

Tail bounds via generic chaining*

Sjoerd Dirksen[†]

Abstract

We modify Talagrand’s generic chaining method to obtain upper bounds for all p -th moments of the supremum of a stochastic process. These bounds lead to an estimate for the upper tail of the supremum with optimal deviation parameters. We use our procedure to improve and extend some known deviation inequalities for suprema of unbounded empirical processes and chaos processes. As an application we give a simplified proof of the restricted isometry property of the subsampled discrete Fourier transform.

Keywords: Generic chaining; deviation inequalities; suprema of empirical processes; restricted isometry property; chaos processes.

AMS MSC 2010: 60E15 ; 60F10 ; 60G50.

Submitted to EJP on September 3, 2014, final version accepted on April 16, 2015.

1 Introduction

This paper is concerned with generic chaining, a method introduced by Talagrand to estimate the expected value of the supremum of a stochastic process. This method grew out of the classical chaining method and the later majorizing measures method, which were developed by, among others, Kolmogorov, Dudley, Fernique and Talagrand, to understand the continuity properties of stochastic processes. Generic chaining yields estimates for the expected value of the supremum in terms of so-called γ -functionals, which measure the metric complexity of the index set of the process [25, 26]. The resulting bounds are known to be sharp in several interesting situations. For instance, the famous majorizing measures theorem [24] states that this is the case for suprema of Gaussian processes, provided that the index set is equipped with the canonical metric induced by the process.

In practical applications of generic chaining in statistics, compressed sensing, and geometric functional analysis, see e.g. [8, 11], it is often not sufficient to have an upper bound for the expected supremum of a process. One also needs to know how probable it is that the supremum of the process exceeds the upper bound. To that aim, a generic chaining bound is typically supplemented with a tail bound for the deviation of the supremum with respect to its expected value. There is an extensive and rapidly growing

*This research was supported by SFB 1060 of the German Research Foundation (DFG)

[†]University of Bonn, Germany. E-mail: dirksen@mathc.rwth-aachen.de

literature on such deviation inequalities, see for instance the monographs [4, 16] for a detailed introduction and a historical overview.

The purpose of this paper is to provide an alternative to the two-step procedure sketched above. We present a simple and general way to directly obtain upper deviation inequalities for the supremum of a stochastic process $(X_t)_{t \in T}$ using generic chaining. The idea is to alter the generic chaining procedure to produce bounds for not only the first, but for all p -th moments of $\sup_{t \in T} |X_t|$. Together with a standard optimization argument using Markov's inequality this yields an upper tail bound. The deviation parameters in the resulting tail bound are sharp up to numerical constants. In particular, the bound is qualitatively as good as the one obtained by combining the usual generic chaining bound for $\mathbb{E} \sup_{t \in T} |X_t|$ with the best possible upper tail bound for the deviation $\sup_{t \in T} |X_t| - \mathbb{E} \sup_{t \in T} |X_t|$.

To give a concrete illustration of these statements, consider the simplest case that $(X_t)_{t \in T}$ is a centered Gaussian process and let $d(s, t) = (\mathbb{E}|X_t - X_s|^2)^{1/2}$ be the canonical metric on T . Under this assumption, we prove that for some universal constants $C, D > 0$ and any $1 \leq p < \infty$,

$$\left(\mathbb{E} \sup_{t \in T} |X_t|^p \right)^{1/p} \leq C \gamma_{2,p}(T, d) + D \sigma \sqrt{p},$$

where $\gamma_{2,p}$ is a truncated version of the γ_2 -functional familiar from generic chaining and $\sigma^2 = \sup_{t \in T} \mathbb{E}(X_t^2)$ is the weak variance of the process. Estimates for the constants C and D are provided in Remark 3.3, although these can certainly be improved. As a direct consequence of the stated L^p -bounds we find

$$\mathbb{P} \left(\sup_{t \in T} |X_t| \geq \sqrt{e} (C \gamma_2(T, d) + u D \sigma) \right) \leq e^{-u^2/2} \quad (u \geq 1). \quad (1.1)$$

The bound (1.1) matches, up to a possibly worse constant D , the upper tail bound obtained by combining the optimal generic chaining estimate for $\mathbb{E} \sup_{t \in T} |X_t|$ with the sharp concentration inequality for suprema of Gaussian processes due to Borell, Ibragimov, Sudakov, and Tsirelson [4, Theorem 5.8]. Moreover, as is shown in Theorem 3.2, the tail bound (1.1) more generally holds for any process which has subgaussian increments with respect to a given metric d .

A first advantage of the method proposed here is its simplicity: an upper tail bound is obtained essentially for free once one uses generic chaining to estimate the expected value. In contrast, the usual proofs of deviation inequalities for suprema of stochastic processes rely on sophisticated tools such as the entropy method, see for example [4, Chapters 6 and 12]. A second advantage is that the method only requires knowledge of the tail behavior of the individual increments of the process. In particular, one can obtain deviation inequalities for processes with dependent increments, see for example the uniform Azuma-Hoeffding inequality in Corollary 3.4. In the context of empirical processes, the method can readily cover situations in which the summands of the empirical process are unbounded and/or dependent. Under these conditions deviation inequalities are still scarcely available, see [1, 9] for notable exceptions.

To demonstrate the wide applicability of our method, we establish an upper tail bound for suprema of stochastic processes in several interesting situations. In Section 3 we consider two 'standard' generic chaining situations. In Theorem 3.2 we investigate processes which have exponentially decaying increments with respect to a single metric. In Theorem 3.5, we consider processes with a mixed subgaussian-subexponential tail, in particular suprema of empirical processes. This result positively answers an open question raised in Talagrand's new book [27]. In the second part of the paper, i.e., Sections 5.1 and 6, we consider two specialized, more involved chaining arguments. In Theorem 5.5 we find L^p -bounds for the supremum of an empirical process which takes

the form of an average of squares. It can be viewed as an L^p -version of a result due to Mendelson, Pajor and Tomczak-Jaegermann [19], see Theorem 5.6 and Corollary 5.7 for a detailed comparison. In the final section we deal with suprema of second order chaos processes.

In Section 4 we use Theorem 3.2 to simplify the proof of the restricted isometry property of the subsampled discrete Fourier transform. This cornerstone result in compressed sensing was originally discovered by Candès and Tao [6] and was later refined by Rudelson and Vershynin [22] and Rauhut [21]. The argument presented in Section 4 more generally applies to matrices obtained by sampling from bounded orthonormal systems, see Theorem 4.2. To keep this paper at a reasonable length we discuss an application of Theorem 5.5 to dimensionality reduction in a separate note [7]. Further applications of our work can be found in [5, 12].

We conclude this introduction with a brief discussion of related work. In [29], Viens and Vizcarra modified a classical, that is, non-generic chaining argument to obtain upper deviation inequalities for so-called sub- n -th chaos processes. Theorem 3.2 below yields an improvement of this result as a special case. We also improve a deviation inequality for suprema of unbounded empirical processes obtained recently by Van de Geer and Lederer [9], see the discussion after Corollary 5.2. In [15], Latała used a procedure related to ours to prove a comparison result for the strong and weak moments of certain log-concave random vectors. Finally, Krahmer, Mendelson, and Rauhut [13] used a chaining argument to prove an upper tail bound for the supremum of a second order chaos process. In Theorem 6.5 we give an improvement of their bound with a simplified proof.

2 Preliminaries

Throughout, we will use $(\Omega, \mathcal{F}, \mathbb{P})$ to denote a probability space and write \mathbb{E} for the expected value. To describe the tail behavior of random variables we consider for every $0 < \alpha < \infty$ the function

$$\psi_\alpha(x) = \exp(x^\alpha) - 1 \quad (x \geq 0).$$

For a complex-valued random variable X we define

$$\|X\|_{\psi_\alpha} = \inf\{C > 0 : \mathbb{E}\psi_\alpha(|X|/C) \leq 1\}.$$

If $\|X\|_{\psi_\alpha} < \infty$ then we call X a ψ_α -random variable. It is common to say that X is *subgaussian* if $\|X\|_{\psi_2} < \infty$ and *subexponential* if $\|X\|_{\psi_1} < \infty$. If $\alpha \geq 1$ then ψ_α is an Orlicz function and the space

$$L_{\psi_\alpha}(\Omega, \mathcal{F}, \mathbb{P}) = \{X : \Omega \rightarrow \mathbb{C} \text{ measurable} : \|X\|_{\psi_\alpha} < \infty\}$$

is an Orlicz space. For $0 < \alpha < 1$ the space L_{ψ_α} is only a quasi-Banach space. We will make use of the following Hölder type inequality, which can be derived from Young's inequality: if X, Y are ψ_2 -random variables, then XY is ψ_1 and

$$\|XY\|_{\psi_1} \leq \|X\|_{\psi_2} \|Y\|_{\psi_2}. \quad (2.1)$$

For more information on Orlicz spaces we refer to [14].

Let us recall some familiar concepts from generic chaining [26]. Let \mathcal{X} be a normed linear space and let (T, d) be a semi-metric space, i.e., $d(x, z) \leq d(x, y) + d(y, z)$ and $d(x, y) = d(y, x)$ for $x, y, z \in T$. To avoid complications with the measurability of suprema of stochastic processes we will always assume that the cardinality $|T|$ of T is *finite*. Criteria for measurability of the supremum of a stochastic process in the case of an

(uncountably) infinite index set can be found in [28, Section 1.7]. We denote the *diameter* of T with respect to d by

$$\Delta_d(T) = \sup_{s,t \in T} d(s,t).$$

We say that an \mathcal{X} -valued process $(X_t)_{t \in T}$ is ψ_α *with respect to* d if for all $s, t \in T$,

$$\mathbb{P}(\|X_t - X_s\| \geq ud(t,s)) \leq 2 \exp(-u^\alpha) \quad (u \geq 0). \tag{2.2}$$

A sequence $\mathcal{T} = (T_n)_{n \geq 0}$ of subsets of T is called *admissible* if $|T_0| = 1$ and $|T_n| \leq 2^{2^n}$ for all $n \geq 1$. For any $0 < \alpha < \infty$, the γ_α -*functional* of (T, d) is defined by

$$\gamma_\alpha(T, d) = \inf_{\mathcal{T}} \sup_{t \in T} \sum_{n=0}^{\infty} 2^{n/\alpha} d(t, T_n),$$

where the infimum is taken over all admissible sequences and we write $d(t, T_n) = \inf_{s \in T_n} d(t, s)$.

For any given $u > 0$ let $N(T, d, u)$ denote the covering number of T , i.e., the smallest number of balls of radius u in (T, d) needed to cover T . One can always estimate

$$\gamma_\alpha(T, d) \lesssim_\alpha \int_0^\infty \left(\log N(T, d, u) \right)^{1/\alpha} du, \tag{2.3}$$

see [26, Section 1.2] for the case $\alpha = 2$ (the other cases are similar). However, the reverse estimate fails in general [26, Section 2.1].

We conclude by fixing some notation. We use $\|\cdot\|_p, 1 \leq p \leq \infty$, to denote the ℓ^p -norms. We will write $A \lesssim_\beta B$ if $A \leq C_\beta B$ for a constant C_β which only depends on a parameter β . Finally, if S is a finite set and $\pi : S \rightarrow \mathbb{R}_+$ is a map, then $\operatorname{argmin}_{s \in S} \pi(s)$ denotes a minimizer of this map, which may not be unique.

3 Suprema of ψ_α and mixed tail processes

We begin our discussion by considering two standard generic chaining situations. First, in Theorem 3.2 we establish tail bounds for suprema of ψ_α processes. At the end of the section, in Theorem 3.5, we do the same for processes with a mixed tail.

In the formulation of our L^p -bounds we will make use of the following truncated version of the γ -functionals. For a given $1 \leq p < \infty$, we will always write $l := \lfloor \log_2(p) \rfloor$, where $\lfloor \cdot \rfloor$ denotes the integer part. We define

$$\gamma_{\alpha,p}(T, d) = \inf_{\mathcal{T}} \sup_{t \in T} \sum_{n \geq l} 2^{n/\alpha} d(t, T_n). \tag{3.1}$$

Clearly, $\gamma_{\alpha,p}(T, d) \leq \gamma_\alpha(T, d)$ for all $1 \leq p < \infty$ and $\gamma_{\alpha,1}(T, d) = \gamma_\alpha(T, d)$. Since we assume T to be finite, the infimum in (3.1) is actually attained. We will call a sequence \mathcal{T} that achieves the infimum *optimal*.

Remark 3.1. If T is an infinite set, then the L^p -bounds presented below continue to hold if we interpret $\mathbb{E} \sup_{t \in T} \|X_t - X_{t_0}\|^p$ as the lattice supremum

$$\mathbb{E} \sup_{t \in T} \|X_t - X_{t_0}\|^p := \sup_{t \in F} \{\mathbb{E} \sup_{t \in F} \|X_t - X_{t_0}\|^p : F \subset T, |F| < \infty\}.$$

Theorem 3.2. *Let $0 < \alpha < \infty$. If $(X_t)_{t \in T}$ is ψ_α , then there exist constants $C_\alpha, D_\alpha > 0$ depending only on α , such that for any $t_0 \in T$ and $1 \leq p < \infty$,*

$$\left(\mathbb{E} \sup_{t \in T} \|X_t - X_{t_0}\|^p \right)^{1/p} \leq C_\alpha \gamma_{\alpha,p}(T, d) + 2 \sup_{t \in T} (\mathbb{E} \|X_t - X_{t_0}\|^p)^{1/p}. \tag{3.2}$$

As a consequence, for any $u \geq 1$,

$$\mathbb{P}\left(\sup_{t \in T} \|X_t - X_{t_0}\| \geq e^{1/\alpha}(C_\alpha \gamma_\alpha(T, d) + u D_\alpha \Delta_d(T))\right) \leq \exp(-u^\alpha/\alpha). \quad (3.3)$$

Remark 3.3. (i) If $(X_t)_{t \in T}$ is a real-valued Gaussian process and d is the canonical distance $d(s, t) := (\mathbb{E}|X_t - X_s|^2)^{1/2}$, then Theorem 3.2 produces a sharp L^p -bound (up to universal constants). Indeed, Talagrand’s majorizing measures theorem [24, 26] states that

$$\gamma_2(T, d) \lesssim \mathbb{E} \sup_{t \in T} |X_t - X_{t_0}|.$$

Moreover, it is of course always true that

$$\sup_{t \in T} (\mathbb{E}|X_t - X_{t_0}|^p)^{1/p} \leq \left(\mathbb{E} \sup_{t \in T} |X_t - X_{t_0}|^p\right)^{1/p}.$$

(ii) Although deviation inequalities are not discussed in [26], it is implicitly used there that

$$\mathbb{P}\left(\sup_{t \in T} \|X_t - X_{t_0}\| \geq uc_1 \gamma_2(T, d)\right) \leq c_2 e^{-u^2/2} \quad (u \geq 2)$$

in the case $\alpha = 2$. Since $\Delta(T) \leq \gamma_2(T, d)$ (and $\Delta(T)$ is potentially much smaller), this bound is qualitatively worse than (3.3). After the first version of this paper was finished, the author learned that (3.3) is proved for $\alpha = 2$ in Talagrand’s new book using a different method (see [27, Theorem 2.2.27]).

