in \mathbb{R}^n$. Then*

$$\varepsilon \sqrt{\log N(B_2^n, \rho_\alpha, \varepsilon)} \leq C \mathbb{E} \alpha(G_n) \quad \textit{for } \varepsilon > 0.$$

Lemma A.4 ([15, Lemma 3])**.** *Let $(G_t)_{t \in T}$ be a centered Gaussian process and $T = \bigcup_{l=1}^m T_l$, $m \geq 1$. Then*

$$\mathbb{E} \sup_{t \in T} G_t \leq \max_{l \leq m} \mathbb{E} \sup_{t \in T_l} G_t + C \sqrt{\ln(m)} \sup_{t, t' \in T} \sqrt{\mathbb{E}(G_t - G_{t'})^2}.$$

Lemma A.5. *Let G be a Gaussian variable in a Banach space $(F, \| \cdot \|)$. Then for any $p \geq 2$,*

$$\frac{1}{C} \left(\|G\|_1 + \sqrt{p} \sup_{\substack{\varphi \in F^* \\ \|\varphi\|_* \leq 1}} \mathbb{E} |\varphi(G)| \right) \leq \|G\|_p \leq \|G\|_1 + C \sqrt{p} \sup_{\substack{\varphi \in F^* \\ \|\varphi\|_* \leq 1}} \mathbb{E} |\varphi(G)|,$$

where $(F^*, \|\cdot\|_*)$ is the dual space to $(F, \|\cdot\|)$.

In particular for $a_1,\ldots,a_n\in F$ *and* $G=\sum_{i=1}^n a_ig_i$ *we have*

$$\Big\|\sum_{i=1}^n a_ig_i\Big\|_p\sim \mathbb{E}\Big\|\sum_i a_ig_i\Big\|+\sqrt{p}\sup_{x\in B_2^n}\Big\|\sum_{i=1}^n a_ix_i\Big\|.$$

Proof. The first part of the Lemma is [15, Lemma 4]. It implies the second part as follows. Since $\varphi(G)$ is a one-dimensional Gaussian vector, we have $\|\varphi(G)\|_2\sim\mathbb{E}|\varphi(G)|$. In particular for $G=\sum_{i=1}^n a_ig_i$, $a_i\in F$ we obtain

$$\begin{aligned}\sup_{\varphi\in F^*,\|\varphi\|\le 1}\mathbb{E}|\varphi(G)|\sim\sup_{\varphi\in F^*,\|\varphi\|\le 1}\|\varphi(G)\|_2&=\sup_{\varphi\in F^*,\|\varphi\|\le 1}\Big(\sum_i\varphi(a_i)^2\Big)^{1/2}\\&=\sup_{\varphi\in F^*,\|\varphi\|\le 1}\sup_{x\in B_2^n}\sum_i\varphi(a_i)x_i=\sup_{x\in B_2^n}\Big\|\sum_i a_ix_i\Big\|.\end{aligned}$$ □

Theorem A.6 ([15, Theorem 1])**.** *For any real-valued matrix* $(a_{i_1,\ldots,i_d})_{i_1,\ldots,i_d}$ *and* $p\ge 2$, *we have*

$$\Big\|\sum_{\mathbf{i}}a_{\mathbf{i}}\prod_{j=1}^d g_{i_j}^{(j)}\Big\|_p\sim^d\sum_{\substack{\mathcal{P}\in\mathcal{P}([d])\\ \mathcal{P}=(I_1,\ldots,I_k)}}p^{|P|/2}\sup\left\{\sum_{\mathbf{i}}a_{\mathbf{i}}\prod_{j=1}^k x_{i_{I_j}}^{(j)}\ \Big|\ \forall_{1\le j\le k}\ \Big\|(x_{i_{I_j}}^{(k)})_{i_{I_j}}\Big\|_2\le 1\right\}.$$

Corollary A.7. *Assume that for any* $i_1,\ldots,i_d,\ a_{i_1,\ldots,i_d}\in\mathbb{R}$. *Then for all* $p\ge 1$

$$\frac{1}{C(d)}\sqrt{p}\sqrt{\sum_{i_1,\ldots,i_d}a_{i_1,\ldots,i_d}^2}\le\Big\|\sum_{i_1,\ldots,i_d}a_{i_1,\ldots,i_d}g_{i_1}^{(1)}\cdots g_{i_d}^{(d)}\Big\|_p\le C(d)p^{d/2}\sqrt{\sum_{i_1,\ldots,i_d}a_{i_1,\ldots,i_d}^2}.$$

Proof. It is an easy consequence of Theorems A.1 and A.6. □

Lemma A.8 ([3, Lemma 9.5])**.** *Let* $Y_i^{(1)}$ *be independent standard symmetric exponential variables (variables with density* $2^{-1}\exp(-|t|)$*) and* $Y_i^{(2)}=g_i^2$, $Y_i^{(3)}=g_ig_i'$, *where* g_i,g_i' *are i.i.d.* $\mathcal{N}(0,1)$ *variables and* ε_i *– i.i.d. Rademacher variables independent of* $(Y^{(1)}),\ (Y^{(2)}),\ (Y^{(3)})$. *Then for any Banach space* $(F,\|\cdot\|)$, *any* $p\ge 1$ *and any vectors* $v_1,\ldots,v_n\in F$ *the quantities*

$$\left(\mathbb{E}\Big\|\sum_i v_i\varepsilon_iY_i^{(j)}\Big\|^p\right)^{1/p},\quad j=1,2,3,$$

are comparable up to universal multiplicative factors.