(iii) Let $n \in \mathbb{N}$. In [29, Theorem 3.1], it was shown that if $(X_t)_{t \in T}$ is a *sub- n -th chaos process*, meaning in the terminology used here that it is $\psi_{2/n}$, then it satisfies the tail bound

$$\mathbb{P}\left(\sup_{t \in T} |X_t - X_{t_0}| \geq C_n M_n + u C'_n \Delta_d(T)\right) \leq 2 \exp\left(-\frac{u^{2/n}}{2}\right), \quad (3.4)$$

where C_n, C'_n are constants depending only on n and M_n is the entropy integral

$$M_n = \int_0^\infty \left(\log N(T, d, u)\right)^{n/2} du.$$

By (2.3) this result is a direct consequence of Theorem 3.2 (with possibly different constants) but not vice versa. Note that already in the subgaussian case $n = 1$ the bound (3.4) is not sharp (see [26, Section 2.1]).

(iv) To keep our exposition clear we do not keep precise track of the numerical constants in the chaining arguments. However, from the proof below it is clear that C_α, D_α are decreasing in α . Moreover, one can decrease C_α at the expense of increasing D_α (and vice versa). To give an idea of their order of magnitude, one can readily deduce from the proof (without making any effort to optimize the constants) that $C_2 \leq (1 + \sqrt{2})(16\sqrt{\pi}e^{1/2e} + \sqrt{2}) \leq 86$ and $D_2 \leq 4\sqrt{2\pi}e^{1/e}e^{-1/2} \leq 9$. Although these estimates can certainly be improved, the method cannot yield optimal numerical constants. In particular losses occur when passing between moment and tail bounds (cf. Lemmas A.1 and A.2).

Proof of Theorem 3.2. Let $\mathcal{T} = (T_n)_{n \geq 0}$ be an optimal admissible sequence for $\gamma_{\alpha,p}(T, d)$ and let $\pi = (\pi_n)_{n \geq 0}$ be a sequence of functions $\pi_n : T \rightarrow T_n$ defined by $\pi_n(t) = \operatorname{argmin}_{s \in T_n} d(s, t)$. Set $l = \lfloor \log_2(p) \rfloor$. We make the decomposition

$$\left(\mathbb{E} \sup_{t \in T} \|X_t - X_{t_0}\|^p\right)^{\frac{1}{p}} \leq \left(\mathbb{E} \sup_{t \in T} \|X_t - X_{\pi_l(t)}\|^p\right)^{\frac{1}{p}} + \left(\mathbb{E} \sup_{t \in T} \|X_{\pi_l(t)} - X_{t_0}\|^p\right)^{\frac{1}{p}}. \quad (3.5)$$

We estimate the second term on the right hand side by Lemma A.3,

$$\begin{aligned} \left(\mathbb{E} \sup_{t \in T} \|X_{\pi_l(t)} - X_{t_0}\|^p\right)^{\frac{1}{p}} &\leq 2 \sup_{t \in T} (\mathbb{E} \|X_{\pi_l(t)} - X_{t_0}\|^p)^{\frac{1}{p}} \\ &\leq 2 \sup_{t \in T} (\mathbb{E} \|X_t - X_{t_0}\|^p)^{\frac{1}{p}}. \end{aligned} \tag{3.6}$$

For the first term, we write the telescoping sum

$$X_t - X_{\pi_l(t)} = \sum_{n>l} X_{\pi_n(t)} - X_{\pi_{n-1}(t)}.$$

Since the increments of X are ψ_α , we have for $n > l$,

$$\mathbb{P}\left(\|X_{\pi_n(t)} - X_{\pi_{n-1}(t)}\| \geq u2^{n/\alpha}d(\pi_n(t), \pi_{n-1}(t))\right) \leq 2 \exp(-u^\alpha 2^n).$$

Note that $|\{(\pi_n(t), \pi_{n-1}(t)); t \in T\}| \leq |T_n| |T_{n-1}| \leq 2^{2^n} 2^{2^{n-1}} \leq 2^{2^{n+1}}$. Therefore, if $\Omega_{u,p}$ denotes the event

$$\forall n > l, \forall t \in T : \|X_{\pi_n(t)} - X_{\pi_{n-1}(t)}\| \leq u2^{n/\alpha}d(\pi_n(t), \pi_{n-1}(t)),$$

then by a union bound (cf. Lemma A.4),

$$\mathbb{P}(\Omega_{u,p}^c) \leq c \exp(-pu^\alpha/4) \quad (u \geq 2^{1/\alpha}).$$

If the event $\Omega_{u,p}$ occurs, then

$$\begin{aligned} \left\| \sum_{n>l} X_{\pi_n(t)} - X_{\pi_{n-1}(t)} \right\| &\leq \sum_{n>l} \|X_{\pi_n(t)} - X_{\pi_{n-1}(t)}\| \\ &\leq u \sum_{n>l} 2^{n/\alpha}d(\pi_n(t), \pi_{n-1}(t)) \leq u(1 + 2^{1/\alpha})\gamma_{\alpha,p}(T, d). \end{aligned}$$

Thus, $\sup_{t \in T} \|X_t - X_{\pi_l(t)}\| \leq u(1 + 2^{1/\alpha})\gamma_{\alpha,p}(T, d)$. In conclusion,

$$\mathbb{P}\left(\sup_{t \in T} \|X_t - X_{\pi_l(t)}\| > u(1 + 2^{1/\alpha})\gamma_{\alpha,p}(T, d)\right) \leq c \exp(-pu^\alpha/4),$$

whenever $u \geq 2^{1/\alpha}$. Lemma A.5 implies that

$$\left(\mathbb{E} \sup_{t \in T} \|X_t - X_{\pi_l(t)}\|^p\right)^{1/p} \leq C_\alpha \gamma_{\alpha,p}(T, d). \tag{3.7}$$

The moment bound (3.2) follows by combining (3.5), (3.6) and (3.7). For the tail bound, note that (2.2) and Lemma A.2 together imply that

$$\sup_{t \in T} (\mathbb{E} \|X_t - X_{t_0}\|^p)^{1/p} \leq D_\alpha \Delta_d(T) p^{1/\alpha}.$$

The final assertion follows by using this estimate in (3.2) and applying Lemma A.1. \square

Note that Theorem 3.2 does not require any independence assumptions on the increments of the process $(X_t)_{t \in T}$. To illustrate this, we recall the Azuma-Hoeffding inequality, see e.g. [16, Lemma 4.1]. If $X = (X_k)_{0 \leq k \leq n}$ is a discrete-time real-valued martingale and $\Delta_k(X) = X_k - X_{k-1}$ denotes its k -th difference, then

$$\mathbb{P}(|X_n - X_0| \geq u) \leq 2 \exp\left(-\frac{u^2}{2 \sum_{k=1}^n \|\Delta_k(X)\|_\infty^2}\right) \quad (u \geq 0).$$

Combined with Theorem 3.2 we immediately obtain the following uniform version of the Azuma-Hoeffding bound.

Corollary 3.4. *Let $X_t = (X_{t,k})_{1 \leq k \leq n}$, $t \in T$, be a family of discrete-time martingales with respect to the same filtration. We consider the metric*

$$d(s, t) = \left(\sum_{k=1}^n \|\Delta_k(X_t - X_s)\|_\infty^2 \right)^{1/2}.$$

For any $u \geq 1$,

$$\mathbb{P} \left(\sup_{t \in T} |X_{t,n} - X_{t,0}| \geq \sqrt{e}(C_2\gamma_2(T, d) + D_2\Delta_d(T)u) \right) \leq e^{-u^2/2}.$$

Let d_1, d_2 be two semi-metrics on T . We say that a process $(X_t)_{t \in T}$ has mixed subgaussian-subexponential increments, or simply has a *mixed tail*, with respect to the pair (d_1, d_2) if for all $s, t \in T$,

$$\mathbb{P}(\|X_t - X_s\| \geq \sqrt{u}d_2(t, s) + ud_1(t, s)) \leq 2e^{-u} \quad (u \geq 0). \tag{3.8}$$

This means that the first part of the tail behaves as the tail of a subgaussian random variable and the second part as the tail of a subexponential random variable. In Theorem 3.5 we prove a tail bound for the supremum of a process with a mixed tail. The result improves [26, Theorem 1.2.9] and, in fact, positively answers an open question in Talagrand’s new book (see the discussion after [27, Theorem 2.2.28]).

In the proof it will be convenient to work with an alternative definition of the γ_α -functionals. Let us say that a sequence $\mathcal{A} = (\mathcal{A}_n)_{n \geq 0}$ of partitions of T is *admissible* if it is increasing with respect to the refinement ordering and $|\mathcal{A}_0| = 1$ and $|\mathcal{A}_n| \leq 2^{2^n}$ for $n \geq 1$. For any $t \in T$, let $A_n(t)$ be the unique element in the partition \mathcal{A}_n containing t . We now set

$$\gamma'_\alpha(T, d) = \inf_{\mathcal{A}} \sup_{t \in T} \sum_{n=0}^{\infty} 2^{n/\alpha} \Delta_d(A_n(t)),$$

where the infimum is taken over all admissible partitions \mathcal{A} of T . It can be shown that

$$\gamma_\alpha(T, d) \leq \gamma'_\alpha(T, d) \lesssim_\alpha \gamma_\alpha(T, d), \tag{3.9}$$

see the discussion following [26, Theorem 1.3.5] for details.

Theorem 3.5. *If $(X_t)_{t \in T}$ has a mixed tail, then there is a constant $C > 0$ such that for any $1 \leq p < \infty$,*

$$\left(\mathbb{E} \sup_{t \in T} \|X_t - X_{t_0}\|^p \right)^{1/p} \leq C(\gamma_2(T, d_2) + \gamma_1(T, d_1)) + 2 \sup_{t \in T} (\mathbb{E} \|X_t - X_{t_0}\|^p)^{1/p}. \tag{3.10}$$

As a consequence, there are constants $c, C > 0$ such that for any $u \geq 1$,

$$\mathbb{P} \left(\sup_{t \in T} \|X_t - X_{t_0}\| \geq C(\gamma_2(T, d_2) + \gamma_1(T, d_1)) + c(\sqrt{u}\Delta_{d_2}(T) + u\Delta_{d_1}(T)) \right) \leq e^{-u}.$$

Proof. We select two admissible sequences of partitions $\mathcal{B} = (\mathcal{B}_n)_{n \geq 0}$ and $\mathcal{C} = (\mathcal{C}_n)_{n \geq 0}$ such that

$$\begin{aligned} \sup_{t \in T} \sum_{n \geq 0} 2^{n/2} \Delta_{d_2}(\mathcal{B}_n(t)) &\leq 2\gamma'_2(T, d_2) \\ \sup_{t \in T} \sum_{n \geq 0} 2^n \Delta_{d_1}(\mathcal{C}_n(t)) &\leq 2\gamma'_1(T, d_1). \end{aligned}$$

Let \mathcal{A}_n be the partition generated by \mathcal{B}_{n-1} and \mathcal{C}_{n-1} , i.e., for $n \geq 0$

$$\mathcal{A}_n = \{B \cap C : B \in \mathcal{B}_{n-1}, C \in \mathcal{C}_{n-1}\},$$

where we set $\mathcal{B}_{-1} := \mathcal{B}_0$ and $\mathcal{C}_{-1} := \mathcal{C}_0$ so that $\mathcal{A}_0 = \{T\}$. Then $\mathcal{A} = (\mathcal{A}_n)_{n \geq 0}$ is increasing and

$$|\mathcal{A}_n| \leq |\mathcal{B}_{n-1}| |\mathcal{C}_{n-1}| \leq 2^{2^{n-1}} 2^{2^{n-1}} = 2^{2^n},$$

so \mathcal{A} is admissible. Observe that $A_n(t) = B_{n-1}(t) \cap C_{n-1}(t)$. For every $n \geq 0$ define a subset T_n of T by selecting exactly one point from each $A \in \mathcal{A}_n$. In this way, we obtain an admissible sequence $\mathcal{T} = (T_n)_{n \geq 0}$ of subsets of T . For every $n \in \mathbb{N}_{\geq 0}$ and $t \in T$ we let $\pi_n(t)$ be the unique element of $T_n \cap A_n(t)$. This yields a sequence $\pi = (\pi_n)_{n \geq 0}$ of maps $\pi_n : T \rightarrow T_n$.

Set $l = \lfloor \log_2(p) \rfloor$. As in the proof of Theorem 3.2, we make the decomposition (3.5) and estimate the second term on the right hand side of (3.5) as in (3.6). For the first term, we write the telescoping sum

$$X_t - X_{\pi_l(t)} = \sum_{n>l} X_{\pi_n(t)} - X_{\pi_{n-1}(t)}.$$

Since X has a mixed tail, we have for $n > l$ and $u \geq 0$,

$$\begin{aligned} \mathbb{P}\left(\|X_{\pi_n(t)} - X_{\pi_{n-1}(t)}\| \geq \sqrt{u}2^{n/2}d_2(\pi_n(t), \pi_{n-1}(t)) + u2^n d_1(\pi_n(t), \pi_{n-1}(t))\right) \\ \leq 2 \exp(-u2^n). \end{aligned}$$

Note that $|\{(\pi_n(t), \pi_{n-1}(t)); t \in T\}| \leq |T_n| |T_{n-1}| \leq 2^{2^n} 2^{2^{n-1}} \leq 2^{2^{n+1}}$. Therefore, if $\Omega_{u,p}$ denotes the event

$$\begin{aligned} \forall n > l, \forall t \in T : \|X_{\pi_n(t)} - X_{\pi_{n-1}(t)}\| \\ \geq \sqrt{u}2^{n/2}d_2(\pi_n(t), \pi_{n-1}(t)) + u2^n d_1(\pi_n(t), \pi_{n-1}(t)), \end{aligned}$$

then by a union bound (cf. Lemma A.4)

$$\mathbb{P}(\Omega_{u,p}^c) \leq c \exp(-pu/4) \quad (u \geq 2).$$

If the event $\Omega_{u,p}$ occurs, then

$$\begin{aligned} \left\| \sum_{n>l} X_{\pi_n(t)} - X_{\pi_{n-1}(t)} \right\| \\ \leq \sum_{n>l} \|X_{\pi_n(t)} - X_{\pi_{n-1}(t)}\| \\ \leq \sqrt{u} \sum_{n>l} 2^{n/2} d_2(\pi_n(t), \pi_{n-1}(t)) + u \sum_{n>l} 2^n d_1(\pi_n(t), \pi_{n-1}(t)). \end{aligned}$$

Observe that for $n \geq 2$ we have $\pi_n(t), \pi_{n-1}(t) \in A_{n-1}(t) \subset B_{n-2}(t)$ and so

$$d_2(\pi_n(t), \pi_{n-1}(t)) \leq \Delta_{d_2}(B_{n-2}(t)).$$

Also,

$$d_2(\pi_1(t), \pi_0(t)) \leq \Delta_{d_2}(B_0(t)) = \Delta_{d_2}(T).$$

Therefore, by our choice of \mathcal{B} ,

$$\begin{aligned} \sum_{n>l} 2^{n/2} d_2(\pi_n(t), \pi_{n-1}(t)) &\leq \sum_{n>l} 2^{n/2} \Delta_{d_2}(B_{n-2}(t)) \\ &\leq 4 \sum_{n \geq 0} 2^{n/2} \Delta_{d_2}(B_n(t)) \leq 8\gamma'_2(T, d_2). \end{aligned}$$

Analogously, by our choice of \mathcal{C} ,

$$\sum_{n>l} 2^n d_1(\pi_n(t), \pi_{n-1}(t)) \leq 12\gamma'_1(T, d_1)$$

Thus, $\sup_{t \in T} \|X_t - X_{\pi_l(t)}\| \leq 8\sqrt{u}\gamma'_2(T, d_2) + 12u\gamma'_1(T, d_1)$. As a consequence, we conclude that

$$\mathbb{P}\left(\sup_{t \in T} \|X_t - X_{\pi_l(t)}\| > 12u(\gamma'_2(T, d_2) + \gamma'_1(T, d_1))\right) \leq c \exp(-pu/4),$$

whenever $u \geq 2$. By Lemma A.5 and (3.9)

$$\left(\mathbb{E} \sup_{t \in T} \|X_t - X_{\pi_l(t)}\|^p\right)^{1/p} \lesssim \gamma'_2(T, d_2) + \gamma'_1(T, d_1) \lesssim \gamma_2(T, d_2) + \gamma_1(T, d_1).$$

This proves the moment bound (3.10). For the tail bound, note that (3.8) and Lemma A.2 together imply that

$$\sup_{t \in T} (\mathbb{E} \|X_t - X_{t_0}\|^p)^{1/p} \lesssim \Delta_{d_2}(T)\sqrt{p} + \Delta_{d_1}(T)p.$$

The assertion follows by using this estimate in (3.10) and applying Lemma A.1. \square

In Corollary 5.2 below we use Theorem 3.5 to derive tail bounds for suprema of empirical processes.