We remark that the above lemma is formulated in [3] for $p=1$ in the real valued case, however the proof presented there (based on the contraction principle) works for arbitrary $p\ge 1$ and arbitrary Banach spaces.

We will also need decoupling inequalities for tetrahedral homogeneous polynomials. Such inequalities were introduced for the first time in [18] for real valued multi-linear forms and since then have been strengthened and generalized by many authors (see the monograph [9]). The following theorem is a special case of results from [13] (treating also general tetrahedral polynomials) and [8, 10] (treating general U-statistics).

Theorem A.9. *Let* $X=(X_1,\ldots,X_n)$ *be a sequence of independent random variables and let* $X^{(l)}=(X_1^{(l)},\ldots,X_n^{(l)})$, $l=1,\ldots,d$, *be i.i.d. copies of* X. *Consider a* d*-indexed symmetric matrix* $(a_{i_1,\ldots,i_d})_{i_1,\ldots,i_d=1}^n$, *with coefficients from a Banach space* F. *Assume* *that* $a_{i_1,\ldots,i_d} = 0$ *whenever there exist* $1 \le k < m \le d$ *such that* $i_k = i_m$*. Then for any* $p \ge 1$*,*

$$\Big\| \sum_{i_1,\ldots,i_d=1}^{n} a_{i_1,\ldots,i_d} X_{i_1}\cdots X_{i_d} \Big\|_p \sim_d \Big\| \sum_{i_1,\ldots,i_d=1}^{n} a_{i_1,\ldots,i_d} X^{(1)}_{i_1}\cdots X^{(d)}_{i_d} \Big\|_p.$$

Moreover, for any $t > 0$*,*

$$\begin{aligned} C_d^{-1}\mathbb{P}\Big(\Big\| \sum_{i_1,\ldots,i_d=1}^{n} a_{i_1,\ldots,i_d} X^{(1)}_{i_1}\cdots X^{(d)}_{i_d} \Big\| \ge C_d t\Big) \\ \le \mathbb{P}\Big(\Big\| \sum_{i_1,\ldots,i_d=1}^{n} a_{i_1,\ldots,i_d} X_{i_1}\cdots X_{i_d} \Big\| \ge t\Big) \\ \le C_d \mathbb{P}\Big(\Big\| \sum_{i_1,\ldots,i_d=1}^{n} a_{i_1,\ldots,i_d} X^{(1)}_{i_1}\cdots X^{(d)}_{i_d} \Big\| \ge t/C_d\Big). \end{aligned}$$

B Glossary

- F – the underlying Banach space, p. 2
- $(g_i^{(k)})_{i,k\ge 1}$ – an array of i.i.d standard Gaussian variables, p. 2
- F^* – the dual of F, p. 12
- B_2^n – the standard unit Euclidean ball, p. 2
- B_1^n – the standard unit ball in $\mathbb{R}^n$ in the norm $\|\cdot\|_1$, p. 30
- $[n]$ – the set $\{1,\ldots,n\}$, p. 5
- $\mathbb{R}^{n^d} = (\mathbb{R}^n)^{\otimes d}$ – the space of d-indexed matrices, p. 5
- $B_2^{n^d}$ – the unit Euclidean (Hilbert-Schmidt) ball in $\mathbb{R}^{n^d}$, p. 5
- $\mathcal{P}(I)$ – the set of partitions of I, p. 5
- $\|\cdot\|_{\mathcal{P}'\mid\mathcal{P}}$ – a norm appearing in general moment estimates, p. 5
- $\|\!|\cdot|\!\|_{\mathcal{P}}$ – a norm appearing in general moment estimates, p. 5
- $G^{(l)} = (g^{(l)}_{i_{I_l}})_{i_{I_l}}$ – independent arrays of i.i.d. standard Gaussian variables, p. 5
- $\|\!|A|\!\|_{L_q,\mathcal{P}}$ – a norm appearing in moment estimates for L_q-values chaoses, p. 10
- $s_k(A)$ – the sum of norms corresponding to partitions of cardinality k, p. 12
- ρ_A – the distance induced by the canonical Gaussian process, p. 13
- α_A – a norm related to ρ_A, p. 13
- $N(S,\rho,\varepsilon)$ – the entropy number, p. 13
- $W_d^U(\alpha_A, s)$ – a quantity appearing in estimates for entropy numbers, p. 13
- $W_I^U(\alpha_A, s)$ – a quantity appearing in estimates for entropy numbers, p. 13
- β_A – an auxiliary norm related to the main Gaussian process, p. 13
- $V_I^U(\beta_A)$ – a quantity appearing in estimates for entropy numbers, p. 13
- $\mu^d_{\varepsilon,T}$ – a measure used in the volumetric argument, p. 14
- $\hat{\alpha}_A$ – an auxiliary norm appearing in estimates for entropy numbers, p. 16
- $F_A(U)$ – the expected supremum of the main Gaussian process, p. 17
- $\Delta_A(U)$ – the diameter of the set U in the distance ρ_A, p. 17
- Δ_d^n – the set of multi-indices corresponding to Hermite polynomials of degree d, p. 26
- $\langle A\rangle_{L_q,\mathcal{M}}$ – an auxiliary norm related to moment estimates for chaoses in exponential variables, p. 29