4 Restricted isometry constants of subsampled unitary matrices

In this section we present an application of Theorem 3.2 in compressed sensing. We use the following terminology. We say that $x \in \mathbb{C}^N$ is s -sparse if

$$\|x\|_0 = |\{i : x_i \neq 0\}| \leq s.$$

For a given $s \in \mathbb{N}$, the s -th restricted isometry constant δ_s of an $m \times N$ matrix A is the smallest constant $\delta \geq 0$ such that

$$(1 - \delta)\|x\|_2^2 \leq \|Ax\|_2^2 \leq (1 + \delta)\|x\|_2^2,$$

for all s -sparse $x \in \mathbb{C}^N$. Equivalently, if we let

$$D_{s,N} = \{x \in \mathbb{C}^N : \|x\|_2 = 1, \|x\|_0 \leq s\},$$

then

$$\delta_s = \sup_{x \in D_{s,N}} \left| \|Ax\|_2^2 - 1 \right|.$$

The restricted isometry constants play an important role in compressed sensing, see [8, Chapter 6] for more information. We restrict ourselves to the task of giving a simpler proof of the fact that the random matrix obtained by uniformly sampling rows of the discrete Fourier transform has small restricted isometry constants with high probability. This result was obtained by Candès and Tao in the influential paper [6]. An improved result was later found by Rudelson and Vershynin [22] using a different method. Finally, by elaborating on this method a better probability estimate was obtained by Rauhut [21].

We consider the following (more general) setup. Let U be a unitary $N \times N$ matrix and suppose that for some constant $K \geq 1$,

$$\sup_{k,l} \sqrt{N}|U_{kl}| \leq K. \tag{4.1}$$

We consider a sequence $(\theta_i)_{1 \leq i \leq N}$ of i.i.d. copies of the random selector $\theta : \Omega \rightarrow \{0, 1\}$ which satisfies

$$\mathbb{P}(\theta = 1) = \frac{m}{N}.$$

Let $I = \{i \in [N] : \theta_i = 1\}$ be the random set of selected indices and note that its expected cardinality is $\mathbb{E}|I| = m$. Let $R_I : \mathbb{C}^N \rightarrow \mathbb{C}^{|I|}$ be the operator which restricts a vector to its entries in I and consider the subsampled and rescaled matrix

$$U_I := \sqrt{\frac{N}{m}} R_I U. \tag{4.2}$$

The subsampled discrete Fourier transform corresponds to taking

$$U_{kl} = \frac{1}{\sqrt{N}} \exp(2\pi i(k-1)(l-1)/N) \quad (k, l = 1, \dots, N).$$

Theorem 4.1. [6, 21, 22] *Let U and I be as above. Set $\delta_s = \delta_s(U_I)$. There exist universal constants $d_1, d_2 > 0$ such that for any given $s \in \mathbb{N}$ and $0 < \delta, \eta < 1$, we have $\mathbb{P}(\delta_s \geq \delta) \leq \eta$, provided that*

$$m \geq sK^2\delta^{-2} \max\{d_1 \log^2 s \log m \log N, d_2 \log(\eta^{-1})\}. \tag{4.3}$$

The proof of Theorem 4.1 in [21] (see also [8, Theorem 12.32]), which refines the approach in [22], consists of two parts: firstly, the expected value of δ_s is estimated using a (classical) chaining argument. Secondly, a deviation inequality for suprema of bounded empirical processes is used to show that δ_s is typically not much larger than its expected value. Here we shorten the proof by merging these two steps, hence dispensing with the concentration inequality. Note that we still use a certain entropy bound obtained in [22] (see (4.6) below), which is nontrivial to prove.

Proof of Theorem 4.1. Let U_i be the i -th row of U . For every $x \in D_{s,N}$ we define $f_x(\theta_i) = \theta_i \frac{1}{\sqrt{m}} \langle \sqrt{N} U_i, x \rangle$. Since U is unitary,

$$\sum_{i=1}^N \mathbb{E} f_x^2(\theta_i) = \frac{N}{m} \sum_{i=1}^N \mathbb{E}(\theta_i) |\langle U_i, x \rangle|^2 = \sum_{i=1}^N |\langle U_i, x \rangle|^2 = \|Ux\|_2^2 = 1, \tag{4.4}$$

and therefore we can write

$$\delta_s = \sup_{x \in D_{s,N}} \left| \sum_{i=1}^N f_x^2(\theta_i) - \mathbb{E} f_x^2(\theta_i) \right|.$$

Fix $1 \leq p < \infty$ and let $(\varepsilon_i)_{i \geq 1}$ be a Rademacher sequence, i.e., a sequence of independent symmetric Bernoulli random variables. By a standard symmetrization argument [17, Lemma 6.3],

$$(\mathbb{E} \delta_s^p)^{1/p} \leq 2 \left(\mathbb{E} \mathbb{E}_\varepsilon \sup_{x \in D_{s,N}} \left| \sum_{i=1}^N \varepsilon_i f_x^2(\theta_i) \right|^p \right)^{1/p}. \tag{4.5}$$

Now we fix $\omega \in \Omega$ and let $t_i = \theta_i(\omega)$. By Hoeffding's inequality,

$$\mathbb{P}_\varepsilon \left(\sum_{i=1}^N \varepsilon_i (f_x^2(t_i) - f_y^2(t_i)) \geq u \left(\sum_{i=1}^N (f_x^2(t_i) - f_y^2(t_i))^2 \right)^{1/2} \right) \leq \exp(-u^2/2).$$

Moreover,

$$\left(\sum_{i=1}^N (f_x^2(t_i) - f_y^2(t_i))^2 \right)^{1/2} = \left(\sum_{i=1}^N (f_x(t_i) - f_y(t_i))^2 (f_x(t_i) + f_y(t_i))^2 \right)^{1/2}$$

Tail bounds via generic chaining

$$\leq 2 \sup_{z \in D_{s,N}} \left(\sum_{i=1}^N f_z^2(t_i) \right)^{1/2} \max_{1 \leq i \leq N} |f_x(t_i) - f_y(t_i)|.$$

In conclusion, the process

$$x \mapsto \sum_{i=1}^N \varepsilon_i f_x^2(t_i)$$

is subgaussian with respect to the metric

$$d(x, y) = \sup_{z \in D_{s,N}} \left(\sum_{i=1}^N f_z^2(t_i) \right)^{1/2} d_t(x, y),$$

where d_t denotes the metric

$$d_t(x, y) = \max_{1 \leq i \leq N} |f_x(t_i) - f_y(t_i)|.$$

By Theorem 3.2,

$$\begin{aligned} & \left(\mathbb{E}_\varepsilon \sup_{x \in D_{s,N}} \left| \sum_{i=1}^N \varepsilon_i f_x^2(t_i) \right|^p \right)^{1/p} \\ & \lesssim \gamma_2(D_{s,N}, d_t) \sup_{x \in D_{s,N}} \left(\sum_{i=1}^N f_x^2(t_i) \right)^{1/2} + \sup_{x \in D_{s,N}} \left(\mathbb{E}_\varepsilon \left| \sum_{i=1}^N \varepsilon_i f_x^2(t_i) \right|^p \right)^{1/p}. \end{aligned}$$

We apply (2.3) with $\alpha = 2$, i.e.,

$$\gamma_2(D_{s,N}, d_t) \lesssim \int_0^\infty \left(\log N(D_{s,N}, d_t, u) \right)^{1/2} du$$

and use that Rudelson and Vershynin already proved that (see inequalities (3.8) and (3.9) in [22], or [8, 21])

$$\int_0^\infty \left(\log N(D_{s,N}, d_t, u) \right)^{1/2} du \lesssim K \sqrt{\frac{s}{m}} \log s \sqrt{\log m} \sqrt{\log N}. \quad (4.6)$$

Moreover, by Khintchine's (or Hoeffding's) inequality,

$$\left(\mathbb{E}_\varepsilon \left| \sum_{i=1}^N \varepsilon_i f_x^2(t_i) \right|^p \right)^{1/p} \leq \sqrt{p} \left(\sum_{i=1}^N f_x^4(t_i) \right)^{1/2} \leq \sqrt{p} \left(\sum_{i=1}^N f_x^2(t_i) \right)^{1/2} \max_{1 \leq i \leq N} |f_x(t_i)|.$$

By Hölder's inequality and (4.1)

$$\max_{1 \leq i \leq N} |f_x(t_i)| \leq \max_{1 \leq i \leq N} \frac{1}{\sqrt{m}} \|\sqrt{N} U_i\|_\infty \|x\|_1 \leq K \sqrt{\frac{s}{m}},$$

where the final inequality follows from the s -sparsity of x . Collecting our estimates we find

$$\begin{aligned} & \left(\mathbb{E}_\varepsilon \sup_{x \in D_{s,N}} \left| \sum_{i=1}^N \varepsilon_i f_x^2(t_i) \right|^p \right)^{1/p} \\ & \lesssim \sup_{x \in D_{s,N}} \left(\sum_{i=1}^N f_x^2(t_i) \right)^{1/2} K \sqrt{\frac{s}{m}} \left(\log s \sqrt{\log m} \sqrt{\log N} + \sqrt{p} \right). \end{aligned}$$

We now take the L^p norm on Ω on both sides and obtain using (4.5)

$$\begin{aligned} & \left(\mathbb{E} \sup_{x \in D_{s,N}} \left| \sum_{i=1}^N f_x^2(\theta_i) - \mathbb{E} f_x^2(\theta_i) \right|^p \right)^{1/p} \\ & \lesssim \left(\mathbb{E} \sup_{x \in D_{s,N}} \left(\sum_{i=1}^N f_x^2(\theta_i) \right)^{p/2} \right)^{1/p} K \sqrt{\frac{s}{m}} \left(\log s \sqrt{\log m} \sqrt{\log N} + \sqrt{p} \right) \\ & \lesssim \left(\mathbb{E} \sup_{x \in D_{s,N}} \left| \sum_{i=1}^N f_x^2(\theta_i) - \mathbb{E} f_x^2(\theta_i) \right|^p \right)^{1/2p} K \sqrt{\frac{s}{m}} \left(\log s \sqrt{\log m} \sqrt{\log N} + \sqrt{p} \right) \\ & \quad + K \sqrt{\frac{s}{m}} \left(\log s \sqrt{\log m} \sqrt{\log N} + \sqrt{p} \right), \end{aligned} \tag{4.7}$$

where in the final step we use (4.4). Note that (4.7) is a quadratic inequality in $(\mathbb{E} \delta_s^p)^{1/2p}$. By solving it and subsequently squaring both sides we find

$$\begin{aligned} (\mathbb{E} \delta_s^p)^{1/p} & \lesssim K \sqrt{\frac{s}{m}} \log s \sqrt{\log m} \sqrt{\log N} + K^2 \frac{s}{m} \log^2 s \log m \log N \\ & \quad + \sqrt{p} K \sqrt{\frac{s}{m}} + p K^2 \frac{s}{m}. \end{aligned}$$

Since $1 \leq p < \infty$ was arbitrary, Lemma A.1 implies that for any $u \geq 1$,

$$\begin{aligned} \mathbb{P} \left(\delta_s \gtrsim K \sqrt{\frac{s}{m}} \log s \sqrt{\log m} \sqrt{\log N} + K^2 \frac{s}{m} \log^2 s \log m \log N \right. \\ \left. + \sqrt{u} K \sqrt{\frac{s}{m}} + u K^2 \frac{s}{m} \right) \leq e^{-u}. \end{aligned}$$

Therefore, if we set $u = \log(\eta^{-1})$ and pick m as in (4.3), then

$$\begin{aligned} \mathbb{P}(\delta_s \geq \delta) & \leq \mathbb{P} \left(\delta_s \gtrsim K \sqrt{\frac{s}{m}} \log s \sqrt{\log m} \sqrt{\log N} + K^2 \frac{s}{m} \log^2 s \log m \log N \right. \\ & \quad \left. + \sqrt{\log(\eta^{-1})} K \sqrt{\frac{s}{m}} + \log(\eta^{-1}) K^2 \frac{s}{m} \right) \leq \eta. \end{aligned}$$

□

A small modification of the proof of Theorem 4.1 yields the following result. It implies in particular the restricted isometry property of matrices obtained by sampling from bounded orthonormal systems, which was established in [21, Theorem 8.4] (see also [8, Theorem 12.32]).

Theorem 4.2. *Let A be an $m \times N$ random matrix with independent rows given by $\frac{1}{\sqrt{m}} X_1, \dots, \frac{1}{\sqrt{m}} X_m$. Suppose that*

$$\frac{1}{m} \sum_{i=1}^m \mathbb{E} |\langle X_i, x \rangle|^2 = \|x\|_2^2 \quad \text{for all } x \in \mathbb{C}^N$$

and

$$\max_{1 \leq i \leq m} \|X_i\|_\infty \leq K.$$

Then there exist universal constants $d_1, d_2 > 0$ such that for any given $s \in \mathbb{N}$ and $0 < \delta, \eta < 1$, we have $\mathbb{P}(\delta_s(A) \geq \delta) \leq \eta$, provided that (4.3) holds.

5 Supremum of an empirical process

In this section we investigate tail bounds for suprema of empirical processes. We begin by applying Theorem 3.5 to these processes. For this purpose we recall Bernstein's inequality. For a proof of this result, see for example [4, Theorem 2.10].

Lemma 5.1. (Bernstein's inequality) *Let X_1, \dots, X_m be real-valued, independent, mean-zero random variables and suppose that for some constants $\sigma, K > 0$,*

$$\frac{1}{m} \sum_{i=1}^m \mathbb{E}|X_i|^q \leq \frac{q!}{2} \sigma^2 K^{q-2}, \quad (q = 2, 3, \dots).$$

Then,

$$\mathbb{P}\left(\left|\frac{1}{m} \sum_{i=1}^m X_i\right| \geq \frac{\sigma}{\sqrt{m}} \sqrt{2u} + \frac{K}{m} u\right) \leq 2 \exp(-u) \quad (u \geq 0). \quad (5.1)$$

In particular, if X_1, \dots, X_m are subexponential, then

$$\mathbb{P}\left(\left|\frac{1}{m} \sum_{i=1}^m X_i\right| \geq \frac{\nu}{\sqrt{m}} \sqrt{2u} + \frac{\kappa}{m} u\right) \leq 2 \exp(-u) \quad (u \geq 0), \quad (5.2)$$

where $\nu^2 = \frac{1}{m} \sum_{i=1}^m \|X_i\|_{\psi_1}^2$ and $\kappa = \max_{1 \leq i \leq m} \|X_i\|_{\psi_1}$.

Consider the following setup. Fix an $m \in \mathbb{N}$ and consider m probability spaces $(\Omega_1, \mathbb{P}_1), \dots, (\Omega_m, \mathbb{P}_m)$. Suppose that we are given a parameter set T consisting of m -tuples $t = (t_1, \dots, t_m)$. For every $t \in T$ we are given an m -tuple $X_t = (X_{t_1}, \dots, X_{t_m})$ of subexponential random variables $X_{t_i} : \Omega_i \rightarrow \mathbb{R}$. We consider the empirical process

$$E_t = \frac{1}{m} \sum_{i=1}^m (X_{t_i} - \mathbb{E}X_{t_i}).$$

In the terminology used here, Bernstein's inequality (5.2) implies that the process $(E_t)_{t \in T}$ has a mixed tail with respect to the metrics $(\frac{1}{m} d_1, \frac{1}{\sqrt{m}} d_2)$, where

$$d_1(s, t) = \max_{1 \leq i \leq m} \|X_{t_i} - X_{s_i}\|_{\psi_1},$$

$$d_2(s, t) = \left(\frac{1}{m} \sum_{i=1}^m \|X_{t_i} - X_{s_i}\|_{\psi_1}^2\right)^{1/2}.$$

Theorem 3.5 can directly be applied to find the following tail bound.

Corollary 5.2. *Let E_t be as above and let $\sigma, K > 0$ be constants such that*

$$\sup_{t \in T} \frac{1}{m} \sum_{i=1}^m \mathbb{E}|X_{t_i} - \mathbb{E}X_{t_i}|^q \leq \frac{q!}{2} \sigma^2 K^{q-2}, \quad (q = 2, 3, \dots).$$

Then, for any $1 \leq p < \infty$,

$$\left(\mathbb{E} \sup_{t \in T} |E_t|^p\right)^{1/p} \lesssim \left(\frac{1}{\sqrt{m}} \gamma_2(T, d_2) + \frac{1}{m} \gamma_1(T, d_1)\right) + \sqrt{p} \frac{\sigma}{\sqrt{m}} + p \frac{K}{m}.$$

In particular, there exist constants $c, C > 0$ such that for any $u \geq 1$,

$$\mathbb{P}\left(\sup_{t \in T} |E_t| \geq C \left(\frac{1}{\sqrt{m}} \gamma_2(T, d_2) + \frac{1}{m} \gamma_1(T, d_1)\right) + c \left(\frac{\sigma}{\sqrt{m}} \sqrt{u} + \frac{K}{m} u\right)\right) \leq e^{-u}. \quad (5.3)$$

Inequality (5.3) can be compared to a deviation inequality in [9, Theorem 8]. In this result the generic chaining estimate

$$\frac{1}{\sqrt{m}}\gamma_2(T, d_2) + \frac{1}{m}\gamma_1(T, d_1)$$

occurring in (5.3) is replaced by an estimate obtained by ‘chaining along a tree’, which is a variation of classical (i.e., non-generic) chaining. As a consequence, the estimate (5.3) is in general better. The parameters σ and K governing the tail behavior in (5.3) are the same in [9]. Note that [9] also contains a tail bound obtained by ‘generic chaining along a tree’ (see Theorem 3 there). However, the parameters governing the tail behavior in the latter result still depend on the metric complexity of the index set T .

5.1 Supremum of an average of squares

We continue in the above setup, but now assume that the random variables $X_{t_i} : \Omega_i \rightarrow \mathbb{R}$ are *subgaussian* instead of subexponential. For every t in T we consider the average

$$A_t = \frac{1}{m} \sum_{i=1}^m (X_{t_i}^2 - \mathbb{E}X_{t_i}^2). \tag{5.4}$$

Clearly, we could use Corollary 5.2 to find L^p -bounds for $\sup_{t \in T} A_t$. In this section, we will however look for a more natural bound involving a metric defined in terms of the X_{t_i} instead of their squares. This type of result turns out to give better bounds in certain applications. The main result of this section, Theorem 5.5, improves a result in this direction of Mendelson, Pajor and Tomczak-Jaegermann [19] (see Theorem 5.6 and Corollary 5.7 for a detailed comparison). We consider the metric d_{ψ_2} on T defined by

$$d_{\psi_2}(s, t) = \max_{i=1, \dots, m} \|X_{s_i} - X_{t_i}\|_{\psi_2}.$$

We define the associated radius of T by

$$\bar{\Delta}_{\psi_2}(T) = \sup_{t \in T} \max_{i=1, \dots, m} \|X_{t_i}\|_{\psi_2}$$

and usually write $\bar{\Delta}_{\psi_2}$ instead of $\bar{\Delta}_{\psi_2}(T)$ for brevity. Finally, we denote by μ_m the normalized counting measure on $\{1, \dots, m\}$. With this notation,

$$\|X_t - X_s\|_{L^2(\mu_m)} = \left(\frac{1}{m} \sum_{i=1}^m (X_{t_i} - X_{s_i})^2 \right)^{1/2}.$$

Lemma 5.3. *Let $s, t \in T$. For any $u \geq 1$,*

$$\mathbb{P}\left(\|X_t - X_s\|_{L^2(\mu_m)} \geq u2(1 + \sqrt{2})d_{\psi_2}(s, t)\right) \leq 2 \exp(-mu^2). \tag{5.5}$$

Proof. Consider Bernstein’s inequality (5.2). If $u \geq m$ then, using that $\nu \leq \kappa$, we have

$$\sqrt{2}\nu \sqrt{\frac{u}{m}} \leq \sqrt{2}\kappa \frac{u}{m}.$$

Thus, for $u \geq m$,

$$\mathbb{P}\left(\left|\frac{1}{m} \sum_{i=1}^m X_i\right| \geq (1 + \sqrt{2})\kappa \frac{u}{m}\right) \leq 2 \exp(-u).$$

We apply this inequality for $X_i = (X_{t_i} - X_{s_i})^2 - \mathbb{E}(X_{t_i} - X_{s_i})^2$. Note that in this case

$$\kappa = \max_{1 \leq i \leq m} \|(X_{t_i} - X_{s_i})^2 - \mathbb{E}(X_{t_i} - X_{s_i})^2\|_{\psi_1} \leq 2 \max_{1 \leq i \leq m} \|X_{t_i} - X_{s_i}\|_{\psi_2}^2 = 2d_{\psi_2}^2(s, t).$$

Therefore, we find for any $u \geq 1$,

$$\begin{aligned} \mathbb{P}\left(\frac{1}{m} \sum_{i=1}^m (X_{t_i} - X_{s_i})^2 - \frac{1}{m} \sum_{i=1}^m \mathbb{E}(X_{t_i} - X_{s_i})^2 \geq u2(1 + \sqrt{2})d_{\psi_2}^2(s, t)\right) \\ \leq 2 \exp(-mu). \end{aligned}$$

Since

$$\frac{1}{m} \sum_{i=1}^m \mathbb{E}(X_{t_i} - X_{s_i})^2 \leq \max_{i=1, \dots, m} \|X_{t_i} - X_{s_i}\|_{\psi_2}^2 = d_{\psi_2}^2(s, t),$$

we deduce that

$$\begin{aligned} \mathbb{P}\left(\|X_t - X_s\|_{L^2(\mu_m)} \geq u2(1 + \sqrt{2})d_{\psi_2}(s, t)\right) \\ = \mathbb{P}\left(\frac{1}{m} \sum_{i=1}^m (X_{t_i} - X_{s_i})^2 - \frac{1}{m} \sum_{i=1}^m \mathbb{E}(X_{t_i} - X_{s_i})^2 \geq u^24(1 + \sqrt{2})^2d_{\psi_2}^2(s, t) - \frac{1}{m} \sum_{i=1}^m \mathbb{E}(X_{t_i} - X_{s_i})^2\right) \\ \leq \mathbb{P}\left(\frac{1}{m} \sum_{i=1}^m (X_{t_i} - X_{s_i})^2 - \frac{1}{m} \sum_{i=1}^m \mathbb{E}(X_{t_i} - X_{s_i})^2 \geq 2(1 + \sqrt{2})\left(2(1 + \sqrt{2})u^2 - \frac{1}{2(1 + \sqrt{2})}\right)d_{\psi_2}^2(s, t)\right) \\ \leq 2 \exp\left(-m\left(2(1 + \sqrt{2})u^2 - \frac{1}{2(1 + \sqrt{2})}\right)\right) \leq 2 \exp(-mu^2), \end{aligned}$$

if $u \geq 1$. □

The following two tail bounds will be used in the proof of Theorem 5.5.

Lemma 5.4. *Let $s, t \in T$ and $n \in \mathbb{N}$. If $2^{n/2} \leq \sqrt{m}$ then for any $u \geq 1$,*

$$\mathbb{P}\left(|A_t - A_s| \geq u2^{n/2} \frac{10\bar{\Delta}_{\psi_2}}{\sqrt{m}} d_{\psi_2}(s, t)\right) \leq 2 \exp(-2^n u).$$

On the other hand, if $2^{n/2} \geq \sqrt{m}$, then for any $u \geq 1$,

$$\mathbb{P}\left(\|X_t - X_s\|_{L^2(\mu_m)} \geq \sqrt{u}2^{n/2} \frac{5}{\sqrt{m}} d_{\psi_2}(s, t)\right) \leq 2 \exp(-2^n u).$$

Proof. Suppose first that $2^{n/2} \leq \sqrt{m}$. We apply Bernstein's inequality (5.2) with $X_i = X_{t_i}^2 - X_{s_i}^2 - \mathbb{E}(X_{t_i}^2 - X_{s_i}^2)$. Let us first estimate the deviation parameters. We use the Hölder type inequality (2.1) to obtain

$$\begin{aligned} \nu &= \left(\frac{1}{m} \sum_{i=1}^m \|X_{t_i}^2 - X_{s_i}^2 - \mathbb{E}(X_{t_i}^2 - X_{s_i}^2)\|_{\psi_1}^2\right)^{1/2} \\ &\leq 2 \max_{1 \leq i \leq m} \|X_{t_i}^2 - X_{s_i}^2\|_{\psi_1} \\ &\leq 2 \max_{1 \leq i \leq m} \|X_{t_i} - X_{s_i}\|_{\psi_2} \|X_{t_i} + X_{s_i}\|_{\psi_2} \\ &\leq 4\bar{\Delta}_{\psi_2} d_{\psi_2}(s, t). \end{aligned}$$

Similarly, $\kappa \leq 4\bar{\Delta}_{\psi_2} d_{\psi_2}(s, t)$. Thus, by (5.2), for any $v \geq 0$,

$$\mathbb{P}\left(|A_t - A_s| \geq \sqrt{2v} \frac{4\bar{\Delta}_{\psi_2} d_{\psi_2}(s, t)}{\sqrt{m}} + v \frac{4\bar{\Delta}_{\psi_2} d_{\psi_2}(s, t)}{m}\right) \leq 2 \exp(-v).$$

Taking $v = 2^n u$ yields

$$\mathbb{P}\left(|A_t - A_s| \geq \sqrt{2u} 2^{n/2} \frac{4\bar{\Delta}_{\psi_2} d_{\psi_2}(s, t)}{\sqrt{m}} + u 2^n \frac{4\bar{\Delta}_{\psi_2} d_{\psi_2}(s, t)}{m}\right) \leq 2 \exp(-2^n u).$$

Now observe that $2^{n/2} \leq \sqrt{m}$ implies that

$$2^n \frac{4\bar{\Delta}_{\psi_2} d_{\psi_2}(s, t)}{m} \leq 2^{n/2} \frac{4\bar{\Delta}_{\psi_2} d_{\psi_2}(s, t)}{\sqrt{m}}.$$

Therefore,

$$\mathbb{P}\left(|A_t - A_s| \geq (\sqrt{2u} + u) 2^{n/2} \frac{4\bar{\Delta}_{\psi_2} d_{\psi_2}(s, t)}{\sqrt{m}}\right) \leq 2 \exp(-2^n u).$$

By using that $u \geq 1$ we obtain the first assertion.

Suppose now that $2^{n/2} \geq \sqrt{m}$. Lemma 5.3 implies that for any $v \geq 1$,

$$\mathbb{P}\left(\|X_t - X_s\|_{L^2(\mu_m)} \geq v 5 d_{\psi_2}(s, t)\right) \leq 2 \exp(-mv^2)$$

Let $v = 2^{n/2} \sqrt{u} \frac{1}{\sqrt{m}}$, then $v \geq 1$ and therefore,

$$\begin{aligned} \mathbb{P}\left(\|X_t - X_s\|_{L^2(\mu_m)} \geq \sqrt{u} 2^{n/2} \frac{5}{\sqrt{m}} d_{\psi_2}(s, t)\right) \\ \leq 2 \exp\left(-m \left(2^{n/2} \sqrt{u} \frac{1}{\sqrt{m}}\right)^2\right) = 2 \exp(-2^n u). \end{aligned}$$

□

We are now ready to prove the main result of this section.