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. MR-3407216

[2] R. Adamczak, R. Latała and R. Meller. *Hanson-Wright inequality in Banach spaces.* Ann. Inst. Henri Poincaré Probab. Stat. **56** (2020), 2356-2376. MR-4164840

[3] 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. MR-3052405

[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. MR-3383337

[5] M. A. Arcones and E. Giné. *On decoupling, series expansions, and tail behavior of chaos processes.* J. Theoret. Probab. **6** (1993), 101–122. MR-1201060

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

[7] A. De, R. Servedio. *A new central limit theorem and decomposition for Gaussian polynomials, with an application to deterministic approximate counting.* Probab. Theory Related Fields **171** (2018), no. 3-4, 981–1044. MR-3827226

[8] V. H. de la Peña. *Decoupling and Khintchine's inequalities for U-statistics.* Ann. Probab. **20** (1992), 1877–1892. MR-1188046

[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. MR-1666908

[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. MR-1334173

[11] 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. MR-3752640

[12] S. Janson. *Gaussian Hilbert spaces.* Cambridge Tracts in Mathematics, **129**, Cambridge University Press, Cambridge, 1997. MR-1474726

[13] S. Kwapień. *Decoupling inequalities for polynomial chaos.* Ann. Probab. **15** (1987), 1062–1071. MR-0893914

[14] K. G. Larsen, J. Nelson. *The Johnson-Lindenstrauss lemma is optimal for linear dimensionality reduction.* 43rd International Colloquium on Automata, Languages, and Programming, Art. No. **82**, LIPIcs. Leibniz Int. Proc. Inform., **55**, Schloss Dagstuhl. Leibniz-Zent. Inform., Wadern, 2016. MR-3577143

[15] R. Latała. *Estimates of moments and tails of Gaussian chaoses.* Ann. Probab. **34** (2006), 2315–2331. MR-2294983

[16] M. Ledoux. *A note on large deviations for Wiener chaos.* Séminaire de Probabilités, XXIV, 1988/89, 1–14, Lecture Notes in Math., **1426**, Springer, Berlin, 1990. MR-1071528

[17] M. Ledoux and M. Talagrand. *Probability in Banach spaces. Isoperimetry and processes.* Springer-Verlag, Berlin, 1991. MR-1102015

[18] T. R. McConnell, M. S. Taqqu. *Decoupling inequalities for multilinear forms in independent symmetric random variables.* Ann. Probab. **14** (1986), 943–954. MR-0841595

[19] E. Mossel, R. O'Donnell, K. Oleszkiewicz. *Noise stability of functions with low influences: invariance and optimality.* Ann. of Math. (2) **171** (2010), no. 1, 295–341. MR-2630040

[20] 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. MR-2421495

[21] E. Nelson. *The free Markoff field.* J. Funct. Anal. **12** (1973), 211–227. MR-0343816

[22] I. Nourdin, G. Peccati. *Normal approximations with Malliavin calculus. From Stein's method to universality.* Cambridge Tracts in Mathematics, **192**, Cambridge University Press, Cambridge, 2012. MR-2962301

[23] D. Nualart. *The Malliavin calculus and related topics.* Probability and its Applications (New York). Springer-Verlag, Berlin, 2006. MR-2200233

[24] R. O'Donnell. *Analysis of Boolean functions.* Cambridge University Press, New York, 2014. MR-3443800

[25] G. Pisier. *Some results on Banach spaces without local unconditional structure.* Compositio Math. **37** (1978), 3–19. MR-0501916

[26] G. Pisier. *The volume of convex bodies and Banach space geometry.* Cambridge Tracts in Mathematics, **94**. Cambridge University Press, Cambridge, 1989. MR-1036275

[27] V. N. Sudakov. *Gaussian measures, Cauchy measures and ε-entropy.* Soviet Math. Dokl. **10** (1969), 310–313. MR-0247034

[28] 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. MR-3184689

[29] M.S. Taqqu. *Weak convergence to fractional Brownian motion and to the Rosenblatt process.* Z. Wahrscheinlichkeitstheorie und Verw. Gebiete **31** (1974/75), 287–302. MR-0400329

[30] M. Wahl. *On the perturbation series for eigenvalues and eigenprojections* (2019), arXiv:1910.08460.