Theorem 5.5. *Let $(A_t)_{t \in T}$ be the process of averages defined in (5.4). Let σ, K be constants satisfying*

$$\sup_{t \in T} \frac{1}{m} \sum_{i=1}^m \mathbb{E}|X_{t_i}^2 - \mathbb{E}X_{t_i}^2|^q \leq \frac{q!}{2} \sigma^2 K^{q-2} \quad (q = 2, 3, \dots). \quad (5.6)$$

For any $1 \leq p < \infty$,

$$\left(\mathbb{E} \sup_{t \in T} |A_t|^p\right)^{\frac{1}{p}} \lesssim \frac{1}{m} \gamma_{2,p}^2(T, d_{\psi_2}) + \frac{1}{\sqrt{m}} \bar{\Delta}_{\psi_2}(T) \gamma_{2,p}(T, d_{\psi_2}) + \sqrt{p} \frac{\sigma}{\sqrt{m}} + p \frac{K}{m}.$$

As a consequence, there are constants $c, C > 0$ such that for all $u \geq 1$,

$$\mathbb{P}\left(\sup_{t \in T} |A_t| \geq C \left(\frac{1}{m} \gamma_2^2(T, d_{\psi_2}) + \frac{\bar{\Delta}_{\psi_2}(T)}{\sqrt{m}} \gamma_2(T, d_{\psi_2})\right) + c \left(\sqrt{u} \frac{\sigma}{\sqrt{m}} + u \frac{K}{m}\right)\right) \leq e^{-u}.$$

Note that we can always take the parameters

$$\begin{aligned} \sigma &= \sup_{t \in T} \left(\frac{1}{m} \sum_{i=1}^m \|X_{t_i}\|_{\psi_2}^4\right)^{1/2} \\ K &= \sup_{t \in T} \max_{1 \leq i \leq m} \|X_i\|_{\psi_2}^2. \end{aligned}$$

Proof. We again write $\bar{\Delta}_{\psi_2} := \bar{\Delta}_{\psi_2}(T)$ for brevity. Set $l = \lfloor \log_2(p) \rfloor$. Let \mathcal{T} be an optimal admissible sequence for $\gamma_{2,p}(T, d_{\psi_2})$ and let $\pi_n(t) = \operatorname{argmin}_{s \in T_n} d_{\psi_2}(s, t)$. We divide $\mathbb{N}_{>l}$ into two disjoint parts given by

$$I_{\text{subg}} = \left\{n \in \mathbb{N}_{>l} : 2^{n/2} \leq \sqrt{m}\right\}, \quad I_{\text{subex}} = \left\{n \in \mathbb{N}_{>l} : 2^{n/2} > \sqrt{m}\right\}.$$

We write the telescoping sum

$$A_t = \sum_{n \in I_{\text{subg}}} A_{\pi_n(t)} - A_{\pi_{n-1}(t)} + \sum_{n \in I_{\text{subex}}} A_{\pi_n(t)} - A_{\pi_{n-1}(t)} + A_{\pi_l(t)}. \quad (5.7)$$

By Lemma 5.4, if $n \in I_{\text{subg}}$ then for all $t \in T$ and $u \geq 1$,

$$\mathbb{P}\left(|A_{\pi_n(t)} - A_{\pi_{n-1}(t)}| \geq u \frac{10\bar{\Delta}_{\psi_2}}{\sqrt{m}} 2^{n/2} d_{\psi_2}(\pi_n(t), \pi_{n-1}(t))\right) \leq 2 \exp(-2^n u),$$

whereas if $n \in I_{\text{subex}}$, then for all $u \geq 1$,

$$\mathbb{P}\left(\|X_{\pi_n(t)} - X_{\pi_{n-1}(t)}\|_{L^2(\mu_m)} \geq \sqrt{u} \frac{5}{\sqrt{m}} 2^{n/2} d_{\psi_2}(\pi_n(t), \pi_{n-1}(t))\right) \leq 2 \exp(-2^n u). \quad (5.8)$$

Let $\Omega_{u,p}$ be the event

$$\forall n \in I_{\text{subg}}, t \in T : |A_{\pi_n(t)} - A_{\pi_{n-1}(t)}| \leq u \frac{10\bar{\Delta}_{\psi_2}}{\sqrt{m}} 2^{n/2} d_{\psi_2}(\pi_n(t), \pi_{n-1}(t)),$$

$$\forall n \in I_{\text{subex}}, t \in T : \|X_{\pi_n(t)} - X_{\pi_{n-1}(t)}\|_{L^2(\mu_m)} \leq \sqrt{u} \frac{5}{\sqrt{m}} 2^{n/2} d_{\psi_2}(\pi_n(t), \pi_{n-1}(t)).$$

Since for any $n > l$ the number of pairs $(\pi_n(t), \pi_{n-1}(t))$ is bounded by $|T_n| |T_{n-1}| \leq 2^{2^n} 2^{2^{n-1}} \leq 2^{2^{n+1}}$, a union bound (see Lemma A.4) implies that there is an absolute constant $c > 0$ such that if $u \geq 2$,

$$\mathbb{P}(\Omega_{u,p}^c) \leq c \exp(-pu/4).$$

If the event $\Omega_{u,p}$ occurs, then for any given $t \in T$,

$$\begin{aligned} \left| \sum_{n \in I_{\text{subg}}} A_{\pi_n(t)} - A_{\pi_{n-1}(t)} \right| &\leq u \frac{10\bar{\Delta}_{\psi_2}}{\sqrt{m}} \sum_{n \in I_{\text{subg}}} 2^{n/2} d_{\psi_2}(\pi_n(t), \pi_{n-1}(t)) \\ &\leq u \frac{10(1 + \sqrt{2})\bar{\Delta}_{\psi_2}}{\sqrt{m}} \gamma_{2,p}(T, d_{\psi_2}). \end{aligned}$$

For the subexponential part we write

$$\begin{aligned} &\left| \sum_{n \in I_{\text{subex}}} A_{\pi_n(t)} - A_{\pi_{n-1}(t)} \right| \\ &= \left| \sum_{n \in I_{\text{subex}}} \frac{1}{m} \sum_{i=1}^m X_{\pi_n(t)_i}^2 - X_{\pi_{n-1}(t)_i}^2 - \mathbb{E}\left(\frac{1}{m} \sum_{i=1}^m X_{\pi_n(t)_i}^2 - X_{\pi_{n-1}(t)_i}^2\right) \right|. \end{aligned} \quad (5.9)$$

By the Cauchy-Schwarz inequality in $L^2(\mu_m)$, we find

$$\begin{aligned} &\left| \frac{1}{m} \sum_{i=1}^m X_{\pi_n(t)_i}^2 - X_{\pi_{n-1}(t)_i}^2 \right| \\ &= \left| \frac{1}{m} \sum_{i=1}^m (X_{\pi_n(t)_i} - X_{\pi_{n-1}(t)_i})(X_{\pi_n(t)_i} + X_{\pi_{n-1}(t)_i}) \right| \\ &\leq \left(\frac{1}{m} \sum_{i=1}^m (X_{\pi_n(t)_i} - X_{\pi_{n-1}(t)_i})^2 \right)^{1/2} \left(\frac{1}{m} \sum_{i=1}^m (X_{\pi_n(t)_i} + X_{\pi_{n-1}(t)_i})^2 \right)^{1/2} \\ &\leq \|X_{\pi_n(t)} - X_{\pi_{n-1}(t)}\|_{L^2(\mu_m)} \left(A_{\pi_n(t)} + \frac{1}{m} \sum_{i=1}^m \mathbb{E} X_{\pi_n(t)_i}^2 \right)^{1/2} \end{aligned}$$

$$\begin{aligned}
 & + \left(A_{\pi_{n-1}(t)} + \frac{1}{m} \sum_{i=1}^m \mathbb{E} X_{\pi_{n-1}(t)_i}^2 \right)^{1/2} \\
 & \leq 2 \left(\sup_{t \in T} |A_t| + \bar{\Delta}_{\psi_2}^2 \right)^{1/2} \sqrt{u} 2^{n/2} \frac{5}{\sqrt{m}} d_{\psi_2}(\pi_n(t), \pi_{n-1}(t)) \\
 & \leq 2 \left(\sup_{t \in T} |A_t|^{1/2} + \bar{\Delta}_{\psi_2} \right) \sqrt{u} 2^{n/2} \frac{5}{\sqrt{m}} d_{\psi_2}(\pi_n(t), \pi_{n-1}(t)).
 \end{aligned}$$

On the other hand,

$$\begin{aligned}
 & \left| \mathbb{E} \left(\frac{1}{m} \sum_{i=1}^m X_{\pi_n(t)_i}^2 - X_{\pi_{n-1}(t)_i}^2 \right) \right| \\
 & \leq \mathbb{E} \left(\left(\frac{1}{m} \sum_{i=1}^m \left(X_{\pi_n(t)_i} - X_{\pi_{n-1}(t)_i} \right)^2 \right)^{1/2} \left(\frac{1}{m} \sum_{i=1}^m \left(X_{\pi_n(t)_i} + X_{\pi_{n-1}(t)_i} \right)^2 \right)^{1/2} \right) \\
 & \leq (\mathbb{E} \|X_{\pi_n(t)} - X_{\pi_{n-1}(t)}\|_{L^2(\mu_m)}^2)^{1/2} \left(\frac{1}{m} \sum_{i=1}^m \mathbb{E} \left(X_{\pi_n(t)_i} + X_{\pi_{n-1}(t)_i} \right)^2 \right)^{1/2}.
 \end{aligned}$$

By (5.8) we can apply Lemma A.5 (with $p = 2$ and $\alpha = 2$) to the first term on the right hand side to obtain

$$\left| \mathbb{E} \left(\frac{1}{m} \sum_{i=1}^m X_{\pi_n(t)_i}^2 - X_{\pi_{n-1}(t)_i}^2 \right) \right| \leq 3\sqrt{2} \bar{\Delta}_{\psi_2} 2^{n/2} \frac{5}{\sqrt{m}} d_{\psi_2}(\pi_n(t), \pi_{n-1}(t)).$$

We now apply these estimates in (5.9) and find (for $u \geq 2$)

$$\begin{aligned}
 & \left| \sum_{n \in I_{\text{subex}}} A_{\pi_n(t)} - A_{\pi_{n-1}(t)} \right| \\
 & \leq \left(2\sqrt{u} \sup_{t \in T} |A_t|^{1/2} + (3\sqrt{2} + 2\sqrt{u}) \bar{\Delta}_{\psi_2} \right) \frac{5}{\sqrt{m}} \sum_{n \in I_{\text{subex}}} 2^{n/2} d_{\psi_2}(\pi_n(t), \pi_{n-1}(t)) \\
 & \leq \left(2 \sup_{t \in T} |A_t|^{1/2} + 5 \bar{\Delta}_{\psi_2} \right) \sqrt{u} \frac{5(1 + \sqrt{2})}{\sqrt{m}} \gamma_{2,p}(T, d_{\psi_2}).
 \end{aligned}$$

In conclusion, if $\Omega_{u,p}$ occurs then we find using (5.7)

$$\begin{aligned}
 \sup_{t \in T} |A_t| & \leq \sqrt{u} \frac{10 + 10\sqrt{2}}{\sqrt{m}} \gamma_{2,p}(T, d_{\psi_2}) \sup_{t \in T} |A_t|^{1/2} \\
 & \quad + u \frac{(35 + 35\sqrt{2}) \bar{\Delta}_{\psi_2}}{\sqrt{m}} \gamma_{2,p}(T, d_{\psi_2}) + \sup_{t \in T} |A_{\pi_l(t)}|,
 \end{aligned}$$

which is a quadratic inequality in $\sup_{t \in T} |A_t|^{1/2}$. By solving this inequality, we obtain

$$\sup_{t \in T} |A_t|^{1/2} \leq \sqrt{u} \frac{25}{\sqrt{m}} \gamma_{2,p}(T, d_{\psi_2}) + \left(u \frac{85 \bar{\Delta}_{\psi_2}}{\sqrt{m}} \gamma_{2,p}(T, d_{\psi_2}) + \sup_{t \in T} |A_{\pi_l(t)}| \right)^{1/2},$$

which implies that

$$\sup_{t \in T} |A_t|^{1/2} - \sup_{t \in T} |A_{\pi_l(t)}|^{1/2} \leq \sqrt{u} \frac{25}{\sqrt{m}} \gamma_{2,p}(T, d_{\psi_2}) + \sqrt{u} \left(\frac{85 \bar{\Delta}_{\psi_2}}{\sqrt{m}} \gamma_{2,p}(T, d_{\psi_2}) \right)^{1/2}$$

In conclusion, if $u \geq 2$, then

$$\mathbb{P} \left(\sup_{t \in T} |A_t|^{1/2} - \sup_{t \in T} |A_{\pi_l(t)}|^{1/2} \geq \sqrt{u} \left(\frac{25}{\sqrt{m}} \gamma_{2,p}(T, d_{\psi_2}) + \left(\frac{85 \bar{\Delta}_{\psi_2}}{\sqrt{m}} \gamma_{2,p}(T, d_{\psi_2}) \right)^{1/2} \right) \right)$$

$$\leq c \exp(-pu/4).$$

Since the random variable

$$\sup_{t \in T} |A_t|^{1/2} - \sup_{t \in T} |A_{\pi_l(t)}|^{1/2}$$

is clearly positive, we can now apply Lemma A.5 (with $\alpha = 2$) to obtain

$$\left(\mathbb{E} \left(\sup_{t \in T} |A_t|^{1/2} - \sup_{t \in T} |A_{\pi_l(t)}|^{1/2} \right)^p \right)^{\frac{1}{p}} \leq C \left(\frac{25}{\sqrt{m}} \gamma_{2,p}(T, d_{\psi_2}) + \left(\frac{85 \bar{\Delta}_{\psi_2}}{\sqrt{m}} \gamma_{2,p}(T, d_{\psi_2}) \right)^{1/2} \right).$$

We use the triangle inequality to write

$$\begin{aligned} \left(\mathbb{E} \sup_{t \in T} |A_t|^{p/2} \right)^{1/p} &\leq C \frac{25}{\sqrt{m}} \gamma_{2,p}(T, d_{\psi_2}) + C \left(\frac{85 \bar{\Delta}_{\psi_2}}{\sqrt{m}} \gamma_{2,p}(T, d_{\psi_2}) \right)^{1/2} \\ &\quad + \left(\mathbb{E} \sup_{t \in T} |A_{\pi_l(t)}|^{p/2} \right)^{1/p}. \end{aligned} \tag{5.10}$$

Finally, we use Lemma A.3, Bernstein's inequality (5.1) and our assumption (5.6) to obtain

$$\left(\mathbb{E} \sup_{t \in T} |A_{\pi_l(t)}|^{p/2} \right)^{2/p} \leq 4 \sup_{t \in T} \left(\mathbb{E} |A_t|^{p/2} \right)^{2/p} \lesssim \sqrt{p} \frac{\sigma}{\sqrt{m}} + p \frac{K}{m}.$$

Taking squares on both sides in (5.10) and applying the latter estimate yields the result. \square

Let us now compare Theorem 5.5 with [19, Corollary 1.9]. We consider the following situation. Let X_1, \dots, X_m be independent copies of a random variable $X : \Omega \rightarrow \Theta$, where Θ is a measurable space. Let μ_X denote the probability distribution of X . Suppose that \mathcal{F} is a set of real-valued measurable functions on Θ and consider the process $(Z(f))_{f \in \mathcal{F}}$ defined by

$$Z(f) = \left| \frac{1}{m} \sum_{i=1}^m (f^2(X_i) - \mathbb{E} f^2(X_i)) \right|.$$

Theorem 5.6. [19] *There exist absolute constants $C, c > 0$ such that the following holds. If $\|f\|_{L^2(\mu_X)} = 1$ for all $f \in \mathcal{F}$, then with probability at least*

$$1 - \exp \left(-c \min \left(m, \gamma_2^2(\mathcal{F}, d_{\psi_2}) / \bar{\Delta}_{\psi_2}^2(\mathcal{F}) \right) \right)$$

we have

$$\sup_{f \in \mathcal{F}} Z(f) \leq C \bar{\Delta}_{\psi_2}(\mathcal{F}) \left(\frac{1}{m} \gamma_2^2(\mathcal{F}, d_{\psi_2}) + \frac{1}{\sqrt{m}} \gamma_2(\mathcal{F}, d_{\psi_2}) \right).$$

Moreover, if \mathcal{F} is symmetric, then

$$\mathbb{E} \sup_{f \in \mathcal{F}} Z(f) \leq C \bar{\Delta}_{\psi_2}(\mathcal{F}) \left(\frac{1}{m} \gamma_2^2(\mathcal{F}, d_{\psi_2}) + \frac{1}{\sqrt{m}} \gamma_2(\mathcal{F}, d_{\psi_2}) \right). \tag{5.11}$$

Theorem 5.5 improves this result in several respects: we can assume the X_i to be only independent instead of i.i.d., we do not need to assume that \mathcal{F} lies on the $L^2(\mu_X)$ -sphere and, most importantly, we get an optimal deviation inequality.

Corollary 5.7. *There exist constants $c, C > 0$ such that the following holds. Let $X_i : \Omega \rightarrow \Theta$, $1 \leq i \leq m$ be independent random variables and let \mathcal{F} be a set of real-valued measurable functions on Θ . Suppose that σ, K are such that*

$$\sup_{f \in \mathcal{F}} \frac{1}{m} \sum_{i=1}^m \mathbb{E} |f^2(X_i) - \mathbb{E} f^2(X_i)|^q \leq \frac{q!}{2} \sigma^2 K^{q-2} \quad (q = 2, 3, \dots).$$

Then, for any $u \geq 1$,

$$\mathbb{P}\left(\sup_{f \in \mathcal{F}} Z(f) \geq C\left(\frac{1}{m}\gamma_2^2(\mathcal{F}, d_{\psi_2}) + \frac{\bar{\Delta}_{\psi_2}(\mathcal{F})}{\sqrt{m}}\gamma_2(\mathcal{F}, d_{\psi_2})\right) + c\left(\sqrt{u}\frac{\sigma}{\sqrt{m}} + u\frac{K}{m}\right)\right) \leq e^{-u}.$$

Corollary 5.7 readily implies a deviation inequality for the sample covariance operator (with respect to the true covariance operator) of a set of i.i.d. subgaussian vectors in a Banach space, see [12, Theorem 9] for details. In [7], Corollary 5.7 is used to give a unified approach to dimensionality reduction with subgaussian matrices (see in particular [7, Theorem 4.8]). In both applications, the set \mathcal{F} consists of linear functionals.

Remark 5.8. The main point of Corollary 5.7 is to establish an improved tail bound for $Z(f)$. We note, however, that substantial improvements have been made to the bound for the expected value in (5.11) [18, 20]. In particular, a special case of [20, Theorem 5.23] (see also [27, Theorem 9.3.1]) shows that for any $q > 4$ there is a constant $C_q > 0$ such that

$$\mathbb{E} \sup_{f \in \mathcal{F}} Z(f) \leq C_q \left(\frac{1}{m} \gamma_2^2(\mathcal{F}, d_{\psi_2}) + \frac{\bar{\Delta}_{L_q(\mu_X)}(\mathcal{F})}{\sqrt{m}} \gamma_2(\mathcal{F}, d_{\psi_2}) \right).$$

We also remark that a few months after this paper was finished, a different proof of Corollary 5.7 was obtained in [3, Theorem 1].

6 Supremum of a second order chaos process

In this section we will make use of the Schatten spaces. For any $m \times n$ matrix A with complex entries A_{ij} we use

$$\|A\|_{S^q} = (\text{Tr}(A^*A)^{q/2})^{1/q} \quad (1 \leq q < \infty), \quad \|A\|_{S^\infty} = \|A\|_{l_n^2 \rightarrow l_m^2}$$

to denote the Schatten norms of A . We use

$$d_q(A_1, A_2) = \|A_1 - A_2\|_{S^q}$$

to denote the associated metrics on the $m \times n$ matrices. Accordingly, for any set \mathcal{A} of $m \times n$ matrices and $1 \leq q \leq \infty$ we define the radius

$$\bar{\Delta}_q(\mathcal{A}) = \sup_{A \in \mathcal{A}} \|A\|_{S^q}.$$

Let ξ be an n -dimensional random vector. For any $n \times n$ matrix B we define the associated *second order chaos* by

$$C_B(\xi) = \xi^* B \xi - \mathbb{E}(\xi^* B \xi) = \sum_{i,j=1}^n B_{ij}(\xi_i \bar{\xi}_j - \mathbb{E}(\xi_i \bar{\xi}_j)).$$

The tail behavior of $C_B(\xi)$ in the case that ξ has subgaussian components was described by Hanson and Wright [10], see also [23] for a modern proof.

Theorem 6.1. *Suppose that ξ_1, \dots, ξ_n are independent, mean-zero, real-valued random variables and $\max_i \|\xi_i\|_{\psi_2} \leq 1$. Then, there is a universal constant $c > 0$ such that for any $u \geq 0$,*

$$\mathbb{P}(|C_B(\xi)| \geq u) \leq 2 \exp\left(-c \min\left(\frac{u^2}{\|B\|_{S^2}^2}, \frac{u}{\|B\|_{S^\infty}}\right)\right). \quad (6.1)$$

In the terminology of Section 3, (6.1) implies that the process $(C_B(\xi))_{B \in \mathcal{B}}$ has a mixed tail with respect to the pair (d_∞, d_2) . Thus, by Theorem 3.5

$$\left(\mathbb{E} \sup_{B \in \mathcal{B}} |C_B(\xi)|^p\right)^{1/p} \lesssim \gamma_1(\mathcal{B}, d_\infty) + \gamma_2(\mathcal{B}, d_2) + \sqrt{p} \bar{\Delta}_2(\mathcal{B}) + p \bar{\Delta}_\infty(\mathcal{B}).$$

As it turns out, the occurrence of the γ_1 -functional in this bound can lead to suboptimal results in certain applications. To mend this, Krahmer, Mendelson and Rauhut proved the following deviation inequality for chaos processes of a special form, which involves only γ_2 -functionals.

Theorem 6.2. [13, Theorem 3.5] *Let \mathcal{A} be a set of $m \times n$ matrices. Set $\xi = (\xi_1, \dots, \xi_n)$, where ξ_1, \dots, ξ_n are independent, mean-zero, unit variance, real-valued, subgaussian random variables. Define*

$$\begin{aligned} E &= \gamma_2^2(\mathcal{A}, d_\infty) + \bar{\Delta}_2(\mathcal{A})\gamma_2(\mathcal{A}, d_\infty) \\ V &= \bar{\Delta}_\infty(\mathcal{A})(\bar{\Delta}_2(\mathcal{A}) + \gamma_2(\mathcal{A}, d_\infty)) \\ U &= \bar{\Delta}_\infty^2(\mathcal{A}) \end{aligned}$$

Then, there exist constants $c_1, c_2 > 0$ depending only on $\|\xi_1\|_{\psi_2}, \dots, \|\xi_n\|_{\psi_2}$ such that for all $u \geq 0$,

$$\mathbb{P}\left(\sup_{A \in \mathcal{A}} \left| \|A\xi\|_2^2 - \mathbb{E}\|A\xi\|_2^2 \right| \geq c_1 E + u\right) \leq 2 \exp\left(-c_2 \min\left\{\frac{u^2}{V}, \frac{u}{U}\right\}\right).$$

As discussed in [13], this result has interesting applications in compressed sensing with structured random matrices.

Note that due to the appearance of the γ_2 -functional in the factor V , the bound does not exhibit the correct tail behavior for large u . In fact, one would expect from Lemma A.3 and the Hanson-Wright bound that V can be replaced by the smaller, and in general optimal, factor $\bar{\Delta}_4^2(\mathcal{A})$. In Theorem 6.5 we show that this is indeed possible. Our proof follows in general lines the proof of [13], with some simplifications. For example, we completely avoid the use of the majorizing measures theorem.

The main chaining argument in the proof is contained in the following lemma. We follow the proof of [13, Lemma 3.2].

Lemma 6.3. *Fix $1 \leq p < \infty$. Let $\xi = (\xi_1, \dots, \xi_n)$ be a random vector of independent, mean-zero random variables with $\max_i \|\xi_i\|_{\psi_2} \leq 1$ and let ξ' be an independent copy of ξ defined on a probability space $(\Omega', \mathcal{F}', \mathbb{P}')$. Set $l = \lfloor \log_2(p) \rfloor$. Let \mathcal{A} be a collection of matrices, let $(\mathcal{A}_n)_{n \geq 0}$ be an optimal admissible sequence for $\gamma_{2,p}(\mathcal{A}, d_\infty)$ and define an associated sequence of maps $\pi_n : \mathcal{A} \rightarrow \mathcal{A}_n$ by $\pi_n(A) = \operatorname{argmin}_{B \in \mathcal{A}_n} d_\infty(A, B)$. Then,*

$$\left(\mathbb{E} \sup_{A \in \mathcal{A}} \left| \xi^*(A^*A - \pi_l(A)^* \pi_l(A)) \xi' \right|^p \right)^{1/p} \leq C \gamma_{2,p}(\mathcal{A}, d_\infty) \left(\mathbb{E} \sup_{A \in \mathcal{A}} \|A\xi\|_2^p\right)^{1/p}.$$

Proof. We make the decomposition

$$\begin{aligned} &\xi^*(A^*A - \pi_l(A)^* \pi_l(A)) \xi' \\ &= \sum_{n>l} \xi^* \pi_n(A)^* \pi_n(A) \xi' - \xi^* \pi_{n-1}(A)^* \pi_{n-1}(A) \xi' \\ &= \sum_{n>l} \xi^* (\pi_n(A) - \pi_{n-1}(A))^* \pi_n(A) \xi' + \sum_{n>l} \xi^* \pi_{n-1}(A)^* (\pi_n(A) - \pi_{n-1}(A)) \xi' \\ &=: S_1(A) + S_2(A). \end{aligned}$$

Let us consider $S_1(A)$. Note that the terms $\xi^* (\pi_n(A) - \pi_{n-1}(A))^* \pi_n(A) \xi'$ are subgaussian in ξ when we condition on ξ' . Thus, for any $n > l$,

$$\begin{aligned} \mathbb{P}(|\xi^* (\pi_n(A) - \pi_{n-1}(A))^* \pi_n(A) \xi'| \geq u 2^{n/2} \|(\pi_n(A) - \pi_{n-1}(A))^* \pi_n(A) \xi'\|_2) \\ \leq 2 \exp(-u^2 2^n). \end{aligned}$$

Note that for any $n > l$,

$$|\{(\pi_n(A) - \pi_{n-1}(A)), \pi_n(A); A \in \mathcal{A}\}| \leq |\mathcal{A}_n| |\mathcal{A}_{n-1}| \leq 2^{2^n} 2^{2^{n-1}} \leq 2^{2^{n+1}}.$$

Let $\Omega_{u,p}$ be the event

$$\begin{aligned} \forall n > l, \forall A \in \mathcal{A} : & |\xi^*(\pi_n(A) - \pi_{n-1}(A))^* \pi_n(A) \xi'| \\ & \leq u 2^{n/2} \|(\pi_n(A) - \pi_{n-1}(A))^* \pi_n(A) \xi'\|_2. \end{aligned}$$

By a union bound (cf. Lemma A.4),

$$\mathbb{P}(\Omega_{u,p}^c) \leq c \exp(-pu^2/4) \quad (u \geq \sqrt{2}).$$

If the event $\Omega_{u,p}$ occurs, then

$$\begin{aligned} |S_1(A)| & \leq \sum_{n>l} u 2^{n/2} \|(\pi_n(A) - \pi_{n-1}(A))^* \pi_n(A) \xi'\|_2 \\ & \leq \sum_{n>l} u 2^{n/2} \|\pi_n(A) - \pi_{n-1}(A)\|_{S^\infty} \|\pi_n(A) \xi'\|_2 \\ & \leq u(1 + \sqrt{2}) \gamma_{2,p}(\mathcal{A}, d_\infty) \sup_{A \in \mathcal{A}} \|A \xi'\|_2. \end{aligned}$$

In conclusion, for any $u \geq \sqrt{2}$,

$$\mathbb{P}\left(\sup_{A \in \mathcal{A}} |S_1(A)| > u(1 + \sqrt{2}) \gamma_{2,p}(\mathcal{A}, d_\infty) \sup_{A \in \mathcal{A}} \|A \xi'\|_2\right) \leq c \exp(-pu^2/4).$$

By Lemma A.5,

$$\left(\mathbb{E} \sup_{A \in \mathcal{A}} |S_1(A)|^p\right)^{\frac{1}{p}} \leq C \gamma_{2,p}(\mathcal{A}, d_\infty) \sup_{A \in \mathcal{A}} \|A \xi'\|_2.$$

Taking the L^p -norm over Ω' yields

$$\left(\mathbb{E}' \mathbb{E} \sup_{A \in \mathcal{A}} |S_1(A)|^p\right)^{\frac{1}{p}} \leq C \left(\mathbb{E}' \sup_{A \in \mathcal{A}} \|A \xi'\|_2^p\right)^{\frac{1}{p}} \gamma_{2,p}(\mathcal{A}, d_\infty).$$

A very similar argument gives

$$\left(\mathbb{E} \mathbb{E}' \sup_{A \in \mathcal{A}} |S_2(A)|^p\right)^{\frac{1}{p}} \leq C \left(\mathbb{E} \sup_{A \in \mathcal{A}} \|A \xi\|_2^p\right)^{\frac{1}{p}} \gamma_{2,p}(\mathcal{A}, d_\infty).$$

The asserted estimate now follows by the triangle inequality. \square

In the proof of the main theorem of this section we use the following decoupling inequality due to Arcones and Giné.

Lemma 6.4. [2] *Let g be an n -dimensional standard gaussian vector, let g' be an independent copy of g and let \mathcal{B} be a collection of self-adjoint $n \times n$ matrices. There is an absolute constant $C > 0$ such that for any $1 \leq p < \infty$,*

$$\left(\mathbb{E} \sup_{B \in \mathcal{B}} |g^* B g - \mathbb{E}(g^* B g)|^p\right)^{1/p} \leq C \left(\mathbb{E} \mathbb{E}' \sup_{B \in \mathcal{B}} |g^* B g'|^p\right)^{1/p}.$$

We will also use the following decoupling inequality, which is elementary to prove (see e.g. [8, Theorem 8.11]). Let ξ_1, \dots, ξ_n be independent, real-valued, mean-zero random variables and let ξ'_1, \dots, ξ'_n be independent copies. Then, for any $1 \leq p < \infty$,

$$\left(\mathbb{E} \sup_{B \in \mathcal{B}} \left| \sum_{i \neq j} \xi_i \xi_j B_{ij} \right|^p\right)^{1/p} \leq 4 \left(\mathbb{E} \mathbb{E}' \sup_{B \in \mathcal{B}} |\xi^* B \xi'|^p\right)^{1/p}. \quad (6.2)$$

Theorem 6.5. *Let \mathcal{A} be a set of $m \times n$ matrices. Suppose that ξ_1, \dots, ξ_n are independent, real-valued, mean-zero random variables, let $\xi = (\xi_1, \dots, \xi_n)$ and set $\|\xi\|_{\psi_2} = \max_i \|\xi_i\|_{\psi_2}$. For any $1 \leq p < \infty$,*

$$\begin{aligned} \left(\mathbb{E} \sup_{A \in \mathcal{A}} \left| \|A\xi\|_2^2 - \mathbb{E} \|A\xi\|_2^2 \right|^p \right)^{1/p} &\lesssim \|\xi\|_{\psi_2}^2 \left(\gamma_{2,p}^2(\mathcal{A}, d_\infty) + \bar{\Delta}_2(\mathcal{A}) \gamma_{2,p}(\mathcal{A}, d_\infty) \right. \\ &\quad \left. + \sqrt{p} \bar{\Delta}_4^2(\mathcal{A}) + p \bar{\Delta}_\infty^2(\mathcal{A}) \right). \end{aligned} \quad (6.3)$$

As a consequence, there are constants $c, C > 0$ such that for any $u \geq 1$,

$$\begin{aligned} \mathbb{P} \left(\sup_{A \in \mathcal{A}} \left| \|A\xi\|_2^2 - \mathbb{E} \|A\xi\|_2^2 \right| \geq C \|\xi\|_{\psi_2}^2 \left(\gamma_2^2(\mathcal{A}, d_\infty) + \bar{\Delta}_2(\mathcal{A}) \gamma_2(\mathcal{A}, d_\infty) \right) \right. \\ \left. + c \|\xi\|_{\psi_2}^2 \left(\sqrt{u} \bar{\Delta}_4^2(\mathcal{A}) + u \bar{\Delta}_\infty^2(\mathcal{A}) \right) \right) \leq e^{-u}. \end{aligned}$$

Proof. By dividing both sides of (6.3) by $\|\xi\|_{\psi_2}$ if necessary, we may assume that $\|\xi\|_{\psi_2} \leq 1$. Let $l = \lfloor \log_2(p) \rfloor$. Let $(\mathcal{A}_n)_{n \geq 0}$ and $(\pi_n)_{n \geq 0}$ be as in Lemma 6.3 and write

$$\begin{aligned} &\sup_{A \in \mathcal{A}} \xi^* A^* A \xi - \mathbb{E}(\xi^* A^* A \xi) \\ &\leq \sup_{A \in \mathcal{A}} \xi^* A^* A \xi - \xi^* \pi_l(A)^* \pi_l(A) \xi - \mathbb{E}(\xi^* A^* A \xi - \xi^* \pi_l(A)^* \pi_l(A) \xi) \\ &\quad + \sup_{A \in \mathcal{A}} \xi^* \pi_l(A)^* \pi_l(A) \xi - \mathbb{E}(\xi^* \pi_l(A)^* \pi_l(A) \xi). \end{aligned}$$

We continue by estimating the first term. We write $B = B(A, l) := A^* A - \pi_l(A)^* \pi_l(A)$ for brevity. By the triangle inequality and the decoupling inequality (6.2),

$$\begin{aligned} &\left(\mathbb{E} \sup_{A \in \mathcal{A}} |\xi^* B \xi - \mathbb{E}(\xi^* B \xi)|^p \right)^{1/p} \\ &\leq \left(\mathbb{E} \sup_{A \in \mathcal{A}} \left| \sum_{i \neq j} \xi_i \xi_j B_{ij} \right|^p \right)^{1/p} + \left(\mathbb{E} \sup_{A \in \mathcal{A}} \left| \sum_i (|\xi_i|^2 - \mathbb{E}|\xi_i|^2) B_{ii} \right|^p \right)^{1/p} \\ &\leq \left(\mathbb{E} \mathbb{E}' \sup_{A \in \mathcal{A}} |\xi^* B \xi'|^p \right)^{1/p} + \left(\mathbb{E} \sup_{A \in \mathcal{A}} \left| \sum_i (|\xi_i|^2 - \mathbb{E}|\xi_i|^2) B_{ii} \right|^p \right)^{1/p}. \end{aligned} \quad (6.4)$$

Let ε be a Rademacher vector and let g be a standard Gaussian vector. By symmetrization [17, Lemma 6.3], the contraction principle [17, Lemma 4.6] and de-symmetrization [17, Lemma 6.3],

$$\begin{aligned} &\left(\mathbb{E} \sup_{A \in \mathcal{A}} \left| \sum_i (|\xi_i|^2 - \mathbb{E}|\xi_i|^2) B_{ii} \right|^p \right)^{1/p} \\ &\leq 2 \left(\mathbb{E} \mathbb{E}_\varepsilon \sup_{A \in \mathcal{A}} \left| \sum_i \varepsilon_i |\xi_i|^2 B_{ii} \right|^p \right)^{1/p} \\ &\lesssim \left(\mathbb{E} \mathbb{E}_\varepsilon \sup_{A \in \mathcal{A}} \left| \sum_i \varepsilon_i g_i^2 B_{ii} \right|^p \right)^{1/p} \\ &\lesssim \left(\mathbb{E} \sup_{A \in \mathcal{A}} \left| \sum_i (g_i^2 - 1) B_{ii} \right|^p \right)^{1/p} + \left(\mathbb{E}_\varepsilon \sup_{A \in \mathcal{A}} \left| \sum_i \varepsilon_i B_{ii} \right|^p \right)^{1/p} \\ &\lesssim \left(\mathbb{E} \sup_{A \in \mathcal{A}} \left| g^* B g - \mathbb{E}(g^* B g) \right|^p \right)^{1/p} + \left(\mathbb{E} \sup_{A \in \mathcal{A}} \left| \sum_{i \neq j} g_i g_j B_{ij} \right|^p \right)^{1/p} \\ &\quad + \left(\mathbb{E}_\varepsilon \sup_{A \in \mathcal{A}} \left| \sum_i \varepsilon_i B_{ii} \right|^p \right)^{1/p} \end{aligned}$$

$$\lesssim \left(\mathbb{E} \mathbb{E}' \sup_{A \in \mathcal{A}} |g^* B g'|^p \right)^{1/p} + \left(\mathbb{E}_\varepsilon \sup_{A \in \mathcal{A}} \left| \sum_i \varepsilon_i B_{ii} \right|^p \right)^{1/p}, \quad (6.5)$$

where in the final step we used the decoupling inequality (6.2) and Lemma 6.4. We first estimate the second term on the far right hand side. By Khintchine's inequality, for any $C, D \in \mathcal{A}$,

$$\left(\mathbb{E}_\varepsilon \left| \sum_i \varepsilon_i (C^* C - D^* D)_{ii} \right|^p \right)^{1/p} \leq \sqrt{p} \left(\sum_i |(C^* C)_{ii} - (D^* D)_{ii}|^2 \right)^{1/2}.$$

Let $C_{(i)}$ denote the i -th column of C . Then we can estimate

$$\begin{aligned} & \left(\sum_i |(C^* C)_{ii} - (D^* D)_{ii}|^2 \right)^{1/2} \\ &= \left(\sum_i \left(\|C_{(i)}\|_2^2 - \|D_{(i)}\|_2^2 \right)^2 \right)^{1/2} \\ &= \left(\sum_i \left(\|C_{(i)}\|_2 - \|D_{(i)}\|_2 \right)^2 \left(\|C_{(i)}\|_2 + \|D_{(i)}\|_2 \right)^2 \right)^{1/2} \\ &\leq \left(\sum_i \|C_{(i)} - D_{(i)}\|_2^2 \left(\|C_{(i)}\|_2 + \|D_{(i)}\|_2 \right)^2 \right)^{1/2}. \end{aligned}$$

Moreover, for any fixed i ,

$$\begin{aligned} \|C_{(i)} - D_{(i)}\|_2^2 &= \|(C - D)_{(i)}\|_2^2 \\ &= ((C - D)^*(C - D))_{ii} \leq \|(C - D)^*(C - D)\|_{S^\infty} = \|C - D\|_{S^\infty}^2. \end{aligned}$$

In conclusion, we find

$$\begin{aligned} & \left(\mathbb{E}_\varepsilon \left| \sum_i \varepsilon_i (C^* C - D^* D)_{ii} \right|^p \right)^{1/p} \\ &\leq 2\sqrt{p} d_\infty(C, D) \sup_{A \in \mathcal{A}} \left(\sum_i \|A_{(i)}\|_2^2 \right)^{1/2} = 2\sqrt{p} \bar{\Delta}_2(\mathcal{A}) d_\infty(C, D). \end{aligned}$$

Thus, by Lemma A.1 the process

$$\left(\sum_i \varepsilon_i (A^* A)_{ii} \right)_{A \in \mathcal{A}}$$

is subgaussian with respect to the metric $\bar{\Delta}_2(\mathcal{A}) d_\infty$ and Theorem 3.2 immediately yields

$$\left(\mathbb{E}_\varepsilon \sup_{A \in \mathcal{A}} \left| \sum_i \varepsilon_i B_{ii} \right|^p \right)^{1/p} \lesssim \bar{\Delta}_2(\mathcal{A}) \gamma_{2,p}(\mathcal{A}, d_\infty).$$

By Lemma 6.3 and the (quasi-)triangle inequality in $L^{p/2}$,

$$\begin{aligned} & \left(\mathbb{E} \mathbb{E}' \sup_{A \in \mathcal{A}} |\xi^* B \xi'|^p \right)^{1/p} \\ &\lesssim \gamma_{2,p}(\mathcal{A}, d_\infty) \left(\mathbb{E} \sup_{A \in \mathcal{A}} \|A \xi\|_2^p \right)^{1/p} \\ &\lesssim \gamma_{2,p}(\mathcal{A}, d_\infty) \left(\mathbb{E} \left| \sup_{A \in \mathcal{A}} \|A \xi\|_2^2 - \mathbb{E} \|A \xi\|_2^2 \right| + \sup_{A \in \mathcal{A}} \mathbb{E} \|A \xi\|_2^2 \right)^{1/p} \\ &\lesssim \gamma_{2,p}(\mathcal{A}, d_\infty) \left(\left(\mathbb{E} \sup_{A \in \mathcal{A}} \left| \|A \xi\|_2^2 - \mathbb{E} \|A \xi\|_2^2 \right|^p \right)^{1/2p} + \bar{\Delta}_2(\mathcal{A}) \right). \end{aligned}$$

This bounds the first term on the right hand side of (6.4) and, specialized to $\xi = g$, the first term on the far right hand side of (6.5). Finally, by Lemma A.3, the Hanson-Wright bound (6.1) and Lemma A.2,

$$\begin{aligned} \left(\mathbb{E} \sup_{A \in \mathcal{A}} \left| \|\pi_l(A)\xi\|_2^2 - \mathbb{E} \|\pi_l(A)\xi\|_2^2 \right|^p\right)^{1/p} &\leq 2 \sup_{A \in \mathcal{A}} \left(\mathbb{E} \left| \|A\xi\|_2^2 - \mathbb{E} \|A\xi\|_2^2 \right|^p\right)^{1/p} \\ &\lesssim \sqrt{p} \bar{\Delta}_4^2(\mathcal{A}) + p \bar{\Delta}_\infty^2(\mathcal{A}). \end{aligned}$$

Collecting our estimates, we find

$$\begin{aligned} &\left(\mathbb{E} \sup_{A \in \mathcal{A}} \left| \|A\xi\|_2^2 - \mathbb{E} \|A\xi\|_2^2 \right|^p\right)^{1/p} \\ &\lesssim \gamma_{2,p}(\mathcal{A}, d_\infty) \left(\mathbb{E} \sup_{A \in \mathcal{A}} \left| \|A\xi\|_2^2 - \mathbb{E} \|A\xi\|_2^2 \right|^p\right)^{1/2p} \\ &\quad + \gamma_{2,p}(\mathcal{A}, d_\infty) \left(\mathbb{E} \sup_{A \in \mathcal{A}} \left| \|Ag\|_2^2 - \mathbb{E} \|Ag\|_2^2 \right|^p\right)^{1/2p} \\ &\quad + \gamma_{2,p}(\mathcal{A}, d_\infty) \bar{\Delta}_2(\mathcal{A}) + \sqrt{p} \bar{\Delta}_4^2(\mathcal{A}) + p \bar{\Delta}_\infty^2(\mathcal{A}). \end{aligned}$$

Since this estimate in particular holds for $\xi = g$, we obtain a quadratic inequality

$$\max\{x, y\} \lesssim a \max\{x, y\}^{1/2} + b,$$

where

$$\begin{aligned} x &= \left(\mathbb{E} \sup_{A \in \mathcal{A}} \left| \|A\xi\|_2^2 - \mathbb{E} \|A\xi\|_2^2 \right|^p\right)^{1/p}, & y &= \left(\mathbb{E} \sup_{A \in \mathcal{A}} \left| \|Ag\|_2^2 - \mathbb{E} \|Ag\|_2^2 \right|^p\right)^{1/p}, \\ a &= \gamma_{2,p}(\mathcal{A}, d_\infty), & b &= \gamma_{2,p}(\mathcal{A}, d_\infty) \bar{\Delta}_2(\mathcal{A}) + \sqrt{p} \bar{\Delta}_4^2(\mathcal{A}) + p \bar{\Delta}_\infty^2(\mathcal{A}). \end{aligned}$$

By solving this inequality, we obtain the result. □

An application of Theorem 6.5 to dimensionality reduction with sparse Johnson-Lindenstrauss transforms can be found in [5] (see in particular section 2 there).

A Appendix

In this appendix we collect some elementary observations that are used throughout the paper. The first lemma states how to pass from moment to tail bounds. The proof is a straightforward consequence of Markov's inequality, see e.g. [8, Propositions 7.11 and 7.15].

Lemma A.1. *If X is a complex-valued random variable satisfying*

$$(\mathbb{E}|X|^p)^{1/p} \leq ap^{1/\alpha} + b, \quad \text{for all } p \geq 1,$$

for some $0 < a, \alpha < \infty$ and $b \geq 0$, then

$$\mathbb{P}(|X| \geq e^{1/\alpha}(au + b)) \leq \exp(-u^\alpha/\alpha) \quad (u \geq 1).$$

If X satisfies

$$(\mathbb{E}|X|^p)^{1/p} \leq a_1p + a_2\sqrt{p} + a_3, \quad \text{for all } p \geq 1,$$

for some $0 \leq a_1, a_2, a_3 < \infty$, then

$$\mathbb{P}(|X| \geq e(a_1u + a_2\sqrt{u} + a_3)) \leq \exp(-u) \quad (u \geq 1).$$

The following observation is a converse statement.

Lemma A.2. *Let $0 < \alpha < \infty$. If a random variable X satisfies*

$$\mathbb{P}(|X| \geq e^{1/\alpha} au) \leq be^{-u^\alpha/\alpha} \quad (u \geq 0),$$

then for any $p \geq 1$,

$$(\mathbb{E}|X|^p)^{1/p} \leq e^{1/2e} a \left(\sqrt{\frac{2\pi}{\alpha}} e^{\alpha/12} b \right)^{1/p} p^{1/\alpha}.$$

If X satisfies

$$\mathbb{P}(|X| \geq a_1 u + a_2 \sqrt{u}) \leq \exp(-u) \quad (u \geq 0) \tag{A.1}$$

for some $0 \leq a_1, a_2 < \infty$, then for all $p \geq 1$

$$(\mathbb{E}|X|^p)^{1/p} \leq a_1 2e^{1/(2e)} (\sqrt{2\pi} e^{1/(12p)})^{1/p} e^{-1} p + a_2 2(2e)^{-1/2} e^{1/(2e)} (\sqrt{\pi} e^{1/(6p)})^{1/p} \sqrt{p}.$$

Proof. For a proof of the first statement, see [8, Proposition 7.13]. For the second assertion, note that (A.1) implies

$$\mathbb{P}(\tfrac{1}{2}|X| \geq u) \leq \begin{cases} e^{-u^2/a_2^2}, & \text{if } 0 \leq u \leq a_2^2/a_1; \\ e^{-u/a_1}, & \text{if } u \geq a_2^2/a_1. \end{cases}$$

Using integration by parts and a change of variable we find

$$\begin{aligned} 2^{-p} \mathbb{E}|X|^p &= p \int_0^\infty u^{p-1} \mathbb{P}(\tfrac{1}{2}|X| \geq u) \, du \\ &\leq p \int_0^{a_2^2/a_1} u^{p-1} e^{-u^2/a_2^2} \, du + p \int_{a_2^2/a_1}^\infty u^{p-1} e^{-u/a_1} \, du \\ &= \tfrac{1}{2} p a_2^p \int_0^{a_2^2/a_1^2} v^{\frac{p}{2}-1} e^{-v} \, dv + p a_1^p \int_{a_2^2/a_1^2}^\infty v^{p-1} e^{-v} \, dv \\ &\leq \tfrac{1}{2} p a_2^p \Gamma(p/2) + p a_1^p \Gamma(p), \end{aligned}$$

where $\Gamma(p) = \int_0^\infty v^{p-1} e^{-v} \, dv$ is the gamma function. The result now readily follows using Stirling's formula, which states that

$$\Gamma(p) = \sqrt{2\pi} p^{p-1/2} e^{-p} e^{\theta(p)/12p}$$

for some $0 \leq \theta(p) \leq 1$. Indeed,

$$p\Gamma(p) \leq p^p \sqrt{2\pi} \sqrt{p} e^{-p} e^{1/(12p)}$$

and therefore

$$(p\Gamma(p))^{1/p} \leq p (\sqrt{2\pi} e^{1/(12p)})^{1/p} e^{-1} e^{1/(2e)},$$

where we used that $p^{1/(2p)} \leq e^{1/(2e)}$ if $p \geq 1$. In the same way,

$$(\tfrac{1}{2} p \Gamma(p/2))^{1/p} \leq (2e)^{-1/2} e^{1/(2e)} (\sqrt{\pi} e^{1/(6p)})^{1/p} \sqrt{p}.$$

□

The following three lemmas are used in every chaining argument in this paper.

Lemma A.3. *Fix $1 \leq p < \infty$, set $l = \lfloor \log_2(p) \rfloor$ and let $(X_t)_{t \in T}$ be a collection of complex-valued random variables. If $|T| \leq 2^{2^l}$, then*

$$\left(\mathbb{E} \sup_{t \in T} |X_t|^p \right)^{1/p} \leq 2 \sup_{t \in T} (\mathbb{E}|X_t|^p)^{1/p}.$$

Proof. Since $|T| \leq 2^p$,

$$\mathbb{E} \sup_{t \in T} |X_t|^p \leq \sum_{t \in T} \mathbb{E} |X_t|^p \leq |T| \sup_{t \in T} \mathbb{E} |X_t|^p \leq 2^p \sup_{t \in T} \mathbb{E} |X_t|^p.$$

□

Lemma A.4. Fix $1 \leq p < \infty$, $0 < \alpha < \infty$, $u \geq 2^{1/\alpha}$ and set $l = \lfloor \log_2(p) \rfloor$. For every $n > l$ let $(\Omega_i^{(n)})_{i \in I_n}$ be a collection of events satisfying

$$\mathbb{P}(\Omega_i^{(n)}) \leq 2 \exp(-2^n u^\alpha), \quad \text{for all } i \in I_n.$$

If $|I_n| \leq 2^{2^{n+1}}$, then for an absolute constant $c \leq 17$,

$$\mathbb{P}\left(\bigcup_{n>l} \bigcup_{i \in I_n} \Omega_i^{(n)}\right) \leq c \exp(-pu^\alpha/4). \tag{A.2}$$

Proof. By a union bound, using that $u^\alpha \geq 2$,

$$\begin{aligned} \mathbb{P}\left(\bigcup_{n>l} \bigcup_{i \in I_n} \Omega_i^{(n)}\right) &\leq \sum_{n>l} 2^{2^{n+1}} 2 \exp(-u^\alpha 2^n) \\ &= 2 \sum_{n>l} \exp(2(\log 2)2^n) \exp(-u^\alpha 2^n) \\ &\leq 2 \sum_{n>l} \exp((\log 2 - 1)u^\alpha 2^n). \end{aligned}$$

Clearly,

$$\begin{aligned} \sum_{n>l} \exp((\log 2 - 1)u^\alpha 2^n) &= \exp(-2^l u^\alpha / 2) \sum_{n>l} \exp((\log 2 - 1)u^\alpha 2^n + 2^l u^\alpha / 2) \\ &\leq \exp(-2^l u^\alpha / 2) \sum_{n \geq 0} \exp((\log 2 - 1)u^\alpha 2^n + 2^n u^\alpha / 4). \end{aligned}$$

Since $\log 2 - \frac{3}{4} < 0$ and $-2^l \leq -\frac{p}{2}$, we conclude that (A.2) holds. Note that

$$\begin{aligned} c &\leq 2 \sum_{n \geq 0} \exp(2^n (2(\log 2 - 1) + \frac{1}{2})) \\ &\leq 2 \sum_{n \geq 1} \exp(n(2(\log 2 - 1) + \frac{1}{2})) \leq \frac{2}{1 - \exp(2(\log 2 - 1) + \frac{1}{2})} - 2 \leq 17. \end{aligned}$$

□

Lemma A.5. Fix $1 \leq p < \infty$ and $0 < \alpha < \infty$. Let $\gamma \geq 0$ and suppose that ξ is a positive random variable such that for some $c \geq 1$ and $u_* > 0$,

$$\mathbb{P}(\xi > \gamma u) \leq c \exp(-pu^\alpha/4) \quad (u \geq u_*).$$

Then, for a constant $\tilde{c}_\alpha > 0$ depending only on α ,

$$(\mathbb{E}\xi^p)^{1/p} \leq \gamma(\tilde{c}_\alpha c + u_*).$$

Proof. By integration by parts and a change of variable,

$$\mathbb{E}\xi^p = \int_0^\infty pu^{p-1} \mathbb{P}(\xi > u) du$$

$$\begin{aligned}
&= \gamma^p \int_0^\infty p v^{p-1} \mathbb{P}(\xi > v\gamma) dv \\
&\leq \gamma^p \left(\int_{u_*}^\infty p v^{p-1} c \exp(-pv^\alpha/4) dv + \int_0^{u_*} p v^{p-1} dv \right) \\
&= \gamma^p \left(c \int_{u_*}^\infty p v^{p-1} \exp(-pv^\alpha/4) dv + u_*^p \right).
\end{aligned}$$

To complete the proof, observe that by another change of variable

$$\begin{aligned}
\int_0^\infty p v^{p-1} e^{-pv^\alpha/4} dv &= p^{-p/\alpha} 2^{p/\alpha} \frac{2p}{\alpha} \int_0^\infty u^{\frac{2p}{\alpha}-1} e^{-u^2/2} du \\
&= p^{-p/\alpha} 2^{p/\alpha} \frac{2p}{\alpha} \frac{\sqrt{2\pi}}{2} \mathbb{E}|g|^{\frac{2p}{\alpha}-1},
\end{aligned}$$

where g is a standard Gaussian. If $p \geq \alpha$ then

$$\mathbb{E}|g|^{\frac{2p}{\alpha}-1} \leq \left(\frac{2p}{\alpha} - 1 \right)^{\frac{p}{\alpha}-\frac{1}{2}},$$

and therefore

$$\int_0^\infty p v^{p-1} e^{-pv^\alpha/4} dv \leq \frac{\sqrt{2\pi}}{2} 2^{p/\alpha} \left(\frac{2}{\alpha} \right)^{\frac{p}{\alpha}+\frac{1}{2}} p^{1/2}.$$

On the other hand, if $p \leq \alpha$, then we can trivially estimate the integral by $2\sqrt{2\pi}$. The result follows by combining all of the above estimates. \square

References

- [1] R. Adamczak, *A tail inequality for suprema of unbounded empirical processes with applications to Markov chains*, Electron. J. Probab. **13** (2008), no. 34, 1000–1034. MR-2424985
- [2] M. Arcones and E. Giné, *On decoupling, series expansions, and tail behavior of chaos processes*, J. Theoret. Probab. **6** (1993), no. 1, 101–122. MR-1201060
- [3] W. Bednorz, *Concentration via chaining method and its applications*, ArXiv:1405.0676.
- [4] S. Boucheron, G. Lugosi, and P. Massart, *Concentration inequalities*, Oxford University Press, Oxford, 2013, A nonasymptotic theory of independence, With a foreword by Michel Ledoux. MR-3185193
- [5] J. Bourgain, S. Dirksen, and J. Nelson, *Toward a unified theory of sparse dimensionality reduction in Euclidean space*, arXiv:1311.2542. To appear in Geom. Funct. Anal. Extended abstract in STOC '15.
- [6] E. Candès and T. Tao, *Near-optimal signal recovery from random projections: universal encoding strategies?*, IEEE Trans. Inform. Theory **52** (2006), no. 12, 5406–5425. MR-2300700
- [7] S. Dirksen, *Dimensionality reduction with subgaussian matrices: a unified theory*, arXiv:1402.3973.
- [8] S. Foucart and H. Rauhut, *A mathematical introduction to compressive sensing*, Birkhäuser/Springer, New York, 2013. MR-3100033
- [9] S. van de Geer and J. Lederer, *The Bernstein–Orlicz norm and deviation inequalities*, Probab. Theory Related Fields **157** (2013), no. 1-2, 225–250. MR-3101846
- [10] D. Hanson and F. Wright, *A bound on tail probabilities for quadratic forms in independent random variables*, Ann. Math. Statist. **42** (1971), 1079–1083. MR-0279864

- [11] V. Koltchinskii, *Oracle inequalities in empirical risk minimization and sparse recovery problems*, Lecture Notes in Mathematics, vol. 2033, Springer, Heidelberg, 2011. MR-2829871
- [12] V. Koltchinskii and K. Lounici, *Concentration inequalities and moment bounds for sample covariance operators*, arXiv:1405.2468.
- [13] F. Krahmer, S. Mendelson, and H. Rauhut, *Suprema of chaos processes and the restricted isometry property*, Comm. Pure Appl. Math. **67** (2014), no. 11, 1877–1904. MR-3263672
- [14] M. Krasnosel'skiĭ and Ja. Rutickiĭ, *Convex functions and Orlicz spaces*, P. Noordhoff Ltd., Groningen, 1961. MR-0126722
- [15] R. Latała, *Weak and strong moments of random vectors*, Marcinkiewicz centenary volume, Banach Center Publ., vol. 95, Polish Acad. Sci. Inst. Math., Warsaw, 2011, pp. 115–121. MR-2918093
- [16] M. Ledoux, *The concentration of measure phenomenon*, American Mathematical Society, Providence, RI, 2001. MR-1849347
- [17] M. Ledoux and M. Talagrand, *Probability in Banach spaces*, Springer-Verlag, Berlin, 1991. MR-1102015
- [18] S. Mendelson, *Empirical processes with a bounded ψ_1 diameter*, Geom. Funct. Anal. **20** (2010), no. 4, 988–1027. MR-2729283
- [19] S. Mendelson, A. Pajor, and N. Tomczak-Jaegermann, *Reconstruction and subgaussian operators in asymptotic geometric analysis*, Geom. Funct. Anal. **17** (2007), no. 4, 1248–1282. MR-2373017
- [20] S. Mendelson and G. Paouris, *On generic chaining and the smallest singular value of random matrices with heavy tails*, J. Funct. Anal. **262** (2012), no. 9, 3775–3811. MR-2899978
- [21] H. Rauhut, *Compressive sensing and structured random matrices*, Theoretical foundations and numerical methods for sparse recovery, Radon Ser. Comput. Appl. Math., vol. 9, Walter de Gruyter, Berlin, 2010, pp. 1–92. MR-2731597
- [22] M. Rudelson and R. Vershynin, *On sparse reconstruction from Fourier and Gaussian measurements*, Comm. Pure Appl. Math. **61** (2008), no. 8, 1025–1045. MR-2417886
- [23] M. Rudelson and R. Vershynin, *Hanson-Wright inequality and sub-Gaussian concentration*, Electron. Commun. Probab. **18** (2013), no. 82, 9. MR-3125258
- [24] M. Talagrand, *Regularity of Gaussian processes*, Acta Math. **159** (1987), no. 1-2, 99–149. MR-906527
- [25] M. Talagrand, *Majorizing measures without measures*, Ann. Probab. **29** (2001), no. 1, 411–417. MR-1825156
- [26] M. Talagrand, *The generic chaining*, Springer-Verlag, Berlin, 2005. MR-2133757
- [27] M. Talagrand, *Upper and lower bounds for stochastic processes*, Springer, Heidelberg, 2014. MR-3184689
- [28] A. van der Vaart and J. Wellner, *Weak convergence and empirical processes*, Springer-Verlag, New York, 1996. MR-1385671
- [29] F. Viens and A. Vizcarra, *Supremum concentration inequality and modulus of continuity for sub- n th chaos processes*, J. Funct. Anal. **248** (2007), no. 1, 1–26. MR-2329681

Acknowledgments. It is a pleasure to thank Holger Rauhut for valuable comments. I thank the reviewers for their helpful comments.

Electronic Journal of Probability

Electronic Communications in Probability

Advantages of publishing in EJP-ECP

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

Economical model of EJP-ECP

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

Help keep the journal free and vigorous

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

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

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

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

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

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

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